(19) Евразийское патентное ведомство

(21) 202192548 (13) A1

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ

(43) Дата публикации заявки
2022.08.25

(22) Дата подачи заявки
2020.03.19

(51) Int. Cl. ***C12N 15/11*** (2006.01)
C12N 9/12 (2006.01)
C12N 9/22 (2006.01)
C12N 15/10 (2006.01)

(54) СПОСОБЫ И КОМПОЗИЦИИ ДЛЯ РЕДАКТИРОВАНИЯ НУКЛЕОТИДНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

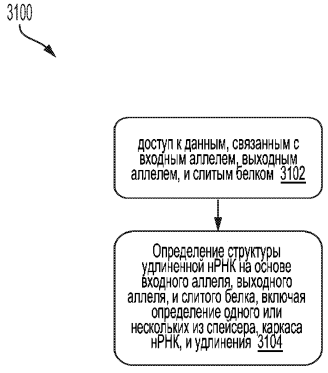
(31) **62/820,813; 62/858,958; 62/889,996; 62/922,654; 62/913,553; 62/973,558; 62/931,195; 62/974,537; 62/944,231; 62/991,069; 63/100,548**
(32) **2019.03.19; 2019.06.07; 2019.08.21; 2019.08.21; 2019.10.10; 2019.10.10; 2019.11.05; 2019.12.05; 2019.12.05; 2020.03.17; 2020.03.17**
(33) **US**
(86) **PCT/US2020/023553**
(87) **WO 2020/191153 2020.09.24**
(88) **2020.12.10**

(71) Заявитель:
ЗЕ БРОУД ИНСТИТЬЮТ, ИНК.; ПРЕЗИДЕНТ ЭНД ФЕЛЛОУЗ ОФ ГАРВАРД КОЛЛЕДЖ; МАССАЧУСЕТС ИНСТИТЬЮТ ОФ ТЕКНОЛОДЖИ (US)

(72) Изобретатель:
Лиу Дэвид Р., Анзалоун Эндрю Вито, Шен Макс Уолт (US)

(74) Представитель:
Гизатуллина Е.М., Угрюмов В.М., Христофоров А.А., Строкова О.В., Гизатуллин Ш.Ф., Костюшенкова М.Ю., Парамонова К.В., Прищепный С.В., Джермакян Р.В. (RU)

(57) Описаны новые направляющие РНК праймерного редактора для праймерного редактирования, конструкты для праймерного редактирования, и способы их применения. Кроме того, описаны композиции и способы осуществления праймерного редактирования молекулы-мишени ДНК (например, генома), которое обеспечивает встраивание нуклеотидного изменения и/или осуществление нацеленного мутагенеза (например, вставки или делеции). Нуклеотидное изменение может включать изменение одного нуклеотида (например, любую транзицию или трансверсию), вставку одного или нескольких нуклеотидов или делецию одного или нескольких нуклеотидов. Более конкретно, описаны слитые белки, включающие программируемые нуклеиновой кислотой ДНК-связывающие белки ("napDNAbp") и полимеразу (например, обратную транскриптазу), которые направляются к определенной последовательности ДНК с помощью РНК праймерного редактора (PEgRNA). Направляющая РНК праймерного редактора включает плечо удлинения, которое включает последовательность матрицы для синтеза ДНК, которая кодирует одноцепочечный ДНК-флэп, который является гомологичным последовательности эндогенной ДНК, но который содержит требуемое одно или несколько нуклеотидных изменений и который после синтеза на основе полимеразы (например, обратной транскриптазы) становится встроенным в молекулу-мишень ДНК.



202192548 A1

СПОСОБЫ И КОМПОЗИЦИИ ДЛЯ РЕДАКТИРОВАНИЯ НУКЛЕОТИДНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

ГОСУДАРСТВЕННАЯ ПОДДЕРЖКА

[0001] Данное изобретение было осуществлено при государственной поддержке в соответствии с грантами U01AI142756, RM1HG009490, R01EB022376, и R35GM118062, выданными Национальными институтами здравоохранения США. Государство обладает определенными правами на данное изобретение.

РОДСТВЕННЫЕ ЗАЯВКИ И ВКЛЮЧЕНИЕ В КАЧЕСТВЕ ССЫЛКИ

[0002] Данная предварительная заявка на патент США ссылается и приводит в качестве ссылки следующие заявки, а именно: предварительную заявку на патент США 62/820,813, поданную 19 марта 2019 г. (внутренний номер дела поверенного B1195.70074US00), предварительную заявку на патент США 62/858,958 (внутренний номер дела поверенного B1195.70074US01), поданную 7 июня 2019 г., предварительную заявку на патент США 62/889,996 (внутренний номер дела поверенного B1195.70074US02), поданную 21 августа 2019 г., предварительную заявку на патент США 62/922,654, поданную 21 августа 2019 г. (внутренний номер дела поверенного B1195.70083US00), предварительную заявку на патент США 62/913,553 (внутренний номер дела поверенного B1195.70074US03), поданную 10 октября 2019 г., предварительную заявку на патент США 62/973,558 (внутренний номер дела поверенного B1195.70083US01), поданную 10 октября 2019 г., предварительную заявку на патент США 62/931,195 (внутренний номер дела поверенного B1195.70074US04), поданную 5 ноября 2019 г., предварительную заявку на патент США 62/944,231 (внутренний номер дела поверенного B1195.70074US05), поданную 5 декабря 2019 г., предварительную заявку на патент США 62/974,537 (внутренний номер дела поверенного B1195.70083US02), поданную 5 декабря 2019 г., предварительную заявку на патент США 62/991,069 (внутренний номер дела поверенного B1195.70074US06), поданную 17 марта 2020 г., и предварительную заявку на патент США (номер не доступен на момент подачи данной заявки) (внутренний номер дела поверенного B1195.70083US03), поданную 17 марта 2020 г.

ВКЛЮЧЕНИЕ ПЕРЕЧНЯ ПОСЛЕДОВАТЕЛЬНОСТЕЙ В КАЧЕСТВЕ ССЫЛКИ

[0003] В соответствии с 37 CFR § 1.52(e) данное Описание включает Перечень последовательностей, поданный одновременно с ним на компакт-диске (2 копии). В соответствии с требованиями 37 CFR § 1.52(e)(5) Заявитель явным образом включает в качестве ссылки всю информацию и материалы, расположенные на компакт-диске в файле с названием "B119570083WO00-SEQ.txt", который был создан 19 марта 2020 года и имеет размер 371.109 MB. На основе этого Перечень последовательностей представляет собой часть настоящего Описания. Компакт-диск не содержит других файлов.

ПРЕДПОСЫЛКИ ИЗОБРЕТЕНИЯ

[0004] Патогенные мутации одиночных нуклеотидов связаны приблизительно с 67% заболеваний человека, при которых присутствует генетический компонент[7]. К сожалению, варианты лечения для пациентов с данными генетическими заболеваниями остаются крайне ограниченными, несмотря на десятилетия исследований в области генной терапии[8]. Вероятно одним из наиболее прямых решений данной терапевтической проблемы является прямое корректирование мутаций одиночных нуклеотидов в геномах пациентов, которое будет направлено на первопричину заболевания и вероятно обеспечит долговременную пользу. Несмотря на то, что подобная стратегия ранее являлась немыслимой, недавние достижения в плане возможностей редактирования генома, принесенные разработкой системы CRISPR/Cas[9], сделали данный терапевтический подход возможным. Путем прямого конструирования последовательности направляющей РНК (нРНК), которая содержит ~20 нуклеотидов, являющихся комплементарными последовательности-мишени ДНК, связанные с CRISPR (Cas) нуклеазы могут получать доступ практически к любому необходимому участку генома[1,2]. На сегодняшний день в рамках геномного редактирования были идентифицированы и адаптированы для применения несколько систем на основе мономерной бактериальной Cas нуклеазы[10]. Данное естественное разнообразие Cas нуклеаз, наряду с растущим набором модифицированных вариантов[11-14], обеспечивает благоприятную почву для разработки новых технологий геномного редактирования.

[0005] В то время как манипулирование генами с помощью CRISPR теперь является зрелым способом, точное редактирование одиночных пар оснований в геноме человека остается серьезной проблемой[3]. Направленная гомологичная репарация (HDR) долго применялась в клетках человека и других организмах для вставки, коррекции, или замены

ДНК последовательностей в участках двуцепочечных разрывов (DSB) с помощью донорских матриц для репарации ДНК, которые кодируют необходимые изменения[15]. Однако традиционная HDR имеет очень низкую эффективность в большинстве типов клеток человека, в частности в неделящихся клетках, при этом конкурентное негомологичное соединение концов (NHEJ) в основном приводит к побочным продуктам в виде вставок-делеций (инделям)[16]. Другие проблемы связаны с осуществлением DSB, которые могут вызывать крупные хромосомные перестроения и делеции в локусах-мишенях[17] или активировать ось p53, что ведет к остановке роста и апоптозу[18,19].

[0006] Были исследованы некоторые подходы для устранения данных недостатков HDR. Например, было показано, что репарация одноцепочечных разрывов ДНК олигонуклеотидными донорами снижает образование инделей, однако выходы желаемых продуктов репарации остаются низкими[20]. В рамках других стратегий осуществляются попытки направить репарацию к HDR вместо NHEJ с помощью низкомолекулярных и биологических реагентов[21-23]. Однако эффективность данных способов зависит от типа клеток, при этом нарушение нормального состояния клеток может вести к нежелательным и непредсказуемым эффектам.

[0007] Недавно Liu *et al.* разработали редактирование оснований в качестве технологии, которая обеспечивает редактирование нуклеотидов-мишеней без создания DSB или использования HDR[4–6,24–27]. Прямое модифицирование оснований ДНК с помощью слитых с Cas деаминаз обеспечивает C•G на T•A или A•T на G•C конверсии пар оснований в рамках короткого окна (~5-7 оснований) с очень высокой эффективностью. В результате этого редакторы оснований стали быстро использоваться научным сообществом. Однако некоторые факторы могут ограничивать их общую применимость в плане точного геномного редактирования.

[0008] Таким образом, разработка программируемых редакторов, которые способны вносить любое необходимое изменение одного или нескольких нуклеотидов, которые способны вносить нуклеотидные вставки или делеции (например, по меньшей мере, 1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, 100 или более вставок или делеций пар оснований), и/или которые могут изменять или модифицировать нуклеотидную последовательность в участке-мишени с высокой специфичностью и

эффективностью, значительно расширит диапазон и терапевтический потенциал технологий геномного редактирования на основе CRISPR.

КРАТКОЕ ОПИСАНИЕ ИЗОБРЕТЕНИЯ

[0009] Настоящее изобретение относится к новым композициям (например PEgRNA и включающим ее PE комплексам) и способам применения праймерного редактирования (PE) для репарации терапевтических мишеней, например мишеней, идентифицированных в базе данных ClinVar, с помощью PEgRNA, сконструированной с помощью специализированного описанного здесь алгоритма. Таким образом, в одном аспекте настоящее изобретение относится к алгоритму предсказания последовательностей PEgRNA в крупном масштабе, которые могут применяться для репарации терапевтических мишеней (например, включенных в базу данных ClinVar). Кроме того, настоящее изобретение относится к предсказанным последовательностям для сконструированных терапевтических PEgRNA, которые могут быть сконструированы с применением описанного алгоритма и которые могут применяться в рамках праймерного редактирования для репарации терапевтических мишеней.

[0010] Описанный здесь алгоритм и предсказанные последовательности PEgRNA в целом относятся к праймерному редактированию. Таким образом, данное изобретение также включает описание различных компонентов и аспектов праймерного редактирования, включая подходящий napDNAbp (например никазу Cas9) и полимеразу (например обратную транскриптазу), а также другие подходящие компоненты (например линкеры, NLS) и слитые белки PE, которые могут применяться с описанной здесь терапевтической PEgRNA.

[0011] Как описано здесь, праймерное редактирование представляет собой универсальный и точный способ геномного редактирования, в рамках которого осуществляется прямая запись новой генетической информации в определенный участок ДНК с помощью программируемого нуклеиновой кислотой ДНК-связывающего белка ("napDNAbp"), работающего совместно с полимеразой (то есть в форме слитого белка или иным образом поступающего *in trans* с napDNAbp), при этом система праймерного редактирования программируется направляющей РНК ("PEgRNA") для праймерного редактирования (PE), которая как указывает участок-мишень, так и служит матрицей для синтеза требуемого

изменения в форме замещающей цепи ДНК в виде удлинения (ДНК или РНК), встроенного в направляющую РНК (например на 5′ или 3′ конце, или во внутренней части направляющей РНК). Замещающая цепь, содержащая требуемое изменение (например замещение одного азотистого основания), имеет ту же последовательность, что и эндогенная цепь подвергаемого редактированию участка-мишени (за исключением того, что она включает требуемое изменение). С помощью механизма репарации ДНК и/или репликации эндогенная цепь участка-мишени замещается новой синтезированной замещающей цепью, содержащей требуемое изменение. В некоторых случаях праймерное редактирование может быть описано как технология геномного редактирования типа "поиск-и-замена", поскольку праймерные редакторы, как описано здесь, не только обеспечивают поиск и нахождение требуемого участка-мишени для редактирования, но в то же время кодируют замещающую цепь, содержащую требуемое изменение, которая встраивается вместо соответствующей эндогенной цепи ДНК участка-мишени. Праймерные редакторы по настоящему изобретению частично связаны с открытием того, что механизм мишень-инициированной обратной транскрипции (TPRT) или "праймерное редактирование" может быть использован или адаптирован для осуществления точного геномного редактирования на основе CRISPR/Cas с высокой эффективностью и генетической гибкостью (например как показано в рамках различных вариантов осуществления на Фиг.1A-1F). TPRT естественным образом применяется в рамках подвижных элементов ДНК, например ретротранспозонов млекопитающих без длинных концевых повторов (LTR) и бактериальных интронов Группы II[28,29]. Здесь авторы использовали продукты слияния белок Cas-обратная транскриптаза или родственные системы для нацеливания на определенную последовательность ДНК с помощью направляющей РНК, осуществления одноцепочечного разрыва в участке-мишени, и применения никированной ДНК в качестве праймера для обратной транскрипции сконструированной матрицы в виде обратной транскриптазы, которая встраивается в направляющую РНК. Однако в то время как данная концепция начинается с праймерных редакторов, в рамках которых используются обратные транскриптазы в качестве компонента ДНК-полимеразы, описанные здесь праймерные редакторы не ограничиваются обратными транскриптазами, но могут включать использование практически любой ДНК-полимеразы. Действительно, в то время как по тексту заявки

могут указываться праймерные редакторы с "обратными транскриптазами", здесь устанавливается, что обратные транскриптазы являются лишь одним из типов ДНК-полимераз, которые могут работать в рамках праймерного редактирования. Таким образом, при указании в данном описании "обратных транскриптаз" специалист в данной области должен понимать, что вместо обратной транскриптазы могут использоваться любые подходящие ДНК-полимеразы. Таким образом, в одном аспекте праймерные редакторы могут включать Cas9 (или эквивалентный napDNAbp), который программируется для нацеливания на ДНК-последовательность путем его связывания со специализированной направляющей РНК (то есть PEgRNA), содержащей спейсерную последовательность, которая отжигается на комплементарном протоспейсере в ДНК-мишени. Специализированная направляющая РНК также содержит новую генетическую информацию в форме удлинения, которое кодирует замещающую цепь ДНК, содержащую требуемое генетическое изменение, которая используется для замены соответствующей эндогенной цепи ДНК в участке-мишени. Для переноса информации из PEgRNA в ДНК-мишень механизм праймерного редактирования включает никирование участка-мишени в одной цепи ДНК для высвобождения 3′-гидроксильной группы. Затем свободная 3′-гидроксильная группа может использоваться для праймирования полимеризации ДНК кодирующего изменение удлинения на PEgRNA напрямую в участок-мишень. В различных вариантах осуществления удлинение, которое обеспечивает матрицу для полимеризации замещающей содержащей изменение цепи, может быть образовано из РНК или ДНК. В случае РНК-удлинения полимераза праймерного редактора может представлять собой РНК-зависимую ДНК-полимеразу (например обратную транскриптазу). В случае ДНК-удлинения полимераза праймерного редактора может представлять собой ДНК-зависимую ДНК-полимеразу.

[0012] Новая синтезированная цепь (то есть замещающая цепь ДНК, содержащая требуемое изменение), которая образуется описанными здесь праймерными редакторами, будет являться гомологичной относительно последовательности геномной мишени (то есть иметь такую же последовательность) за исключением включения требуемого нуклеотидного изменения (например изменения отдельного нуклеотида, делеции, или вставки, или их комбинации). Новая синтезированная (или замещающая) цепь ДНК также может быть обозначена как одноцепочечный ДНК-флэп, который будет конкурировать за

гибридизацию с комплементарной гомологичной эндогенной цепью ДНК, тем самым вытесняя соответствующую эндогенную цепь. В определенных вариантах осуществления система может комбинироваться с использованием фермента в виде подверженной ошибкам обратной транскриптазы (например поступающей в виде слитого белка с доменом Cas9 или поступающей *in trans* с доменом Cas9). Фермент в виде подверженной ошибкам обратной транскриптазы может вносить изменения при синтезе одноцепочечного ДНК-флэпа. Таким образом, в определенных вариантах осуществления подверженная ошибкам обратная транскриптаза может использоваться для внесения нуклеотидных изменений в ДНК-мишень. В зависимости от подверженной ошибкам обратной транскриптазы, которая используется в системе, изменения могут являться случайными или неслучайными.

[0013] Коррекция гибридизированного промежуточного продукта (включающего одноцепочечный ДНК-флэп, синтезированный обратной транскриптазой, гибридизированной с эндогенной цепью ДНК) может включать удаление образованного вытесненного флэпа эндогенной ДНК (например с помощью эндонуклеазы 5′-концевого ДНК-флэпа, FEN1), лигирование синтезированного одноцепочечного ДНК-флэпа с ДНК-мишенью, и ассимилирование требуемого нуклеотидного изменения в результате процессов репарации и/или репликации клеточной ДНК Поскольку синтез ДНК на основе матриц обеспечивает точность на уровне отдельных нуклеотидов при модификации любого нуклеотида, включая вставки и делеции, объем данного подхода является очень широким и, как представляется, он может использоваться в рамках множества вариантов применения в фундаментальной науке и терапии.

Алгоритм и способы конструирования терапевтических PEgRNA

[0014] В одном аспекте настоящее изобретение относится к новому алгоритму конструирования терапевтической PEgRNA, в частности реализуемому в крупном масштабе вместо единичного осуществления конструирования PEgRNA.

[0015] Соответственно, некоторые аспекты относятся к компьютеризированному способу определения последовательности направляющей РНК праймерного редактора (PEgRNA). Способ включает применение, по меньшей мере, одного компьютерного аппаратного процессора для доступа к данным, связанным с входным аллелем, выходным аллелем, и слитыми белками, включающими программируемый нуклеиновой кислотой ДНК-

связывающий белок и полимеразу (например обратную транскриптазу). Способ включает определение последовательности PEgRNA на основе входного аллеля, выходного аллеля, и слитого белка, при этом последовательность PEgRNA конструируется для ассоциирования со слитым белком для изменения входного аллеля с получением выходного аллеля, включая определение для последовательности PEgRNA одного или нескольких из следующих свойств: спейсера, комплементарного нуклеотидной последовательности-мишени во входном аллеле (то есть спейсера, как определено на Фиг.27); каркаса нРНК для взаимодействия со слитым белком (то есть ядра нРНК, как определено на Фиг.27); и удлинения (то есть плеча удлинения, как показано на Фиг.27), включающего одну или несколько из: матрицы для синтеза ДНК (как показано на Фиг.27), включающей требуемое нуклеотидное изменение для изменения входного аллеля с получением выходного аллеля; участка связывания праймера (то есть участка связывания праймера, как показано на Фиг.27). PEgRNA также может включать 3' сигнал терминации, который терминирует транскрипцию из промотора. Кроме того, PEgRNA может включать первый модификатор на 5'-конце плеча удлинения и второй модификатор на 3'-конце плеча удлинения. Такие последовательности (показанные как "e1" и "e2" на Фиг.27) могут включать последовательности типа "стебель-петля", которые могут повышать стабильность PEgRNA.

[0016] В некоторых примерах способ включает определение спейсера и удлинения, а также определение того, что спейсер расположен на 5'-конце PEgRNA, а удлинение расположено на 3'-конце структуры PEgRNA.

[0017] В некоторых примерах способ включает определение спейсера и удлинения, а также определение того, что спейсер расположен на 5'-конце PEgRNA, а удлинение расположено 3' относительно спейсера.

[0018] В некоторых примерах доступ к данным, связанным с входным аллелем и выходным аллелем, включает доступ к базе данных, включающей набор входных аллелей и ассоциированных выходных аллелей. Доступ к базе данных может включать доступ к базе данных ClinVar Национального центра биотехнологической информации (www.ncbi.nlm.nih.gov/clinvar/), включающей множество записей, при этом каждая запись включает входной аллель из набора входных аллелей и выходной аллель из набора выходных аллелей (например относящихся к дикому типу или аллелей с требуемой

активностью). Определение последовательности PEgRNA может включать определение одной или нескольких последовательностей PEgRNA для каждого входного аллеля и ассоциированного выходного аллеля из набора.

[0019] В некоторых примерах доступ к данным, связанным со слитым белком, включает определение слитого белка из множества слитых белков.

[0020] В некоторых примерах слитый белок включает белок Cas9. Слитый белок может включать белок Cas9-NG, Cas9-NGG, saCas9-KKH, или белок SpCas9.

[0021] В некоторых примерах изменение входного аллеля с получением выходного аллеля включает изменение одного нуклеотида, вставку одного или нескольких нуклеотидов, делецию одного или нескольких нуклеотидов, или их комбинацию.

[0022] В некоторых вариантах осуществления способ включает определение спейсера, при этом спейсер включает нуклеотидную последовательность длиной от 1 до 40 нуклеотидов. В некоторых вариантах осуществления способ включает определение спейсера, при этом спейсер включает нуклеотидную последовательность длиной от 5 до 35 нуклеотидов. В некоторых вариантах осуществления способ включает определение спейсера, при этом спейсер включает нуклеотидную последовательность длиной от 10 до 30 нуклеотидов. В некоторых вариантах осуществления способ включает определение спейсера, при этом спейсер включает нуклеотидную последовательность длиной от 15 до 25 нуклеотидов. В некоторых вариантах осуществления способ включает определение спейсера, при этом спейсер включает нуклеотидную последовательность длиной приблизительно 20 нуклеотидов. Способ может включать определение спейсера на основе положения изменения в соответствующей нуклеотидной последовательности протоспейсера. Изменение может быть внесено в окне редактирования, которое располагается приблизительно от положения -15 протоспейсера до положения +39 протоспейсера. Изменение может быть внесено в окне редактирования, которое располагается приблизительно от положения -10 протоспейсера до положения +34 протоспейсера. Изменение может быть внесено в окне редактирования, которое располагается приблизительно от положения -5 протоспейсера до положения +29 протоспейсера. Изменение может быть внесено в окне редактирования, которое располагается приблизительно от положения -1 протоспейсера до положения +27 протоспейсера.

[0023] В некоторых примерах способ может включать: определение набора исходных протоспейсеров-кандидатов на основе входного аллеля и слитого белка, при этом каждый исходный протоспейсер-кандидат включает PAM слитого белка во входном аллеле; определение одного или нескольких исходных протоспейсеров-кандидатов из набора исходных протоспейсеров-кандидатов, каждый из которых включает несовместимое положение разрыва; удаление определенного одного или нескольких исходных протоспейсеров-кандидатов из набора для получения набора оставшихся протоспейсеров-кандидатов; и при этом где определение структуры PEgRNA включает определение нескольких структур PEgRNA, при этом каждая структура PEgRNA включает различный спейсер, определенный на основе соответствующего протоспейсера из набора оставшихся протоспейсеров-кандидатов.

[0024] В некоторых примерах способ включает определение удлинения и матрицы для синтеза ДНК (например последовательности RT матрицы), при этом матрица для синтеза ДНК (например последовательность RT матрицы) включает приблизительно от 1 до 40 нуклеотидов. В некоторых примерах способ включает определение удлинения и матрицы для синтеза ДНК (например последовательности RT матрицы), при этом матрица для синтеза ДНК (например последовательность RT матрицы) включает приблизительно от 3 до 38 нуклеотидов. В некоторых примерах способ включает определение удлинения и матрицы для синтеза ДНК (например последовательности RT матрицы), при этом матрица для синтеза ДНК (например последовательность RT матрицы) включает приблизительно от 5 до 36 нуклеотидов. В некоторых примерах способ включает определение удлинения и матрицы для синтеза ДНК (например последовательности RT матрицы), при этом матрица для синтеза ДНК (например последовательность RT матрицы) включает приблизительно от 7 до 34 нуклеотидов.

[0025] В некоторых примерах определение PEgRNA включает определение спейсера на основе входного аллеля и/или слитого белка и определение матрицы для синтеза ДНК (например последовательности RT матрицы) на основе спейсера.

[0026] В некоторых примерах матрица для синтеза ДНК (например последовательность RT матрицы) кодирует одноцепочечный ДНК-флэп, который комплементарен последовательности эндогенной ДНК, примыкающей к участку разрыва, при этом одноцепочечный ДНК-флэп включает требуемое нуклеотидное изменение.

Одноцепочечный ДНК-флэп может гибридизироваться с последовательностью эндогенной ДНК, примыкающей к участку разрыва, тем самым обеспечивая внесение требуемого нуклеотидного изменения. Одноцепочечный ДНК-флэп может вытеснять последовательность эндогенной ДНК, примыкающую к участку разрыва. Клеточная репарация одноцепочечного ДНК-флэпа может приводить к внесению требуемого нуклеотидного изменения, тем самым обеспечивая образование требуемого продукта.

[0027] В некоторых примерах слитый белок при образовании комплекса с PEgRNA способен связываться с последовательностью-мишенью ДНК. Последовательность-мишень ДНК может включать цепь-мишень, в которой осуществляется изменение, и комплементарную не являющуюся мишенью цепь.

[0028] В некоторых примерах входной аллель включает патогенную мутацию ДНК, а выходной аллель включает скорректированную последовательность ДНК.

[0029] Некоторые варианты осуществления связаны с системой, включающей, по меньшей мере, один процессор и, по меньшей мере, один машиночитаемый носитель данных, на котором закодированы инструкции, которые при выполнении побуждают указанный, по меньшей мере, один процессор выполнять компьютеризированные способы для определения структуры PEgRNA.

[0030] Некоторые варианты осуществления связаны с, по меньшей мере, одним машиночитаемым носителем данных, на котором закодированы инструкции, которые при выполнении побуждают, по меньшей мере, один процессор выполнять компьютеризированные способы для определения последовательности PEgRNA. Некоторые варианты осуществления связаны со способом редактирования оснований с применением PEgRNA, определенной в соответствии с компьютеризированными способами для определения PEgRNA.

Терапевтическая PEgRNA

[0031] В другом аспекте настоящее изобретение предусматривает терапевтическую PEgRNA, которая была сконструирована с помощью описанного здесь алгоритма, как указано на Фиг.27 и Фиг.28.

[0032] Например, PEgRNA, которые могут использоваться в рамках приведенного здесь описания, приведены в качестве примера на Фиг.27. На Фиг.3D приведена структура варианта осуществления предусмотренной здесь PEgRNA, которая может быть

сконструирована в соответствии с методологией, определенной в рамках Примера 2. PEgRNA включает три основных компонента в направлении от 5′ к 3′, а именно: спейсер, ядро нРНК, и плечо удлинения на 3′-конце. Плечо удлинения может быть дополнительно разделено на следующие структурные элементы в направлении от 5′ к 3′, а именно: участок связывания праймера (A), матрицу для редактирования (B), и плечо гомологии (C). Кроме того, PEgRNA может включать необязательный 3′-концевой модификаторный участок (e1) и необязательный 5′-концевой модификаторный участок (e2). Кроме того, PEgRNA может включать сигнал терминации транскрипции на 3′-конце PEgRNA (не показан). Данные структурные элементы дополнительно определены здесь. Изображение структуры PEgRNA не является ограничивающим и охватывает вариации в плане расположения элементов. Например, необязательные модификаторы последовательности (e1) и (e2) могут быть расположены внутри или между любыми из показанных участков без ограничения в виде расположения на 3′- или 5′-концах. PEgRNA, показанная на Фиг.27, может быть сконструирована с помощью описанного здесь алгоритма.

[0033] В другом примере на Фиг.28 приведена структура другого варианта осуществления предусмотренной здесь PEgRNA, которая может быть сконструирована в соответствии с методологией, определенной в рамках Примера 2. PEgRNA включает три основных компонента в направлении от 5′ к 3′, а именно: спейсер, ядро нРНК, и плечо удлинения на 3′-конце. Плечо удлинения может быть дополнительно разделено на следующие структурные элементы в направлении от 5′ к 3′, а именно: участок связывания праймера (A), матрицу для редактирования (B), и плечо гомологии (C). Кроме того, PEgRNA может включать необязательный 3′-концевой модификаторный участок (e1) и необязательный 5′-концевой модификаторный участок (e2). Кроме того, PEgRNA может включать сигнал терминации транскрипции на 3′-конце PEgRNA (не показан). Данные структурные элементы дополнительно определены здесь. Изображение структуры PEgRNA не является ограничивающим и охватывает вариации в плане расположения элементов. Например, необязательные модификаторы последовательности (e1) и (e2) могут быть расположены внутри или между любыми из показанных участков без ограничения в виде расположения на 3′- или 5′-концах. PEgRNA, показанная на Фиг.27, может быть сконструирована с помощью описанного здесь алгоритма.

[0034] В различных вариантах осуществления изобретение предусматривает PEgRNA согласно SEQ ID NO: 1-135514 и 813085-880462, сконструированную с помощью описанного здесь алгоритма в соответствии с записями из базы данных ClinVar.

[0035] В различных других вариантах осуществления примеры PEgRNA, сконструированной в соответствии с базой данных ClinVar с использованием описанного здесь алгоритма, включены в Перечень последовательностей, который является частью настоящего описания. Перечень последовательностей включает последовательности полной PEgRNA SEQ ID NO: 1-135514 и 813085-880462. Каждая из данных полных PEgRNA состоит из спейсера (SEQ ID NO: 135515 – 271028 и 880463-947840) и плеча удлинения (SEQ ID NO: 271029 – 406542 и 947841-1015218). Кроме того, каждая PEgRNA включает ядро нРНК, например как определено согласно SEQ ID NO: 1361579-1361580. Каждое плечо удлинения согласно SEQ ID NO: 271029 – 406542 и 947841-1015218 дополнительно состоит из участка связывания праймера (SEQ ID NO: 406543 – 542056 и 1015219-1082596), матрицы для редактирования (SEQ ID NO.: 542057 – 677570 и 1082597-1149974), и плеча гомологии (SEQ ID NO.: 677571 – 813084 и 1149975-1217352). PEgRNA может необязательно включать 5′-концевой модификаторный участок и/или 3′-концевой модификаторный участок. PEgRNA может также включать сигнал терминации обратной транскрипции (например SEQ ID NO: 1361560-1361566) на 3'-конце PEgRNA. Изобретение охватывает конструирование и применение всех данных последовательностей.

[0036] В различных вариантах осуществления направляющая РНК праймерного редактора включает (a) направляющую РНК и (b) удлинение РНК на 5'- или 3'-конце направляющей РНК или во внутримолекулярном положении в направляющей РНК, примеры которой изображены на Фиг.3A-C. Удлинение РНК может включать (i) матрицу для синтеза ДНК, включающую требуемое нуклеотидное изменение, (ii) участок связывания праймера обратной транскрипции, и (iii) необязательно последовательность линкера. В различных вариантах осуществления матрица для синтеза ДНК кодирует одноцепочечный ДНК-флэп, который комплементарен последовательности эндогенной ДНК, примыкающей к участку разрыва, при этом одноцепочечный ДНК-флэп включает требуемое нуклеотидное изменение.

[0037] В различных вариантах осуществления плечо удлинения РНК имеет длину, по меньшей мере, 5 нуклеотидов, по меньшей мере, 6 нуклеотидов, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, по меньшей мере, 20 нуклеотидов, по меньшей мере, 21 нуклеотид, по меньшей мере, 22 нуклеотида, по меньшей мере, 23 нуклеотида, по меньшей мере, 24 нуклеотида, или, по меньшей мере, 25 нуклеотидов.

[0038] В определенных вариантах осуществления направляющая РНК праймерного редактора включает нуклеотидную последовательность SEQ ID NO: 1361548-1361581 или нуклеотидную последовательность, имеющую, по меньшей мере, 85%, или, по меньшей мере, 90%, или, по меньшей мере, 95%, или, по меньшей мере, 98%, или, по меньшей мере, 99% идентичность последовательности относительно любой одной из SEQ ID NO: 1361548-1361581.

[0039] В некоторых вариантах осуществления направляющая РНК праймерного редактора (PEgRNA) включает вариант нуклеотидной последовательности SEQ ID NO: 1361548-1361581, включающий, по меньшей мере, одну мутацию по сравнению с нуклеотидной последовательностью SEQ ID NO: 1361548-1361581. В некоторых вариантах осуществления вариант включает более чем 1 (например 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 или более) мутаций по сравнению с нуклеотидной последовательностью SEQ ID NO: 1361548-1361581.

[0040] В другом аспекте настоящее изобретение предусматривает направляющую РНК праймерного редактора, включающую направляющую РНК и, по меньшей мере, одно удлинение РНК (то есть плечо удлинения согласно Фиг.27). Удлинение РНК расположено на 3′-конце направляющей РНК. В других вариантах осуществления удлинение РНК расположено на 5′-конце направляющей РНК. В других вариантах осуществления удлинение РНК расположено во внутримолекулярном положении в направляющей РНК, при этом внутримолекулярное положение удлиненной части предпочтительно не нарушает функционирование протоспейсера.

[0041] В различных вариантах осуществления направляющая РНК праймерного редактора (PEgRNA) способна связываться с napDNAbp и направлять napDNAbp к последовательности-мишени ДНК. Последовательность-мишень ДНК может включать цепь-мишень и комплементарную не являющуюся мишенью цепь, при этом направляющая РНК гибридизируется с цепью-мишенью для образования РНК-ДНК гибрида и R-петли.

[0042] В различных вариантах осуществления направляющей РНК праймерного редактора указанное, по меньшей мере, одно удлинение РНК включает матрицу для синтеза ДНК. В различных других вариантах осуществления удлинение РНК дополнительно включает участок связывания праймера обратной транскрипции. В других вариантах осуществления удлинение РНК включает линкер или спейсер, который присоединяет удлинение РНК к направляющей РНК.

[0043] В различных вариантах осуществления удлинение РНК может иметь длину, по меньшей мере, 5 нуклеотидов, по меньшей мере, 6 нуклеотидов, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, по меньшей мере, 20 нуклеотидов, по меньшей мере, 21 нуклеотидов, по меньшей мере, 22 нуклеотидов, по меньшей мере, 23 нуклеотидов, по меньшей мере, 24 нуклеотидов, по меньшей мере, 25 нуклеотидов, по меньшей мере, 30 нуклеотидов, по меньшей мере, 40 нуклеотидов, по меньшей мере, 50 нуклеотидов, по меньшей мере, 60 нуклеотидов, по меньшей мере, 70 нуклеотидов, по меньшей мере, 80 нуклеотидов, по меньшей мере, 90 нуклеотидов, по меньшей мере, 100 нуклеотидов, по меньшей мере, 150 нуклеотидов, по меньшей мере, 200 нуклеотидов, по меньшей мере, 300 нуклеотидов, по меньшей мере, 400 нуклеотидов, или, по меньшей мере, 500 нуклеотидов.

[0044] В других вариантах осуществления матрица для синтеза ДНК (то есть матрица для редактирования согласно Фиг.27) имеет длину, по меньшей мере, 3 нуклеотида, по меньшей мере, 4 нуклеотида, по меньшей мере, 5 нуклеотидов, по меньшей мере, 6 нуклеотидов, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по

меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, по меньшей мере, 20 нуклеотидов, по меньшей мере, 30 нуклеотидов, по меньшей мере, 40 нуклеотидов, по меньшей мере, 50 нуклеотидов, по меньшей мере, 60 нуклеотидов, по меньшей мере, 70 нуклеотидов, по меньшей мере, 80 нуклеотидов, по меньшей мере, 90 нуклеотидов, по меньшей мере, 100 нуклеотидов, по меньшей мере, 200 нуклеотидов, по меньшей мере, 300 нуклеотидов, по меньшей мере, 400 нуклеотидов, или, по меньшей мере, 500 нуклеотидов.

[0045] В других вариантах осуществления последовательность участка связывания праймера обратной транскрипции (то есть участка связывания праймера согласно Фиг.27) имеет длину, по меньшей мере, 3 нуклеотида, по меньшей мере, 4 нуклеотида, по меньшей мере, 5 нуклеотидов, по меньшей мере, 6 нуклеотидов, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, по меньшей мере, 20 нуклеотидов, по меньшей мере, 30 нуклеотидов, по меньшей мере, 40 нуклеотидов, по меньшей мере, 50 нуклеотидов, по меньшей мере, 60 нуклеотидов, по меньшей мере, 70 нуклеотидов, по меньшей мере, 80 нуклеотидов, по меньшей мере, 90 нуклеотидов, по меньшей мере, 100 нуклеотидов, по меньшей мере, 200 нуклеотидов, по меньшей мере, 300 нуклеотидов, по меньшей мере, 400 нуклеотидов, или, по меньшей мере, 500 нуклеотидов.

[0046] В других вариантах осуществления необязательный линкер или спейсер имеет длину, по меньшей мере, 3 нуклеотида, по меньшей мере, 4 нуклеотида, по меньшей мере, 5 нуклеотидов, по меньшей мере, 6 нуклеотидов, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15

нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, по меньшей мере, 20 нуклеотидов, по меньшей мере, 30 нуклеотидов, по меньшей мере, 40 нуклеотидов, по меньшей мере, 50 нуклеотидов, по меньшей мере, 60 нуклеотидов, по меньшей мере, 70 нуклеотидов, по меньшей мере, 80 нуклеотидов, по меньшей мере, 90 нуклеотидов, по меньшей мере, 100 нуклеотидов, по меньшей мере, 200 нуклеотидов, по меньшей мере, 300 нуклеотидов, по меньшей мере, 400 нуклеотидов, или, по меньшей мере, 500 нуклеотидов.

[0047] Описанная здесь сконструированная PEgRNA может образовывать комплекс со слитым белком праймерного редактора.

[0048] В одном аспекте изобретение предусматривает слитый белок праймерного редактора, включающий программируемый нуклеиновой кислотой ДНК-связывающий белок (napDNAbp) и обратную транскриптазу. В различных вариантах осуществления слитый белок способен осуществлять геномное редактирование с помощью мишень-инициированной обратной транскрипции в присутствии направляющей РНК праймерного редактора (PEgRNA).

[0049] В некоторых вариантах осуществления napDNAbp выбран из группы, состоящей из: Cas9, CasX, CasY, Cpf1, C2c1, C2c2, C2C3, и Argonaute, и необязательно имеет никазную активность.

[0050] В других вариантах осуществления слитый белок при образовании комплекса с направляющей РНК праймерного редактора как описано здесь способен связываться с последовательностью-мишенью ДНК (например геномной ДНК).

[0051] В других вариантах осуществления последовательность-мишень ДНК включает цепь-мишень и комплементарную не являющуюся мишенью цепь.

[0052] В других вариантах осуществления связывание слитого белка в виде комплекса с направляющей РНК праймерного редактора образует R-петлю. R-петля может включать (i) РНК-ДНК гибрид, включающий направляющую РНК праймерного редактора и цепь-мишень, и (ii) комплементарную не являющуюся мишенью цепь.

[0053] В других вариантах осуществления комплементарная не являющаяся мишенью цепь никируется для образования праймерной последовательности обратной транскриптазы, имеющей свободный 3´-конец.

[0054] В других вариантах осуществления одноцепочечный ДНК-флэп гибридизируется с последовательностью эндогенной ДНК, примыкающей к участку разрыва, тем самым обеспечивая внесение требуемого нуклеотидного изменения. В других вариантах осуществления одноцепочечный ДНК-флэп замещает последовательность эндогенной ДНК, примыкающую к участку разрыва и имеющую свободный 5′-конец. В некоторых вариантах осуществления вытесненная эндогенная ДНК, имеющая 5′-конец, подвергается эксцизии клеткой.

[0055] В различных вариантах осуществления клеточная репарация одноцепочечного ДНК-флэпа приводит к внесению требуемого нуклеотидного изменения, тем самым обеспечивая образование требуемого продукта.

[0056] В различных других вариантах осуществления требуемое нуклеотидное изменение вносится в окне редактирования, которое соответствует от -4 до +10 относительно последовательности PAM.

[0057] В других вариантах осуществления требуемое нуклеотидное изменение вносится в окне редактирования, которое соответствует приблизительно от -5 до +5 относительно участка разрыва, или приблизительно от -10 до +10 относительно участка разрыва, или приблизительно от -20 до +20 относительно участка разрыва, или приблизительно от -30 до +30 относительно участка разрыва, или приблизительно от -40 до +40 относительно участка разрыва, или приблизительно от -50 до +50 относительно участка разрыва, или приблизительно от -60 до +60 относительно участка разрыва, или приблизительно от -70 до +70 относительно участка разрыва, или приблизительно от -80 до +80 относительно участка разрыва, или приблизительно от -90 до +90 относительно участка разрыва, или приблизительно от -100 до +100 относительно участка разрыва, или приблизительно от -200 до +200 относительно участка разрыва.

[0058] В различных вариантах осуществления napDNAbp включает аминокислотную последовательность SEQ ID NO: 1361421. В различных других вариантах осуществления napDNAbp включает аминокислотную последовательность, которая, по меньшей мере, на 80%, 85%, 90%, 95%, 98%, или 99% идентична аминокислотной последовательности любой одной из SEQ ID NO: 1361421-1361484, и 1361593-1361596.

[0059] В других вариантах осуществления обратная транскриптаза описанных слитых белков и/или композиций может включать любую одну из аминокислотных

последовательностей SEQ ID NO: 1361485-1361514, и 1361597-1361598. В других вариантах осуществления обратная транскриптаза может включать аминокислотную последовательность, которая, по меньшей мере, на 80%, 85%, 90%, 95%, 98%, или 99% идентична аминокислотной последовательности любой одной из SEQ ID NO: 1361485-1361514, и 1361597-1361598. Данные последовательности могут являться встречающимися в природе последовательностями обратной транскриптазы, например из ретровируса или ретротранспозона, или данные последовательности могут являться не встречающимися в природе или сконструированными.

[0060] В различных других вариантах осуществления описанные здесь слитые белки могут включать различные структурные конфигурации. Например, слитые белки могут включать структуру NH_2-[napDNAbp]-[обратная транскриптаза]-COOH; или NH_2-[обратная транскриптаза]-[napDNAbp]-COOH, при этом каждый экземпляр "]-[" обозначает наличие последовательности необязательного линкера.

[0061] В различных вариантах осуществления последовательность линкера включает аминокислотную последовательность SEQ ID NO: 1361520-1361530, 1361585, и 1361603, или аминокислотную последовательность, которая, по меньшей мере, на 80%, 85%, или 90%, или 95%, или 99% идентична любой одной аминокислотной последовательности линкера SEQ ID NO: 1361520-1361530, 1361585, и 1361603.

[0062] В различных вариантах осуществления требуемое нуклеотидное изменение, которое вносится в ДНК-мишень, может представлять собой изменение одного нуклеотида (например транзицию или трансверсию), вставку одного или нескольких нуклеотидов, делецию одного или нескольких нуклеотидов, или их комбинацию.

[0063] В определенных случаях вставка имеет длину, по меньшей мере, 1, по меньшей мере, 2, по меньшей мере, 3, по меньшей мере, 4, по меньшей мере, 5, по меньшей мере, 6, по меньшей мере, 7, по меньшей мере, 8, по меньшей мере, 9, по меньшей мере, 10, по меньшей мере, 11, по меньшей мере, 12, по меньшей мере, 13, по меньшей мере, 14, по меньшей мере, 15, по меньшей мере, 16, по меньшей мере, 17, по меньшей мере, 18, по меньшей мере, 19, по меньшей мере, 20, по меньшей мере, 30, по меньшей мере, 40, по меньшей мере, 50, по меньшей мере, 60, по меньшей мере, 70, по меньшей мере, 80, по меньшей мере, 90, по меньшей мере, 100, по меньшей мере, 200, по меньшей мере, 300, по меньшей мере, 400, или, по меньшей мере, 500 нуклеотидов.

[0064] В определенных случаях делеция имеет длину, по меньшей мере, 1, по меньшей мере, 2, по меньшей мере, 3, по меньшей мере, 4, по меньшей мере, 5, по меньшей мере, 6, по меньшей мере, 7, по меньшей мере, 8, по меньшей мере, 9, по меньшей мере, 10, по меньшей мере, 11, по меньшей мере, 12, по меньшей мере, 13, по меньшей мере, 14, по меньшей мере, 15, по меньшей мере, 16, по меньшей мере, 17, по меньшей мере, 18, по меньшей мере, 19, по меньшей мере, 20, по меньшей мере, 30, по меньшей мере, 40, по меньшей мере, 50, по меньшей мере, 60, по меньшей мере, 70, по меньшей мере, 80, по меньшей мере, 90, по меньшей мере, 100, по меньшей мере, 200, по меньшей мере, 300, по меньшей мере, 400, или, по меньшей мере, 500 нуклеотидов.

[0065] В различных вариантах осуществления направляющих РНК праймерного редактора матрица для синтеза ДНК (то есть матрица для редактирования согласно Фиг.27) может кодировать одноцепочечный ДНК-флэп, который комплементарен последовательности эндогенной ДНК, примыкающей к участку разрыва, при этом одноцепочечный ДНК-флэп включает требуемое нуклеотидное изменение. Одноцепочечный ДНК-флэп может вытеснять эндогенную одноцепочечную ДНК в участке разрыва. Вытесненная эндогенная одноцепочечная ДНК в участке разрыва может иметь 5′-конец и образовывать эндогенный флэп, который может быть подвергнут эксцизии клеткой. В различных вариантах осуществления эксцизия 5′-концевого эндогенного флэпа может обеспечивать стимулирование образования продукта, поскольку удаление 5′-концевого эндогенного флэпа стимулирует гибридизирование одноцепочечного 3′ ДНК-флэпа с соответствующей комплементарной цепью ДНК и встраивание или ассимилирование требуемого нуклеотидного изменения, которое несет одноцепочечный 3′ ДНК-флэп, в ДНК-мишень.

[0066] В различных вариантах осуществления направляющих РНК праймерного редактора клеточная репарация одноцепочечного ДНК-флэпа приводит к внесению требуемого нуклеотидного изменения, тем самым обеспечивая образование требуемого продукта.

[0067] В другом аспекте изобретения предусмотрены комплексы, включающие описанный здесь слитый белок и любую описанную выше направляющую РНК праймерного редактора (PEgRNA).

[0068] В других аспектах изобретения предусмотрен комплекс, включающий napDNAbp (например Cas9) и направляющую РНК праймерного редактора. napDNAbp может представлять собой никазу Cas9 (например spCas9) или может представлять собой

аминокислотную последовательность SEQ ID NO: 1361421 или аминокислотную последовательность, которая, по меньшей мере, на 80%, по меньшей мере, на 85%, по меньшей мере, на 90%, по меньшей мере, на 95%, по меньшей мере, на 98%, или, по меньшей мере, на 99% идентично аминокислотной последовательности любой одной из SEQ ID NO: 1361421-1361484, и 1361593-1361596.

[0069] В различных вариантах осуществления, включающих комплекс, направляющая РНК праймерного редактора способна направлять napDNAbp к последовательности-мишени ДНК. В различных вариантах осуществления обратная транскриптаза может поступать *in trans*, то есть поступать из другого источника, отличного от самого комплекса. Например, обратная транскриптаза может поступать в ту же клетку, имеющую комплекс, путем введения отдельного вектора, который отдельным образом кодирует обратную транскриптазу.

[0070] В другом аспекте изобретение предусматривает фармацевтические композиции (например описанные здесь слитые белки, PEgRNA согласно SEQ ID NO: 1-135,514). В некоторых вариантах осуществления фармацевтические композиции включают один или несколько из napDNAbp, слитого белка, обратной транскриптазы, и направляющей РНК праймерного редактора. В некоторых вариантах осуществления описанный здесь слитый белок и фармацевтически приемлемый наполнитель. В других вариантах осуществления фармацевтические композиции включают любую описанную здесь удлиненную направляющую РНК и фармацевтически приемлемый наполнитель. В других вариантах осуществления фармацевтические композиции включают любую описанную здесь удлиненную направляющую РНК в комбинации с любым описанным здесь слитым белком и фармацевтически приемлемым наполнителем. В других вариантах осуществления фармацевтические композиции включают любую полинуклеотидную последовательность, кодирующую один или несколько из napDNAbp, слитого белка, обратной транскриптазы, и направляющей РНК праймерного редактора. В других вариантах осуществления различные описанные здесь компоненты могут быть разделены на одну или несколько фармацевтических композиций. Например, первая фармацевтическая композиция может включать слитый белок или napDNAbp, вторая фармацевтическая композиция может включать обратную транскриптазу, а третья

фармацевтическая композиция может включать направляющую РНК праймерного редактора.

[0071] В другом аспекте настоящее изобретение предусматривает наборы. В одном варианте осуществления набор включает один или несколько полинуклеотидов, кодирующих один или несколько компонентов, включая слитый белок, napDNAbp, обратную транскриптазу, и направляющую РНК праймерного редактора (например любую из SEQ ID NO: 1-135514 или 813085-880462). Наборы могут также включать векторы, клетки, и выделенные препараты полипептидов, включая любой описанный здесь слитый белок, napDNAbp, или обратную транскриптазу.

[0072] В другом аспекте настоящее изобретение предусматривает способы применения описанных композиций на основе PEgRNA.

[0073] В одном варианте осуществления способы относятся к способу внесения требуемого нуклеотидного изменения в двуцепочечную ДНК с применением описанной здесь PEgRNA. Способ сначала включает контактирование последовательности двуцепочечной ДНК с комплексом, включающим слитый белок и направляющую РНК праймерного редактора, как описано здесь, при этом слитый белок включает napDNAbp и обратную транскриптазу, и при этом направляющая РНК праймерного редактора включает матрицу для синтеза ДНК, включающую требуемое нуклеотидное изменение. napDNAbp обеспечивает никирование последовательности двуцепочечной ДНК в не являющейся мишенью цепи, тем самым обеспечивая образование свободной одноцепочечной ДНК, имеющей 3´-конец. После никирования 3'-конец свободной одноцепочечной ДНК гибридизируется с матрицей для синтеза ДНК, тем самым обеспечивая праймирование домена обратной транскриптазы. Затем обратная транскриптаза обеспечивает полимеризацию ДНК с 3'-конца, тем самым обеспечивая образование одноцепочечного ДНК-флэпа, включающего требуемое нуклеотидное изменение. Затем одноцепочечный ДНК-флэп замещает цепь эндогенной ДНК, примыкающую к участку разрыва, тем самым обеспечивая внесение требуемого нуклеотидного изменения в последовательность двуцепочечной ДНК.

[0074] В других вариантах осуществления в изобретении предусмотрен способ внесения одного или нескольких изменений в нуклеотидную последовательность молекулы ДНК в локусе-мишени, включающий контактирование молекулы ДНК с программируемым

нуклеиновой кислотой ДНК-связывающим белком (napDNAbp) и направляющей РНК, которая нацеливает napDNAbp на локус-мишень, при этом направляющая РНК включает последовательность матрицы в виде обратной транскриптазы (RT), включающую, по меньшей мере, одно требуемое нуклеотидное изменение. napDNAbp высвобождает 3'-конец цепи ДНК в локусе-мишени, который гибридизируется с матрицей для синтеза ДНК (например последовательностью RT матрицы) для праймирования обратной транскрипции. Далее одноцепочечный ДНК-флэп, включающий указанное, по меньшей мере, одно требуемое нуклеотидной изменение на основе матрицы для синтеза ДНК (например последовательности RT матрицы), синтезируется или полимеризируется обратной транскриптазой. Наконец, указанное, по меньшей мере, одно требуемое нуклеотидное изменение встраивается в соответствующую эндогенную ДНК, тем самым обеспечивая внесение одного или нескольких изменений в нуклеотидную последовательность молекулы ДНК в локусе-мишени.

[0075] В других вариантах осуществления изобретение предусматривает способ внесения одного или нескольких изменений в нуклеотидную последовательность молекулы ДНК в локусе-мишени с помощью мишень-инициированной обратной транскрипции, при этом способ включает: контактирование молекулы ДНК в локусе-мишени с (i) слитым белком, включающим программируемый нуклеиновой кислотой ДНК-связывающий белок (napDNAbp) и обратную транскриптазу, и (ii) направляющей РНК, включающей RT матрицу, включающую требуемое нуклеотидное изменение (например любую из SEQ ID NO: 1-135514 или 813085-880462); при этом данное контактирование обеспечивает осуществление мишень-инициированной обратной транскрипции RT матрицы для получения одноцепочечной ДНК, включающей требуемое нуклеотидное изменение, и встраивание требуемого нуклеотидного изменения в молекулу ДНК в локусе-мишени с помощью процесса репарации и/или репликации ДНК.

[0076] В некоторых вариантах осуществления этап замещения цепи эндогенной ДНК включает: (i) гибридизацию одноцепочечного ДНК-флэпа с цепью эндогенной ДНК, примыкающей к участку разрыва, для создания несовпадения последовательности; (ii) эксцизию цепи эндогенной ДНК; и (iii) репарацию несовпадения для образования требуемого продукта, включающего требуемое нуклеотидное изменение в обеих цепях ДНК.

[0077] Описанные здесь способы могут включать слитые белки, имеющие napDNAbp, который представляет собой не имеющий нуклеазной активности Cas9 (dCas9), никазу Cas9 (nCas9), или имеющий нуклеазную активность Cas9. В других вариантах осуществления napDNAbp и обратная транскриптаза не кодируются в виде единого слитого белка, но могут поступать в рамках отдельных конструктов. Таким образом, в некоторых вариантах осуществления обратная транскриптаза может поступать *in trans* относительно napDNAbp (а не в виде слитого белка).

[0078] В различных вариантах осуществления, относящихся к способам, napDNAbp может включать аминокислотную последовательность SEQ ID NO: 1361421 (Cas9). napDNAbp может также включать аминокислотную последовательность, которая, по меньшей мере, на 80%, 85%, 90%, 95%, 98%, или 99% идентична аминокислотной последовательности любой одной из SEQ ID NO: 1361421.

[0079] В различных вариантах осуществления, относящихся к способам, обратная транскриптаза может включать любую одну из аминокислотных последовательностей SEQ ID NO: 1361485-1361514, и 1361597-1361598. Обратная транскриптаза может также включать аминокислотную последовательность, которая, по меньшей мере, на 80%, 85%, 90%, 95%, 98%, или 99% идентична аминокислотной последовательности любой одной из SEQ ID NO: 1361485-1361514, и 1361597-1361598.

[0080] Способы могут включать применение удлиненной РНК, имеющей нуклеотидную последовательность SEQ ID NO: 271029 – 406542 и 947841-1015218, или нуклеотидную последовательность, имеющую, по меньшей мере, 80%, или, по меньшей мере, 85%, или, по меньшей мере, 90%, или, по меньшей мере, 95%, или, по меньшей мере, 99% идентичность последовательности относительно них.

[0081] Способы могут включать применение направляющих РНК праймерного редактора, которые включают удлинение РНК на 3'-конце, при этом удлинение РНК включает матрицу для синтеза ДНК, например PEgRNA, показанная на Фиг.3B (со следующими компонентами, как описано, в направлении от 5' к 3': спейсером; ядром нРНК, матрицей обратной транскрипции; участком связывания праймера), имеет плечо удлинения, включающее в направлении от 5' к 3' матрицу обратной транскрипции и участок связывания праймера.

[0082] Способы могут включать применение направляющих РНК праймерного редактора, которые включают удлинение РНК на 5'-конце, при этом удлинение РНК включает матрицу для синтеза ДНК, например PEgRNA, показанная на Фиг.3А (со следующими компонентами, как описано, в направлении от 5' к 3': матрицей обратной транскрипции; участком связывания праймера; линкером; спейсером; ядром нРНК), имеет плечо удлинения, включающее в направлении от 5' к 3' матрицу обратной транскрипции, участок связывания праймера, и линкер длиной 5-20 нуклеотидов.

[0083] Способы могут включать применение направляющих РНК праймерного редактора, включающих удлинение РНК во внутримолекулярном расположении в направляющей РНК, при этом удлинение РНК включает матрицу для синтеза ДНК.

[0084] Способы могут включать применение направляющих РНК праймерного редактора, имеющих одно или несколько удлинений РНК длиной, по меньшей мере, 1, по меньшей мере, 2, по меньшей мере, 3, по меньшей мере, 4, по меньшей мере, 5, по меньшей мере, 6, по меньшей мере, 7, по меньшей мере, 8, по меньшей мере, 9, по меньшей мере, 10, по меньшей мере, 11, по меньшей мере, 12, по меньшей мере, 13, по меньшей мере, 14, по меньшей мере, 15, по меньшей мере, 16, по меньшей мере, 17, по меньшей мере, 18, по меньшей мере, 19, по меньшей мере, 20, по меньшей мере, 30, по меньшей мере, 40, по меньшей мере, 50, по меньшей мере, 60, по меньшей мере, 70, по меньшей мере, 80, по меньшей мере, 90, по меньшей мере, 100, по меньшей мере, 200, по меньшей мере, 300, по меньшей мере, 400, или, по меньшей мере, 500 нуклеотидов.

[0085] Следует понимать, что вышеуказанные концепции и дополнительные концепции, которые обсуждаются ниже, могут быть сгруппированы в рамках любой подходящей комбинации, поскольку настоящее изобретение не является ограниченным в данном отношении. Дополнительно, другие преимущества и новые признаки настоящего изобретения станут очевидны на основе нижеследующего подробного описания различных неограничивающих вариантов осуществления при их рассмотрении совместно с прилагаемыми фигурами.

ОПИСАНИЕ ФИГУР

[0086] Нижеследующие фигуры образуют часть настоящего описания и включены для дополнительной демонстрации определенных аспектов настоящего изобретения, которое может быть лучше понятно со ссылкой на одну или несколько из данных фигур в

комбинации с подробным описанием конкретных приведенных здесь вариантов осуществления.

[0087] На **Фиг.1А.1** приведена схема примера процесса внесения изменения одного нуклеотида, вставки и/или делеции в молекулу ДНК (например геном) с помощью слитого белка, включающего обратную транскриптазу, слитую с белком napDNAbp (например Cas9), в комплексе с направляющей РНК праймерного редактора. В данном варианте осуществления направляющая РНК удлиняется на 3′-конце для включения матрицы для синтеза ДНК. Схема показывает, как обратная транскриптаза (RT), слитая с никазой Cas9, в комплексе с направляющей РНК (нРНК) связывается с участком-мишенью ДНК и осуществляет разрыв PAM-содержащей цепи ДНК, примыкающей к нуклеотиду-мишени. RT матрица использует никированную ДНК в качестве праймера для синтеза ДНК из нРНК, которая используется в качестве матрицы для синтеза новой цепи ДНК, которая кодирует требуемое изменение. Показанный процесс редактирования может быть обозначен как редактирование на основе мишень-инициированной обратной транскрипции (праймерное редактирование). На **Фиг.1А.2** показано то же, что и на Фиг.1А.1, за исключением того, что комплекс праймерного редактора обозначен более общим образом как [napDNAbp]-[P]:PEgRNAPEgRNA или [P]-[napDNAbp]:PEgRNAPEgRNA, где "P" обозначает любую полимеразу (например обратную транскриптазу), "napDNAbp" обозначает программируемый нуклеиновой кислотой ДНК-связывающий белок (например SpCas9), а "PEgRNAPEgRNA" обозначает направляющую РНК для праймерного редактирования, а "]-[" обозначает необязательный линкер. Как описано в других разделах, например на Фиг.3А-3G, PEgRNAPEgRNA включает 5′ плечо удлинения, включающее участок связывания праймера и матрицу для синтеза ДНК. Несмотря на то, что это не показано, подразумевается, что плечо удлинения PEgRNAPEgRNA (то есть которое включает участок связывания праймера и матрицу для синтеза ДНК) может представлять собой ДНК или РНК. Конкретная полимераза, предполагаемая в данной конфигурации, будет зависеть от природы матрицы для синтеза ДНК. Например, если матрица для синтеза ДНК представляет собой РНК, то полимераза будет представлять собой РНК-зависимую ДНК-полимеразу (например обратную транскриптазу). Если матрица для синтеза ДНК представляет собой ДНК, то полимераза может представлять собой ДНК-зависимую ДНК-полимеразу. В различных вариантах

осуществления PEgRNA может быть сконструирована или синтезирована для включения матрицы для синтеза ДНК на основе ДНК.

[0088] На **Фиг.1B.1** приведена схема примера процесса внесения изменения одного нуклеотида, вставки и/или делеции в молекулу ДНК (например геном) с помощью слитого белка, включающего обратную транскриптазу, слитую с белком napDNAbp (например Cas9), в комплексе с направляющей РНК праймерного редактора. В данном варианте осуществления направляющая РНК удлиняется на 5′-конце для включения матрицы для синтеза ДНК. Схема показывает, как обратная транскриптаза (RT), слитая с никазой Cas9, в комплексе с направляющей РНК (нРНК) связывается с участком-мишенью ДНК и осуществляет разрыв PAM-содержащей цепи ДНК, примыкающей к нуклеотиду-мишени. Стандартная последовательность PAM представляет собой 5′-NGG-3′, однако различные последовательности PAM могут быть ассоциированы с различными белками Cas9 или эквивалентными белками из различных организмов. Кроме того, любая нуклеаза Cas9, например SpCas9, может быть модифицирована для изменения специфичности белка в отношении PAM для распознавания альтернативной последовательности PAM. Фермент RT использует никированную ДНК в качестве праймера для синтеза ДНК из нРНК, которая используется в качестве матрицы для синтеза новой цепи ДНК, которая кодирует требуемое изменение. Показанный процесс редактирования может быть обозначен как редактирование на основе мишень-инициированной обратной транскрипции (TPRT редактор или праймерный редактор). На **Фиг.1B.2** показано то же, что и на **Фиг.1B.1**, за исключением того, что комплекс праймерного редактора обозначен более общим образом как [napDNAbp]-[P]:PEgRNAPEgRNA или [P]-[napDNAbp]:PEgRNAPEgRNA, где "P" обозначает любую полимеразу (например обратную транскриптазу), "napDNAbp" обозначает программируемый нуклеиновой кислотой ДНК-связывающий белок (например SpCas9), а "PEgRNAPEgRNA" обозначает направляющую РНК для праймерного редактирования, а "]-[" обозначает необязательный линкер. Как описано в других разделах, например на Фиг.3A-3G, PEgRNAPEgRNA включает 3′ плечо удлинения, включающее участок связывания праймера и матрицу для синтеза ДНК. Несмотря на то, что это не показано, подразумевается, что плечо удлинения PEgRNAPEgRNA (то есть которое включает участок связывания праймера и матрицу для синтеза ДНК) может представлять собой ДНК или РНК. Конкретная полимераза, предполагаемая в данной

конфигурации, будет зависеть от природы матрицы для синтеза ДНК. Например, если матрица для синтеза ДНК представляет собой РНК, то полимераза будет представлять собой РНК-зависимую ДНК-полимеразу (например обратную транскриптазу). Если матрица для синтеза ДНК представляет собой ДНК, то полимераза может представлять собой ДНК-зависимую ДНК-полимеразу.

[0089] **Фиг.1С** представляет собой схему, на которой показан пример процесса того как синтезированная одиночная цепь ДНК (которая включает требуемое нуклеотидное изменение) подвергается коррекции, так что требуемое нуклеотидное изменение, вставка и/или делеция вносится в ДНК. Как показано, после синтеза подвергнутой редактированию цепи (или "мутагенной цепи") выравнивание с эндогенной цепью, расщепление флэпа эндогенной цепи и лигирование ведет к встраиванию изменения ДНК после коррекция несовпадающего ДНК дуплекса за счет процессов репарации и/или репликации эндогенной ДНК.

[0090] **Фиг.1D** представляет собой схему, показывающую, что "никирование противоположной цепи" может быть встроено в способ коррекции в соответствии с Фиг.1С для стимулирования образования требуемого продукта вместо продукта, связанного с возвратом. При никировании противоположной цепи второй комплекс napDNAbp/нРНК (например комплекс Cas9/нРНК) используется для внесения второго разрыва в противоположной цепи относительно исходной подвергнутой никированию цепи. Это индуцирует процессы репарации и/или репликации эндогенной клеточной ДНК для предпочтительного замещения не подвергавшейся редактированию цепи (то есть цепи, содержащей второй подвергнутый разрыву участок).

[0091] На **Фиг.1Е** показана другая схема примера процесса внесения, по меньшей мере, одного нуклеотидного изменения (например 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 или более), вставки и/или делеции в молекулу ДНК (например геном) локуса-мишени с помощью программируемого нуклеиновой кислотой ДНК-связывающего белка (napDNAbp) в виде комплекса с направляющей РНК праймерного редактора (например в рамках праймерного редактирования). Направляющая РНК праймерного редактора включает удлинение на 3′- или 5′-конце направляющей РНК или во внутримолекулярном положении в направляющей РНК. На этапе (а) комплекс napDNAbp/нРНК контактирует с молекулой ДНК, при этом нРНК направляет napDNAbp для связывания с локусом-мишенью. На

этапе (b) осуществляется разрыв в одной из цепей ДНК (цепи R-петли, или PAM-содержащей цепи, или не являющейся мишенью цепи ДНК, или протоспейсерной цепи) локуса-мишени (например с помощью нуклеазы или химического агента), тем самым обеспечивая создание свободного 3′-конца в одной из цепей локуса-мишени. В определенных вариантах осуществления разрыв создается в цепи ДНК, которая соответствует цепи R-петли, то есть цепи, которая не гибридизируется с последовательностью направляющей РНК. На этапе (c) 3′-концевая цепь ДНК взаимодействует с удлиненной частью направляющей РНК для праймирования обратной транскрипции. В некоторых вариантах осуществления 3′-концевая цепь ДНК гибридизируется с определенным участком связывания праймера на удлиненной части направляющей РНК. На этапе (d) вносится обратная транскриптаза, которая синтезирует одиночную цепь ДНК от 3′-конца праймированного участка к 3′-концу направляющей РНК. Это образует одноцепочечный ДНК-флэп, включающий требуемое нуклеотидное изменение (например изменение (изменения), вставку (вставки), делецию (делеции), или их комбинацию одного или нескольких оснований). На этапе (e) napDNAbp и направляющая РНК высвобождаются. Этапы (f) и (g) связаны с коррекцией одноцепочечного ДНК-флэпа, так что требуемое нуклеотидное изменение становится встроенным в локус-мишень. Данный процесс может быть смещен в сторону образования требуемого продукта путем удаления соответствующего 5′ эндогенного ДНК-флэпа, который образуется когда 3′ одноцепочечный ДНК-флэп вносится и гибридизируется с комплементарной последовательностью на другой цепи. Процесс также может быть смещен в сторону образования продукта с никированием второй цепи, как показано в качестве примера на Фиг.1D. Данный процесс может обеспечивать внесение, по меньшей мере, одного или нескольких из следующих генетических изменений: трансверсий, транзиций, делеций, и вставок.

[0092] Фиг.1F представляет собой схему, на которой показаны типы генетических изменений, которые являются возможными в рамках описанных здесь процессов редактирования на основе мишень-инициированной обратной транскрипции (праймерного редактирования). Типы нуклеотидных изменений, которые могут быть реализованы с помощью праймерного редактирования, выключают делеции (включая короткие и

длинные делеции), изменения одного или нескольких нуклеотидов, и вставки (включая короткие и длинные вставки).

[0093] **Фиг.1G** представляет собой схему, изображающую пример временно́го никирования второй цепи на основе примера в виде комплекса праймерного редактора. Временно́е никирование второй цепи представляет собой вариант никирования второй цепи для реализации образования требуемого подвергнутого редактированию продукта. Термин "временно́е" означает то, что второй разрыв не подвергнутой редактированию цепи осуществляется только после встраивания требуемого изменения в подвергнутую редактированию цепь. Это обеспечивает возможность избежать одновременных разрывов на обеих цепях, которые могут приводить к двуцепочечным разрывам ДНК.

[0094] На **Фиг.1H** показан вариант предполагаемого здесь праймерного редактирования, в рамках которого осуществляется замещение napDNAbp (например никазы SpCas9) любым доменом программируемой нуклеазы, например нуклеазами цинкового пальца (ZFN) или подобными активатору транскрипции эффекторными нуклеазами (TALEN). Таким образом, предполагается, что подходящие нуклеазы не обязательно должны "программироваться" нацеленной на нуклеиновую кислоту молекулой (например направляющей РНК), но вместо этого могут программироваться путем задания специфичности ДНК-связывающего домена, например (и в частности) нуклеазы. Как и при праймерном редактировании с использованием фрагментов napDNAbp, является предпочтительным, чтобы такие альтернативные программируемые нуклеазы были модифицированы таким образом, что только одна цепь ДНК-мишени подвергается разрыву. Другими словами, программируемые нуклеазы предпочтительно должны функционировать в качестве никаз. После выбора программируемой нуклеазы (например ZFN или TALEN) в систему могут быть добавлены дополнительные виды функциональности для обеспечения ее работы в соответствии с подобным праймерному редактированию механизмом. Например, программируемые нуклеазы могут быть модифицированы путем присоединения к ним (например через химический линкер) РНК или ДНК плеча удлинения, при этом плечо удлинения включает участок связывания праймера (PBS) и матрицу для синтеза ДНК. Программируемая нуклеаза также может быть присоединена (например через химический или аминокислотный линкер) к полимеразе, природа которой будет зависеть от того, является ли плечо удлинения ДНК

или РНК. В случае плеча удлинения в виде РНК полимераза может представлять собой РНК-зависимую ДНК-полимеразу (например обратную транскриптазу). В случае плеча удлинения в виде ДНК полимераза может представлять собой ДНК-зависимую ДНК-полимеразу (например прокариотическую полимеразу, включая Pol I, Pol II, или Pol III, или эукариотическую полимеразу, включая Pol a, Pol b, Pol g, Pol d, Pol e, или Pol z). Система может также включать другие виды функциональности, добавленные в виде продуктов слияния с программируемыми нуклеазами или добавленные *in trans* для реализации реакции в целом (например (a) хеликазу для разматывания ДНК в участке разрыва, чтобы сделать цепь с разрывом с 3'-концом доступной в качестве праймера, (b) FEN1 для обеспечения удаления эндогенной цепи на цепи с разрывом для смещения реакции в сторону замещения эндогенной цепи синтезированной цепью, или (c) комплекс nCas9:нРНК для создания второго участка разрыва на противоположной цепи, что может обеспечивать интеграцию синтезированной коррекции за счет предпочтительной клеточной репарации не подвергнутой редактированию цепи). Аналогично праймерному редактированию с использованием napDNAbp, подобный комплекс с иным образом программируемой нуклеазой может использоваться для синтеза и последующего встраивания новой синтезированной замещающей цепи ДНК, несущей представляющее интерес изменение, перманентно в участок-мишень ДНК.

[0095] На **Фиг.1I** в рамках одного варианта осуществления показаны структурные свойства ДНК-мишени, которая может подвергаться редактированию с помощью праймерного редактирования. ДНК-мишень включает "не являющуюся мишенью цепь" и "цепь-мишень". Цепь-мишень представляет собой цепь, которая отжигается на спейсере PEgRNA комплекса праймерного редактора, который распознает участок PAM (в этом случае — NGG, который распознается праймерными редакторами на основе стандартного SpCas9). Цепь-мишень может быть также обозначена как "не включающая PAM цепь" или "не подвергаемая редактированию цепь". С другой стороны, не являющаяся мишенью цепь (то есть цепь, содержащая протоспейсер и PAM-последовательность NGG) может быть обозначена как "включающая PAM цепь" или "подвергаемая редактированию цепь". В различных вариантах осуществления участок разрыва PE комплекса будет располагаться в протоспейсере включающей PAM цепи (например в случае PE на основе SpCas9). Расположение разрыва будет соответствовать конкретному Cas9, который

образует PE. Например, в случае PE на основе SpCas9 участок разрыва в фосфодиэфирной связи располагается между основаниями три (положение "-3" относительно положения 1 последовательности PAM) и четыре (положение "-4" относительно положения 1 последовательности PAM). Участок разрыва в протоспейсере образует свободную 3' гидроксильную группу, которая, как видно на последующих фигурах, образует комплекс с участком связывания праймера на плече удлинения PEgRNA и обеспечивает субстрат для начала полимеризации одиночной цепи кода ДНК с помощью матрицы для синтеза ДНК плеча удлинения PEgRNA. Данная реакция полимеризации катализируется полимеразой (например обратной транскриптазой) слитого белка PE в направлении от 5' до 3'. Полимеризации терминируется до достижения ядра нРНК (например за счет включения сигнала терминации полимеризации, или вторичной структуры, которая функционирует для терминирования полимеризационной активности PE), обеспечивая продуцирование одноцепочечного ДНК-флэпа, который простирается от исходной 3' гидроксильной группы никированной включающей PAM цепи. Матрица для синтеза ДНК кодирует одноцепочечную ДНК, которая гомологична эндогенной 5'-концевой одиночной цепи ДНК, которая немедленно следует за участком разрыва на включающей PAM цепи и обеспечивает встраивание требуемого нуклеотидного изменения (например замещения одного основания, вставки, делеции, инверсии). Положение требуемого изменения может соответствовать любому положению ниже участка разрыва на включающей PAM цепи, которое может включать положение +1, +2, +3, +4 (начало участка PAM), +5 (положение 2 участка PAM), +6 (положение 3 участка PAM), +7, +8, +9, +10, +11, +12, +13, +14, +15, +16, +17, +18, +19, +20, +21, +22, +23, +24, +25, +26, +27, +28, +29, +30, +31, +32, +33, +34, +35, +36, +37, +38, +39, +40, +41, +42, +43, +44, +45, +46, +47, +48, +49, +50, +51, +52, +53, +54, +55, +56, +57, +58, +59, +60, +61, +62, +63, +64, +65, +66, +67, +68, +69, +70, +71, +72, +73, +74, +75, +76, +77, +78, +79, +80, +81, +82, +83, +84, +85, +86, +87, +88, +89, +90, +91, +92, +93, +94, +95, +96, +97, +98, +99, +100, +101, +102, +103, +104, +105, +106, +107, +108, +109, +110, +111, +112, +113, +114, +115, +116, +117, +118, +119, +120, + 121, +122, +123, +124, +125, +126, +127, +128, +129, +130, +131, +132, +133, +134, +135, +136, +137, +138, +139, +140, +141, +142, +143, +144, +145, +146, +147, +148, +149, или +150, или более (относительно положения ниже участка разрыва). При замещении 3'-концевой одноцепочечной ДНК (содержащей

представляющее интерес изменение) эндогенной 5'-концевой одноцепочечной ДНК процессы репарации и репликации ДНК будут приводить к перманентному внесению участка изменения во включающую PAM цепь и затем к коррекции несовпадения в не включающей PAM цепи, которое будет присутствовать в участке редактирования. За счет этого изменение будет присутствовать на обоих цепях ДНК в участке-мишени ДНК. Очевидно, что определения "подвергнутая редактированию цепь" и "не подвергнутая редактированию" цепь служат лишь для обозначения цепей ДНК, затрагиваемых механизмом PE. "Подвергнутая редактированию" цепь представляет собой цепь, которая сначала подвергается редактированию путем замещения 5'-концевой одноцепочечной ДНК непосредственно ниже участка разрыва синтезированной 3'-концевой одноцепочечной ДНК, содержащей требуемое изменение. "Не подвергаемая редактированию" цепь представляет собой парную цепь подвергаемой редактированию цепи, но которая сама по себе также становится отредактированной за счет репарации и/или репликации для сохранения комплементарности относительно подвергнутой редактированию цепи, в частности представляющего интерес изменения.

[0096] На **Фиг.1J** изображен механизм праймерного редактирования, демонстрирующий структурные свойства ДНК-мишени, комплекса праймерного редактора, и взаимодействие между PEgRNA и ДНК-мишенью. Сначала праймерный редактор, включающий слитый белок, имеющий полимеразу (например обратную транскриптазу) и napDNAbp (например никазу SpCas9, например SpCas9, имеющую деактивирующую мутацию в домене нуклеазы HNH (например H840A) или деактивирующую мутацию в домене нуклеазы RuvC (D10A)) образует комплекс с PEgRNA и ДНК, имеющей подвергаемую редактированию ДНК-мишень. PEgRNA включает спейсер, ядро нРНК (также известное как каркас нРНК или основа нРНК) (которое связывается с napDNAbp), и плечо удлинения. Плечо удлинения может быть расположено на 3'-конце, 5'-конце, или где-то в промежуточном месте в молекуле PEgRNA. Как показано, плечо удлинения расположено на 3'-конце PEgRNA. Плечо удлинения включает в направлении от 3' к 5' участок связывания праймера и матрицу для синтеза ДНК (включающую как представляющее интерес изменение, так и участки гомологии (то есть плечи гомологии), которые являются гомологичными 5'-концевой одноцепочечной ДНК, следующей сразу же за участком разрыва включающей PAM цепи. Как показано, при внесении разрыва с созданием, таким

образом, свободной 3' гидроксильной группы непосредственно выше участка разрыва, участок непосредственно выше участка разрыва на включающей PAM цепи отжигается на комплементарной последовательности на 3'-конце плеча удлинения, которая обозначается как "участок связывания праймера", создавая короткий двуцепочечный участок со свободным 3' гидроксильным концом, который образует субстрат для полимеразы комплекса праймерного редактора. Полимераза (например обратная транскриптаза) затем полимеризируется в виде цепи ДНК от 3' карбоксильного конца к концу плеча удлинения. Последовательность одноцепочечной ДНК кодируется матрицей для синтеза ДНК, которая представляет собой часть плеча удлинения (то есть за исключением участка связывания праймера), которая "считывается" полимеразой для синтеза новой ДНК. Данная полимеризация фактически обеспечивает удлинение последовательности исходного 3' карбоксильного конца первоначального участка разрыва. Матрица для синтеза ДНК кодирует одиночную цепь ДНК, которая включает не только требуемое изменение, но также участки, которые являются гомологичными относительно эндогенной одноцепочечной ДНК непосредственно ниже участка разрыва на включающей PAM цепи. Далее копируемая 3'-концевая одноцепочечная ДНК (то есть 3' одноцепочечный ДНК-флэп) вытесняет соответствующую гомологичную эндогенную 5'-концевую одиночную цепь ДНК непосредственно ниже участка разрыва на включающей PAM цепи, образуя промежуточную ДНК, имеющую 5'-концевой одноцепочечный ДНК-флэп, который удаляется клеткой (например с помощью флэп-эндонуклеазы). 3'-концевой одноцепочечный ДНК-флэп, который отжигается на комплементе эндогенного 5'-концевого одноцепочечного ДНК-флэпа, лигируется с эндогенной цепью после удаления 5' ДНК-флэпа. Требуемое изменение в 3'-концевом одноцепочечном ДНК-флэпе, который был подвергнут отжигу и лигированию, образует несовпадение с комплементарной цепью, которое подвергается репарации ДНК и/или раунду репликации, тем самым обеспечивая перманентное внесение требуемого изменение в обе цепи.

[0097] На **Фиг.2** показаны три комплекса Cas, которые будут протестированы, а также их свойства в плане PAM, нРНК, и расщепления ДНК. На фигуре показаны конструкции комплексов, включающих SpCas9, SaCas9, и LbCas12a.

[0098] На **Фиг.3A-3C** показаны конструкции сконструированной 5'-удлиненной нРНК (Фиг.3A), 3'-удлиненной нРНК (Фиг.3B), и внутримолекулярного удлинения (Фиг.3C),

каждая из которых может использоваться для праймерного редактирования. Данные варианты осуществления изображают примеры конструкций матрицы для синтеза ДКН, участка связывания праймера, и последовательности необязательного линкера в удлиненных частях нРНК с 3'-, 5'-, и внутримолекулярными удлинениями, а также конструкцию протоспейсерных и относящихся к ядру участков. Описанный процесс TPRT не ограничивается данными конфигурациями направляющих РНК праймерного редактора.

[0099] На **Фиг.4A-4E** показаны *in vitro* TPRT анализы. Фиг.4A представляет собой схему удлинения на основе нРНК матрицы флуоресцентно меченных субстратов ДНК с помощью фермента RT при анализе на основе электрофореза с использованием полиакриламидного геля (PAGE) продуктов обратной транскриптазы. На Фиг.4B показана TPRT с использованием подвергнутых предварительному никированию субстратов, dCas9, и 5′-удлиненных нРНК с матрицей для редактирования различной длины. На Фиг.4C показана RT реакция с использованием подвергнутых предварительному никированию субстратов ДНК при отсутствии Cas9. На Фиг.4D показана TPRT полных субстратов дцДНК с использованием Cas9(H840A) и 5′-удлиненных нРНК. На Фиг.4E показана 3′-удлиненная нРНК матрица с подвергнутыми предварительному никированию и полными субстратами дцДНК. Все реакции осуществлялись с использованием RT M-MLV.

[0100] На **Фиг.5** показаны *in vitro* анализы с использованием 5′-удлиненных нРНК с матрицами для редактирования различной длины. В данном наборе экспериментов флуоресцентно меченные (Cy5) ДНК-мишени использовали в качестве субстратов и подвергали предварительному никированию. Cas9, использованный в этих экспериментах, представляет собой каталитически мертвый Cas9 (dCas9), а использованная RT представляет собой Superscript III, коммерчески доступную RT, полученную из вируса лейкоза мышей Молони (M-MLV). Комплексы dCas9:нРНК были образованы из очищенных компонентов. Затем добавляли флуоресцентно меченный субстрат ДНК совместно с dNTP и ферментом RT. Спустя 1 час инкубирования при 37 °C продукты реакции анализировали с помощью денатурирующего электрофореза на основе полиакриламидного геля с применением мочевины (PAGE). Изображение на геле

показывает удлинение исходной цепи ДНК до показателей длины, которые соответствуют длине матрицы обратной транскрипции.

[0101] На **Фиг.6** показаны *in vitro* анализы с использованием 5′-удлиненных нРНК с матрицами для синтеза различной длины, аналогичные показанным на Фиг.5. Однако субстраты ДНК не подвергались предварительному никированию в данном наборе экспериментов. Cas9, используемый в этих экспериментах, представляет собой никазу Cas9 (мутант SpyCas9 H840A), использованная RT представляет собой Superscript III, коммерчески доступную RT, полученную из вируса лейкоза мышей Молони (M-MLV). Продукты реакции анализировали с помощью денатурирующего электрофореза на основе полиакриламидного геля с применением мочевины (PAGE). Как показано на геле, никаза эффективно расщепляет цепь ДНК при использовании нРНК (нРНК_0, дорожка 3).

[0102] На **Фиг.7** показано, что 3′-удлинения поддерживают синтез ДНК и не оказывают значительного влияния на активность никазы Cas9. Подвергнутые предварительному никированию субстраты (черная стрелка) практически полностью конвертируются в продукты RT при использовании никазы dCas9 или Cas9 (дорожки 4 и 5). Более чем 50% конверсия в продукт RT (красная стрелка) наблюдается с полными субстратами (дорожка 3). Использовали никазу Cas9 (мутант SpyCas9 H840A), каталитически мертвый Cas9 (dCas9), и Superscript III, коммерчески доступную RT, полученную из вируса лейкоза мышей Молони (M-MLV).

[0103] На **Фиг.8** показаны эксперименты на основе двойного цвета, которые использовались для определения того, осуществляется ли предпочтительным образом RT реакция с нРНК in cis (связывание с тем же комплексом). Два отдельных эксперимента проводили для 5′-удлиненных и 3′-удлиненных нРНК. Продукты анализировали с помощью PAGE. Соотношение продуктов вычисляли как (Cy3cis/Cy3trans) / (Cy5trans/Cy5cis).

[0104] На **Фиг.9A-9D** показана модель в виде флэп-субстрата. На Фиг.9A показан репортер с двойным флуоресцентным белком для флэп-направленного мутагенеза. На Фиг.9B показана репарация стоп-кодона в клетках HEK. На Фиг.9C показаны секвенированные клоны дрожжей после репарации флэпа. На Фиг.9D показано тестирование различных свойств флэпа в клетках человека.

[0105] На **Фиг.10** демонстрируется праймерное редактирование на плазмидных субстратах. Репортерную плазмиду с двойной флуоресценцией конструировали для экспрессии в дрожжах (*S. cerevisiae*). Экспрессия данного конструкта в дрожжах продуцирует только GFP. *In vitro* реакция TRT обеспечивает внесение точечной мутации и трансформирует дрожжи исходной плазмидой или *in vitro* Cas9(H840A) никированной плазмидой. Колонии визуализированы с помощью флуоресцентной визуализации. Показаны дрожжевые плазмидные трансформанты с двойным флуоресцентным белком. Трансформирование исходной плазмидой или *in vitro* Cas9(H840A) никированной плазмидой приводит к получению колоний, экспрессирующих только зеленый GFP. Реакция TRT с 5′-удлиненными или 3′-удлиненными нРНК продуцирует смесь зеленых и желтых колоний. Последние экспрессируют как GFP, так и mCherry. Еще больше желтых колоний наблюдается при использовании 3′-удлиненной нРНК. Также показан положительный контроль, который не содержит стоп-кодон.

[0106] На **Фиг.11** показано праймерное редактирование на плазмидных субстратах аналогично эксперименту на Фиг.10, но вместо внесения точечной мутации в стоп-кодон праймерное редактирование обеспечивает внесение вставки (слева) или делеции (справа) в виде одного нуклеотида, которая корректирует мутацию сдвига рамки считывания и обеспечивает возможность синтеза mCherry ниже. В обоих экспериментах использовались 3′-удлиненные нРНК.

[0107] На **Фиг.12** показаны продукты редактирования в рамках праймерного редактирования на плазмидных субстратах, проанализированные с помощью секвенирования по Сэнгеру. Колонии из TRT трансформантов отдельным образом отбирали и анализировали с помощью секвенирования по Сэнгеру. Точные изменения наблюдали на основе секвенирования выбранных колоний. Зеленые колонии содержали плазмиды с исходной последовательность ДНК, в то время как желтые колонии содержали точную мутацию, осуществленную с помощью нРНК для праймерного редактирования. Другие точечные мутации или индели не наблюдались.

[0108] На **Фиг.13** показан потенциальный объем новой технологии праймерного редактирования и сравнен с технологиями редактирования оснований на основе дезаминазы.

[0109] На **Фиг.14** показана схема редактирования в клетках человека.

[0110] На **Фиг.15** демонстрируется удлинение участка связывания праймера в нРНК.

[0111] На **Фиг.16** показаны усеченные нРНК для смежного нацеливания.

[0112] **Фиг.17A-17C** представляют собой графики, показывающие % конверсии T в A в нуклеотиде-мишени после трансфекции компонентов в эмбриональные клетки почки человека (HEK). На Фиг.17A показаны данные, которые демонстрируют результаты с использованием N-концевого слияния обратной транскриптазы MLV дикого типа и никазы Cas9(H840A) (32-аминокислотный линкер). Фиг.17B аналогична Фиг.17A, но для C-концевого слияния фермента RT. Фиг.17C аналогична Фиг.17A, но линкер между RT MLV и Cas9 имеет длину 60 аминокислот вместо 32 аминокислот.

[0113] На **Фиг.18** показана высокая чистота редактирования T в A в участке HEK3 на основе высокопроизводительного секвенирования ампликонов. Вывод анализа на основе секвенирования демонстрирует наиболее часто встречающиеся генотипы подвергнутых редактированию клеток.

[0114] На **Фиг.19** показана эффективность редактирования в нуклеотиде-мишени (левый столбец в каждой паре столбцов) наряду с показателями инделей (правый столбец в каждой паре столбцов). WT обозначает фермент RT MLV дикого типа. Мутантные ферменты (от M1 до M4) содержат перечисленные справа мутации. Показатели редактирования количественно анализировали с помощью высокопроизводительного секвенирования ампликонов геномной ДНК.

[0115] На **Фиг.20** показана эффективность редактирования нуклеотида-мишени при внесении одноцепочечного разрыва в цепь комплементарной ДНК поблизости от нуклеотида-мишени. Было протестировано никирование на различном расстоянии от нуклеотида-мишени (оранжевые треугольники). Показана эффективность редактирования в паре азотистых оснований-мишеней (синие столбцы) наряду с показателями образования инделей (оранжевые столбцы). Пример "нет" не содержит никирующей комплементарную цепь направляющей РНК. Показатели редактирования количественно анализировали с помощью высокопроизводительного секвенирования ампликонов геномной ДНК.

[0116] На **Фиг.21** демонстрируются обработанные данные высокопроизводительного секвенирования, показывающие требуемую трансверсивную мутацию T на A и общее отсутствие других значительных побочных продуктов геномного редактирования.

[0117] На **Фиг.22** показана схема примера процесса осуществления направленного мутагенеза с использованием подверженной ошибкам обратной транскриптазы в локусе-мишени с помощью программируемого нуклеиновой кислотой ДНК-связывающего белка (napDNAbp) в виде комплекса с направляющей РНК праймерного редактора. Данный процесс может быть обозначен как вариант осуществления праймерного редактирования для направленного мутагенеза. Направляющая РНК праймерного редактора включает удлинение на 3′- или 5′-конце направляющей РНК или во внутримолекулярном положении в направляющей РНК. На этапе (a) комплекс napDNAbp/нРНК контактирует с молекулой ДНК, при этом нРНК направляет napDNAbp для связывания с подвергаемым мутагенезу локусом-мишенью. На этапе (b) осуществляется разрыв в одной из цепей ДНК локуса-мишени (например с помощью нуклеазы или химического агента), тем самым обеспечивая создание свободного 3′-конца в одной из цепей локуса-мишени. В определенных вариантах осуществления разрыв создается в цепи ДНК, которая соответствует цепи R-петли, то есть цепи, которая не гибридизируется с последовательностью направляющей РНК. На этапе (c) 3′-концевая цепь ДНК взаимодействует с удлиненной частью направляющей РНК для праймирования обратной транскрипции. В некоторых вариантах осуществления 3′-концевая цепь ДНК гибридизируется с определенным участком связывания праймера на удлиненной части направляющей РНК. На этапе (d) вносится подверженная ошибкам обратная транскриптаза, которая синтезирует подвергнутую мутагенезу одиночную цепь ДНК от 3′-конца праймированного участка к 3′-концу направляющей РНК. Примеры мутаций указаны с помощью звездочки "*". Это обеспечивает образование одноцепочечного ДНК-флэпа, включающего требуемый подвергнутый мутагенезу участок. На этапе (e) napDNAbp и направляющая РНК высвобождаются. Этапы (f) и (g) связаны с коррекцией одноцепочечного ДНК-флэпа (включающего подвергнутый мутагенезу участок), так что требуемый подвергнутый мутагенезу участок становится встроенным в локус-мишень. Данный процесс может быть смещен в сторону образования требуемого продукта путем удаления соответствующего 5′ эндогенного ДНК-флэпа, который образуется когда 3′ одноцепочечный ДНК-флэп вносится и гибридизируется с комплементарной последовательностью на другой цепи. Процесс также может быть смещен в сторону образования продукта с никированием второй цепи, как показано в качестве примера на

Фиг.1D. После процессов репарации и/или репликации эндогенной ДНК мутагенизированный участок становится встроенным в обе цепи ДНК локуса ДНК.

[0118] Фиг.23 представляет собой схему конструкции нРНК для сокращения последовательностей тринуклеотидных повторов и сокращения тринуклеотидных повторов с помощью редактирования генома на основе TPRT. Экспансия тринуклеотидных повторов связана с рядом заболеваний человека, включая болезнь Хантингтона, синдром ломкой Х-хромосомы и атаксию Фридрейха. Наиболее часто встречающийся тринуклеотидный повтор содержит триплеты CAG, хотя также встречаются триплеты GAA (атаксия Фридрейха), и триплеты CGG (синдром ломкой Х-хромосомы). Унаследование предрасположенности к экспансии или приобретение уже экспансированного родительского аллеля повышает вероятность возникновения заболевания. Патогенные экспансии тринуклеотидных повторов гипотетически могут быть скорректированы с помощью праймерного редактирования. Участок выше участка повтора может быть никирован с помощью РНК-направляемой нуклеазы и затем использован для праймирования синтеза новой цепи ДНК, которая содержит здоровое количество повторов (которое зависит от конкретного гена и заболевания). После последовательности повтора добавляется короткий фрагмент гомологии, который соответствует последовательности, примыкающей к другому концу повтора (красная цепь). Внесение новой синтезированной цепи и последующее замещение эндогенной ДНК новым синтезированным флэпом ведет к сокращению повторяющегося аллеля.

[0119] Фиг.24 представляет собой схему, показывающую точную 10-нуклеотидную делецию с помощью праймерного редактирования. Направляющая РНК, нацеленная на локус HEK3, была сконструирована с помощью матрицы обратной транскрипции, которая кодирует 10-нуклеотидную делецию после участка разрыва. Эффективность редактирования в трансфицированных HEK клетках анализировали с помощью секвенирования ампликонов.

[0120] Фиг.25 представляет собой схему, показывающую конструкцию нРНК для генов пептидного мечения в эндогенных геномных локусах и пептидное мечение с помощью геномного редактирования на основе TPRT. Системы мечения FlAsH и ReAsH включают две части: (1) флуорофор-бимышьяковый зонд и (2) генетически кодируемый пептид, содержащий мотив тетрацистеина, например последовательность FLNCCPGCCMEP (SEQ

ID NO: 1361586). При экспрессии в клетках белки, содержащие мотив тетрацистеина, могут подвергаться флуоресцентному окрашиванию флуорофор-мышьяковыми зондами (см. *J. Am. Chem. Soc.*, 2002, 124 (21), стр. 6063–6076. DOI: 10.1021/ja017687n). В рамках системы "sortagging" (на основе опосредованного сортазой лигирования) используются бактериальные ферменты сортазы, которые ковалентно конъюгируют меченые пептидные зонды с белками, содержащими подходящие пептидные субстраты (см. *Nat. Chem. Biol.* 2007 Nov;3(11):707-8. DOI: 10.1038/nchembio.2007.31). FLAG-метка (DYKDDDDK (SEQ ID NO: 1361587)), V5-метка (GKPIPNPLLGLDST (SEQ ID NO: 1361588)), GCN4-метка (EELLSKNYHLENEVARLKK (SEQ ID NO: 1361589)), HA-метка (YPYDVPDYA (SEQ ID NO: 1361590)), и Myc-метка (EQKLISEEDL (SEQ ID NO: 1361591)) обычно используются в качестве эпитопных меток для иммуноанализов. π-клэмп кодирует пептидную последовательность (FCPF (SEQ ID NO: 1361592)), которая может быть помечена пентафтор-ароматическими субстратами (*Nat. Chem.* 2016 Feb;8(2):120-8. doi: 10.1038/nchem.2413).

[0121] На **Фиг.26** показано точное внесение His_6-метки и FLAG-метки в геномную ДНК. С помощью матрицы обратной транскрипции была сконструирована направляющая РНК, нацеленная на локус HEK3, которая кодирует либо вставку 18 nt His-метки, или вставку 24 nt FLAG-метки. Эффективность редактирования в трансфицированных HEK клетках анализировали с помощью секвенирования ампликонов. Следует обратить внимание, что полная 24 nt последовательность FLAG-метки располагается вне наблюдаемого изображения (секвенирование подтвердило осуществление полной и точной вставки).

[0122] На **Фиг.27** приведена структура варианта осуществления предусмотренной здесь PEgRNA, которая может быть сконструирована в соответствии с методологией, определенной в рамках Примера 2. PEgRNA включает три основных компонента в направлении от 5′ к 3′, а именно: спейсер, ядро нРНК, и плечо удлинения на 3′-конце. Плечо удлинения может быть дополнительно разделено на следующие структурные элементы в направлении от 5′ к 3′, а именно: участок связывания праймера (A), матрицу для редактирования (B), и плечо гомологии (C). Кроме того, PEgRNA может включать необязательный 3′-концевой модификаторный участок (e1) и необязательный 5′-концевой модификаторный участок (e2). Кроме того, PEgRNA может включать сигнал терминации транскрипции на 3′-конце PEgRNA (не показан). Данные структурные элементы

дополнительно определены здесь. Изображение структуры PEgRNA не является ограничивающим и охватывает вариации в плане расположения элементов. Например, необязательные модификаторы последовательности (e1) и (e2) могут быть расположены внутри или между любыми из показанных участков без ограничения в виде расположения на 3′- или 5′-концах. PEgRNAPEgRNA может включать в определенных вариантах осуществления вторичную структуру РНК, например, но без ограничения, шпильки, стебли/петли, опорные петли, домены рекрутирования РНК-связывающего белка (например аптамер MS2, который рекрутирует и связывает белок MS2cp). Например, подобные вторичные структуры могут быть расположены в спейсере, ядре нРНК, или плече удлинения, и, в частности, в модификаторных участках e1 и/или e2. Помимо вторичных структур РНК, PEgRNAPEgRNA могут включать (например в модификаторных участках e1 и/или e2) химический линкер или поли(N) линкер или хвост, где "N" может представлять собой любое азотистое основание. В некоторых вариантах осуществления (например как показано на Фиг.72(c)), химический линкер может функционировать для предотвращения обратной транскрипции каркаса или ядра sgРНК. Кроме того, в определенных вариантах осуществления (например см. Фиг.72(c)), плечо удлинения (3) может состоять из РНК или ДНК, и/или может включать один или несколько аналогов азотистых оснований (например которые могут добавлять функциональность, например температурную устойчивость). Кроме того, ориентация плеча удлинения (3) может соответствовать естественному направлению от 5′ к 3′, или оно может быть синтезировано с противоположной ориентаций в направлении от 3′ к 5′ (относительно общей ориентации молекулы PEgRNAPEgRNA). Следует также отметить, что средний специалист в данной области будет способен выбрать подходящую ДНК-полимеразу в зависимости от природы материалов нуклеиновой кислоты плеча удлинения (то есть ДНК или РНК) для применения в рамках праймерного редактирования, которая может быть реализована либо в виде продукта слияния с napDNAbp или поступать *in trans* в виде отдельного фрагмента для синтеза требуемого кодируемого матрицей 3′ одноцепочечного ДНК-флэпа, который включает требуемое изменение. Например, если плечо удлинения представляет собой РНК, то ДНК-полимераза может представлять собой обратную транскриптазу или другую подходящую РНК-зависимую ДНК-полимеразу. Однако если плечо удлинения представляет собой ДНК, то ДНК-полимераза может

представлять собой ДНК-зависимую ДНК-полимеразу. В различных вариантах осуществления поступление ДНК-полимеразы может осуществляться *in trans*, например путем применения домена рекрутирования РНК-белка (например MS2 шпильки, расположенной на PEgRNAPEgRNA (например в участке e1 или e2, или где-либо еще), и белка MS2cp, слитого с ДНК-полимеразой, тем самым обеспечивая солокализацию ДНК-полимеразы с PEgRNAPEgRNA). Следует также отметить, что участок связывания праймера в целом не образует часть матрицы, которая используется ДНК-полимеразой (например обратной транскриптазой) для кодирования полученного 3′ одноцепочечного ДНК-флэпа, который включает требуемое изменение. Таким образом, обозначение "матрица для синтеза ДНК" обозначает участок или часть плеча удлинения (3), которая используется в качестве матрицы ДНК-полимеразой для кодирования требуемого 3′ одноцепочечного ДНК-флэпа, содержащего изменение. В некоторых вариантах осуществления матрица для синтеза ДНК включает "матрицу для редактирования" и "плечо гомологии". В других вариантах осуществления матрица для синтеза ДНК также может включать участок e2 или его часть. Например, если участок e2 включает вторичную структуру, которая вызывает терминацию активности ДНК-полимеразы, то является возможным, что функция ДНК-полимеразы будет терминироваться до фактического кодирования в ДНК какой-либо части участка e2. Также возможно, что некоторая часть или даже весь участок e2 будет кодироваться в ДНК. То, какое количество e2 будет фактически использовано в качестве матрицы, будет зависеть от его структуры и того, нарушает ли эта структура функцию ДНК-полимеразы.

[0123] На **Фиг.28** приведена структура другого варианта осуществления предусмотренной здесь PEgRNA, которая может быть сконструирована в соответствии с методологией, определенной в рамках Примера 2. PEgRNA включает три основных компонента в направлении от 5′ к 3′, а именно: спейсер, ядро нРНК, и плечо удлинения на 3′-конце. Плечо удлинения может быть дополнительно разделено на следующие структурные элементы в направлении от 5′ к 3′, а именно: участок связывания праймера (A), матрицу для редактирования (B), и плечо гомологии (C). Кроме того, PEgRNA может включать необязательный 3′-концевой модификаторный участок (e1) и необязательный 5′-концевой модификаторный участок (e2). Кроме того, PEgRNA может включать сигнал терминации транскрипции на 3′-конце PEgRNA (не показан). Данные структурные элементы

дополнительно определены здесь. Изображение структуры PEgRNA не является ограничивающим и охватывает вариации в плане расположения элементов. Например, необязательные модификаторы последовательности (e1) и (e2) могут быть расположены внутри или между любыми из показанных участков без ограничения в виде расположения на 3′- или 5′-концах. PEgRNAPEgRNA может включать в определенных вариантах осуществления вторичную структуру РНК, например, но без ограничения, шпильки, стебли/петли, опорные петли, домены рекрутирования РНК-связывающего белка (например аптамер MS2, который рекрутирует и связывает белок MS2cp). Данные вторичные структуры могут быть расположены в любом месте в рамках молекулы PEgRNA. Например, подобные вторичные структуры могут быть расположены в спейсере, ядре нРНК, или плече удлинения, и, в частности, в модификаторных участках e1 и/или e2. Помимо вторичных структур РНК, PEgRNAPEgRNA могут включать (например в модификаторных участках e1 и/или e2) химический линкер или поли(N) линкер или хвост, где "N" может представлять собой любое азотистое основание. В некоторых вариантах осуществления (например как показано на Фиг.27), химический линкер может функционировать для предотвращения обратной транскрипции каркаса или ядра sgРНК. Кроме того, в определенных вариантах осуществления (например см. Фиг.28), плечо удлинения (3) может состоять из РНК или ДНК, и/или может включать один или несколько аналогов азотистых оснований (например которые могут добавлять функциональность, например температурную устойчивость). Кроме того, ориентация плеча удлинения (3) может соответствовать естественному направлению от 5′ к 3′, или оно может быть синтезировано с противоположной ориентаций в направлении от 3′ к 5′ (относительно общей ориентации молекулы PEgRNAPEgRNA). Следует также отметить, что средний специалист в данной области будет способен выбрать подходящую ДНК-полимеразу в зависимости от природы материалов нуклеиновой кислоты плеча удлинения (то есть ДНК или РНК) для применения в рамках праймерного редактирования, которая может быть реализована либо в виде продукта слияния с napDNAbp или поступать *in trans* в виде отдельного фрагмента для синтеза требуемого кодируемого матрицей 3′ одноцепочечного ДНК-флэпа, который включает требуемое изменение. Например, если плечо удлинения представляет собой РНК, то ДНК-полимераза может представлять собой обратную транскриптазу или другую подходящую РНК-зависимую ДНК-полимеразу.

Однако если плечо удлинения представляет собой ДНК, то ДНК-полимераза может представлять собой ДНК-зависимую ДНК-полимеразу. В различных вариантах осуществления поступление ДНК-полимеразы может осуществляться *in trans*, например путем применения домена рекрутирования РНК-белка (например MS2 шпильки, расположенной на PEgRNAPEgRNA (например в участке e1 или e2, или где-либо еще), и белка MS2cp, слитого с ДНК-полимеразой, тем самым обеспечивая солокализацию ДНК-полимеразы с PEgRNAPEgRNA). Следует также отметить, что участок связывания праймера в целом не образует часть матрицы, которая используется ДНК-полимеразой (например обратной транскриптазой) для кодирования полученного 3′ одноцепочечного ДНК-флэпа, который включает требуемое изменение. Таким образом, обозначение "матрица для синтеза ДНК" обозначает участок или часть плеча удлинения (3), которая используется в качестве матрицы ДНК-полимеразой для кодирования требуемого 3′ одноцепочечного ДНК-флэпа, содержащего изменение. В некоторых вариантах осуществления матрица для синтеза ДНК включает "матрицу для редактирования" и "плечо гомологии". В других вариантах осуществления матрица для синтеза ДНК также может включать участок e2 или его часть. Например, если участок e2 включает вторичную структуру, которая вызывает терминацию активности ДНК-полимеразы, то является возможным, что функция ДНК-полимеразы будет терминироваться до фактического кодирования в ДНК какой-либо части участка e2. Также возможно, что некоторая часть или даже весь участок e2 будет кодироваться в ДНК. То, какое количество e2 будет фактически использовано в качестве матрицы, будет зависеть от его структуры и того, нарушает ли эта структура функцию ДНК-полимеразы.

[0124] Схема на **Фиг.29** показывает взаимодействие стандартной PEgRNA с участком-мишенью двуцепочечной РНК и сопутствующее получение 3′ одноцепочечного ДНК-флэпа, содержащего представляющее интерес генетическое изменение. Двуцепочечная ДНК показана с верхней цепью (то есть цепью-мишенью) в направлении от 3′ к 5′ и нижней цепью (то есть включающей PAM цепью или не являющейся мишенью цепью) в направлении от 5′ к 3′. Верхняя цепь включает комплемент "протоспейсера" и комплемент последовательности PAM и обозначена как "цепь-мишень", поскольку она представляет собой цепь, на которую нацелен спейсер PEgRNA и на котором она отжигается. Комплементарная нижняя цепь обозначена как "не являющаяся мишенью

цепь" или "включающая PAM цепь" или "протоспейсерная цепь", поскольку она содержит последовательность PAM (например NGG) и протоспейсер. Хотя это и не показано, изображенная PEgRNA будет образовывать комплекс с Cas9 или эквивалентным доменом слитого белка праймерного редактора. Как показано на схеме, спейсер PEgRNA отжигается на комплементарном участке протоспейсера на цепи-мишени, который обозначен как протоспейсер, который расположен непосредственно ниже последовательности PAM и имеет длину приблизительно 20 нуклеотидов. Это взаимодействие образует ДНК/РНК гибрид между РНК спейсера и комплементом ДНК протоспейсера и индуцирует образование R-петли в участке напротив протоспейсера. Как описано в других разделах, затем белок Cas9 (не показан) индуцирует разрыв в не являющейся мишенью цепи, как показано. Это затем ведет к образованию участка в виде 3′ оцДНК-флэпа (одноцепочечной ДНК) непосредственно выше участка разрыва, который, в соответствии с *z*, взаимодействует с 3′-концом PEgRNA в участке связывания праймера. 3′-конец оцДНК-флэпа (то есть последовательность праймера в виде обратной транскриптазы) отжигается на участке связывания праймера (A) на PEgRNA, тем самым обеспечивая праймирование обратной транскриптазы. Далее обратная транскриптаза (например поступающая in trans или поступающая cis в виде слитого белка, присоединенного к конструкту Cas9) затем обеспечивает полимеризацию одиночной цепи ДНК, которая кодируется матрицей для синтеза ДНК (включая матрицу для редактирования (B) и плечо гомологии (C)). Полимеризация продолжается по направлению к 5′-концу плеча удлинения. Полимеризованная цепь оцДНК образует 3′-концевой оцДНК-флэп, который, как описано в других разделах (например как показано на Фиг.1E) вносится в эндогенную ДНК, вытесняя соответствующую эндогенную цепь (которая удаляется в виде 5′-концевого ДНК-флэпа эндогенной ДНК) и обеспечивая внесение требуемого нуклеотидного изменения (изменения одной пары азотистых оснований нуклеотида, делеций, вставок (включая целые гены) с помощью раундов естественной репарации/репликации ДНК).

[0125] Фиг.30 служит для облегчения понимания описания PEgRNA в соответствии с Перечнем последовательностей. На фигуре показаны два примера последовательности PEgRNA (SEQ ID NO: 135529 (сверху) и SEQ ID NO: 135880 (снизу)) и расположение на них различных описанных подмножеств последовательностей. Для SEQ ID NO: 135529

соответствующие последовательности представляют собой спейсер (SEQ ID NO: 271043), плечо удлинения (SEQ ID NO: 406557), участок связывания праймера (SEQ ID NO: 542071), матрицу для редактирования (SEQ ID NO: 677585), и плечо гомологии (SEQ ID NO: 813099). Для SEQ ID NO: 135880 соответствующие последовательности представляют собой спейсер (SEQ ID NO: 880463), плечо удлинения (SEQ ID NO: 947841), участок связывания праймера (SEQ ID NO: 1015219), матрицу для редактирования (SEQ ID NO:1082597), и плечо гомологии (SEQ ID NO: 1149975).

[0126] Фиг.31 представляет собой блок-схему, показывающую пример высокоуровневого компьютеризированного способа 3100 для определения структуры удлиненной нРНК в соответствии с некоторыми вариантами осуществления изобретения. На этапе 3102 вычислительное устройство (например вычислительное устройство 3400, описанное на Фиг.34) осуществляет доступ к данным, связанным с входным аллелем, выходным аллелем, и слитым белком, который включает программируемый нуклеиновой кислотой ДНК-связывающий белок и обратную транскриптазу. В то время как этап 3102 описывает доступ ко всем трем из входного аллеля, выходного аллеля, и слитого белка в рамках одного этапа, это приведено для целей иллюстрирования, при этом является очевидным, что доступ к таким данным может осуществляться с помощью одного или нескольких этапов без выхода за рамки объема описанных здесь способов. Осуществление доступа к данным может включать получение данных, хранение данных, доступ к базе данных и/или им подобные.

[0127] Фиг.32 представляет собой блок-схему, показывающую пример компьютеризированного способа 3200 для определения компонентов структуры удлиненной нРНК, включая компоненты удлинения, в соответствии с некоторыми вариантами осуществления. Является очевидным, что Фиг.32 предназначена для иллюстрирования и, таким образом, способы, применяемые для определения удлиненной нРНК, могут включать больше или меньше этапов, чем те, которые показаны на Фиг.32.

[0128] Фиг.33 представляет собой блок-схему, показывающую пример компьютеризированного способа 3300 для определения наборов структур удлиненной нРНК для каждой связанной с мутацией записи в базе данных в соответствии с некоторыми вариантами осуществления изобретения. На этапе 3302 вычислительное устройство осуществляет доступ к базе данных (например базе данных ClinVar, которая

доступна по адресу www.ncbi.nlm.nih.gov/clinvar/), которая включает набор связанных с мутацией записей, каждая из которых включает входной аллель, соответствующий мутации, и выходной аллель, соответствующей скорректированной последовательности дикого типа.

[0129] **Фиг.34** представляет собой пример реализации компьютерной системы 3400, которая может использоваться для реализации любых аспектов описанных здесь способов и вариантов осуществления. Компьютерная система 3400 может включать один или несколько процессоров 3410 и один или несколько энергонезависимых машиночитаемых носителей данных (например память 3420 и один или несколько энергонезависимых носителей данных 3430) и дисплей 3440. Процессор 3410 может управлять записью и чтением данных из памяти 3420 и энергонезависимого носителя данных 3430 любым подходящим образом, поскольку аспекты описанного здесь изобретения не являются ограниченными в данном отношении.

[0130] **Фиг.35A** представляет собой схему основанной на PE вставки последовательностей, кодирующих мотивы РНК, в рамках Примера 3.

[0131] **Фиг.35B** представляет собой перечень (не исчерпывающий) примеров некоторых мотивов, которые потенциально могут быть вставлены, и их функций, в рамках Примера 3.

[0132] На **Фиг.36** приведена гистограмма для сравнения эффективности (то есть "% общих считываний секвенирования с указанным изменением или инделями") PE2, PE2-trunc, PE3, и PE3-trunc в различных участках-мишенях в различных линиях клеток. Данные показывают, что праймерные редакторы, включающие усеченные варианты RT, были примерно настолько же эффективны, что и праймерные редакторы, включающие неусеченные RT белки.

[0133] На **Фиг.37A** показана нуклеотидная последовательность молекулы PEgRNA SpCas9 (сверху), которая терминируется на 3′-конце в "UGU" и не содержит элемент в виде опорной петли. Нижняя часть фигуры изображает ту же молекулу PEgRNA SpCas9, но которая дополнительно модифицирована для включения элемента в виде опорной петли, имеющей последовательность 5′-"GAAANNNNN"-3′, вставленную непосредственно перед "UUU" 3′-концом. N может представлять собой любое азотистое основание.

[0134] На **Фиг.37B** показаны результаты Примера 4, который демонстрирует, что эффективность праймерного редактирования в клетках HEK или клетках EMX повышается с применением PEgRNA, содержащей элементы в виде опорной петли, в то время как процентная величина образования инделей в значительной степени не меняется.

[0135] На **Фиг.38** изображен один вариант осуществления праймерного редактора, представленного в виде двух половин белка PE, которые соединяются в виде цельного праймерного редактора за счет аутосплайсингового действия половин разделенного интеина, расположенных в конце или начале каждой из половин белка праймерного редактора.

[0136] На **Фиг.39** изображен механизм удаления интеина из полипептидной последовательности и восстановление пептидной связи между последовательностями N-концевого и C-концевого экстеина. (a) изображен общий механизм двух половин белка, каждая из которых содержит половину последовательности интеина, которые при контакте в клетке приводят к образованию полностью функционального интеина, который затем подвергается аутосплайсингу и эксцизии. Процесс эксцизии приводит к образованию пептидной связи между N-концевой половиной белка (или "N-экстеином") и C-концевой половиной белка (или "C-экстеином") для образования цельного единого полипептида, включающего части в виде N-экстеина и C-экстеина. В различных вариантах осуществления N-экстеин может соответствовать N-концевой половине разделенного слитого белка праймерного редактора, а C-экстеин может соответствовать C-концевой половине разделенного праймерного редактора. (b) показан химический механизм эксцизии интеина и восстановление пептидной связи, которая соединяет половину в виде N-экстеина (обозначенная красным половина) и половину в виде C-экстеина (обозначенная синим половина) Эксцизия разделенных интеинов (то есть N-интеина и C-интеина в рамках конфигурации с разделенным интеином) также может быть обозначена как "транс-сплайсинг", поскольку она включает действие в виде сплайсинга двух отдельных компонентов, поступающих *in trans*.

ОПРЕДЕЛЕНИЯ

Антисмысловая цепь

[0137] В генетике "антисмысловая" цепь сегмента в рамках двуцепочечной ДНК представляет собой матричную цепь, которая, как считается, проходит в направлении от 3' к 5'. С другой стороны, "смысловая" цепь представляет собой сегмент в рамках двуцепочечной ДНК, которая проходит от 5' к 3' и которая является комплементарной антисмысловой цепи ДНК, или матричной цепи, которая проходит от 3' к 5'. В случае сегмента ДНК, который кодирует белок, смысловая цепь представляет собой цепь ДНК, которая имеет ту же последовательность, что и иРНК, которая использует антисмысловую цепь в качестве ее матрицы при транскрипции и в конечном итоге (как правило, но не всегда) подвергается трансляции в белок. Антисмысловая цепь, таким образом, отвечает за РНК, которая позже транслируется в белок, в то время как смысловая цепь имеет практически такой же состав, что и иРНК. Следует обратить внимание, что для каждого сегмента дцДНК вероятно будет присутствовать два набора смысловых и антисмысловых цепей в зависимости от направления чтения (поскольку смысл и антисмысл зависят от перспективы). В конечном итоге то, какая цепь одного сегмента дцДНК считается смысловой или антисмысловой, определяется генным продуктом или иРНК.

Cas9

[0138] Термин "Cas9" или "нуклеаза Cas9" обозначает РНК-направляемую нуклеазу, включающую домен Cas9 или его фрагмент (например белок, включающий активный или неактивный домен расщепления ДНК Cas9, и/или домен связывания нРНК Cas9). Термин "домен Cas9" в соответствии с используемым здесь значением обозначает фрагмент белка, включающий активный или неактивный домен расщепления Cas9 и/или домен связывания нНРК Cas9. Белок Cas9 представляет собой полноразмерный белок Cas9. Нуклеазу Cas9 также иногда обозначают как нуклеаза casn1 или CRISPR-ассоциированная нуклеаза (Короткие Регулярно Расположенные Группами Палиндромные Повторы, Clustered Regularly Interspaced Short Palindromic Repeat). CRISPR представляет собой адаптивную иммунную систему, которая обеспечивает защиту против подвижных генетических элементов (вирусов, элементов с непостоянной локализацией, и конъюгативных плазмид). Кластеры CRISPR содержат спейсеры, которые представляют собой последовательности, комплементарные предшествующим подвижным элементам, атакующие мишени нуклеиновые кислоты. Кластеры CRISPR транскрибируются и процессируются в РНК CRISPR (crРНК). В системах CRISPR типа II корректное процессирование пре-crРНК

требует транс-кодируемую малую РНК (tracrРНК), эндогенную рибонуклеазу 3 (rnc) и домен Cas9. tracrРНК служит в качестве направляющей для осуществляемого с помощью рибонуклеазы 3 процессирования пре-crРНК. Далее Cas9/crРНК/tracrРНК обеспечивает расщепление эндонуклеазой линейной или циркулярной дцДНК-мишени, являющейся комплементарной спейсеру. Цепь-мишень, которая не является комплементарной crРНК, сначала разрезается эндонуклеазой, затем подрезается 3′-5′ экзонуклеазой. В природе связывание и расщепление ДНК обычно требует белок и обе РНК. Однако одиночные направляющие РНК ("sgРНК" или просто "нРНК") могут быть сконструированы для включения аспектов как crРНК, так и tracrРНК в виде одиночной РНК. См., например, Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012), полное содержание которого приведено здесь в качестве ссылки. Cas9 распознает короткий мотив в последовательностях повторов CRISPR (PAM или смежный с протоспейсером мотив), чтобы помогать отличать себя от чужого. Последовательности и структуры нуклеазы Cas9 хорошо известны специалистам в данной области (см., например, "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*." Ferretti *et al.*, J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E., Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); “CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.” Deltcheva E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., Nature 471:602-607(2011); и “A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.” Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012), полное содержание которых приведено здесь в качестве ссылки). Ортологи Cas9 были описаны у различных видов, включая, но без ограничения , *S. pyogenes* and *S. thermophilus*. Дополнительные подходящие нуклеазы Cas9 и последовательности будут очевидны специалистам в данной области на основе данного описания, при этом такие нуклеазы Cas9 и последовательности включают последовательности Cas9 из организмов и локусов, описанных в Chylinski, Rhun, and Charpentier, “The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems” (2013) *RNA Biology* 10:5, 726-737; полное содержание которого приведено здесь в качестве ссылки. В некоторых вариантах осуществления

нуклеаза Cas9 включает одну или несколько мутаций, которые частично нарушают или деактивируют домен расщепления ДНК.

[0139] Домен деактивированной нуклеазы Cas9 может взаимозаменяемым образом быть обозначен как белок "dCas9" (что означает "мертвый" в качестве нуклеазы Cas9). Способы получения домена Cas9 (или его фрагмента), имеющего деактивированный домен расщепления ДНК, являются известными (см., например, Jinek *et al.*, Science. 337:816-821(2012); Qi *et al.*, "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression" (2013) Cell. 28;152(5):1173-83, полное содержание каждого из которых приведено здесь в качестве ссылки). Например, известно, что домен расщепления ДНК Cas9 включает два субдомена: субдомен нуклеазы HNH и субдомен RuvC1. Субдомен HNH расщепляет цепь, комплементарную нРНК, в то время как субдомен RuvC1 расщепляет некомплементарную цепь. Мутации в данных субдоменах могут заглушать нуклеазную активность Cas9. Например, мутации D10A и H840A полностью деактивируют нуклеазную активность Cas9 *S. pyogenes* (Jinek et al., *Science*. 337:816-821(2012); Qi et al., *Cell.* 28;152(5):1173-83 (2013)). В некоторых вариантах осуществления представлены белки, включающие фрагменты Cas9. Например, в некоторых вариантах осуществления белок включает один из двух доменов Cas9: (1) нРНК-связывающий домен Cas9; или (2) домен расщепления ДНК Cas9. В некоторых вариантах осуществления белки, включающие Cas9 или его фрагменты, обозначаются как "варианты Cas9". Вариант Cas9 имеют такую же гомологию, что и Cas9 или его фрагмент. Например, вариант Cas9 является, по меньшей мере, приблизительно на 70% идентичным, по меньшей мере, приблизительно на 80% идентичным, по меньшей мере, приблизительно на 90% идентичным, по меньшей мере, приблизительно на 95% идентичным, по меньшей мере, приблизительно на 96% идентичным, по меньшей мере, приблизительно на 97% идентичным, по меньшей мере, приблизительно на 98% идентичным, по меньшей мере, приблизительно на 99% идентичным, по меньшей мере, приблизительно на 99,5% идентичным, по меньшей мере, приблизительно на 99,8% идентичным, или, по меньшей мере, приблизительно на 99,9% идентичным Cas9 дикого типа (например SpCas9 согласно SEQ ID NO: 1361421). В некоторых вариантах осуществления вариант Cas9 может иметь 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 или больше

аминокислотных изменений по сравнению с Cas9 дикого типа (например SpCas9 согласно SEQ ID NO: 1361421). В некоторых вариантах осуществления вариант Cas9 включает фрагмент Cas9 SEQ ID NO: X (например нРНК-связывающий домен или домен расщепления ДНК), так что фрагмент является, по меньшей мере, приблизительно на 70% идентичным, по меньшей мере, приблизительно на 80% идентичным, по меньшей мере, приблизительно на 90% идентичным, по меньшей мере, приблизительно на 95% идентичным, по меньшей мере, приблизительно на 96% идентичным, по меньшей мере, приблизительно на 97% идентичным, по меньшей мере, приблизительно на 98% идентичным, по меньшей мере, приблизительно на 99% идентичным, по меньшей мере, приблизительно на 99,5% идентичным, или, по меньшей мере, приблизительно на 99,9% идентичным соответствующему фрагменту Cas9 дикого типа (например SpCas9 согласно SEQ ID NO: 1361421). В некоторых вариантах осуществления фрагмент имеет длину, соответствующую, по меньшей мере, на 30%, по меньшей мере, на 35%, по меньшей мере, на 40%, по меньшей мере, на 45%, по меньшей мере, на 50%, по меньшей мере, на 55%, по меньшей мере, на 60%, по меньшей мере, на 65%, по меньшей мере, на 70%, по меньшей мере, на 75%, по меньшей мере, на 80%, по меньшей мере, на 85%, по меньшей мере, на 90%, по меньшей мере, на 95%, по меньшей мере, на 96%, по меньшей мере, на 97%, по меньшей мере, на 98%, по меньшей мере, на 99%, или, по меньшей мере, на 99,5% длине аминокислоты соответствующего Cas9 дикого типа (например SpCas9 согласно SEQ ID NO: 1361421).

кДНК

[0140] Термин "кДНК" обозначает цепь ДНК, скопированную из РНК матрицы. кДНК является комплементарной РНК матрице.

Циркулярный пермутант

[0141] В соответствии с используемым здесь значением термин "циркулярный пермутант" обозначает белок или полипептид (например Cas9), включающий циркулярную пермутацию, которая представляет собой изменение в структурной конфигурации белка, включающую изменение в порядке аминокислот, следующих в рамках аминокислотной последовательности белка. Другими словами, циркулярные пермутанты представляют собой белки, которые имеют измененные N- и C-концы по сравнению с аналогами дикого типа, например C-концевая половина белка дикого типа становится новой N-концевой

половиной. Циркулярная пермутация (или CP) по существу представляет собой топологическую перестройку первичной последовательности белка, соединяющей его N- и C-конец, зачастую с пептидным линкером, в то же время одновременно обеспечивая разделение его последовательности в другом положении для создания новых смежных N- и C-концов. Результатом является структура белка с иной связностью, но которая зачастую может иметь ту же общую аналогичную трехмерную (3D) форму и возможно включать улучшенные или измененные характеристики, включая сниженную протеолитическую чувствительность, повышенную каталитическую активность, измененное связывание субстрата или лиганда, и/или улучшенную термостабильность. Белки в виде циркулярных пермутантов могут возникать в природе (например конканавалин A и лектин). Кроме того, циркулярная пермутация может возникать в результате посттрансляционных модификаций или может являться результатом модификации с использованием рекомбинантных способов.

Циркулярно пермутированный Cas9

[0142] Термин "циркулярно пермутированный Cas9" обозначает любой белок Cas9 или его вариант в виде циркулярного пермутанта, если его N- и C-концы были перестроены за счет перестройки первичной последовательности белка. Подобные циркулярно пермутированные белки Cas9 ("CP-Cas9") или их варианты сохраняют способность связывать ДНК при образовании комплексов с направляющей РНК (нРНК). См. Oakes *et al.*, "Protein Engineering of Cas9 for enhanced function," *Methods Enzymol*, 2014, 546: 491–511 и Oakes *et al.*, "CRISPR-Cas9 Circular Permutants as Programmable Scaffolds for Genome Modification," *Cell*, January 10, 2019, 176: 254-267, каждый из которых приведен здесь в качестве ссылки. Настоящее изобретение охватывает любые ранее известные CP-Cas9 или предусматривает использование нового CP-Cas9 при условии, если полученный циркулярно пермутированный белок сохраняет способность связывать ДНК при образовании комплексов с направляющей РНК (нРНК). Примеры белков CP-Cas9 приведены в SEQ ID NO: 1361475-1361484.

Матрица для синтеза ДНК

[0143] В соответствии с используемым здесь значением термин "матрица для синтеза ДНК" обозначает участок или часть плеча удлинения PEgRNA, которая используется в качестве матричной цепи полимеразой праймерного редактора для кодирования 3′

одноцепочечного ДНК-флэпа, который содержит требуемое изменение и который затем за счет механизма праймерного редактирования замещает соответствующую эндогенную цепь ДНК в участке-мишени. В различных вариантах осуществления матрица для синтеза ДНК показана на Фиг.3A (в контексте PEgRNA, включающей 5′ плечо удлинения), Фиг.3B (в контексте PEgRNA, включающей 3′ плечо удлинения), Фиг.3C (в контексте внутреннего плеча удлинения), Фиг.3D (в контексте 3′ плеча удлинения), и Фиг.3E (в контексте 5′ плеча удлинения). Плечо удлинения, включая матрицу для синтеза ДНК, может состоять из ДНК или РНК. В случае плеча удлинения в виде РНК полимераза праймерного редактора может представлять собой РНК-зависимую ДНК-полимеразу (например обратную транскриптазу). В случае плеча удлинения в виде ДНК, полимераза праймерного редактора может представлять собой ДНК-зависимую ДНК-полимеразу. В различных вариантах осуществления (например как показано на Фиг.3D-3E), матрица для синтеза ДНК (4) может включать "матрицу для редактирования" и "плечо гомологии", а также весь или часть необязательного 5'-концевого модификаторного участка e2. То есть в зависимости от природы участка e2 (например того, включает ли он шпильку, опорную петлю, или вторичную структуру в виде стебля/петли), полимераза может также кодировать ни одну часть, некоторую часть, или весь участок e2. Другими словами, в случае 3′ плеча удлинения матрица для синтеза ДНК (3) может включать часть плеча удлинения (3), которая проходит от 5′-конца участка связывания праймера (PBS) до 3′-конца ядра нРНК, которая может функционировать в качестве матрицы для синтеза одноцепочечной ДНК полимеразой (например обратной транскриптазой). В случае 5′ плеча удлинения матрица для синтеза ДНК (3) может включать часть плеча удлинения (3), которая проходит от 5′-конца молекулы PEgRNA до 3′-конца матрицы для редактирования. Матрица для синтеза ДНК предпочтительно исключает участок связывания праймера (PBS) PEgRNA, имеющих 3′ плечо удлинения или 5′ плечо удлинения. Определенные описанные здесь варианты осуществления (например Фиг.71A) ссылаются на "RT матрицу", которая включает матрицу для редактирования и плечо гомологии, то есть последовательность плеча удлинения PEgRNA, которая фактически используется в качестве матрицы при синтезе ДНК. Термин "RT матрица" эквивалентен термину "матрица для синтеза ДНК".

Ниже

[0144] В соответствии с используемым здесь значением термины "выше" и "ниже" представляют собой термины для указания относительного положения, которые определяют линейное положение, по меньшей мере, двух элементов, расположенных в молекуле нуклеиновой кислоты (одноцепочечной или двуцепочечной), которая ориентирована в направлении от 5′ к 3′. В частности, первый элемент расположен выше второго элемента в молекуле нуклеиновой кислоты, если первый элемент расположен где-либо 5' относительно второго элемента. Например, SNP расположен выше Cas9-индуцированного участка разрыва, если SNP расположен на 5' стороне участка разрыва. С другой стороны, первый элемент расположен ниже второго элемента в молекуле нуклеиновой кислоты, если первый элемент расположен где-либо 3' относительно второго элемента. Например, SNP расположен ниже Cas9-индуцированного участка разрыва, если SNP расположен на 3' стороне участка разрыва. Молекула нуклеиновой кислоты может представлять собой ДНК (двуцепочечную или одноцепочечную). РНК (двуцепочечную или одноцепочечную), или гибрид ДНК и РНК. Анализ является таким же для одноцепочечной молекулы нуклеиновой кислоты, что и для двуцепочечной молекулы, поскольку термины "выше" и "ниже" обозначают только одну цепь молекулы нуклеиновой кислоты, за исключением случаев, когда необходимо выбрать указываемую цепь двуцепочечной молекулы. Зачастую цепь двуцепочечной ДНК, которая может применяться для определения относительности положения, по меньшей мере, двух элементов, представляет собой "смысловую" или "кодирующую" цепь. В генетике "смысловая" цепь представляет собой сегмент в рамках двуцепочечной ДНК, которая проходит от 5' к 3' и которая является комплементарной антисмысловой цепи ДНК, или матричной цепи, которая проходит от 3' к 5'. Таким образом, в качестве примера, азотистое основание SNP находится ниже последовательности промотора в геномной ДНК (которая является двуцепочечной), если азотистое основание SNP располагается на 3' стороне промотора на смысловой или кодирующей цепи.

CRISPR

[0145] CRISPR представляет собой семейство последовательностей ДНК (то есть кластеров CRISPR) в бактериях и археях, которые соответствуют фрагментам, связанным с предшествующим заражением вирусом, который вторгся в прокариоту. Фрагменты ДНК используются прокариотической клеткой для детектирования и устранения ДНК при

последующих атаках аналогичными вирусами и по существу составляет, наряду с массивом CRISPR-ассоциированных белков (включая Cas9 и его гомологи) и CRISPR-ассоциированной РНК, прокариотическую иммунную защитную систему. В природе кластеры CRISPR транскрибируются и процессируются в РНК CRISPR (crРНК). В системах CRISPR определенного типа (например системах CRISPR типа II) корректное процессирование пре-crРНК требует транс-кодируемую малую РНК (tracrРНК), эндогенную рибонуклеазу 3 (rnc) и белок Cas9. tracrРНК служит в качестве направляющей для осуществляемого с помощью рибонуклеазы 3 процессирования пре-crРНК. Далее Cas9/crРНК/tracrРНК обеспечивает расщепление эндонуклеазой линейной или циркулярной дцДНК-мишени, являющейся комплементарной РНК. В частности, цепь-мишень, которая не является комплементарной crРНК, сначала разрезается эндонуклеазой, затем подрезается 3′-5′ экзонуклеазой. В природе связывание и расщепление ДНК обычно требует белок и обе РНК. Однако одиночные направляющие РНК ("sgРНК" или просто "нРНК") могут быть сконструированы для включения аспектов как crРНК, так и tracrРНК в виде одиночной РНК - направляющей РНК. См, например Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012), полное содержание которого приведено здесь в качестве ссылки. Cas9 распознает короткий мотив в последовательностях повторов CRISPR (PAM или смежный с протоспейсером мотив), чтобы помогать отличать себя от чужого. Биология CRISPR, а также последовательности и структуры нуклеазы Cas9 хорошо известны специалистам в данной области (см., например, "Complete genome sequence of an M1 strain of Streptococcus pyogenes." Ferretti *et al.*, J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E., *Proc. Natl. Acad. Sci. U.S.A.* 98:4658-4663(2001); “CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.” Deltcheva E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., *Nature* 471:602-607(2011); и “A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.” Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012), полное содержание каждого из которых приведено здесь в качестве ссылки). Ортологи Cas9 были описаны у различных видов, включая, но без ограничения, *S. pyogenes* и *S.*

thermophilus. Дополнительные подходящие нуклеазы Cas9 и последовательности будут очевидны специалистам в данной области на основе данного описания, при этом такие нуклеазы Cas9 и последовательности включают последовательности Cas9 из организмов и локусов, описанных в Chylinski, Rhun, and Charpentier, “The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems” (2013) RNA Biology 10:5, 726-737; полное содержание которого приведено здесь в качестве ссылки.

[0146] Термин "Cas9" или "нуклеаза Cas9" обозначает РНК-направляемую нуклеазу, включающую домен Cas9 или его фрагмент (например белок, включающий активный или неактивный домен расщепления ДНК Cas9, и/или домен связывания нРНК Cas9). Термин "домен Cas9" в соответствии с используемым здесь значением обозначает фрагмент белка, включающий активный или неактивный домен расщепления Cas9 и/или домен связывания нНРК Cas9. Белок Cas9 представляет собой полноразмерный белок Cas9. Нуклеазу Cas9 также иногда обозначают как нуклеаза casn1 или CRISPR-ассоциированная нуклеаза (Короткие Регулярно Расположенные Группами Палиндромные Повторы, Clustered Regularly Interspaced Short Palindromic Repeat). CRISPR представляет собой адаптивную иммунную систему, которая обеспечивает защиту против подвижных генетических элементов (вирусов, элементов с непостоянной локализацией, и конъюгативных плазмид). Кластеры CRISPR содержат спейсеры, которые представляют собой последовательности, комплементарные предшествующим подвижным элементам, атакующие мишени нуклеиновые кислоты. Кластеры CRISPR транскрибируются и процессируются в РНК CRISPR (crРНК). В системах CRISPR типа II корректное процессирование пре-crРНК требует транс-кодируемую малую РНК (tracrРНК), эндогенную рибонуклеазу 3 (rnc) и домен Cas9. tracrРНК служит в качестве направляющей для осуществляемого с помощью рибонуклеазы 3 процессирования пре-crРНК. Далее Cas9/crРНК/tracrРНК обеспечивает расщепление эндонуклеазой линейной или циркулярной дцДНК-мишени, являющейся комплементарной спейсеру. Цепь-мишень, которая не является комплементарной crРНК, сначала разрезается эндонуклеазой, затем подрезается 3′-5′ экзонуклеазой. В природе связывание и расщепление ДНК обычно требует белок и обе РНК. Однако одиночные направляющие РНК ("sgРНК" или просто "нРНК") могут быть сконструированы для включения аспектов как crРНК, так и tracrРНК в виде одиночной РНК. См., например, Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-

821(2012), полное содержание которого приведено здесь в качестве ссылки. Cas9 распознает короткий мотив в последовательностях повторов CRISPR (PAM или смежный с протоспейсером мотив), чтобы помогать отличать себя от чужого. Последовательности и структуры нуклеазы Cas9 хорошо известны специалистам в данной области (см., например, , "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*." Ferretti *et al.*, J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E., Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., Nature 471:602-607(2011); и "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012), полное содержание каждого из которых приведено здесь в качестве ссылки). Ортологи Cas9 были описаны у различных видов, включая, но без ограничения, *S. pyogenes* и *S. thermophilus*. Дополнительные подходящие нуклеазы Cas9 и последовательности будут очевидны специалистам в данной области на основе данного описания, при этом такие нуклеазы Cas9 и последовательности включают последовательности Cas9 из организмов и локусов, описанных в Chylinski, Rhun, and Charpentier, "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems" (2013) RNA Biology 10:5, 726-737; полное содержание которого приведено здесь в качестве ссылки. В некоторых вариантах осуществления нуклеаза Cas9 включает одну или несколько мутаций, которые частично нарушают или деактивируют домен расщепления ДНК.

[0147] Домен деактивированной нуклеазы Cas9 может взаимозаменяемым образом быть обозначен как белок "dCas9" (что означает "мертвый" в качестве нуклеазы Cas9). Способы получения домена Cas9 (или его фрагмента), имеющего деактивированный домен расщепления ДНК, являются известными (см., например, Jinek et al., Science. 337:816-821(2012); Qi *et al.*, "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression" (2013) *Cell.* 28;152(5):1173-83, полное содержание которого приведено здесь в качеств ссылки). Например, известно, что домен расщепления ДНК Cas9 включает два субдомена: субдомен нуклеазы HNH и субдомен RuvC1. Субдомен

HNH расщепляет цепь, комплементарную нРНК, в то время как субдомен RuvC1 расщепляет некомплементарную цепь. Мутации в данных субдоменах могут заглушать нуклеазную активность Cas9. Например, мутации D10A и H840A полностью деактивируют нуклеазную активность Cas9 *S. pyogenes* (Jinek et al., *Science*. 337:816-821(2012); Qi et al., *Cell*. 28;152(5):1173-83 (2013)). В некоторых вариантах осуществления представлены белки, включающие фрагменты Cas9. Например, в некоторых вариантах осуществления белок включает один из двух доменов Cas9: (1) нРНК-связывающий домен Cas9; или (2) домен расщепления ДНК Cas9. В некоторых вариантах осуществления белки, включающие Cas9 или его фрагменты, обозначаются как "варианты Cas9". Вариант Cas9 имеют такую же гомологию, что и Cas9 или его фрагмент. Например, вариант Cas9 является, по меньшей мере, приблизительно на 70% идентичным, по меньшей мере, приблизительно на 80% идентичным, по меньшей мере, приблизительно на 90% идентичным, по меньшей мере, приблизительно на 95% идентичным, по меньшей мере, приблизительно на 96% идентичным, по меньшей мере, приблизительно на 97% идентичным, по меньшей мере, приблизительно на 98% идентичным, по меньшей мере, приблизительно на 99% идентичным, по меньшей мере, приблизительно на 99,5% идентичным, по меньшей мере, приблизительно на 99,8% идентичным, или, по меньшей мере, приблизительно на 99,9% идентичным Cas9 дикого типа (например SpCas9 согласно SEQ ID NO: 1361421). В некоторых вариантах осуществления вариант Cas9 может иметь 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 или больше аминокислотных изменений по сравнению с Cas9 дикого типа (например SpCas9 согласно SEQ ID NO: 1361421). В некоторых вариантах осуществления вариант Cas9 включает фрагмент SEQ ID NO: 1361421 (например нРНК-связывающий домен или домен расщепления ДНК), так что фрагмент является, по меньшей мере, приблизительно на 70% идентичным, по меньшей мере, приблизительно на 80% идентичным, по меньшей мере, приблизительно на 90% идентичным, по меньшей мере, приблизительно на 95% идентичным, по меньшей мере, приблизительно на 96% идентичным, по меньшей мере, приблизительно на 97% идентичным, по меньшей мере, приблизительно на 98% идентичным, по меньшей мере, приблизительно на 99% идентичным, по меньшей мере, приблизительно на 99,5% идентичным, или, по меньшей мере, приблизительно на 99,9%

идентичным соответствующему фрагменту Cas9 дикого типа (например SpCas9 согласно SEQ ID NO: 1361421). В некоторых вариантах осуществления фрагмент имеет длину, соответствующую, по меньшей мере, на 30%, по меньшей мере, на 35%, по меньшей мере, на 40%, по меньшей мере, на 45%, по меньшей мере, на 50%, по меньшей мере, на 55%, по меньшей мере, на 60%, по меньшей мере, на 65%, по меньшей мере, на 70%, по меньшей мере, на 75%, по меньшей мере, на 80%, по меньшей мере, на 85%, по меньшей мере, на 90%, по меньшей мере, на 95%, по меньшей мере, на 96%, по меньшей мере, на 97%, по меньшей мере, на 98%, по меньшей мере, на 99%, или, по меньшей мере, на 99,5% длине аминокислоты соответствующего Cas9 дикого типа (например SpCas9 согласно SEQ ID NO: 1361421).

Матрица для редактирования

[0148] Термин "матрица для редактирования" обозначает часть плеча удлинения, которая кодирует требуемое изменение в одноцепочечном 3′ ДНК-флэпе, который синтезируется полимеразой, например ДНК-зависимой ДНК-полимеразой, РНК-зависимой ДНК-полимеразой (например обратной транскриптазой). Определенные описанные здесь варианты осуществления (например Фиг.71A) ссылаются на "RT матрицу", которая совместно обозначает как матрицу для редактирования, так и плечо гомологии, то есть последовательность плеча удлинения PEgRNA, которая фактически используется в качестве матрицы при синтезе ДНК. Термин "RT матрица для редактирования" также эквивалентен термину "матрица для синтеза ДНК", однако "RT матрица для редактирования" обозначает использование праймерного редактора, имеющего полимеразу, которая представляет собой обратную транскриптазу, а "матрица для синтеза ДНК" более широким образом обозначает использование праймерного редактора, имеющего любую полимеразу.

Подверженная ошибкам

[0149] В соответствии с используемым здесь значением термин "подверженная ошибкам" обратная транскриптаза (или, более широко, любая полимераза) обозначает обратную транскриптазу (или, более широко, любую полимеразу), которая встречается в природе или которая была получена из другой обратной транскриптазы (например обратной транскриптазы M-MLV дикого типа), которая имеет частоту ошибок, которая меньше, чем частота ошибок обратной транскриптазы M-MLV дикого типа. Сообщается, что частота

ошибок обратной транскриптазы M-MLV дикого типа соответствует диапазону в виде одной ошибки на от 15000 (верхняя граница) до 27000 (нижняя граница). Частота ошибок 1 на 15000 соответствует частоте ошибок 6,7 x 10^{-5}. Частота ошибок 1 на 27000 соответствует частоте ошибок 3,7 x 10^{-5}. См. Boutabout et al. (2001) "DNA synthesis fidelity by the reverse transcriptase of the yeast retrotransposon Ty1," *Nucleic Acids Res* 29(11):2217 - 2222, который приведен здесь в качестве ссылки. Таким образом, для целей данного изобретения термин "подверженная ошибкам" означает RT, которые имеют частоту ошибок в виде более чем одной ошибки на 15000 встраиваний азотистых оснований (6,7 x 10^{-5} или более), например 1 ошибку в 14000 азотистых оснований (7,14 x 10^{-5} или более), 1 ошибку в 13000 азотистых оснований или менее (7,7 x 10^{-5} или более), 1 ошибку в 12000 азотистых оснований или менее (7,7 x 10^{-5} или более), 1 ошибку в 11000 азотистых оснований или менее (9,1 x 10^{-5} или более), 1 ошибку в 10000 азотистых оснований или менее (1 x 10^{-4} или 0,0001 или более), 1 ошибку в 9000 азотистых оснований или менее (0,00011 или более), 1 ошибку в 8000 азотистых оснований или менее (0,00013 или более), 1 ошибку в 7000 азотистых оснований или менее (0,00014 или более), 1 ошибку в 6000 азотистых оснований или менее (0,00016 или более), 1 ошибку в 5000 азотистых оснований или менее (0,0002 или более), 1 ошибку в 4000 азотистых оснований или менее (0,00025 или более), 1 ошибку в 3000 азотистых оснований или менее (0,00033 или более), 1 ошибку в 2000 азотистых оснований или менее (0,00050 или более), или 1 ошибку в 1000 азотистых оснований или менее (0,001 или более), или 1 ошибку в 500 азотистых основаниях или менее (0,002 или более), или 1 ошибку в 250 азотистых основаниях или менее (0,004 или более).

Плечо удлинения

[0150] Термин "плечо удлинения" обозначает компонент нуклеотидной последовательности PEgRNA, который обеспечивает несколько функций, включая участок связывания праймера и матрицу для редактирования для обратной транскриптазы. В других вариантах осуществления, например на Фиг.3D, плечо удлинения расположено на 3′-конце направляющей РНК. В других вариантах осуществления, например на Фиг.3E, плечо удлинения расположено на 5′-конце направляющей РНК. В некоторых вариантах осуществления плечо удлинения также включает плечо гомологии. В различных вариантах осуществления плечо удлинения включает следующие компоненты в

направлении от 5′ к 3′: плечо гомологии, матрицу для редактирования, и участок связывания праймера. Поскольку полимеризационная активность обратной транскриптазы осуществляется в направлении от 5′ к 3′, предпочтительная группировка плеча гомологии, матрицы для редактирования и участка связывания праймера соответствует направлению от 5′ к 3′, так что обратная транскриптаза после праймирования отожженной праймерной последовательностью обеспечивает полимеризацию одиночной цепи ДНК с использованием матрицы для редактирования в качестве цепи комплементарной матрицы. Дополнительные подробности, такие как длина плеча удлинения, описаны в других разделах описания.

[0151] Плечо удлинения также может быть описано как включающее в целом два участка: участок связывания праймера (PBS) и матрицу для синтеза ДНК, как показано, например, на Фиг.3G (сверху). Участок связывания праймера связывается с праймерной последовательностью, которая образуется из цепи эндогенной ДНК участка-мишени, когда он подвергается никированию комплексом праймерного редактора, тем самым высвобождая 3′-конец на эндогенной никированной цепи. Как описано здесь, связывание праймерной последовательности с участком связывания праймера на плече удлинения PEgRNA создает дуплексный участок со свободным 3′-концом (то есть 3′ праймерной последовательности), который затем обеспечивает субстрат для полимеразы для начала полимеризации одиночной цепи ДНК из свободного 3′-конца в рамках длины матрицы для синтеза ДНК. Последовательность продукта в виде одиночной цепи ДНК представляет собой комплемент матрицы для синтеза ДНК. Полимеризация продолжается по направлению к 5′ матрицы для синтеза ДНК (или плеча удлинения) до терминации полимеризации. Таким образом, матрица для синтеза ДНК соответствует части плеча удлинения, которая закодирована в продукте в виде одноцепочечной ДНК (то есть 3′ одноцепочечного ДНК-флэпа, содержащего информацию в виде требуемого генетического изменения) полимеразой комплекса праймерного редактора и которая в конечном итоге замещает соответствующую эндогенную цепь ДНК участка-мишени, который располагается непосредственно ниже PE-индуцированного участка разрыва. Без связи с теорией, полимеризация матрицы для синтеза ДНК продолжается по направлению к 5′-концу плеча удлинения до осуществления терминации. Полимеризация может терминироваться рядом способов, включая, но без ограничения, (а) достижение 5′-конца

PEgRNA (например в случае 5′ плеча удлинения, где у ДНК-полимеразы просто заканчивается матрица), (b) достижение непроходимой вторичной структуры РНК (например шпильки или стебля/петли), или (c) достижение сигнала терминации репликации, например определенной нуклеотидной последовательности, которая блокирует или ингибирует полимеразу, или топологического сигнала нуклеиновой кислоты, например суперспиральной ДНК или РНК. Эффективное количество

[0152] В соответствии с используемым здесь значением термин "эффективное количество" обозначает количество биологически активного агента, которое является достаточным для реализации требуемого биологического ответа. Например, в некоторых вариантах осуществления эффективное количество праймерного редактора может обозначать количество редактора, которое является достаточным для редактирования участка-мишени нуклеотидной последовательности, например генома. В некоторых вариантах осуществления указанное здесь эффективное количество праймерного редактора, например слитого белка, включающего домен никазы Cas9 и обратную транскриптазу, может обозначать количество слитого белка, которое является достаточным для индуцирования редактирования участка-мишени, специфично связанного и подвернутого редактирования слитым белком. Как будет очевидно специалисту в данной области, эффективное количество агента, например слитого белка, нуклеазы, гибридного белка, димера белка, комплекса белка (или димера белка) и полинуклеотида, или полинуклеотида может варьироваться в зависимости от различных факторов, например требуемого биологического ответа, например подвигаемого редактированию определенного аллеля, генома, или участка-мишени, являющейся мишенью клетки или ткани, и используемого агента.

Функциональный эквивалент

[0153] Термин "функциональный эквивалент" обозначает вторую биомолекулу, которая является эквивалентной в плане функции, но не обязательно эквивалентной в плане структуры относительно первой биомолекулы. Например, "эквивалент Cas9" обозначает белок, который имеет такие же или по существу такие же функции, что и Cas9, но не обязательно такую же аминокислотную последовательность. В контексте изобретения в тексте описания указывается "белок X или его функциональный эквивалент". В данном контексте "функциональный эквивалент" белка X охватывает любую гомологичную,

паралогичную, представляющую собой фрагмент, встречающуюся в природе, сконструированную, мутированную или синтетическую версию белка X, которая несет эквивалентную функцию.

Слитый белок

[0154] В соответствии с используемым здесь значением термин "слитый белок" обозначает гибридный полипептид, который включает домены белка из, по меньшей мере, двух различных белков. Один белок может быть расположен на амино-концевой (N-концевой) части слитого белка или карбокси-концевой (C-концевой) части слитого белка, тем самым образуя "амино-концевой слитый белок" или "карбокси-концевой слитый белок", соответственно. Белок может включать различные домены, например домен связывания нуклеиновой кислоты (например домен связывания нРНК Cas9, который направляет связывание белка с участком-мишенью) и домен расщепления нуклеиновой кислоты или каталитический домен белка для редактирования нуклеиновой кислоты. Другой пример включает Cas9 или его эквивалент, слитый с обратной транскриптазой. Любой из указанных здесь белков может быть получен с помощью любого способа, известного в данной области. Например, указанные здесь белки могут быть получены с помощью экспрессии рекомбинантного белка и очистки, которые являются особенно подходящими для слитых белков, включающих пептидный линкер. Способы экспрессии рекомбинантных белков и очистки хорошо известны и включают способы, описанные Green and Sambrook, Molecular Cloning: A Laboratory Manual (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)), полное содержание которого приведено здесь в качестве ссылки.

Генный продукт

[0155] В соответствии с используемым здесь значением термин "генный продукт" обозначает любой продукт, кодируемый последовательностью нуклеиновой кислоты. Соответственно, генный продукт может представлять собой, например, первичный транскрипт, зрелый транскрипт, процессированный транскрипт, или белок или пептид, кодируемый транскриптом. Примеры генных продуктов, соответственно, включают, иРНК, рРНК, тРНК, образующие шпильки РНК, микроРНК (миРНК), shРНК (короткие образующие шпильки РНК), siРНК (короткие интерферирующие РНК), и пептиды и белки, например репортерные белки или терапевтические белки.

Представляющий интерес ген (GOI)

[0156] Термин "представляющий интерес ген" или "GOI" обозначает ген, который кодирует представляющую интерес биомолекулу (например белок или молекулу РНК). Представляющий интерес белок может включать любой внутриклеточный белок, белок мембраны, или внеклеточный белок, например ядерный белок, фактор транскрипции, транспортер ядерной мембраны, ассоциированный с внутриклеточной органеллой белок, мембранный рецептор, каталитический белок, и фермент, терапевтический белок, мембранный белок, мембранный транспортерный белок, белок передачи сигнала, или иммунологический белок (например IgG или другой белок в виде антитела), и так далее. Представляющий интерес ген может также кодировать молекулу РНК, включая, но без ограничения, информационную РНК (иРНК), транспортную РНК (тРНК), рибисомную РНК (рРНК), малую ядерную РНК (мяРНК), антисмысловую РНК, направляющую РНК, микроРНК (мРНК), малую интерферирующую РНК (миРНК) и бесклеточную РНК (бкРНК).

Направляющая РНК ("нРНК")

[0157] В соответствии с используемым здесь значением термин "направляющая РНК" обозначает определенный тип направляющей нуклеиновой кислоты, которая наиболее часто ассоциирована с белком Cas CRISPR-Cas9 и которая ассоциируется с Cas9, направляя белок Cas9 к определенной последовательности в молекуле ДНК, которая включает комплементарность относительно последовательности протоспейсера направляющей ДНК. Как описано в других разделах, PEgRNA представляют собой подкатегорию направляющей РНК, которая дополнительно включает плечо удлинения на 3'- или 5'-конце направляющей, которое обеспечивает возможность использования молекулы с описанными здесь праймерными редакторами. Термин "направляющая РНК" также охватывает эквивалентные молекулы направляющей нуклеиновой кислоты, которые ассоциируется с эквивалентами Cas9, гомологами, ортологами, или паралогами, которые встречаются или не встречаются в природе (например являются сконструированными или рекомбинантными) и которые иным образом программируют эквивалент Cas9 для локализации в определенной нуклеотидной последовательности-мишени. Эквиваленты Cas9 могут включать другой napDNAbp из системы CRISPR любого типа (например типа II, V, VI), включая Cpf1 (системы CRISPR-Cas типа V), C2c1 (системы CRISPR-Cas типа

V), C2c2 (системы CRISPR-Cas типа VI) и C2c3 (системы CRISPR-Cas типа V). Дополнительные эквиваленты Cas описаны в Makarova *et al.*, "C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector," *Science* 2016; 353(6299), содержание которого приведено здесь в качестве ссылки. Здесь указаны примеры последовательностей и структуры направляющих РНК. Кроме того, здесь указаны способы конструирования подходящих последовательностей направляющих РНК. В соответствии с используемым здесь значением термин "направляющая РНК" может также обозначать "традиционную направляющую РНК" для ее противопоставления модифицированным формам направляющей РНК, обозначаемым как "направляющие РНК праймерного редактора" (или "PEgRNA"), которые были изобретены для использования в рамках описанных здесь способов праймерного редактирования и композиций.

[0158] Направляющие РНК или PEgRNA могут включать различные структурные элементы, которые включают, но без ограничения:

[0159] Спейсерную последовательность — последовательность в направляющей РНК или PEgRNA (имеющую длину от около 10 до около 40 (например около 10, около 15, около 20, около 25, около 30) нуклеотидов), которая связывается с протоспейсером (как определено здесь ниже) в ДНК-мишени.

[0160] Ядро нРНК (или последовательность каркаса или основы нРНК) — обозначает последовательность в нРНК, которая отвечает за связывание napDNAbp (например Cas9), она не включает спейсер/нацеливающую последовательность, которая используется для направления napDNAbp (например Cas9) к ДНК-мишени.

[0161] Плечо удлинения — обозначает удлиненную часть направляющей РНК на 5'- или 3'-конце, выключающую плечо гомологии, матрицу для редактирования, и участок связывания праймера. Данный компонент дополнительно описан здесь.

[0162] Плечо гомологии — обозначает часть (части) плеча удлинения, которые кодируют часть получаемого кодируемого обратной транскриптазой одноцепочечного ДНК-флэпа, который подвергается встраиванию в участок-мишень ДНК путем замещения эндогенной цепи. Часть одноцепочечного ДНК-флэпа, кодируемая плечом гомологии, является комплементарной не подвергаемой редактированию цепи последовательности-мишени ДНК, которая обеспечивает вытеснение эндогенной цепи и отжиг одноцепочечного ДНК-

флэпа в его месте, тем самым обеспечивая внесение изменения. Данный компонент дополнительно описан здесь.

[0163] Матрицу для редактирования — обозначает часть плеча удлинения, которая кодирует требуемое изменение в одноцепочечном ДНК-флэпе, который синтезируется обратной транскриптазой. Данный компонент дополнительно описан здесь.

[0164] Участок связывания праймера — обозначает часть плеча удлинения, которая отжигается на праймерной последовательности, которая образуется из цепи ДНК-мишени после ее никирования, опосредованного Cas9. Данный компонент дополнительно описан здесь.

[0165] Терминатор транскрипции — направляющая РНК или PEgRNA может включать последовательность терминации транскрипции на 3′-конце молекулы. Как правило, последовательности терминатора транскрипции (например SEQ ID NO: 1361560-1361565) имеют длину от около 70 до около 125 нуклеотидов, однако охватываются и могут использоваться любые известные в данной области более короткие и более длинные последовательности терминатора транскрипции.

Флэп-эндонуклеаза (например FEN1)

[0166] В соответствии с используемым здесь значением термин "флэп-эндонуклеаза" обозначает фермент, который катализирует удаление 5′ одноцепочечных ДНК-флэпов. Они представляют собой встречающиеся в природе ферменты, которые процессируют удаление 5′ флэпов, образованных в рамках клеточных процессов, включая репликацию ДНК. В рамках описанных здесь способов праймерного редактирования могут применяться эндогенно или *in trans* поступающие флэп-эндонуклеазы для удаления 5′ флэпа эндогенной ДНК, образованного в участке-мишени при праймерном редактировании. Флэп-эндонуклеазы известны в данной области и описаны в Patel et al., "Flap endonucleases pass 5′-flaps through a flexible arch using a disorder-thread-order mechanism to confer specificity for free 5′-ends," *Nucleic Acids Research*, 2012, 40(10): 4507-4519 и Tsutakawa et al., "Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily," *Cell*, 2011, 145(2): 198-211, и Balakrishnan et al., "Flap Endonuclease 1," Annu Rev Biochem, 2013, Vol 82: 119-138 (каждый из которых приведен здесь в качестве ссылки). Пример флэп-эндонуклеазы представляет собой

FEN1, которая может быть описана в виде следующей аминокислотной последовательности:

Описание	Последовательность	SEQ ID NO
FEN1 дикого типа	MGIQGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSIYQFLIAVRQ GGDVLQNEEGETTSHLMGMFYRTIRMMENGIKPVYVFDGKPPQLKSGE LAKRSERRAEAEKQLQQAQAAGAEQEVEKFTKRLVKVTKQHNDECKHL LSLMGIPYLDAPSEAEASCAALVKAGKVYAAATEDMDCLTFGSPVLMR HLTASEAKKLPIQEFHLSRILQELGLNQEQFVDLCILLGSDYCESIRG IGPKRAVDLIQKHKSIEEIVRRLDPNKYPVPENWLHKEAHQLFLEPEV LDPESVELKWSEPNEEELIKFMCGEKQFSEERIRSGVKRLSKSRQGST QGRLDDFFKVTGSLSSAKRKEPEPKGSTKKKAKTGAAGKFKRGK	SEQ ID NO: 1361542

Слитый белок

[0167] В соответствии с используемым здесь значением термин "слитый белок" обозначает гибридный полипептид, который включает домены белка из, по меньшей мере, двух различных белков. Один белок может быть расположен на амино-концевой (N-концевой) части слитого белка или карбокси-концевой (C-концевой) части слитого белка, тем самым образуя "амино-концевой слитый белок" или "карбокси-концевой слитый белок", соответственно. Белок может включать различные домены, например домен связывания нуклеиновой кислоты (например домен связывания нРНК Cas9, который обеспечивает связывание белка с участком-мишенью) и домен расщепления нуклеиновой кислоты (например никазу Cas9, napDNAbp) или каталитический домен белка для редактирования нуклеиновой кислоты (например домен RT). Другой пример включает napDNAbp (например Cas9) или его эквивалент, слитый с обратной транскриптазой. Любой из указанных здесь белков может быть получен с помощью любого способа, известного в данной области. Например, указанные здесь белки могут быть получены с помощью экспрессии рекомбинантного белка и очистки, которые являются особенно подходящими для слитых белков, включающих пептидный линкер. Способы экспрессии рекомбинантных белков и очистки хорошо известны и включают способы, описанные Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)), полное содержание которого приведено здесь в качестве ссылки.

Плечо гомологии

[0168] Термин "плечо гомологии" обозначает часть плеча удлинения, которая включает последовательность получаемого кодируемого обратной транскриптазой одноцепочечного ДНК-флэпа, который подвергается встраиванию в участок-мишень ДНК путем замещения

эндогенной цепи. Часть одноцепочечного ДНК-флэпа, кодируемая плечом гомологии, является комплементарной не подвергаемой редактированию цепи последовательности-мишени ДНК, которая обеспечивает вытеснение эндогенной цепи и отжиг одноцепочечного ДНК-флэпа в его месте, тем самым обеспечивая внесение изменения. Данный компонент дополнительно описан здесь.

Клетка-хозяин

[0169] В соответствии с используемым здесь значением термин "клетка-хозяин" обозначает клетку, которая может являться хозяином, обеспечивать репликацию и экспрессию описанного здесь вектора, например вектора, включающего молекулу нуклеиновой кислоты, кодирующей слитый белок, включающий napDNAbp или эквивалент napDNAbp (например Cas9 или эквивалент) и обратную транскриптазу.

Выделенный

[0170] Термин "выделенный" обозначает измененный или удаленный из естественной среды. Например, нуклеиновая кислота или пептид, естественным образом присутствующий в живом животном не является "выделенным", но та же нуклеиновая кислота или пептид, который частично или полностью отделен от сопутствующих материалов его естественной среды, является "выделенным". Выделенная нуклеиновая кислота или белок может существовать в по существу очищенной форме или может существовать в ненативной среде, такой как, например, клетка-хозяин.

[0171] В некоторых вариантах осуществления представляющий интерес ген кодируется выделенной нуклеиновой кислотой. В соответствии с используемым здесь значением термин "выделенный" обозначает характеристику материала, как указано здесь, являющегося удаленным из его исходной или нативной среды (например природной среды, если он встречается в природе). Таким образом, встречающийся в природе полинуклеотид или белок или полипептид, присутствующий в живом животном, не является выделенным, но тот же полинуклеотид или полипептид, отделенный за счет вмешательства человека от некоторых или всех сопутствующих материалов в рамках природной системы, является выделенным. Искусственный или сконструированный материал, например не встречающийся в природе конструкт нуклеиновой кислоты, например описанные здесь конструкты экспрессии и векторы, соответственно также обозначается в качестве выделенного. Материал не обязательно должен быть

очищенным, чтобы быть выделенным. Соответственно, материал может представлять собой часть вектора и/или часть композиции и все равно являться выделенным, поскольку данный вектор или композиция не является частью среды, в которой материал находится в природе.

napDNAbp

[0172] В соответствии с используемым здесь значением термин "программируемый нуклеиновой кислотой ДНК-связывающий белок" или "napDNAbp", примером которого является Cas9, обозначает белки, включающие использование РНК:ДНК гибридизации для нацеливания и связывания определенных последовательностей в молекуле ДНК. Каждый napDNAbp связан, по меньшей мере, с одной направляющей нуклеиновой кислотой (например направляющей РНК), которая локализует napDNAbp в последовательности ДНК, которая включает цепь ДНК (то есть цепь-мишень), которая комплементарна направляющей нуклеиновой кислоте или ее части (например протоспейсеру направляющей РНК). Другими словами, направляющая нуклеиновая кислота "программирует" napDNAbp (например Cas9 или его эквивалент) для локализации и связывания с комплементарной последовательностью.

[0173] Без связи с теорией, механизм связывания комплекса napDNAbp-направляющая РНК в целом включает этап образования R-петли, при котором napDNAbp индуцирует разматывание двуцепочечной ДНК-мишени, тем самым разделяя цепи в участке, связываемом napDNAbp. Затем протоспейсер направляющей РНК гибридизируется с "цепью-мишенью". Это обеспечивает вытеснение "не являющейся мишенью цепи", которая комплементарна цепи-мишени, которая образует одноцепочечный участок R-петли. В некоторых вариантах осуществления napDNAbp включает один или несколько видов нуклеазной активности, которые затем обеспечивают разрезание ДНК, оставляя различные типы разрезов. Например napDNAbp может включать нуклеазную активность, обеспечивающую разрезание не являющейся мишенью цепи в первом месторасположении и/или разрезание цепи-мишени во втором месторасположении. В зависимости от нуклеазной активности, ДНК-мишень может быть разрезана с образованием "двуцепочечного разрыва", при котором обе цепи являются разрезанными. В других вариантах осуществления ДНК-мишень может быть разрезана только в одном участке, то есть ДНК "подвергается никированию" на одной цепи. Пример napDNAbp с различными

видами нуклеазной активности включает "никазу Cas9" ("nCas9") и деактивированный Cas9 без нуклеазной активности ("мертвый Cas9" или "dCas9"). Здесь приведены примеры последовательностей для данных и других napDNAbp.

Линкер

[0174] В соответствии с используемым здесь значением термин "линкер" обозначает молекулу, связывающую две другие молекулы или фрагмента. Линкеры хорошо известны в данной области и могут включать любую подходящую комбинацию нуклеиновых кислот или аминокислот для обеспечения правильного функционирования структур, которые они соединяют. Линкер может представлять собой группу аминокислот. Линкер может представлять собой любую аминокислотную последовательность в случае, если линкер соединяет два слитых белка. Например, napDNAbp (например Cas9) может быть слит с обратной транскриптазой с помощью аминокислотной последовательности линкера. Линкер также может представлять собой нуклеотидную последовательность в случае соединения вместе двух нуклеотидных последовательностей. Например, в данном случае традиционная направляющая РНК соединяется через спейсерную или линкерную нуклеотидную последовательность с РНК удлинением направляющей РНК праймерного редактора, которая может включать последовательность матрицы для синтеза ДНК (например последовательность RT матрицы) и участок связывания праймера. В других вариантах осуществления линкер представляет собой органическую молекулу, группу, полимер, или химический фрагмент. В некоторых вариантах осуществления линкер имеет длину 5-100 аминокислот, например длину 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, или 150-200 аминокислот. В некоторых вариантах осуществления линкер имеет длину 5-100 нуклеотидов, например 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, 150-200, 200-300, 300-500, 500-1000, 1000-2000, или 2000-5000 нуклеотидов. Также охватываются более длинные или короткие линкеры.

Никаза

[0175] Термин "никаза" обозначает napDNAbp (например Cas9) с деактивированным одним из двух нуклеазных доменов. Данный фермент способен расщеплять только одну цепь ДНК-мишени.

Последовательность ядерной локализации (NLS)

[0176] Термин "последовательность ядерной локализации" или "NLS" обозначает аминокислотную последовательность, которая промотирует импорт белка в ядро клетки, например за счет ядерного транспорта. Последовательности ядерной локализации известны в данной области и будут очевидны специалисту в данной области. Например, последовательности NLS описаны Plank *et al.*, в международной PCT заявке PCT/EP2000/011690, поданной 23 ноября 23 2000 г., опубликованной как WO2001/038547 31 мая 2001 г., содержание которой приведено здесь в качестве ссылки в части описания примеров последовательностей ядерной локализации. В некоторых вариантах осуществления NLS включает аминокислотную последовательность PKKKRKV (SEQ ID NO: 1361531) или MDSLLMNRRKFLYQFKNVRWAKGRRETYLC (SEQ ID NO: 1361533).

Молекула нуклеиновой кислоты

[0177] В соответствии с используемым здесь значением термин "нуклеиновая кислота" обозначает полимер (то есть несколько, более одного (например 2, 3, 4, и так далее) нуклеотида). Полимер может включать природные нуклеозиды (то есть аденозин, тимидин, гуанозин, цитидин, уридин, дезоксиаденозин, дезокситимидин, дезоксигуанозин и дезоксицитидин), аналоги нуклеозидов (например 2-аминоаденозин, 2-тиотимидин, инозин, пирроло-пиримидин, 3-метиладенозин, 5-метилцитидин, C5-бромуридин, C5-фторуридин, C5-йодоуридин, C5-пропинилуридин, C5-пропинилцитидин, C5-метилцитидин, 7-деазааденозин, 7-деазагуанозин, 8-оксоаденозин, 8-оксогуанозин, O(6)метилгуанин, 4-ацетилцитидин, 5-(карбоксигидроксиметил)уридин, дигидроуридин, метилпсевдоуридин, 1-метиладенозин, 1-метилгуанозин, N6-метиладенозин и 2-тиоцитидин), химически модифицированные основания, биологически модифицированные основания (например, метилированные основания), интеркалированные основания, модифицированные сахара (например, 2'-фторорибозу, рибозу, 2'-дезоксирибозу, 2′-O-метилцитидин, арабинозу и гексозу) или модифицированные фосфатные группы (например, фосфоротиоаты и 5′-N фосфорамидитные связи).

Азотистое основание

[0178] В соответствии с используемым здесь значением термин "азотистое основание" обозначает содержащие азот биологические соединения, которые образуют нуклеозиды,

которые в свою очередь представляют собой компоненты нуклеотидов, при этом все данные мономеры представляют собой основные строительные блоки нуклеиновых кислот. Способность азотистых оснований образовывать пары оснований и связываться друг с другом приводит к непосредственному образованию спиральных структур в виде длинных цепей, например рибонуклеиновой кислоты (РНК) и дезоксирибонуклеиновой кислоты (ДНК).

[0179] Пять азотистых оснований, которые представляют собой аденин (A), цитозин (C), гуанин (G), тимин (T), и урацил (U), могут быть обозначены как первичные или стандартные. Они функционируют в качестве фундаментальных единиц генетического кода, при этом основания A, G, C, и T находятся в ДНК, в то время как A, G, C, и U находятся в РНК. Тимин и урацил являются идентичными за исключением того, что T включает метильную группу, которая отсутствует у U. ДНК и РНК также могут включать модифицированные азотистые основания. Например, для азотистых оснований в виде аденозина и гуанозина альтернативные азотистые основания могут включать гипоксантин, ксантин, или 7-метилгуанин, которые соответствуют альтернативным нуклеозидам в виде инозина, ксантозина, и 7-метилгуанозина, соответственно. Кроме того, например, для азотистых оснований в виде цитозина, тимина или уридина альтернативные азотистые основания могут включать 5,6-дигидроурацил, 5-метилцитозин или 5-гидроксиметилцитозин, которые соответствуют альтернативным нуклеозидам в виде дигидроуридина, 5-метилцитидина и 5-гидроксиметилцитидин соответственно. Азотистые основания также могут включать аналоги азотистых оснований, большое количество которых известны в данной области. Как правило, аналоги азотистых оснований придают, помимо прочего, различные свойства, связанные со связыванием оснований и группированием оснований. Примеры включают универсальные основания, которые могут образовывать пары со всеми четырьмя стандартными основаниями, и аналоги с фосфат-сахарным каркасом, например PNA, которые изменяют свойства цепи (PNA может образовывать даже тройную спираль). Аналоги нуклеиновых кислот также обозначаются как "ксенонуклеиновая кислота" и представляют собой одну из двух основ ксенобиологии — конструирования новых для природы форм жизни на основе альтернативных биохимических принципов. Искусственные нуклеиновые кислоты включают пептидную нуклеиновую кислоту (PNA), морфолино и заблокированную

нуклеиновую кислоту (LNA), а также гликолевую нуклеиновую кислоту (GNA) и треозную нуклеиновую кислоту (TNA). Каждая из них отличается от встречающейся в природе ДНК или РНК изменениями в каркасе молекулы. Примеры аналогов представляют собой, например, 2-аминоаденозин, 2-тиотимидин, инозин, пирроло-пиримидин, 3-метиладенозин, 5-метилцитидин, C5-бромуридин, C5-фторуридин, C5-йодоуридин, C5-пропинилуридин, C5-пропинилцитидин, C5-метилцитидин, 7-деазааденозин, 7-деазагуанозин, 8-оксоаденозин, 8-оксогуанозин, O(6)метилгуанин, 4-ацетилцитидин, 5-(карбоксигидроксиметил)уридин, дигидроуридин, метилпсевдоуридин, 1-метиладенозин, 1-метилгуанозин, N6-метиладенозин и 2-тиоцитидин), химически модифицированные основания, биологически модифицированные основания (например, метилированные основания), интеркалированные основания, модифицированные сахара (например, 2'-фторорибозу, рибозу, 2'-дезоксирибозу, 2′-O-метилцитидин, арабинозу и гексозу) или модифицированные фосфатные группы (например, фосфоротиоаты и 5′-N фосфорамидитные связи).

PEgRNA

[0180] В соответствии с используемым здесь значением термины "направляющая РНК праймерного редактора" или "PEgRNA" или "удлиненная направляющая РНК" обозначает специализированную форму направляющей РНК, которая была модифицирована для включения одной или нескольких дополнительных последовательностей для использования в рамках описанных здесь способов праймерного редактирования, композиций, и систем. Как описано здесь, направляющая РНК праймерного редактора включает один или несколько "удлиненных участков" последовательности нуклеиновой кислоты. Удлиненные участки могут включать, но без ограничения, одноцепочечную РНК. Дополнительно, удлиненные участки могут присутствовать на 3′-конце традиционной направляющей РНК. В рамках других вариантов удлиненные участки могут присутствовать на 5′-конце традиционной направляющей РНК. В других вариантах удлиненный участок может располагаться во внутримолекулярном участке вместо одного из кондов традиционной направляющей РНК, например в участке в виде ядра нРНК, который ассоциируется и/или связывается с napDNAbp. Удлиненный участок включает "последовательность матрицы в виде обратной транскриптазы", которая представляет собой одноцепочечную молекулу РНК, которая кодирует одноцепочечную

комплементарную ДНК (кДНК), которая, в свою очередь, была сконструирована для того, чтобы (a) являться гомологичной подвергаемой редактированию эндогенной ДНК-мишени и (b) включать, по меньшей мере, одну требуемое нуклеотидное изменение (например транзицию, трансверсию, делецию, вставку, или их комбинацию) для внесения или встраивания в эндогенную ДНК-мишень. Удлиненный участок может также включать другие функциональные элементы последовательности, например, но без ограничения, "участок связывания праймера" и/или "спейсерную или линкерную" последовательность. В соответствии с используемым здесь значением "участок связывания праймера" включает последовательность, которая гибридизируется с одноцепочечной последовательностью ДНК, имеющей 3′-конец, образованный из никированной ДНК R-петли, которая включает праймер для обратной транскриптазы.

[0181] В некоторых вариантах осуществления PEgRNA соответствуют Фиг.3A, на которой показана PEgRNA, имеющая 5′ плечо удлинения, спейсер, и ядро нРНК. 5′ удлинение дополнительно включает в направлении от 5′ к 3′ матрицу в виде обратной транскриптазы, участок связывания праймера, и линкер.

[0182] В некоторых вариантах осуществления PEgRNA соответствуют Фиг.3B, на которой показана PEgRNA, имеющая 3′ плечо удлинения, спейсер, и ядро нРНК. 3′ удлинение дополнительно включает в направлении от 5′ к 3′ матрицу в виде обратной транскриптазы и участок связывания праймера.

[0183] В других вариантах осуществления PEgRNA соответствуют Фиг.27, на которой показана PEgRNA, имеющая в направлении от 5′ к 3′ спейсер (1), ядро нРНК (2), и плечо удлинения (3). Плечо удлинения (3) расположено на 3'-конце PEgRNA. Плечо удлинения (3) дополнительно включает в направлении от 5′ к 3′ "участок связывания праймера" (A), "матрицу для редактирования" (B), и "плечо гомологии" (C). Плечо удлинения (3) также может включать необязательный модификаторный участок на 3′- и 5′-концах, который может представлять собой одинаковые или различающиеся последовательности. Кроме того, 3′-конец PEgRNA может включать последовательность терминатора транскрипции. Данные элементы последовательности PEgRNA дополнительно описаны и определены здесь. Кроме того, в описании раскрыт пример PEgRNA, которая была сконструирована в соответствии с описанными здесь способами, в рамках прилагаемого Перечня последовательностей.

[0184] В других вариантах осуществления PEgRNA соответствуют Фиг.28, на которой показана PEgRNA, имеющая в направлении от 5′ к 3′ плечо удлинения (3), спейсер (1), и ядро нРНК (2). Плечо удлинения (3) расположено на 5'-конце PEgRNA. Плечо удлинения (3) дополнительно включает в направлении от 3′ к 5′ "участок связывания праймера" (A), "матрицу для редактирования" (B), и "плечо гомологии" (C). Плечо удлинения (3) также может включать необязательный модификаторный участок на 3′ и 5′-концах, который может представлять собой одинаковые или различающиеся последовательности. PEgRNA может также включать последовательность терминатора транскрипции на 3′-конце. Данные элементы последовательности PEgRNA дополнительно описаны и определены здесь.

Пептидная метка

[0185] Термин "пептидная метка" обозначает аминокислотную последовательность пептида, которая генетически слита с последовательностью белка для придания одной или нескольких функций белкам, которые обеспечивают возможность манипуляции белком для различных целей, например визуализации, идентификации, локализации, очистки, солюбилизации, сепарации, и так далее. Пептидные метки могут включать различные типы меток, разделенные на категории на основе цели или функции, которые могут включать "аффинные метки" (для реализации очистки белка), "метки солюбилизации" (для обеспечения правильного фолдинга белков), "хроматографические метки" (для изменения хроматографических свойств белков), "эпитопные метки" (для связывания с антителами с высокой аффинностью), "флуоресцентные метки" (для обеспечения визуализации белков в клетке или *in vitro*).

PE1

[0186] В соответствии с используемым здесь значением "PE1" обозначает комплекс PE, включающий слитый белок, включающий Cas9(H840A) и RT MMLV дикого типа, имеющий следующую структуру: [NLS]-[Cas9(H840A)]-[линкер]-[MMLV_RT (дикого типа)] + требуемая PEgRNA, где продукт слияния PE имеет аминокислотную последовательность SEQ ID NO: 1361515, которая является следующей:

MKRTADGSEFESPKKKRKVDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALL
FDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDA
KLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDL

TLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSG
EQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN
EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDM
YVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLI
TQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKL
VSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKA
TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY
EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHL
FTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD<u>SGGSSGGSSGSETP</u>
<u>GTSESATPESSGGSSGGSS</u>*TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIP*
LKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKR
VEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQG
FKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQI
CQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTG
TLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDP
VAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQ
FGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVT
TETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEI
KNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP<u>SGGS</u>
<u>KRTADGSEFEPKKKRKV</u> (SEQ ID NO: 1361515)

обозначения:
<u>Последовательность ядерной локализации (NLS)</u> Сверху:(SEQ ID NO: 1361532), Снизу: (SEQ ID NO: 1361541)
Cas9(H840A) (SEQ ID NO: 1361454)
<u>Линкер длиной 33 аминокислоты</u> (SEQ ID NO: 1361528)
Обратная транскриптаза M-MLV (SEQ ID NO: 1361485).

<u>PE2</u>

[0187] В соответствии с используемым здесь значением "PE2" обозначает комплекс PE, включающий слитый белок, включающий Cas9(H840A) и вариант RT MMLV, имеющий следующую структуру: [NLS]-[Cas9(H840A)]-[линкер]-[MMLV_RT(D200N)(T330P)(L603W)(T306K)(W313F)] + требуемая PEgRNA, где продукт слияния PE имеет аминокислотную последовательность SEQ ID NO: 1361516, которая является следующей:

<u>MKRTADGSEFESPKKKRKV</u>**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALL**
FDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDA
KLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDL
TLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ

RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSG
EQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN
EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDM
YVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLI
TQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKL
VSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKA
TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY
EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHL
FTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSSGGSSGSETP
GTSESATPESSGGSSGGSS*TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIP*
LKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKR
VEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQG
FKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQI
CQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPG
TLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDP
VAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQ
FGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVT
TETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEI
*KNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP*SGGS
KRTADGSEFEPKKKRKV (SEQ ID NO: 1361516)

обозначения:
Последовательность ядерной локализации (NLS) Сверху:(SEQ ID NO: 1361532), Снизу: (SEQ ID NO: 1361541)
Cas9(H840A) (SEQ ID NO: 1361454)
Линкер длиной 33 аминокислоты (SEQ ID NO: 1361528)
Обратная транскриптаза M-MLV (SEQ ID NO: 1361514).

PE3

[0188] В соответствии с используемым здесь значением термин "PE3" обозначает PE2 плюс никирующую вторую цепь направляющую РНК, которая образует комплекс с PE2 и вносит разрыв в не подвергаемой редактированию цепи ДНК для индуцирования предпочтительного замещения подвергнутой редактированию цепи.

PE3b

[0189] В соответствии с используемым здесь значением термин "PE3b" обозначает PE3, но где никирующая вторую цепь направляющая РНК сконструирована для временно́го контроля, так что разрыв второй цепи не осуществляется до реализации внесения требуемого изменения. Это достигается за счет конструирования нРНК со спейсерной последовательностью, которая соответствует только подвергаемой редактированию цепи,

но не исходному аллелю. За счет использования данной стратегии, обозначаемой далее как PE3b, несовпадения между протоспейсером и не подвергаемым редактированию аллелем должны способствовать отсутствию никирования за счет sgРНК до реализации редактирования на включающей PAM цепи.

PE-short

[0190] В соответствии с используемым здесь значением термин "PE-short" обозначает конструкт PE, который слит с усеченной по С-концу обратной транскриптазой и имеет следующую аминокислотную последовательность:

MKRTADGSEFESPKKKRKV**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALL**
FDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDA
KLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDL
TLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSG
EQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN
EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDM
YVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLI
TQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKL
VSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKA
TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY
EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHL
FTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSSGGSSGSETP
GTSESATPESSGGSSGGSS*TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIP*
LKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKR
VEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQG
FKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQI
CQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPG
TLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDP
VAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQ
*FGPVVALNPATLLPLPEEGLQHNCLDNSRLIN*SGGSKRTADGSEFEPKKKRKV (SEQ ID NO: 1361602)

обозначения:
Последовательность ядерной локализации (NLS) Сверху:(SEQ ID NO: 1361532), Снизу: (SEQ ID NO: 1361541)
Cas9(H840A) (SEQ ID NO: 1361454)
Линкер 1 длиной 33 аминокислоты (SEQ ID NO: 1361528)
Усеченная обратная транскриптаза M-MLV
(SEQ ID NO: 1361597)

Процентная доля идентичности

[0191] Термины "процентная доля идентичности", "идентичность последовательности", "% идентичности", или "% идентичности последовательности" (в соответствии с тем, как они могут взаимозаменяемым образом использоваться здесь) последовательности (например нуклеиновой кислоты или аминокислоты) обозначают количественную меру сходства между двумя последовательностями (например нуклеиновой кислотой или аминокислотой). Процентная доля идентичности последовательности геномной ДНК, последовательности интрона и экзона, и аминокислотной последовательности между человеком и другими видами варьируется в зависимости от типа вида, при этом шимпанзе имеют наибольшую процентную долю идентичности относительно человека из всех видов в рамках каждой категории. Процентная доля идентичности может быть определена с помощью алгоритмов Karlin and Altschul, Proc. Natl. Acad. Sci. USA 87:2264-68, 1990, модифицированных согласно Karlin and Altschul, Proc. Natl. Acad. Sci. USA 90:5873-77, 1993. Подобные алгоритмы встроены в программы NBLAST и XBLAST (version 2.0) Altschul *et al.*, J. Mol. Biol. 215:403-10, 1990. Поиски белка BLAST могут быть осуществлены с использованием программы XBLAST, показатель=50, длина слова=3, для получения аминокислотных последовательностей, являющихся гомологичными представляющим интерес молекулам белка. При наличии гэпов между двумя последовательностями может быть использована программа Gapped BLAST, как описано в Altschul *et al.*, Nucleic Acids Res. 25(17):3389-3402, 1997. При использовании программ BLAST и Gapped BLAST могут использоваться параметры по умолчанию соответствующих программ (например XBLAST и NBLAST). При указании или обозначении процентной доли идентичности или ее диапазона (например в виде "по меньшей мере", "более чем", "между", и так далее), если иное не указано, то границы должны включаться и диапазон (например, по меньшей мере, 70% идентичность) будет включать все диапазоны в рамках указанного диапазона (например, по меньшей мере, 71%, по меньшей мере, 72%, по меньшей мере, 73%, по меньшей мере, 74%, по меньшей мере, 75%, по меньшей мере, 76%, по меньшей мере, 77%, по меньшей мере, 78%, по меньшей мере, 79%, по меньшей мере, 80%, по меньшей мере, 81%, по меньшей мере, 82%, по меньшей мере, 83%, по меньшей мере, 84%, по меньшей мере, 85%, по меньшей мере, 86%, по меньшей мере, 87%, по меньшей мере, 88%, по меньшей мере, 89%, по меньшей мере, 90%, по меньшей мере, 91%, по меньшей мере, 92%, по меньшей мере,

93%, по меньшей мере, 94%, по меньшей мере, 95%, по меньшей мере, 95.5%,по меньшей мере, 96%, по меньшей мере, 96,5%, по меньшей мере, 97%, по меньшей мере, 97,5%, по меньшей мере, 98%, по меньшей мере, 98,5%, по меньшей мере, 99%, по меньшей мере, 99,5%, по меньшей мере, 99,6%, по меньшей мере, 99,7%, по меньшей мере, 99,8%, по меньшей мере, 99,9% идентичность) и все их инкременты (например десятые доли процента (то есть 0,1%), сотые доли процента (то есть 0,01%), и так далее).

Праймерный редактор

[0192] Термин "праймерный редактор" обозначает описанные здесь слитые конструкты, включающие napDNAbp (например никазу Cas9) и обратную транскриптазу, способные осуществлять праймерное редактирование в нуклеотидной последовательности-мишени в присутствии PEgRNA (или "удлиненной направляющей РНК"). Термин "праймерный редактор" может обозначать слитый белок или слитый белок в виде комплекса с PEgRNA. В некоторых вариантах осуществления термин "праймерный редактор" может также обозначать комплекс, включающий слитый белок (обратную транскриптазу, слитую с napDNAbp), PEgRNA, и обычную направляющую РНК, способную направлять этап никирования второго участка не подвергаемой редактированию цепи, как описано здесь. В определенных вариантах осуществления компонент "праймерного редактора" в виде обратной транскриптазы поступает *in trans*.

Участок связывания праймера

[0193] Термин "участок связывания праймера" или "PBS" обозначает нуклеотидную последовательность, расположенную на PEgRNA в качестве компонента плеча удлинения (обычно на 3′-конце плеча удлинения) и служащую для связывания с праймерной последовательностью, которая образуется после никирования последовательности-мишени за счет napDNAbp (например Cas9) с помощью праймерного редактора. Как подробно описано в других разделах, когда компонент праймерного редактора в виде никазы Cas9 осуществляет никирование одной цепи последовательности-мишени ДНК, образуется 3′-концевой оцДНК-флэп, который служит праймерной последовательностью, которая отжигается на участке связывания праймера на PEgRNA для праймирования обратной транскрипции. На Фиг.27 и 28 показаны варианты осуществления участка связывания праймера, расположенного на 3′ и 5′ плече удлинения, соответственно.

Белок, пептид, и полипептид

[0194] Термины "белок", "пептид", и "полипептид" используются здесь взаимозаменяемым образом и обозначают полимер из аминокислотных остатков, которые связаны вместе пептидными (амидными) связями. Данные термины обозначают белок, пептид, или полипептид с любым размером, структурой или функцией. Как правило, белок, пептид, или полипептид будет иметь длину, по меньшей мере, в виде трех аминокислот. Термин "белок", "пептид", или "полипептид" может обозначать отдельный белок или группу белков. Одна или несколько аминокислот в белке, пептиде, или полипептиде могут быть модифицированы, например, путем добавления химической сущности, например углеводородной группы, гидроксильной группы, фосфатной группы, фарнезильной группы, изофарнезильной группы, группы в виде жирной кислоты, линкера для конъюгирования, функционализации или другой модификации, и так далее. Белок, пептид, или пептид также может представлять собой отдельную молекулу или представлять собой мультимолекулярный комплекс. Белок, пептид, или полипептид может представлять собой лишь фрагмент встречающегося в природе белка или пептида. Белок, пептид, или полипептид может являться встречающимся в природе, рекомбинантным, или синтетическим, или любой их комбинацией. Любой из указанных здесь белков может быть получен с помощью любого способа, известного в данной области. Например, указанные здесь белки могут быть получены с помощью экспрессии рекомбинантного белка и очистки, которые являются особенно подходящими для слитых белков, включающих пептидный линкер. Способы экспрессии рекомбинантных белков и очистки хорошо известны и включают способы, описанные Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)), полное содержание которого приведено здесь в качестве ссылки.

Функционально связанный

[0195] В соответствии с используемым здесь значением термин "функционально связанный" обозначает функциональную связь между регуляторной последовательностью и гетерологичной последовательностью нуклеиновой кислоты (например трансгеном), которая приводит к экспрессии гетерологичной последовательности нуклеиновой кислоты (например трансгена). Например, первая последовательность нуклеиновой кислоты является функционально связанной со второй последовательностью нуклеиновой кислоты когда первая последовательность нуклеиновой кислоты находится в состоянии

функциональной связи со второй последовательностью нуклеиновой кислоты. Например, промотор является функционально связанным с кодирующей последовательностью, если промотор влияет на транскрипцию или экспрессию кодирующей последовательности. В целом, функционально связанные последовательности нуклеиновой кислоты являются смежными и, в случае необходимости соединения двух участков кодирования белка, находятся в рамках одной рамки считывания.

Промотор

[0196] Термин "промотор" используется в данной области и обозначает молекулу нуклеиновой кислоты с последовательностью, распознаваемой механизмом клеточной транскрипции и способной инициировать транскрипцию расположенного ниже гена. Промотор может являться постоянно активным, что означает, что промотор является всегда активным в указанном клеточном контексте, или условно активным, что означает, что промотор является активным лишь при наличии определенного условия. Например, условный промотор может являться активным лишь в присутствии определенного белка, который соединяет белок, ассоциированный с регуляторным элементом в промоторе, с механизмом транскрипции оснований, или лишь при отсутствии ингибиторной молекулы. Подклассом условно активных промоторов являются индуцируемые промоторы, которые требуют наличия низкомолекулярного "индуктора" для активности. Примеры индуцируемых промоторов включают, но без ограничения, индуцируемые арабинозой промоторы, Tet-on промоторы, и индуцируемые тамоксифеном промоторы. Различные постоянные, условные, и индуцируемые промоторы хорошо известны специалисту в данной области, при этом специалист в данной области будет способен определить ряд подобных промоторов, которые могут применяться при реализации настоящего изобретения, которое не является ограниченным в данном отношении.

Смежный с протоспейсером мотив (PAM)

[0197] В соответствии с используемым здесь значением термин "смежная с протоспейсером последовательность" или "PAM" обозначает последовательность ДНК длиной приблизительно 2-6 пар оснований, которая является важным нацеливающим компонентом никазы Cas9. Как правило, последовательность PAM располагается на любой одной из цепей и расположена ниже в направлении от 5' к 3' относительно участка разреза Cas9. Стандартная последовательность PAM (то есть последовательность PAM,

которая ассоциирована с нуклеазой Cas9 *Streptococcus pyogenes* или SpCas9) представляет собой 5′-NGG-3′, где "N" представляет собой любое азотистое основание с последующими двумя азотистыми основаниями в виде гуанина ("G"). Различные последовательности PAM могут быть ассоциированы с различными нуклеазами Cas9 или эквивалентными белками из других организмов, например 5′-NG-3′, где "N" представляет собой любое азотистое основание с последующим одним азотистым основанием в виде гуанина ("G"), или 5′-KKH-3′, где после двух азотистых оснований в виде лизина ("K") следует один гистидин ("H"). Кроме того, любая нуклеаза Cas9, например SpCas9, может быть модифицирована для изменения специфичности нуклеазы в отношении PAM, так что нуклеаза будет распознавать альтернативную последовательность PAM.

[0198] Например, с учетом того, что аминокислотная последовательность стандартной SpCas9 представляет собой SEQ ID NO: 1361421 (SpCas9 M1 QQ99ZW2 дикого типа), последовательность PAM может быть модифицирована путем внесения одной или нескольких мутаций, включая (a) D1135V, R1335Q, и T1337R "вариант VQR", который изменяет специфичность в отношении PAM на NGAN или NGNG, (b) D1135E, R1335Q, и T1337R "вариант EQR", который изменяет специфичность в отношении PAM на NGAG, и (c) D1135V, G1218R, R1335E, и T1337R "вариант VRER", который изменяет специфичность в отношении PAM на NGCG. Кроме того, вариант D1135E стандартной SpCas9 все еще распознает NGG, но является более избирательным по сравнению с белком SpCas9 дикого типа.

[0199] Также будет очевидно, что ферменты Cas9 из бактерий различных видов (то есть ортологов Cas9) могут иметь различные специфичности в отношении PAM. Например, Cas9 из *Staphylococcus aureus* (SaCas9) распознает NGRRT или NGRRN. Кроме того, Cas9 из *Neisseria meningitis* (NmCas) распознает NNNNGATT. В другом примере Cas9 из *Streptococcus thermophilis* (StCas9) распознает NNAGAAW. В другом примере Cas9 из *Treponema denticola* (TdCas) распознает NAAAAC. Эти примеры не являются ограничивающими. Также будет очевидно, что отличные от SpCas9 нуклеазы связываются с рядом последовательностей PAM, что делает их полезными при отсутствии подходящей последовательности PAM для SpCas9 в требуемом участке-мишени разрыва. Более того, отличные от SpCas9 нуклеазы могут иметь другие характеристики, которые делают их более полезными, чем SpCas9. Например Cas9 из *Staphylococcus aureus* (SaCas9)

имеет на 1 килооснование меньший размер, чем SpCas9, поэтому она может быть упакована в аденоассоциированный вирус (AAV). Дополнительные ссылки могут быть сделаны на Shah *et al.*, "Protospacer recognition motifs: mixed identities and functional diversity," *RNA Biology*, 10(5): 891-899 (который приведен здесь в качестве ссылки).

Протоспейсер

[0200] В соответствии с используемым здесь значением термин "протоспейсер" обозначает последовательность (~20 bp) в ДНК, являющуюся смежной относительно последовательности PAM (смежного с протоспейсером мотива), которая имеет такую же последовательность, что и спейсерная последовательность направляющей РНК. Направляющая РНК отжигается на комплементе протоспейсерной последовательности на ДНК-мишени (в частности, одной ее цепи, то есть "цепи-мишени" против "не являющейся мишенью цепи" последовательности-мишени). Для работы Cas9 ей также требуется определенный смежный с протоспейсером мотив (PAM), который варьируется в зависимости от вида бактерий, к которому относится ген Cas9. Наиболее часто используемая нуклеаза Cas9, относящаяся к *S. pyogenes*, распознает последовательность PAM NGG, которая располагается непосредственно ниже последовательности-мишени в геномной ДНК на не являющейся мишенью цепи. Специалисту в данной области будет известно, что "протоспейсер" в литературе в данной области иногда указывается в виде мишень-специфичной направляющей последовательности длиной ~20 nt на самой направляющей РНК вместо указания его как "спейсера". Таким образом, в некоторых случаях термин "протоспейсер" в соответствии с используемым здесь значением может использоваться взаимозаменяемым образом с термином "спейсер". Контекст описания, связанный с указанием "протоспейсера" или "спейсера", поможет читателю определить, указывает ли данный термин на нРНК или мишень в виде ДНК. Оба варианта использования данных терминов являются приемлемыми, поскольку оба термина используются в данной области в соответствии с обоими данными значениями.

Обратная транскриптаза

[0201] Термин "обратная транскриптаза" обозначает класс полимераз, характеризующихся как РНК-зависимые ДНК-полимеразы. Все известные обратные транскриптазы требуют праймер для синтеза ДНК транскрипта из РНК матрицы. Исторически обратную

транскриптазу использовали в основном для транскрибирования иРНК в кДНК, которая затем может быть клонирована в вектор для дальнейшего манипулирования. Обратная транскриптаза вируса миобластоза птиц (AMV) была первой широко используемой РНК-зависимой ДНК-полимеразой (Verma, *Biochim. Biophys. Acta* 473:1 (1977)). Данный фермент имеет активность 5′-3′ РНК-направленной ДНК-полимеразы, активность 5′-3′ ДНК-направленной ДНК-полимеразы, и активность РНКазы H. РНКаза H представляет собой процессивную 5' и 3' рибонуклеазу, специфичную в отношении цепи РНК в РНК-ДНК гибридах (Perbal, *A Practical Guide to Molecular Cloning*, New York: Wiley & Sons (1984)). Ошибки транскрипции не могут быть скорректированы обратной транскриптазой, поскольку известные вирусные обратные транскриптазы не имеют 3′-5′ экзонуклеазную активность, необходимую для проверки (Saunders and Saunders, Microbial Genetics Applied to Biotechnology, London: Croom Helm (1987)). Подробное исследование активности обратной транскриптазы AMV и связанной с ней активности РНКазы H было представлено Berger *et al.*, Biochemistry 22:2365-2372 (1983). Другая обратная транскриптаза, которая широко применяется в молекулярной биологии, представляет собой обратную транскриптазу из вируса лейкоза мышей Молони (M-MLV). См., например, Gerard, G. R., DNA 5:271-279 (1986) и Kotewicz, M. L., *et al.*, *Gene* 35:249-258 (1985). Также была описана обратная транскриптаза M-MLV, по существу не имеющая активности РНКазы H. См., например, патент США 5244797. Изобретение охватывает применение любых подобных обратных транскриптаз или их вариантов или мутантов.

[0202] Кроме того, изобретение охватывает применение обратных транскриптаз, которые подвержены ошибкам, то есть которые могут быть обозначены как подверженные ошибкам обратные транскриптазы или обратные транскриптазы, которые не поддерживают встраивание нуклеотидов с высокой точностью при полимеризации. При синтезе одноцепочечного ДНК-флэпа на основе RT матрицы, встроенной в направляющую ДНК, подверженная ошибкам обратная транскриптаза может вносить один или несколько нуклеотидов, которые не совпадают с матрицей для синтеза ДНК (например последовательностью RT матрицы), тем самым внося изменения в нуклеотидную последовательность за счет ошибочной полимеризации одноцепочечного ДНК-флэпа. Данные ошибки, внесенные при синтезе одноцепочечного ДНК-флэпа, затем становятся встроенными в двуцепочечную молекулу за счет гибридизации с

соответствующей эндогенной цепью-мишенью, удаления эндогенной вытесненной цепи, лигирования, и затем за счет еще одного или нескольких раундов репарации и/или репликации эндогенной ДНК.

Обратная транскрипция

[0203] В соответствии с используемым здесь значением термин "обратная транскрипция" обозначает способность фермента синтезировать цепь ДНК (то есть комплементарную ДНК или кДНК) с помощью РНК в качестве матрицы. В некоторых вариантах осуществления обратная транскрипция может представлять собой "подверженную ошибкам обратную транскрипцию", что обозначает свойства определенных ферментов в виде обратной транскриптазы, которые являются подверженными ошибкам в плане их активности, связанной с полимеризацией ДНК.

Смысловая цепь

[0204] В генетике "смысловая" цепь представляет собой сегмент в рамках двуцепочечной ДНК, которая проходит от 5' к 3' и которая является комплементарной антисмысловой цепи ДНК, или матричной цепи, которая проходит от 3' к 5'. В случае сегмента ДНК, который кодирует белок, смысловая цепь представляет собой цепь ДНК, которая имеет ту же последовательность, что и иРНК, которая использует антисмысловую цепь в качестве ее матрицы при транскрипции и в конечном итоге (как правило, но не всегда) подвергается трансляции в белок. Антисмысловая цепь, таким образом, отвечает за РНК, которая позже транслируется в белок, в то время как смысловая цепь имеет практически такой же состав, что и иРНК. Следует обратить внимание, что для каждого сегмента дцДНК вероятно будет присутствовать два набора смысловых и антисмысловых цепей в зависимости от направления чтения (поскольку смысл и антисмысл зависят от перспективы). В конечном итоге то, какая цепь одного сегмента дцДНК считается смысловой или антисмысловой, определяется генным продуктом или иРНК.

[0205] В контексте PEgRNA первый этап представляет собой синтез одноцепочечной комплементарной ДНК (то есть 3′ оцДНК-флэпа, который становится встроенным), ориентированной в направлении от 5′ к 3′, которая использует в качестве матрицы плечо удлинения PEgRNA. То, следует ли считать 3′ оцДНК-флэп смысловой или антисмысловой цепью, зависит от направления транскрипции, поскольку считается, что обе цепи ДНК могут служить в качестве матрицы для транскрипции (но не

одновременно). Таким образом, в некоторых вариантах осуществления 3′ оцДНК-флэп (который в целом проходит в направлении от 5′ к 3′) будет служить в качестве смысловой цепи, поскольку он представляет собой кодирующую цепь. В других вариантах осуществления 3′ оцДНК-флэп (который в целом проходит в направлении от 5′ к 3′) будет служить в качестве антисмысловой цепи и, таким образом, матрицы для транскрипции.

Никирование второй цепи

[0206] В соответствии с используемым здесь значением данная концепция обозначает внесение второго разрыва в участке ниже первого разрыва (то есть первоначального участка разрыва, который обеспечивает свободный 3′-конец для использования при праймировании обратной транскриптазы на удлиненной части направляющей РНК). В некоторых вариантах осуществления первый разрыв и второй разрыв располагаются на противоположных цепях. В других вариантах осуществления первый разрыв и второй разрыв располагаются на противоположных цепях. В другом варианте осуществления первый разрыв располагается на не подвергаемой редактированию цепи (то есть цепи, которая образует одноцепочечную часть R-петли), а второй разрыв располагается на цепи-мишени. Второй разрыв расположен, по меньшей мере, на 5 нуклеотидов ниже первого разрыва, или, по меньшей мере, на 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, или 30 или более нуклеотидов ниже первого разрыва. Без связи с теорией, второй разрыв индуцирует клеточные процессы репарации и репликации эндогенной ДНК с замещением не подвергаемо редактированию цепи. В некоторых вариантах осуществления подвергаемая редактированию цепь представляет собой не являющуюся мишенью цепь, а не подвергаемая редактированию цепь представляет собой цепь-мишень. В других вариантах осуществления подвергаемая редактированию цепь представляет собой цепь-мишень, а не подвергаемая редактированию цепь представляет собой не являющуюся мишенью цепь.

Спейсерная последовательность

[0207] В соответствии с используемым здесь значением термин "спейсерная последовательность" при использовании в контексте направляющей нРНК или PEgRNA обозначает часть направляющей нРНК или PEgRNA длиной от около 10 до около 40 (например около 10, около 15, около 20, около 25, около 30) нуклеотидов, которая содержит нуклеотидную последовательность, которая является комплементарной

протоспейсерной последовательности в последовательности-мишени ДНК. Спейсерная последовательность отжигается на протоспейсерной последовательности для образования структуры в виде оцРНК/оцДНК гибрида в участке-мишени и соответствующей структуры в виде оцДНК R-петли эндогенной цепи ДНК, которая комплементарна протоспейсерной последовательности.

Субъект

[0208] В соответствии с используемым здесь значением термин "субъект" обозначает отдельный организм, например отдельное млекопитающее. В некоторых вариантах осуществления субъект представляет собой человека. В некоторых вариантах осуществления субъект представляет собой отличное от человека млекопитающее. В некоторых вариантах осуществления субъект представляет собой отличного от человека примата. В некоторых вариантах осуществления субъект представляет собой грызуна. В некоторых вариантах осуществления субъект представляет собой овцу, козу, крупный рогатый скот, кошку, или собаку. В некоторых вариантах осуществления субъект представляет собой позвоночное, амфибию, рептилию, рыбу, насекомое, муху, или нематоду. В некоторых вариантах осуществления субъект представляет собой лабораторное животное. В некоторых вариантах осуществления субъект является генетически модифицированным, например генетически модифицированным отличным от человека субъектом. Субъект может быть любого пола и на любой стадии развития.

Участок-мишень

[0209] Термин "участок-мишень" обозначает последовательность в молекуле нуклеиновой кислоты, которая подвергается редактированию с помощью описанного здесь праймерного редактора. Термин "участок-мишень" также обозначает последовательность в молекуле нуклеиновой кислоты, с которой связывается комплекс редактора оснований и нРНК.

Временно́й разрыв второй цепи

[0210] В соответствии с используемым здесь значением термин "временно́й разрыв второй цепи" обозначает вариант никирования второй цепи, при котором внесение второго разрыва в не подвергаемую редактированию цепь осуществляется только после внесения требуемого изменения в подвергаемую редактированию цепь. Это обеспечивает возможность избежать одновременных разрывов на обеих цепях, которые могут

приводить к двуцепочечным разрывам ДНК. Никирующая вторую цепь направляющая РНК сконструирована для временно́го контроля, так что разрыв второй цепи не осуществляется до реализации внесения требуемого изменения. Это достигается за счет конструирования нРНК со спейсерной последовательностью, которая соответствует только подвергаемой редактированию цепи, но не исходному аллелю. За счет использования данной стратегии несовпадения между протоспейсером и не подвергаемым редактированию аллелем должны способствовать отсутствию никирования за счет sgРНК до реализации редактирования на включающей PAM цепи.

tPERT

[0211] См. определение "РНК-матрица *trans* праймерного редактора (tPERT)".

Временно́й разрыв второй цепи

[0212] В соответствии с используемым здесь значением термин "временно́й разрыв второй цепи" обозначает вариант никирования второй цепи, при котором внесение второго разрыва в не подвергаемую редактированию цепь осуществляется только после внесения требуемого изменения в подвергаемую редактированию цепь. Это обеспечивает возможность избежать одновременных разрывов на обеих цепях, которые могут приводить к двуцепочечным разрывам ДНК. Никирующая вторую цепь направляющая РНК сконструирована для временно́го контроля, так что разрыв второй цепи не осуществляется до реализации внесения требуемого изменения. Это достигается за счет конструирования нРНК со спейсерной последовательностью, которая соответствует только подвергаемой редактированию цепи, но не исходному аллелю. За счет использования данной стратегии несовпадения между протоспейсером и не подвергаемым редактированию аллелем должны способствовать отсутствию никирования за счет sgРНК до реализации редактирования на включающей PAM цепи.

***Trans* праймерное редактирование**

[0213] В соответствии с используемым здесь значением термин "*trans* праймерное редактирование" обозначает модифицированную форму праймерного редактирования, в рамках которой используется разделенная PEgRNA, то есть где PEgRNA разделяется на две отдельные молекулы: sgРНК и РНК-матрицу для *trans* праймерного редактирования (tPERT). sgРНК служит для нацеливания праймерного редактора (или в более общем плане для нацеливания компонента праймерного редактора в виде napDNAbp) на

требуемый геномный участок-мишень, в то время как tPERT используется полимеразой (например обратной транскриптазой) для записи новой последовательности ДНК в локус-мишень, когда tPERT рекрутируется *in trans* к праймерному редактору за счет взаимодействия связывающих доменов, расположенных на праймерном редактора и tPERT. В одном варианте осуществления домены связывания могут включать фрагменты для рекрутирования РНК-белка, например аптамер MS2, расположенный на tPERT, и белок MS2cp, слитый с праймерным редактором. Преимущество *trans* праймерного редактирования заключается в том, что за счет отделения матрицы для синтеза ДНК от направляющей РНК является потенциально возможным использовать матрицы большей длины.

[0214] Вариант осуществления *trans* праймерного редактирования показан на Фиг.3G и 3H. На Фиг.3G показана композиция комплекса *trans* праймерного редактора слева ("комплекс RP-PE:нРНК"), который включает napDNAbp, слитый с каждой полимеразой (например обратной транскриптазой) и rPERT-рекрутирующим белком (например MS2sc), который образует комплекс с направляющей РНК. На Фиг.3G дополнительно показана отдельная молекула tPERT, которая включает элементы плеча удлинения PEgRNA, включая матрицу для синтеза ДНК и последовательность связывания праймера. Молекула tPERT также включает домен рекрутирования РНК-белок (который, в данном случае, представляет собой структуру типа "стебель-петля" и может являться, например, аптамером MS2). Как показано в рамках процесса, изображенного на Фиг.3H, комплекс RP-PE:нРНК связывает и никирует последовательность-мишень ДНК. Затем рекрутирующий белок (RP) рекрутирует tPERT для солокализации с комплексом праймерного редактора, связанным с участком-мишенью ДНК, тем самым обеспечивая связывание участка связывания праймера с праймерной последовательностью на никированной цепи и далее обеспечивая синтез полимеразой (например RT) одиночной цепи ДНК на матрице для синтеза ДНК на 5′ tPERT.

[0215] В то время как tPERT показана на Фиг.3G и Фиг.3H в качестве включающей PBS и матрицу для синтеза ДНК на 5′-конце домена рекрутирования РНК-белка, tPERT в других конфигурациях может быть сконструирована с PBS и матрицей для синтеза ДНК, расположенными на 3′-конце домена рекрутирования РНК-белка. Однако tPERT с 5' удлинением имеет преимущество, заключающееся в том, что синтез одиночной цепи ДНК

будет естественным образом терминирован на 5'-конце tPERT и, таким образом, не будет связан с риском использования какой-либо части домена рекрутирования РНК-белка в качестве матрицы на этапе праймерного редактирования, связанного с синтезом ДНК.

РНК-матрица *trans* праймерного редактора (tPERT)

[0216] В соответствии с используемым здесь значением термин "РНК-матрица *trans* праймерного редактора (tPERT)" обозначает компонент, используемый при *trans* праймерном редактировании, представляющим собой модифицированную версию праймерного редактирования, которая работает на основе разделения PEgRNA на две отдельные молекулы: направляющую РНК и молекулу tPERT. Молекула tPERT программируется для солокализации с комплексом праймерного редактора в участке-мишени ДНК, направляя участок связывания праймера и матрицу для синтеза ДНК к праймерному редактору *in trans*. Например, см. вариант осуществления *trans* праймерного редактора (tPE) на Фиг.3G в виде двухкомпонентной системы, включающей (1) комплекс RP-PE:нРНК и (2) tPERT, которая включает участок связывания праймера и матрицу для синтеза ДНК, присоединенные к домену рекрутирования РНК-белка, при этом компонент в виде RT (рекрутирующего белка) комплекса RP-PE:нРНК рекрутирует tPERT к подвергаемому редактированию участку-мишени, тем самым связывая PBS и матрицу для синтеза ДНК с праймерным редактором *in trans*. Другими словами, tPERT конструируют для включения (полностью или частично) плеча удлинения PEgRNA, которое включает участок связывания праймера и матрицу для синтеза ДНК.

Транзиции

[0217] В соответствии с используемым здесь значением термин "транзиции" обозначает замену пуриновых азотистых оснований (A ↔ G) или замену пиримидиновых азотистых оснований (C ↔ T). Данный класс замен затрагивает азотистые основания одинаковой формы. Описанные здесь композиции и способы способны обеспечивать индуцирование одной или нескольких транзиций в молекуле-мишени ДНК. Описанные здесь композиции и способы также способны обеспечивать индуцирование как транзиций, так и трансверсии в одной и той же молекуле-мишени ДНК. Эти изменения включают A ↔ G, G ↔ A, C ↔ T, или T ↔ C. В контексте двуцепочечной ДНК с парными азотистыми основаниями Уотсона-Крика трансверсии обозначают следующие замены пар оснований: A:T ↔ G:C, G:G ↔ A:T, C:G ↔ T:A, или T:A↔ C:G. Описанные здесь композиции и способы

способны обеспечивать индуцирование одной или нескольких транзиций в молекуле-мишени ДНК. Описанные здесь композиции и способы также способны обеспечивать индуцирование как транзиций, так и трансверсий в одной и той же молекуле-мишени ДНК, а также других нуклеотидных изменений, включая делеции и вставки.

Трансверсии

[0218] В соответствии с используемым здесь значением термин "трансверсии" обозначает замену пуриновых азотистых оснований пиримидиновыми азотистыми основаниями или наоборот и, таким образом, включает взаимную замену азотистых оснований с различающейся формой. Данные изменения включают T ↔ A, T↔ G, C ↔ G, C ↔ A, A ↔ T, A ↔ C, G ↔ C, и G ↔ T. В контексте двуцепочечной ДНК с парными азотистыми основаниями Уотсона-Крика трансверсии обозначают следующие замены пар оснований: T:A ↔ A:T, T:A ↔ G:C, C:G ↔ G:C, C:G ↔ A:T, A:T ↔ T:A, A:T ↔ C:G, G:C ↔ C:G, и G:C ↔ T:A. Описанные здесь композиции и способы способны обеспечивать индуцирование одной или нескольких трансверсий в молекуле-мишени ДНК. Описанные здесь композиции и способы также способны обеспечивать индуцирование как транзиций, так и трансверсий в одной и той же молекуле-мишени ДНК, а также других нуклеотидных изменений, включая делеции и вставки.

Лечение

[0219] Термины "лечение", "лечить" и "лечащий" обозначают клиническое вмешательство, направленное на обращение, облегчение, отсрочку развития или ингибирования прогрессирования заболевания или расстройства, или одного или нескольких его симптомов, как описано здесь. В соответствии с используемым здесь значением термины "лечение", "лечить" и "лечащий" обозначают клиническое вмешательство, направленное на обращение, облегчение, отсрочку развития или ингибирования прогрессирования заболевания или расстройства, или одного или нескольких его симптомов, как описано здесь. В некоторых вариантах осуществления лечение может осуществляться после развития одного или нескольких симптомов и/или после диагностирования заболевания. В других вариантах осуществления лечение может осуществляться при отсутствии симптомов, например для предотвращения или задержки развития симптома или ингибирования развития или прогрессирования заболевания. Например, лечение может осуществляться в отношении уязвимого субъекта до развития

симптомов (например в свете истории симптомов и/или в свете генетических или иных факторов предрасположенности). Лечение может также продолжаться после устранения симптомов, например для предотвращения или замедления их повторного возникновения.

Связанное с тринуклеотидными повторами заболевание

[0220] В соответствии с используемым здесь значением термин "связанное с тринуклеотидными повторами заболевание" (или, альтернативно, "связанное экспансией повторов заболевание" или "связанное с повторной экспансией заболевание") обозначает набор генетических заболеваний, которые вызываются "экспансией тринуклеотидных повторов", которая представляет собой разновидность мутации, при которой определенный тринуклеотид повторяется в определенных генах или интронах. Ранее считалось, что тринуклеотидные повторы являются обычными итерациями в рамках генома, однако в 1990-х данные заболевания были выявлены. Данные на первый взгляд "доброкачественные" последовательности ДНК иногда могут подвергаться экспансии и вызывать заболевание. Заболевания, вызываемые экспансией тринуклеотидных повторов, разделяют некоторые определяющие признаки. Во-первых, мутантные повторы демонстрируют нестабильность как соматической, так и зародышевой линии, и они более часто подвергаются удлинению, а не сжатию при последующих передачах. Во-вторых, более ранний возраст развития и повышающаяся тяжесть фенотипа в последующих поколениях (ожидающаяся) в целом коррелируют с большей длиной повторов. Наконец, происхождение связанного с заболеванием аллеля от родителя зачастую может влиять на ожидание вследствие того, что передача от родителя связана с повышающимся риском экспансии для многих из данных заболеваний.

[0221] Считается, что экспансия триплетов вызывается проскальзыванием при репликации ДНК. Вследствие повторяющейся природы последовательности ДНК в данных участках могут образовываться "выступающие в виде петли" структуры при репликации ДНК с одновременным сохранением связывания комплементарных оснований между исходной цепью и дочерней синтезируемой цепью. Если выступающая в виде петли структура образуется из последовательности на дочерней цепи, это приведет к повышению количества повторов. Однако если выступающая в виде петли структура образуется на исходной цепи, то это приводит к снижению количества повторов. Похоже, что экспансия данных повторов является более частой, чем сжатие. В целом чем больше

экспансия, тем с большей вероятностью вызывается заболевание или повышается тяжесть заболевания. Это свойство приводит к характеристике в виде ожидания при заболеваниях, связанных с тринуклеотидными повторами. Ожидание описывает тенденцию, связанную с тем, что возраст возникновения заболевания понижается, а тяжесть симптомов повышается у последующих поколений семьи вследствие экспансии данных повторов.

[0222] Связанные с нуклеотидными повторами заболевания могут включать заболевания, при которых триплетный повтор присутствует в некодирующем участке (то есть связанное с некодирующими тринуклеотидными повторами заболевание) или в кодирующем участке.

[0223] Описанная здесь система праймерного редактора (PE) может использоваться для лечения связанных с нуклеотидными повторами заболеваний, которые могут включать синдром ломкой X-хромосомы (FRAXA), ломкую XE MR (FRAXE), атаксию Фрейдрейха (FRDA), миотоническую дистрофию (DM), спиноцеребеллярную атаксию 8 типа (SCA8) и спиноцеребеллярную атаксию 12 типа (SCA12), среди прочих.

Праймерное редактирование или "праймерное редактирование (PE)"

[0224] В соответствии с используемым здесь значением термин "праймерное редактирование" или "праймерное редактирование (PE)" обозначает новый подход к редактированию генов с использованием napDNAbp и специализированных направляющих РНК, как описано в настоящем описании, примеры которого приведены в рамках вариантов осуществления на Фиг.1A-1J. TPRT обозначает "мишень-инициированную обратную транскрипцию", поскольку в одном варианте осуществления используется молекула-мишень ДНК для праймирования синтеза цепи ДНК обратной транскриптазой (или другой полимеразой). В различных вариантах осуществления праймерное редактирование осуществляется путем контактирования молекулы-мишени ДНК (для которое требуется внести изменение в нуклеотидной последовательности) с программируемым нуклеиновой кислотой ДНК-связывающим белком (napDNAbp) в виде комплекса с направляющей РНК праймерного редактора. Со ссылкой на Фиг.1E, направляющая РНК праймерного редактора включает удлинение на 3′- или 5′-конце направляющей РНК или во внутримолекулярном положении в направляющей РНК и кодирует требуемое нуклеотидное изменение (например изменение одного нуклеотида, вставку, или делецию). На этапе (a) комплекс в виде napDNAbp/удлиненной нРНК

контактирует с молекулой ДНК, при этом удлиненная нРНК направляет napDNAbp для связывания с локусом-мишенью. На этапе (b) осуществляется разрыв в одной из цепей ДНК локуса-мишени (например с помощью нуклеазы или химического агента), тем самым обеспечивая создание свободного 3′-конца в одной из цепей локуса-мишени. В определенных вариантах осуществления разрыв создается в цепи ДНК, которая соответствует цепи R-петли, то есть цепи, которая не гибридизируется с последовательностью направляющей РНК, то есть "не являющейся мишенью цепи". Однако разрыв может быть осуществлен в любой одной из цепей. То есть разрыв может быть осуществлен в "цепи-мишени" (то есть цепи, гибридизированной со спейсером удлиненной нРНК) или "не являющейся мишенью цепи" (то есть цепи, образующей одноцепочечную часть R-петли, которая является комплементарной цепи-мишени). На этапе (c) 3′-конец цепи ДНК (образованный разрывом) взаимодействует с удлиненной частью направляющей РНК для праймирования обратной транскрипции (то есть "праймированной мишенью RT"). В некоторых вариантах осуществления 3′-конец цепи ДНК гибридизируется с определенным участком связывания праймера на удлиненной части направляющей РНК, то есть "праймирующей последовательностью обратной транскриптазы". На этапе (d) вносится обратная транскриптаза, которая синтезирует одиночную цепь ДНК от 3′-конца праймированного участка к 3′-концу направляющей РНК праймерного редактора. Это образует одноцепочечный ДНК-флэп, включающий требуемое нуклеотидное изменение (например изменение одного основания, вставку, или делецию, или их комбинацию) и являющийся иным образом гомологичным эндогенной ДНК в участке разрыва или являющейся смежной с ним. На этапе (e) napDNAbp и направляющая РНК высвобождаются. Этапы (f) и (g) связаны с коррекцией одноцепочечного ДНК-флэпа, так что требуемое нуклеотидное изменение становится встроенным в локус-мишень. Данный процесс может быть смещен в сторону образования требуемого продукта путем удаления соответствующего 5′ эндогенного ДНК-флэпа, который образуется когда 3′ одноцепочечный ДНК-флэп вносится и гибридизируется с последовательностью эндогенной ДНК. Без связи с теорией, процессы репарации и репликации клеточной эндогенной ДНК обеспечивают коррекцию несовпадающей ДНК для встраивания нуклеотидного изменения (изменений) с образованием требуемого измененного продукта. Процесс также может быть смещен в сторону образования

продукта с "никированием второй цепи", как показано в качестве примера на Фиг.1D. Данный процесс может обеспечивать внесение, по меньшей мере, одного или нескольких из следующих генетических изменений: трансверсий, транзиций, делеций, и вставок.

[0225] Термин "система праймерного редактора (PE)" или "праймерный редактор" или "система PE" или "система редактирования PE" обозначает композиции, используемые в рамках способа геномного редактирования с применением описанной здесь мишень-инициированной обратной транскрипции (TPRT), включая, но без ограничения, napDNAbp, обратные транскриптазы, слитые белки (например включающие napDNAbp и обратные транскриптазы), направляющие РНК праймерного редактора, и комплексы, включающие слитые белки и направляющие РНК праймерного редактора, а также вспомогательные элементы, например компоненты для никирования второй цепи и эндонуклеазы для удаления 5′ эндогенного ДНК-флэпа для обеспечения сдвига процесса праймерного редактирования в сторону образования подвергнутого редактированию продукта.

Выше

[0226] В соответствии с используемым здесь значением термины "выше" и "ниже" представляют собой термины для указания относительного положения, которые определяют линейное положение, по меньшей мере, двух элементов, расположенных в молекуле нуклеиновой кислоты (одноцепочечной или двуцепочечной), которая ориентирована в направлении от 5′ к 3′. В частности, первый элемент расположен выше второго элемента в молекуле нуклеиновой кислоты, если первый элемент расположен где-либо 5' относительно второго элемента. Например, SNP расположен выше Cas9-индуцированного участка разрыва, если SNP расположен на 5' стороне участка разрыва. С другой стороны, первый элемент расположен ниже второго элемента в молекуле нуклеиновой кислоты, если первый элемент расположен где-либо 3' относительно второго элемента. Например, SNP расположен ниже Cas9-индуцированного участка разрыва, если SNP расположен на 3' стороне участка разрыва. Молекула нуклеиновой кислоты может представлять собой ДНК (двуцепочечную или одноцепочечную), РНК (двуцепочечную или одноцепочечную), или гибрид ДНК и РНК. Анализ является таким же для одноцепочечной молекулы нуклеиновой кислоты, что и для двуцепочечной молекулы, поскольку термины "выше" и "ниже" обозначают только одну цепь молекулы

нуклеиновой кислоты, за исключением случаев, когда необходимо выбрать указываемую цепь двуцепочечной молекулы. Зачастую цепь двуцепочечной ДНК, которая может применяться для определения относительности положения, по меньшей мере, двух элементов, представляет собой "смысловую" или "кодирующую" цепь. В генетике "смысловая" цепь представляет собой сегмент в рамках двуцепочечной ДНК, которая проходит от 5' к 3' и которая является комплементарной антисмысловой цепи ДНК, или матричной цепи, которая проходит от 3' к 5'. Таким образом, в качестве примера, азотистое основание SNP находится ниже последовательности промотора в геномной ДНК (которая является двуцепочечной), если азотистое основание SNP располагается на 3' стороне промотора на смысловой или кодирующей цепи.

Вариант

[0227] В соответствии с используемым здесь значением термин "вариант" должен пониматься в качестве обозначающего демонстрирование качеств, которые имеют паттерн, являющийся отличным от находящегося в природе, например вариант Cas9 представляет собой Cas9, включающую одно или несколько изменений в аминокислотных остатках по сравнению с аминокислотной последовательностью Cas9 дикого типа. Термин "вариант" охватывает гомологичные белки, имеющие, по меньшей мере, 75%, или, по меньшей мере, 80%, или, по меньшей мере, 90%, или, по меньшей мере, 95%, или, по меньшей мере, 99% идентичность относительно исходной последовательности и имеющие такую же или по существу такую же функциональную активность или активности, что и исходная последовательность. Термин также охватывает мутанты, усечения, или домены исходной последовательности, которые демонстрируют такую же или по существу такую же функциональную активность или активности, что и исходная последовательность.

Вектор

[0228] В соответствии с используемым здесь значением термин "вектор" обозначает нуклеиновую кислоту, которая может быть модифицирована для кодирования представляющего интерес гена и которая способна проникать в клетку-хозяина, мутировать и реплицироваться в клетке-хозяине, и затем переносить реплицированную форму вектора в другую клетку-хозяин. Примеры подходящих векторов включают вирусные векторы, например ретровирусные векторы или бактериофаги и нитчатый фаг, и

конъюгативные плазмиды. Дополнительные подходящие векторы будут очевидны специалистам в данной области на основе настоящего описания.

Дикий тип

[0229] В соответствии с используемым здесь значением термин "дикий тип" представляет собой термин в данной области, который понимается специалистами в данной области и обозначает обычную форму организма, штамма, гена или характеристику в том виде, в котором она встречается в природе, для различения от мутантных или вариантных форм.

Удаление 5′ эндогенного ДНК-флэпа

[0230] В соответствии с используемым здесь значением термин "удаление 5′ эндогенного ДНК-флэпа" обозначает удаление 5′ эндогенного ДНК-флэпа, который образуется когда синтезированный RT одноцепочечный ДНК-флэп конкурентно вносится и гибридизируется с эндогенной ДНК, вытесняя при этом эндогенную цепь. Удаление данной эндогенной вытесненной цепи может склонять реакцию к образованию требуемого продукта, включающего требуемое нуклеотидное изменение. Собственные ферменты клетки для репарации ДНК могут катализировать удаление или эксцизию 5′ эндогенного флэпа (например флэп-нуклеаза, такая как EXO1 или FEN1). Кроме того, клетки-хозяева могут быть трансформированы для экспрессии одного или нескольких ферментов, которые катализируют удаление указанных 5′ эндогенных флэпов, тем самым склоняя процесс к образованию продукта (например флэп-эндонуклеазы). Флэп-эндонуклеазы известны в данной области и описаны в Patel et al., "Flap endonucleases pass 5′-flaps through a flexible arch using a disorder-thread-order mechanism to confer specificity for free 5′-ends," Nucleic Acids Research, 2012, 40(10): 4507-4519 и Tsutakawa et al., "Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily," Cell, 2011, 145(2): 198-211 (каждый из которых приведен здесь в качестве ссылки).

5′ эндогенный ДНК-флэп

[0231] В соответствии с используемым здесь значением термин "5′ эндогенный ДНК-флэп" обозначает цепь ДНК, расположенную непосредственно ниже индуцированного PE участка разрыва в ДНК-мишени. Никирование цепи ДНК-мишени с помощью PE высвобождает 3′ гидроксильную группу на лежащей выше стороне участка разрыва и 5′ гидроксильную группу на лежащей ниже стороне участка разрыва. Эндогенная цепь, оканчивающаяся 3′ гидроксильной группой, используется для праймирования ДНК-

полимеразы праймерного редактора (например когда ДНК-полимераза представляет собой обратную транскриптазу). Эндогенная цепь на расположенной ниже стороне участка разрыва, которая начинается со свободной 5′ гидроксильной группы, обозначается как "5′ эндогенный ДНК-флэп" и в конечном итоге удаляется и замещается новой синтезированной замещающей цепью (то есть "3′ замещающим ДНК-флэпом"), кодируемой удлинением PEgRNA.

3′ замещающий ДНК-флэп

[0232] В соответствии с используемым здесь значением термин "3′ замещающий ДНК-флэп" или просто "замещающий ДНК-флэп" обозначает цепь ДНК, которая синтезируется праймерным редактором и которая кодируется плечом удлинения PEgRNA праймерного редактора. Более конкретно, 3′ замещающий ДНК-флэп кодируется полимеразной матрицей PEgRNA. 3′ замещающий ДНК-флэп включает ту же последовательность, что и 5′ эндогенный ДНК-флэп за исключением того, что он также содержит отредактированную последовательность (например изменение одного нуклеотида). 3′ замещающий ДНК-флэп отжигается на ДНК-мишени, вытесняя или замещая 5′ эндогенный ДНК-флэп (который может быть подвергнут эксцизии, например с помощью 5′ флэп-эндонуклеазы, такой как FEN1 или EXO1), и затем лигируется для присоединения 3′-конца 3′ замещающего ДНК-флэпа к высвобожденному 5′ гидроксильному концу эндогенной ДНК (высвобожденному после эксцизии 5′ эндогенного ДНК-флэпа), тем самым восстанавливая фосфодиэфирную связь и обеспечивая внесение 3′ замещающего ДНК-флэпа для образования гетеродуплексной ДНК, содержащей одну подвергнутую редактированию цепь и одну не подвергнутую редактированию цепь. Процессы репарации ДНК корректируют гетеродуплекс за счет копирования информации из подвергнутой редактированию цепи в комплементарную цепь, обеспечивая перманентное внесение изменения в ДНК. Данный процесс корректирования может быть дополнительно смещен в сторону завершения путем никирования не подвергнутой редактированию цепи, то есть путем "никирования второй цепи", как описано здесь.

ПОДРОБНОЕ ОПИСАНИЕ НЕКОТОРЫХ ВАРИАНТОВ ОСУЩЕСТВЛЕНИЯ

[0233] Настоящее изобретение относится к новым композициям (например новым PEgRNA и включающим их PE комплексам) и способам для применения праймерного

редактирования (PE) для репарации терапевтических мишеней, например мишеней, идентифицированных в базе данных ClinVar, с помощью PEgRNA, сконструированной с использованием специализированного алгоритма, который описан здесь. Таким образом, настоящее изобретение относится к алгоритму для крупномасштабного предсказывания последовательностей PEgRNA, которые могут использоваться для репарации терапевтических мишеней (например включенных в базу данных ClinVar). Кроме того, настоящее изобретение относится к предсказанным последовательностям для терапевтической PEgRNA, сконструированной с помощью описанного алгоритма, которая может применяться с рамках праймерного редактирования для репарации терапевтических мишеней.

[0234] Описанный здесь алгоритм и предсказанные последовательности PEgRNA в целом относятся к праймерному редактированию. Таким образом, данное изобретение также включает описание различных компонентов и аспектов праймерного редактирования, включая подходящий napDNAbp (например никазу Cas9) и обратные транскриптазы, а также другие подходящие компоненты (например линкеры, NLS) и слитые белки PE, которые могут применяться с описанной здесь терапевтической PEgRNA.

[0235] Использование системы на основе коротких регулярно расположенных группами палиндромных повторов (CRISPR) для геномного редактирования произвело революцию в биологических науках[1-3]. Несмотря на то, что манипулирование генами с помощью CRISPR теперь является обыденным, точное внесение изменений отдельных нуклеотидов остается трудной задачей, несмотря на то, что оно является необходимым для исследования или корректирования большого количества вызывающих заболевания мутаций. Направленная гомологичная репарация (HDR) способна обеспечивать реализацию подобных изменений, но связана с низкой эффективностью (зачастую <5%), потребностью в донорных матрицах для репарации ДНК, и негативными эффектами образования двуцепочечных разрывов ДНК (DSB). Недавно в лаборатории профессора David Liu *et al.* было разработано редактирование оснований, которое обеспечивает достижение эффективного редактирования отдельных нуклеотидов без DSB. В рамках редакторов оснований (BE) система CRISPR комбинируется с ферментами в виде модифицирующей основания дезаминазы для конвертации пар-мишеней оснований C•G или A•T в A•T или G•C, соответственно[4-6]. Несмотря на то, что они уже широко

используются исследователями по всему миру, текущие BE обеспечивают конверсию лишь четырех из двенадцати возможных пар оснований и не способны корректировать малые вставки или делеции. Более того, объем нацеливания при редактировании оснований ограничен редактированием не являющихся мишенями оснований C или A, являющихся смежными с основанием-мишенью ("побочное редактирование"), а также необходимостью наличия последовательности PAM на расстоянии 15±2 bp от основания-мишени. Таким образом, преодоление данных ограничений значительно расширит варианты исследовательского и терапевтического использования редактирования оснований для геномного редактирования.

[0236] В рамках настоящего изобретения предлагается новый подход точного редактирования, который обеспечивает множество преимуществ редактирования оснований, а именно устранение двуцепочечных разрывов и потребности в донорских матрицах для репарации ДНК, в то же время обеспечивая преодоление его основных ограничений. Предложенный описанный здесь подход обеспечивает достижение прямого внесения подвергнутых редактированию цепей ДНК в геномные участки-мишени с помощью мишень-инициированной обратной транскрипции (TPRT). В рамках описанной здесь конструкции направляющая РНК (нРНК) CRISPR сконструирована для включения последовательности матрицы в виде обратной транскриптазы (RT), кодирующей одноцепочечную ДНК, включающую требуемое нуклеотидное изменение. Никированный нуклеазой (Cas9) CRISPR участок-мишень ДНК будет служить в качестве праймера для обратной транскрипции последовательности матрицы на модифицированной нРНК, обеспечивая прямое встраивание любого требуемого нуклеотидного изменения.

[0237] Соответственно, настоящее изобретение частично связано с открытием того, что механизм мишень-инициированной обратной транскрипции (TPRT) может быть использован или адаптирован для осуществления точного геномного редактирования на основе CRISPR/Cas с высокой эффективностью и генетической гибкостью (например как показано в рамках различных вариантов осуществления на Фиг.1A-1G). Авторы предложили здесь использовать продукты слияния в виде napDNAbp-полимеразы (например никазы Cas9, слитой с обратной транскриптазой) для нацеливания на определенную последовательность ДНК с помощью модифицированной направляющей РНК ("удлиненная направляющая РНК" или PEgRNA), осуществления одноцепочечного

разрыва в участке-мишени, и использования никированной ДНК в качестве праймера для синтеза ДНК полимеразой (например обратной транскриптазой) на основе матрицы для синтеза ДНК, которая представляет собой компонент PEgRNA. Новая синтезированная цепь будет являться гомологичной относительно последовательности геномной мишени за исключением включения требуемого нуклеотидного изменения (например изменения отдельного нуклеотида, делеции, или вставки, или их комбинации). Новая синтезированная цепь ДНК также может быть обозначена как одноцепочечный ДНК-флэп, который будет конкурировать за гибридизацию с комплементарной гомологичной эндогенной цепью ДНК, тем самым вытесняя соответствующую эндогенную цепь. Корректирование данного гибридизированного промежуточного продукта может включать удаление полученного вытесненного флэпа из эндогенной ДНК (например с помощью эндонуклеазы 5′-концевого ДНК-флэпа, FEN1), лигирование синтезированного одноцепочечного ДНК-флэпа с ДНК-мишенью, и ассимилирование требуемого нуклеотидного изменения в результате процессов репарации и/или репликации клеточной ДНК. Поскольку синтез ДНК на основе матриц обеспечивает точность на уровне отдельных нуклеотидов, объем данного подхода является очень широким и, как представляется, он может использоваться в рамках множества вариантов применения в фундаментальной науке и терапии.

I. Терапевтические PEgRNA

[0238] Описанная здесь система праймерного редактора (PE) включает применение любых подходящих направляющих РНК или PEgRNA праймерного редактора. Авторы обнаружили, что механизм мишень-инициированной обратной транскрипции (TPRT) может быть использован или адаптирован для осуществления точного и универсального геномного редактирования на основе CRISPR/Cas путем применения специальным образом сконфигурированной направляющей РНК, включающей матрицу для синтеза ДНК, которая кодирует требуемое нуклеотидное изменение, с помощью полимеразы (например обратной транскриптазы). В рамках изобретения данная специальным образом сконфигурированная направляющая РНК обозначается как "направляющая РНК праймерного редактора" (или PEgRNA), поскольку матрица для синтеза ДНК может быть представлена в виде удлинения стандартной или традиционной молекулы направляющей

РНК. Изобретение охватывает любую подходящую конфигурацию или строение направляющей РНК праймерного редактора.

[0239] В различных вариантах осуществления изобретение предусматривает PEgRNA согласно SEQ ID NO: 1-135514 и 813085-880462, сконструированную с помощью описанного здесь алгоритма в соответствии с записями из базы данных ClinVar.

[0240] В различных других вариантах осуществления примеры PEgRNA, сконструированной в соответствии с базой данных ClinVar с использованием описанного здесь алгоритма, включены в Перечень последовательностей, который является частью настоящего описания. Перечень последовательностей включает последовательности полной PEgRNA SEQ ID NO: 1-135514 и 813085-880462. Каждая из данных полных PEgRNA состоит из спейсера (SEQ ID NO: 135515 – 271028 и 880463-947840) и плеча удлинения (SEQ ID NO: 271029 – 406542 и 947841-1015218). Кроме того, каждая PEgRNA включает ядро нРНК, например как определено согласно SEQ ID NO: 1361579-1361580. Каждое плечо удлинения согласно SEQ ID NO: 271029 – 406542 и 947841-1015218 дополнительно состоит из участка связывания праймера (SEQ ID NO: 406543 – 542056 и 1015219-1082596), матрицы для редактирования (SEQ ID NO.: 542057 – 677570 и 1082597-1149974), и плеча гомологии (SEQ ID NO.: 677571 – 813084 и 1149975-1217352). PEgRNA может необязательно включать 5′-концевой модификаторный участок и/или 3′-концевой модификаторный участок. PEgRNA может также включать сигнал терминации обратной транскрипции (например SEQ ID NO: 1361560-1361566) на 3'-конце PEgRNA. Изобретение охватывает конструирование и применение всех данных последовательностей.

[0241] На Фиг.3А показан один вариант осуществления направляющей РНК праймерного редактора (обозначаемой как "PEgRNA" или "удлиненная нРНК"), которая может применяться в рамках описанной здесь системы праймерного редактора (PE), в рамках которой традиционная направляющая РНК (зеленая часть) включает спейсер и участок в виде ядра нРНК, который связывается с napDNAbp. В данном варианте осуществления направляющая РНК включает удлиненный сегмент РНК на 5′-конце, то есть 5′ удлинение. В данном варианте осуществления 5′ удлинение включает матрицу для синтеза ДНК, участок связывания праймера, и необязательную последовательность линкера длиной 5-20 нуклеотидов. Как показано на Фиг.1А, участок связывания праймера гибридизируется со

свободным 3′-концом, который образуется после осуществления разрыва в не являющейся мишенью цепи R-петли, тем самым обеспечивая праймирование полимеразы (например обратной транскриптазы) для полимеризации ДНК в направлении от 5′ к 3′.

[0242] На Фиг.3B показан другой вариант осуществления направляющей РНК праймерного редактора, которая может применяться в рамках описанной здесь системы праймерного редактора (PE), где традиционная направляющая РНК (зеленая часть) включает последовательность спейсера длиной ~20 nt и ядро нРНК, которое связывается с napDNAbp. В данном варианте осуществления направляющая РНК включает удлиненный сегмент РНК на 3′-конце, то есть 3′ удлинение. В данном варианте осуществления 3′ удлинение включает матрицу для синтеза ДНК и участок связывания праймера. Как показано на Фиг.1B, участок связывания праймера гибридизируется со свободным 3′-концом, который образуется после осуществления разрыва в не являющейся мишенью цепи R-петли, тем самым обеспечивая праймирование полимеразы для полимеризации ДНК в направлении от 5′ к 3′.

[0243] На Фиг.3C показан другой вариант осуществления удлиненной направляющей РНК, которая может применяться в рамках описанной здесь системы праймерного редактора (PE), где традиционная направляющая РНК (зеленая часть) включает спейсер длиной ~20 nt и ядро нРНК, которое связывается с napDNAbp. В данном варианте осуществления направляющая РНК включает удлиненный сегмент РНК во внутримолекулярном положении в ядре нРНК, то есть внутримолекулярное удлинение. В данном варианте осуществления внутримолекулярное удлинение включает матрицу для синтеза ДНК и участок связывания праймера. Участок связывания праймера гибридизируется со свободным 3′-концом, который образуется после осуществления разрыва в не являющейся мишенью цепи R-петли, тем самым обеспечивая праймирование полимеразы для полимеризации ДНК в направлении от 5′ к 3′.

[0244] В одном варианте осуществления положение внутримолекулярного удлинения РНК соответствует спейсеру направляющей РНК. В другом варианте осуществления положение внутримолекулярного удлинения РНК соответствует ядру нРНК. В другом варианте осуществления положение внутримолекулярного удлинения РНК соответствует любому положению в рамках молекулы направляющей РНК за исключением спейсера или положения, которое нарушает спейсер.

[0245] В одном варианте осуществления внутримолекулярное удлинение РНК подвергается вставке ниже 3′-конца спейсера. В другом варианте осуществления внутримолекулярное удлинение РНК подвергается вставке, по меньшей мере, на 1 нуклеотид, по меньшей мере, на 2 нуклеотида, по меньшей мере, на 3 нуклеотида, по меньшей мере, на 4 нуклеотида, по меньшей мере, на 5 нуклеотидов, по меньшей мере, на 6 нуклеотидов, по меньшей мере, на 7 нуклеотидов, по меньшей мере, на 8 нуклеотидов, по меньшей мере, на 9 нуклеотидов, по меньшей мере, на 10 нуклеотидов, по меньшей мере, на 11 нуклеотидов, по меньшей мере, на 12 нуклеотидов, по меньшей мере, на 13 нуклеотидов, по меньшей мере, на 14 нуклеотидов, по меньшей мере, на 15 нуклеотидов, по меньшей мере, на 16 нуклеотидов, по меньшей мере, на 17 нуклеотидов, по меньшей мере, на 18 нуклеотидов, по меньшей мере, на 19 нуклеотидов, по меньшей мере, на 20 нуклеотидов, по меньшей мере, на 21 нуклеотид, по меньшей мере, на 22 нуклеотида, по меньшей мере, на 23 нуклеотида, по меньшей мере, на 24 нуклеотида, по меньшей мере, на 25 нуклеотидов ниже 3′-конца спейсера.

[0246] В других вариантах осуществления внутримолекулярное удлинение РНК подвергается вставке в нРНК, что соответствует части направляющей РНК, соответствующей или включающей tracrРНК, которая связывает и/или взаимодействует с белком Cas9 или его эквивалентом (то есть другим napDNAbp). Вставка внутримолекулярного удлинения РНК предпочтительно не нарушает или минимально нарушает взаимодействие между частью в виде tracrРНК и napDNAbp.

[0247] Длина удлинения РНК может представлять собой любую подходящую длину. В различных вариантах осуществления удлинение РНК может иметь длину, по меньшей мере, 5 нуклеотидов, по меньшей мере, 6 нуклеотидов, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, по меньшей мере, 20 нуклеотидов, по меньшей мере, 21 нуклеотидов, по меньшей мере, 22 нуклеотида, по меньшей мере, 23 нуклеотида, по меньшей мере, 24 нуклеотидов, по меньшей мере, 25 нуклеотидов, по меньшей мере, 30 нуклеотидов, по меньшей мере, 40 нуклеотидов, по

меньшей мере, 50 нуклеотидов, по меньшей мере, 60 нуклеотидов, по меньшей мере, 70 нуклеотидов, по меньшей мере, 80 нуклеотидов, по меньшей мере, 90 нуклеотидов, по меньшей мере, 100 нуклеотидов, по меньшей мере, 200 нуклеотидов, по меньшей мере, 300 нуклеотидов, по меньшей мере, 400 нуклеотидов, или, по меньшей мере, 500 нуклеотидов.

[0248] Матрица для синтеза ДНК (например последовательность RT матрицы) также может иметь любую подходящую длину. Например, матрица для синтеза ДНК (например последовательность RT матрицы) может иметь длину, по меньшей мере, 3 нуклеотида, по меньшей мере, 4 нуклеотида, по меньшей мере, 5 нуклеотидов, по меньшей мере, 6 нуклеотидов, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, по меньшей мере, 20 нуклеотидов, по меньшей мере, 30 нуклеотидов, по меньшей мере, 40 нуклеотидов, по меньшей мере, 50 нуклеотидов, по меньшей мере, 60 нуклеотидов, по меньшей мере, 70 нуклеотидов, по меньшей мере, 80 нуклеотидов, по меньшей мере, 90 нуклеотидов, по меньшей мере, 100 нуклеотидов, по меньшей мере, 200 нуклеотидов, по меньшей мере, 300 нуклеотидов, по меньшей мере, 400 нуклеотидов, или, по меньшей мере, 500 нуклеотидов.

[0249] В других вариантах осуществления последовательность участка связывания праймера обратной транскрипции имеет длину, по меньшей мере, 3 нуклеотида, по меньшей мере, 4 нуклеотида, по меньшей мере, 5 нуклеотидов, по меньшей мере, 6 нуклеотидов, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, по меньшей мере, 20 нуклеотидов, по меньшей мере, 30 нуклеотидов, по меньшей мере, 40 нуклеотидов, по меньшей мере, 50 нуклеотидов, по

меньшей мере, 60 нуклеотидов, по меньшей мере, 70 нуклеотидов, по меньшей мере, 80 нуклеотидов, по меньшей мере, 90 нуклеотидов, или, по меньшей мере, 100 нуклеотидов.

[0250] В других вариантах осуществления необязательный линкер или спейсер имеет длину, по меньшей мере, 3 нуклеотида, по меньшей мере, 4 нуклеотида, по меньшей мере, 5 нуклеотидов, по меньшей мере, 6 нуклеотидов, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, по меньшей мере, 20 нуклеотидов, по меньшей мере, 30 нуклеотидов, по меньшей мере, 40 нуклеотидов, по меньшей мере, 50 нуклеотидов, по меньшей мере, 60 нуклеотидов, по меньшей мере, 70 нуклеотидов, по меньшей мере, 80 нуклеотидов, по меньшей мере, 90 нуклеотидов, по меньшей мере, 100 нуклеотидов, по меньшей мере, 200 нуклеотидов, по меньшей мере, 300 нуклеотидов, или, по меньшей мере, 400 нуклеотидов.

[0251] Матрица для синтеза ДНК (например последовательность RT матрицы) в определенных вариантах осуществления кодирует молекулу одноцепочечной ДНК, которая является гомологичной не являющейся мишенью цепи (и, таким образом, комплементарной соответствующему участку цепи-мишени), но включает одно или несколько нуклеотидных изменений. Указанное нуклеотидное изменение может включать одно или несколько нуклеотидных изменений одного основания, одну или несколько делеций, одну или несколько вставок, и их комбинации.

[0252] Как изображено на Фиг.1E, синтезированный продукт в виде одноцепочечной ДНК матрицы для синтеза ДНК (например последовательности RT матрицы) является гомологичным не являющейся мишенью цепи и содержит одно или несколько нуклеотидных изменений. Продукт в виде одноцепочечной ДНК матрицы для синтеза ДНК (например последовательности RT матрицы) гибридизируется равновесным образом с комплементарной последовательностью цепи-мишени, тем самым вытесняя гомологичную последовательность эндогенной цепи-мишени. Вытесненная эндогенная цепь может быть обозначена в некоторых вариантах осуществления как 5′ эндогенный ДНК-флэп (например, см. Фиг.1C). Данный 5′ эндогенный ДНК-флэп может быть удален

с помощью 5′ флэп-эндонуклеазы (например FEN1), при этом продукт в виде одноцепочечной ДНК, являющийся теперь гибридизированным с эндогенной цепью-мишенью, может быть лигирован, тем самым создавая несовпадение между эндогенной последовательностью и новой синтезированной цепью. Несовпадение может быть подвергнуто коррекции с помощью внутренних процессов репарации и/или репликации ДНК клетки.

[0253] В различных вариантах осуществления нуклеотидная последовательность матрицы для синтеза ДНК (например последовательности RT матрицы) соответствует нуклеотидной последовательность не являющейся мишенью цепи, которая становится вытесненной в виде 5′ флэпа и которая перекрывает подвергаемый редактированию участок.

[0254] В различных вариантах осуществления направляющих РНК праймерного редактора матрица для синтеза ДНК может кодировать одноцепочечный ДНК-флэп, который комплементарен последовательности эндогенной ДНК, примыкающей к участку разрыва, при этом одноцепочечный ДНК-флэп включает требуемое нуклеотидное изменение. Одноцепочечный ДНК-флэп может вытеснять эндогенную одноцепочечную ДНК в участке разрыва. Вытесненная эндогенная одноцепочечная ДНК в участке разрыва может иметь 5′-конец и образовывать эндогенный флэп, который может быть подвергнут эксцизии клеткой. В различных вариантах осуществления эксцизия 5′-концевого эндогенного флэпа может обеспечивать стимулирование образования продукта, поскольку удаление 5′-концевого эндогенного флэпа стимулирует гибридизирование одноцепочечного 3′ ДНК-флэпа с соответствующей комплементарной цепью ДНК и встраивание или ассимилирование требуемого нуклеотидного изменения, которое несет одноцепочечный 3′ ДНК-флэп, в ДНК-мишени.

[0255] В различных вариантах осуществления направляющих РНК праймерного редактора клеточная репарация одноцепочечного ДНК-флэпа приводит к внесению требуемого нуклеотидного изменения, тем самым обеспечивая образование требуемого продукта.

[0256] В других вариантах осуществления требуемое нуклеотидное изменение вносится в окне редактирования, которое соответствует приблизительно от -5 до +5 относительно участка разрыва, или приблизительно от -10 до +10 относительно участка разрыва, или приблизительно от -20 до +20 относительно участка разрыва, или приблизительно от -30

до +30 относительно участка разрыва, или приблизительно от -40 до +40 относительно участка разрыва, или приблизительно от -50 до +50 относительно участка разрыва, или приблизительно от -60 до +60 относительно участка разрыва, или приблизительно от -70 до +70 относительно участка разрыва, или приблизительно от -80 до +80 относительно участка разрыва, или приблизительно от -90 до +90 относительно участка разрыва, или приблизительно от -100 до +100 относительно участка разрыва, приблизительно от -200 до +200 относительно участка разрыва.

[0257] В различных аспектах направляющие РНК праймерного редактора представляют собой модифицированные версии направляющей РНК. Направляющие РНК могут являться встречающимися в природе, экспрессируемыми из кодирующей нуклеиновой кислоты, или химически синтезированными. В данной области хорошо известны способы получения или иным образом синтезирования направляющих РНК и определения подходящей последовательности направляющей РНК, включая спейсер, который взаимодействует и гибридизируется с цепью-мишенью представляющего интерес геномного участка-мишени.

[0258] В различных вариантах осуществления конкретные аспекты конструирования последовательности направляющей РНК будут зависеть от нуклеотидной последовательности представляющего интерес геномного участка-мишени (то есть требуемого подвергаемого редактированию участка) и типа napDNAbp (например белка Cas9), присутствующего в рамках описанной здесь системы праймерного редактора (PE), среди других факторов, таких как расположение последовательностей PAM, процентная величина содержания G/C в последовательности-мишени, показатель участков микрогомологии, вторичные структуры и так далее.

[0259] В целом, направляющая последовательность представляет собой любую полинуклеотидную последовательность, имеющую достаточную комплементарность относительно полинуклеотидной последовательности-мишени для гибридизирования с последовательностью-мишенью и прямого специфичного в плане последовательности связывания napDNAbp (например Cas9, гомолога Cas9, или варианта Cas9) с последовательностью-мишенью. В некоторых вариантах осуществления степень комплементарности между направляющей последовательностью и соответствующей ей последовательностью-мишенью при оптимальном выравнивании с помощью подходящего

алгоритма выравнивания составляет приблизительно или более чем приблизительно 50%, 60%, 75%, 80%, 85%, 90%, 95%, 99% или более. Оптимальное выравнивание может быть определено с помощью любого подходящего алгоритма для выравнивания последовательностей, неограничивающие примеры которых включают алгоритм Смита-Уотермана, алгоритм Нидлмана-Вунша, алгоритмы на основе преобразования Барроуза-Уилера (например выравнивание Барроуза-Уилера), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies, ELAND (Illumina, San Diego, Calif.), SOAP (доступен на soap.genomics.org.cn), и Maq (доступен на maq.sourceforge.net) В некоторых вариантах осуществления направляющая последовательность имеет длину приблизительно или более чем приблизительно 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, или более нуклеотидов.

[0260] В некоторых вариантах осуществления направляющая последовательность имеет длину менее чем приблизительно 75 50, 45, 40, 35, 30, 25, 20, 15, 12 или менее нуклеотидов. Способность направляющей последовательности обеспечивать специфичное в плане последовательности связывание редактора оснований с последовательностью-мишенью может быть проанализирована с помощью любого подходящего анализа. Например, компоненты редактора оснований, включая подвергаемую тестированию направляющую последовательность, могут быть перенесены в клетку-хозяина, имеющую соответствующую последовательность-мишень, например с помощью трансфекции с использованием векторов, кодирующих компоненты описанного здесь редактора оснований, с последующим анализом предпочтительного расщепления в рамках последовательности-мишени, например с помощью анализа Surveyor, как описано здесь. Аналогичным образом, расщепление полинуклеотидной последовательности-мишени может быть проанализировано в аналитической пробирке путем добавления последовательности-мишени, компонентов редактора оснований, включая подвергаемую тестированию направляющую последовательность и исходную направляющую последовательность, отличную от подвергаемой тестированию направляющей последовательности, и сравнения связывания или скорости расщепления в последовательности-мишени между реакциями подвергаемой тестированию и исходной направляющей последовательности. Также возможны другие анализы, при этом они будут очевидны специалистам в данной области.

[0261] Направляющая последовательность может быть выбрана для нацеливания на любую последовательность-мишень. В некоторых вариантах осуществления последовательность-мишень представляет собой последовательность в рамках генома клетки. Примеры последовательностей-мишеней включают последовательности, которые являются уникальными в геноме-мишени. Например для Cas9 S. pyogenes уникальная последовательность-мишень в геноме может включать участок-мишень Cas9 в виде MMMMMMMMNNNNNNNNNNNNXGG (SEQ ID NO: 1361548), где NNNNNNNNNNNNXGG (SEQ ID NO: 1361549) (N представляет собой A, G, T, или C; а X может быть любым) встречается один раз в геноме. Уникальная последовательность-мишень в геноме может включать участок-мишень Cas9 S. pyogenes в виде MMMMMMMMMNNNNNNNNNNNXGG (SEQ ID NO: 1361550), где NNNNNNNNNNNXGG (SEQ ID NO: 1361551) (N представляет собой A, G, T, или C; а X может быть любым) встречается один раз в геноме. Для CRISPR 1 Cas9 S. thermophilus уникальная последовательность-мишень в геноме может включать участок-мишень Cas9 в виде MMMMMMMMNNNNNNNNNNNNXXAGAAW (SEQ ID NO: 1361552), где NNNNNNNNNNNNXXAGAAW (SEQ ID NO: 1361553) (N представляет собой A, G, T, или C; X может быть любым; а W представляет собой A или T) встречается один раз в геноме. Уникальная последовательность-мишень в геноме может включать участок-мишень CRISPR 1 Cas9 S. thermophilus в виде MMMMMMMMMNNNNNNNNNNNNXXAGAAW (SEQ ID NO: 1361554), где NNNNNNNNNNNNXXAGAAW (SEQ ID NO: 1361555) (N представляет собой A, G, T, или C; X может быть любым; а W представляет собой A или T) встречается один раз в геноме. Например для Cas9 S. pyogenes уникальная последовательность-мишень в геноме может включать участок-мишень Cas9 в виде MMMMMMMMNNNNNNNNNNNNXGGXG (SEQ ID NO: 1361556), где NNNNNNNNNNNNXGGXG (SEQ ID NO: 1361557) (N представляет собой A, G, T, или C; а X может быть любым) встречается один раз в геноме. Уникальная последовательность-мишень в геноме может включать участок-мишень Cas9 S. pyogenes в виде MMMMMMMMMNNNNNNNNNNNXGGXG (SEQ ID NO: 1361558), где NNNNNNNNNNNXGGXG (SEQ ID NO: 1361559) (N представляет собой A, G, T, или C; а X может быть любым) встречается один раз в геноме. В каждой из данных последовательностей "M" может представлять собой A, G, T, или C, при этом он не должен рассматриваться при идентификации последовательности в качестве уникальной.

[0262] В некоторых вариантах осуществления направляющая последовательность выбрана для снижения величины вторичной структуры в направляющей последовательности. Вторичная структура может быть определена с помощью любого подходящего алгоритма полинуклеотидного фолдинга. Некоторые программы основаны на вычислении минимальной свободной энергии Гиббса. Пример одного подобного алгоритма представляет собой mFold, как описано Zuker и Stiegler (Nucleic Acids Res. 9 (1981), 133-148). Другой пример алгоритма фолдинга представляет собой интернет-вебсервер RNAfold, разработанный в Институте теоретической химии Венского университета, основанный на использовании алгоритма предсказания центроидной структуры (см., например, A. R. Gruber *et al.*, 2008, Cell 106(1): 23-24; и PA Carr and GM Church, 2009, Nature Biotechnology 27(12): 1151-62). Дополнительные алгоритмы описаны в заявке на патент США 61/836080; широкая ссылка BI-2013/004A, которая приведена здесь в качестве ссылки.

[0263] В целом, последовательность tracr mate включает любую последовательность, которая имеет достаточную комплементарность относительно последовательности tracr для промотирования одного или нескольких из: (1) эксцизии направляющей последовательности, фланкируемой последовательностями tracr mate, в клетке, содержащей соответствующую последовательность tracr; и (2) образования комплекса в последовательности-мишени, при этом комплекс включает последовательность tracr mate, гибридизированную с последовательностью tracr. В целом, степень комплементарности соответствует оптимальному выравниванию последовательности tracr mate и последовательности tracr в рамках длины более короткой из данных двух последовательностей. Оптимальное выравнивание может быть определено с помощью любого подходящего алгоритма выравнивания и может дополнительно учитывать вторичные структуры, например самокомплементарность в последовательности tracr или последовательности tracr mate. В некоторых вариантах осуществления степень комплементарности между последовательностью tracr и последовательностью tracr mate в рамках длины более короткой из них при оптимальном выравнивании составляет приблизительно или более чем приблизительно 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97,5%, 99%, или более. В некоторых вариантах осуществления последовательность tracr имеет длину приблизительно или более чем приблизительно 5, 6,

7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50 или более нуклеотидов. В некоторых вариантах осуществления последовательность tracr и последовательность tracr mate содержатся в рамках одного транскрипта, так что гибридизация между ними двумя продуцирует транскрипт, имеющий вторичную структуру, например шпильку. Предпочтительные образующие петли последовательности для применения в рамках структур в виде шпилек имеют длину четыре нуклеотида и, наиболее предпочтительно, имеют последовательность GAAA. Однако могут применяться более длинные или короткие последовательности петель, а также альтернативные последовательности. Последовательности предпочтительно включают нуклеотидный триплет (например AAA) и дополнительный нуклеотид (например C или G). Примеры образующих петли последовательностей включают CAAA и AAAG. В одном варианте осуществления изобретения транскрипт или транскрибированная полинуклеотидная последовательность имеет, по меньшей мере, две или более шпильки. В предпочтительных вариантах осуществления транскрипт имеет две, три, четыре или пять шпилек. В дополнительном варианте осуществления транскрипт имеет не более пяти шпилек. В некоторых вариантах осуществления отдельный транскрипт дополнительно включает последовательности терминации транскрипции; она предпочтительно представляет собой последовательность polyT, например шесть нуклеотидов T. Дополнительные неограничивающие примеры отдельных полинуклеотидов, включающих направляющую последовательность, последовательность tracr mate, и последовательность tracr, являются следующими (перечислены по направлению от 5' к 3'), где "N" обозначает основание направляющей последовательности, первый блок строчных букв обозначает последовательность tracr mate, а второй блок строчных букв обозначает последовательность tracr, и конечная последовательность poly-T соответствует терминатору транскрипции: (1) NNNNNNNNgtttttgtactctcaagatttaGAAAtaaatcttgcagaagctacaaagataaggcttcatgccgaaatcaacaccctgtcattttatggcagggtgttttcgttatttaaTTTTTT (SEQ ID NO: 1361560); (2) NNNNNNNNNNNNNNNNNNgtttttgtactctcaGAAAtgcagaagctacaaagataaggcttcatgccgaaatcaacaccctgtcattttatggcagggtgttttcgttatttaaTTTTTT (SEQ ID NO: 1361561); (3) NNNNNNNNNNNNNNNNNNNNgtttttgtactctcaGAAAtgcagaagctacaaagataaggctt

catgccgaaatcaacaccctgtcattttatggcagggtgtTTTTT (SEQ ID NO: 1361562); (4) NNNNNNNNNNNNNNNNNNNNgttttagagctaGAAAtagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgcTTTTTT (SEQ ID NO: 1361563); (5) NNNNNNNNNNNNNNNNNNNNgttttagagctaGAAATAGcaagttaaaataaggctagtccgttatcaacttgaaaaagtgTTTTTTT (SEQ ID NO: 1361564 и (6) NNNNNNNNNNNNNNNNNNNNgttttagagctagAAATAGcaagttaaaataaggctagtccgttatcaTTTTTTTT (SEQ ID NO: 1361565). В некоторых вариантах осуществления последовательности от (1) до (3) используются в комбинации с Cas9 из CRISPR1 S. thermophilus. В некоторых вариантах осуществления последовательности от (4) до (6) используются в комбинации с Cas9 из S. pyogenes. В некоторых вариантах осуществления последовательность tracr представляет собой отдельный транскрипт от транскрипта, включающего последовательность tracr mate.

[0264] Специалистам в данной области будет очевидно, что для нацеливания на любые слитые белки, включающие домен Cas9 и связывающий одноцепочечную ДНК белок, как описано здесь, для нацеливания на участок, например участок, включающий подвергаемую редактированию точечную мутацию, обычно необходимо совместно экспрессировать слитый белок вместе с направляющей РНК, например sgРНК. Как более подробно описано в других разделах, направляющая РНК обычно включает каркас tracrРНК, обеспечивающий связывание Cas9, и направляющую последовательность, которая придает специфичность в плане последовательности слитому белку Cas9:фермент/домен для редактирования нуклеиновой кислоты.

[0265] В некоторых вариантах осуществления направляющая РНК включает структуру 5′-[направляющая последовательность]-guuuuagagcuagaaauagcaaguuaaaauaaaggcuaguccguuaucaacuugaaaaaguggcaccgagucggugcuuuuu-3′ (SEQ ID NO: 1361566), где направляющая последовательность включает последовательность, которая комплементарна последовательности-мишени. Направляющая последовательность обычно имеет длину 20 нуклеотидов. Последовательности подходящих направляющих РНК для нацеливания слитых белков Cas9:фермент/домен для редактирования нуклеиновой кислоты на определенные геномные участки-мишени будут очевидны специалистам в данной области на основе

настоящего описания. Такие последовательности подходящих направляющих РНК обычно включают направляющие последовательности, которые комплементарны нуклеотидной последовательности на расстоянии до 50 нуклеотидов выше или ниже подвергаемого редактированию нуклеотида-мишени. Некоторые примеры последовательностей направляющих ДНК, подходящих для нацеливания любых указанных здесь слитых белков на определенные последовательности-мишени, приведены здесь. Дополнительные направляющие последовательности хорошо известны в данной области и могут применяться с описанными здесь редакторами оснований.

[0266] В других вариантах осуществления PEgRNA могут включать PEgRNA, изображенные в виде структуры, показанной на Фиг.27, которая включает направляющую РНК и 3' плечо удлинения.

[0267] На Фиг.27 приведена структура варианта осуществления предусмотренной здесь PEgRNA, которая может быть сконструирована в соответствии с методологией, определенной в рамках Примера 2. PEgRNA включает три основных компонента в направлении от 5′ к 3′, а именно: спейсер, ядро нРНК, и плечо удлинения на 3′-конце. Плечо удлинения может быть дополнительно разделено на следующие структурные элементы в направлении от 5′ к 3′, а именно: участок связывания праймера (A), матрицу для редактирования (B), и плечо гомологии (C). Кроме того, PEgRNA может включать необязательный 3′-концевой модификаторный участок (e1) и необязательный 5′-концевой модификаторный участок (e2). Кроме того, PEgRNA может включать сигнал терминации транскрипции на 3′-конце PEgRNA (не показан). Данные структурные элементы дополнительно определены здесь. Изображение структуры PEgRNA не является ограничивающим и охватывает вариации в плане расположения элементов. Например, необязательные модификаторы последовательности (e1) и (e2) могут быть расположены внутри или между любыми из показанных участков без ограничения в виде расположения на 3′- или 5′-концах.

[0268] В других вариантах осуществления PEgRNA могут включать PEgRNA, изображенные в виде структуры, показанной на Фиг.28, которая включает направляющую РНК и 5' плечо удлинения.

[0269] На Фиг.28 приведена структура другого варианта осуществления предусмотренной здесь PEgRNA, которая может быть сконструирована в соответствии с методологией,

определенной в рамках Примера 2. PEgRNA включает три основных компонента в направлении от 5′ к 3′, а именно: спейсер, ядро нРНК, и плечо удлинения на 3′-конце. Плечо удлинения может быть дополнительно разделено на следующие структурные элементы в направлении от 5′ к 3′, а именно: участок связывания праймера (A) (SEQ ID NO: 406543-542056 и 1015219-1082596), матрицу для редактирования (B) (SEQ ID NO: 542057-677570 и 1082597-1149974), и плечо гомологии (C) (SEQ ID NO: 677571-813084 и 1149975-1217352). Кроме того, PEgRNA может включать необязательный 3′-концевой модификаторный участок (e1) и необязательный 5′-концевой модификаторный участок (e2). Кроме того, PEgRNA может включать сигнал терминации транскрипции на 3′-конце PEgRNA (не показан). Данные структурные элементы дополнительно определены здесь. Изображение структуры PEgRNA не является ограничивающим и охватывает вариации в плане расположения элементов. Например, необязательные модификаторы последовательности (e1) и (e2) могут быть расположены внутри или между любыми из показанных участков без ограничения в виде расположения на 3′- или 5′-концах.

[0270] PEgRNA также могут включать дополнительные усовершенствования конструкции, связанные с модификацией свойств и/или характеристик PEgRNA, тем самым повышая эффективность праймерного редактирования. В различных вариантах осуществления данные усовершенствования могут относится к одной или нескольким из ряда различных категорий, включая, но без ограничения: (1) конструкции для обеспечения эффективной экспрессии функциональных PEgRNA из не относящихся к полимеразе III (pol III) промоторов, которые будут обеспечивать экспрессию более длинных PEgRNA без обременительных требований в плане последовательности; (2) усовершенствования ядра, Cas9-связывающего каркаса PEgRNA, которые могут повысить эффективность; (3) модификации PEgRNA для повышения процессирования RT, обеспечивающие вставку более длинных последовательностей в геномных локусах-мишенях; и (4) добавление мотивов РНК на 5'- или 3'-концы PEgRNA, которые повышают стабильность PEgRNA, стимулируют процессирование RT, предотвращают неправильный фолдинг PEgRNA, или рекрутируют дополнительные факторы, являющиеся важными для геномного редактирования.

[0271] В одном варианте осуществления PEgRNA может быть сконструирована с pol III промоторами для повышения экспрессии более длинной PEgRNA с более крупными

плечами удлинения. sgРНК обычно экспрессируются из промотора мяРНК U6. Данный промотор рекрутирует pol III для экспрессии ассоциированной РНК и может применяться для экспрессии коротких РНК, которые остаются в ядре. Однако pol III не является высокопроцессивной и не способна экспрессировать РНК с длиной более чем несколько сотен нуклеотидов на уровнях, требуемых для эффективного геномного редактирования. Кроме того, pol III может вызывать остановку или терминирование на длинных участках в виде U, что потенциально ограничивает разнообразие цепей, которые могут быть вставлены с помощью PEgRNA. Были проанализированы другие промоторы, которые рекрутируют полимеразу II (например pCMV) или полимеразу I (например промотор мяРНК U1) на их способность экспрессировать более длинные sgРНК. Однако данные промоторы обычно частично транскрибируются, что будет приводить к дополнительной последовательности 5' спейсера в экспрессируемой PEgRNA, что, как было показано, приводит к заметному снижению активности Cas9:sgРНК в зависимости от участка. Кроме того, в то время как транскрибированные из pol III PEgRNA могут просто приводить к терминации на участке из 6-7 U, PEgRNA, транскрибированные из pol II или pol I, будут требовать другого сигнала терминации. Зачастую такие сигналы также приводят к полиаденилированию, которое приведет к нежелательному транспорту PEgRNA из ядра. Аналогичным образом, РНК, экспрессируемые из pol II промоторов, например pCMV, обычно являются 5'-кэпированными, что также приводит к их экспорту из ядра.

[0272] Ранее Rinn и коллеги провели скрининг ряда платформ экспрессии на предмет продуцирования sgРНК, меченных длинными некодирующими РНК (днкРНК)[183]. Данные платформы включают РНК, экспрессированные из pCMV, которые приводят к терминации в элементе ENE из нкРНК MALAT1 человека[184], элементе PAN ENE из KSHV[185], или 3' боксе из sgРНК U1[186]. Следует отметить, что нкРНК MALAT1 и PAN ENE образуют тройные спирали, защищающие polyA-хвост [184, 187]. Данные конструкты также могут повышать стабильность РНК. Предполагается, что данные системы экспрессии также обеспечат возможность экспрессии более длинных PEgRNA.

[0273] Кроме того, был разработан ряд способов для расщепления части pol II промотора, которая будет транскрибирована в виде части PEgRNA, обеспечивая добавление либо саморасщепляемого рибозима, например рибозимов типа "головка молотка" (hammerhead)[188], "пистолет" (pistol)[189], "топор" (hatchet)[189], "шпилька" (hairpin)[190], VS[191],

"твистер"[192], или "твистер-систер" (twister sister)[192], или других саморасщепляемых элементов для процессирования транскрибированной направляющей, или шпильки, которая распознается Csy4[193] и также ведет к процессированию направляющей. Кроме того, существует гипотеза, что встраивание нескольких мотивов ENE может приводить к повышенной экспрессии PEgRNA и стабильности, как ранее было продемонстрировано для РНК и элемента KSHV PAN[185]. Также ожидается, что циркуляризация PEgRNA в форме циркулярной интронной РНК (циРНК) также может приводить к повышенной экспрессии РНК и стабильности, а также ядерной локализации[194].

[0274] В различных вариантах осуществления PEgRNA может включать различные указанные выше элементы, как показано на примере следующей последовательности.

[0275] Неограничивающий пример 1 – платформа экспрессии PEgRNA, состоящая из pCMV, шпильки Csy4, PEgRNA, и MALAT1 ENE

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTT
ACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAA
TAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTA
TTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATT
GACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTC
CTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTA
CATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTC
AATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCC
CATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGT
GAACCGTCAGATCGTTCACTGCCGTATAGGCAGGGCCCAGACTGAGCACGTGAGTTTTAGAGCT
AGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTC
CTCTGCCATCAAAGCGTGCTCAGTCTGTTTTAGGGTCATGAAGGTTTTTCTTTTCCTGAGAAAA
CAACACGTATTGTTTTCTCAGGTTTTGCTTTTTGGCCTTTTTCTAGCTTAAAAAAAAAAAAAGC
AAAAGATGCTGGTGGTTGGCACTCCTGGTTTCCAGGACGGGGTTCAAATCCCTGCGGCGTCTTT
GCTTTGACT (SEQ ID NO: 1361567)

[0276] Неограничивающий пример 2 – платформа экспрессии PEgRNA, состоящая из pCMV, шпильки Csy4, PEgRNA, и PAN ENE

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTT
ACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAA
TAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTA
TTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATT
GACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTC
CTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTA
CATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTC
AATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCC
CATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGT
GAACCGTCAGATCGTTCACTGCCGTATAGGCAGGGCCCAGACTGAGCACGTGAGTTTTAGAGCT
AGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTC

CTCTGCCATCAAAGCGTGCTCAGTCTGTTTTGTTTTGGCTGGGTTTTTCCTTGTTCGCACCGGA
CACCTCCAGTGACCAGACGGCAAGGTTTTTATCCCAGTGTATATTGGAAAAACATGTTATACTT
TTGACAATTTAACGTGCCTAGAGCTCAAATTAAACTAATACCATAACGTAATGCAACTTACAAC
ATAAATAAAGGTCAATGTTTAATCCATAAAAAAAAAAAAAAAAAAA (SEQ ID NO:
1361568)

[0277] Неограничивающий пример 3 – платформа экспрессии PEgRNA, состоящая из pCMV, шпильки Csy4, PEgRNA, и 3xPAN ENE

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTT
ACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAA
TAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTA
TTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATT
GACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTC
CTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTA
CATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTC
AATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCC
CATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGT
GAACCGTCAGATCGTTCACTGCCGTATAGGCAGGGCCCAGACTGAGCACGTGAGTTTTAGAGCT
AGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTC
CTCTGCCATCAAAGCGTGCTCAGTCTGTTTTGTTTTGGCTGGGTTTTTCCTTGTTCGCACCGGA
CACCTCCAGTGACCAGACGGCAAGGTTTTTATCCCAGTGTATATTGGAAAAACATGTTATACTT
TTGACAATTTAACGTGCCTAGAGCTCAAATTAAACTAATACCATAACGTAATGCAACTTACAAC
ATAAATAAAGGTCAATGTTTAATCCATAAAAAAAAAAAAAAAAAAAACACACTGTTTTGGCTGG
GTTTTTCCTTGTTCGCACCGGACACCTCCAGTGACCAGACGGCAAGGTTTTTATCCCAGTGTAT
ATTGGAAAAACATGTTATACTTTTGACAATTTAACGTGCCTAGAGCTCAAATTAAACTAATACC
ATAACGTAATGCAACTTACAACATAAATAAAGGTCAATGTTTAATCCATAAAAAAAAAAAAAAA
AAAATCTCTCTGTTTTGGCTGGGTTTTTCCTTGTTCGCACCGGACACCTCCAGTGACCAGACGG
CAAGGTTTTTATCCCAGTGTATATTGGAAAAACATGTTATACTTTTGACAATTTAACGTGCCTA
GAGCTCAAATTAAACTAATACCATAACGTAATGCAACTTACAACATAAATAAAGGTCAATGTTT
AATCCATAAAAAAAAAAAAAAAAAA (SEQ ID NO: 1361569)

[0278] Неограничивающий пример 4 – платформа экспрессии PEgRNA, состоящая из pCMV, шпильки Csy4, PEgRNA, и 3′ бокса

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTT
ACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAA
TAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTA
TTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATT
GACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTC
CTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTA
CATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTC
AATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCC
CATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGT
GAACCGTCAGATCGTTCACTGCCGTATAGGCAGGGCCCAGACTGAGCACGTGAGTTTTAGAGCT
AGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTC
CTCTGCCATCAAAGCGTGCTCAGTCTGTTTGTTTCAAAAGTAGACTGTACGCTAAGGGTCATAT
CTTTTTTTGTTTGGTTTGTGTCTTGGTTGGCGTCTTAAA (SEQ ID NO: 1361570)

[0279] Неограничивающий пример 5 - платформа экспрессии PEgRNA, состоящая из pU1, шпильки Csy4, PEgRNA, и 3′ бокса

CTAAGGACCAGCTTCTTTGGGAGAGAACAGACGCAGGGGCGGGAGGGAAAAAGGGAGAGGCAGA
CGTCACTTCCCCTTGGCGGCTCTGGCAGCAGATTGGTCGGTTGAGTGGCAGAAAGGCAGACGGG
GACTGGGCAAGGCACTGTCGGTGACATCACGGACAGGGCGACTTCTATGTAGATGAGGCAGCGC
AGAGGCTGCTGCTTCGCCACTTGCTGCTTCACCACGAAGGAGTTCCCGTGCCCTGGGAGCGGGT
TCAGGACCGCTGATCGGAAGTGAGAATCCCAGCTGTGTGTCAGGGCTGGAAAGGGCTCGGGAGT
GCGCGGGGCAAGTGACCGTGTGTGTAAAGAGTGAGGCGTATGAGGCTGTGTCGGGGCAGAGGCC
CAAGATCTCAGTTCACTGCCGTATAGGCAGGGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGA
AATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTC
TGCCATCAAAGCGTGCTCAGTCTGTTTCAGCAAGTTCAGAGAAATCTGAACTTGCTGGATTTTT
GGAGCAGGGAGATGGAATAGGAGCTTGCTCCGTCCACTCCACGCATCGACCTGGTATTGCAGTA
CCTCCAGGAACGGTGCACCCACTTTCTGGAGTTTCAAAAGTAGACTGTACGCTAAGGGTCATAT
CTTTTTTTGTTTGGTTTGTGTCTTGGTTGGCGTCTTAAA (SEQ ID NO: 1361571)

[0280] В различных других вариантах осуществления PEgRNA может быть усовершенствована путем введения усовершенствований в последовательности каркаса или ядра. Это может быть осуществлено путем внесения известных

[0281] Ядро, представляющее собой Cas9-связывающий каркас PEgRNA, может быть вероятно усовершенствовано для повышения активности PE. Некоторые подобные подходы уже были продемонстрированы. Например, первый парный элемент каркаса (P1) содержит парный элемент GTTTT-AAAAC. Было показано, что такие повторы T приводят к остановке pol III и преждевременной терминации РНК транскрипта. Было показано, что разумная мутация одной из пар T-A на пару G-C в этой части P1 повышает активность sgРНК, что свидетельствует о том, что данный подход также может применяться для PEgRNA[195]. Кроме того, также было показано, что повышение длины P1 стимулирует фолдинг sgРНК и ведет к повышенной активности[195], что свидетельствует о том, что оно является другой возможностью повысить активность PEgRNA. Примеры усовершенствований ядра могут включать:

[0282] PEgRNA, содержащую удлинение P1 длиной 6 nt

GGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGCTCATGAAAATGAGCTAGCAAGTTAAAATAA
GGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAGCGTGCTC
AGTCTGTTTTTTT (SEQ ID NO: 1361572)

[0283] PEgRNA, содержащую мутацию T-A на G-C в P1

GGCCCAGACTGAGCACGTGAGTTTGAGAGCTAGAAATAGCAAGTTTAAATAAGGCTAGTCCGTT
ATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAGCGTGCTCAGTCTGTTTTTT
T (SEQ ID NO: 1361573)

[0284] В различных других вариантах осуществления PEgRNA может быть усовершенствована путем внесения модификаций в участок матрицы для редактирования. С увеличением размера вставки, матрицей которого служит PEgRNA, она с большей вероятностью будет расщепляться эндонуклеазами, подвергаться спонтанному гидролизу или фолдингу во вторичные структуры, не способные подвергаться обратной транскрипции за счет RT или нарушающие фолдинг каркаса PEgRNA и последующее связывание Cas9-RT. Соответственно, является вероятным, что для реализации крупных вставок может быть необходимым модифицирование матрицы PEgRNA, например для вставок целых генов. Некоторые стратегии для этого включают встраивание модифицированных нуклеотидов в синтетическую или полусинтетическую PEgRNA, которые делают РНК более устойчивой к расщеплению или гидролизу или менее подверженной включению ингибиторных вторичных структур[196]. Подобные модификации могут включать 8-аза-7-деазагуанозин, который будет снижать содержание вторичной структуры РНК в последовательностях с высоким содержанием G; "заблокированные" нуклеиновые кислоты (LNA), которые снижают расщепление и повышают определенные виды вторичной структуры РНК; модификации в виде 2'-O-метил, 2'-фтор, или 2'-O-метоксиэтокси, которые повышают стабильность РНК. Такие модификации также могут быть включены в другие части PEgRNA для повышения стабильности и активности. В качестве альтернативы или дополнительно, матрица PEgRNA может быть сконструирована таким образом, что она как кодирует требуемый продукт в виде белка, так и с большей вероятностью включает простые вторичные структуры, которые способны подвергаться обратному фолдингу за счет RT. Такие простые структуры будут действовать в качестве термодинамического стока, делая менее вероятным появление более сложных структур, предотвращающих обратную транскрипцию. Наконец, также является возможным разделить матрицу на две отдельные PEgRNA. При такой конструкции PE будет использоваться для инициирования транскрипции, а также рекрутирования отдельной матричной РНК к являющемуся мишенью участку с помощью РНК-связывающего белка, слитого с Cas9, или элемента распознавания РНК на самой PEgRNA, например аптамера MS2. RT может либо напрямую связываться с данной отдельной матричной РНК, или инициировать обратную транскрипцию на исходной PEgRNA перед заменой на вторую матрицу. Такой подход может обеспечивать

возможность осуществления длинных вставок за счет как предотвращения неправильного фолдинга PEgRNA при добавлении длинной матрицы, так и отсутствия необходимости в диссоциации Cas9 из генома для реализации длинных вставок, что вероятно может ингибировать длинные вставки на основе PE.

(iv) Внесение дополнительных мотивов РНК на 5′- или 3′-концах

[0285] В других вариантах осуществления PEgRNA может быть усовершенствована путем внесения дополнительных мотивов РНК на 5′ и 3′-концах PEgRNA. Некоторые подобные мотивы, например PEN ENE из KSHV и ENE из MALAT1, были описаны выше в качестве возможных способов терминации экспрессии длинных PEgRNA из не относящихся к pol III промоторов. Эти элементы образуют тройные спирали РНК, которые окружают polyA-хвост, что приводит к их удержанию в ядре[184,187]. Однако за счет образования сложных структур на 3′-конце PEgRNA, которые закрывают концевой нуклеотид, данные структуры также вероятно будут помогать предотвращать опосредованное экзонуклеазой расщепление PEgRNA.

[0286] Другие структурные элементы, вставленные на 3′-конце, также могут повышать стабильность РНК, хотя и без обеспечения терминации из не относящихся к pol III промоторов. Такие мотивы могут включать шпильки или квадруплексы РНК, которые будут закрывать 3′-конец[197], или саморасщепляющиеся рибозимы, например HDV, которые будут приводить к образованию 2'-3′-циклического фосфата на 3′-конце, а также потенциально делать PEgRNA менее подверженной расщеплению экзонуклеазами[198]. Индуцирование циклизации PEgRNA за счет неполного сплайсинга - для образования циРНК - также может повышать стабильность PEgRNA и приводить к удержанию PEgRNA в ядре[194].

[0287] Дополнительные мотивы РНК также могут повышать процессирование RT или повышать активность PEgRNA за счет повышения связывания RT с дуплексом ДНК-РНК. Добавление нативной последовательности, связанной RT в ее родственном ретровирусном геноме, может повышать активность RT[199]. Это может включать нативный участок связывания праймера (PBS), полипуриновый тракт (PPT), или целующиеся петли, принимающие участие в димеризациии ретровирусного генома и инициировании транскрипции[199].

[0288] Добавление мотивов димеризации, таких как целующиеся петли или пары тетрапетля GNRA/рецептор тетрапетли[200], на 5′- и 3′-концах PEgRNA также может приводить к эффективной циркуляризации PEgRNA, что повышает стабильность. Кроме того, предполагается, что добавление данных мотивов может обеспечивать физическое разделение спейсера и праймера PEgRNA, предотвращать закрытие спейсера, ухудшающее активность PE. Короткие 5′ удлинения PEgRNA, которые образуют малую опорную шпильку в спейсерном участке, также могут благоприятным образом конкурировать против участка отжига PEgRNA, связывающего спейсер. Наконец, целующиеся петли также могут использоваться для рекрутирования других матричных РНК в участок генома и обеспечения обмена RT активностью от одной РНК к другой. Примеры усовершенствований включают, но без ограничения:

[0289] Продукт слияния PEgRNA-HDV

GGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTT
ATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAGCGTGCTCAGTCTGGGCCGG
CATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTTCGGCATGGCGAATGGGACTT
TTTTT (SEQ ID NO: 1361574)

[0290] Целующуюся петлю PEgRNA-MMLV

GGTGGGAGACGTCCCACCGGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTA
AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAGC
TTCGACCGTGCTCAGTCTGGTGGGAGACGTCCCACCTTTTTTT (SEQ ID NO: 1361575)

[0291] Целующуюся петлю PEgRNA-VS рибозим

GAGCAGCATGGCGTCGCTGCTCACGGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGC
AAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCAT
CAAAGCTTCGACCGTGCTCAGTCTCCATCAGTTGACACCCTGAGGTTTTTTT (SEQ ID NO:
1361576)

[0292] PEgRNA-GNRA тетрапетля/рецептор тетрапетли

GCAGACCTAAGTGGUGACATATGGTCTGGGCCCAGACTGAGCACGTGAGTTTTAGAGCTAUACG
TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTUACGAAGTGGGACCGAGTCGGTCCTCTG
CCATCAAAGCTTCGACCGTGCTCAGTCTGCATGCGATTAGAAATAATCGCATGTTTTTTT
(SEQ ID NO: 1361577)

[0293] PEgRNA матрицу с заменой вторичного продукта слияния в виде РНК-HDV

TCTGCCATCAAAGCTGCGACCGTGCTCAGTCTGGTGGGAGACGTCCCACCGGCCGGCATGGTCC
CAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTTCGGCATGGCGAATGGGACTTTTTTT
(SEQ ID NO: 1361578)

[0294] Каркас PEgRNA может быть дополнительно усовершенствован с помощью направленной эволюции аналогично тому, как был усовершенствован SpCas9 и редакторы оснований. Направленная эволюция может повышать степень распознавания PEgRNA

Cas9 или подвергнутыми эволюции вариантами Cas9. Кроме того, является возможным, что различные последовательности каркаса PEgRNA будут оптимальными в рамках различных геномных локусов, обеспечивая повышение активности PE в рассматриваемом участке, снижая нецелевые виды активности, или осуществляя оба из данных действий. Наконец, эволюция каркасов PEgRNA, к которым были добавлены другие мотивы РНК, будет практически наверняка повышать активность слитой PEgRNA по сравнению с не подвергавшейся эволюции слитой РНК. Например, эволюция аллостерических рибозимов, состоящих из аптамеров c-di-GMP-I и рибозимов типа "головка молотка" (hammerhead), приводила к значительному повышению активности[202], что свидетельствует о том, что эволюция также будет повышать активность продуктов слияния PEgRNA и рибозимов типа "головка молотка" (hammerhead). Кроме того, в то время как Cas9 в настоящее время в целом не совместим с 5′ удлинением sgРНК, направленная эволюция вероятно обеспечит реализацию мутаций, которые устранят данную несовместимость, обеспечивая возможность использования дополнительных мотивов РНК.

[0295] Настоящее изобретение охватывает любые подобные подходы к дальнейшему усовершенствованию эффективности описанных здесь систем праймерного редактирования.

II. Алгоритм и способ конструирования терапевтической PEgRNA

[0296] Как описано здесь, авторы открыли и определили, что праймерное редактирование с использованием PEgRNA может использоваться для внесения широкого ряда изменений нуклеотидов, включая вставки (любой длины, включая целые гены или кодирующие белки участки), делеции (любой длины), и коррекции патогенных мутаций. Однако на сегодняшний день пока не существуют способы для определения и/или предсказания структур PEgRNA, включая определение различных компонентов PEgRNA, таких как спейсер, ядро нРНК, и плечо удлинения (и компонентов удлинения, как описано здесь). Авторы разработали компьютеризированные способы для определения PEgRNA, включая определение структур удлиненной нРНК. Каждая структура удлиненной нРНК может быть определена на основе входного аллеля (например соответствующего патогенной мутации), выходного аллеля (например соответствующего скорректированной последовательности дикого типа), и слитого белка (например системы CRISPR для праймерного редактирования, включая мотив PAM и соответствующее положение

разрывов праймерных редакторов). Различие между входным аллелем и выходным аллелем соответствует требуемому изменению (например изменению одного нуклеотида, вставке, делеции, и/или им подобным). Определенные структуры могут созданы и использованы для осуществления редактирования оснований для изменения входного аллеля с получением выходного аллеля, как описано далее здесь.

[0297] Фиг.31 представляет собой блок-схему, показывающую пример высокоуровневого компьютеризированного способа 3100 для определения структуры удлиненной нРНК в соответствии с некоторыми вариантами осуществления. На этапе 3102 вычислительное устройство (например вычислительное устройство 3400, описанное на Фиг.34) осуществляет доступ к данным, связанным с входным аллелем, выходным аллелем, и слитым белком, который включает программируемый нуклеиновой кислотой ДНК-связывающий белок и обратную транскриптазу. В то время как этап 3102 описывает доступ ко всем трем из входного аллеля, выходного аллеля, и слитого белка в рамках одного этапа, это приведено для целей иллюстрирования, при этом является очевидным, что доступ к таким данным может осуществляться с помощью одного или нескольких этапов без выхода за рамки объема описанных здесь способов. Осуществление доступа к данным может включать получение данных, хранение данных, доступ к базе данных и/или им подобные.

[0298] На этапе 3104 вычислительное устройство определяет структуру удлиненной нРНК на основе входного аллеля, выходного аллеля, и слитого белка, доступ к которым осуществляется на этапе 3102. Структура удлиненной нРНК конструируется для ассоциирования со слитым белком для изменения входного аллеля с получением выходного аллеля. Слитый белок при образовании комплекса с удлиненной нРНК способен связываться с последовательностью-мишенью ДНК, которая включает цепь-мишень, в которой осуществляется изменение, и комплементарную не являющуюся мишенью цепь. Как описано здесь, входной аллель может соответствовать патогенной мутации ДНК, а выходной аллель может соответствовать скорректированной последовательности ДНК.

[0299] Изменение входного аллеля с получением выходного аллеля может включать изменение одного нуклеотида, вставку одного или нескольких нуклеотидов, делецию одного или нескольких нуклеотидов, и/или любое другое изменение, предусмотренное для

получения выходного аллеля. В частности, примеры классов изменений, которые могут быть индуцированы одиночной PEgRNA, включают замещения одного нуклеотида, вставки от 1 nt до приблизительно 40 nt, делеции от 1 nt до приблизительно 30 nt, и их комбинации. Например, праймерное редактирование может поддерживать изменения данных типов от положения спейсера -3 (то есть непосредственно 3' разрыва) до положения спейсера +27 (например 30 nt 3' разрыва во входном аллеле). Также могут использоваться другие положения. Например, изменения в положении спейсера -4 могут быть осуществлены с помощью системы на основе SpCas9 в рамках праймерного редактирования (например те, которые могут быть вызваны за счет случайного расщепления RuvC между положениями спейсера -5 и -4). Тип изменения, длина изменений в nt, и/или положение изменения может представлять собой конфигурируемый параметр (параметры), которые могут использоваться в рамках компьютеризированных способов для определения структур удлиненной нРНК.

[0300] Как описано на основе Фиг.3А-3В и Фиг.27-28, удлиненная нРНК может включать различные компоненты, например спейсер для удлиненной нРНК, который является комплементарным нуклеотидной последовательности-мишени во входном аллеле, каркас нРНК для взаимодействия со слитым белком, и удлинение. Со ссылкой на этап 3104, вычислительное устройство определяет один или несколько из спейсера, каркаса нРНК, и удлинения. В некоторых вариантах осуществления в то время как способы могут включать определение любой комбинации из спейсера, каркаса нРНК, и/или удлинения, в некоторых вариантах осуществления известен один или несколько подобных компонентов и/или аспектов таких компонентов (например являются предварительно определенными, предварительно указанными, фиксированными, и так далее), и, таким образом, они могут не определяться в рамках части этапа 3104.

[0301] Как описано здесь, удлинение нРНК может включать различные компоненты. Например, как показано на Фиг.3А-3В и 27-28, удлинение может включать одну или несколько из RT матрицы (которая включает RT матрицу для редактирования и плечо гомологии), участок связывания праймера, сигнал терминации RT, необязательный 5'-концевой модификаторный участок, и необязательный 3'-концевой модификаторный участок. Фиг.32 представляет собой блок-схему, показывающую пример компьютеризированного способа 3200 для определения компонентов структуры

удлиненной нРНК, включая компоненты удлинения, в соответствии с некоторыми вариантами осуществления. Является очевидным, что Фиг.32 предназначена для иллюстрирования и, таким образом, способы, применяемые для определения удлиненной нРНК, могут включать больше или меньше этапов, чем те, которые показаны на Фиг.32.

[0302] На этапе 3202 вычислительное устройство определяет набор протоспейсеров, которые совместимы с мотивом PAM выбранной системы CRISPR во входном аллеле на обеих цепях. В некоторых вариантах осуществления вычислительное устройство определяет начальный набор протоспейсеров и отфильтровывает протоспейсеры, соответствующие положения разрывов которых не совместимы с праймерным редактированием выходного аллеля, для получения набора оставшихся протоспейсеров-кандидатов. Например вычислительное устройство может определять, что протоспейсер является несовместимым, поскольку разрыв располагается на 3' стороне требуемого изменения цепи. В качестве другого примера, вычислительное устройство может определять, что расстояние между разрывом и требуемым изменением является слишком большим (например больше, чем определенное пользователем пороговое значение, например 30 nt, 35 nt, и так далее).

[0303] На этапе 3204 вычислительное устройство выбирает протоспейсер из набора определенных протоспейсеров. На этапе 3206 вычислительное устройство определяет спейсерная последовательность и последовательность матрицы для редактирования с помощью последовательности протоспейсера входного аллеля, положение разрыва, и последовательность требуемого изменения. Спейсер может включать нуклеотидную последовательность длиной приблизительно 20 нуклеотидов.

[0304] На этапе 3208 вычислительное устройство выбирает один ли несколько наборов параметров, при этом каждый набор параметров включает значение длины участка связывания праймера (например которое может варьироваться в плане количества nt, например приблизительно от 8 до 17 nt), длины плеча гомологии (например которое может варьироваться в плане количества nt, например приблизительно от 2 до 33 nt), и последовательность каркаса нРНК. Например, последовательность каркаса нРНК может представлять собой GTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC (SEQ ID NO: 1361579),

GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCA CCGAGTCGGTGC (SEQ ID NO: 1361580), и/или другие последовательности каркаса нРНК, например последовательности каркаса нРНК, которые сохраняют вторичную структуру РНК дикого типа.

[0305] На этапе 3210 вычислительное устройство выбирает набор параметров, определенных на этапе 3208. На этапе 3212 вычислительное устройство определяет плечо гомологии, последовательность участка связывания праймера, и каркас нРНК с помощью выбранного набора параметров. На этапе 3214 вычислительное устройство затем образует полученную последовательность PEgRNA путем конкатенирования спейсера, каркаса нРНК, плеча удлинения PEgRNA (которое включает плечо гомологии и матрицу для редактирования). Кроме того, плечо удлинения может включать сигнал терминации, который представляет собой последовательность, которая вызывает терминацию обратной транскрипции. Такая терминаторная последовательность может включать, например, TTTTTTGTTTT (SEQ ID NO: 1361581). В некоторых вариантах осуществления плечо удлинения PEgRNA может рассматриваться в качестве включающего сигнал терминации. В других вариантах осуществления плечо удлинения PEgRNA может рассматриваться в качестве не включающего сигнал терминации, но вместо этого где плечо удлинения присоединено к сигналу терминации в качестве элемента, расположенного за пределами плеча удлинения.

[0306] Способ 3200 переходит к этапе 3216, при этом вычислительное устройство определяет наличие других наборов параметров. В случае их наличия способ переходит к этапу 3210 и вычислительное устройство выбирает другой набор параметров. В случае их отсутствия способ переходит к этапу 3218, при этом вычислительное устройство определяет наличие других протоспейсеров. В случае их наличия способ переходит обратно к этапу 3204 и вычислительное устройство выбирает другой протоспейсер из набора протоспейсеров. В случае их отсутствия способ переходит к этапу 3220 и завершается.

[0307] Как описано здесь, матрица для синтеза ДНК (например последовательность RT матрицы) удлинения включает требуемое нуклеотидное изменение для изменения входного аллеля с получением выходного аллеля и включает RT матрицу для редактирования (например определенную на этапе 3206) и плечо гомологии (например

определенное на этапе 3212). Как также описано здесь, матрица для синтеза ДНК (например последовательность RT матрицы) кодирует одноцепочечный ДНК-флэп, который комплементарен последовательности эндогенной ДНК, примыкающей к участку разрыва. Одноцепочечный ДНК-флэп включает требуемое нуклеотидное изменение (например изменение одного нуклеотида, одну или несколько вставок нуклеотидов, одну или несколько делеций нуклеотидов, и/или им подобные). В рамках некоторых вариантов редактирования оснований одноцепочечный ДНК-флэп может гибридизироваться с последовательностью эндогенной ДНК, примыкающей к участку разрыва, для внесения требуемого нуклеотидного изменения. В рамках некоторых вариантов редактирования оснований одноцепочечный ДНК-флэп вытесняет последовательность эндогенной ДНК, примыкающую к участку разрыва. Клеточная репарация одноцепочечного ДНК-флэпа может приводить к внесению требуемого нуклеотидного изменения для образования требуемого продукта в виде выходного аллеля. Матрица для синтеза ДНК (например последовательность RT матрицы) может иметь различное количество нуклеотидов, при этом ее длина может варьироваться приблизительно от 7 до 34 нуклеотидов.

[0308] Хотя это не показано на Фиг.32, вычислительное устройство может быть выполнено с возможностью определять другие компоненты удлиненной нРНК. Например, в некоторых вариантах осуществления вычислительное устройство выполнено с возможностью определять сигнал терминации RT, являющегося смежным относительно RT матрицы. В некоторых вариантах осуществления вычислительное устройство может быть выполнено с возможностью определять первый модификатор, являющийся смежным относительно сигнала терминации RT. В некоторых вариантах осуществления вычислительное устройство выполнено с возможностью определять второй модификатор, являющийся смежным относительно участка связывания праймера.

[0309] Компоненты удлиненной нРНК могут быть сгруппированы в рамках различных конфигураций, например показанных на Фиг.3А-3В и Фиг.27-28. Например, со ссылкой на Фиг.3А, удлинение расположено на 5'-конце структуры удлиненной нРНК, при этом спейсер расположен 3' относительно удлинения и 5' относительно ядра нРНК. В качестве другого примера, со ссылкой на Фиг.3В, спейсер расположен на 5' конце структуры удлиненной нРНК (и 5' относительно ядра нРНК), а удлинение расположено на 3'-конце структуры удлиненной нРНК (и 3' относительно ядра нРНК).

[0310] В некоторых вариантах осуществления вычислительное устройство осуществляет доступ к базе данных, которая включает набор входных аллелей и ассоциированных с ними выходных аллелей. Например, вычислительное устройство может осуществлять доступ к базе данных, предоставленной ClinVar, которая включает сотни тысяч мутаций, каждая из которых включает аллель, соответствующий патогенной мутации, и аллель, соответствующих скорректированной последовательности дикого типа. Данные способы могут применяться для определения одной или нескольких структур удлиненной нРНК для каждой записи в базе данных. Фиг.33 представляет собой блок-схему, показывающую пример компьютеризированного способа 3300 для определения наборов структур удлиненной нРНК для каждой связанной с мутацией записи в базе данных, в соответствии с некоторыми вариантами осуществления. На этапе 3302 вычислительное устройство осуществляет доступ к базе данных (например базе данных ClinVar), которая включает набор связанных с мутацией записей, каждая из которых включает входной аллель, соответствующий мутации, и выходной аллель, соответствующей скорректированной последовательности дикого типа.

[0311] На этапе 3304 вычислительное устройство осуществляет доступ к набору из одного или нескольких слитых белков. В некоторых вариантах осуществления данные способы могут включать генерирование наборов структур удлиненной нРНК для одного слитого белка и/или для комбинации различных слитых белков (например для различных белков Cas9). Вычислительное устройство может быть выполнено с возможностью осуществлять доступ к данным, связанным с множеством слитых белков, и возможностью создавать набор структур удлиненной нРНК для каждого слитого белка (например белка Cas9-NG и белка SpCas9), как описано здесь.

[0312] На этапе 3306 вычислительное устройство осуществляет выбор слитого белка из набора слитых белков. На этапе 3308 вычислительное устройство осуществляет выбор связанной с мутацией записи из набора записей в базе данных. Вычислительное устройство может быть выполнено с возможностью, например, осуществлять итерацию по каждой записи в базе данных и создавать набор структур удлиненной нРНК для записи (например один набор для определенного слитого белка и/или несколько наборов для каждого из множества слитых белков). В некоторых вариантах осуществления вычислительное устройство может быть выполнено с возможностью генерировать

структуры удлиненной нРНК для подмножества записей в базе данных, например предварительно определенного набора, набора мутаций с наиболее высокой значимостью (например связанных с известными терапевтическими преимуществами), и/или им подобных. В некоторых вариантах осуществления в случае, если база данных включает записи, которые не являются совместимыми с некоторыми слитыми белками для праймерного редактирования, то вычислительное устройство может быть выполнено с возможностью определять то, какие записи в базе данных являются совместимыми с праймерным редактированием, с помощью слитого белка, выбранного на этапе 3304, и осуществлять выбор записей, которые являются совместимыми со слитым белком, выбранным на этапе 3308.

[0313] На этапе 3310 вычислительное устройство определяет набор из одной или нескольких структур удлиненной нРНК с помощью описанных здесь способов. Способ переходит от этапа 3310 к этапу 3312, при этом вычислительное устройство определяет наличие дополнительных записей в базе данных. В случае их наличия вычислительное устройство переходит обратно к этапу 3308 и осуществляет выбор другой записи. В случае их отсутствия способ переходит к этапу 3314 и определяет наличие других слитых белков. В случае их наличия вычислительное устройство переходит обратно к этапу 3306 и осуществляет выбор другого слитого белка. В случае их отсутствия вычислительное устройство переходит к этапу 3316 и осуществляет завершение способа 3300.

[0314] В некоторых вариантах осуществления данные способы могут обеспечивать конструирование PEgRNA с удлинениями нРНК, которые содержат некомплементарные последовательности, например некомплементарные последовательности, расположенные 5' относительно плеча гомологии, 3' относительно участка связывания праймера, или и те, и другие. Например, некомплементарные последовательности могут быть сконструированы для реализации взаимодействия на основе целующихся петель для функционирования в качестве защитной шпильки для стабильности РНК, и/или им подобных.

[0315] В некоторых вариантах осуществления PEgRNA может быть сконструирована с помощью стратегий на основе приоритизации в рамках нескольких конструкций-кандидатов. Например, данные способы могут быть модифицированы для устранения удлинений PEgRNA, где самый крайний 5' нуклеотид представляет собой цитозин

(например вследствие прерывания нативных взаимодействий нуклеотид-белок в комплексе sgРНК:Cas9). В качестве другого примера, данные способы могут включать применение инструментов для предсказания вторичной структуры РНК для выбора предпочтительной длины PBS, длины флэпа, и/или им подобных на основе других параметров удлиненной нРНК, например протоспейсера, требуемого изменения, и/или им подобных.

[0316] Пример реализации описанных здесь компьютеризированных способов для определения структур удлиненной нРНК является следующим:

```
# Python 3
# b_design_PEgRNA .py
from __future__ import division
import _config
import sys, os, fnmatch, datetime, subprocess
sys.path.append('/home/unix/maxwshen/')
import numpy as np
from collections import defaultdict
from mylib import util, compbio
import pandas as pd

# Параметры по умолчанию
inp_dir = _config.OUT_PLACE + 'a_annotate/'
NAME = util.get_fn(__file__)
out_dir = _config.OUT_PLACE + NAME + '/'
util.ensure_dir_exists(out_dir)

SPLIT = None

# Гиперпараметры
grna_nick_pos = 17
grna_len = 20

max_dist_nick_to_edit = 20

primer_binding_len = 13
homology_arm_len = 13

grna_hairpin =
'GTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTG
AAAAAGTGGCACCGAGTCGGTGC' (SEQ ID NO: 1361579)
terminator = 'TTTTTTGTTTT' (SEQ ID NO: 1361581)

castypes = {
  'SpCas9 (NGG)': 'NGG',
```

```
  'SpCas9-NG (NG)': 'NG',
}

assert grna_nick_pos < grna_len
assert primer_binding_len < grna_nick_pos

##
# Поиск нРНК
##
iupac_nt = {
  'A': list('A'),
  'C': list('C'),
  'G': list('G'),
  'T': list('T'),
  'Y': list('CT'),
  'R': list('AG'),
  'W': list('AT'),
  'S': list('GC'),
  'K': list('TG'),
  'M': list('AC'),
  'D': list('AGT'),
  'V': list('ACG'),
  'H': list('ACT'),
  'B': list('CGT'),
  'N': list('ACGT'),
}
def match(template, dna):
  if len(dna) != len(template):
    return False
  for char, t in zip(dna, template):
    if char not in iupac_nt[t]:
      return False
  return True

def pam_match(seq, grna_pos1, pam):
  flag, stats = None, dict()
  cand_pam = seq[grna_pos1 + grna_len : grna_pos1 + grna_len +
len(pam)]
  if match(pam, cand_pam):
    flag = True
    stats['Designed gRNA (NGG orientation)'] = seq[grna_pos1 :
grna_pos1 + grna_len]
    stats['PAM'] = cand_pam
    stats['gRNA pos1 within sequence'] = grna_pos1
  return flag, stats
```

```
def find_grnas(seq, alt_start, alt_len, ref_allele, path_idx,
orient):
  min_grna_pos1 = alt_start - grna_nick_pos -
max_dist_nick_to_edit
  max_grna_pos1 = alt_start - grna_nick_pos

  '''
    gRNA nick site must be on the 5' side of the edit. Limit up
to 10 nt away and consider gRNAs on both strands
  '''

  all_grnas = defaultdict(list)
  for grna_pos1 in range(min_grna_pos1, max_grna_pos1 + 1):
    for castype in castypes:
      pam = castypes[castype]
      flag, grna_details = pam_match(seq, grna_pos1, pam)
      if flag:
        for key in grna_details:
          all_grnas[key].append(grna_details[key])
        all_grnas['Cas type'].append(castype)

        grna = grna_details['Designed gRNA (NGG orientation)']
        grna_pos1 = grna_details['gRNA pos1 within sequence']

        primer_binding = grna[grna_nick_pos -
primer_binding_len : grna_nick_pos]
        edit_template =  seq[grna_pos1 + grna_nick_pos :
path_idx] + ref_allele
        homology_arm = seq[path_idx + alt_len : path_idx +
alt_len + homology_arm_len]

        grna_extension = primer_binding + edit_template +
homology_arm

        all_grnas['Designed primer
binder'].append(primer_binding)
        all_grnas['Designed edit
template'].append(edit_template)
        all_grnas['Designed homology arm'].append(homology_arm)

        all_grnas['Designed gRNA
extension'].append(grna_extension)
        all_grnas['Designed orientation'].append(orient)

        all_grnas['Designed gRNA full (NGG
orientation)'].append(grna + grna_hairpin +
compbio.reverse_complement(grna_extension) + terminator)
```

```
  return all_grnas

##
#
##
def process_row(row):
  '''
    Find gRNAs in sequence at a single row.
  '''

  dis_seq = row['Sequence - alternate']
  ref_seq = row['Sequence - reference']
  path_start = row['buffer_length_bp']

  # [, ) interval
  alt_len = len(row['AlternateAllele'])
  path_end = path_start + alt_len

  fwd_grnas = find_grnas(
    dis_seq,
    path_start,
    alt_len,
    row['ReferenceAllele'],
    path_start,
    '+'
  )

  rev_grnas = find_grnas(
    compbio.reverse_complement(dis_seq),
    len(dis_seq) - path_end,
    alt_len,
    compbio.reverse_complement(row['ReferenceAllele']),
    path_start,
    '-'
  )

  fwd_df = pd.DataFrame(fwd_grnas)
  rev_df = pd.DataFrame(rev_grnas)
  df = fwd_df.append(rev_df, ignore_index = True)
  for col in row.index:
    df[col] = row[col]

  return df

def process_df():
```

```
  df = pd.read_csv(inp_dir + f'clinvar_{SPLIT}.csv', index_col =
0)

  mdf = pd.DataFrame()
  timer = util.Timer(total = len(df))
  for idx, row in df.iterrows():
    d = process_row(row)
    mdf = mdf.append(d, ignore_index = True)
    timer.update()

  mdf.to_csv(out_dir + f'clinvar_{SPLIT}.csv')
  return

##
# qsub
##
def gen_qsubs():
  # Генерирование qsub скриптов оболочки и команд для облегчения
параллелизации
  print('Generating qsub scripts...')
  qsubs_dir = _config.QSUBS_DIR + NAME + '/'
  util.ensure_dir_exists(qsubs_dir)
  qsub_commands = []

  num_scripts = 0
  for idx in range(0, 60):
    command = 'python %s.py %s' % (NAME, idx)
    script_id = NAME.split('_')[0]

    # Запись скриптов оболочки
    sh_fn = qsubs_dir + 'q_%s_%s.sh' % (script_id, idx)
    with open(sh_fn, 'w') as f:
      f.write('#!/bin/bash\n%s\n' % (command))
    num_scripts += 1

    # Запись команд qsub
    qsub_commands.append('qsub -V -l
h_rt=12:00:00,h_vmem=1G,os=RedHat7 -wd %s %s &' %
(_config.SRC_DIR, sh_fn))

  # Сохранение команд
  commands_fn = qsubs_dir + '_commands.sh'
  with open(commands_fn, 'w') as f:
    f.write('\n'.join(qsub_commands))

  subprocess.check_output('chmod +x %s' % (commands_fn), shell =
True)
```

```
  print('Wrote %s shell scripts to %s' % (num_scripts,
qsubs_dir))
  return

##
# Функция main
##
@util.time_dec
def main(argv):
  print(NAME)

  # Вызовы функции
  global SPLIT
  SPLIT = int(argv[0])

  process_df()

  return


if __name__ == '__main__':
  if len(sys.argv) > 1:
    main(sys.argv[1:])
  else:
    gen_qsubs()
```

[0317] Прилагаемый пример перечня последовательностей был сгенерирован с помощью описанных здесь способов с использованием базы данных ClinVar для получения входных аллелей и соответствующих выходных аллелей. Записи в базе данных ClinVar были сначала отфильтрованы для определения мутаций зародышевых линий, обозначенных как патогенные или вероятно патогенные. Для данных примеров использовали Cas9-NG и SpCas9 для идентификации совместимых мутаций. Из подвергнутых фильтрации мутаций приблизительно 72020 уникальных мутаций на основе ClinVar были идентифицированы в качестве совместимых с праймерным редактированием при использовании Cas9-NG, при этом приблизительно 63496 уникальных мутаций на основе ClinVar были идентифицированы в качестве совместимых с праймерным редактированием при использовании SpCas9 с PAM NGG. Является очевидным, что другие и/или дополнительные мутации могут являться корректируемыми при использовании

праймерного редактора, содержащего другой вариант Cas9 с другой совместимостью в плане PAM.

[0318] В различных вариантах осуществления данный алгоритм использовали для конструирования терапевтической PEgRNA согласно SEQ ID NO: 1-135514 и 813085-880462, сконструированной с помощью описанного здесь алгоритма в соответствии с записями из базы данных ClinVar.

[0319] В различных других вариантах осуществления данный алгоритм использовали для конструирования терапевтических PEgRNA в соответствии с базой данных ClinVar с использованием описанного здесь алгоритма, которые включены в Перечень последовательностей, который является частью настоящего описания. Перечень последовательностей включает последовательности полной PEgRNA SEQ ID NO: 1-135514 и 813085-880462. Каждая из данных полных PEgRNA состоит из спейсера (SEQ ID NO: 135515 – 271028 и 880463-947840) и плеча удлинения (SEQ ID NO: 271029 – 406542 и 947841-1015218). Кроме того, каждая PEgRNA включает ядро нРНК, например как определено согласно SEQ ID NO: 1361579-1361580. Каждое плечо удлинения согласно SEQ ID NO: 271029 – 406542 и 947841-1015218 дополнительно состоит из участка связывания праймера (SEQ ID NO: 406543 – 542056 и 1015219-1082596), матрицы для редактирования (SEQ ID NO.: 542057 – 677570 и 1082597-1149974), и плеча гомологии (SEQ ID NO.: 677571 – 813084 и 1149975-1217352). PEgRNA может необязательно включать 5′-концевой модификаторный участок и/или 3′-концевой модификаторный участок. PEgRNA может также включать сигнал терминации обратной транскрипции (например SEQ ID NO: 1361560-1361566) на 3'-конце PEgRNA. Изобретение охватывает конструирование и применение всех данных последовательностей.

[0320] Мутации были разделены на четыре класса согласно клинической значимости на основе малой частоты аллелей, количества отправителей, наличия противоречий в интерпретации отправителей, и наличия или отсутствия обзора мутации, выполненного группой экспертов. Среди 63496 SpCas9-совместимых мутаций: 4627 мутации были идентифицированы на наиболее высоком уровне значимости (четыре); 13943 мутации были идентифицирована на уровнях значимости три или четыре; и 44385 мутации были идентифицирована на уровнях значимости два, три, или четыре.

[0321] В приведенном перечне последовательностей указана одна PEgRNA на каждую уникальную мутацию, выбранная в виде PEgRNA с наиболее коротким расстоянием между разрывом и изменением. PEgRNA были сконструированы с плечом гомологии длиной 13 nt, длиной участка связывания праймера 13 nt, положением разрыва нРНК 17 nt, и длиной нРНК 20 nt. Протоспейсеры с участками разрыва на расстоянии более 20 nt от изменения исключались. Использованная последовательность каркаса нРНК являлась следующей: GTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGA AAAAGTGGCACCGAGTCGGTGC (SEQ ID NO: 1361579). Использованная терминаторная последовательность являлась следующей: TTTTTTGTTTT (SEQ ID NO: 1361581).

[0322] Как описано здесь, приведенные примеры перечней последовательностей не являются ограничивающими. Является очевидным, что вариации в рамках указанных конструкций PEgRNA могут включать описанные здесь вариации, включая различную последовательность каркаса нРНК, длину участка связывания праймера, длину флэпа, и/или им подобные.

[0323] Пример реализации компьютерной системы 3400, которая может использоваться для реализации любых аспектов описанных здесь способов и вариантов осуществления, показан на Фиг.34. Компьютерная система 3400 может включать один или несколько процессоров 3410 и один или несколько энергонезависимых машиночитаемых носителей данных (например память 3420 и один или несколько энергонезависимых носителей данных 3430) и дисплей 3440. Процессор 3410 может управлять записью и чтением данных из памяти 3420 и энергонезависимого носителя данных 3430 любым подходящим образом, поскольку аспекты описанного здесь изобретения не являются ограниченными в данном отношении. Для реализации описанной здесь функциональности и/или способов процессор 3410 может выполнять одну или несколько инструкций, хранящихся на одном или нескольких машиночитаемых носителях данных (например в памяти 3420, на носителе данных и так далее), которые могут служить в качестве энергонезависимых машиночитаемых носителей данных, на которых хранятся инструкции для выполнения процессором 3410.

[0324] В соответствии с описанными здесь способами, код, который используется, например, для определения структур удлиненной нРНК, может храниться на одном или

нескольких машиночитаемых носителях данных компьютерной системы 3400. Процессор 3410 может выполнять любой подобный код для реализации любых способов для планирования выполнения, как описано здесь. Любое другое описанное здесь программное обеспечение, программы или инструкции также могут храниться и выполнять компьютерной системой 3400. Является очевидным, что компьютерный код может применяться в рамках любых аспектов описанных здесь способов. Например, компьютерный код может применяться для взаимодействия с операционной системой для определения структур удлиненной нРНК в рамках стандартных процессов операционной системы.

[0325] Различные описанные здесь способы или процессы могут быть закодированы в виде программного обеспечения, которое выполняется на одном или нескольких процессорах, которые используют любую из множества операционных систем или платформ. Кроме того, подобное программное обеспечение может быть написано с помощью любого из множества подходящих языков программирования и/или инструментов для написания программ или скриптов, и также может быть скомпилировано в виде машинных кодов или промежуточного кода, который выполняется на виртуальной машине или в рамках подходящего фреймворка.

[0326] В данном отношении могут быть реализованы различные связанные с изобретением концепции в виде, по меньшей мере, одного энергонезависимого машиночитаемого носителя данных (например компьютерной памяти, одного или нескольких флоппи-дисков, компакт-дисков, оптических дисков, магнитных кассет, флеш-памяти, интегральных конфигураций в рамках программируемых вентильных массивов или других полупроводниковых устройствах, и т. д.) с кодом одной или нескольких программ, которые при выполнении на одном или нескольких компьютерах или других устройствах обеспечивают реализацию различных вариантов осуществления настоящего изобретения. Энергонезависимый машиночитаемый носитель или носители данных могут являться переносными, так что программа или программы, которые на них хранятся, могут быть загружена на любой компьютерный ресурс для реализации различных аспектов настоящего изобретения, как описано выше.

[0327] Термины "программа", "программное обеспечение", и/или "приложение" используются здесь в общем смысле для обозначения любого типа компьютерного кода

или набора машиночитаемых инструкций, которые могут быть использованы для программирования компьютера или другого вычислительного устройства для реализации различных аспектов вариантов осуществления, как описано выше. Кроме того, является очевидным, что в соответствии с одним аспектом, одна или несколько компьютерных программ, которые при выполнении обеспечивают реализацию способов по настоящему изобретению, не обязательно должны располагаться на одном компьютере или вычислительном устройстве, но могут быть распределены модулярным образом среди различных компьютеров или вычислительных устройств для реализации различных аспектов настоящего изобретения.

[0328] Машинные инструкции могут быть в различной форме, например в виде программных модулей, выполняемых одним или несколькими компьютерами или другими устройства. В целом, программные модули включают подпрограммы, программы, объекты, компоненты, структуры данных и так далее, которые обеспечивают выполнение определенных задач или реализацию определенных абстрактных типов данных. Как правило, функциональность программных модулей может быть скомбинирована или распределена в соответствии с необходимостью в рамках различных вариантов осуществления.

[0329] Кроме того структуры данных могут храниться на энергонезависимых машиночитаемых носителях данных в любой подходящей форме. Структуры данных могут иметь поля, которые являются связанными в плане местоположения в рамках структуры данных. Такие связи могут быть аналогичным образом реализованы путем присваивания пространства для хранения полям с месторасположением в рамках энергонезависимого машиночитаемого носителя данных, которое определяют связь между полями. Однако для установления связей между информацией в полях структуры данных может использоваться любой подходящий механизм, включая использование указателей, тегов, или других механизмов, которые обеспечивают установление связей между элементами данных.

[0330] Различные концепции, связанные с изобретением, могут быть реализованы в виде одного или нескольких способов, примеры которых были показаны. Действия, осуществленные в качестве части способа, могут быть реализованы в любом подходящем порядке. Соответственно, могут быть сформулированы варианты осуществления, в рамках

которых действия реализуются в отличном от указанного порядке, который может включать осуществление некоторых действий одновременно, даже если они показаны в виде последовательных действий в рамках примеров вариантов осуществления.

III. Праймерные редакторы для использования с терапевтической PEgRNA

[0331] Терапевтическая PEgRNA, сконструированная в соответствии с описанным здесь алгоритмом, может использоваться для осуществления праймерного редактирования в виде комплекса с праймерным редактором. Праймерные редакторы включают napDNAbp, слитый с полимеразой (например обратной транскриптазой) (или поступающей *in trans*), необязательно где два домена соединены линкерами, и могут дополнительно включать одну или несколько NLS. Эти аспекты дополнительно описаны ниже следующим образом.

A. napDNAbp

[0332] Описанные здесь праймерные редакторы могут включать программируемый нуклеиновой кислотой ДНК-связывающий белок (napDNAbp).

[0333] В одном аспекте napDNAbp может быть связан или образовывать комплекс, по меньшей мере, с одной направляющей нуклеиновой кислотой (например направляющей РНК или PEgRNA), которая локализует napDNAbp в последовательности ДНК, которая включает цепь ДНК (то есть цепь-мишень), которая комплементарна направляющей нуклеиновой кислоте или ее части (например спейсеру направляющей РНК, который отжигается на протоспейсере ДНК-мишени). Другими словами, направляющая нуклеиновая кислота "программирует" napDNAbp (например Cas9 или его эквивалент) для локализации и связывания с комплементарной последовательностью протоспейсера в ДНК.

[0334] В описанных здесь праймерных редакторах может применяться любой подходящий napDNAbp. В различных вариантах осуществления napDNAbp может представлять собой любую систему CRISPR-Cas класса 2, включая любой фермент CRISPR-Cas типа II, типа V, или типа VI. С учетом быстрого развития CRISPR-Cas в качестве инструмента для геномного редактирования происходили постоянные изменения в номенклатуре, используемой для описания и/или идентификации ферментов CRISPR-Cas, например Cas9 и ортологов Cas9. Данная заявка содержит указания ферментов CRISPR-Cas в соответствии с номенклатурой, которая может являться старой и/или новой. Специалист в

данной области будет способен идентифицировать конкретный указанный в данной заявке фермент CRISPR-Cas на основе используемой номенклатуры, независимо от того, является ли она старой (то есть "устаревшей") или новой номенклатурой. Номенклатура CRISPR-Cas широко описана в Makarova et al., "Classification and Nomenclature of CRISPR-Cas Systems: Where from Here?," *The CRISPR Journal*, Vol. 1. No. 5, 2018, полное содержание которого приведено здесь в качестве ссылки. Конкретная номенклатура CRISPR-Cas, используемая в любом случае в рамках данной заявки, не является ограниченной каким-либо образом и специалист в данной области будет способен идентифицировать указываемый фермент CRISPR-Cas.

[0335] Например, следующие ферменты CRISPR-Cas класса 2 типа II, типа V, и типа VI имеют следующие известные в данной области старые (то есть устаревшие) и новые названия. Каждый из данных ферментов и/или их вариантов может применяться в рамках описанных здесь праймерных редакторов.

Устаревшая номенклатура	**Текущая номенклатура ***
Ферменты CRISPR-Cas типа II	
Cas9	Та же
Ферменты CRISPR-Cas типа V	
Cpf1	Cas12a
CasX	Cas12e
C2c1	Cas12b1
Cas12b2	Та же
C2c3	Cas12c
CasY	Cas12d
C2c4	Та же
C2c8	Та же
C2c5	Та же
C2c10	Та же
C2c9	Та же
Ферменты CRISPR-Cas типа VI	
C2c2	Cas13a
Cas13d	Та же
C2c7	Cas13c
C2c6	Cas13b

* См. Makarova et al., *The CRISPR Journal*, Vol. 1, No. 5, 2018.

[0336] Без связи с теорией, механизм действия определенных предусмотренных здесь napDNAbp включает этап образования R-петли, при котором napDNAbp индуцирует разматывание двуцепочечной ДНК-мишени, тем самым разделяя цепи в участке,

связываемом napDNAbp. Затем спейсер направляющей РНК гибридизируется с "цепью-мишенью" на протоспейсерной последовательности. Это обеспечивает вытеснение "не являющейся мишенью цепи", которая комплементарна цепи-мишени, которая образует одноцепочечный участок R-петли. В некоторых вариантах осуществления napDNAbp включает один или несколько видов нуклеазной активности, которые затем обеспечивают разрезание ДНК, оставляя различные типы разрезов. Например napDNAbp может включать нуклеазную активность, обеспечивающую разрезание не являющейся мишенью цепи в первом месторасположении и/или разрезание цепи-мишени во втором месторасположении. В зависимости от нуклеазной активности, ДНК-мишень может быть разрезана с образованием "двуцепочечного разрыва", при котором обе цепи являются разрезанными. В других вариантах осуществления ДНК-мишень может быть разрезана только в одном участке, то есть ДНК "подвергается никированию" на одной цепи. Пример napDNAbp с различными видами нуклеазной активности включает "никазу Cas9" ("nCas9") и деактивированный Cas9 без нуклеазной активности ("мертвый Cas9" или "dCas9").

[0337] Приведенное ниже описание различных napDNAbp, которые могут применяться совместно с описываемыми здесь праймерными редакторами, не является ограничивающим каким-либо образом. Праймерные редакторы могут включать стандартный SpCas9 или любой ортолог белка Cas9, или любой вариант белка Cas9, включая любой встречающийся в природе вариант, мутант, или иным образом модифицированную версию Cas9, которая известна или которая может быть получена или может быть подвергнута эволюции с помощью связанного с направленной эволюцией или иного мутагенезного процесса. В различных вариантах осуществления Cas9 или варианты Cas9 обладают никазной активностью, то есть расщепляют только цепь последовательности-мишени ДНК. В других вариантах осуществления Cas9 или варианты Cas9 имеют деактивированные нуклеазы, то есть представляют собой белки в виде "мертвого" Cas9. Другие белки в виде вариантов Cas9, которые могут применяться, представляют собой белки с более низкой молекулярной массой, в отличие от стандартного SpCas9 (например для более легкой доставки) или имеющие модифицированную или перестроенную первичную аминокислотную структуру (например циркулярные пермутантные форматы).

[0338] Описанные здесь праймерные редакторы также могут включать эквиваленты Cas9, включая белки Cas12a (Cpf1) и Cas12b1, которые представляют собой результат конвергентной эволюции. Используемые здесь napDNAbp (например SpCas9, вариант Cas9, или эквиваленты Cas9) могут также содержать различные модификации, которые изменяют/повышают их специфичность в отношении PAM. Наконец, в рамках заявки предусмотрен любой Cas9, вариант Cas9, или эквивалент Cas9, который имеет, по меньшей мере, 70%, по меньшей мере, 75%, по меньшей мере, 80%, по меньшей мере, 85%, по меньшей мере, 90%, по меньшей мере, 91%, по меньшей мере, 92%, по меньшей мере, 93%, по меньшей мере, 94%, по меньшей мере, 95%, по меньшей мере, 96%, по меньшей мере, 97%, по меньшей мере, 98%, по меньшей мере, 99%, или, по меньшей мере, 99,9% идентичность последовательности относительно исходной последовательности Cas9, например исходной стандартной последовательности SpCas9 или исходного эквивалента Cas9 (например Cas12a (Cpf1)).

[0339] napDNAbp может представлять собой нуклеазу, ассоциированную с CRISPR (короткими регулярно расположенными группами палиндромными повторами). Как указано выше, CRISPR представляет собой адаптивную иммунную систему, которая обеспечивает защиту против подвижных генетических элементов (вирусов, переносимых элементов, и конъюгативных плазмид). Кластеры CRISPR содержат спейсеры, которые представляют собой последовательности, комплементарные предшествующим подвижным элементам, атакующие мишени нуклеиновые кислоты. Кластеры CRISPR транскрибируются и процессируются в РНК CRISPR (crРНК). В системах CRISPR типа II корректное процессирование пре-crРНК требует транс-кодируемую малую РНК (tracrРНК), эндогенную рибонуклеазу 3 (rnc) и белок Cas9. tracrРНК служит в качестве направляющей для осуществляемого с помощью рибонуклеазы 3 процессирования пре-crРНК. Далее Cas9/crРНК/tracrРНК обеспечивает расщепление эндонуклеазой линейной или циркулярной дцДНК-мишени, являющейся комплементарной спейсеру. Цепь-мишень, которая не является комплементарной crРНК, сначала разрезается эндонуклеазой, затем подрезается 3′-5′ экзонуклеазой. В природе связывание и расщепление ДНК обычно требует белок и обе РНК. Однако одиночные направляющие РНК ("sgРНК" или просто "нРНК") могут быть сконструированы для включения аспектов как crРНК, так и tracrРНК

в виде одиночной РНК. См, например Jinek M. et al., Science 337:816-821(2012), полное содержание которого приведено здесь в качестве ссылки.

[0340] В некоторых вариантах осуществления napDNAbp направляет расщепление одной или обеих цепей в месторасположении последовательности-мишени, например в последовательности-мишени и/или в комплементе последовательности-мишени. В некоторых вариантах осуществления napDNAbp направляет расщепление одной или обеих цепей на расстоянии около 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500 или более пар оснований от первого или последнего нуклеотида последовательности-мишени. В некоторых вариантах осуществления вектор кодирует napDNAbp, который подвергается мутированию с учетом соответствующего фермента дикого типа, так что мутированный napDNAbp не имеет способности расщеплять одну или обе цепи полинуклеотида-мишени, содержащего последовательность-мишень. Например, замещение аспартата на аланин (D10A) в каталитическом домене RuvC I Cas9 из *S. pyogenes* обеспечивает конвертацию Cas9 из нуклеазы, которая расщепляет обе цепи, в никазу (расщепляет одну цепь). Другие примеры мутаций, которые превращают Cas9 в никазу, включают, без ограничения, H840A, N854A и N863A для стандартной последовательности SpCas9 или для эквивалентных аминокислотных положений в других вариантах Cas9 или эквивалентах Cas9.

[0341] В соответствии с используемым здесь значением термин "белок Cas" обозначает полноразмерный белок Cas, полученный из природы, рекомбинантный белок Cas, имеющий последовательности, которые отличаются от встречающегося в природе белка Cas, или любой фрагмент белка Cas, который тем не менее сохраняет все или значительную часть требуемых основных функций, необходимых для описанных способов, то есть (i) программируемое нуклеиновой кислотой связывание белка Cas с ДНК-мишенью и (ii) способность никировать последовательность-мишень ДНК на одной цепи. Предусмотренные здесь белки Cas охватывают белки Cas9 CRISPR, а также эквиваленты Cas9, варианты (например никазу Cas9 (nCas9) или не имеющий нуклезной активности Cas9 (dCas9)), гомологи, ортологи, или паралоги, встречающиеся или не встречающиеся в природе (например сконструированные или рекомбинантные), и могут включать эквивалент Cas9 из любой системы CRISPR класса 2 (например типа II, V, VI), включая Cas12a (Cpf1), Cas12e (CasX), Cas12b1 (C2c1), Cas12b2, Cas12c (C2c3), C2c4,

C2c8, C2c5, C2c10, C2c9 Cas13a (C2c2), Cas13d, Cas13c (C2c7), Cas13b (C2c6), и Cas13b. Дополнительные эквиваленты Cas описаны в Makarova et al., "C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector," Science 2016; 353(6299) и Makarova et al., "Classification and Nomenclature of CRISPR-Cas Systems: Where from Here?," *The CRISPR Journal*, Vol. 1. No. 5, 2018, полное содержание которых приведено здесь в качестве ссылки.

[0342] Термины "Cas9" или "нуклеаза Cas9" или "фрагмент Cas9" или "домен Cas9" охватывают любой встречающийся в природе Cas9 из любого организма, любой встречающийся в природе эквивалент Cas9 или его функциональный фрагмент, любой гомолог Cas9, ортолог, или паралог из любого организма, и любой мутант или вариант Cas9, являющийся встречающимся в природе или сконструированным. Термин "Cas9" не является конкретно ограничивающим и может рассматриваться как "Cas9 или эквивалент". Примеры белков Cas9 дополнительно описаны здесь и/или описаны в данной области и приведены здесь в качестве ссылки. Настоящее изобретение не ограничено в плане конкретного Cas9, который используется в рамках праймерного редактора (PE) по изобретению.

[0343] Как указано здесь, последовательности и структуры нуклеазы Cas9 хорошо известны специалистам в данной области (см., например, "Complete genome sequence of an M1 strain of Streptococcus pyogenes." Ferretti et al., J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E., *Proc. Natl. Acad. Sci.* U.S.A. 98:4658-4663(2001); “CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.” Deltcheva E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., Nature 471:602-607(2011); и “A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.” Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012), полное содержание которых приведено здесь в качестве ссылки).

[0344] Примеры Cas9 и эквивалентов Cas9 приведены ниже; однако данные конкретные примеры не являются ограничивающими. В рамках праймерного редактора по

настоящему изобретению может использоваться любой подходящий napDNAbp, включая любой подходящий Cas9 или эквивалент Cas9.

(i) Стандартный SpCas9 дикого типа

[0345] В одном варианте осуществления описанный здесь конструкт праймерного редактора может включать "стандартную нуклеазу SpCas9" из *S. pyogenes*, которая широко использовалась в качестве инструмента для геномной инженерии и относится к подгруппе ферментов типа II систем CRISPR-Cas класса 2. Данный белок Cas9 представляет собой крупный имеющий несколько доменов белок, содержащий два отдельных нуклеазных домена. В Cas9 могут быть внесены точечные мутации для устранения одного или обоих видов нуклеазной активности, что приводит к получению никазы Cas9 (nCas9) или мертвого Cas9 (dCas9), соответственно, при этом он все же сохраняет способность связываться с ДНК за счет программирования sgРНК. В принципе, при слиянии с другим белком или доменом Cas9 или его вариант (например nCas9) может нацеливать данный белок практически на любую последовательность ДНК просто за счет совместной экспрессии с подходящей sgРНК. В соответствии с используемым здесь значением термин "стандартный белок SpCas9" обозначает белок дикого типа из *Streptococcus pyogenes*, имеющий следующую аминокислотную последовательность:

Описание	Последовательность	SEQ ID NO:
SpCas9 *Streptococcus pyogenes* M1 SwissProt кат. # Q99ZW2 Дикий тип	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDS GETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEED KKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFR GHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDL DNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDG TEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKI EKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKD FLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSG QGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVD QELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMK NYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQIL DSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNF FKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSK KLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR KRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHY LDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	SEQ ID NO: 1361421
SpCas9 Обратная	ATGGATAAAAAATATAGCATTGGCCTGGATATTGGCACCAACAGCGTGGGCTGGG CGGTGATTACCGATGAATATAAAGTGCCGAGCAAAAAATTTAAAGTGCTGGGCAA	SEQ ID NO: 1361422

трансляция SwissProt кат. # Q99ZW2 *Streptococcus pyogenes*	CACCGATCGCCATAGCATTAAAAAAAACCTGATTGGCGCGCTGCTGTTTGATAGC GGCGAAACCGCGGAAGCGACCCGCCTGAAACGCACCGCGCGCCGCCGCTATACCC GCCGCAAAAACCGCATTTGCTATCTGCAGGAAATTTTTAGCAACGAAATGGCGAA AGTGGATGATAGCTTTTTTCATCGCCTGGAAGAAAGCTTTCTGGTGGAAGAAGAT AAAAAACATGAACGCCATCCGATTTTTGGCAACATTGTGGATGAAGTGGCGTATC ATGAAAAATATCCGACCATTTATCATCTGCGCAAAAAACTGGTGGATAGCACCGA TAAAGCGGATCTGCGCCTGATTTATCTGGCGCTGGCGCATATGATTAAATTTCGC GGCCATTTTCTGATTGAAGGCGATCTGAACCCGGATAACAGCGATGTGGATAAAC TGTTTATTCAGCTGGTGCAGACCTATAACCAGCTGTTTGAAGAAAACCCGATTAA CGCGAGCGGCGTGGATGCGAAAGCGATTCTGAGCGCGCGCCTGAGCAAAAGCCGC CGCCTGGAAAACCTGATTGCGCAGCTGCCGGGCGAAAAAAAAAACGGCCTGTTTG GCAACCTGATTGCGCTGAGCCTGGGCCTGACCCCGAACTTTAAAAGCAACTTTGA TCTGGCGGAAGATGCGAAACTGCAGCTGAGCAAAGATACCTATGATGATGATCTG GATAACCTGCTGGCGCAGATTGGCGATCAGTATGCGGATCTGTTTCTGGCGGCGA AAAACCTGAGCGATGCGATTCTGCTGAGCGATATTCTGCGCGTGAACACCGAAAT TACCAAAGCGCCGCTGAGCGCGAGCATGATTAAACGCTATGATGAACATCATCAG GATCTGACCCTGCTGAAAGCGCTGGTGCGCCAGCAGCTGCCGGAAAAATATAAAG AAATTTTTTTTGATCAGAGCAAAAACGGCTATGCGGGCTATATTGATGGCGGCGC GAGCCAGGAAGAATTTTATAAATTTATTAAACCGATTCTGGAAAAAATGGATGGC ACCGAAGAACTGCTGGTGAAACTGAACCGCGAAGATCTGCTGCGCAAACAGCGCA CCTTTGATAACGGCAGCATTCCGCATCAGATTCATCTGGGCGAACTGCATGCGAT TCTGCGCCGCCAGGAAGATTTTTATCCGTTTCTGAAAGATAACCGCGAAAAAATT GAAAAAATTCTGACCTTTCGCATTCCGTATTATGTGGGCCCGCTGGCGCGCGGCA ACAGCCGCTTTGCGTGGATGACCCGCAAAAGCGAAGAAACCATTACCCCGTGGAA CTTTGAAGAAGTGGTGGATAAAGGCGCGAGCGCGCAGAGCTTTATTGAACGCATG ACCAACTTTGATAAAAACCTGCCGAACGAAAAAGTGCTGCCGAAACATAGCCTGC TGTATGAATATTTTACCGTGTATAACGAACTGACCAAAGTGAAATATGTGACCGA AGGCATGCGCAAACCGGCGTTTCTGAGCGGCGAACAGAAAAAAGCGATTGTGGAT CTGCTGTTTAAAACCAACCGCAAAGTGACCGTGAAACAGCTGAAAGAAGATTATT TTAAAAAAATTGAATGCTTTGATAGCGTGGAAATTAGCGGCGTGGAAGATCGCTT TAACGCGAGCCTGGGCACCTATCATGATCTGCTGAAAATTATTAAAGATAAAGAT TTTCTGGATAACGAAGAAAACGAAGATATTCTGGAAGATATTGTGCTGACCCTGA CCCTGTTTGAAGATCGCGAAATGATTGAAGAACGCCTGAAAACCTATGCGCATCT GTTTGATGATAAAGTGATGAAACAGCTGAAACGCCGCCGCTATACCGGCTGGGGC CGCCTGAGCCGCAAACTGATTAACGGCATTCGCGATAAACAGAGCGGCAAAACCA TTCTGGATTTTCTGAAAAGCGATGGCTTTGCGAACCGCAACTTTATGCAGCTGAT TCATGATGATAGCCTGACCTTTAAAGAAGATATTCAGAAAGCGCAGGTGAGCGGC CAGGGCGATAGCCTGCATGAACATATTGCGAACCTGGCGGGCAGCCCGGCGATTA AAAAAGGCATTCTGCAGACCGTGAAAGTGGTGGATGAACTGGTGAAAGTGATGGG CCGCCATAAACCGGAAAACATTGTGATTGAAATGGCGCGCGAAAACCAGACCACC CAGAAAGGCCAGAAAAACAGCCGCGAACGCATGAAACGCATTGAAGAAGGCATTA AAGAACTGGGCAGCCAGATTCTGAAAGAACATCCGGTGGAAAACACCCAGCTGCA GAACGAAAAACTGTATCTGTATTATCTGCAGAACGGCCGCGATATGTATGTGGAT CAGGAACTGGATATTAACCGCCTGAGCGATTATGATGTGGATCATATTGTGCCGC AGAGCTTTCTGAAAGATGATAGCATTGATAACAAAGTGCTGACCCGCAGCGATAA AAACCGCGGCAAAAGCGATAACGTGCCGAGCGAAGAAGTGGTGAAAAAAATGAAA AACTATTGGCGCCAGCTGCTGAACGCGAAACTGATTACCCAGCGCAAATTTGATA ACCTGACCAAAGCGGAACGCGGCGGCCTGAGCGAACTGGATAAAGCGGGCTTTAT TAAACGCCAGCTGGTGGAAACCCGCCAGATTACCAAACATGTGGCGCAGATTCTG GATAGCCGCATGAACACCAAATATGATGAAAACGATAAACTGATTCGCGAAGTGA AAGTGATTACCCTGAAAAGCAAACTGGTGAGCGATTTTCGCAAAGATTTTCAGTT TTATAAAGTGCGCGAAATTAACAACTATCATCATGCGCATGATGCGTATCTGAAC GCGGTGGTGGGCACCGCGCTGATTAAAAAATATCCGAAACTGGAAAGCGAATTTG TGTATGGCGATTATAAAGTGTATGATGTGCGCAAAATGATTGCGAAAAGCGAACA GGAAATTGGCAAAGCGACCGCGAAATATTTTTTTTATAGCAACATTATGAACTTT TTTAAAACCGAAATTACCCTGGCGAACGGCGAAATTCGCAAACGCCCGCTGATTG AAACCAACGGCGAAACCGGCGAAATTGTGTGGGATAAAGGCCGCGATTTTGCGAC CGTGCGCAAAGTGCTGAGCATGCCGCAGGTGAACATTGTGAAAAAAACCGAAGTG CAGACCGGCGGCTTTAGCAAAGAAAGCATTCTGCCGAAACGCAACAGCGATAAAC TGATTGCGCGCAAAAAAGATTGGGATCCGAAAAAATATGGCGGCTTTGATAGCCC GACCGTGGCGTATAGCGTGCTGGTGGTGGCGAAAGTGGAAAAAGGCAAAAGCAAA AAACTGAAAAGCGTGAAAGAACTGCTGGGCATTACCATTATGGAACGCAGCAGCT TTGAAAAAAACCCGATTGATTTTCTGGAAGCGAAAGGCTATAAAGAAGTGAAAAA	

	AGATCTGATTATTAAACTGCCGAAATATAGCCTGTTTGAACTGGAAAACGGCCGC AAACGCATGCTGGCGAGCGCGGGCGAACTGCAGAAAGGCAACGAACTGGCGCTGC CGAGCAAATATGTGAACTTTCTGTATCTGGCGAGCCATTATGAAAAACTGAAAGG CAGCCCGGAAGATAACGAACAGAAACAGCTGTTTGTGGAACAGCATAAACATTAT CTGGATGAAATTATTGAACAGATTAGCGAATTTAGCAAACGCGTGATTCTGGCGG ATGCGAACCTGGATAAAGTGCTGAGCGCGTATAACAAACATCGCGATAAACCGAT TCGCGAACAGGCGGAAAACATTATTCATCTGTTTACCCTGACCAACCTGGGCGCG CCGGCGGCGTTTAAATATTTTGATACCACCATTGATCGCAAACGCTATACCAGCA CCAAAGAAGTGCTGGATGCGACCCTGATTCATCAGAGCATTACCGGCCTGTATGA AACCCGCATTGATCTGAGCCAGCTGGGCGGCGAT	

[0346] Описанные здесь праймерные редакторы могут включать стандартный SpCas9 или любой его вариант, имеющий, по меньшей мере 80%, по меньшей мере, 85%, по меньшей мере, 90%, по меньшей мере, 95%, или, по меньшей мере, 99% идентичность последовательности относительно последовательности Cas9 дикого типа, приведенной выше. Данные варианты могут включать варианты SpCas9, содержащие одну или несколько мутаций, включая любую известную мутацию, указанную в записи SwissProt кат. # Q99ZW2, которая включает:

Мутацию SpCas9 (относительно аминокислотной последовательности стандартной последовательности SpCas9, SEQ ID NO: 1361421)	Функция/Характеристика (как указано) (см. запись – Q99ZW2 (CAS9_STRPT1) – приведена здесь в качестве ссылки)
D10A	Мутант никазы, который расщепляет протоспейсерную цепь (но не расщепляет непротоспейсерную цепь)
S15A	Сниженная активность в плане расщепления ДНК
R66A	Сниженная активность в плане расщепления ДНК
R70A	Расщепление ДНК отсутствует
R74A	Сниженное расщепление ДНК
R78A	Сниженное расщепление ДНК
Делеция 97-150	Нуклеазная активность отсутствует
R165A	Сниженное расщепление ДНК
Делеция 175-307	Сниженное приблизительно на 50% расщепление ДНК
Делеция 312-409	Нуклеазная активность отсутствует
E762A	Никаза
H840A	Мутант никазы, который расщепляет непротоспейсерную цепь, но не расщепляет протоспейсерную цепь
N854A	Никаза
N863A	Никаза
H982A	Сниженное расщепление ДНК
D986A	Никаза
Делеция 1099-1368	Нуклеазная активность отсутствует
R1333A	Сниженное связывание ДНК

[0347] Другие последовательности SpCas9 дикого типа, которые могут использоваться в настоящем изобретении, включают:

Описание	Последовательность	SEQ ID NO:
SpCas9	ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGT CGGATGGGCGGTGATCACTGATGATTATAAGGTTCCGTCTAAAAAGT	SEQ ID NO: 1361423

Streptococcus pyogenes MGAS1882 дикий тип NC_017053.1 —	TCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAAATCTT ATAGGGGCTCTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACTCG TCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTA TTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGAT GATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGA CAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAG TTGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAA TTGGCAGATTCTACTGATAAAGCGGATTTGCGCTTAATCTATTTGGC CTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAG ATTTAAATCCTGATAATAGTGATGTGGACAAACTATTTATCCAGTTG GTACAAATCTACAATCAATTATTTGAAGAAAACCCTATTAACGCAAG TAGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAA GACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAGAAAT GGCTTGTTTGGGAATCTCATTGCTTTGTCATTGGGATTGACCCCTAA TTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTT CAAAAGATACTTACGATGATGATTTAGATAATTTATTGGCGCAAATT GGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGA TGCTATTTTACTTTCAGATATCCTAAGAGTAAATAGTGAAATAACTA AGGCTCCCCTATCAGCTTCAATGATTAAGCGCTACGATGAACATCAT CAAGACTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGA AAAGTATAAAGAAATCTTTTTTTGATCAATCAAAAAACGGATATGCAG GTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATC AAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAA ACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACG GCTCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTG AGAAGACAAGAAGACTTTTATCCATTTTTAAAAGACAATCGTGAGAA GATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCAT TGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAA GAAACAATTACCCCATGGAATTTTGAAGAAGTTGTCGATAAAGGTGC TTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGATAAAAATC TTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTAT TTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAGGG AATGCGAAAACCAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTG TTGATTTACTCTTCAAAACAAATCGAAAAGTAACCGTTAAGCAATTA AAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAAT TTCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGCGCCTACCATG ATTTGCTAAAAATTATTAAAGATAAAGATTTTTTGGATAATGAAGAA AATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGA AGATAGGGGGATGATTGAGGAAAGACTTTAAAACATATGCTCACCTCT TTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGT TGGGGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCA ATCTGGCAAAACAATATTAGATTTTTTGAAATCAGATGGTTTTGCCA ATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTTAAA GAAGATATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACA TGAACAGATTGCTAACTTAGCTGGCAGTCCTGCTATTAAAAAAGGTA TTTTACAGACTGTAAAAATTGTTGATGAACTGGTCAAAGTAATGGGG CATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATCAGAC AACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCG AAGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCT GTTGAAAATACTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCT ACAAAATGGAAGAGACATGTATGTGGACCAAGAATTAGATATTAATC GTTTAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTCATT AAAGACGATTCAATAGACAATAAGGTACTAACGCGTTCTGATAAAAA TCGTGGTAAATCGGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAGA TGAAAAACTATTGGAGACAACTTCTAAACGCCAAGTTAATCACTCAA CGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGA ACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCC AAATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACT AAATACGATGAAAATGATAAACTTATTCGAGAGGTTAAAGTGATTAC CTTAAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCT ATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGATGCGTAT CTAAATGCCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAAACT TGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTA AAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAA TATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTAC ACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAAACTAATG GGGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGCCACA GTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAAAAC AGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAA GAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAA AAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGT GGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAATCCGTTA AAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAA AATCCGATTGACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAA AGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGTTAGAAA ACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTACAAAAAGGA AATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGC TAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAA AACAATTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTATT GAGCAAATCAGTGAATTTTCTAAGCGTGTTATTTTAGCAGATGCCAA TTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCAA TACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAAT CTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCG TAAACGATATACGTCTACAAAAGAAGTTTTAGATGCCACTCTTATCC ATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAG CTAGGAGGTGACTGA	
SpCas9 *Streptococcus pyogenes*	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNL IGALLFGSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKK LADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQL VQIYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRN GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI	SEQ ID NO: 1361424

MGAS1882 дикий тип NC_017053.1	GDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHH QDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFI KPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSE ETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKIIKDKDFLDNEE NEDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMKQLKRRRYTG WGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK EDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMG HKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFI KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQ RKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAY LNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAK YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEK NPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKG NELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ LGGD	
SpCas9 *Streptococcus pyogenes* дикий тип SWBC2D7W014	ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCCGT TGGATGGGCTGTCATAACCGATGAATACAAAGTACCTTCAAAGAAAT TTAAGGTGTTGGGGAACACAGACCGTCATTCGATTAAAAAGAATCTT ATCGGTGCCCTCCTATTCGATAGTGGCGAAACGGCAGAGGCGACTCG CCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGCAAGAACCGAA TATGTTACTTACAAGAAATTTTTAGCAATGAGATGGCCAAAGTTGAC GATTCTTTCTTTCACCGTTTGGAAGAGTCCTTCCTTGTCGAAGAGGA CAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGATGAGG TGGCATATCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAAG CTAGTTGACTCAACTGATAAAGCGGACCTGAGGTTAATCTACTTGGC TCTTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTG ATCTAAATCCGGACAACTCGGATGTCGACAAACTGTTCATCCAGTTA GTACAAACCTATAATCAGTTGTTTGAAGAGAACCCTATAAATGCAAG TGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCC GACGGCTAGAAAACCTGATCGCACAATTACCCGGAGAGAAGAAAAAT GGGTTGTTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCAAA TTTTAAGTCGAACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTA GTAAGGACACGTACGATGACGATCTCGACAATCTACTGGCACAAATT GGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGA TGCAATCCTCCTATCTGACATACTGAGAGTTAATACTGAGATTACCA AGGCGCCGTTATCCGCTTCAATGATCAAAAGGTACGATGAACATCAC CAAGACTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGA GAAATATAAGGAAATATTCTTTGATCAGTCGAAAAACGGGTACGCAG GTTATATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGTTTATC AAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAAA ACTCAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTCGACAACG GTAGCATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTT AGAAGGCAGGAGGATTTTTATCCGTTCCTCAAAGACAATCGTGAAAA GATTGAGAAAATCCTAACCTTTCGCATACCTTACTATGTGGGACCCC TGGCCCGAGGGAACTCTCGGTTCGCATGGATGACAAGAAAGTCCGAA GAAACGATTACTCCATGGAATTTTGAGGAAGTTGTCGATAAAGGTGC GTCAGCTCAATCGTTCATCGAGAGGATGACCAACTTTGACAAGAATT TACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTAT TTCACAGTGTACAATGAACTCACGAAAGTTAAGTATGTCACTGAGGG CATGCGTAAACCCGCCTTTCTAAGCGGAGAACAGAAGAAAGCAATAG TAGATCTGTTATTCAAGACCAACCGCAAAGTGACAGTTAAGCAATTG AAAGAGGACTACTTTAAGAAAATTGAATGCTTCGATTCTGTCGAGAT CTCCGGGGTAGAAGATCGATTTAATGCGTCACTTGGTACGTATCATG ACCTCCTAAAGATAATTAAAGATAAGGACTTCCTGGATAACGAAGAG AATGAAGATATCTTAGAAGATATAGTGTTGACTCTTACCCTCTTTGA AGATCGGGAAATGATTGAGGAAAGACTAAAAACATACGCTCACCTGT TCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGC TGGGGACGATTGTCGCGGAAACTTATCAACGGGATAAGAGACAAGCA AAGTGGTAAAACTATTCTCGATTTTCTAAAGAGCGACGGCTTCGCCA ATAGGAACTTTATGCAGCTGATCCATGATGACTCTTTAACCTTCAAA GAGGATATACAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTGCA CGAACATATTGCGAATCTTGCTGGTTCGCCAGCCATCAAAAAGGGCA TACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTCATGGGA CGTCACAAACCGGAAAACATTGTAATCGAGATGGCACGCGAAAATCA AACGACTCAGAAGGGGCAAAAAAACAGTCGAGAGCGGATGAAGAGAA TAGAAGAGGGTATTAAAGAACTGGGCAGCCAGATCTTAAAGGAGCAT CCTGTGGAAAATACCCAATTGCAGAACGAGAAACTTTACCTCTATTA CCTACAAAATGGAAGGGACATGTATGTTGATCAGGAACTGGACATAA ACCGTTTATCTGATTACGACGTCGATCACATTGTACCCCAATCCTTT TTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAA GAACCGAGGGAAAAGTGACAATGTTCCAAGCGAGGAAGTCGTAAAGA AAATGAAGAACTATTGGCGGCAGCTCCTAAATGCGAAACTGATAACG CAAAGAAAGTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTC TGAACTTGACAAGGCCGGATTTATTAAACGTCAGCTCGTGGAAACCC GCCAAATCACAAAGCATGTTGCACAGATACTAGATTCCCGAATGAAT ACGAAATACGACGAGAACGATAAGCTGATTCGGGAAGTCAAAGTAAT CACTTTAAAGTCAAAATTGGTGTCGGACTTCAGAAAGGATTTTCAAT TCTATAAAGTTAGGGAGATAAATAACTACCACCATGCGCACGACGCT TATCTTAATGCCGTCGTAGGGACCGCACTCATTAAGAAATACCCGAA GCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCC GTAAGATGATCGCGAAAAGCGAACAGGAGATAGGCAAGGCTACAGCC AAATACTTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAAT CACTCTGGCAAACGGAGAGATACGCAAACGACCTTTAATTGAAACCA	SEQ ID NO: 1361425

	ATGGGGAGACAGGTGAAATCGTATGGGATAAGGGCCGGGACTTCGCG ACGGTGAGAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTAAAGAA AACTGAGGTGCAGACCGGAGGGTTTTCAAAGGAATCGATTCTTCCAA AAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGACTGGGACCCG AAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCTATTCTGTCCT AGTAGTGGCAAAAGTTGAGAAGGGAAAATCCAAGAAACTGAAGTCAG TCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAA AAGAACCCCATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAA AAAGGATCTCATAATTAAACTACCAAAGTATAGTCTGTTTGAGTTAG AAAATGGCCGAAAACGGATGTTGGCTAGCGCCGGAGAGCTTCAAAAG GGGAACGAACTCGCACTACCGTCTAAATACGTGAATTTCCTGTATTT AGCGTCCCATTACGAGAAGTTGAAAGGTTCACCTGAAGATAACGAAC AGAAGCAACTTTTTGTTGAGCAGCACAAACATTATCTCGACGAAATC ATAGAGCAAATTTCGGAATTCAGTAAGAGAGTCATCCTAGCTGATGC CAATCTGGACAAAGTATTAAGCGCATACAACAAGCACAGGGATAAAC CCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACC AACCTCGGCGCTCCAGCCGCATTCAAGTATTTTGACACAACGATAGA TCGCAAACGATACACTTCTACCAAGGAGGTGCTAGACGCGACACTGA TTCACCAATCCATCACGGGATTATATGAAACTCGGATAGATTTGTCA CAGCTTGGGGGTGACGGATCCCCCAAGAAGAAGAGGAAAGTCTCGAG CGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATT ACAAGGATGACGATGACAAGGCTGCAGGA	
SpCas9 *Streptococcus pyogenes* дикий тип Кодируемый продукт SWBC2D7W014	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKK LVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQL VQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI GDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHH QDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFI KPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSE ETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTG WGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK EDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEH PVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLIT QRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMN TKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFA TVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFE KNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQK GNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLS QLGGDGSPKKKRKVSSDYKDHDGDYKDHDIDYKDDDDKAAG	SEQ ID NO: 1361426
SpCas9 *Streptococcus pyogenes* M1GAS дикий тип NC_002737.2	ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGT CGGATGGGCGGTGATCACTGATGAATATAAGGTTCCGTCTAAAAAGT TCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAAATCTT ATAGGGGCTCTTTTATTTGACAGTGGAGAGACAGCGGAAGCGACTCG TCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTA TTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGAT GATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGA CAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAG TTGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAA TTGGTAGATTCTACTGATAAAGCGGATTTGCGCTTAATCTATTTGGC CTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAG ATTTAAATCCTGATAATAGTGATGTGGACAAACTATTTATCCAGTTG GTACAAACCTACAATCAATTATTTGAAGAAAACCCTATTAACGCAAG TGGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAA GACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAAAAAT GGCTTATTTGGGAATCTCATTGCTTTGTCATTGGGTTTGACCCCTAA TTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTT CAAAAGATACTTACGATGATGATTTAGATAATTTATTGGCGCAAATT GGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGA TGCTATTTTACTTTCAGATATCCTAAGAGTAAATACTGAAATAACTA AGGCTCCCCTATCAGCTTCAATGATTAAACGCTACGATGAACATCAT CAAGACTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGA AAAGTATAAAGAAATCTTTTTTGATCAATCAAAAAACGGATATGCAG GTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATC AAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAA ACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACG GCTCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTG AGAAGACAAGAAGACTTTTATCCATTTTTAAAAGACAATCGTGAGAA GATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCAT TGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAA GAAACAATTACCCCATGGAATTTTGAAGAAGTTGTCGATAAAGGTGC TTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGATAAAAATC TTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTAT TTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAAGG AATGCGAAAACCAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTG TTGATTTACTCTTCAAAACAAATCGAAAAGTAACCGTTAAGCAATTA AAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAAT TTCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGTACCTACCATG ATTTGCTAAAAATTATTAAAGATAAAGATTTTTTGGATAATGAAGAA AATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGA	SEQ ID NO: 1361427

	AGATAGGGAGATGATTGAGGAAAGACTTAAAACATATGCTCACCTCT TTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGT TGGGGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCA ATCTGGCAAAACAATATTAGATTTTTTGAAATCAGATGGTTTTGCCA ATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTTAAA GAAGACATTCAAAAAGCACAAGTGTCTGGACAAGGCGATAGTTTACA TGAACATATTGCAAATTTAGCTGGTAGCCCTGCTATTAAAAAAGGTA TTTTACAGACTGTAAAAGTTGTTGATGAATTGGTCAAAGTAATGGGG CGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATCA GACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAA TCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCAT CCTGTTGAAAATACTCAATTGCAAAATGAAAAGCTCTATCTCTATTA TCTCCAAAATGGAAGAGACATGTATGTGGACCAAGAATTAGATATTA ATCGTTTAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTC CTTAAAGACGATTCAATAGACAATAAGGTCTTAACGCGTTCTGATAA AAATCGTGGTAAATCGGATAACGTTCCAAGTGAAGAAGTAGTCAAAA AGATGAAAAACTATTGGAGACAACTTCTAAACGCCAAGTTAATCACT CAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAG TGAACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTC GCCAAATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAAT ACTAAATACGATGAAAATGATAAACTTATTCGAGAGGTTAAAGTGAT TACCTTAAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAAT TCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGATGCG TATCTAAATGCCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAA ACTTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTC GTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCA AAATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAAT TACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAAACTA ATGGGGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGCC ACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAA AACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAA AAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCA AAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCT AGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAATCCG TTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAA AAAAATCCGATTGACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAA AAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGTTAG AAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTACAAAAA GGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTT AGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAAC AAAAACAATTGTTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATT ATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTTTAGCAGATGC CAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAAC CAATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACG AATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGA TCGTAAACGATATACGTCTACAAAAGAAGTTTTAGATGCCACTCTTA TCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGT CAGCTAGGAGGTGACTGA	
SpCas9 *Streptococcus pyogenes* M1GAS дикий тип Кодируемый продукт NC_002737.2 (100% идентичен стандартному Q99ZW2 дикого типа)	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKK LVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQL VQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI GDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHH QDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFI KPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSE ETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTG WGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK EDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEH PVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLIT QRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMN TKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFA TVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFE KNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQK GNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLS QLGGD	SEQ ID NO: 1361428

[0348] Описанные здесь праймерные редакторы могут включать любую из указанных выше последовательностей SpCas9 или любой его вариант, имеющий, по меньшей мере 80%, по меньшей мере, 85%, по меньшей мере, 90%, по меньшей мере, 95%, или, по меньшей мере, 99% идентичность относительно них.

(ii) Ортологи Cas9 дикого типа

[0349] В других вариантах осуществления белок Cas9 может представлять собой ортолог Cas9 дикого типа из другого вида бактерий, отличный от стандартного Cas9 из S. pyogenes. Например, с описанными в данном описании конструктами праймерного редактора могут использоваться следующие ортологи Cas9. Кроме того, любой вариант ортологов Cas9, имеющий, по меньшей мере, 80%, по меньшей мере, 85%, по меньшей мере, 90%, по меньшей мере, 95%, или, по меньшей мере, 99% идентичность последовательности относительно любого из указанных ниже ортологов, также может использоваться с настоящими праймерными редакторами.

Описание	Последовательность
LfCas9 *Lactobacillus fermentum* дикий тип *GenBank: SNX31424.1* 1	MKEYHIGLDIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFEEGKTAAERRTFRTTRRRLKRRKWRLHYLDEIFAPHLQEVD ENFLRRLKQSNIHPEDPTKNQAFIGKLLFPDLLKKNERGYPTLIKMRDELPVEQRAHYPVMNIYKLREAMINEDRQFDLRE VYLAVHHIVKYRGHFLNNASVDKFKVGRIDFDKSFNVLNEAYEELQNGEGSFTIEPSKVEKIGQLLLDTKMRKLDRQKAVA KLLEVKVADKEETKRNKQIATAMSKLVLGYKADFATVAMANGNEWKIDLSSETSEDEIEKFREELSDAQNDILTEITSLFS QIMLNEIVPNGMSISESMMDRYWTHERQLAEVKEYLATQPASARKEFDQVYNKYIGQAPKERGFDLEKGLKKILSKKENWK EIDELLKAGDFLPKQRTSANGVIPHQMHQQELDRIIEKQAKYYPWLATENPATGERDRHQAKYELDQLVSFRIPYYVGPLV TPEVQKATSGAKFAWAKRKEDGEITPWNLWDKIDRAESAEAFIKRMTVKDTYLLNEDVLPANSLLYQKYNVLNELNNVRVN GRRLSVGIKQDIYTELFKKKKTVKASDVASLVMAKTRGVNKPSVEGLSDPKKFNSNLATYLDLKSIVGDKVDDNRYQTDLE NIIEWRSVFEDGEIFADKLTEVEWLTDEQRSALVKKRYKGWGRLSKKLLTGIVDENGQRIIDLMWNTDQNFKEIVDQPVFK EQIDQLNQKAITNDGMTLRERVESVLDDAYTSPQNKKAIWQVVRVVEDIVKAVGNAPKSISIEFARNEGNKGEITRSRRTQ LQKLFEDQAHELVKDTSLTEELEKAPDLSDRYYFYFTQGGKDMYTGDPINFDEISTKYDIDHILPQSFVKDNSLDNRVLTS RKENNKKSDQVPAKLYAAKMKPYWNQLLKQGLITQRKFENLTKDVDQNIKYRSLGFVKRQLVETRQVIKLTANILGSMYQE AGTEIIETRAGLTKQLREEFDLPKVREVNDYHHAVDAYLTTFAGQYLNRRYPKLRSFFVYGEYMKFKHGSDLKLRNFNFFH ELMEGDKSQGKVVDQQTGELITTRDEVAKSFDRLLNMKYMLVSKEVHDRSDQLYGATIVTAKESGKLTSPIEIKKNRLVDL YGAYTNGTSAFMTIIKFTGNKPKYKVIGIPTTSAASLKRAGKPGSESYNQELHRIIKSNPKVKKGFEIVVPHVSYGQLIVD GDCKFTLASPTVQHPATQLVLSKKSLETISSGYKILKDKPAIANERLIRVFDEVVGQMNRYFTIFDQRSNRQKVADARDKF LSLPTESKYEGAKKVQVGKTEVITNLLMGLHANATQGDLKVLGLATFGFFQSTTGLSLSEDTMIVYQSPTGLFERRICLKD I(SEQ ID NO: 1361429)
SaCas9 *Staphylococcus aureus* дикий тип GenBank: AYD60528.1	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICY LQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIA LSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKR YDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGA SAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKED YFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVM KQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAG SPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEK LYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKL ITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLA NGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF DSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLA SAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNK HRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD(SEQ ID NO: 1361430)
SaCas9 *Staphylococcus aureus*	MGKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDH SELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKD GEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPE ELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEF TNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLIL DELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSK DAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFD NSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLV DTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVM ENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNNLNGLY DKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAH LDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYKNDL IKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKK

Описание	Последовательность
	(SEQ ID NO: 1361431)
StCas9 *Streptococcus thermophilus* UniProtKB/Swiss-Prot: G3ECR1.2 Дикий тип	MLFNKCIISINLDFSNKEKCMTKPYSIGLDIGTNSVGWAVITDNYKVPSKKMKVLGNTSKKYIKKNLLGVLLFDSGITAE GRRLKRTARRRYTRRRNRILYLQEIFSTEMATLDDAFFQRLDDSFLVPDDKRDSKYPIFGNLVEEKVYHDEFPTIYHLRKY LADSTKKADLRLVYLALAHMIKYRGHFLIEGEFNSKNNDIQKNFQDFLDTYNAIFESDLSLENSKQLEEIVKDKISKLEKK DRILKLFPGEKNSGIFSEFLKLIVGNQADFRKCFNLDEKASLHFSKESYDEDLETLLGYIGDDYSDVFLKAKKLYDAILLS GFLTVTDNETEAPLSSAMIKRYNEHKEDLALLKEYIRNISLKTYNEVFKDDTKNGYAGYIDGKTNQEDFYVYLKNLLAEFE GADYFLEKIDREDFLRKQRTFDNGSIPYQIHLQEMRAILDKQAKFYPFLAKNKERIEKILTFRIPYYVGPLARGNSDFAWS IRKRNEKITPWNFEDVIDKESSAEAFINRMTSFDLYLPEEKVLPKHSLLYETFNVYNELTKVRFIAESMRDYQFLDSKQKK DIVRLYFKDKRKVTDKDIIEYLHAIYGYDGIELKGIEKQFNSSLSTYHDLLNIINDKEFLDDSSNEAIIEEIIHTLTIFED REMIKQRLSKFENIFDKSVLKKLSRRHYTGWGKLSAKLINGIRDEKSGNTILDYLIDDGISNRNFMQLIHDDALSFKKKIQ KAQIIGDEDKGNIKEVVKSLPGSPAIKKGILQSIKIVDELVKVMGGRKPESIVVEMARENQYTNQGKSNSQQRLKRLEKSL KELGSKILKENIPAKLSKIDNNALQNDRLYLYYLQNGKDMYTGDDLDIDRLSNYDIDHIIPQAFLKDNSIDNKVLVSSASN RGKSDDFPSLEVVKKRKTFWYQLLKSKLISQRKFDNLTKAERGGLLPEDKAGFIQRQLVETRQITKHVARLLDEKFNNKKD ENNRAVRTVKIITLKSTLVSQFRKDFELYKVREINDFHHAHDAYLNAVIASALLKKYPKLEPEFVYGDYPKYNSFRERKSA TEKVYFYSNIMNIFKKSISLADGRVIERPLIEVNEETGESVWNKESDLATVRRVLSYPQVNVVKKVEEQNHGLDRGKPKGL FNANLSSKPKPNSNENLVGAKEYLDPKKYGGYAGISNSFAVLVKGTIEKGAKKKITNVLEFQGISILDRINYRKDKLNFLL EKGYKDIELIIELPKYSLFELSDGSRRMLASILSTNNKRGEIHKGNQIFLSQKFVKLLYHAKRISNTINENHRKYVENHKK EFEELFYYILEFNENYVGAKKNGKLLNSAFQSWQNHSIDELCSSFIGPTGSERKGLFELTSRGSAADFEFLGVKIPRYRDY TPSSLLKDATLIHQSVTGLYETRIDLAKLGEG(SEQ ID NO: 1361432)
LcCas9 *Lactobacillus crispatus* NCBI Исходная последовательность: WP_133478044.1 Дикий тип	MKIKNYNLALTPSTSAVGHVEVDDDLNILEPVHHQKAIGVAKFGEGETAEARRLARSARRTTKRRANRINHYFNEIMKPEI DKVDPLMFDRIKQAGLSPLDERKEFRTVIFDRPNIASYYHNQFPTIWHLQKYLMITDEKADIRLIYWALHSLLKHRGHFFN TTPMSQFKPGKLNLKDDMLALDDYNDLEGLSFAVANSPEIEKVIKDRSMHKKEKIAELKKLIVNDVPDKDLAKRNNKIITQ IVNAIMGNSFHLNFIFDMDLDKLTSKAWSFKLDDPELDTKFDAISGSMTDNQIGIFETLQKIYSAISLLDILNGSSNVVDA KNALYDKHKRDLNLYFKFLNTLPDEIAKTLKAGYTLYIGNRKKDLLAARKLLKVNVAKNFSQDDFYKLINKELKSIDKQGL QTRFSEKVGELVAQNNFLPVQRSSDNVFIPYQLNAITFNKILENQGKYYDFLVKPNPAKKDRKNAPYELSQLMQFTIPYYV GPLVTPEEQVKSGIPKTSRFAWMVRKDNGAITPWNFYDKVDIEATADKFIKRSIAKDSYLLSELVLPKHSLLYEKYEVFNE LSNVSLDGKKLSGGVKQILFNEVFKKTNKVNTSRILKALAKHNIPGSKITGLSNPEEFTSSLQTYNAWKKYFPNQIDNFAY QQDLEKMIEWSTVFEDHKILAKKLDEIEWLDDDQKKFVANTRLRGWGRLSKRLLTGLKDNYGKSIMQRLETTKANFQQIVY KPEFREQIDKISQAAAKNQSLEDILANSYTSPSNRKAIRKTMSVVDEYIKLNHGKEPDKIFLMFQRSEQEKGKQTEARSKQ LNRILSQLKADKSANKLFSKQLADEFSNAIKKSKYKLNDKQYFYFQQLGRDALTGEVIDYDELYKYTVLHIIPRSKLTDDS QNNKVLTKYKIVDGSVALKFGNSYSDALGMPIKAFWTELNRLKLIPKGKLLNLTTDFSTLNKYQRDGYIARQLVETQQIVK LLATIMQSRFKHTKIIEVRNSQVANIRYQFDYFRIKNLNEYYRGFDAYLAAVVGTYLYKVYPKARRLFVYGQYLKPKKTNQ ENQDMHLDSEKKSQGFNFLWNLLYGKQDQIFVNGTDVIAFNRKDLITKMNTVYNYKSQKISLAIDYHNGAMFKATLFPRND RDTAKTRKLIPKKKDYDTDIYGGYTSNVDGYMLLAEIIKRDGNKQYGFYGVPSRLVSELDTLKKTRYTEYEEKLKEIIKPE LGVDLKKIKKIKILKNKVPFNQVIIDKGSKFFITSTSYRWNYRQLILSAESQQTLMDLVVDPDFSNHKARKDARKNADERL IKVYEEILYQVKNYMPMFVELHRCYEKLVDAQKTFKSLKISDKAMVLNQILILLHSNATSPVLEKLGYHTRFTLGKKHNLI SENAVLVTQSITGLKENHVSIKQML(SEQ ID NO: 1361433)
PdCas9 *Pedicoccus damnosus* NCBI Исходная последовательность: WP_062913273.1 Дикий тип	MTNEKYSIGLDIGTSSIGFAVVNDNNRVIRVKGKNAIGVRLFDEGKAAADRRSFRTTRRSFRTTRRRLSRRRWRLKLLREI FDAYITPVDEAFFIRLKESNLSPKDSKKQYSGDILFNDRSDKDFYEKYPTIYHLRNALMTEHRKFDVREIYLAIHHIMKFR GHFLNATPANNFKVGRLNLEEKFEELNDIYQRVFPDESIEFRTDNLEQIKEVLLDNKRSRADRQRTLVSDIYQSSEDKDIE KRNKAVATEILKASLGNKAKLNVITNVEVDKEAAKEWSITFDSESIDDDLAKIEGQMTDDGHEIIEVLRSLYSGITLSAIV PENHTLSQSMVAKYDLHKDHLKLFKKLINGMTDTKKAKNLRAAYDGYIDGVKGKVLPQEDFYKQVQVNLDDSAEANEIQTY IDQDIFMPKQRTKANGSIPHQLQQQELDQIIENQKAYYPWLAELNPNPDKKRQQLAKYKLDELVTFRVPYYVGPMITAKDQ KNQSGAEFAWMIRKEPGNITPWNFDQKVDRMATANQFIKRMTTTDTYLLGEDVLPAQSLLYQKFEVLNELNKIRIDHKPIS IEQKQQIFNDLFKQFKNVTIKHLQDYLVSQGQYSKRPLIEGLADEKRFNSSLSTYSDLCGIFGAKLVEENDRQEDLEKIIE WSTIFEDKKIYRAKLNDLTWLTDDQKEKLATKRYQGWGRLSRKLLVGLKNSEHRNIMDILWITNENFMQIQAEPDFAKLVT DANKGMLEKTDSQDVINDLYTSPQNKKAIRQILLVVHDIQNAMHGQAPAKIHVEFARGEERNPRRSVQRQRQVEAAYEKVS NELVSAKVRQEFKEAINNKRDFKDRLFLYFMQGGIDIYTGKQLNIDQLSSYQIDHILPQAFVKDDSLTNRVLTNENQVKAD SVPIDIFGKKMLSVWGRMKDQGLISKGKYRNLTMNPENISAHTENGFINRQLVETRQVIKLAVNILADEYGDSTQIISVKA DLSHQMREDFELLKNRDVNDYHHAFDAYLAAFIGNYLLKRYPKLESYFVYGDFKKFTQKETKMRRFNFIYDLKHCDQVVNK ETGEILWTKDEDIKYIRHLFAYKKILVSHEVREKRGALYNQTIYKAKDDKGSGQESKKLIRIKDDKETKIYGGYSGKSLAY MTIVQITKKNKVSYRVIGIPTLALARLNKLENDSTENNGELYKIIKPQFTHYKVDKKNGEIIETTDDFKIVVSKVRFQQLI DDAGQFFFMLASDTYKNNAQQLVISNNALKAINNTNITDCPRDDLERLDNLRLDSAFDEIVKKMDKYFSAYDANNFREKIRN SNLIFYQLPVEDQWENNKITELGKRTVLTRILQGLHANATTTDMSIFKIKTPFGQLRQRSGISLSENAQLIYQSPTGLFER RVQLNKIK(SEQ ID NO: 1361434)
FnCas9 *Fusobacterium nucleatum* NCBI Исходная последовательность: WP_060798984.1	MKKQKFSDYYLGFDIGTNSVGWCVTDLDYNVLRFNKKDMWGSRLFEEAKTAAERRVQRNSRRRLKRRKWRLNLLEEIFSNE ILKIDSNFFRRLKESSLWLEDKSSKEKFTLFNDDNYKDYDFYKQYPTIFHLRNELIKNPEKKDIRLVYLAIHSIFKSRGHF LFEGQNLKEIKNFETLYNNLIAFLEDNGINKIIDKNNIEKLEKIVCDSKKGLKDKEKEFKEIFNSDKQLVAIFKLSVGSSV SLNDLFDTDEYKKGEVEKEKISFREQIYEDDKPIYYSILGEKIELLDIAKTFYDFMVLNNILADSQYISEAKVKLYEEHKK DLKNLKYIIRKYNKGNYDKLFKDKNENNYSAYIGLNKEKSKKEVIEKSRLKIDDLIKNIKGYLPKVEEIEEKDKAIFNKIL NKIELKTILPKQRISDNGTLPYQIHEAELEKILENQSKYYDFLNYEENGIITKDKLLMTFKFRIPYYVGPLNSYHKDKGGN SWIVRKEEGKILPWNFEQKVDIEKSAEEFIKRMTNKCTYLNGEDVIPKDTFLYSEYVILNELNKVQVNDEFLNEENKRKII DELFKENKKVSEKKFKEYLLVKQIVDGTIELKGVKDSFNSNYISYIRFKDIFGEKLNLDIYKEISEKSILWKCLYGDDKKI FEKKIKNEYGDILTKDEIKKINTFKFNNWGRLSEKLLTGIEFINLETGECYSSVMDALRRTNYNLMELLSSKFTLQESINN ENKEMNEASYRDLIEESYVSPSLKRAIFQTLKIYEEIRKITGRVPKKVFIEMARGGDESMKNKKIPARQEQLKKLYDSCGN DIANFSIDIKEMKNSLISYDNNSLRQKKLYLYYLQFGKCMYTGREIDLDRLLQNNDTYDIDHIYPRSKVIKDDSFDNLVLV LKNENAEKSNEYPVKKEIQEKMKSFWRFLKEKNFISDEKYKRLTGKDDFELRGFMARQLVNVRQTTKEVGKILQQIEPEIK IVYSKAEIASSFREMFDFIKVRELNDTHHAKDAYLNIVAGNVYNTKFTEKPYRYLQEIKENYDVKKIYNYDIKNAWDKENS LEIVKKNMEKNTVNITRFIKEKKGQLFDLNPIKKGETSNEIISIKPKVYNGKDDKLNEKYGYYKSLNPAYFLYVEHKEKNK

Описание	Последовательность
	RIKSFERVNLVDVNNIKDEKSLVKYLIENKKLVEPRVIKKVYKRQVILINDYPYSIVTLDSNKLMDFENLKPLFLENKYEK ILKNVIKFLEDNQGKSEENYKFIYLKKKDRYEKNETLESVKDRYNLEFNEMYDKFLEKLDSKDYKNYMNNKKYQELLDVKE KFIKLNLFDKAFTLKSFLDLFNRKTMADFSKVGLTKYLGKIQKISSNVLSKNELYLLEESVTGLFVKKIKL(SEQ ID NO: 1361435)
EcCas9 *Enterococcus cecorum* NCBI Исходная последовательность: WP_047338501.1 Дикий тип	RRKQRIQILQELLGEEVLKTDPGFFHRMKESRYVVEDKRTLDGKQVELPYALFVDKDYTDKEYYKQFPTINHLIVYLMTTS DTPDIRLVYLALHYYMKNRGNFLHSGDINNVKDINDILEQLDNVLETFLDGWNLKLKSYVEDIKNIYNRDLGRGERKKAFV NTLGAKTKAEKAFCSLISGGSTNLAELFDDSSLKEIETPKIEFASSSLEDKIDGIQEALEDRFAVIEAAKRLYDWKTLTDI LGDSSSLAEARVNSYQMHHEQLLELKSLVKEYLDRKVFQEVFVSLNVANNYPAYIGHTKINGKKKELEVKRTKRNDFYSYV KKQVIEPIKKKVSDEAVLTKLSEIESLIEVDKYLPLQVNSDNGVIPYQVKLNELTRIFDNLENRIPVLRENRDKIIKTFKF RIPYYVGSLNGVVKNGKCTNWMVRKEEGKIYPWNFEDKVDLEASAEQFIRRMTNKCTYLVNEDVLPKYSLLYSKYLVLSEL NNLRIDGRPLDVKIKQDIYENVFKKNRKVTLKKIKKYLLKEGIITDDDELSGLADDVKSSLTAYRDFKEKLGHLDLSEAQM ENIILNITLFGDDKKLLKKRLAALYPFIDDKSLNRIATLNYRDWGRLSERFLSGITSVDQETGELRTIIQCMYETQANLMQ LLAEPYHFVEAIEKENPKVDLESISYRIVNDLYVSPAVKRQIWQTLLVIKDIKQVMKHDPERIFIEMARENQESKKTKSRK QVLSEVYKKAKEYEHLFEKLNSLTEEQLRSKKIYLYFTQLGKCMYSGEPIDFENLVSANSNYDIDHIYPQSKTIDDSFNNI VLVKKSLNAYKSNHYPIDKNIRDNEKVKTLWNTLVSKGLITKEKYERLIRSTPFSDEELAGFIARQLVETRQSTKAVAEIL SNWFPESEIVYSKAKNVSNFRQDFEILKVRELNDCHHAHDAYLNIVVGNAYHTKFTNSPYRFIKNKANQEYNLRKLLQKVN KIESNGVVAWVGQSENNPGTIATVKKVIRRNTVLISRMVKEVDGQLFDLTLMKKGKGQVPIKSSDERLTDISKYGGYNKAT GAYFTFVKSKKRGKVVRSFEYVPLHLSKQFENNNELLKEYIEKDRGLTDVEILIPKVLINSLFRYNGSLVRITGRGDTRLL LVHEQPLYVSNSFVQQLKSVSSYKLKKSENDNAKLTKTATEKLSNIDELYDGLLRKLDLPIYSYWFSSIKEYLVESRTKYI KLSIEEKALVIFEILHLFQSDAQVPNLKILGLSTKPSRIRIQKNLKDTDKMSIIHQSPSGIFEHEIELTSL(SEQ ID NO: 1361436)
AhCas9 *Anaerostipes hadrus* NCBI Исходная последовательность: WP_044924278.1 Дикий тип	MQNGFLGITVSSEQVGWAVTNPKYELERASRKDLWGVRLFDKAETAEDRRMFRTNRRLNQRKKNRIHYLRDIFHEEVNQKD PNFFQQLDESNFCEDDRTVEFNFDTNLYKNQFPTVYHLRKYLMETKDKPDIRLVYLAFSKFMKNRGHFLYKGNLGEVMDFE NSMKGFCESLEKFNIDFPTLSDEQVKEVRDILCDHKIAKTVKKKNIITITKVKSKTAKAWIGLFCGCSVPVKVLFQDIDEE IVTDPEKISFEDASYDDYIANIEKGVGIYYEAIVSAKMLFDWSILNEILGDHQLLSDAMIAEYNKHHDDLKRLQKIIKGTG SRELYQDIFINDVSGNYVCYVGHAKTMSSADQKQFYTFLKNRLKNVNGISSEDAEWIDTEIKNGTLLPKQTKRDNSVIPHQ LQLREFELILDNMQEMYPFLKENREKLLKIFNFVIPYYVGPLKGVVRKGESTNWMVPKKDGVIHPWNFDEMVDKEASAECF ISRMTGNCSYLFNEKVLPKNSLLYETFEVLNELNPLKINGEPISVELKQRIYEQLFLTGKKVTKKSLTKYLIKNGYDKDIE LSGIDNEFHSNLKSHIDFEDYDNLSDEEVEQIILRITVFEDKQLLKDYLNREFVKLSEDERKQICSLSYKGWGNLSEMLLN GITVTDSNGVEVSVMDMLWNTNLNLMQILSKKYGYKAEIEHYNKEHEKTIYNREDLMDYLNIPPAQRRKVNQLITIVKSLK KTYGVPNKIFFKISREHQDDPKRTSSRKEQLKYLYKSLKSEDEKHLMKELDELNDHELSNDKVYLYFLQKGRCIYSGKKLN LSRLRKSNYQNDIDYIYPLSAVNDRSMNNKVLTGIQENRADKYTYFPVDSEIQKKMKGFWMELVLQGFMTKEKYFRLSREN DFSKSELVSFIERELSDNQQSGRMIASVLQYYFPESKIVFVKEKLISSFKRDFHLISSYGHNHLQAAKDAYITIVVGNVYH TKFTMDPAIYFKNHKRKDYDLNRLFLENISRDGQIAWESGPYGSIQTVRKEYAQNHIAVTKRVVEVKGGLFKQMPLKKGHG EYPLKTNDPRFGNIAQYGGYTNVTGSYFVLVESMEKGKKRISLEYVPVYLHERLEDDPGHKLLKEYLVDHRKLNHPKILLA KVRKNSLLKIDGFYYRLNGRSGNALILTNAVELIMDDWQTKTANKISGYMKRRAIDKKARVYQNEFHIQELEQLYDFYLDK LKNGVYKNRKNNQAELIHNEKEQFMELKTEDQCVLLTEIKKLFVCSPMQADLTLIGGSKHTGMIAMSSNVTKADFAVIAED PLGLRNKVIYSHKGEK(SEQ ID NO: 1361437)
KvCas9 *Kandleria vitulina* NCBI Исходная последовательность: WP_031589969.1 Дикий тип	MSQNNNKIYNIGLDIGDASVGWAVVDEHYNLLKRHGKHMWGSRLFTQANTAVERRSSRSTRRRYNKRRERIRLLREIMEDM VLDVDPTFFIRLANVSFLDQEDKKDYLKENYHSNYNLFIDKDFNDKTYYDKYPTIYHLRKHLCESKEKEDPRLIYLALHHI VKYRGNFLYEGQKFSMDVSNIEDKMIDVLRQFNEINLFEYVEDRKKIDEVLNVLKEPLSKKHKAEKAFALFDTTKDNKAAY KELCAALAGNKFNVTKMLKEAELHDEDEKDISFKFSDATFDDAFVEKQPLLGDCVEFIDLLLHDIYSWVELQNILGSAHTSE PSISAAMIQRYEDHKNDLKLLKDVIRKYLPKKYFEVFRDEKSKKNNYCNYINHPSKTPVDEFYKYIKKLIEKIDDPDVKTI LNKIELESFMLKQNSRTNGAVPYQMQLDELNKILENQSVYYSDLKDNEDKIRSILTFRIPYYFGPLNITKDRQFDWIIKKE GKENERILPWNANEIVDVDKTADEFIKRMRNFCTYFPDEPVMAKNSLTVSKYEVLNEINKLRINDHLIKRDMKDKMLHTLF MDHKSISANAMKKWLVKNQYFSNTDDIKIEGFQKENACSTSLTPWIDFTKIFGKINESNYDFIEKIIYDVTVFEDKKILRR RLKKEYDLDEEKIKKILKLKYSGWSRLSKKLLSGIKTKYKDSTRTPETVLEVMERTNMNLMQVINDEKLGFKKTIDDANST SVSGKFSYAEVQELAGSPAIKRGIWQALLIVDEIKKIMKHEPAHVYIEFARNEDEKERKDSFVNQMLKLYKDYDFEDETEK EANKHLKGEDAKSKIRSERLKLYYTQMGKCMYTGKSLDIDRLDTYQVDHIVPQSLLKDDSIDNKVLVLSSENQRKLDDLVI PSSIRNKMYGFWEKLFNNKIISPKKFYSLIKTEFNEKDQERFINRQIVETRQITKHVAQIIDNHYENTKVVTVRADLSHQF RERYHIYKNRDINDFHHAHDAYIATILGTYIGHRFESLDAKYIYGEYKRIFRNQKNKGKEMKKNNDGFILNSMRNIYADKD TGEIVWDPNYIDRIKKCFYYKDCFVTKKLEENNGTFFNVTVLPNDTNSDKDNTLATVPVNKYRSNVNKYGGFSGVNSFIVA IKGKKKKGKKVIEVNKLTGIPLMYKNADEEIKINYLKQAEDLEEVQIGKEILKNQLIEKDGGLYYIVAPTEIINAKQLILN ESQTKLVCEIYKAMKYKNYDNLDSEKIIDLYRLLINKMELYYPEYRKQLVKKFEDRYEQLKVISIEEKCNIIKQILATLHC NSSIGKIMYSDFKISTTIGRLNGRTISLDDISFIAESPTGMYSKKYKL(SEQ ID NO: 1361438)
EfCas9 *Enterococcus faecalis* NCBI Исходная последовательность: WP_016631044.1 Дикий тип	MRLFEEGHTAEDRRLKRTARRRISRRRNRLRYLQAFFEEAMTDLDENFFARLQESFLVPEDKKWHRHPIFAKLEDEVAYHE TYPTIYHLRKKLADSSEQADLRLIYLALAHIVKYRGHFLIEGKLSTENTSVKDQFQQFMVIYNQTFVNGESRLVSAPLPES VLIEEELTEKASRTKKSEKVLQQFPQEKANGLFGQFLKLMVGNKADFKKVFGLEEEAKITYASESYEEDLEGILAKVGDEY SDVFLAAKNVYDAVELSTILADSDKKSHAKLSSSMIVRFTEHQEDLKKFKRFIRENCPDEYDNLFKNEQKDGYAGYIAHAG KVSQLKFYQYVKKIIQDIAGAEYFLEKIAQENFLRKQRTFDNGVIPHQIHLAELQAIIHRQAAYYPFLKENQEKIEQLVTF RIPYYVGPLSKGDASTFAWLKRQSEEPIRPWNLQETVDLDQSATAFIERMTNFDTYLPSEKVLPKHSLLYEKFMVFNELTK ISYTDDRGIKANFSGKEKEKIFDYLFKTRRKVKKKDIIQFYRNEYNTEIVTLSGLEEDQFNASFSTYQDLLKCGLTRAELD HPDNAEKLEDIIKILTIFEDRQRIRTQLSTFKGQFSAEVLKKLERKHYTGWGRLSKKLINGIYDKESGKTILDYLVKDDGV SKHYNRNFMQLINDSQLSFKNAIQKAQSSEHEETLSETVNELAGSPAIKKGIYQSLKIVDELVAIMGYAPKRIVVEMAREN QTTSTGKRRSIQRLKIVEKAMAEIGSNLLKEQPTTNEQLRDTRLFLYYMQNGKDMYTGDELSLHRLSHYDIDHIIPQSFMK DDSLDNLVLVGSTENRGKSDDVPSKEVVKDMKAYWEKLYAAGLISQRKFQRLTKGEQGGLTLEDKAHFIQRQLVETRQITK NVAGILDQRYNAKSKEKKVQIITLKASLTSQFRSIFGLYKVREVNDYHHGQDAYLNCVVATTLLKVYPNLAPEFVYGEYPK FQTFKENKATAKAIIYTNLLRFFTEDEPRFTKDGEILWSNSYLKTIKKELNYHQMNIVKKVEVQKGGFSKESIKPKGPSNK LIPVKNGLDPQKYGGFDSPVVAYTVLFTHEKGKKPLIKQEILGITIMEKTRFEQNPILFLEEKGFLRPRVLMKLPKYTLYE

Описание	Последовательность
	FPEGRRRLLASAKEAQKGNQMVLPEHLLTLLYHAKQCLLPNQSESLAYVEQHQPEFQEILERVVDFAEVHTLAKSKVQQIV KLFEANQTADVKEIAASFIQLMQFNAMGAPSTFKFFQKDIERARYTSIKEIFDATIIYQSPTGLYETRRKVVD(SEQ ID NO: 1361439)
Staphylococcus aureus Cas9	KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSE LSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGE VRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEEL RSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTN LKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDE LWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDA QKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNS FNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDT RYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMEN QMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDK DNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLD ITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIK INGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG(SEQ ID NO: 1361440)
Geobacillus thermodenitrificans Cas9	MKYKIGLDIGITSIGWAVINLDIPRIEDLGVRIFDRAENPKTGESLALPRRLARSARRRLRRRKHRLERIRRLFVREGILT KEELNKLFEKKHEIDVWQLRVEALDRKLNNDELARILLHLAKRRGFRSNRKSERTNKENSTMLKHIEENQSILSSYRTVAE MVVKDPKFSLHKRNKEDNYTNTVARDDLEREIKLIFAKQREYGNIVCTEAFEHEYISIWASQRPFASKDDIEKKVGFCTFE PKEKRAPKATYTFQSFTVWEHINKLRLVSPGGIRALTDDERRLIYKQAFHKNKITFHDVRTLLNLPDDTRFKGLLYDRNTT LKENEKVRFLELGAYHKIRKAIDSVYGKGAAKSFRPIDFDTFGYALTMFKDDTDIRSYLRNEYEQNGKRMENLADKVYDEE LIEELLNLSFSKFGHLSLKALRNILPYMEQGEVYSTACERAGYTFTGPKKKQKTVLLPNIPPIANPVVMRALTQARKVVNA IIKKYGSPVSIHIELARELSQSFDERRKMQKEQEGNRKKNETAIRQLVEYGLTLNPTGLDIVKFKLWSEQNGKCAYSLQPI EIERLLEPGYTEVDHVIPYSRSLDDSYTNKVLVLTKENREKGNRTPAEYLGLGSERWQQFETFVLTNKQFSKKKRDRLLRL HYDENEENEFKNRNLNDTRYISRFLANFIREHLKFADSDDKQKVYTVNGRITAHLRSRWNFNKNREESNLHHAVDAAIVAC TTPSDIARVTAFYQRREQNKELSKKTDPQFPQPWPHFADELQARLSKNPKESIKALNLGNYDNEKLESLQPVFVSRMPKRS ITGAAHQETLRRYIGIDERSGKIQTVVKKKLSEIQLDKTGHFPMYGKESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKK NGELGPIIRTIKIIDTTNQVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCVPIYTIDMMKGILPNKAIEPNKPYSEWKEMTE DYTFRFSLYPNDLIRIEFPREKTIKTAVGEEIKIKDLFAYYQTIDSSNGGLSLVSHDNNFSLRSIGSRTLKRFEKYQVDVL GNIYKVRGEKRVGVASSSHSKAGETIRPL(SEQ ID NO: 1361441)
ScCas9 *S. canis* 1375 AA 159.2 кДа	MEKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTNRKSIKKNLMGALLFDSGETAEATRLKRTARRRYTRRKNRIRY LQEIFANEMAKLDDSFFQRLEESFLVEEDKKNERHPIFGNLADEVAYHRNYPTIYHLRKKLADSPEKADLRLIYLALAHII KFRGHFLIEGKLNAENSDVAKLFYQLIQTYNQLFEESPLDEIEVDAKGILSARLSKSKRLEKLIAVFPNEKKNGLFGNIIA LALGLTPNFKSNFDLTEDAKLQLSKDTYDDDLDELLGQIGDQYADLFSAAKNLSDAILLSDILRSNSEVTKAPLSASMVKR YDEHHQDLALLKTLVRQQFPEKYAEIFKDDTKNGYAGYVGIGIKHRKRTTKLATQEEFYKFIKPILEKMDGAEELLAKLNR DDLLRKQRTFDNGSIPHQIHLKELHAILRRQEEFYPFLKENREKIEKILTFRIPYYVGPLARGNSRFAWLTRKSEEAITPW NFEEVVDKGASAQSFIERMTNFDEQLPNEKVLPKHSLLYEYFTVYNELTKVKYVTERMRKPEFLSGEQKKAIVDLLFKTNR KVTVKQLKEDYFKKIECFDSVEIIGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRHYTGWGRLSRKMINGIRDKQSGKTILDFLKSDGFSNRNFMQLIHDDSLTFKEEIEKAQVSGQGDS LHEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTTKGLQQSRERKKRIEEGIKELESQILKENPV ENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSVENRGKSDNVPSEEVVKKMKNY WRQLLNAKLITQRKFDNLTKAERGGLSEADKAGFIKRQLVETRQITKHVARILDSRMNTKRDKNDKPIREVKVITLKSKLV SDFRKDFQLYKVRDINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKRFFYSNIMN FFKTEVKLANGEIRKRPLIETNGETGEVVWNKEKDFATVRKVLAMPQVNIVKKTEVQTGGFSKESILSKRESAKLIPRKKG WDTRKYGGFGSPTVAYSILVVAKVEKGKAKKLKSVKVLVGITIMEKGSYEKDPIGFLEAKGYKDIKKELIFKLPKYSLFEL ENGRRRMLASATELQKANELVLPQHLVRLLYYTQNISATTGSNNLGYIEQHREEFKEIFEKIIDFSEKYILKNKVNSNLKS SFDEQFAVSDSILLSNSFVSLLKYTSFGASGGFTFLDLDVKQGRLRYQTVTEVLDATLIYQSITGLYETRTDLSQLGGD (SEQ ID NO: 1361442)

[0350] Описанные здесь праймерные редакторы могут включать любую из указанных выше последовательностей ортолога Cas9 или любой его вариант, имеющий, по меньшей мере 80%, по меньшей мере, 85%, по меньшей мере, 90%, по меньшей мере, 95%, или, по меньшей мере, 99% идентичность относительно них.

[0351] napDNAbp может включать любые подходящие гомологи и/или ортологи или встречающиеся в природе ферменты, например Cas9. Гомологи и/или ортологи Cas9 были описаны у различных видов, включая, но без ограничения, *S. pyogenes* и *S. thermophilus*. Предпочтительно, фрагмент Cas имеет конфигурацию (например мутагенизирован, рекомбинантно сконструирован, или в ином случае получен из природы) в виде никазы, то

есть способен расщеплять только одну цепь мишени. Дополнительные подходящие нуклеазы и последовательности Cas9 будут очевидны специалистам в данной области на основании данного описания, при этом подобные нуклеазы и последовательности Cas9 включают последовательности Cas9 из организмов и локусов, описанных в Chylinski, Rhun, and Charpentier, “The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems” (2013) RNA Biology 10:5, 726-737, полное содержание которого приведено здесь в качестве ссылки. В некоторых вариантах осуществления нуклеаза Cas9 имеет неактивный (например деактивированный) домен расщепления ДНК, то есть Cas9 представляет собой никазу. В некоторых вариантах осуществления белок Cas9 включает аминокислотную последовательность, которая, по меньшей мере, на 80% идентична аминокислотной последовательности белка Cas9, представленной в рамках любого одного из вариантов в Таблице 3. В некоторых вариантах осуществления белок Cas9 включает аминокислотную последовательность, которая, по меньшей мере, на 85%, по меньшей мере, на 90%, по меньшей мере, на 92%, по меньшей мере, на 95%, по меньшей мере, на 96%, по меньшей мере, на 97%, по меньшей мере, на 98%, по меньшей мере, на 99%, или, по меньшей мере, на 99,5% идентична аминокислотной последовательности белка Cas9 любого одного из ортологов Cas9 в приведенных выше таблицах.

(iii)Вариант в виде мертвого Cas9

[0352] В некоторых вариантах осуществления описанные здесь праймерные редакторы могут включать мертвый Cas9, например мертвый SpCas9, который не имеет нуклеазной активности вследствие одной или нескольких мутаций, которые деактивируют оба нуклеазных домена Cas9, а именно домен RuvC (который расщепляет непротоспейсерную цепь ДНК) и домен HNH (который расщепляет протоспейсерную цепь ДНК). Нуклеазная деактивация может быть вызвана одной или несколькими мутациями, которые приводят к одному или нескольким замещениям и/или делециям в аминокислотной последовательности кодируемого белка или любого его варианта, имеющего, по меньшей мере 80%, по меньшей мере, 85%, по меньшей мере, 90%, по меньшей мере 95%, или, по меньшей мере, 99% идентичность последовательности относительно нее.

[0353] В соответствии с используемым здесь значением термин "dCas9" обозначает Cas9 с деактивированной нуклеазной активностью Cas9 или не имеющий нуклеазной активности Cas9 или его функциональный фрагмент и охватывает любой встречающийся в природе dCas9 из любого организма, любой эквивалент встречающегося в природе dCas9 или его

функциональный фрагмент, любой гомолог dCas9, ортолог, или паралог из любого организма и любой мутант или вариант dCas9, являющийся встречающимся в природе или сконструированным. Термин "dCas9" не является конкретно ограничивающим и может рассматриваться как "dCas9 или эквивалент". Примеры белков dCas9 и способ получения белков dCas9 дополнительно описаны здесь и/или описаны в данной области и приведены здесь в качестве ссылки.

[0354] В других вариантах осуществления dCas9 соответствует или включает частично или полностью аминокислотную последовательность Cas9, имеющую одну или несколько мутаций, которые деактивируют нуклеазную активность Cas9. В других вариантах осуществления предусмотрены варианты Cas9, имеющие мутации, отличные от D10A и H840A, которые могут приводить к полной или частичной деактивации эндогенной нуклеазной активности Cas9 (например nCas9 или dCas9, соответственно). В качестве примера такие мутации включают другие аминокислотные замещения в D10 и H820 или другие замещения в нуклеазных доменах Cas9 (например замещения в нуклеазном субдомене HNH и/или субдомене RuvC1) со ссылкой на последовательность дикого типа, например Cas9 из *Streptococcus pyogenes* (кат. # NCBI: NC_017053.1 (SEQ ID NO: 1361424)). В некоторых вариантах осуществления предусмотрены варианты или гомологи Cas9 (например варианты Cas9 из Streptococcus pyogenes (кат. # NCBI: NC_017053.1 (SEQ ID NO: 1361424))), которые, по меньшей мере, на 70% идентичны, по меньшей мере, на 80% идентичны, по меньшей мере, на 90% идентичны, по меньшей мере, на 95% идентичны, по меньшей мере, на 98% идентичны, по меньшей мере, на 99% идентичны, по меньшей мере, на 99,5% идентичны, или, по меньшей мере, на 99,9% идентичны последовательности NCBI NC_017053.1 (SEQ ID NO: 1361424). В некоторых вариантах осуществления предусмотрены варианты dCas9 (например варианты кат. # NCBI NC_017053.1 (SEQ ID NO: 1361424)), имеющие аминокислотные последовательности, которые являются более короткими или более длинными, чем NC_017053.1 (SEQ ID NO: 1361424) приблизительно на 5 аминокислот, приблизительно на 10 аминокислот, приблизительно на 15 аминокислот, приблизительно на 20 аминокислот, приблизительно на 25 аминокислот, приблизительно на 30 аминокислот, приблизительно на 40 аминокислот, приблизительно на 50 аминокислот, приблизительно на 75 аминокислот, приблизительно на 100 аминокислот или более.

[0355] В одном варианте осуществления мертвый Cas9 может быть основан на стандартной последовательности SpCas9 Q99ZW2 и может иметь следующую последовательность, которая включает замещения D10A и H810A (подчеркнуты или указаны жирным шрифтом), или являться вариантом SEQ ID NO: 1361444, имеющим, по меньшей мере, 80%, по меньшей мере, 85%, по меньшей мере, 90%, по меньшей мере, 95%, или, по меньшей мере, 99% идентичность последовательности относительно нее:

Описание	Последовательность	SEQ ID NO:
мертвый Cas9 или dCas9 *Streptococcus pyogenes* Q99ZW2 Cas9 с D10**X** и H810**X** Где "X" представляет собой любую аминокислоту	MDKKYSIGL**X**IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKK LVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQL VQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI GDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHH QDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFI KPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSE ETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTG WGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK EDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEH PVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD**X**IVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLIT QRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMN TKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFA TVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFE KNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQK GNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLS QLGGD	SEQ ID NO: 1361443
мертвый Cas9 или dCas9 *Streptococcus pyogenes* Q99ZW2 Cas9 с D10**A** и H810**A**	MDKKYSIGL**A**IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKK LVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQL VQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI GDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHH QDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFI KPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSE ETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTG WGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK EDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEH PVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD**A**IVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLIT QRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMN TKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFA TVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFE KNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQK GNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLS QLGGD	SEQ ID NO: 1361444

(iv) Вариант в виде никазы Cas9

[0356] В одном варианте осуществления описанные здесь праймерные редакторы включают никазу Cas9. Термин "никаза Cas9" или "nCas9" обозначает вариант Cas9, который способен вносить одноцепочечный разрыв в двуцепочечную молекулу-мишень

ДНК. В некоторых вариантах осуществления никаза Cas9 включает только один функционирующий нуклеазный домен. Cas9 дикого типа (например стандартный SpCas9) включает два отдельных нуклеазных домена, а именно домен RuvC (который расщепляет непротоспейсерную цепь ДНК) и домен HNH (который расщепляет протоспейсерную цепь ДНК). В одном варианте осуществления никаза Cas9 включает мутацию в домене RuvC, которая деактивирует нуклеазную активность RuvC. Например, сообщалось о мутациях в аспартате (D) 10, гистидине (H) 983, аспартате (D) 986, или глутамате (E) 762 как о связанных с потерей функции мутациях нуклеазного домена RuvC и создании функциональной никазы Cas9 (например Nishimasu *et al.*, "Crystal structure of Cas9 in complex with guide RNA and target DNA," *Cell* 156(5), 935 - 949, который приведен здесь в качестве ссылки). Таким образом, никазные мутации в домене RuvC могут включать D10X, H983X, D986X, или E762X, где X представляет собой любую аминокислоту, отличную от аминокислоты дикого типа. В некоторых вариантах осуществления никаза может представлять собой D10A, или H983A, или D986A, или E762A, или их комбинацию.

[0357] В различных вариантах осуществления никаза Cas9 может иметь мутацию в нуклеазном домене RuvC и иметь одну из следующих аминокислотных последовательностей или являться ее вариантом, имеющим аминокислотную последовательность, которая имеет, по меньшей мере, 80%, по меньшей мере, 85%, по меньшей мере 90%, по меньшей мере, 95%, или, по меньшей мере, 99% идентичность последовательности относительно нее.

Описание	Последовательность	SEQ ID NO:
Никаза Cas9 *Streptococcus pyogenes* Q99ZW2 Cas9 с D10**X**, где X представляет собой любую альтернативную аминокислоту	MDKKYSIGL**X**IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEAT**R**LKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTI**Y**HLRKKLVDSTDKADLRLIYLALA HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLENLIAQLPGEKKNGLFGN**L**IA**L**SLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILR RQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYV TEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGV EDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVK VVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK DDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESE FVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKR PLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPK RNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGE LQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQ ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAEN**I**IH**L**FTLTNLGAPAAFK YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	SEQ ID NO: 1361445

Описание	Последовательность	SEQ ID NO:
Никаза Cas9 *Streptococcus pyogenes* Q99ZW2 Cas9 с E762X, где X представляет собой любую альтернативную аминокислоту	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILR RQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYV TEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGV EDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVK VVDELVKVMGRHKPENIVI**X**MARENQTTQKGQKNSRERMKRIEEGIKELGSQI LKEHPVENTQLQNEKLYLY**Y**LQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK DDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESE FVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKR PLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPK RNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGE LQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQ ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	SEQ ID NO: 1361446
Никаза Cas9 *Streptococcus pyogenes* Q99ZW2 Cas9 с H983X, где X представляет собой любую альтернативную аминокислоту	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILR RQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYV TEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGV EDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVK VVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK DDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK VITLKSKLVSDFRKDFQFYKVREINNYH**X**AHDAYLNAVVGTALIKKYPKLESE FVYGDYKVYDVRKMIAKSEQEIGKATAK**Y**FFYSNIMNFFKTEITLANGEIRKR PLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPK RNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGE LQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQ ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	SEQ ID NO: 1361447
Никаза Cas9 *Streptococcus pyogenes* Q99ZW2 Cas9 с D986X, где X представляет собой любую альтернативную аминокислоту	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILR RQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYV TEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGV EDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVK VVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK DDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK VITLKSKLVSDFRKDFQFYKVREINNYHHAH**X**AYLNAVVGTALIKKYPKLESE FVYGDYKVYDVRKMIAKSEQEIGKATAKYFF**Y**SNIMNFFKTEITLANGEIRKR PLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPK RNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGE LQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQ ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	SEQ ID NO: 1361448
Никаза Cas9 *Streptococcus pyogenes* Q99ZW2 Cas9 с D10**A**	MDKKYSIGL**A**IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILR RQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYV TEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGV EDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL	SEQ ID NO: 1361449

Описание	Последовательность	SEQ ID NO:
	KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVK VVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK DDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESE FVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKR PLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPK RNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGE LQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQ ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	
Никаза Cas9 *Streptococcus pyogenes* Q99ZW2 Cas9 с E762A	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILR RQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYV TEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGV EDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVK VVDELVKVMGRHKPENIVI**E**MARENQTTQKGQKNSRERMKRIEEGIKELGSQI LKEHPVENTQLQNEKLYLY**Y**LQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK DDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESE FVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKR PLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPK RNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGE LQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQ ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	SEQ ID NO: 1361450
Никаза Cas9 *Streptococcus pyogenes* Q99ZW2 Cas9 с H983A	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILR RQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYV TEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGV EDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVK VVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK DDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK VITLKSKLVSDFRKDFQFYKVREINNYH**A**AHDAYLNAVVGTALIKKYPKLESE FVYGDYKVYDVRKMIAKSEQEIGKATAK**Y**FFYSNIMNFFKTEITLANGEIRKR PLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPK RNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGE LQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQ ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	SEQ ID NO: 1361451
Никаза Cas9 *Streptococcus pyogenes* Q99ZW2 Cas9 с D986A	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILR RQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYV TEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGV EDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVK VVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK DDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK VITLKSKLVSDFRKDFQFYKVREINNYHHAH**A**AYLNAVVGTALIKKYPKLESE FVYGDYKVYDVRKMIAKSEQEIGKATAKYFF**Y**SNIMNFFKTEITLANGEIRKR PLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPK RNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGE LQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQ	SEQ ID NO: 1361452

Описание	Последовательность	SEQ ID NO:
	ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	

[0358] В другом варианте осуществления никаза Cas9 включает мутацию в домене HNH, которая деактивирует нуклеазную активность HNH. Например, сообщалось о мутациях в гистидине (H) 840 или аспарагине (R) 863 как о связанных с потерей функции мутациях нуклеазного домена HNH и создании функциональной никазы Cas9 (например Nishimasu *et al.*, "Crystal structure of Cas9 in complex with guide RNA and target DNA," *Cell* 156(5), 935–949, который приведен здесь в качестве ссылки). Таким образом, никазные мутации в домене HNH могут включать H840X и R863X, где X представляет собой любую аминокислоту, отличную от аминокислоты дикого типа. В некоторых вариантах осуществления никаза может представлять собой H840A или R863A, или их комбинацию.

[0359] В различных вариантах осуществления никаза Cas9 может иметь мутацию в нуклеазном домене HNH и иметь одну из следующих аминокислотных последовательностей или являться ее вариантом, имеющим аминокислотную последовательность, которая имеет, по меньшей мере, 80%, по меньшей мере, 85%, по меньшей мере 90%, по меньшей мере, 95%, или, по меньшей мере, 99% идентичность последовательности относительно нее.

Описание	Последовательность	SEQ ID NO:
Никаза Cas9 *Streptococcus pyogenes* Q99ZW2 Cas9 с H840**X**, где X представляет собой любую альтернативную аминокислоту	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGN TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTR RKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLF IQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLA EDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQE EFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN GSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKIL TFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEV VDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKT NRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLG TYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDR EMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKL INGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLT FKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLY YLQNGRDMYVDQELDINRLSDYDVD**X**IVPQSFLKDDS IDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKL VSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIK KYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKY FFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESI LPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQ KGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLF VEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYN KHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID	SEQ ID NO: 1361453

Описание	Последовательность	SEQ ID NO:
	RKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	
Никаза Cas9 *Streptococcus pyogenes* Q99ZW2 Cas9 с H840**A**, где A представляет собой любую альтернативную аминокислоту	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGN TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTR RKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLF IQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLA EDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQE EFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN GSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKIL TFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEV VDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKT NRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLG TYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDR EMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKL INGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLT FKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLY YLQNGRDMYVDQELDINRLSDYDVD**A**IVPQSFLKDDS IDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKL VSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIK KYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKY FFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESI LPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQ KGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLF VEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYN KHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID RKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	SEQ ID NO: 1361454
Никаза Cas9 *Streptococcus pyogenes* Q99ZW2 Cas9 с R863X, где X представляет собой любую альтернативную аминокислоту	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGN TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTR RKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLF IQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLA EDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQE EFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN GSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKIL TFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEV VDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKT NRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLG TYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDR EMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKL INGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLT FKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLY YLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDS IDNKVLTRSDKN**X**GKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNL**T**KAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKL VSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIK KYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKY FFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESI LPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQ KGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLF VEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYN KHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID RKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	SEQ ID NO: 1361455
Никаза Cas9 *Streptococcus pyogenes* Q99ZW2 Cas9 с R863**A**, где A	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGN TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTR RKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLF IQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLA EDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQE EFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN GSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKIL	SEQ ID NO: 1361456

Описание	Последовательность	SEQ ID NO:
представляет собой любую альтернативную аминокислоту	TFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEV VDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKT NRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLG TYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDR EMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKL INGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLT FKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLY YLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDS IDNKVLTRSDKN**A**GKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKL VSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIK KYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKY FFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESI LPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQ KGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLF VEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYN KHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID RKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	

(v) Другие варианты Cas9

[0360] Помимо мертвого Cas9 и вариантов никазы Cas9, используемые здесь белки Cas9 могут также включать другие "варианты Cas9", являющиеся, по меньшей мере, приблизительно на 70% идентичными, по меньшей мере, приблизительно на 80% идентичными, по меньшей мере, приблизительно на 90% идентичными, по меньшей мере, приблизительно на 95% идентичными, по меньшей мере, приблизительно на 96% идентичными, по меньшей мере, приблизительно на 97% идентичными, по меньшей мере, приблизительно на 98% идентичными, по меньшей мере, приблизительно на 99% идентичными, по меньшей мере, приблизительно на 99,5% идентичными, или, по меньшей мере, приблизительно на 99,9% идентичными любому исходному белку Cas9, включая любой Cas9 дикого типа или мутант Cas9 (например мертвый Cas9 или никазу Cas9), или фрагмент Cas9, или циркулярный пермутант Cas9, или другой описанный здесь или известный в данной области вариант Cas9. В некоторых вариантах осуществления вариант Cas9 может иметь 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 или более аминокислотных изменений по сравнению с исходным Cas9. В некоторых вариантах осуществления вариант Cas9 включает фрагмент исходного Cas9 (например связывающий нРНК домен или домен расщепления ДНК), так что фрагмент является, по меньшей мере, приблизительно на 70% идентичным, по меньшей мере, приблизительно на 80% идентичным, по меньшей мере, приблизительно на 90% идентичным, по меньшей мере, приблизительно на 95% идентичным, по меньшей мере, приблизительно на 96% идентичным, по меньшей мере, приблизительно на 97%

идентичным, по меньшей мере, приблизительно на 98% идентичным, по меньшей мере, приблизительно на 99% идентичным, по меньшей мере, приблизительно на 99,5% идентичным, или, по меньшей мере, приблизительно на 99,9% идентичным соответствующему фрагменту Cas9 дикого типа. В некоторых вариантах осуществления фрагмент имеет длину, соответствующую, по меньшей мере, на 30%, по меньшей мере, на 35%, по меньшей мере, на 40%, по меньшей мере, на 45%, по меньшей мере, на 50%, по меньшей мере, на 55%, по меньшей мере, на 60%, по меньшей мере, на 65%, по меньшей мере, на 70%, по меньшей мере, на 75%, по меньшей мере, на 80%, по меньшей мере, на 85%, по меньшей мере, на 90%, по меньшей мере, на 95%, по меньшей мере, на 96%, по меньшей мере, на 97%, по меньшей мере, на 98%, по меньшей мере, на 99%, или, по меньшей мере, на 99,5% длине аминокислоты соответствующего Cas9 дикого типа (например SEQ ID NO: 1361421).

[0361] В некоторых вариантах осуществления в изобретении также могут использоваться фрагменты Cas9, которые сохраняют свою функциональность и которые представляют собой фрагменты любого описанного здесь белка Cas9. В некоторых вариантах осуществления фрагмент Cas9 имеет длину, по меньшей мере, 100 аминокислот. В некоторых вариантах осуществления фрагмент имеет длину, по меньшей мере, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, или, по меньшей мере, 1300 аминокислот.

[0362] В различных вариантах осуществления описанные здесь праймерные редакторы могут включать один из вариантов Cas9, описанных далее, или его вариант, который является, по меньшей мере, приблизительно на 70% идентичным, по меньшей мере, приблизительно на 80% идентичным, по меньшей мере, приблизительно на 90% идентичным, по меньшей мере, приблизительно на 95% идентичным, по меньшей мере, приблизительно на 96% идентичным, по меньшей мере, приблизительно на 97% идентичным, по меньшей мере, приблизительно на 98% идентичным, по меньшей мере, приблизительно на 99% идентичным, по меньшей мере, приблизительно на 99,5% идентичным, или, по меньшей мере, приблизительно на 99,9% идентичным любым исходным вариантам Cas9.

(vi) Варианты Cas9 малого размера

[0363] В некоторых вариантах осуществления указанные здесь праймерные редакторы могут включать белок Cas9, который имеет меньшую молекулярную массу, чем стандартная последовательность SpCas9. В некоторых вариантах осуществления варианты Cas9 малого размера могут обеспечивать доставку в клетки, например с помощью вектора экспрессии, наночастицы, или других средств доставки. В определенных вариантах осуществления варианты Cas9 малого размера могут включать ферменты, относящиеся к ферментам типа II систем CRISPR-Cas Класса 2. В некоторых вариантах осуществления варианты Cas9 малого размера могут включать ферменты, относящиеся к ферментам типа V систем CRISPR-Cas Класса 2. В других вариантах осуществления варианты Cas9 малого размера могут включать ферменты, относящиеся к ферментам типа VI систем CRISPR-Cas Класса 2.

[0364] Стандартный белок SpCas9 имеет длину 1368 аминокислот и имеет предполагаемую молекулярную массу 158 кДа. В соответствии с используемым здесь значением термин "вариант Cas9 малого размера" обозначает любой вариант Cas9 — встречающийся в природе, сконструированный или полученный иным образом — длиной менее чем, по меньшей мере, 1300 аминокислот, или менее чем, по меньшей мере, 1290 аминокислот, или менее чем 1280 аминокислот, или менее чем 1270 аминокислот, или менее чем 1260 аминокислот, или менее чем 1250 аминокислот, или менее чем 1240 аминокислот, или менее чем 1230 аминокислот, или менее чем 1220 аминокислот, или менее чем 1210 аминокислот, или менее чем 1200 аминокислот, или менее чем 1190 аминокислот, или менее чем 1180 аминокислот, или менее чем 1170 аминокислот, или менее чем 1160 аминокислот, или менее чем 1150 аминокислот, или менее чем 1140 аминокислот, или менее чем 1130 аминокислот, или менее чем 1120 аминокислот, или менее чем 1110 аминокислот, или менее чем 1100 аминокислот, или менее чем 1050 аминокислот, или менее чем 1000 аминокислот, или менее чем 950 аминокислот, или менее чем 900 аминокислот, или менее чем 850 аминокислот, или менее чем 800 аминокислот, или менее чем 750 аминокислот, или менее чем 700 аминокислот, или менее чем 650 аминокислот, или менее чем 600 аминокислот, или менее чем 550 аминокислот, или менее чем 500 аминокислот, но, по меньшей мере, более чем приблизительно 400 аминокислот, сохраняющий требуемые функции белка Cas9. Варианты Cas9 могут

включать варианты, относящиеся к ферментам типа II, типа V, или типа VI системы CRISPR-Cas Класса 2.

[0365] В различных вариантах осуществления описанные здесь праймерные редакторы могут включать один из вариантов Cas9 малого размера, описанных далее, или вариант Cas9, который является, по меньшей мере, приблизительно на 70% идентичным, по меньшей мере, приблизительно на 80% идентичным, по меньшей мере, приблизительно на 90% идентичным, по меньшей мере, приблизительно на 95% идентичным, по меньшей мере, приблизительно на 96% идентичным, по меньшей мере, приблизительно на 97% идентичным, по меньшей мере, приблизительно на 98% идентичным, по меньшей мере, приблизительно на 99% идентичным, по меньшей мере, приблизительно на 99,5% идентичным, или, по меньшей мере, приблизительно на 99,9% идентичным любому исходному белку Cas9 малого размера.

Описание	Последовательность	SEQ ID NO:
SaCas9 *Staphylococcus aureus* 1053 AA 123 кДа	MGKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRG ARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEF SAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERL KKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTY YEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNN LVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTS TGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTN LNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKL VPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIII ELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDM QEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSK KGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRF SVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKW KFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESM PEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKD DKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQY GDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNS RNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKL KKISNQAEFIASFYKNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLE NMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKK	SEQ ID NO: 1361457, тот же, что и 1361431
NmeCas9 *N. meningitidis* 1083 AA 124,5 кДа	MAAFKPNSINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGVRVFERAEVPK TGDSLAMARRLARSVRRLTRRRAHRLLRTRRLLKREGVLQAANFDENGLIKSL PNTPWQLRAAALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALL KGVAGNAHALQTGDFRTPAELALNKFEKESGHIRNQRSDYSHTFSRKDLQAEL ILLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDAVQKMLGHCTFEPAEP KAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSKLTYA QARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSP LNLSPELQDEIGTAFSLFKTDEDITGRLKDRIQPEILEALLKHISFDKFVQIS LKALRRIVPLMEQGKRYDEACAEIYGDHYGKKNTEEKIYLPPIPADEIRNPVV LRALSQARKVINGVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEENRKDR EKAAAKFREYFPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLGRLNEKGY VEIDAALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFK ARVETSRFPRSKKQRILLQKFDEDGFKERNLNDTRYVNRFLCQFVADRMRLTG KGKKRVFASNGQITNLLRGFWGLRKVRAENDRHHALDAVVVACSTVAMQQKIT RFVRYKEMNAFDGKTIDKETGEVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKP EFEEADTLEKLRTLLAEKLSSRPEAVHEYVTPLFVSRAPNRKMSGQGHMETVK SAKRLDEGVSVLRVPLTQLKLKDLEKMVNREREPKLYEALKARLEAHKDDPAK AFAEPFYKYDKAGNRTQQVKAVRVEQVQKTGVWVRNHNGIADNATMVRVDVFE KGDKYYLVPIYSWQVAKGILPDRAVVQGKDEEDWQLIDDSFNFKFSLHPNDLV EVITKKARMFGYFASCHRGTGNINIRIHDLDHKIGKNGILEGIGVKTALSFQK YQIDELGKEIRPCRLKKRPPVR	SEQ ID NO: 1361458
CjCas9 *C. jejuni* 984 AA 114,9 кДа	MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRLARS ARKRLARRKARLNHLKHLIANEFKLNYEDYQSFDESLAKAYKGSLISPYELRF RALNELLSKQDFARVILHIAKRRGYDDIKNSDDKEKGAILKAIKQNEEKLANY QSVGEYLYKEYFQKFKENSKEFTNVRNKKESYERCIAQSFLKDELKLIFKKQR EFGFSFSKKFEEEVLSVAFYKRALKDFSHLVGNCSFFTDEKRAPKNSPLAFMF VALTRIINLLNNLKNTEGILYTKDDLNALLNEVLKNGTLTYKQTKKLLGLSDD YEFKGEKGTYFIEFKKYKEFIKALGEHNLSQDDLNEIAKDITLIKDEIKLKKA LAKYDLNQNQIDSLSKLEFKDHLNISFKALKLVTPLMLEGKKYDEACNELNLK VAINEDKKDFLPAFNETYYKDEVTNPVVLRAIKEYRKVLNALLKKYGKVHKIN IELAREVGKNHSQRAKIEKEQNENYKAKKDAELECEKLGLKINSKNILKLRLF KEQKEFCAYSGEKIKISDLQDEKMLEIDHIYPYSRSFDDSYMNKVLVFTKQNQ EKLNQTPFEAFGNDSAKWQKIEVLAKNLPTKKQKRILDKNYKDKEQKNFKDRN LNDTRYIARLVLNYTKDYLDFLPLSDDENTKLNDTQKGSKVHVEAKSGMLTSA	SEQ ID NO: 1361459

Описание	Последовательность	SEQ ID NO:
	LRHTWGFSAKDRNNHLHHAIDAVIIAYANNSIVKAFSDFKKEQESNSAELYAK KISELDYKNKRKFFEPFSGFRQKVLDKIDEIFVSKPERKKPSGALHEETFRKE EEFYQSYGGKEGVLKALELGKIRKVNGKIVKNGDMFRVDIFKHKKTNKFYAVP IYTMDFALKVLPNKAVARSKKGEIKDWILMDENYEFCFSLYKDSLILIQTKDM QEPEFVYYNAFTSSTVSLIVSKHDNKFETLSKNQKILFKNANEKEVIAKSIGI QNLKVFEKYIVSALGEVTKAEFRQREDFKK	
GeoCas9 *G. stearotherm ophilus* 1087 AA 127 кДа	MRYKIGLDIGITSVGWAVMNLDIPRIEDLGVRIFDRAENPQTGESLALPRRLA RSARRRLRRRKHRLERIRRLVIREGILTKEELDKLFEEKHEIDVWQLRVEALD RKLNNDELARVLLHLAKRRGFKSNRKSERSNKENSTMLKHIEENRAILSSYRT VGEMIVKDPKFALHKRNKGENYTNTIARDDLEREIRLIFSKQREFGNMSCTEE FENEYITIWASQRPVASKDDIEKKVGFCTFEPKEKRAPKATYTFQSFIAWEHI NKLRLISPSGARGLTDEERRLLYEQAFQKNKITYHDIRTLLHLPDDTYFKGIV YDRGESRKQNENIRFLELDAYHQIRKAVDKVYGKGKSSSFLPIDFDTFGYALT LFKDDADIHSYLRNEYEQNGKRMPNLANKVYDNELIEELLNLSFTKFGHLSLK ALRSILPYMEQGEVYSSACERAGYTFTGPKKKQKTMLLPNIPPIANPVVMRAL TQARKVVNAIIKKYGSPVSIHIELARDLSQTFDERRKTKKEQDENRKKNETAI RQLMEYGLTLNPTGHDIVKFKLWSEQNGRCAYSLQPIEIERLLEPGYVEVDHV IPYSRSLDDSYTNKVLVLTRENREKGNRIPAEYLGVGTERWQQFETFVLTNKQ FSKKKRDRLLRLHYDENEETEFKNRNLNDTRYISRFFANFIREHLKFAESDDK QKVYTVNGRVTAHLRSRWEFNKNREESDLHHAVDAVIVACTTPSDIAKVTAFY QRREQNKELAKKTEPHFPQPWPHFADELRARLSKHPKESIKALNLGNYDDQKL ESLQPVFVSRMPKRSVTGAAHQETLRRYVGIDERSGKIQTVVKTKLSEIKLDA SGHFPMYGKESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKKNGEPGPVIRT VKIIDTKNQVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCVPVYTMDIMKGILP NKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIELPREKTVKTAAGEEINVK DVFVYYKTIDSANGGLELISHDHRFSLRGVGSRTLKRFEKYQVDVLGNIYKVR GEKRVGLASSAHSKPGKTIRPLQSTRD	SEQ ID NO: 1361460
LbaCas12a *L. bacterium* 1228 AA 143,9 кДа	MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYKGVKKL LDRYYLSFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIA KAFKGNEGYKSLFKKDIIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRE NMFSEEAKSTSIAFRCINENLTRYISNMDIFEKVDAIFDKHEVQEIKEKILNS DYDVEDFFEGEFFNFVLTQEGIDVYNAIIGGFVTESGEKIKGLNEYINLYNQK TKQKLPKFKPLYKQVLSDRESLSFYGEGYTSDEEVLEVFRNTLNKNSEIFSSI KKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDIHL KKKAVVTEKYEDDRRKSFKKIGSFSLEQLQEYADADLSVVEKLKEIIIQKVDE IYKVYGSSEKLFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEG KETNRDESFYGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQF MGGWDKDKETDYRATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINY KLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNLNDCHKLI DFFKDSISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKKE VDKLVEEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGA ELFMRRASLKKEELVVHPANSPIANKNPDNPKKTTTLSYDVYKDKRFSEDQYE LHIPIAINKCPKNIFKINTEVRVLLKHDDNPYVIGIDRGERNLLYIVVVDGKG NIVEQYSLNEIINNFNGIRIKTDYHSLLDKKEKERFEARQNWTSIENIKELKA GYISQVVHKICELVEKYDAVIALEDLNSGFKNSRVKVEKQVYQKFEKMLIDKL NYMVDKKSNPCATGGALKGYQITNKFESFKSMSTQNGFIFYIPAWLTSKIDPS TGFVNLLKTKYTSIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYI KKWKLYSYGNRIRIFRNPKKNNVFDWEEVCLTSAYKELFNKYGINYQQGDIRA LLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDFLISPVKNSDGIFYDSRN YEAQENAILPKNADANGAYNIARKVLWAIGQFKKAEDEKLDKVKIAISNKEWL EYAQTSVKH	SEQ ID NO: 1361461
BhCas12b *B. hisashii* 1108 AA 130,4 кДа	MATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHHEQD PKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSS VEKKGEANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEK KKWEEDKKDPLAKILGKLAEYGLIPLFIPYTDSNEPIVKEIKWMEKSRNQSV RRLDKDMFIQALERFLSWESWNLKVKEEYEKVEKEYKTLEERIKEDIQALKAL EQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDENEPSEKYLEV FKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKKKDA KQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDR LIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPL KGTLGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIH RDDFPKVVNFKPKELTEWIKDSKGKKLKSGIESLEIGLRVMSIDLGQRQAAAA SIFEVVDQKPDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREVLRKAR EDNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDEL IQIRELMYKPYKDWVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLK NIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLKKM ANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYNPYEERSRFENSKLM KWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQ DNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTHADIN AAQNLQKRFWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYFI LKDGVYEWVNAGKLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYR DPSGNVFPSDKWMAAGVFFGKLERILISKLTNQYSISTIEDDSSKQSM	SEQ ID NO: 1361462

(vii) Эквиваленты Cas9

[0366] В некоторых вариантах осуществления описанные здесь праймерные редакторы могут включать любой эквивалент Cas9. В соответствии с используемым здесь значением термин "эквивалент Cas9" является широким термином, который охватывает любой белок napDNAbp, который служит той же функции, что и Cas9 в настоящих праймерных редакторах, несмотря на то, что его первичная аминокислотная

последовательность и/или его трехмерная структура может отличаться и/или не являться родственной с эволюционной точки зрения. Таким образом, в то время как эквиваленты Cas9 включают любой ортолог, гомолог, мутант или вариант Cas9, описанный или охватываемый здесь, который является эволюционно родственным, эквиваленты Cas9 также охватывают белки, которые могли эволюционировать за счет конвергентных эволюционных процессов, имеющие такую же или аналогичную функцию, что и Cas9, но которые не обязательно имеют какое-либо сходство в плане аминокислотной последовательности и/или трехмерной структуры. Описанные здесь праймерные редакторы охватывают любой эквивалент Cas9, обеспечивающий предоставление такой же или аналогичной функции, что и Cas9, несмотря на то, что эквивалент Cas9 может быть основан на белке, который образовался за счет конвергентной эволюции. Например, если Cas9 обозначает фермент типа II системы CRISPR-Cas, эквивалент Cas9 может обозначать фермент типа V или типа VI системы CRISPR-Cas.

[0367] Например Cas12e (CasX) представляет собой эквивалент Cas9, который, как сообщается, имеет такую же функцию, что и Cas9, но который эволюционировал за счет конвергентной эволюции. Таким образом, предполагается, что белок Cas12e (CasX), описанный в Liu et al., "CasX enzymes comprises a distinct family of RNA-guided genome editors," *Nature*, 2019, Vol.566: 218-223, будет использоваться с описанными здесь праймерными редакторами. Кроме того, в рамках объема настоящего изобретения может рассматриваться любой вариант или модификация Cas12e (CasX).

[0368] Cas9 представляет собой бактериальный фермент, который эволюционировал у широкого ряда видов. Однако рассматриваемые здесь эквиваленты Cas9 также могут быть получены из архей, которые составляют домен и царство одноклеточных прокариотических микробов, отличных от бактерий.

[0369] В некоторых вариантах осуществления эквиваленты Cas9 могут обозначать Cas12e (CasX) или Cas12d (CasY), которые были описаны, например, в Burstein et al., “New CRISPR–Cas systems from uncultivated microbes.” *Cell Res.* 2017 Feb 21. doi: 10.1038/cr.2017.21, полное содержание которого приведено здесь в качестве ссылки. С помощью метагеномики с геномным корректированием был идентифицирован ряд систем CRISPR-Cas, включая первый обнаруженный Cas9 в домене жизни архей. Данный отличающийся белок Cas9 был обнаружен у малоизученных наноархей в виде части

активной системы CRISPR-Cas. В бактериях были открыты две ранее неизвестные системы, CRISPR-Cas12e и CRISPR-Cas12d, которые являются одними из наиболее компактных систем, открытых на сегодняшний день. В некоторых вариантах осуществления Cas9 обозначает Cas12e или вариант Cas12e. В некоторых вариантах осуществления Cas9 обозначает Cas12d или вариант Cas12d. Следует понимать, что в качестве программируемого нуклеиновой кислотой ДНК-связывающего белка могут использоваться другие РНК-направляемые связывающие ДНК белки, при этом они находятся в рамках объема данного изобретения. Также см. Liu et al., "CasX enzymes comprises a distinct family of RNA-guided genome editors," *Nature*, 2019, Vol.566: 218-223. Охватываются любые из данных эквивалентов Cas9.

[0370] В некоторых вариантах осуществления эквивалент Cas9 включает аминокислотную последовательность, которая, по меньшей мере, на 85%, по меньшей мере, на 90%, по меньшей мере, на 91%, по меньшей мере, на 92%, по меньшей мере, на 93%, по меньшей мере, на 94%, по меньшей мере, на 95%, по меньшей мере, на 96%, по меньшей мере, на 97%, по меньшей мере, на 98%, по меньшей мере, на 99%, или, по меньшей мере, на 99,5% идентична встречающемуся в природе белку Cas12e (CasX) или Cas12d (CasY). В некоторых вариантах осуществления napDNAbp представляет собой встречающийся в природе белок Cas12e (CasX) или Cas12d (CasY). В некоторых вариантах осуществления napDNAbp включает аминокислотную последовательность, которая, по меньшей мере, на 85%, по меньшей мере, на 90%, по меньшей мере, на 91%, по меньшей мере, на 92%, по меньшей мере, на 93%, по меньшей мере, на 94%, по меньшей мере, на 95%, по меньшей мере, на 96%, по меньшей мере, на 97%, по меньшей мере, на 98%, по меньшей мере, на 99%, или, по меньшей мере, на 99,5% идентична фрагменту Cas дикого типа или любому указанному здесь фрагменту Cas.

[0371] В различных вариантах осуществления программируемые нуклеиновой кислотой ДНК-связывающие белки включают, без ограничения, Cas9 (например dCas9 и nCas9), Cas12e (CasX), Cas12d (CasY), Cas12a (Cpf1), Cas12b1 (C2c1), Cas13a (C2c2), Cas12c (C2c3), Argonaute, и Cas12b1. Один пример программируемого нуклеиновой кислотой ДНК-связывающего белка, который имеет другую специфичность в отношении PAM, чем Cas9, представляет собой короткие регулярно расположенные группами палиндромные повторы из Prevotella и Francisella 1 (то есть Cas12a (Cpf1)). Аналогично Cas9, Cas12a

(Cpf1) также представляет собой эффектор CRISPR Класса 2, но он является членом подгруппы ферментов типа V, а не подгруппы типа II. Было показано, что Cas12a (Cpf1) опосредует устойчивую интерференцию ДНК со свойствами, отличающимися от Cas9. Cas12a (Cpf1) представляет собой направляемую одной РНК эндонуклеазу без tracrРНК и использует смежный с протоспейсером мотив с высоким содержанием T (TTN, TTTN, или YTN). Более того, Cpf1 расщепляет ДНК за счет зигзагообразного двуцепочечного разрыва ДНК. Показано, что из 16 белков семейства Cpf1 два фермента из *Acidaminococcus* и *Lachnospiraceae* имеют эффективную связанную с редактированием генома активность в клетках человека. Белки Cpf1 известны в данной области и были ранее описаны, например, в Yamano *et al.*, "Crystal structure of Cpf1 in complex with guide RNA and target DNA." *Cell (165) 2016, p. 949-962*; полное содержание которого приведено здесь в качестве ссылки.

[0372] В других вариантах осуществления белок Cas может включать любой ассоциированный с CRISPR белок, включая, но без ограничения, Cas12a, Cas12b1, Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (также известный как Csn1 и Csx12), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, их гомологи или модифицированные версии, и, предпочтительно, включающие никазную мутацию (например мутацию, соответствующую мутации D10A полипептида Cas9 дикого типа согласно SEQ ID NO: 1361421).

[0373] В различных других вариантах осуществления napDNAbp может представлять собой любой из следующих белков: Cas9, Cas12a (Cpf1), Cas12e (CasX), Cas12d (CasY), Cas12b1 (C2c1), Cas13a (C2c2), Cas12c (C2c3), GeoCas9, CjCas9, Cas12g, Cas12h, Cas12i, Cas13b, Cas13c, Cas13d, Cas14, Csn2, xCas9, SpCas9-NG, циркулярно пермутированный Cas9, или домен Argonaute (Ago), или их вариант.

[0374] Примеры последовательностей белка-эквивалента Cas9 могут включать следующие:

Описание	Последовательность
Cas12a (ранее известный как Cpf1) *Acidaminococcus*	MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTY ADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAIN KRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAE DISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEVFSFP FYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPL FKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFI SHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAA

Описание	Последовательность
sp. (штамм *BV3L6)* UniProtKB U2UMQ6	GKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDE SNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTLASGWDVNKE KNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFPDAAKMIPKCST QLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKTGDQKGYRE ALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIM DAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKS RMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEV SHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLI YITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVI HEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVG GVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFL EGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKR IVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIR SVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHL KESKDLKLQNGISNQDWLAYIQELRN (SEQ ID NO: 1361463)
Cas12a (ранее известный как Cpf1) *Lachnospiraceae bacterium GAM79* Исх. посл. WP_119623382.1	MNYKTGLEDFIGKESLSKTLRNALIPTESTKIHMEEMGVIRDDELRAEKQQELKEIMDDYY RTFIEEKLGQIQGIQWNSLFQKMEETMEDISVRKDLDKIQNEKRKEICCYFTSDKRFKDLF NAKLITDILPNFIKDNKEYTEEEKAEKEQTRVLFQRFATAFTNYFNQRRNNFSEDNISTAI SFRIVNENSEIHLQNMRAFQRIEQQYPEEVCGMEEEYKDMLQEWQMKHIYSVDFYDRELTQ PGIEYYNGICGKINEHMNQFCQKNRINKNDFRMKKLHKQILCKKSSYYEIPFRFESDQEVY DALNEFIKTMKKKEIIRRCVHLGQECDDYDLGKIYISSNKYEQISNALYGSWDTIRKCIKE EYMDALPGKGEKKEEKAEAAAKKEEYRSIADIDKIISLYGSEMDRTISAKKCITEICDMAG QISIDPLVCNSDIKLLQNKEKTTEIKTILDSFLHVYQWGQTFIVSDIIEKDSYFYSELEDV LEDFEGITTLYNHVRSYVTQKPYSTVKFKLHFGSPTLANGWSQSKEYDNNAILLMRDQKFY LGIFNVRNKPDKQIIKGHEKEEKGDYKKMIYNLLPGPSKMLPKVFITSRSGQETYKPSKHI LDGYNEKRHIKSSPKFDLGYCWDLIDYYKECIHKHPDWKNYDFHFSDTKDYEDISGFYREV EMQGYQIKWTYISADEIQKLDEKGQIFLFQIYNKDFSVHSTGKDNLHTMYLKNLFSEENLK DIVLKLNGEAELFFRKASIKTPIVHKKGSVLVNRSYTQTVGNKEIRVSIPEEYYTEIYNYL NHIGKGKLSSEAQRYLDEGKIKSFTATKDIVKNYRYCCDHYFLHLPITINFKAKSDVAVNE RTLAYIAKKEDIHIIGIDRGERNLLYISVVDVHGNIREQRSFNIVNGYDYQQKLKDREKSR DAARKNWEEIEKIKELKEGYLSMVIHYIAQLVVKYNAVVAMEDLNYGFKTGRFKVERQVYQ KFETMLIEKLHYLVFKDREVCEEGGVLRGYQLTYIPESLKKVGKQCGFIFYVPAGYTSKID PTTGFVNLFSFKNLTNRESRQDFVGKFDEIRYDRDKKMFEFSFDYNNYIKKGTILASTKWK VYTNGTRLKRIVVNGKYTSQSMEVELTDAMEKMLQRAGIEYHDGKDLKGQIVEKGIEAEII DIFRLTVQMRNSRSESEDREYDRLISPVLNDKGEFFDTATADKTLPQDADANGAYCIALKG LYEVKQIKENWKENEQFPRNKLVQDNKTWFDFMQKKRYL(SEQ ID NO: 1361464)
Cas12a – ранее известный как Cpf1 *Prevotella copri* Исх. посл. WP_119227726.1	MAKNFEDFKRLYSLSKTLRFEAKPIGATLDNIVKSGLLDEDEHRAASYVKVKKLIDEYHKV FIDRVLDDGCLPLENKGNNNSLAEYYESYVSRAQDEDAKKKFKEIQQNLRSVIAKKLTEDK AYANLFGNKLIESYKDKEDKKKIIDSDLIQFINTAESTQLDSMSQDEAKELVKEFWGFVTY FYGFFDNRKNMYTAEEKSTGIAYRLVNENLPKFIDNIEAFNRAITRPEIQENMGVLYSDFS EYLNVESIQEMFQLDYYNMLLTQKQIDVYNAIIGGKTDDEHDVKIKGINEYINLYNQQHKD DKLPKLKALFKQILSDRNAISWLPEEFNSDQEVLNAIKDCYERLAENVLGDKVLKSLLGSL ADYSLDGIFIRNDLQLTDISQKMFGNWGVIQNAIMQNIKRVAPARKHKESEEDYEKRIAGI FKKADSFSISYINDCLNEADPNNAYFVENYFATFGAVNTPTMQRENLFALVQNAYTEVAAL LHSDYPTVKHLAQDKANVSKIKALLDAIKSLQHFVKPLLGKGDESDKDERFYGELASLWAE LDTVTPLYNMIRNYMTRKPYSQKKIKLNFENPQLLGGWDANKEKDYATIILRRNGLYYLAI MDKDSRKLLGKAMPSDGECYEKMVYKFFKDVTTMIPKCSTQLKDVQAYFKVNTDDYVLNSK AFNKPLTITKEVFDLNNVLYGKYKKFQKGYLTATGDNVGYTHAVNVWIKFCMDFLNSYDST CIYDFSSLKPESYLSLDAFYQDANLLLYKLSFARASVSYINQLVEEGKMYLFQIYNKDFSE YSKGTPNMHTLYWKALFDERNLADVVYKLNGQAEMFYRKKSIENTHPTHPANHPILNKNKD NKKKESLFDYDLIKDRRYTVDKFMFHVPITMNFKSVGSENINQDVKAYLRHADDMHIIGID RGERHLLYLVVIDLQGNIKEQYSLNEIVNEYNGNTYHTNYHDLLDVREEERLKARQSWQTI ENIKELKEGYLSQVIHKITQLMVRYHAIVVLEDLSKGFMRSRQKVEKQVYQKFEKMLIDKL NYLVDKKTDVSTPGGLLNAYQLTCKSDSSQKLGKQSGFLFYIPAWNTSKIDPVTGFVNLLD THSLNSKEKIKAFFSKFDAIRYNKDKKWFEFNLDYDKFGKKAEDTRTKWTLCTRGMRIDTF RNKEKNSQWDNQEVDLTTEMKSLLEHYYIDIHGNLKDAISAQTDKAFFTGLLHILKLTLQM RNSITGTETDYLVSPVADENGIFYDSRSCGNQLPENADANGAYNIARKGLMLIEQIKNAED LNNVKFDISNKAWLNFAQQKPYKNG(SEQ ID NO: 1361465)
Cas12a – ранее известный как Cpf1 *Eubacterium rectale* Исх. посл. WP_119223642.1	MFSAKLISDILPEFVIHNNNYSASEKEEKTQVIKLFSRFATSFKDYFKNRANCFSANDISS SSCHRIVNDNAEIFFSNALVYRRIVKNLSNDDINKISGDMKDSLKEMSLEEIYSYEKYGEF ITQEGISFYNDICGKVNLFMNLYCQKNKENKNLYKLRKLHKQILCIADTSYEVPYKFESDE EVYQSVNGFLDNISSKHIVERLRKIGENYNGYNLDKIYIVSKFYESVSQKTYRDWETINTA LEIHYNNILPGNGKSKADKVKKAVKNDLQKSITEINELVSNYKLCPDDNIKAETYIHEISH ILNNFEAQELKYNPEIHLVESELKASELKNVLDVIMNAFHWCSVFMTEELVDKDNNFYAEL EEIYDEIYPVISLYNLVRNYVTQKPYSTKKIKLNFGIPTLADGWSKSKEYSNNAIILMRDN LYYLGIFNAKNKPDKKIIEGNTSENKGDYKKMIYNLLPGPNKMIPKVFLSSKTGVETYKPS AYILEGYKQNKHLKSSKDFDITFCHDLIDYFKNCIAIHPEWKNFGFDFSDTSTYEDISGFY REVELQGYKIDWTYISEKDIDLLQEKGQLYLFQIYNKDFSKKSSGNDNLHTMYLKNLFSEE NLKDIVLKLNGEAEIFFRKSSIKNPIIHKKGSILVNRTYEAEEKDQFGNIQIVRKTIPENI YQELYKYFNDKSDKELSDEAAKLKNVVGHHEAATNIVKDYRYTYDKYFLHMPITINFKANK TSFINDRILQYIAKEKDLHVIGIDRGERNLIYVSVIDTCGNIVEQKSFNIVNGYDYQIKLK QQEGARQIARKEWKEIGKIKEIKEGYLSLVIHEISKMVIKYNAIIAMEDLSYGFKKGRFKV ERQVYQKFETMLINKLNYLVFKDISITENGGLLKGYQLTYIPDKLKNVGHQCGCIFYVPAA YTSKIDPTTGFVNIFKFKDLTVDAKREFIKKFDSIRYDSDKNLFCFTFDYNNFITQNTVMS KSSWSVYTYGVRIKRRFVNGRFSNESDTIDITKDMEKTLEMTDINWRDGHDLRQDIIDYEI VQHIFEIFKLTVQMRNSLSELEDRDYDRLISPVLNENNIFYDSAKAGDALPKDADANGAYC IALKGLYEIKQITENWKEDGKFSRDKLKISNKDWFDFIQNKRYL(SEQ ID NO: 1361466)
Cas12a – ранее известный как Cpf1 *Clostridium sp. AF34-10BH* Исх. посл.	MNYKTGLEDFIGKESLSKTLRNALIPTESTKIHMEEMGVIRDDELRAEKQQELKEIMDDYY RAFIEEKLGQIQGIQWNSLFQKMEETMEDISVRKDLDKIQNEKRKEICCYFTSDKRFKDLF NAKLITDILPNFIKDNKEYTEEEKAEKEQTRVLFQRFATAFTNYFNQRRNNFSEDNISTAI SFRIVNENSEIHLQNMRAFQRIEQQYPEEVCGMEEEYKDMLQEWQMKHIYLVDFYDRVLTQ PGIEYYNGICGKINEHMNQFCQKNRINKNDFRMKKLHKQILCKKSSYYEIPFRFESDQEVY DALNEFIKTMKEKEIICRCVHLGQKCDDYDLGKIYISSNKYEQISNALYGSWDTIRKCIKE EYMDALPGKGEKKEEKAEAAAKKEEYRSIADIDKIISLYGSEMDRTISAKKCITEICDMAG QISTDPLVCNSDIKLLQNKEKTTEIKTILDSFLHVYQWGQTFIVSDIIEKDSYFYSELEDV LEDFEGITTLYNHVRSYVTQKPYSTVKFKLHFGSPTLANGWSQSKEYDNNAILLMRDQKFY LGIFNVRNKPDKQIIKGHEKEEKGDYKKMIYNLLPGPSKMLPKVFITSRSGQETYKPSKHI LDGYNEKRHIKSSPKFDLGYCWDLIDYYKECIHKHPDWKNYDFHFSDTKDYEDISGFYREV

Описание	Последовательность
WP_118538418.1	EMQGYQIKWTYISADEIQKLDEKGQIFLFQIYNKDFSVHSTGKDNLHTMYLKNLFSEENLK DIVLKLNGEAELFFRKASIKTPVVHKKGSVLVNRSYTQTVGDKEIRVSIPEEYYTEIYNYL NHIGRGKLSTEAQRYLEERKIKSFTATKDIVKNYRYCCDHYFLHLPITINFKAKSDIAVNE RTLAYIAKKEDIHIIGIDRGERNLLYISVVDVHGNIREQRSFNIVNGYDYQQKLKDREKSR DAARKNWEEIEKIKELKEGYLSMVIHYIAQLVVKYNAVVAMEDLNYGFKTGRFKVERQVYQ KFETMLIEKLHYLVFKDREVCEEGGVLRGYQLTYIPESLKKVGKQCGFIFYVPAGYTSKID PTTGFVNLFSFKNLTNRESRQDFVGKFDEIRYDRDKKMFEFSFDYNNYIKKGTMLASTKWK VYTNGTRLKRIVVNGKYTSQSMEVELTDAMEKMLQRAGIEYHDGKDLKGQIVEKGIEAEII DIFRLTVQMRNSRSESEDREYDRLISPVLNDKGEFFDTATADKTLPQDADANGAYCIALKG LYEVKQIKENWKENEQFPRNKLVQDNKTWFDFMQKKRYL(SEQ ID NO: 1361467)
Cas12b *Bacillus hisashii* *Исх. посл.* WP_095142515.1	MATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHHEQDPKNPKKVS KAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGEANQLSNKFLY PLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILGKLAEYGL IPLFIPYTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKV EKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLK MDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDK KKKDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIY PTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQF DRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWI KDSKGKKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPDIEGKLFFPIKGTELYA VHRASFNIKLPGETLVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTK WISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFLKQLHKRLEVEIGKEVKHWRKSLSDG RKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLK KMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYNPYEERSRFENSKLMKWSRRE IPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREGRL TLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCK AYQVDGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWVNAGKLKIKKGSSKQSSSELVD SDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFGKLERILISKLTNQYSIST IEDDSSKQSM(SEQ ID NO: 1361468)
Cas12b *Thermomonas hydrothermalis* *Исх. посл.* WP_072754838	MSEKTTQRAYTLRLNRASGECAVCQNNSCDCWHDALWATHKAVNRGAKAFGDWLLTLRGGL CHTLVEMEVPAKGNNPPQRPTDQERRDRRVLLALSWLSVEDEHGAPKEFIVATGRDSADDR AKKVEEKLREILEKRDFQEHEIDAWLQDCGPSLKAHIREDAVWVNRRALFDAAVERIKTLT WEEAWDFLEPFFGTQYFAGIGDGKDKDDAEGPARQGEKAKDLVQKAGQWLSARFGIGTGAD FMSMAEAYEKIAKWASQAQNGDNGKATIEKLACALRPSEPPTLDTVLKCISGPGHKSATRE YLKTLDKKSTVTQEDLNQLRKLADEDARNCRKKVGKKGKKPWADEVLKDVENSCELTYLQD NSPARHREFSVMLDHAARRVSMAHSWIKKAEQRRRQFESDAQKLKNLQERAPSAVEWLDRF CESRSMTTGANTGSGYRIRKRAIEGWSYVVQAWAEASCDTEDKRIAAARKVQADPEIEKFG DIQLFEALAADEAICVWRDQEGTQNPSILIDYVTGKTAEHNQKRFKVPAYRHPDELRHPVF CDFGNSRWSIQFAIHKEIRDRDKGAKQDTRQLQNRHGLKMRLWNGRSMTDVNLHWSSKRLT ADLALDQNPNPNPTEVTRADRLGRAASSAFDHVKIKNVFNEKEWNGRLQAPRAELDRIAKL EEQGKTEQAEKLRKRLRWYVSFSPCLSPSGPFIVYAGQHNIQPKRSGQYAPHAQANKGRAR LAQLILSRLPDLRILSVDLGHRFAAACAVWETLSSDAFRREIQGLNVLAGGSGEGDLFLHV EMTGDDGKRRTVVYRRIGPDQLLDNTPHPAPWARLDRQFLIKLQGEDEGVREASNEELWTV HKLEVEVGRTVPLIDRMVRSGFGKTEKQKERLKKLRELGWISAMPNEPSAETDEKEGEIRS ISRSVDELMSSALGTLRLALKRHGNRARIAFAMTADYKPMPGGQKYYFHEAKEASKNDDET KRRDNQIEFLQDALSLWHDLFSSPDWEDNEAKKLWQNHIATLPNYQTPEEISAELKRVERN KKRKENRDKLRTAAKALAENDQLRQHLHDTWKERWESDDQQWKERLRSLKDWIFPRGKAED NPSIRHVGGLSITRINTISGLYQILKAFKMRPEPDDLRKNIPQKGDDELENFNRRLLEARD RLREQRVKQLASRIIEAALGVGRIKIPKNGKLPKRPRTTVDTPCHAVVIESLKTYRPDDLR TRRENRQLMQWSSAKVRKYLKEGCELYGLHFLEVPANYTSRQCSRTGLPGIRCDDVPTGDF LKAPWWRRAINTAREKNGGDAKDRFLVDLYDHLNNLQSKGEALPATVRVPRQGGNLFIAGA QLDDTNKERRAIQADLNAAANIGLRALLDPDWRGRWWYVPCKDGTSEPALDRIEGSTAFND VRSLPTGDNSSRRAPREIENLWRDPSGDSLESGTWSPTRAYWDTVQSRVIELLRRHAGLPT S(SEQ ID NO: 1361469)
Cas12b *Laceyella sacchari* WP_132221894.1	MSIRSFKLKLKTKSGVNAEQLRRGLWRTHQLINDGIAYYMNWLVLLRQEDLFIRNKETNEI EKRSKEEIQAVLLERVHKQQQRNQWSGEVDEQTLLQALRQLYEEIVPSVIGKSGNASLKAR FFLGPLVDPNNKTTKDVSKSGPTPKWKKMKDAGDPNWVQEYEKYMAERQTLVRLEEMGLIP LFPMYTDEVGDIHWLPQASGYTRTWDRDMFQQAIERLLSWESWNRRVRERRAQFEKKTHDF ASRFSESDVQWMNKLREYEAQQEKSLEENAFAPNEPYALTKKALRGWERVYHSWMRLDSAA SEEAYWQEVATCQTAMRGEFGDPAIYQFLAQKENHDIWRGYPERVIDFAELNHLQRELRRA KEDATFTLPDSVDHPLWVRYEAPGGTNIHGYDLVQDTKRNLTLILDKFILPDENGSWHEVK KVPFSLAKSKQFHRQVWLQEEQKQKKREVVFYDYSTNLPHLGTLAGAKLQWDRNFLNKRTQ QQIEETGEIGKVFFNISVDVRPAVEVKNGRLQNGLGKALTVLTHPDGTKIVTGWKAEQLEK WVGESGRVSSLGLDSLSEGLRVMSIDLGQRTSATVSVFEITKEAPDNPYKFFYQLEGTEMF AVHQRSFLLALPGENPPQKIKQMREIRWKERNRIKQQVDQLSAILRLHKKVNEDERIQAID KLLQKVASWQLNEEIATAWNQALSQLYSKAKENDLQWNQAIKNAHHQLEPVVGKQISLWRK DLSTGRQGIAGLSLWSIEELEATKKLLTRWSKRSREPGVVKRIERFETFAKQIQHHINQVK ENRLKQLANLIVMTALGYKYDQEQKKWIEVYPACQVVLFENLRSYRFSFERSRRENKKLME WSHRSIPKLVQMQGELFGLQVADVYAAYSSRYHGRTGAPGIRCHALTEADLRNETNIIHEL IEAGFIKEEHRPYLQQGDLVPWSGGELFATLQKPYDNPRILTLHADINAAQNIQKRFWHPS MWFRVNCESVMEGEIVTYVPKNKTVHKKQGKTFRFVKVEGSDVYEWAKWSKNRNKNTFSSI TERKPPSSMILFRDPSGTFFKEQEWVEQKTFWGKVQSMIQAYMKKTIVQRMEE(SEQ ID NO: 1361470)
Cas12b *Dsulfonatronum thiodismutans* *WP_031386437*	MVLGRKDDTAELRRALWTTHEHVNLAVAEVERVLLRCRGRSYWTLDRRGDPVHVPESQVAE DALAMAREAQRRNGWPVVGEDEEILLALRYLYEQIVPSCLLDDLGKPLKGDAQKIGTNYAG PLFDSDTCRRDEGKDVACCGPFHEVAGKYLGALPEWATPISKQEFDGKDASHLRFKATGGD DAFFRVSIEKANAWYEDPANQDALKNKAYNKDDWKKEKDKGISSWAVKYIQKQLQLGQDPR TEVRRKLWLELGLLPLFIPVFDKTMVGNLWNRLAVRLALAHLLSWESWNHRAVQDQALARA KRDELAALFLGMEDGFAGLREYELRRNESIKQHAFEPVDRPYVVSGRALRSWTRVREEWLR HGDTQESRKNICNRLQDRLRGKFGDPDVFHWLAEDGQEALWKERDCVTSFSLLNDADGLLE KRKGYALMTFADARLHPRWAMYEAPGGSNLRTYQIRKTENGLWADVVLLSPRNESAAVEEK TFNVRLAPSGQLSNVSFDQIQKGSKMVGRCRYQSANQQFEGLLGGAEILFDRKRIANEQHG ATDLASKPGHVWFKLTLDVRPQAPQGWLDGKGRPALPPEAKHFKTALSNKSKFADQVRPGL RVLSVDLGVRSFAACSVFELVRGGPDQGTYFPAADGRTVDDPEKLWAKHERSFKITLPGEN PSRKEEIARRAAMEELRSLNGDIRRLKAILRLSVLQEDDPRTEHLRLFMEAIVDDPAKSAL NAELFKGFGDDRFRSTPDLWKQHCHFFHDKAEKVVAERFSRWRTETRPKSSSWQDWRERRG YAGGKSYWAVTYLEAVRGLILRWNMRGRTYGEVNRQDKKQFGTVASALLHHINQLKEDRIK TGADMIIQAARGFVPRKNGAGWVQVHEPCRLILFEDLARYRFRTDRSRRENSRLMRWSHRE IVNEVGMQGELYGLHVDTTEAGFSSRYLASSGAPGVRCRHLVEEDFHDGLPGMHLVGELDW

Описание	Последовательность
	LLPKDKDRTANEARRLLGGMVRPGMLVPWDGGELFATLNAASQLHVIHADINAAQNLQRRF WGRCGEAIRIVCNQLSVDGSTRYEMAKAPKARLLGALQQLKNGDAPFHLTSIPNSQKPENS YVMTPTNAGKKYRAGPGEKSSGEEDELALDIVEQAEELAQGRKTFFRDPSGVFFAPDRWLP SEIYWSRIRRRIWQVTLERNSSGRQERAEMDEMPY(SEQ ID NO: 1361471)

[0375] Описанные здесь праймерные редакторы также могут включать варианты Cas12a/Cpf1 (dCpf1), которые могут применяться в качестве домена программируемого направляющей нуклеотидной последовательностью ДНК-связывающего белка. Белок Cas12a/Cpf1 имеет RuvC-подобный эндонуклеазный домен, который аналогичен домену RuvC Cas9, но не имеет эндонуклеазного домена HNH, при этом N-конец Cpf1 не имеет доли распознавания альфа-цепи Cas9. В Zetsche *et al.*, *Cell*, 163, 759 - 771, 2015 (который включен здесь в качестве ссылки) было показано, что RuvC-подобный домен Cpf1 отвечает за расщепление обеих цепей ДНК, а деактивация RuvC-подобного домена обеспечивает деактивацию нуклеазной активности Cpf1. Описанные здесь праймерные редакторы также могут включать варианты Cas12a (Cpf1) (dCpf1), которые могут применяться в качестве домена программируемого направляющей нуклеотидной последовательностью ДНК-связывающего белка. Белок Cas12a Cpf1 имеет RuvC-подобный эндонуклеазный домен, который аналогичен домену RuvC Cas9, но не имеет эндонуклеазного домена HNH, при этом N-конец Cas12a (Cpf1) не имеет доли распознавания альфа-цепи Cas9. В Zetsche *et al.*, *Cell*, 163, 759 - 771, 2015 (который включен здесь в качестве ссылки) было показано, что RuvC-подобный домен Cas12a (Cpf1) отвечает за расщепление обеих цепей ДНК, а деактивация RuvC-подобного домена обеспечивает деактивацию нуклеазной активности Cas12a (Cpf1).

[0376] В некоторых вариантах осуществления napDNAbp представляет собой одиночный эффектор микробной системы CRISPR-Cas. Одиночные эффекторы микробных систем CRISPR-Cas включают, без ограничения, Cas9, Cas12a (Cpf1), Cas12b1 (C2c1), Cas13a (C2c2), и Cas12c (C2c3). Как правило, микробные системы CRISPR-Cas разделяются на системы Класса 1 и Класса 2. Системы Класса 1 имеют мультисубъединичные эффекторные комплексы, в то время как системы Класса 2 имеют одиночный белковый эффектор. Например, Cas9 и Cas12a (Cpf1) представляют собой эффекторы Класса 2. Помимо Cas9 и Cas12a (Cpf1), в Shmakov et al., “Discovery and Functional Characterization of Diverse Class 2 CRISPR Cas Systems”, *Mol. Cell*, 2015 Nov 5; 60(3): 385–397, полное содержание которого приведено здесь в качестве ссылки, были описаны три различных системы CRISPR-Cas Класса 2 (Cas12b1, Cas13a, и Cas12c).

[0377] Эффекторы двух систем, Cas12b1 и Cas12c, содержат RuvC-подобные эндонуклеазные домены, являющиеся родственными Cas12a. Третья система, Cas13a, содержит эффектор с двумя предсказанными доменами HEPN РНКазы. Продуцирование зрелой РНК CRISPR не зависит от tracrРНК, в отличие от продуцирования РНК CRISPR за счет Cas12b1. Cas12b1 зависит как от РНК CRISPR, так и tracrРНК для расщепления ДНК. Было показано, что бактериальный Cas13a обладает уникальной РНКазной активностью в плане созревания РНК CRISPR, отличной от его РНК-активируемой связанной с расщеплением одноцепочечной РНК активности. Данные РНКазные функции отличаются друг от друга и от связанного с процессированием РНК CRISPR поведения Cas12a. См., например, East-Seletsky, et al., "Two distinct RNase activities of CRISPR-Cas13a enable guide-RNA processing and RNA detection", *Nature,* 2016 Oct 13;538(7624):270-273, полное содержание которого приведено здесь в качестве ссылки. *In vitro* биохимический анализ Cas13a в *Leptotrichia shahii* показал, что Cas13a направляется одиночной РНК CRISPR и может быть запрограммирован для расщепления мишеней в виде оцРНК, несущих комплементарные протоспейсеры. Каталитические остатки в двух консервативных доменах HEPN опосредуют расщепление. Мутации в каталитических остатках продуцируют каталитически деактивированные РНК-связывающие белки. См., например, Abudayyeh et al., "C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector", *Science,* 2016 Aug 5; 353(6299), полное содержание которого приведено здесь в качестве ссылки.

[0378] Сообщалось о кристаллической структуре Cas12b1 (AacC2c1) *Alicyclobaccillus acidoterrastris* в комплексе с химерной направляющей РНК в виде одиночной молекулы (sgРНК). См., например, Liu et al., “C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism”, *Mol. Cell,* 2017 Jan 19;65(2):310-322, полное содержание которого приведено здесь в качестве ссылки. Также сообщалось о кристаллической структуре в C2c1 *Alicyclobacillus acidoterrestris*, связанном с ДНК-мишенями в виде третичных комплексов. См., например,Yang et al., "PAM-dependent Target DNA Recognition and Cleavage by C2C1 CRISPR-Cas endonuclease", *Cell,* 2016 Dec 15;167(7):1814-1828, полное содержание которого приведено здесь в качестве ссылки. Каталитически компетентные конформации AacC2c1 как с являющимися, так и не являющимися мишенью цепями ДНК были обнаружены в виде независимо

расположенных в одном каталитическом кармане RuvC с C2c1-опосредованным расщеплением, приводящим к зигзагообразному семинуклеотидному разрыву ДНК-мишени. Структурные сравнения третичных комплексов C2c1 и ранее идентифицированных Cas9 и Cpf1 демонстрируют разнообразие механизмов, используемых в рамках систем CRISPR-Cas9.

[0379] В некоторых вариантах осуществления napDNAbp может представлять собой белок C2c1, 2c2, или C2c3. В некоторых вариантах осуществления napDNAbp представляет собой белок C2c1. В некоторых вариантах осуществления napDNAbp представляет собой белок Cas13a. В некоторых вариантах осуществления napDNAbp представляет собой белок Cas12c. В некоторых вариантах осуществления napDNAbp включает аминокислотную последовательность, которая, по меньшей мере, на 85%, по меньшей мере, на 90%, по меньшей мере, на 91%, по меньшей мере, на 92%, по меньшей мере, на 93%, по меньшей мере, на 94%, по меньшей мере, на 95%, по меньшей мере, на 96%, по меньшей мере, на 97%, по меньшей мере, на 98%, по меньшей мере, на 99%, или, по меньшей мере, на 99,5% идентична встречающемуся в природе белку Cas12b1 (C2c1), Cas13a (C2c2), или Cas12c (C2c3). В некоторых вариантах осуществления napDNAbp представляет собой встречающийся в природе белок Cas12b1 (C2c1), Cas13a (C2c2), или Cas12c (C2c3).

[0380]

(viii) Циркулярные пермутанты Cas9

[0381] В различных вариантах осуществления описанные здесь праймерные редакторы могут включать циркулярный пермутант Cas9.

[0382] Термин "циркулярно пермутированный Cas9" или "циркулярный пермутант" Cas9 или "CP-Cas9" обозначает любой белок Cas9 или его вариант, который возникает или был сконструирован в виде циркулярного пермутантного варианта, что означает, что N-конец и C-конец белка Cas9 (например белка Cas9 дикого типа) были локально перенесены. Подобные циркулярно пермутированные белки Cas9 или их варианты сохраняют способность связывать ДНК при образовании комплексов с направляющей РНК (нРНК). См. Oakes et al., "Protein Engineering of Cas9 for enhanced function," *Methods Enzymol*, 2014, 546: 491–511 и Oakes et al., "CRISPR-Cas9 Circular Permutants as Programmable Scaffolds for Genome Modification," *Cell*, January 10, 2019, 176: 254-267, каждый из которых приведен

здесь в качестве ссылки. Настоящее изобретение охватывает любые ранее известные CP-Cas9 или предусматривает использование нового CP-Cas9 при условии, если полученный циркулярно пермутированный белок сохраняет способность связывать ДНК при образовании комплексов с направляющей РНК (нРНК).

[0383] Любой из описанных здесь белков Cas9, включая любой вариант, ортолог, или встречающийся в природе Cas9 или его эквивалент может быть подвергнут реконфигурированию в виде циркулярного пермутантного варианта.

[0384] В различных вариантах осуществления циркулярные пермутанты Cas9 могут иметь следующую структуру:

N-конец-[исходный C-конец] - [необязательный линкер] - [исходный N-конец]-C-конец.

[0385] В качестве примера, настоящее изобретение включает следующие циркулярные пермутанты стандартного Cas9 *S. pyogenes* (1368 аминокислот UniProtKB - Q99ZW2 (CAS9_STRP1) (нумерация основана на аминокислотном положении в SEQ ID NO: 1361421)):

N-конец-[1268-1368]-[необязательный линкер]-[1-1267]-C-конец;

N-конец-[1168-1368]-[необязательный линкер]-[1-1167]-C-конец;

N-конец-[1068-1368]-[необязательный линкер]-[1-1067]-C-конец;

N-конец-[968-1368]-[необязательный линкер]-[1-967]-C-конец;

N-конец-[868-1368]-[необязательный линкер]-[1-867]-C-конец;

N-конец-[768-1368]-[необязательный линкер]-[1-767]-C-конец;

N-конец-[668-1368]-[необязательный линкер]-[1-667]-C-конец;

N-конец-[568-1368]-[необязательный линкер]-[1-567]-C-конец;

N-конец-[468-1368]-[необязательный линкер]-[1-467]-C-конец;

N-конец-[368-1368]-[необязательный линкер]-[1-367]-C-конец;

N-конец-[268-1368]-[необязательный линкер]-[1-267]-C-конец;

N-конец-[168-1368]-[необязательный линкер]-[1-167]-C-конец;

N-конец-[68-1368]-[необязательный линкер]-[1-67]-C-конец; or

N-конец-[10-1368]-[необязательный линкер]-[1-9]-C-конец, или соответствующие циркулярные пермутанты других белков Cas9 (включая другие ортологи Cas9, варианты, и так далее).

[0386] В определенных вариантах осуществления циркулярный пермутант Cas9 имеет следующую структуру (на основе *S. pyogenes* Cas9 (1368 аминокислот UniProtKB - Q99ZW2 (CAS9_STRP1) (нумерация основана на аминокислотном положении в SEQ ID NO: 1361421):

N-конец-[102-1368]-[необязательный линкер]-[1-101]-C-конец;

N-конец-[1028-1368]-[необязательный линкер]-[1-1027]-C-конец;

N-конец-[1041-1368]-[необязательный линкер]-[1-1043]-C-конец;

N-конец-[1249-1368]-[необязательный линкер]-[1-1248]-C-конец; or

N-конец-[1300-1368]-[необязательный линкер]-[1-1299]-C-конец, или соответствующие циркулярные пермутанты других белков Cas9 (включая другие ортологи Cas9, варианты, и так далее).

[0387] В других вариантах осуществления циркулярный пермутант Cas9 имеет следующую структуру (на основе Cas9 *S. pyogenes* (1368 аминокислот UniProtKB - Q99ZW2 (CAS9_STRP1) (нумерация основана на аминокислотном положении в SEQ ID NO: 1361421):

N-конец-[103-1368]-[необязательный линкер]-[1-102]-C-конец;

N-конец-[1029-1368]-[необязательный линкер]-[1-1028]-C-конец;

N-конец-[1042-1368]-[необязательный линкер]-[1-1041]-C-конец;

N-конец-[1250-1368]-[необязательный линкер]-[1-1249]-C-конец; or

N-конец-[1301-1368]-[необязательный линкер]-[1-1300]-C-конец, или соответствующие циркулярные пермутанты других белков Cas9 (включая другие ортологи Cas9, варианты, и так далее).

[0388] В некоторых вариантах осуществления циркулярный пермутант может быть образован путем связывания C-концевого фрагмент Cas9 с N-концевым фрагментом Cas9, либо напрямую, или с помощью линкера, например аминокислотного линкера. В некоторых вариантах осуществления C-концевой фрагмент может соответствовать 95% или более C-концевым аминокислотам Cas9 (например приблизительно 1300-1368 аминокислотам) или 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 50%, 45%, 40%, 35%, 30%, 25%, 20%, 15%, 10%, или 5% или более C-концевым аминокислотам Cas9 (например любой одной из SEQ ID NO: 1361421-1361484, и 1361593-1361596). N-концевая часть может соответствовать 95% или более N-концевым аминокислотам Cas9 (например

приблизительно 1-1300 аминокислотам) или 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 50%, 45%, 40%, 35%, 30%, 25%, 20%, 15%, 10%, или 5% или более N-концевым аминокислотам Cas9 (например SEQ ID NO: 1361421).

[0389] В некоторых вариантах осуществления циркулярный пермутант может быть образован путем связывания С-концевого фрагмент Cas9 с N-концевым фрагментом Cas9, либо напрямую, или с помощью линкера, например аминокислотного линкера. В некоторых вариантах осуществления С-концевой фрагмент, который переносится на N-конец, включает или соответствует 30% или менее С-концевым аминокислотам на Cas9 (например аминокислотам 1012-1368 SEQ ID NO: 1361421). В некоторых вариантах осуществления С-концевой фрагмент, который переносится на N-конец, включает или соответствует 30%, 29%, 28%, 27%, 26%, 25%, 24%, 23%, 22%, 21%, 20%, 19%, 18%, 17%, 16%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, или 1% С-концевым аминокислотам Cas9 (например Cas9 согласно SEQ ID NO: 1361421). В некоторых вариантах осуществления С-концевой фрагмент, который переносится на N-конец, включает или соответствует 410 или менее С-концевым остаткам Cas9 (например Cas9 SEQ ID NO: 1361421). В некоторых вариантах осуществления С-концевая часть, которая переносится на N-конец, включает или соответствует 410, 400, 390, 380, 370, 360, 350, 340, 330, 320, 310, 300, 290, 280, 270, 260, 250, 240, 230, 220, 210, 200, 190, 180, 170, 160, 150, 140, 130, 120, 110, 100, 90, 80, 70, 60, 50, 40, 30, 20, или 10 С-концевым остаткам Cas9 (например Cas9 согласно SEQ ID NO: 1361421). В некоторых вариантах осуществления С-концевая часть, которая переносится на N-конец, включает или соответствует 357, 341, 328, 120, или 69 С-концевым остаткам Cas9 (например Cas9 SEQ ID NO: 1361421).

[0390] В других вариантах осуществления циркулярные пермутантные варианты Cas9 могут быть определены как топологическое перестроение первичной структуры Cas9 на основе следующего способа, который основан на Cas9 *S. pyogenes* согласно SEQ ID NO: 1361421: (a) выборе циркулярного пермутантного (CP) участка, соответствующего внутреннему аминокислотному остатку первичной структуры Cas9, который рассекает исходный белок на две половины: N-концевой участок и С-концевой участок; (b) модифицировании последовательности белка Cas9 (например с помощью способов генной инженерии) путем переноса исходного С-концевого участка (включающего аминокислоту

CP участка) для расположения перед исходным N-концевым участком, тем самым образуя новый N-конец белка Cas9, который теперь начинается с аминокислотного остатка CP участка. CP участок может быть расположен в любом домене белка Cas9, включая, например, домен спирали II, домен RuvCIII, или домен CTD. Например, CP участок может быть расположен (относительно Cas9 *S. pyogenes* согласно SEQ ID NO: 1361421) в исходном аминокислотном остатке 181, 199, 230, 270, 310, 1010, 1016, 1023, 1029, 1041, 1247, 1249, или 1282. Таким образом, после переноса на N-конец исходная аминокислота 181, 199, 230, 270, 310, 1010, 1016, 1023, 1029, 1041, 1247, 1249, или 1282 становится новой N-концевой аминокислотой. Номенклатура данных белков CP-Cas9 может быть обозначена как Cas9-CP^{181}, Cas9-CP^{199}, Cas9-CP^{230}, Cas9-CP^{270}, Cas9-CP^{310}, Cas9-CP^{1010}, Cas9-CP^{1016}, Cas9-CP^{1023}, Cas9-CP^{1029}, Cas9-CP^{1041}, Cas9-CP^{1247}, Cas9-CP^{1249}, и Cas9-CP^{1282}, соответственно. Данное описание не ограничивается получением вариантов CP на основе SEQ ID NO: 1361421, но может быть реализовано для получения вариантов CP в любой последовательности Cas9, либо в CP участках, которые соответствуют данным положениям, или в полностью других CP участках. Данное описание не является ограничивающим конкретные CP участки каким-либо образом. Практически любой CP участок может быть использован для образования варианта CP-Cas9.

[0391] Ниже приведены примеры аминокислотных последовательностей CP-Cas9 на основе Cas9 согласно SEQ ID NO: 1361421, в рамках которых последовательности линкера отмечены подчеркиванием, а необязательные остатки метионина (M) указаны жирным шрифтом. Следует понимать, что в рамках изобретения предусмотрены последовательности CP-Cas9, которые не включают последовательность линкера или которые включают различные последовательности линкера. Следует понимать, что последовательности CP-Cas9 могут быть основаны на последовательностях Cas9, отличных от SEQ ID NO: 1361421, при этом любые представленные здесь примеры не являются ограничивающими.

CP1012	DYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETN GETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEK NPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSK YVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADAN LDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKE VLDATLIHQSITGLYETRIDLSQLGGDGGSGGSGGSGGSGGSGGSGGDKKYSIGL AIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL	SEQ ID NO: 1361475

	NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQL PGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGD QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALV RQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLN REDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIP YYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPN EKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENED ILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLING IRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHI ANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRE RMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNA KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIK KYPKLESEFVYG	
CP1028	EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSP TVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYE TRIDLSQLGGDGGSGGSGGSGGSGGSGGMDKKYSIGLAIGTNSVGWAVITDE YKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPT IYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIAL SLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDA ILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGS IPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIEC FDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDR EMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQ TVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKD DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYK VYDVRKMIAKSEQ	SEQ ID NO: 1361476
CP1041	NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIV KKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTL TNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGG SGGSGGSGGSGGSGGSGGDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNT DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINA SGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDL AEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEIT KAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGAS QEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNF EEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEG MRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFN ASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLF DDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIH DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGR	SEQ ID NO: 1361477

	HKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN EKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKN RGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQE IGKATAKYFFYS	
CP1249	PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIR EQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYET RIDLSQLGGDGGSGGSGGSGGSGGSGGSGGMDKKYSIGLAIGTNSVGWAVITDEY KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRIC YLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI YHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQ TYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWM TRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRE MIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKV YDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD WDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF LYLASHYEKLKGS	SEQ ID NO: 1361478
CP1300	KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG LYETRIDLSQLGGDGGSGGSGGSGGSGGSGGDKKYSIGLAIGTNSVGWAVIT DEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKN RICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQ LVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLI ALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN GSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKI ECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFE DREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGI LQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTK AERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVIT LKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGD YKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNG ETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIAR KKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKY VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANL DKVLSAYNKHRD	SEQ ID NO: 1361479

[0392] Здесь описаны циркулярные пермутанты Cas9, которые могут применяться в рамках конструктов для праймерного редактирования. Примеры С-концевых фрагментов Cas9 на основе Cas9 согласно SEQ ID NO: 1361421, которые могут быть перенесены на N-конец Cas9, приведены ниже. Следует понимать, то подобные С-концевые фрагменты Cas9 представляют собой примеры и не являются ограничивающими.

С-концевой фрагмент CP1012	DYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETN GETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEK NPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSK YVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADAN LDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKE VLDATLIHQSITGLYETRIDLSQLGGD	SEQ ID NO: 1361480
С-концевой фрагмент CP1028	EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSP TVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYE TRIDLSQLGGD	SEQ ID NO: 1361481
С-концевой фрагмент CP1041	NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIV KKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTL TNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	SEQ ID NO: 1361482
С-концевой фрагмент CP1249	PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIR EQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYET RIDLSQLGGD	SEQ ID NO: 1361483
С-концевой фрагмент CP1300	KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG LYETRIDLSQLGGD	SEQ ID NO: 1361484

(ix) Варианты Cas9 с модифицированной специфичностью в отношении PAM

[0393] Праймерные редакторы по настоящему изобретению также могут включать варианты Cas9 с модифицированной специфичностью в отношении PAM. Некоторые аспекты данного изобретения предусматривают белки Cas9, которые демонстрируют активность на последовательности-мишени, которая не включает стандартный PAM (5′-NGG-3′, где N представляет собой A, C, G, или T) на ее 3′-конце. В некоторых вариантах осуществления белок Cas9 демонстрирует активность на последовательности-мишени, включающей последовательность PAM 5′-NGG-3′ на ее 3′-конце. В некоторых вариантах осуществления белок Cas9 демонстрирует активность на последовательности-мишени, включающей последовательность PAM 5′-NNG-3′ на ее 3′-конце. В некоторых вариантах осуществления белок Cas9 демонстрирует активность на последовательности-мишени, включающей последовательность PAM 5′-NNA-3′ на ее 3′-конце. В некоторых вариантах осуществления белок Cas9 демонстрирует активность на последовательности-мишени,

включающей последовательность PAM 5′-NNC-3′ на ее 3′-конце. В некоторых вариантах осуществления белок Cas9 демонстрирует активность на последовательности-мишени, включающей последовательность PAM 5′-NNT-3′ на ее 3′-конце. В некоторых вариантах осуществления белок Cas9 демонстрирует активность на последовательности-мишени, включающей последовательность PAM 5′-NGT-3′ на ее 3′-конце. В некоторых вариантах осуществления белок Cas9 демонстрирует активность на последовательности-мишени, включающей последовательность PAM 5′-NGA-3′ на ее 3′-конце. В некоторых вариантах осуществления белок Cas9 демонстрирует активность на последовательности-мишени, включающей последовательность PAM 5′-NGC-3′ на ее 3′-конце. В некоторых вариантах осуществления белок Cas9 демонстрирует активность на последовательности-мишени, включающей последовательность PAM 5′-NAA-3′ на ее 3′-конце. В некоторых вариантах осуществления белок Cas9 демонстрирует активность на последовательности-мишени, включающей последовательность PAM 5′-NAC-3′ на ее 3′-конце. В некоторых вариантах осуществления белок Cas9 демонстрирует активность на последовательности-мишени, включающей последовательность PAM 5′-NAT-3′ на ее 3′-конце. В некоторых вариантах осуществления белок Cas9 демонстрирует активность на последовательности-мишени, включающей последовательность PAM 5′-NAG-3′на ее 3′-конце.

[0394] Следует понимать, что любые описанные здесь аминокислотные мутации (например A262T) из первого аминокислотного остатка (например A) во второй аминокислотный остаток (например T) также могут включать мутации из первого аминокислотного остатка в аминокислотный остаток, который аналогичен второму аминокислотному остатку (например является консервативным). Например, мутация аминокислоты с гидрофобной боковой цепью (например аланина, валина, изолейцина, лейцина, метионина, фенилаланина, тирозина, или триптофана) может представлять собой мутацию второй аминокислоты с другой гидрофобной боковой цепью (например аланина, валина, изолейцина, лейцина, метионина, фенилаланина, тирозина, или триптофана). Например, мутация аланина на треонин (например мутация A262T) также может представлять собой мутацию из аланина в аминокислоту, которая имеет размер и химические свойства, аналогичные треонину, например серину. В качестве другого примера мутация аминокислоты с положительно заряженной боковой цепью (например аргинина, гистидина или лизина) может представлять собой мутацию во вторую

аминокислоту с другой положительно заряженной боковой цепью (например аргинин, гистидин, или лизин). В качестве другого примера мутация аминокислоты с полярной боковой цепью (например серина. треонина, аспарагина, или глутамина) может представлять собой мутацию во вторую аминокислоту с другой полярной боковой цепью (например серин, треонин, аспарагин, или глутамин). Дополнительные аналогичные аминокислотные пары включают, но без ограничения, следующие: фенилаланин и тирозин; аспарагин и глутамин; метионин и цистеин; аспарагиновую кислоту и глутаминовую кислоту; и аргинин и лизин. Специалист в данной области определит, что подобные консервативные аминокислотные замещения скорее всего будут иметь минимальное влияние не структуру белка и скорее всего будут хорошо переноситься без ухудшения функции. В некоторых вариантах осуществления любые указанные здесь аминокислотные мутации из одной аминокислоты в треонин могут представлять собой аминокислотную мутацию в серин. В некоторых вариантах осуществления любые указанные здесь аминокислотные мутации из одной аминокислоты в аргинин могут представлять собой аминокислотную мутацию в лизин. В некоторых вариантах осуществления любые указанные здесь аминокислотные мутации из одной аминокислоты в изолейцин могут представлять собой аминокислотную мутацию в аланин, валин, метионин, или лейцин. В некоторых вариантах осуществления любые указанные здесь аминокислотные мутации из одной аминокислоты в лизин могут представлять собой аминокислотную мутацию в аргинин. В некоторых вариантах осуществления любые указанные здесь аминокислотные мутации из одной аминокислоты в аспарагиновую кислоту могут представлять собой аминокислотную мутацию в глутаминовую кислоту или аспарагин. В некоторых вариантах осуществления любые указанные здесь аминокислотные мутации из одной аминокислоты в валин могут представлять собой аминокислотную мутацию в аланин, изолейцин, метионин, или лейцин. В некоторых вариантах осуществления любые указанные здесь аминокислотные мутации из одной аминокислоты в глицин могут представлять собой аминокислотную мутацию в аланин. Следует понимать, однако, что дополнительные консервативные аминокислотные остатки будут очевидны специалисту в данной области и любые аминокислотные мутации в другие консервативные аминокислотные остатки также входят в объем данного изобретения.

[0395] В некоторых вариантах осуществления в рамках настоящего изобретения могут быть использованы любые варианты Cas9, описанные здесь в разделе Перечень последовательностей.

SpCas9 H840A
DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYH
EKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYN
QLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFD
LAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASM
IKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDG
TEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIP
YYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL
LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS
VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK
EDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQT
TQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL
SDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGK
SKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA
GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVI
LADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA
TLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 1361593)

Cas9-NG_H840A
DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYH
EKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYN
QLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFD
LAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASM
IKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDG
TEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIP
YYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL
LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS
VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK
EDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQT
TQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL
SDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESIRPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGK
SKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA
RFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVI
LADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIDRKVYRSTKEVLDA
TLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 1361595)

KKH-Cas9 N580A
GKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRI
QRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDT
GNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQ
LDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLY
NALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGK
PEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQIS
NLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSP
VVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTT
GKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVK
QEEASKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKD
FINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAED
ALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKD
YKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHH
DPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDD
YPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQA
EFIASFYKNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPHIIKTIASKT
QSIKKYSTDILGNLYEVKSKKHPQIIKKG (SEQ ID NO: 1361596)

[0396] В некоторых вариантах осуществления белок Cas9 включает комбинацию мутаций, которые демонстрируют активность на последовательности-мишени, включающей последовательность PAM 5′-NAA-3′ на ее 3′-конце. В некоторых вариантах осуществления комбинация мутаций присутствует в любом одном из клонов, перечисленных в Таблице 1. В некоторых вариантах осуществления комбинация мутаций представляет собой консервативные мутации клонов, перечисленных в Таблице 1. В некоторых вариантах осуществления белок Cas9 включает комбинацию мутаций любого одного из клонов Cas9, перечисленных в Таблице X.

[0397] Таблица X: Клоны NAA PAM

Мутации из SpCas9 дикого типа (например SEQ ID NO: 1361421)
D177N, K218R, D614N, D1135N, P1137S, E1219V, A1320V, A1323D, R1333K
D177N, K218R, D614N, D1135N, E1219V, Q1221H, H1264Y, A1320V, R1333K
A10T, I322V, S409I, E427G, G715C, D1135N, E1219V, Q1221H, H1264Y, A1320V, R1333K
A367T, K710E, R1114G, D1135N, P1137S, E1219V, Q1221H, H1264Y, A1320V, R1333K
A10T, I322V, S409I, E427G, R753G, D861N, D1135N, K1188R, E1219V, Q1221H, H1264H, A1320V, R1333K
A10T, I322V, S409I, E427G, R654L, V743I, R753G, M1021T, D1135N, D1180G, K1211R, E1219V, Q1221H, H1264Y, A1320V, R1333K
A10T, I322V, S409I, E427G, V743I, R753G, E762G, D1135N, D1180G, K1211R, E1219V, Q1221H, H1264Y, A1320V, R1333K
A10T, I322V, S409I, E427G, R753G, D1135N, D1180G, K1211R, E1219V, Q1221H, H1264Y, S1274R, A1320V, R1333K
A10T, I322V, S409I, E427G, A589S, R753G, D1135N, E1219V, Q1221H, H1264H, A1320V, R1333K
A10T, I322V, S409I, E427G, R753G, E757K, G865G, D1135N, E1219V, Q1221H, H1264Y, A1320V, R1333K
A10T, I322V, S409I, E427G, R654L, R753G, E757K, D1135N, E1219V, Q1221H, H1264Y, A1320V, R1333K
A10T, I322V, S409I, E427G, K599R, M631A, R654L, K673E, V743I, R753G, N758H, E762G,

D1135N, D1180G, E1219V, Q1221H, Q1256R, H1264Y, A1320V, A1323D, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, N869S, N1054D, R1114G, D1135N, D1180G, E1219V, Q1221H, H1264Y, A1320V, A1323D, R1333K
A10T, I322V, S409I, E427G, R654L, L727I, V743I, R753G, E762G, R859S, N946D, F1134L, D1135N, D1180G, E1219V, Q1221H, H1264Y, N1317T, A1320V, A1323D, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, N803S, N869S, Y1016D, G1077D, R1114G, F1134L, D1135N, D1180G, E1219V, Q1221H, H1264Y, V1290G, L1318S, A1320V, A1323D, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, N803S, N869S, Y1016D, G1077D, R1114G, F1134L, D1135N, K1151E, D1180G, E1219V, Q1221H, H1264Y, V1290G, L1318S, A1320V, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, N803S, N869S, Y1016D, G1077D, R1114G, F1134L, D1135N, D1180G, E1219V, Q1221H, H1264Y, V1290G, L1318S, A1320V, A1323D, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, F693L, V743I, R753G, E762G, N803S, N869S, L921P, Y1016D, G1077D, F1080S, R1114G, D1135N, D1180G, E1219V, Q1221H, H1264Y, L1318S, A1320V, A1323D, R1333K
A10T, I322V, S409I, E427G, E630K, R654L, K673E, V743I, R753G, E762G, Q768H, N803S, N869S, Y1016D, G1077D, R1114G, F1134L, D1135N, D1180G, E1219V, Q1221H, H1264Y, L1318S, A1320V, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, F693L, V743I, R753G, E762G, Q768H, N803S, N869S, Y1016D, G1077D, R1114G, F1134L, D1135N, D1180G, E1219V, Q1221H, G1223S, H1264Y, L1318S, A1320V, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, F693L, V743I, R753G, E762G, N803S, N869S, L921P, Y1016D, G1077D, F1801S, R1114G, D1135N, D1180G, E1219V, Q1221H, H1264Y, L1318S, A1320V, A1323D, R1333K
A10T, I322V, S409I, E427G, R654L, V743I, R753G, M1021T, D1135N, D1180G, K1211R, E1219V, Q1221H, H1264Y, A1320V, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, M673I, N803S, N869S, G1077D, R1114G, D1135N, V1139A, D1180G, E1219V, Q1221H, A1320V, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, N803S, N869S, R1114G, D1135N, E1219V, Q1221H, A1320V, R1333K

[0398] В некоторых вариантах осуществления белок Cas9 включает аминокислотную последовательность, которая, по меньшей мере, на 80% идентична аминокислотной последовательности белка Cas9, представленной в рамках любого одного из вариантов в Таблице 1. В некоторых вариантах осуществления белок Cas9 включает аминокислотную последовательность, которая, по меньшей мере, на 85%, по меньшей мере, на 90%, по меньшей мере, на 92%, по меньшей мере, на 95%, по меньшей мере, на 96%, по меньшей мере, на 97%, по меньшей мере, на 98%, по меньшей мере, на 99%, или, по меньшей мере, на 99,5% идентична аминокислотной последовательности белка Cas9, представленной в рамках любого одного из вариантов в Таблице X.

[0399] В некоторых вариантах осуществления белок Cas9 демонстрирует повышенную активность на последовательности-мишени, которая не включает стандартный PAM (5′-NGG-3′) на ее 3′-конце в отличие от Cas9 *Streptococcus pyogenes* согласно SEQ ID NO: 1361421. В некоторых вариантах осуществления белок Cas9 демонстрирует активность на последовательности-мишени, имеющей 3′-конец, которая не является непосредственно

смежной со стандартной последовательностью PAM (5′-NGG-3′), которая является, по меньшей мере, 5-кратно повышенной по сравнению с активностью Cas9 *Streptococcus pyogenes* согласно SEQ ID NO: 1361421 на той же последовательности-мишени. В некоторых вариантах осуществления белок Cas9 демонстрирует активность на последовательности-мишени, которая не является непосредственно смежной со стандартной последовательностью PAM (5′-NGG-3′), которая является, по меньшей мере, 10-кратно, по меньшей мере, 50-кратно, по меньшей мере, 100-кратно, по меньшей мере, 500-кратно, по меньшей мере, 1000-кратно, по меньшей мере, 5000-кратно, по меньшей мере, 10000-кратно, по меньшей мере, 50000-кратно, по меньшей мере, 100000-кратно, по меньшей мере, 500000-кратно, или, по меньшей мере, 1000000-кратно повышенной по сравнению с активностью Cas9 *Streptococcus pyogenes* согласно SEQ ID NO: 1361421 на той же последовательности-мишени. В некоторых вариантах осуществления 3′-конец последовательности-мишени является непосредственно смежным с последовательностью AAA, GAA, CAA, или TAA. В некоторых вариантах осуществления белок Cas9 включает комбинацию мутаций, которые демонстрируют активность на последовательности-мишени, включающей последовательность PAM 5′-NAC-3′ на ее 3′-конце. В некоторых вариантах осуществления комбинация мутаций присутствует в любом одном из клонов, перечисленных в Таблице 2. В некоторых вариантах осуществления комбинация мутаций представляет собой консервативные мутации клонов, перечисленных в Таблице 2. В некоторых вариантах осуществления белок Cas9 включает комбинацию мутаций любого одного из клонов Cas9, перечисленных в Таблице Y.

[0400] Таблица Y: Клоны NAC PAM

Мутации из SpCas9 дикого типа (например SEQ ID NO: 1361421)
T472I, R753G, K890E, D1332N, R1335Q, T1337N
I1057S, D1135N, P1301S, R1335Q, T1337N
T472I, R753G, D1332N, R1335Q, T1337N
D1135N, E1219V, D1332N, R1335Q, T1337N
T472I, R753G, K890E, D1332N, R1335Q, T1337N
I1057S, D1135N, P1301S, R1335Q, T1337N
T472I, R753G, D1332N, R1335Q, T1337N
T472I, R753G, Q771H, D1332N, R1335Q, T1337N
E627K, T638P, K652T, R753G, N803S, K959N, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
E627K, T638P, K652T, R753G, N803S, K959N, R1114G, D1135N, K1156E, E1219V, D1332N, R1335Q, T1337N
E627K, T638P, V647I, R753G, N803S, K959N, G1030R, I1055E, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N

E627K, E630G, T638P, V647A, G687R, N767D, N803S, K959N, R1114G, D1135N, E1219V, D1332G, R1335Q, T1337N
E627K, T638P, R753G, N803S, K959N, R1114G, D1135N, E1219V, N1266H, D1332N, R1335Q, T1337N
E627K, T638P, R753G, N803S, K959N, I1057T, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
E627K, T638P, R753G, N803S, K959N, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
E627K, M631I, T638P, R753G, N803S, K959N, Y1036H, R1114G, D1135N, E1219V, D1251G, D1332G, R1335Q, T1337N
E627K, T638P, R753G, N803S, V875I, K959N, Y1016C, R1114G, D1135N, E1219V, D1251G, D1332G, R1335Q, T1337N, I1348V
K608R, E627K, T638P, V647I, R654L, R753G, N803S, T804A, K848N, V922A, K959N, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
K608R, E627K, T638P, V647I, R753G, N803S, V922A, K959N, K1014N, V1015A, R1114G, D1135N, K1156N, E1219V, N1252D, D1332N, R1335Q, T1337N
K608R, E627K, R629G, T638P, V647I, A711T, R753G, K775R, K789E, N803S, K959N, V1015A, Y1036H, R1114G, D1135N, E1219V, N1286H, D1332N, R1335Q, T1337N
K608R, E627K, T638P, V647I, T740A, R753G, N803S, K948E, K959N, Y1016S, R1114G, D1135N, E1219V, N1286H, D1332N, R1335Q, T1337N
K608R, E627K, T638P, V647I, T740A, N803S, K948E, K959N, Y1016S, R1114G, D1135N, E1219V, N1286H, D1332N, R1335Q, T1337N
I670S, K608R, E627K, E630G, T638P, V647I, R653K, R753G, I795L, K797N, N803S, K866R, K890N, K959N, Y1016C, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
K608R, E627K, T638P, V647I, T740A, G752R, R753G, K797N, N803S, K948E, K959N, V1015A, Y1016S, R1114G, D1135N, E1219V, N1266H, D1332N, R1335Q, T1337N
I570T, A589V, K608R, E627K, T638P, V647I, R654L, Q716R, R753G, N803S, K948E, K959N, Y1016S, R1114G, D1135N, E1207G, E1219V, N1234D, D1332N, R1335Q, T1337N
K608R, E627K, R629G, T638P, V647I, R654L, Q740R, R753G, N803S, K959N, N990S, T995S, V1015A, Y1036D, R1114G, D1135N, E1207G, E1219V, N1234D, N1266H, D1332N, R1335Q, T1337N
I562F, V565D, I570T, K608R, L625S, E627K, T638P, V647I, R654I, G752R, R753G, N803S, N808D, K959N, M1021L, R1114G, D1135N, N1177S, N1234D, D1332N, R1335Q, T1337N
I562F, I570T, K608R, E627K, T638P, V647I, R753G, E790A, N803S, K959N, V1015A, Y1036H, R1114G, D1135N, D1180E, A1184T, E1219V, D1332N, R1335Q, T1337N
I570T, K608R, E627K, T638P, V647I, R654H, R753G, E790A, N803S, K959N, V1015A, R1114G, D1127A, D1135N, E1219V, D1332N, R1335Q, T1337N
I570T, K608R, L625S, E627K, T638P, V647I, R654I, T703P, R753G, N803S, N808D, K959N, M1021L, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
I570S, K608R, E627K, E630G, T638P, V647I, R653K, R753G, I795L, N803S, K866R, K890N, K959N, Y1016C, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
I570T, K608R, E627K, T638P, V647I, R654H, R753G, E790A, N803S, K959N, V1016A, R1114G, D1135N, E1219V, K1246E, D1332N, R1335Q, T1337N
K608R, E627K, T638P, V647I, R654L, K673E, R753G, E790A, N803S, K948E, K959N, R1114G, D1127G, D1135N, D1180E, E1219V, N1286H, D1332N, R1335Q, T1337N
K608R, L625S, E627K, T638P, V647I, R654I, I670T, R753G, N803S, N808D, K959N, M1021L, R1114G, D1135N, E1219V, N1286H, D1332N, R1335Q, T1337N
E627K, M631V, T638P, V647I, K710E, R753G, N803S, N808D, K948E, M1021L, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N, S1338T, H1349R

[0401] В некоторых вариантах осуществления белок Cas9 включает аминокислотную последовательность, которая, по меньшей мере, на 80% идентична аминокислотной последовательности белка Cas9, представленной в рамках любого одного из вариантов в Таблице 2. В некоторых вариантах осуществления белок Cas9 включает аминокислотную последовательность, которая, по меньшей мере, на 85%, по меньшей мере, на 90%, по меньшей мере, на 92%, по меньшей мере, на 95%, по меньшей мере, на 96%, по меньшей мере, на 97%, по меньшей мере, на 98%, по меньшей мере, на 99%, или, по меньшей мере, на 99,5% идентична аминокислотной последовательности белка Cas9, представленной в рамках любого одного из вариантов в Таблице Y.

[0402] В некоторых вариантах осуществления белок Cas9 демонстрирует повышенную активность на последовательности-мишени, которая не включает стандартный PAM (5′-NGG-3′) на ее 3′-конце в отличие от Cas9 *Streptococcus pyogenes* согласно SEQ ID NO: 1361421. В некоторых вариантах осуществления белок Cas9 демонстрирует активность на последовательности-мишени, имеющей 3′-конец, которая не является непосредственно смежной со стандартной последовательностью PAM (5′-NGG-3′), которая является, по меньшей мере, 5-кратно повышенной по сравнению с активностью Cas9 *Streptococcus pyogenes* согласно SEQ ID NO: 1361421 на той же последовательности-мишени. В некоторых вариантах осуществления белок Cas9 демонстрирует активность на последовательности-мишени, которая не является непосредственно смежной со стандартной последовательностью PAM (5′-NGG-3′), которая является, по меньшей мере, 10-кратно, по меньшей мере, 50-кратно, по меньшей мере, 100-кратно, по меньшей мере, 500-кратно, по меньшей мере, 1000-кратно, по меньшей мере, 5000-кратно, по меньшей мере, 10000-кратно, по меньшей мере, 50000-кратно, по меньшей мере, 100000-кратно, по меньшей мере, 500000-кратно, или, по меньшей мере, 1000000-кратно повышенной по сравнению с активностью Cas9 *Streptococcus pyogenes* согласно SEQ ID NO: 1361421 на той же последовательности-мишени. В некоторых вариантах осуществления 3′-конец последовательности-мишени является непосредственно смежным с последовательностью AAC, GAC, CAC, или TAC.

[0403] В некоторых вариантах осуществления белок Cas9 включает комбинацию мутаций, которые демонстрируют активность на последовательности-мишени, включающей последовательность PAM 5′-NAT-3′ на ее 3′-конце. В некоторых вариантах

осуществления комбинация мутаций присутствует в любом одном из клонов, перечисленных в Таблице 3. В некоторых вариантах осуществления комбинация мутаций представляет собой консервативные мутации клонов, перечисленных в Таблице 3. В некоторых вариантах осуществления белок Cas9 включает комбинацию мутаций любого одного из клонов Cas9, перечисленных в Таблице Z.

[0001] Таблица Z: Клоны PAM NAT

Мутации из SpCas9 дикого типа (например SEQ ID NO: 1361421)
K961E, H985Y, D1135N, K1191N, E1219V, Q1221H, A1320A, P1321S, R1335L
D1135N, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L
V743I, R753G, E790A, D1135N, G1218S, E1219V, Q1221H, A1227V, P1249S, N1286K, A1293T, P1321S, D1322G, R1335L, T1339I
F575S, M631L, R654L, V748I, V743I, R753G, D853E, V922A, R1114G D1135N, G1218S, E1219V, Q1221H, A1227V, P1249S, N1286K, A1293T, P1321S, D1322G, R1335L, T1339I
F575S, M631L, R654L, R664K, R753G, D853E, V922A, R1114G D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, N1286K, P1321S, D1322G, R1335L
M631L, R654L, R753G, K797E, D853E, V922A, D1012A, R1114G D1135N, G1218S, E1219V, Q1221H, P1249S, N1317K, P1321S, D1322G, R1335L
F575S, M631L, R654L, R664K, R753G, D853E, V922A, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L
F575S, M631L, R654L, R664K, R753G, D853E, V922A, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L
F575S, D596Y, M631L, R654L, R664K, R753G, D853E, V922A, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, Q1256R, P1321S, D1322G, R1335L
F575S, M631L, R654L, R664K, K710E, V750A, R753G, D853E, V922A, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L
F575S, M631L, K649R, R654L, R664K, R753G, D853E, V922A, R1114G, Y1131C, D1135N, K1156E, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L
F575S, M631L, R654L, R664K, R753G, D853E, V922A, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L
F575S, M631L, R654L, R664K, R753G, D853E, V922A, I1057G, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, N1308D, P1321S, D1322G, R1335L
M631L, R654L, R753G, D853E, V922A, R1114G, Y1131C, D1135N, E1150V, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1332G, R1335L
M631L, R654L, R664K, R753G, D853E, I1057V, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1332G, R1335L
M631L, R654L, R664K, R753G, I1057V, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1332G, R1335L

[0404] Приведенное выше описание различных napDNAbp, которые могут применяться совместно с описываемыми здесь праймерными редакторами, не является ограничивающим каким-либо образом. Праймерные редакторы могут включать стандартный SpCas9 или любой ортолог белка Cas9, или любой вариант белка Cas9,

включая любой встречающийся в природе вариант, мутант, или иным образом модифицированную версию Cas9, которая известна или которая может быть получена или может быть подвергнута эволюции с помощью связанного с направленной эволюцией или иного мутагенезного процесса. В различных вариантах осуществления Cas9 или варианты Cas9 обладают никазной активностью, то есть только расщепляют цепь последовательности-мишени ДНК. В других вариантах осуществления Cas9 или варианты Cas9 имеют деактивированные нуклеазы, то есть представляют собой белки в виде "мертвого" Cas9. Другие белки в виде вариантов Cas9, которые могут применяться, представляют собой белки с более низкой молекулярной массой, в отличие от стандартного SpCas9 (например для более легкой доставки) или имеющие модифицированную или перестроенную первичную аминокислотную структуру (например циркулярные пермутантные форматы). Описанные здесь праймерные редакторы также могут включать эквиваленты Cas9, включая белки Cas12a/Cpf1 и Cas12b, которые представляют собой результат конвергентной эволюции. Используемые здесь napDNAbp (например SpCas9, вариант Cas9, или эквиваленты Cas9) могут также содержать различные модификации, которые изменяют/повышают их специфичность в отношении PAM. Наконец, изобретение охватывает любой Cas9, вариант Cas9, или эквивалент Cas9, который имеет, по меньшей мере, 70%, по меньшей мере, 75%, по меньшей мере, 80%, по меньшей мере, 85%, по меньшей мере, 90%, по меньшей мере, 91%, по меньшей мере, 92%, по меньшей мере, 93%, по меньшей мере, 94%, по меньшей мере, 95%, по меньшей мере, 96%, по меньшей мере, 97%, по меньшей мере, 98%, по меньшей мере, 99%, или, по меньшей мере, 99,9% идентичность последовательности относительно исходной последовательности Cas9, например исходной стандартной последовательности SpCas9 или исходного эквивалента Cas9 (например Cas12a (Cpf1).

[0405] Кроме того, могут использоваться любые доступные способы для получения или конструирования варианта или мутанта белка Cas9. В соответствии с используемым здесь значением термин "мутация" обозначает замещение остатка в последовательности, например нуклеиновой кислоты или аминокислотной последовательности, другим остатком, или делецию или вставку одного или нескольких остатков в последовательности. Мутации обычно описаны здесь на основе указания исходного остатка и далее положения остатка в последовательности и на основе указания нового

замещенного остатка. Различные способы осуществления указанных здесь аминокислотных замещений (мутаций) хорошо известны в данной области и описаны, например, в Green and Sambrook, Molecular Cloning: A Laboratory Manual (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)). Мутации могут включать различные категории, например полиморфизмы одиночных оснований, участки микродупликации, индели и инверсии, и не являются ограниченными каким-либо образом. Мутации могут включать мутации типа "потеря функции", что является нормальным результатом мутации, которая снижает или устраняет активность белка. Большинство мутаций потери функции являются рецессивными, поскольку в гетерозиготе вторая копия хромосомы несет не подверженную мутации версию гена, кодирующего полностью функциональный пептид, чье присутствие компенсирует эффект мутации. Мутации также охватывают мутации типа "приобретение функции", которые придают нехарактерную активность белку или клетке, которая иным образом не присутствует при нормальных условиях. Многие мутации, связанные с приобретением функции, располагаются в регуляторных последовательностях, а не в кодирующих участках, и, таким образом, имеют ряд последствий. Например, мутация может приводить к тому, что один или несколько генов экспрессируются в неправильных тканях, при этом эти ткани приобретают функции, которые они обычно не имеют. Вследствие их природы, мутации, связанные с приобретением функции, обычно являются доминантными.

[0406] Мутации могут быть внесены в исходный белок Cas9 с помощью сайт-направленного мутагенеза. Старые известные в данной области способы сайт-направленного мутагенеза основаны на субклонировании подвергаемой мутированию последовательности в вектор, например вектор на основе бактериофага M13, который обеспечивает изолирование одноцепочечной матрицы ДНК. В рамках данных способов осуществляется отжиг мутагенного праймера (то есть праймера, способного отжигаться на подвергаемом мутированию участке, но несущего один или несколько несовпадающих нуклеотидов в подвергаемом мутированию участке) на одноцепочечной матрице и затем полимеризация комплемента матрицы начиная с 3′-конца мутагенного праймера. Получаемые в результате этого дуплексы затем переносятся в бактерии-хозяева и осуществляется скрининг бляшек на предмет требуемой мутации. Недавно в рамках сайт-направленного мутагенеза стали использоваться способы ПЦР, которые имеют

преимущество, заключающееся в отсутствии потребности в одноцепочечной матрице. Кроме того, были разработаны способы, не требующие субклонирования. При осуществлении сайт-направленного мутагенеза на основе ПЦР необходимо учитывать несколько нюансов. Во-первых, в рамках этих способов является желательным снижать количество циклов ПЦР для предотвращения экспансии нежелательных мутаций, вносимых полимеразой. Во-вторых, должен осуществляться отбор для снижения количества не подвергнутых мутированию исходных молекул, остающихся в реакции. В-третьих, ПЦР способ с повышенной длиной является предпочтительным для обеспечения возможности использования одного набора праймеров ПЦР. И в-четвертых, вследствие активности не зависящего от матрицы концевого удлинения некоторых термостабильных полимераз зачастую является необходимым включать конечный финишный этап в процедуру перед лигированием тупого конца полученного с помощью ПЦР мутантного продукта.

[0407] Мутации также могут быть внесены с помощью процессов направленной эволюции, например непрерывной эволюции с помощью фагов (PACE) или прерывистой эволюции с помощью фагов (PANCE). В соответствии с используемым здесь значением термин "непрерывная эволюция с помощью фагов (PACE)" обозначает непрерывную эволюцию, в рамках которой используется фаг в качестве вирусных векторов. Общая концепция технологии PACE была описана, например, в международной PCT заявке PCT/US2009/056194, поданной 8 сентября 2009 г., опубликованной как WO 2010/028347 11 марта 2010 г.; международной PCT заявке PCT/US2011/066747, поданной 22 декабря 2011 г., опубликованной как WO 2012/088381 28 июня 2012 г.; заявке на патент США, патенте США 9023594, выданном 5 мая 2015 г., международной PCT заявке PCT/US2015/012022, поданной 20 января 2015 г., опубликованной как WO 2015/134121 11 сентября 2015 г., и международной PCT заявке PCT/US2016/027795, поданной 15 апреля 2016 г., опубликованной как WO 2016/168631 20 октября 2016 г., полное содержание каждой из которых приведено здесь в качестве ссылки. Подверженные ошибкам обратные транскриптазы также могут быть получены с помощью "прерывистой эволюции с помощью фагов (PANCE)", которая в соответствии с используемым здесь значением обозначает прерывистую эволюцию, в рамках которой используется фаг в качестве вирусных векторов. PANCE представляет собой упрощенный способ быстрой *in vivo*

направленной эволюции с применением последовательных переносов во флаконах подвергаемого эволюции "фага отбора" (SP), который содержит представляющий интерес подвергаемый эволюции ген, между свежими клетками-хозяевами *E. coli*, тем самым обеспечивая сохранение генов внутри *E. coli* неизменными, в то время как гены, содержащиеся в SP, непрерывно подвергаются эволюции. Последовательные переносы во флаконах долго служили в качестве широко доступного подхода для лабораторной эволюции микробов, при этом в недавнее время были разработаны аналогичные подходы для эволюции бактериофагов. Системе PANCE присуща более низкая строгость, чем системе PACE.

[0408] Любые из приведенных выше ссылок, которые относятся к Cas9 или эквивалентам Cas9, полностью приведены здесь в качестве ссылки, если не указано иное.

[0409] В некоторых вариантах осуществления napDNAbp представляет собой программируемый нуклеиновой кислотой ДНК-связывающий белок, который не требует стандартной (NGG) последовательности PAM. В некоторых вариантах осуществления napDNAbp представляет собой белок Argonaute. Один пример такого программируемого нуклеиновой кислотой ДНК-связывающего белка представляет собой белок Argonaute из *Natronobacterium gregoryi* (NgAgo). NgAgo представляет собой оцДНК-направляемую эндонуклеазу. NgAgo связывает 5' фосфорилированную оцДНК из ~24 нуклеотидов (нДНК) для ее направления к ее участку-мишени и осуществляет двуцепочечные разрывы ДНК в участке нДНК. В отличие от Cas9, система NgAgo-нДНК не требует смежного с протоспейсером мотива (PAM). Использование NgAgo без нуклеазной активности (dNgAgo) может значительно расширить перечень оснований, на которое может осуществляться нацеливание. Характеристики и применение NgAgo было описано в Gao et al., *Nat Biotechnol.*, 2016 Jul;34(7):768-73. PubMed PMID: 27136078; Swarts et al., Nature. 507(7491) (2014):258-61; и Swarts et al., Nucleic Acids Res. 43(10) (2015):5120-9, каждый из которых приведен здесь в качестве ссылки.

[0410] В некоторых вариантах осуществления napDNAbp представляет собой прокариотический гомолог белка Argonaute. Прокариотические гомологи белков Argonaute известны и были описаны, например, в Makarova K., *et al.*, "Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements", *Biol Direct.* 2009 Aug 25;4:29. doi: 10.1186/1745-6150-4-29,

полное содержание которого приведено здесь в качестве ссылки. В некоторых вариантах осуществления napDNAbp представляет собой белок Argunaute Marinitoga piezophila (MpAgo). Ассоциированный с CRISPR белок Argunaute Marinitoga piezophila (MpAgo) расщепляет одноцепочечные последовательности-мишени с помощью 5′-фосфорилированных направляющих. 5′ направляющие используются всеми известными Argonaute. Кристаллическая структура комплекса MpAgo-РНК демонстрирует участок связывания направляющей цепи, включающий остатки, которые блокируют 5′ фосфатные взаимодействия. Эти данные свидетельствуют об эволюции подкласса Argonaute с нестандартной специфичностью в отношении 5′-гидроксилированной направляющей. См., например, Kaya *et al.*, "A bacterial Argonaute with noncanonical guide RNA specificity", *Proc Natl Acad Sci U S A.* 2016 Apr 12;113(15):4057-62, полное содержание которого приведено здесь в качестве ссылки). Является очевидным, что могут использоваться другие белки Argonaute, при этом они находятся в рамках объема данного изобретения

[0411] В некоторых вариантах осуществления napDNAbp представляет собой одиночный эффектор микробной системы CRISPR-Cas. Одиночные эффекторы микробных систем CRISPR-Cas включают, без ограничения, Cas9, Cpf1, C2c1, C2c2, и C2c3. Как правило, микробные системы CRISPR-Cas разделяются на системы Класса 1 и Класса 2. Системы Класса 1 имеют мультисубъединичные эффекторные комплексы, в то время как системы Класса 2 имеют одиночный белковый эффектор. Например, Cas9 и Cpf1 представляют собой эффекторы Класса 2. Помимо Cas9 и Cpf1, в Shmakov *et al.*, “Discovery and Functional Characterization of Diverse Class 2 CRISPR Cas Systems”, *Mol. Cell*, 2015 Nov 5; 60(3): 385–397, полное содержание которого приведено здесь в качестве ссылки, были описаны три различных системы CRISPR-Cas Класса 2 (C2c1, C2c2, и C2c3). Эффекторы двух систем, C2c1 и C2c3, содержат RuvC-подобные эндонуклеазные домены, являющиеся родственными Cpf1. Третья система, C2c2, содержит эффектор с двумя предсказанными доменами HEPN РНКазы. Продуцирование зрелой РНК CRISPR не зависит от tracrРНК, в отличие от продуцирования РНК CRISPR за счет C2c1. C2c1 зависит как от РНК CRISPR, так и tracrРНК для расщепления ДНК. Было показано, что бактериальный C2c2 обладает уникальной РНКазной активностью в плане созревания РНК CRISPR, отличной от его РНК-активируемой связанной с расщеплением одноцепочечной РНК активности. Данные РНКазные функции отличаются друг от друга и

от связанного с процессированием РНК CRISPR поведения Cpf1. См., например, East-Seletsky, *et al.*, "Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection", *Nature,* 2016 Oct 13;538(7624):270-273, полное содержание которого приведено здесь в качестве ссылки. *In vitro* биохимический анализ C2c2 в Leptotrichia shahii показал, что C2c2 направляется одиночной РНК CRISPR и может быть запрограммирован для расщепления мишеней в виде оцРНК, несущих комплементарные протоспейсеры. Каталитические остатки в двух консервативных доменах HEPN опосредуют расщепление. Мутации в каталитических остатках продуцируют каталитически деактивированные РНК-связывающие белки. См., например, Abudayyeh *et al.*, "C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector", *Science,* 2016 Aug 5; 353(6299), полное содержание которого приведено здесь в качестве ссылки.

[0412] Сообщалось о кристаллической структуре C2c1 (AacC2c1) *Alicyclobaccillus acidoterrastris* в комплексе с химерной направляющей РНК в виде одиночной молекулы (sgРНК). См., например, Liu *et al.*, "C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism", *Mol. Cell,* 2017 Jan 19;65(2):310-322, полное содержание которого приведено здесь в качестве ссылки. Также сообщалось о кристаллической структуре в C2c1 *Alicyclobacillus acidoterrestris*, связанном с ДНК-мишенями в виде третичных комплексов. См., например,Yang *et al.*, "PAM-dependent Target DNA Recognition and Cleavage by C2C1 CRISPR-Cas endonuclease", *Cell,* 2016 Dec 15;167(7):1814-1828, полное содержание которого приведено здесь в качестве ссылки. Каталитически компетентные конформации AacC2c1 как с являющимися, так и не являющимися мишенью цепями ДНК были обнаружены в виде независимо расположенных в одном каталитическом кармане RuvC с C2c1-опосредованным расщеплением, приводящим к зигзагообразному семинуклеотидному разрыву ДНК-мишени. Структурные сравнения третичных комплексов C2c1 и ранее идентифицированных Cas9 и Cpf1 демонстрируют разнообразие механизмов, используемых в рамках систем CRISPR-Cas9.

[0413] В некоторых вариантах осуществления napDNAbp может представлять собой белок C2c1, 2c2, или C2c3. В некоторых вариантах осуществления napDNAbp представляет собой белок C2c1. В некоторых вариантах осуществления napDNAbp представляет собой

белок C2c2. В некоторых вариантах осуществления napDNAbp представляет собой белок C2c3. В некоторых вариантах осуществления napDNAbp включает аминокислотную последовательность, которая, по меньшей мере, на 85%, по меньшей мере, на 90%, по меньшей мере, на 91%, по меньшей мере, на 92%, по меньшей мере, на 93%, по меньшей мере, на 94%, по меньшей мере, на 95%, по меньшей мере, на 96%, по меньшей мере, на 97%, по меньшей мере, на 98%, по меньшей мере, на 99%, или, по меньшей мере, на 99,5% идентична встречающемуся в природе белку C2c1, C2c2, или C2c3. В некоторых вариантах осуществления napDNAbp представляет собой встречающийся в природе белок C2c1, C2c2, или C2c3.

[0414] Некоторые аспекты изобретения предусматривают домены Cas9, которые имеют различную специфичность в отношении PAM. Как правило, белки Cas9, такие как Cas9 из *S. pyogenes* (spCas9), требуют стандартную последовательность PAM NGG для связывания с определенным участком нуклеиновой кислоты. Это может ограничивать возможность редактировать требуемые основания в рамках генома. В некоторых вариантах осуществления может требоваться расположение предусмотренных здесь слитых белков для редактирования оснований в точном месторасположении, например где основание-мишень располагается в рамках участка длиной 4 основания (например "окна редактирования"), что приблизительно соответствует 15 основаниям выше PAM. См. Komor, A.C., *et al.*, "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" *Nature* 533, 420-424 (2016), полное содержание которого приведено здесь в качестве ссылки. Соответственно, в некоторых вариантах осуществления любой из предусмотренных здесь слитых белков может включать домен Cas9, который способен связывать нуклеотидную последовательность, которая не содержит стандартной (например NGG) последовательности PAM. Домены Cas9, которые связываются с нестандартными последовательностями PAM, были описаны в данной области и будут очевидны специалисту в данной области. Например, домены Cas9, которые связываются с нестандартными последовательностями PAM, были описаны в Kleinstiver, B. P., et al., "Engineered CRISPR-Cas9 nucleases with altered PAM specificities" Nature 523, 481-485 (2015); и Kleinstiver, B. P., et al., "Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition" *Nature Biotechnology* 33, 1293-1298 (2015); содержание каждого из которых приведено здесь в качестве ссылки.

[0415] Например, может использоватьсядомен napDNAbp с измененной специфичностью в отношении PAM, например домен, имеющий, по меньшей мере, 80%, по меньшей мере, 85%, по меньшей мере, 90%, по меньшей мере, 95%, или, по меньшей мере, 99% идентичность последовательности относительно Cpf1 *Francisella novicida* дикого типа (SEQ ID NO: 1361472) (D917, E1006, и D1255 выделены жирным шрифтом и подчеркнуты):

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEE
ILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLI
DAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSS
NDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQR
VFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYK
MSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQ
KLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKA
KYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKK
DLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIV
PLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKI
FDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQ
KGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFEN
ISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFY
RKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKF
NDEINLLLKEKANDVHILSIDRGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAI
EKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFEDLNFGFKRGRFKVEKQVY
QKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPV
TGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLI
NFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQ
MRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGK
KLNLVIKNEEYFEFVQNRNN (SEQ ID NO: 1361472)

[0416] Может использоваться дополнительный домен napDNAbp с измененной специфичностью в отношении PAM, например домен, имеющий, по меньшей мере, 80%, по меньшей мере, 85%, по меньшей мере, 90%, по меньшей мере, 95%, или, по меньшей мере, 99% идентичность последовательности относительно Cas9 *Geobacillus thermodenitrificans* дикого типа (SEQ ID NO: 1361473).

MKYKIGLDIGITSIGWAVINLDIPRIEDLGVRIFDRAENPKTGESLALPRRLARSARRRLRRRK
HRLERIRRLFVREGILTKEELNKLFEKKHEIDVWQLRVEALDRKLNNDELARILLHLAKRRGFR
SNRKSERTNKENSTMLKHIEENQSILSSYRTVAEMVVKDPKFSLHKRNKEDNYTNTVARDDLER
EIKLIFAKQREYGNIVCTEAFEHEYISIWASQRPFASKDDIEKKVGFCTFEPKEKRAPKATYTF
QSFTVWEHINKLRLVSPGGIRALTDDERRLIYKQAFHKNKITFHDVRTLLNLPDDTRFKGLLYD
RNTTLKENEKVRFLELGAYHKIRKAIDSVYGKGAAKSFRPIDFDTFGYALTMFKDDTDIRSYLR
NEYEQNGKRMENLADKVYDEELIEELLNLSFSKFGHLSLKALRNILPYMEQGEVYSTACERAGY
TFTGPKKKQKTVLLPNIPPIANPVVMRALTQARKVVNAIIKKYGSPVSIHIELARELSQSFDER
RKMQKEQEGNRKKNETAIRQLVEYGLTLNPTGLDIVKFKLWSEQNGKCAYSLQPIEIERLLEPG

YTEVDHVIPYSRSLDDSYTNKVLVLTKENREKGNRTPAEYLGLGSERWQQFETFVLTNKQFSKK
KRDRLLRLHYDENEENEFKNRNLNDTRYISRFLANFIREHLKFADSDDKQKVYTVNGRITAHLR
SRWNFNKNREESNLHHAVDAAIVACTTPSDIARVTAFYQRREQNKELSKKTDPQFPQPWPHFAD
ELQARLSKNPKESIKALNLGNYDNEKLESLQPVFVSRMPKRSITGAAHQETLRRYIGIDERSGK
IQTVVKKKLSEIQLDKTGHFPMYGKESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKKNGELG
PIIRTIKIIDTTNQVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCVPIYTIDMMKGILPNKAIEP
NKPYSEWKEMTEDYTFRFSLYPNDLIRIEFPREKTIKTAVGEEIKIKDLFAYYQTIDSSNGGLS
LVSHDNNFSLRSIGSRTLKRFEKYQVDVLGNIYKVRGEKRVGVASSSHSKAGETIRPL (SEQ
ID NO: 1361473)

[0417] В некоторых вариантах осуществления программируемый нуклеиновой кислотой ДНК-связывающий белок (napDNAbp) представляет собой программируемый нуклеиновой кислотой ДНК-связывающий белок, который не требует стандартной (NGG) последовательности PAM. В некоторых вариантах осуществления napDNAbp представляет собой белок Argonaute. Один пример такого программируемого нуклеиновой кислотой ДНК-связывающего белка представляет собой белок Argonaute из *Natronobacterium gregoryi* (NgAgo). NgAgo представляет собой оцДНК-направляемую эндонуклеазу. NgAgo связывает 5' фосфорилированную оцДНК из ~24 нуклеотидов (нДНК) для ее направления к ее участку-мишени и осуществляет двуцепочечные разрывы ДНК в участке нДНК. В отличие от Cas9, система NgAgo-нДНК не требует смежного с протоспейсером мотива (PAM). Использование NgAgo без нуклеазной активности (dNgAgo) может значительно расширить перечень оснований, на которое может осуществляться нацеливание. Характеристики и применение NgAgo было описано в Gao et al., *Nat Biotechnol.*, 34(7): 768-73 (2016), PubMed PMID: 27136078; Swarts et al., Nature, 507(7491): 258-61 (2014); и Swarts et al., Nucleic Acids Res. 43(10) (2015): 5120-9, каждый из которых приведен здесь в качестве ссылки. Последовательность Argonaute *Natronobacterium gregoryi* приведена в SEQ ID NO: 1361474.

[0418] Описанные слитые белки могут включать домен napDNAbp, имеющий, по меньшей мере, 80%, по меньшей мере, 85%, по меньшей мере, 90%, по меньшей мере, 95%, или, по меньшей мере, 99% идентичность последовательности относительно Argonaute *Natronobacterium gregoryi* дикого типа (SEQ ID NO: 1361474).

MTVIDLDSTTTADELTSGHTYDISVTLTGVYDNTDEQHPRMSLAFEQDNGERRYITLWKNTTPK
DVFTYDYATGSTYIFTNIDYEVKDGYENLTATYQTTVENATAQEVGTTDEDETFAGGEPLDHHL
DDALNETPDDAETESDSGHVMTSFASRDQLPEWTLHTYTLTATDGAKTDTEYARRTLAYTVRQE
LYTDHDAAPVATDGLMLLTPEPLGETPLDLDCGVRVEADETRTLDYTTAKDRLLARELVEEGLK
RSLWDDYLVRGIDEVLSKEPVLTCDEFDLHERYDLSVEVGHSGRAYLHINFRHRFVPKLTLADI
DDDNIYPGLRVKTTYRPRRGHIVWGLRDECATDSLNTLGNQSVVAYHRNNQTPINTDLLDAIEA

ADRRVVETRRQGHGDDAVSFPQELLAVEPNTHQIKQFASDGFHQQARSKTRLSASRCSEKAQAF
AERLDPVRLNGSTVEFSSEFFTGNNEQQLRLLYENGESVLTFRDGARGAHPDETFSKGIVNPPE
SFEVAVVLPEQQADTCKAQWDTMADLLNQAGAPPTRSETVQYDAFSSPESISLNVAGAIDPSEV
DAAFVVLPPDQEGFADLASPTETYDELKKALANMGIYSQMAYFDRFRDAKIFYTRNVALGLLAA
AGGVAFTTEHAMPGDADMFIGIDVSRSYPEDGASGQINIAATATAVYKDGTILGHSSTRPQLGE
KLQSTDVRDIMKNAILGYQQVTGESPTHIVIHRDGFMNEDLDPATEFLNEQGVEYDIVEIRKQP
QTRLLAVSDVQYDTPVKSIAAINQNEPRATVATFGAPEYLATRDGGGLPRPIQIERVAGETDIE
TLTRQVYLLSQSHIQVHNSTARLPITTAYADQASTHATKGYLVQTGAFESNVGFL (SEQ ID
NO: 1361474)

[0419] В некоторых вариантах осуществления домен Cas9 представляет собой домен Cas9 из *Staphylococcus aureus* (SaCas9). В некоторых вариантах осуществления домен SaCas9 представляет собой имеющий нуклеазную активность SaCas9, не имеющий нуклеазной активности SaCas9 (SaCas9d), или никазу SaCas9 (SaCas9n).

(x) Разделенные домены napDNAbp для доставки разделенного PE

[0420] В различных вариантах осуществления описанные здесь праймерные редакторы могут быть доставлены в клетки в виде двух или более фрагментов, которые становятся собранными внутри клетки (либо за счет пассивной сборки, или за счет активной сборки, например использования последовательностей разделенного интеина) с образованием восстановленного праймерного редактора. В некоторых случаях самосборка может являться пассивной, когда два или более фрагмента праймерного редактора ковалентно или нековалентно связываются внутри клетки для восстановления праймерного редактора. В других случаях самосборка может катализироваться доменами димеризации, располагаемыми на каждом из фрагментов. Примеры доменов димеризации описаны здесь. В других случаях самосборка может катализироваться последовательностями разделенного интеина, располагаемыми на каждом из фрагментов праймерного редактора.

[0421] Доставка разделенного PE может быть полезной для преодоления различных связанных с размером ограничений разных подходов к доставке. Например, подходы к доставке могут включать способы доставки на основе вирусов, способы доставки на основе информационной РНК, или доставку на основе RNP (доставку на основе рибонуклеопротеина). При этом каждый из данных способов доставки может быть более эффективным и/или действенным за счет разделения праймерного редактора на части более малого размера. При попадании внутрь клетки части малого размера могут собираться в функциональный праймерный редактор. В зависимости от способа разделения разделенные фрагменты праймерного редактора могут быть собраны обратно

нековалентным способом или ковалентным способом для восстановления праймерного редактора. В одном варианте осуществления праймерный редактор может быть разделен в одном или нескольких участках разделения на два или более фрагмента. Фрагменты могут являться немодифицированными (не считая того, что они являются разделенными). После доставки фрагментов в клетку (например с помощью прямой доставки рибонуклеопротеинового комплекса или с помощью ядерной доставки - например доставки с помощью иРНК или доставки на основе вирусного вектора) фрагменты могут ковалентно или нековалентно связываться для восстановления праймерного редактора. В другом варианте осуществления праймерный редактор может быть разделен в одном или нескольких участках разделения на два или более фрагмента. Каждый из фрагментом может быть модифицирован для включения домена димеризации, при этом каждый образованный фрагмент соединяется с доменом димеризации. При доставке или экспрессии в клетке домены димеризации различных фрагментов ассоциируются и связываются друг с другом, объединяя различные фрагменты праймерного редактора вместе для восстановления функционального праймерного редактора. В другом варианте осуществления фрагмент праймерного редактора может быть модифицирован для включения разделенного интеина. При доставке или экспрессии в клетке домены разделенного интеина различных фрагментов ассоциируются и связываются друг с другом и затем подвергаются транс-сплайсингу, который приводит к эксцизии доменов разделенного интеина из каждого из фрагментов и сопутствующему образованию пептидной связи между фрагментами, тем самым обеспечивая восстановление праймерного редактора.

[0422] В одном варианте осуществления праймерный редактор может быть доставлен с применением подхода на основе разделенного интеина.

[0423] Участок разделения может быть расположен между любой одной или несколькими парами остатков в праймерном редакторе и в любом из доменов в нем, включая домен napDNAbp, полимеразный домен (например домен RT), линкерный домен, который соединяет домен napDNAbp и полимеразный домен.

[0424] В одном варианте осуществления, изображенном на Фиг.66, праймерный редактор (PE) разделен в участке разделения в napDNAbp.

[0425] В определенных вариантах осуществления napDNAbp представляет собой стандартный полипептид SpCas9 согласно SEQ ID NO: 1361421, а именно:

SpCas9 *Streptococcus pyogenes* M1 SwissProt кат. # Q99ZW2 дикий тип 1368 AA	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDS GETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEED KKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFR GHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDL DNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDG TEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKI EKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKD FLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSG QGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVD QELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMK NYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQIL DSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNF FKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSK KLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR KRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHY LDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	SEQ ID NO: 1361421

[0426] В определенных вариантах осуществления SpCas9 разделен на два фрагмента в участке разделения, расположенном между остатками 1 и 2, или 2 и 3, или 3 и 4, или 4 и 5, или 5 и 6, или 6 и 7, или 7 и 8, или 8 и 9, или 9 и 10, или между любыми двумя парами остатков, расположенными между остатками 1-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, 1000-1100, 1100-1200, 1200-1300, или 1300-1368 стандартного SpCas9 согласно SEQ ID NO: 1361421.

[0427] В определенных вариантах осуществления napDNAbp разделен на два фрагмента в участке разделения, расположенном в паре остатков, которые соответствуют любым двум парам остатков, расположенным где-либо между положениями 1-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, 1000-1100, 1100-1200, 1200-1300, или 1300-1368 стандартного SpCas9 согласно SEQ ID NO: 1361421.

[0428] В определенных вариантах осуществления SpCas9 разделен на два фрагмента в участке разделения, расположенном между остатками 1 и 2, или 2 и 3, или 3 и 4, или 4 и 5, или 5 и 6, или 6 и 7, или 7 и 8, или 8 и 9, или 9 и 10, или между любыми двумя парами

остатков, расположенными между остатками 1-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, 1000-1100, 1100-1200, 1200-1300, или 1300-1368 стандартного SpCas9 согласно SEQ ID NO: 1361421.

[0429] В определенных вариантах осуществления участок разделения расположен в одном или нескольких участках полипептидных связей (то есть "участок разделения или участок разделения разделенного интеина"), слит с разделенным интеином и затем доставляется в клетки в виде отдельно кодируемых слитых белков. При экспрессии слитых белков в виде разделенного интеина (то есть половин белка) в клетке белки подвергаются транс-сплайсингу с образованием полного или цельного PE с сопутствующим удалением последовательностей объединенного разделенного интеина.

[0430] Например, как показано на Фиг.38, N-концевой экстеин может быть слит с первым разделенным интеином (например N-интеином), а C-концевой экстеин может быть слит со вторым разделенным интеином (например C-интеином). N-концевой экстеин становится слитым с C-концевым экстеином для восстановления цельного слитого белка праймерного редактора, включающего домен napDNAbp и полимеразный домен (например домен RT), при самоассоциации N-интеина и C-интеина внутри клетки и далее их самоэксцизии с сопутствующим образованием пептидной связи между частями N-концевого экстеина и C-концевого экстеина цельного праймерного редактора (PE).

[0431] Для использования преимущества, связанного со стратегией доставки разделенного PE с использованием разделенных интеинов, праймерный редактор должен быть разделен в одном или нескольких участках разделения для создания, по меньшей мере, двух отдельных половин праймерного редактора, каждая из которых может быть соединена внутри клетки, если каждая половина слита с последовательностью разделенного интеина.

[0432] В определенных вариантах осуществления праймерный редактор разделен в одном участке разделения. В определенных других вариантах осуществления праймерный редактор разделен в двух участках разделения, или трех участках разделения, или четырех участках разделения, или более.

[0433] В предпочтительном варианте осуществления праймерный редактор разделен в одном участке разделения для создания двух отдельных половин праймерного редактора, каждая из которых может быть слита с последовательностью разделенного интеина.

[0434] Примером разделенного интеина является интеин *Ssp* DnaE, который включает две субъединицы, а именно DnaE-N и DnaE-C. Две различных субъединицы кодируются отдельными генами, а именно *dnaE-n* и *dnaE-c*, которые кодируют субъединицы DnaE-N и DnaE-C, соответственно. DnaE представляет собой встречающийся в природе разделенный интеин в *Synechocytis sp.* PCC6803 и способен направлять транс-сплайсинг двух отдельных белков, каждый из которых включает продукт слияния с DnaE-N или DnaE-C.

[0435] Дополнительные встречающиеся в природе или сконструированные последовательности разделенного интеина известны или могут быть реализованы на основе описанных здесь или доступных в данной области последовательностей цельного интеина. Примеры последовательностей разделенного интеина можно найти в Stevens et al., "A promiscuous split intein with expanded protein engineering applications," PNAS, 2017, Vol.114: 8538-8543; Iwai et al., "Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostc punctiforme, FEBS Lett, 580: 1853-1858, каждый из которых приведен здесь в качестве ссылки. Дополнительные последовательности разделенного интеина можно найти, например, в WO 2013/045632, WO 2014/055782, WO 2016/069774, и EP2877490, содержание каждой из которых приведено здесь в качестве ссылки.

[0436] Кроме того, сплайсинг белка *in trans* был описан *in vivo* и *in vitro* (Shingledecker, et al., [0076] *Gene* 207:187 (1998), Southworth, et al., *EMBO J.* 17:918 (1998); Mills, et al., *Proc. Natl. Acad. Sci. USA,* 95:3543-3548 (1998); Lew, et al., *J. Biol. Chem.,* 273:15887-15890 (1998); Wu, et al., *Biochim. Biophys. Acta* 35732:1 (1998b), Yamazaki, et al., *J. Am. Chem. Soc.* 120:5591 (1998), Evans, et al., *J. Biol. Chem.* 275:9091 (2000); Otomo, et al., *Biochemistry* 38:16040-16044 (1999); Otomo, et al., *J. Biolmol. NMR* 14:105-114 (1999); Scott, et al., *Proc. Natl. Acad. Sci. USA* 96:13638-13643 (1999)), при этом он обеспечивает возможность экспрессировать белок в виде двух неактивных фрагментов, которые далее подвергаются лигированию для образования функционального продукта, например как показано на Фиг.38 и 39 в отношении образования полного слитого белка PE из двух отдельно экспрессируемых половин.

[0437] В различных описанных здесь вариантах осуществления могут использоваться способы непрерывной эволюции (например PACE) для подвергания эволюции первой части редактора оснований. Первая часть может включать один компонент или домен,

например домен Cas9, домен дезаминазы, или домен UGI. Отдельно подвергнутый эволюции компонент или домен затем может быть слит с остальными частями редактора оснований в клетке путем отдельной экспрессии как подвергнутой эволюции части, так и остальных не подвергнутых эволюции частей с полипептидными доменами разделенного интеина. Первая часть может более широким образом включать любую первую аминокислотную часть редактора оснований, которую требуется подвергнуть эволюции с использованием описанного здесь способа непрерывной эволюции. Вторая часть в рамках данного варианта осуществления будет обозначать оставшуюся аминокислотную часть редактора оснований, которая не подвергается эволюции с помощью описанных здесь способов. Подвергнутая эволюции первая часть и вторая часть редактора оснований могут экспрессироваться с полипептидными доменами разделенного интеина в клетке. Природные механизмы сплайсинга белка клетки будут обеспечивать сборку подвергнутой эволюции первой части и не подвергнутой эволюции второй части для образования подвергнутого эволюции редактора оснований в виде единого слитого белка. Подвергнутая эволюции первая часть может включать N- или C-концевую часть единого слитого белка. Аналогичным образом применение второй ортогональной подвергаемой транс-сплайсингу пары интеина может обеспечивать включение подвергнутой эволюции первой частью внутренней части единого слитого белка.

[0438] Таким образом, любые из подвергнутых или не подвернутых эволюции компонентов описанных здесь редакторов оснований могут экспрессироваться с метками в виде разделенного интеина для реализации образования полного редактора оснований, включающего подвергнутый и не подвергнутый эволюции компонент в клетке.

[0439] Механизм процесса сплайсинга белка был подробно изучен (Chong, et al., J. Biol. Chem. 1996, 271, 22159-22168; Xu, M-Q & Perler, F. B. EMBO Journal, 1996, 15, 5146-5153), при этом консервативные аминокислоты были обнаружены в точках сплайсинга интеина и экстеина (Xu, et al., EMBO Journal, 1994, 13 5517-522). Описанные здесь конструкты содержат последовательность интеина, слитую с 5′-концом первого гена (например подвергнутой эволюции частью редактора оснований). Подходящие последовательности интеина могут быть выбраны из любых белков, в отношении которых известно, что они содержат элементы сплайсинга белка. Базу данных, содержащую все известные интеины, можно найти в интернете (Perler, F. B. Nucleic Acids Research, 1999, 27, 346-347).

Последовательность интеина слита на 3′ конце с 5′ концом второго гена. Для нацеливания данного гена на определенную органеллу пептидный сигнал может быть слит с кодирующей последовательностью гена. После второго гена последовательность гена интеина может быть повторена необходимое количество раз для экспрессии нескольких белков в одной клетке. Для содержащих несколько интеинов конструктов может являться полезным использовать элементы в виде интеина из разных источников. После последовательности последнего подвергаемого экспрессии гена должна быть вставлена последовательность терминации транскрипции. В одном варианте осуществления сконструирована единица для сплайсинга в виде модифицированного интеина, так что она может как катализировать эксцизию экстеинов из интеинов, так и предотвращать лигирование экстеинов. Было обнаружено, что мутагенез С-концевого соединения экстеина в ДНК-полимеразе GB-D вида Pyrococcus продуцирует измененный элемент сплайсинга, который индуцирует расщепление экстеинов и интеинов, но предотвращает последующее лигирование экстеинов (Xu, M-Q & Perler, F. B. EMBO Journal, 1996, 15, 5146-5153). Мутация серина 538 в аланин или глицин индуцировала расщепление, но предотвращала лигирование. Мутация эквивалентных остатков в других единицах для сплайсинга в виде интеина также должна предотвращать лигирование экстеина вследствие консервации аминокислот в С-концевом соединении экстеина с интеином. Предпочтительный интеин, не содержащий эндонуклеазный домен, представляет собой белок GyrA Mycobacterium xenopi (Telenti, et al. J. Bacteriol. 1997, 179, 6378-6382). Другие из них были обнаружены в природе или были искусственно созданы путем удаления эндонуклеазных доменов из содержащих эндонуклеазу интеинов (Chong, et al. J. Biol. Chem. 1997, 272, 15587-15590). В предпочтительном варианте осуществления интеин выбран таким образом, что он состоит из минимального количества аминокислот, необходимых для реализации функции сплайсинга, например интеин из белка GyrA Mycobacterium xenopi (Telenti, A., et al., J. Bacteriol. 1997, 179, 6378-6382). В альтернативном варианте осуществления выбирают интеин без эндонуклеазной активности, например интеин из белка GyrA Mycobacterium xenopi или интеин VMA Saccharaomyces cerevisiae, который был модифицирован для удаления эндонуклеазных доменов (Chong, 1997).Дополнительная модификация единицы для сплайсинга в виде интеина может обеспечивать изменение скорости реакции расщепления, обеспечивающее

управление дозировкой белка путем простого модифицирования генной последовательности единицы для сплайсинга.

[0440] Интеины также могут существовать в виде двух фрагментов, кодируемых двумя отдельно транскрибируемыми и транслируемыми генами. Данные так называемые разделенные интеины самоассоциируются и катализируют активность сплайсинга белка in trans. Разделенные интеины были идентифицированы у многих цианобактерий и архей (Caspi et al, Mol Microbiol. 50: 1569-1577 (2003); Choi J. et al, J Mol Biol. 556: 1093-1106 (2006.); Dassa B. et al, Biochemistry. 46:322-330 (2007.); Liu X. and Yang J., J Biol Chem. 275:26315-26318 (2003); Wu H. et al.

[0441] Proc Natl Acad Sci USA. Pound5:9226-9231 (1998.); и Zettler J. et al, FEBS Letters. 553:909-914 (2009)), но до сих пор не были обнаружены у эукариотов. Недавно биоинформационный анализ природных метагеномных данных выявил 26 различных локусов с новым геномным строением. В каждом локусе консервативный кодирующий фермент участок прерывается разделенным интеином со свободным эндонуклеазным геном, вставленным между участками, кодирующими субдомены интеина. Среди них пять локусов были полностью собраны: ДНК хеликазы (gp41-l, gp41-8); инозин-5 '-монофосфатдегидрогеназа (IMPDH-1); и каталитические субъединицы рибонуклеотид-редуктазы (NrdA-2 и NrdJ-1). Похоже, что данная фрагментированная организация генов присутствует в основном у фагов (Dassa et al, Nucleic Acids Research. 57:2560-2573 (2009)).

[0442] Разделенный интеин Npu DnaE был описан как имеющий наиболее высокую отмеченную скорость реакции транс-сплайсинга белка. Кроме того, реакция сплайсинга белка Npu DnaE считается устойчивой и связанной с высоким выходом в плане различных последовательностей экстеина, температурами от 6 до 37°C, и наличием до 6M мочевины (Zettler J. et al, FEBS Letters. 553:909-914 (2009); Iwai I. et al, FEBS Letters 550: 1853-1858 (2006)). Как ожидается, при внесении мутации Cysl Ala в N-домене данных интеинов первоначальный N-C-ацильный сдвиг и, таким образом, сплайсинг белка блокируется. К сожалению, реакция C-концевого расщепления также практически полностью ингибировалась. Зависимость циклизации аспарагина в C-концевом сплайсинговом соединении от ацильного сдвига в N-концевой неустойчивой пептидной связи вероятно является уникальным свойством, присущим аллелям естественным образом разделенного интеина DnaE (Zettler J. et al. FEBS Letters. 555:909-914 (2009).

[0443] Механизм сплайсинга белка обычно имеет четыре этапа [29-30]: 1) N-S или N-O ацильный сдвиг на N-конце интеина, который разрывает расположенную выше пептидную связь и образует эфирную связь между N-экстеином и боковой цепью первой аминокислоты интеина (Cys или Ser); 2) трансэстерификацию с переносом N-экстеина на C-конец интеина, образующую новую эфирную связь, присоединяющую N-экстеин к боковой цепи первой аминокислоты C-экстеина (Cys, Ser, или Thr); 4) циклизацию Asn, разрывающую пептидную связь между интеином и C-экстеином; и 4) S-N или O-N ацильный сдвиг, который замещает эфирную связь пептидной связью между N-экстеином и C-экстеином.

[0444] Транс-сплайсинг белка, катализируемый разделенными интеинами, обеспечивает полностью ферментативный способ лигирования белка [31]. Разделенный интеин представляет собой по существу непрерывный интеин (например мини-интеин), разделенный на две части, называемые N-интеином и C-интеином, соответственно. N-интеин и C-интеин разделенного интеина могут нековалентно связываться для образования активного интеина и катализировать реакцию сплайсинга по существу так же, как и непрерывный интеин. Разделенные интеины были обнаружены в природе, а также сконструированы в лабораториях [31-35]. В соответствии с используемым здесь значением термин "разделенный интеин" обозначает любой интеин, в котором присутствует один или несколько разрывов пептидных связей между N-концевыми и C-концевыми последовательностями аминокислот, так что N-концевые и C-концевые последовательности становятся отдельными молекулами, которые могут нековалентно связываться или восстанавливаться с образованием интеина, который является функциональным для реакций транс-сплайсинга белка. Для получения разделенного интеина для использования в рамках способов по изобретению может применяться любой каталитически активный интеин или его фрагмент. Например, в одном аспекте разделенный интеин может быть получен из эукариотического интеина. В другом аспекте разделенный интеин может быть получен из бактериального интеина. В другом аспекте разделенный интеин может быть получен из архейного интеина. Предпочтительно, полученный таким образом разделенный интеин будет иметь только аминокислотные последовательности, необходимые для катализирования реакций транс-сплайсинга.

[0445] В соответствии с используемым здесь значением термин "N-концевой разделенный интеин (In)" обозначает любую последовательность интеина, которая включает N-концевую аминокислотную последовательность, которая является функциональной для реакций транс-сплайсинга. Таким образом, In также включает последовательность, которая устраняется при осуществлении транс-сплайсинга. In может включать последовательность, которая представляет собой модификацию N-концевой части встречающейся в природе последовательности интеина. Например, In может включать дополнительные аминокислотные остатки и/или мутированные остатки, если включение подобных дополнительных и/или мутированных остатков не делает In нефункциональным в плане транс-сплайсинга. Предпочтительно, включение дополнительных и/или мутированных остатков повышает или стимулирует активность In, связанную с транс-сплайсингом.

[0446] В соответствии с используемым здесь значением термин "C-концевой разделенный интеин (Ic)" обозначает любую последовательность интеина, которая включает C-концевую аминокислотную последовательность, которая является функциональной для реакций транс-сплайсинга. В одном аспекте Ic включает от 4 до 7 последовательных аминокислотных остатков, по меньшей мере, 4 аминокислоты из которых происходят из последней β-цепи интеина, из которого он был получен. Таким образом, Ic также включает последовательность, которая устраняется при осуществлении транс-сплайсинга. Ic может включать последовательность, которая представляет собой модификацию C-концевой части встречающейся в природе последовательности интеина. Например, Ic может включать дополнительные аминокислотные остатки и/или мутированные остатки, если включение подобных дополнительных и/или мутированных остатков не делает Ic нефункциональным в плане транс-сплайсинга. Предпочтительно, включение дополнительных и/или мутированных остатков повышает или стимулирует активность Ic, связанную с транс-сплайсингом.

[0447] В некоторых вариантах осуществления изобретения пептид, связанный с Ic или In, может включать дополнительный химический фрагмент, включая, помимо прочено, флуоресцентные группы, биотин, полиэтиленгликоль (ПЭГ), аналоги аминокислот, искусственные аминокислоты, фосфатные группы, гликозильные группы, радиоизотопные метки, и фармацевтические молекулы. В других вариантах осуществления пептид,

связанный с Ic, может включать одну или несколько химически реактивных групп, включая, среди прочих, кетон, альдегид, остатки Cys и остатки Lys. N-интеин и C-интеин разделенного интеина могут нековалентно связываться для образования активного интеина и катализировать реакцию сплайсинга при присутствии "полипептида сплайсинга интеина (ISP)". В соответствии с используемым здесь значением термин "полипептид сплайсинга интеина (ISP)" обозначает часть аминокислотной последовательности разделенного интеина, которая сохраняется, когда Ic, In, или они оба удаляются из разделенного интеина. В определенных вариантах осуществления In включает ISP. В другом варианте осуществления Ic включает ISP. В другом варианте осуществления ISP представляет собой отдельный пептид, который не является ковалентно связанным ни с In, ни с Ic.

[0448] Разделенные интеины могут быть созданы из непрерывных интеинов путем конструирования одного или нескольких участков разделения в неструктурированной петле или внесения аминокислотной последовательности между 12 консервативными бета-цепями, находящимися в структуре мини-интеинов [25-28]. Может присутствовать некоторая гибкость в плане расположения участка разрыва в участках между бета-цепями при условии, что создание разрыва не будет нарушать структуру интеина, в частности структурированных бета-цепей, до значительной степени, при которой теряется активность, связанная со сплайсингом белка.

[0449] При транс-сплайсинге белка один белок-предшественник состоит из части в виде N-экстеина и далее N-интеина, другой белок-предшественник состоит из C-интеина и далее части в виде C-экстеина, при этом реакция транс-сплайсинга (катализируемая совместно N- и C-интеинами) обеспечивает эксцизию данных двух последовательностей интеина и соединение данных двух последовательностей экстеина с помощью пептидной связи. Транс-сплайсинг белка, являясь ферментативной реакцией, может работать при очень низких (например микромолярных) концентрациях белков и может осуществляться при физиологических условиях.

B. Программируемые нуклеазы (отличные от napDNAbp)

[0450] В различных описанных здесь вариантах осуществления праймерные редакторы включают napDNAbp, например белок Cas9. Данные белки "программируются" за счет образования комплексов с направляющей РНК (или PEgRNA, как может быть в этом

случае), которая направляет белок Cas9 к участку-мишени на ДНК, которая обладает последовательностью, являющейся комплементарной спейсерной части нНРК (или PEgRNA), и которая также обладает требуемой последовательностью PAM. Однако в определенных предполагаемых здесь вариантах осуществления napDNAbp может быть замещен другим типом программируемого белка, например нуклеазой цинкового пальца или подобной активатору транскрипции эффекторной нуклеазой (TALEN).

[0451] На Фиг.1H показан вариант предполагаемого здесь праймерного редактирования, в рамках которого осуществляется замещение napDNAbp (например никазы SpCas9) любым доменом программируемой нуклеазы, например нуклеазами цинкового пальца (ZFN) или подобными активатору транскрипции эффекторными нуклеазами (TALEN). Таким образом, предполагается, что подходящие нуклеазы не обязательно должны "программироваться" нацеленной на нуклеиновую кислоту молекулой (например направляющей РНК), но вместо этого могут программироваться путем задания специфичности ДНК-связывающего домена, например (и в частности) нуклеазы. Как и при праймерном редактировании с использованием фрагментов napDNAbp, является предпочтительным, чтобы такие альтернативные программируемые нуклеазы были модифицированы таким образом, что только одна цепь ДНК-мишени подвергается разрыву. Другими словами, программируемые нуклеазы предпочтительно должны функционировать в качестве никаз. После выбора программируемой нуклеазы (например ZFN или TALEN) в систему могут быть добавлены дополнительные виды функциональности для обеспечения ее работы в соответствии с подобным праймерному редактированию механизмом. Например, программируемые нуклеазы могут быть модифицированы путем присоединения к ним (например через химический линкер) РНК или ДНК плеча удлинения, при этом плечо удлинения включает участок связывания праймера (PBS) и матрицу для синтеза ДНК. Программируемая нуклеаза также может быть присоединена (например через химический или аминокислотный линкер) к полимеразе, природа которой будет зависеть от того, является ли плечо удлинения ДНК или РНК. В случае плеча удлинения в виде РНК полимераза может представлять собой РНК-зависимую ДНК-полимеразу (например обратную транскриптазу). В случае плеча удлинения в виде ДНК полимераза может представлять собой ДНК-зависимую ДНК-полимеразу (например прокариотическую полимеразу, включая Pol I, Pol II, или Pol III,

или эукариотическую полимеразу, включая Pol a, Pol b, Pol g, Pol d, Pol e, или Pol z). Система может также включать другие виды функциональности, добавленные в виде продуктов слияния с программируемыми нуклеазами или добавленные *in trans* для реализации реакции в целом (например (a) хеликазу для разматывания ДНК в участке разрыва, чтобы сделать цепь с разрывом с 3'-концом доступной в качестве праймера, (b) FEN1 для обеспечения удаления эндогенной цепи на цепи с разрывом для смещения реакции в сторону замещения эндогенной цепи синтезированной цепью, или (c) комплекс nCas9:нРНК для создания второго участка разрыва на противоположной цепи, что может обеспечивать интеграцию синтезированной коррекции за счет предпочтительной клеточной репарации не подвергнутой редактированию цепи). Аналогично праймерному редактированию с использованием napDNAbp, подобный комплекс с иным образом программируемой нуклеазой может использоваться для синтеза и последующего встраивания новой синтезированной замещающей цепи ДНК, несущей представляющее интерес изменение, перманентно в участок-мишень ДНК.

[0452] Подходящие альтернативно программируемые нуклеазы хорошо известны в данной области и могут использоваться вместо комплекса napDNAbp:нРНК для конструирования альтернативной системы праймерного редактирования, которая может быть запрограммирована для избирательного связывания с участком-мишенью ДНК и которая может быть дополнительно модифицирована описанным выше образом для солокализации полимеразы и РНК или ДНК плеча удлинения, включающего участок связывания праймера и матрицу для синтеза ДНК, в определенном участке разрыва. В качестве примера и как показано на Фиг.1H подобные активатору транскрипции эффекторные нуклеазы (TALEN) могут применяться в качестве программируемой нуклеазы в рамках описанных здесь способов праймерного редактирования и соответствующих композиций. TALEN представляют собой искусственные ферменты рестрикции, полученные путем слиянии эффекторного связывающего ДНК домена TAL с доменом расщепления ДНК. Данные реагенты обеспечивают эффективное, программируемое, и направленное расщепление ДНК и представляют собой мощные инструменты для геномного редактирования *in situ*. Подобные активатору транскрипции эффекторы (TALE) могут быть быстро сконструированы для связывания практически с любой последовательностью ДНК. В соответствии с используемым здесь значением

термин "TALEN" является широким и включает мономерную TALEN, которая может расщеплять двуцепочечную ДНК без помощи со стороны другой TALEN. Термин "TALEN" также используется для обозначения одного или обоих членов пары TALEN, которые конструируются совместной работы для расщепления ДНК в одном и том же участке. TALEN, которые работают совместно, могут быть обозначены как левая TALEN и правая TALEN, что обозначает хиральность ДНК. См. заявку на патент США 12/965590; заявку на патент США 13/426991 (патент США. 8450471); заявку на патент США 13/427040 (патент США 8440431); заявку на патент США 13/427137 (патент США 8440432); и заявку на патент США 13/738381, все из которых полностью приведены здесь в качестве ссылки. Кроме того, TALEN описаны в WO 2015/027134, патенте США 9181535, Boch et al., "Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors", Science, vol. 326, pp. 1509-1512 (2009), Bogdanove et al., TAL Effectors: Customizable Proteins for DNA Targeting, Science, vol. 333, pp. 1843-1846 (2011), Cade et al., "Highly efficient generation of heritable zebrafish gene mutations using homo- and heterodimeric TALENs", Nucleic Acids Research, vol. 40, pp. 8001-8010 (2012), и Cermak et al., "Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting", Nucleic Acids Research, vol. 39, No. 17, e82 (2011), каждый из которых приведен здесь в качестве ссылки.

[0453] Как показано на Фиг.1H, нуклеазы цинкового пальца также могут использоваться в качестве альтернативных программируемых нуклеаз в рамках праймерного редактирования вместо napDNAb, например никаз Cas9. Подобно TALEN, белки ZFN могут быть модифицированы таким образом, что они функционируют в качестве никаз, то есть за счет конструирования ZFN, так что он расщепляет только одну цепь ДНК-мишени таким же образом, что и napDNAbp, используемый с описанными здесь праймерными редакторами. Белки ZFN были обширно описаны в данной области, например в Carroll et al., "Genome Engineering with Zinc-Finger Nucleases," *Genetics*, Aug 2011, Vol. 188: 773-782; Durai et al., "Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells," *Nucleic Acids Res*, 2005, Vol. 33: 5978-90; и Gaj et al., "ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering," *Trends Biotechnol.* 2013, Vol.31: 397-405, каждый из которых полностью приведен здесь в качестве ссылки.

C. <u>Полимеразы (например обратные транскриптазы)</u>

[0454] В различных вариантах осуществления описанная здесь система праймерного редактора (PE) включает полимеразу (например ДНК-зависимую ДНК-полимеразу или РНК-зависимую ДНК-полимеразу, например обратную транскриптазу) или ее вариант, который может поступать в виде слитого белка с napDNAbp или другой программируемой нуклеазой или поступать *in trans*.

[0455] В рамках описанных здесь праймерных редакторов может использоваться любая полимераза. Полимеразы могут представлять собой полимеразы дикого типа, функциональные фрагменты, мутанты, варианты, или усеченные варианты, и им подобные. Полимеразы могут включать полимеразы дикого типа из эукариотических, прокариотических, архейных, или вирусных организмов, и/или полимеразы могут быть модифицированы с помощью генной инженерии, мутагенеза, процессов на основе направленной эволюции. Полимеразы могут включать ДНК-полимеразу T7, ДНК-полимеразу T5, ДНК-полимеразу T4, фрагмент ДНК-полимеразы Klenow, ДНК-полимеразу III, и им подобные. Полимеразы также могут являться термостабильными и могут включать Taq, Tne, Tma, Pfu, Tfl, Tth, фрагмент Stoffel, ДНК-полимеразы VENT® и DEEPVENT®, KOD, Tgo, JDF3, и их мутанты, варианты, и производные (см. патент США 5436149; патент США 4889818; патент США 4965185; патент США 5079352; патент США 5614365; патент США 5374553; патент США 5270179; патент США 5047342; патент США 5512462; WO 92/06188; WO 92/06200; WO 96/10640; Barnes, W. M., Gene 112:29-35 (1992); Lawyer, F. C., et al., PCR Meth. Appl. 2:275-287 (1993); Flaman, J.-M, et al., Nuc. Acids Res. 22(15):3259-3260 (1994), каждый из которых приведен здесь в качестве ссылки). Для синтеза более длинных молекул нуклеиновой кислоты (например молекул нуклеиновой кислоты длиной более чем около 3-5 kb) могут использоваться, по меньшей мере, две ДНК-полимеразы. В определенных вариантах осуществления одна из полимераз может по существу не иметь 3' экзонуклеазной активности, а другая может иметь 3' экзонуклеазную активность. Подобные пары могут включать полимеразы, которые являются одинаковыми или различающимися. Примеры ДНК-полимераз, которые по существу не имеют 3' экзонуклеазной активности, включают, но без ограничения, Taq, Tne(exo-), Tma(exo-), Pfu(exo-), Pwo(exo-), exo-KOD и Tth ДНК-полимеразы, и их мутанты, варианты и производные.

[0456] Предпочтительно, полимеразы, которые могут использоваться в рамках описанных здесь праймерных редакторов, представляют собой "зависимую от матрицы" полимеразу (поскольку полимеразы должны полагаться на матрицу для синтеза ДНК для указания последовательности цепи ДНК при синтезе во время праймерного редактирования). В соответствии с используемым здесь значением термин "молекула матричной ДНК" обозначает цепь нуклеиновой кислоты, из которой синтезируется комплементарная цепь нуклеиновой кислоты ДНК-полимеразой, например при реакции праймерного удлинения матрицы для синтеза ДНК PEgRNAPEgRNA.

[0457] В соответствии с используемым здесь значением термин "зависимым от матрицы образом" обозначает процесс, который включает зависимое от матрицы удлинение праймерной молекулы (например синтез ДНК ДНК-полимеразой). Термин "зависимым от матрицы образом" обозначает полинуклеотидный синтез РНК или ДНК, при котором последовательность новой синтезированной цепи полинуклеотида определяется хорошо известными правилами связывания комплементарных оснований (см., например, Watson, J. D. et al., In: Molecular Biology of the Gene, 4th Ed., W. A. Benjamin, Inc., Menlo Park, Calif. (1987)). Термин "комплементарный" обозначает широкую концепцию комплементарности последовательностей между участками двух полинуклеотидных цепей или между двумя нуклеотидами за счет спаривания оснований. Известно, что нуклеотид аденин способен образовывать определенные водородные связи ("спаривание оснований") с нуклеотидом, который представляет собой тимин или урацил. Аналогичным образом, известно, что нуклеотид цитозин способен обеспечивать спаривание оснований с нуклеотидом гуанином. Таким образом, в случае праймерного редактирования может быть сказано, что одна цепь ДНК, синтезированная полимеразой праймерного редактора на основе матрицы для синтеза ДНК, является "комплементарной" последовательности матрицы для синтеза ДНК.

(i) Примеры полимераз

[0458] В различных вариантах осуществления описанные здесь праймерные редакторы включают полимеразу. Изобретение охватывает любую полимеразу дикого типа, полученную из любого встречающегося в природе организма или вируса, или полученную из коммерческого или некоммерческого источника. Кроме того, полимеразы, которые могут применяться в рамках праймерных редакторов по изобретению, могут включать

любую встречающуюся в природе мутантную полимеразу, сконструированную мутантную полимеразу, или другую представляющую собой вариант полимеразу, включая усеченные варианты, которые сохраняют функцию. Полимеразы, которые могут использоваться здесь, также могут быть сконструированы для включения определенных аминокислотных замещений, например тех, которые конкретно указаны здесь. В определенных предпочтительных вариантах осуществления полимеразы, которые могут использоваться в рамках праймерных редакторов по изобретению, представляют собой полимеразы на основе матрицы, то есть они синтезируют нуклеотидные последовательности зависимым от матрицы образом.

[0459] Полимераза представляет собой фермент, который синтезирует нуклеотидную цепь и который может применяться совместно с описанными здесь системами праймерных редакторов. Полимеразы предпочтительно являются "зависимыми от матрицы" полимеразами (то есть полимеразой, которая синтезирует нуклеотидную цепь на основе порядка нуклеотидных оснований матричной цепи). В определенных конфигурациях полимеразы также могут являться "независимыми от матрицы" (то есть полимеразой, которая синтезирует нуклеотидную цепь без необходимости в матричной цепи). Полимераза также может быть дополнительно разделена на категории в виде "ДНК-полимеразы" или "РНК-полимеразы". В различных вариантах осуществления система праймерного редактора включает ДНК-полимеразу. В различных вариантах осуществления ДНК-полимераза может представлять собой "ДНК-зависимую ДНК-полимеразу (где молекула матрицы представляет собой цепь ДНК). В таких случаях молекула матрицы ДНК может представлять собой PEgRNAPEgRNA, где плечо удлинения включает цепь ДНК. В таких случаях PEgRNAPEgRNA может быть обозначена как химерная или гибридная PEgRNAPEgRNA, которая включает часть в виде РНК (то есть компоненты направляющей РНК, включая спейсер и ядро нРНК) и часть в виде ДНК (например плечо удлинения). В различных других вариантах осуществления ДНК-полимераза может представлять собой "РНК-зависимую ДНК-полимеразу" (где молекула матрицы представляет собой цепь РНК). В таких случаях PEgRNAPEgRNA представляет собой РНК, то есть включает удлинение РНК. Термин "полимераза" также может обозначать фермент, который катализирует полимеразу нуклеотида (то есть полимеразную активность). В целом, данный фермент будет инициировать синтез на 3'-

конце праймера, отжигаемого на последовательности полинуклеотидной матрицы (например последовательности праймера, отжигаемой на участке связывания праймера PEgRNAPEgRNA) по направлению к 5'-концу цепи матрицы. ДНК-полимераза катализирует полимеризацию дезоксинуклеотидов. В соответствии с используемым здесь значением в отношении ДНК-полимеразы термин "ДНК-полимераза" включает "ее функциональный фрагмент". Термин "ее функциональный фрагмент" обозначает любую часть относящейся к дикому типу или мутантной ДНК-полимеразы, которая охватывает менее чем целую аминокислотную последовательность полимеразы и которая сохраняет способность, по меньшей мере, при одном наборе условий катализировать полимеризацию полинуклеотида. Подобный функциональный фрагмент может присутствовать в виде отдельной сущности или может представлять собой часть более крупного полипептида, например слитого белка.

[0460] В некоторых вариантах осуществления полимеразы могут быть получены из бактериофага. ДНК-полимеразы бактериофагов в целом лишены экзонуклеазной активности от 5' к 3', поскольку данная активность кодируется отдельным полипептидом. Примеры подходящих ДНК-полимераз представляют собой ДНК-полимеразу T4, T7, и phi29. Коммерчески доступные ферменты представляют собой: T4 (доступный из многих источников, например Epicentre) и T7 (доступный из многих источников, например Epicentre для немодифицированного и USB для ДНК-полимеразы T7 exo от 3' к 5' "Sequenase").

[0461] В других вариантах осуществления полимеразы представляют собой архейные полимеразы. Существует 2 различных класса ДНК-полимераз, которые были идентифицированы у архей: 1. Семейство B/pol типа I (гомологи Pfu из Pyrococcus furiosus) и 2. pol типа II (гомологи P. furiosus DP1/2-субъединичной полимеразы DP2). Было показано, что ДНК-полимеразы обоих классов естественным образом не имеют ассоциированной экзонуклеазной активности от 5' к 3' и обладают экзонуклеазной (проверочной) активностью от 3' к 5'. Подходящие ДНК-полимеразы (pol I или pol II) могут быть получены из архей с оптимальными температурами роста, которые аналогичны требуемым температурам в рамках анализов.

[0462] Термостабильные архейные ДНК-полимеразы выделяют из вида Pyrococcus (furiosus, вид GB-D, woesii, abysii, horikoshii), вида Thermococcus (kodakaraensis KOD1,

litoralis, вид 9 degrees North-7, вид JDF-3, gorgonarius), Pyrodictium occultum, и Archaeoglobus fulgidus.

[0463] Полимеразы также могут быть получены из эубактериальных видов. Существует 3 класса эубактериальных ДНК-полимераз: pol I, II, и III. Ферменты семейства ДНК-полимераз pol I обладают экзонуклеазной активностью от 5' к 3', при этом определенные участники также демонстрируют экзонуклеазную активность от 3' к 3'. ДНК-полимеразы pol II естественным образом не имеют экзонуклеазную активность от 5' к 3', но демонстрируют экзонуклеазную активность от 3' к 5'. ДНК-полимеразы pol III представляют собой крупную репликативную ДНК-полимеразу клетки и состоят из нескольких субъединиц. Каталитическая субъединица pol III не имеет экзонуклеазную активность от 5' к 3', но в некоторых случаях экзонуклеазная активность от 3' к 5' располагается в том же полипептиде.

[0464] Существует множество коммерчески доступных ДНК-полимераз pol I, некоторые из которых были модифицированы для снижения или устранения экзонуклеазной активности от 5' к 3'.

[0465] Подходящие термостабильные ДНК-полимеразы pol I могут быть выделены из множества термофильных эубактерий, включая виды Thermus и Thermotoga maritime, например Thermus aquaticus (Taq), Thermus thermophilus (Tth) и Thermotoga maritima (Tma UlTma).

[0466] Дополнительные эубактерии, являющиеся родственными тем, которые перечислены выше, описаны в Thermophilic Bacteria (Kristjansson, J. K., ed.) CRC Press, Inc., Boca Raton, Fla., 1992.

[0467] Изобретение дополнительно включает химерные и нехимерные ДНК-полимеразы, которые химически модифицированы в соответствии со способами, описанными в патентах США 5677152, 6479264 и 6183998, содержание которых полностью приведено здесь в качестве ссылки.

[0468] Дополнительные архейные ДНК-полимеразы, являющиеся родственными тем, которые перечислены выше, описаны в следующих ссылках: Archaea: A Laboratory Manual (Robb, F. T. and Place, A. R., eds.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1995 и Thermophilic Bacteria (Kristjansson, J. K., ed.) CRC Press, Inc., Boca Raton, Fla., 1992.

(ii) B. Примеры обратных транскриптаз

[0469] В различных вариантах осуществления описанная здесь система праймерного редактора (PE) включает обратную транскриптазу или ее вариант.

[0470] Обратные транскриптазы представляют собой многофункциональные ферменты обычно с тремя видами ферментативной активности, включая активность в виде РНК- и ДНК-зависимой полимеризации ДНК, и активность РНКазы H, которая катализирует расщепление РНК в РНК-ДНК гибридах. Некоторые мутанты обратной транскриптазы вызвали устранение способности фрагмента в виде РНКазы H предотвращать ненамеренное повреждение иРНК. Данные ферменты, которые синтезируют комплементарную ДНК (кДНК) с использованием иРНК в качестве матрицы, были сначала идентифицированы в РНК-вирусах. После этого обратные транскриптазы были выделены и очищены напрямую из вирусных частиц, клеток, или тканей. (см., например, Kacian et al., 1971, Biochim. Biophys. Acta 46: 365-83; Yang et al., 1972, Biochem. Biophys. Res. Comm. 47: 505-11; Gerard et al., 1975, J. Virol. 15: 785-97; Liu et al., 1977, Arch. Virol. 55 187-200; Kato et al., 1984, J. Virol. Methods 9: 325-39; Luke et al., 1990, Biochem. 29: 1764-69 и Le Grice et al., 1991, J. Virol. 65: 7004-07, каждый из которых приведен здесь в качестве ссылки). Недавно были созданы мутанты и слитые белки в попытке улучшить свойства, такие как термостабильность, точность и активность. Здесь охватываются любые относящиеся к дикому типу, вариантные, и/или мутантные формы обратной транскриптазы, которые известны в данной области или которые могут быть созданы с помощью известных в данной области способов.

[0471] Ген обратной транскриптазы (RT) (или содержащаяся в нем генетическая информация) может быть получены из ряда различных источников. Например ген может быть получен из эукариотических клеток, которые инфицированы ретровирусом, или из ряда плазмид, которые содержат либо часть, либо полный геном ретровируса. Кроме того, из ретровирусов может быть получена РНК, подобная информационной РНК, которая содержит ген RT. Примеры источников RT включают, но без ограничения, вирус лейкоза мышей Молони (M-MLV или MLVRT); вирус Т-клеточного лейкоза человека типа 1 (HTLV-1); вирус лейкоза крупного рогатого скота (BLV); вирус саркомы Рауса (RSV); вирус иммунодефицита человека (ВИЧ); дрожжи, включая Saccharomyces, Neurospora, Drosophila; приматов; и грызунов. См., например, Weiss, et al., патент США 4663290

(1987); Gerard, G. R., DNA:271-79 (1986); Kotewicz, M. L., et al., Gene 35:249-58 (1985); Tanese, N., et al., Proc. Natl. Acad. Sci. (USA):4944-48 (1985); Roth, M. J., at al., J. Biol. Chem. 260:9326-35 (1985); Michel, F., et al., Nature 316:641-43 (1985); Akins, R. A., et al., Cell 47:505-16 (1986), EMBO J. 4:1267-75 (1985); и Fawcett, D. F., Cell 47:1007-15 (1986) (каждый из которых полностью приведен здесь в качестве ссылки).

(a) RT дикого типа

[0472] Примеры ферментов для применения в рамках описанных здесь праймерных редакторов могут включать, но без ограничения, обратную транскриптазу M-MLV и обратную транскриптазу RSV. Ферменты, имеющие активность обратной транскриптазы, являются коммерчески доступными. В определенных вариантах осуществления обратная транскриптаза предоставляется *in trans* другим компонентам системы праймерного редактора (PE). То есть обратная транскриптаза экспрессируется или иным образом предоставляется в виде отдельного компонента, то есть не в виде слитого белка с napDNAbp.

[0473] Среднему специалисту в данной области будет очевидно, что обратные транскриптазы дикого типа, включая, но без ограничения, обратную транскриптазу вируса лейкемии мышей Молони (M-MLV), обратную транскриптазу вируса иммунодефицита человека (ВИЧ) и обратную транскриптазу вируса саркомы-лейкоза птиц (ASLV), которая включает, но без ограничения, обратную транскриптазу вируса саркомы Рауса (RSV), обратную транскриптазу вируса миелобластоза птиц (AMV), обратную транскриптазу хелперного вируса MCAV вируса эритробластоза птиц (AEV), обратную транскриптазу хелперного вируса MCAV вируса миелоцитоматоза птиц MC29, обратную транскриптазу хелперного вируса REV-A вируса ретикулоэндотелиоза птиц (REV-T), обратную транскриптазу хелперного вируса UR2AV вируса саркомы птиц UR2, обратную транскриптазу хелперного вируса YAV вируса саркомы птиц Y73, обратную транскриптазу Раус-ассоциированного вируса (RAV) и обратную транскриптазу миелобластоз-ассоциированного вируса (MAV), могут подходящим образом использоваться в рамках настоящих описанных здесь способов и композиций.

[0474] Примеры ферментов RT дикого типа приведены ниже:

Описание	Последовательность
Обратная транскрипт	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYP MSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVP

аза(RT M-MLV RT) дикого типа Вирус лейкемии мышей Молони Используется в PE1 (описанный здесь слитый белок праймерного редактора 1)	NPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNS PTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQ ICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLY PLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRR PVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNA RMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHT WYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSR YAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNR MADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 1361485)
Обратная транскриптаза Вирус лейкемии мышей Молони Исх. посл. AAA66622.1	AFPLERPDWD YTTQAGRNHL VHYRQLLLAG LQNAGRSPTN LAKVKGITQG PNESPSAFLE RLKEAYRRYT PYDPEDPGQE TNVSMSFIWQ SAPDIGRKLG RLEDLKSKTL GDLVREAEKI FNKRETPEER EERIRRETEE KEERRRTVDE QKEKERDRRR HREMSKLLAT VVIGQEQDRQ EGERKRPQLD KDQCAYCKEK GHWAKDCPKK PRGPRGPRPQ TSLLTLGDXG GQGQDPPPEP RITLKVGGQP VTFLVDTGAQ HSVLTQNPGP LSDKSAWVQG ATGGKRYRWT TDRKVHLATG KVTHSFLHVP DCPYPLLGRD LLTKLKAQIH FEGSGAQVVG PMGQPLQVLT LNIEDEYRLH ETSKEPDVSL GFTWLSDFPQ AWAESGGMGL AVRQAPLIIP LKATSTPVSI KQYPMSQEAR LGIKPHIQRL LDQGILVPCQ SPWNTPLLPV KKPGTNDYRP VQDLREVNKR VEDIHPTVPN PYNLLSGLPP SHQWYTVLDL KDAFFCLRLH PTSQPLFAFE WRDPEMGISG QLTWTRLPQG FKNSPTLFDE ALHRDLADFR (SEQ ID NO: 1361486)
Обратная транскриптаза Вирус лейкемии кошек Исх. посл. NP955579.1	TLQLEEEYRL FEPESTQKQE MDIWLKNFPQ AWAETGGMGT AHCQAPVLIQ LKATATPISI RQYPMPHEAY QGIKPHIRRM LDQGILKPCQ SPWNTPLLPV KKPGTEDYRP VQDLREVNKR VEDIHPTVPN PYNLLSTLPP SHPWYTVLDL KDAFFCLRLH SESQLLFAFE WRDPEIGLSG QLTWTRLPQG FKNSPTLFDE ALHSDLADFR VRYPALVLLQ YVDDLLLAAA TRTECLEGTK ALLETLGNKG YRASAKKAQI CLQEVTYLGY SLKDGQRWLT KARKEAILSI PVPKNSRQVR EFLGTAGYCR LWIPGFAELA APLYPLTRPG TLFQWGTEQQ LAFEDIKKAL LSSPALGLPD ITKPFELFID ENSGFAKGVL VQKLGPWKRP VAYLSKKLDT VASGWPPCLR MVAAIAILVK DAGKLTLGQP LTILTSHPVE ALVRQPPNKW LSNARMTHYQ AMLLDAERVH FGPTVSLNPA TLLPLPSGGN HHDCLQILAE THGTRPDLTD QPLPDADLTW YTDGSSFIRN GEREAGAAVT TESEVIWAAP LPPGTSAQRA ELIALTQALK MAEGKKLTVY TDSRYAFATT HVHGEIYRRR GLLTSEGKEI KNKNEILALL EALFLPKRLS IIHCPGHQKG DSPQAKGNRL ADDTAKKAAT ETHSSLTVL (SEQ ID NO: 1361487)
Обратная транскриптаза RT ВИЧ-1, цепь A Исх. посл. ITL3-A	PISPIETVPV KLKPGMDGPK VKQWPLTEEK IKALVEICTE MEKEGKISKI GPENPYNTPV FAIKKKDSTK WRKLVDFREL NKRTQDFWEV QLGIPHPAGL KKKKSVTVLD VGDAYFSVPL DEDFRKYTAF TIPSINNETP GIRYQYNVLP QGWKGSPAIF QSSMTKILEP FRKQNPDIVI YQYMDDLYVG SDLEIGQHRT KIEELRQHLL RWGLTTPDKK HQKEPPFLWM GYELHPDKWT VQPIVLPEKD SWTVNDIQKL VGKLNWASQI YPGIKVRQLX KLLRGTKALT EVIPLTEEAE LELAENREIL KEPVHGVYYD PSKDLIAEIQ KQGQGQWTYQ IYQEPFKNLK TGKYARMRGA HTNDVKQLTE AVQKITTESI VIWGKTPKFK LPIQKETWET WWTEYWQATW IPEWEFVNTP PLVKLWYQLE KEPIVGAETF YVDGAANRET KLGKAGYVTN RGRQKVVTLT DTTNQKTELQ AIYLALQDSG LEVNIVTDSQ YALGIIQAQP DQSESELVNQ IIEQLIKKEK VYLAWVPAHK GIGGNEQVDK LVSAGIRKVL (SEQ ID NO: 1361488) См. Martinelli *et al.*, Virology, 1990, 174(1): 135-144, который приведен в качестве ссылки

Обратная транскриптаза RT ВИЧ-1, цепь B Исх. посл. ITL3-B	PISPIETVPV KLKPGMDGPK VKQWPLTEEK IKALVEICTE MEKEGKISKI GPENPYNTPV FAIKKKDSTK WRKLVDFREL NKRTQDFWEV QLGIPHPAGL KKKKSVTVLD VGDAYFSVPL DEDFRKYTAF TIPSINNETP GIRYQYNVLP QGWKGSPAIF QSSMTKILEP FRKQNPDIVI YQYMDDLYVG SDLEIGQHRT KIEELRQHLL RWGLTTPDKK HQKEPPFLWM GYELHPDKWT VQPIVLPEKD SWTVNDIQKL VGKLNWASQI YPGIKVRQLC KLLRGTKALT EVIPLTEEAE LELAENREIL KEPVHGVYYD PSKDLIAEIQ KQGQGQWTYQ IYQEPFKNLK TGKYARMRGA HTNDVKQLTE AVQKITTESI VIWGKTPKFK LPIQKETWET WWTEYWQATW IPEWEFVNTP PLVKLWYQLE KEPIVGAETF (SEQ ID NO: 1361489) См. Stammers *et al.*, J. Mol. Biol., 1994, 242(4): 586-588, который приведен в качестве ссылки
Обратная транскриптаза RT вируса саркомы Рауса Исх. посл. ACL14945	TVALHLAIPL KWKPNHTPVW IDQWPLPEGK LVALTQLVEK ELQLGHIEPS LSCWNTPVFV IRKASGSYRL LHDLRAVNAK LVPFGAVQQG APVLSALPRG WPLMVLDLKD CFFSIPLAEQ DREAFAFTLP SVNNQAPARR FQWKVLPQGM TCSPTICQLI VGQILEPLRL KHPSLRMLHY MDDLLLAASS HDGLEAAGEE VISTLERAGF TISPDKVQKE PGVQYLGYKL GSTYAAPVGL VAEPRIATLW DVQKLVGSLQ WLRPALGIPP RLRGPFYEQL RGSDPNEARE WNLDMKMAWR EIVQLSTTAA LERWDPALPL EGAVARCEQG AIGVLGQGLS THPRPCLWLF STQPTKAFTA WLEVLTLLIT KLRASAVRTF GKEVDILLLP ACFRDELPLP EGILLALRGF AGKIRSSDTP SIFDIARPLH VSLKVRVTDH PVPGPTVFTD ASSSTHKGVV VWREGPRWEI KEIADLGASV QQLEARAVAM ALLLWPTTPT NVVTDSAFVA KMLLKMGQEG VPSTAAAFIL EDALSQRSAM AAVLHVRSHS EVPGFFTEGN DVADSQATFQ AYPLREAKDL HTALHIGPRA LSKACNISMQ QAREVVQTCP HCNSAPALEA GVNPRGLGPL QIWQTDFTLE PRMAPRSWLA VTVDTASSAI VVTQHGRVTS VAAQHHWATV IAVLGRPKAI KTDNGSCFTS KSTREWLARW GIAHTTGIPG NSQGQAMVER ANRLLKDKIR VLAEGDGFMK RIPTSKQGEL LAKAMYALNH FERGENTKTP IQKHWRPTVL TEGPPVKIRI ETGEWEKGWN VLVWGRGYAA VKNRDTDKVI WVPSRKVKPD IAQKDEVTKK DEASPLFA (SEQ ID NO: 1361490) См. Yasukawa *et al.*, J. Biochem. 2009, 145(3): 315-324, который приведен в качестве ссылки
Обратная транскриптаза RT вируса мозаики цветной капусты Исх. посл. AGT42196	MMDHLLQKTQ IQNQTEQVMN ITNPNSIYIK GRLYFKGYKK IELHCFVDTG ASLCIASKFV IPEEHWINAE RPIMVKIADG SSITINKVCR DIDLIIAGEI FHIPTVYQQE SGIDFIIGNN FCQLYEPFIQ FTDRVIFTKD RTYPVHIAKL TRAVRVGTEG FLESMKKRSK TQQPEPVNIS TNKIAILSEG RRLSEEKLFI TQQRMQKIEE LLEKVCSENP LDPNKTKQWM KASIKLSDPS KAIKVKPMKY SPMDREEFDK QIKELLDLKV IKPSKSPHMA PAFLVNNEAE KRRGKKRMVV NYKAMNKATV GDAYNLPNKD ELLTLIRGKK IFSSFDCKSG FWQVLLDQDS RPLTAFTCPQ GHYEWNVVPF GLKQAPSIFQ RHMDEAFRVF RKFCCVYVDD ILVFSNNEED HLLHVAMILQ KCNQHGIILS KKKAQLFKKK INFLGLEIDE GTHKPQGHIL EHINKFPDTL EDKKQLQRFL GILTYASDYI PKLAQIRKPL QAKLKENVPW KWTKEDTLYM QKVKKNLQGF PPLHHPLPEE KLIIETDASD DYWGGMLKAI KINEGTNTEL ICRYASGSFK AAEKNYHSND KETLAVINTI KKFSIYLTPV HFLIRTDNTH FKSFVNLNYK GDSKLGRNIR WQAWLSHYSF DVEHIKGTDN HFADFLSREF NRVNS (SEQ ID NO: 1361491) См. Farzadfar *et al.*, Virus Genes, 2013, 47(2): 347-356, который приведен в качестве ссылки
Обратная транскриптаза Клебсиелла пневмонии Исх. посл. RFF81513.1	MKEKISKIDK NFYTDIFIKT SFQNEFEAGG VIPPIAKNQV STISNKNKTF YSLAHSSPHY SIQTRIEKFL LKNIPLSASS FAFRKERSYL HYLEPHTQNV KYCHLDIVSF FHSIDVNIVR DTFSVYFSDE FLVKEKQSLL DAFMASVTLT AELDGVEKTF IPMGFKSSPS ISNIIFRKID ILIQKFCDKN KITYTRYADD LLFSTKKENN ILSSTFFINE ISSILSINKF KLNKSKYLYK EGTISLGGYV IENILKDNSS GNIRLSSSKL NPLYKALYEI KKGSSSKHIC IKVFNLKLKR FIYKKNKEKF EAKFYSSQLK NKLLGYRSYL LSFVIFHKKY KCINPIFLEK CVFLISEIES IMNRKF (SEQ ID NO: 1361492)
Обратная транскриптаза	MKITSNNVTA VINGKGWHSI NWKKCHQHVK TIQTRIAKAA CNQQWRTVGR LQRLLVRSFS ARALAVKRVT ENSGRKTPGV DGQIWSTPES KWEAIFKLRR KGYKPLPLKR VFIPKSNGKK RPLGIPVMLD RAMQALHLLG LEPVSETNAD HNSYGFRPAR CTADAIQQVC NMYSSRNASK

RT Escerichia coli Исх. посл. TGH57013	WVLEGDIKGC FEHISHEWLL ENIPMDKQIL RNWLKAGIIE KSIFSKTLSG TPQGGIISPV LANMALDGLE RLLQNRFGRN RLI (SEQ ID NO: 1361493)
Обратная транскриптаза RT *Bacillus subtilis* Исх. посл. QBJ66766	MSKIKINYEK YHIKPFPHFD QRIKVNKKVK ENLQNPFYIA AHSFYPFIHY KKISYKFKNG TLSSPKERDI FYSGHMDGYI YKHYGEILNH KYNNTCIGKG IDHVSLAYRN NKMGKSNIHF AAEVINFISE QQQAFIFVSD FSSYFDSLDH AILKEKLIEV LEEQDKLSKD WWNVFKHITR YNWVEKEEVI SDLECTKEKI ARDKKSRERY YTPAEFREFR KRVNIKSNDT GVGIPQGTAI SAVLANVYAI DLDQKLNQYA LKYGGIYRRY SDDIIMVLPM TSDGQDPSND HVSFIKSVVK RNKVTMGDSK TSVLYYANNN IYEDYQRKRE SKMDYLGFSF DGMTVKIREK SLFKYYHRTY KKINSINWAS VKKEKKVGRK KLYLLYSHLG RNYKGHGNFI SYCKKAHAVF EGNKKIESLI NQQIKRHWKK IQKRLVDV (SEQ ID NO: 1361494)
RT интрона группы II *Eubacterium rectale*	DTSNLMEQILSSDNLNRAYLQVVRNKGAEGVDGMKYTELKEHLAKNGETIKGQLRTRKYKPQPAR RVEIPKPDGGVRNLGVPTVTDRFIQQAIAQVLTPIYEEQFHDHSYGFRPNRCAQQAILTALNIMN DGNDWIVDIDLEKFFDTVNHDKLMTLIGRTIKDGDVISIVRKYLVSGIMIDDEYEDSIVGTPQGG NLSPLLANIMLNELDKEMEKRGLNFVRYADDCIIMVGSEMSANRVMRNISRFIEEKLGLKVNMTK SKVDRPSGLKYLGFGFYFDPRAHQFKAKPHAKSVAKFKKRMKELTCRSWGVSNSYKVEKLNQLIR GWINYFKIGSMKTLCKELDSRIRYRLRMCIWKQWKTPQNQEKNLVKLGIDRNTARRVAYTGKRIA YVCNKGAVNVAISNKRLASFGLISMLDYYIEKCVTC (SEQ ID NO: 1361495)
RT интрона группы II *Geobacillus stearothermophilus*	ALLERILARDNLITALKRVEANQGAPGIDGVSTDQLRDYIRAHWSTIHAQLLAGTYRPAPVRRVE IPKPGGGTRQLGIPTVVDRLIQQAILQELTPIFDPDFSSSSFGFRPGRNAHDAVRQAQGYIQEGY RYVVDMDLEKFFDRVNHDILMSRVARKVKDKRVLKLIRAYLQAGVMIEGVKVQTEEGTPQGGPLS PLLANILLDDLDKELEKRGLKFCRYADDCNIYVKSLRAGQRVKQSIQRFLEKTLKLKVNEEKSAV DRPWKRAFLGFSFTPERKARIRLAPRSIQRLKQRIRQLTNPNWSISMPERIHRVNQYVMGWIGYF RLVETPSVLQTIEGWIRRRLRLCQWLQWKRVRTRIRELRALGLKETAVMEIANTRKGAWRTTKTP QLHQALGKTYWTAQGLKSLTQR (SEQ ID NO: 1361496)

(b) Варианты RT

[0475] В различных вариантах осуществления обратная транскриптаза может представлять собой вариант обратной транскриптазы. В соответствии с используемым здесь значением термин "вариант обратной транскриптазы" включает любой встречающийся в природе или генетически сконструированный вариант, включающий одну или несколько мутаций (включая единичные мутации, инверсии, делеции, вставки, и перестроения) относительно исходных последовательностей (например исходной последовательности дикого типа). RT естественным образом имеет несколько видов активности, включая активность РНК-зависимой ДНК-полимеразы, активность рибонуклеазы H, и активность ДНК-зависимой ДНК-полимеразы. Данные виды активности совместно обеспечивают возможность фермента конвертировать одноцепочечную РНК в двуцепочечную кДНК. В ретровирусах и ретротранспозонах

данная кДНК затем может встраиваться в геном хозяина, из которого могут быть получены новые копии РНК за счет транскрипции в клетке-хозяине. Варианты RT могут включать мутацию, которая влияет на один или несколько данных видов активности (которая снижает или повышает данные виды активности или которая полностью устраняет данные виды активности). Кроме того, варианты RT могут включать одну или несколько мутаций, которые делают RT более или менее стабильной, более склонной к агрегации, и обеспечивают очистку и/или детектирование, и/или другое модифицирование свойств или характеристик.

[0476] Среднему специалисту в данной области будет очевидно, что варианты обратных транскриптаз, являющиеся производными из других обратных транскриптаз, включая, но без ограничения, обратную транскриптазу вируса лейкемии мышей Молони (M-MLV), обратную транскриптазу вируса иммунодефицита человека (ВИЧ) и обратную транскриптазу вируса саркомы-лейкоза птиц (ASLV), которая включает, но без ограничения, обратную транскриптазу вируса саркомы Рауса (RSV), обратную транскриптазу вируса миелобластоза птиц (AMV), обратную транскриптазу хелперного вируса MCAV вируса эритробластоза птиц (AEV), обратную транскриптазу хелперного вируса MCAV вируса миелоцитоматоза птиц MC29, обратную транскриптазу хелперного вируса REV-A вируса ретикулоэндотелиоза птиц (REV-T), обратную транскриптазу хелперного вируса UR2AV вируса саркомы птиц UR2, обратную транскриптазу хелперного вируса YAV вируса саркомы птиц Y73, обратную транскриптазу Раус-ассоциированного вируса (RAV) и обратную транскриптазу миелобластоз-ассоциированного вируса (MAV), могут подходящим образом использоваться в рамках настоящих описанных здесь способов и композиций.

[0477] Один способ получения вариантов RT заключается в генетическом модифицировании (например путем модифицирования последовательности ДНК обратной транскриптазы дикого типа). В данной области известен ряд способов, которые обеспечивают случайное, а также направленное мутирование последовательностей ДНК (см., например, Ausubel et. al. *Short Protocols in Molecular Biology (1995) 3.sup.rd Ed.* John Wiley & Sons, Inc.). Кроме того, существует ряд коммерчески доступных наборов для сайт-направленного мутагенеза, включая как стандартные, так и основанные на ПЦР способы. Примеры включают наборы для сайт-направленного мутагенеза QuikChange

(AGILENT®), набор для сайт-направленного мутагенеза Q5® (NEW ENGLAND BIOLABS®), и систему для сайт-направленного мутагенеза GeneArt™ (THERMOFISHER SCIENTIFIC®).

[0478] Кроме того, мутантные обратные транскриптазы могут быть получены с помощью мутации в виде вставки или усечения (N-концевых, внутренних, или С-концевых вставок или усечений) в соответствии со способами, известными специалисту в данной области. В соответствии с используемым здесь значением термин "мутация" обозначает замещение остатка в последовательности, например нуклеиновой кислоты или аминокислотной последовательности, другим остатком, или делецию или вставку одного или нескольких остатков в последовательности. Мутации обычно описаны здесь на основе указания исходного остатка и далее положения остатка в последовательности и на основе указания нового замещенного остатка. Различные способы осуществления указанных здесь аминокислотных замещений (мутаций) хорошо известны в данной области и описаны, например, в Green and Sambrook, Molecular Cloning: A Laboratory Manual (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)). Мутации могут включать различные категории, например полиморфизмы одиночных оснований, участки микродупликации, индели и инверсии, и не являются ограниченными каким-либо образом. Мутации могут включать мутации типа "потеря функции", что является нормальным результатом мутации, которая снижает или устраняет активность белка. Большинство мутаций потери функции являются рецессивными, поскольку в гетерозиготе вторая копия хромосомы несет не подверженную мутации версию гена, кодирующего полностью функциональный пептид, чье присутствие компенсирует эффект мутации. Мутации также охватывают мутации типа "приобретение функции", которые придают нехарактерную активность белку или клетке, которая иным образом не присутствует при нормальных условиях. Многие мутации, связанные с приобретением функции, располагаются в регуляторных последовательностях, а не в кодирующих участках, и, таким образом, имеют ряд последствий. Например, мутация может приводить к тому, что один или несколько генов экспрессируются в неправильных тканях, при этом эти ткани приобретают функции, которые они обычно не имеют. Вследствие их природы, мутации, связанные с приобретением функции, обычно являются доминантными.

[0479] Старые известные в данной области способы сайт-направленного мутагенеза основаны на субклонировании подвергаемой мутированию последовательности в вектор, например вектор на основе бактериофага M13, который обеспечивает изолирование одноцепочечной матрицы ДНК. В рамках данных способов осуществляется отжиг мутагенного праймера (то есть праймера, способного отжигаться на подвергаемом мутированию участке, но несущего один или несколько несовпадающих нуклеотидов в подвергаемом мутированию участке) на одноцепочечной матрице и затем полимеризация комплемента матрицы начиная с 3'-конца мутагенного праймера. Получаемые в результате этого дуплексы затем переносятся в бактерии-хозяева и осуществляется скрининг бляшек на предмет требуемой мутации.

[0480] Недавно в рамках сайт-направленного мутагенеза стали использоваться способы ПЦР, которые имеют преимущество, заключающееся в отсутствии потребности в одноцепочечной матрице. Кроме того, были разработаны способы, не требующие субклонирования. При осуществлении сайт-направленного мутагенеза на основе ПЦР необходимо учитывать несколько нюансов. Во-первых, в рамках этих способов является желательным снижать количество циклов ПЦР для предотвращения экспансии нежелательных мутаций, вносимых полимеразой. Во-вторых, должен осуществляться отбор для снижения количества не подвергнутых мутированию исходных молекул, остающихся в реакции. В-третьих, ПЦР способ с повышенной длиной является предпочтительным для обеспечения возможности использования одного набора праймеров ПЦР. И в-четвертых, вследствие активности не зависящего от матрицы концевого удлинения некоторых термостабильных полимераз зачастую является необходимым включать конечный финишный этап в процедуру перед лигированием тупого конца полученного с помощью ПЦР мутантного продукта.

[0481] В данной области существуют способы случайного мутагенеза, которые будут приводить к получению группы мутантов, несущих одну или несколько случайно расположенных мутаций. Такая группа мутантов затем может быть подвергнута скринингу на предмет тех из них, которые демонстрируют требуемые свойства, например повышенную стабильность, по сравнению с обратной транскриптазой дикого типа.

[0482] Пример способа случайного мутагенеза представляет собой так называемый "способ на основе подверженной ошибкам ПЦР". Как следует из названия, способ

обеспечивает амплификацию заданной последовательности в условиях, при которых ДНК-полимераза не поддерживает встраивание с высокой точностью. Несмотря на то, что условия, стимулирующие подверженное ошибкам встраивание, варьируются для различных ДНК-полимераз, специалист в данной области может определить такие условия для заданного фермента. Ключевая переменная для многих ДНК-полимераз в плане точности амплификации представляет собой, например, тип и концентрацию двухвалентного иона металла в буфере. Применение иона марганца и/или варьирование концентрации иона магния или марганца может, таким образом, применяться для управления величиной ошибок полимеразы.

[0483] В рамках различных аспектов RT праймерных редакторов может представлять собой вариант "подверженной ошибкам" обратной транскриптазы. Могут применяться подверженные ошибкам обратные транскриптазы, которые известны и/или доступны в данной области. Кроме того, RT может быть получена с помощью любого из ранее указанных способов мутагенеза включая процессы направленной эволюции, такие как непрерывная эволюция с помощью фагов (PACE) или прерывистая эволюция с помощью фагов (PANCE). В соответствии с используемым здесь значением термин "непрерывная эволюция с помощью фагов (PACE)" обозначает непрерывную эволюцию, в рамках которой используется фаг в качестве вирусных векторов. Общая концепция технологии PACE была описана, например, в международной PCT заявке PCT/US2009/056194, поданной 8 сентября 2009 г., опубликованной как WO 2010/028347 11 марта 2010 г.; международной PCT заявке PCT/US2011/066747, поданной 22 декабря 2011 г., опубликованной как WO 2012/088381 28 июня 2012 г.; заявке на патент США, патенте США 9023594, выданном 5 мая 2015 г., международной PCT заявке PCT/US2015/012022, поданной 20 января 2015 г., опубликованной как WO 2015/134121 11 сентября 2015 г., и международной PCT заявке PCT/US2016/027795, поданной 15 апреля 2016 г., опубликованной как WO 2016/168631 20 октября 2016 г., полное содержание каждой из которых приведено здесь в качестве ссылки. Подверженные ошибкам обратные транскриптазы также могут быть получены с помощью "прерывистой эволюции с помощью фагов (PANCE)", которая в соответствии с используемым здесь значением обозначает прерывистую эволюцию, в рамках которой используется фаг в качестве вирусных векторов. PANCE представляет собой упрощенный способ быстрой *in vivo*

направленной эволюции с применением последовательных переносов во флаконах подвергаемого эволюции "фага отбора" (SP), который содержит представляющий интерес подвергаемый эволюции ген, между свежими клетками-хозяевами *E. coli* тем самым обеспечивая сохранение генов внутри *E. coli* неизменными, в то время как гены, содержащиеся в SP, непрерывно подвергаются эволюции. Последовательные переносы во флаконах долго служили в качестве широко доступного подхода для лабораторной эволюции микробов, при этом в недавнее время были разработаны аналогичные подходы для эволюции бактериофагов. Системе PANCE присуща более низкая строгость, чем системе PACE.

[0484] Гены для требуемых мутантных обратных транскриптаз, полученных с помощью мутагенеза или эволюционных процессов, могут быть подвергнуты секвенированию для идентификации участков и количеств мутаций. Для мутантов, которые включают более чем одну мутацию, эффект определенной мутации может быть проанализирован путем внесения идентифицированной мутации в ген дикого типа с помощью сайт-направленного мутагенеза с изолированием от других мутаций, которые несет конкретный мутант. Затем скрининговые анализы одного полученного таким образом мутанта обеспечат определение эффекта данной мутации в отдельном виде.

[0485] Используемые здесь ферменты в виде вариантов RT могут также включать другие "варианты RT", являющиеся, по меньшей мере, приблизительно на 70% идентичными, по меньшей мере, приблизительно на 80% идентичными, по меньшей мере, приблизительно на 90% идентичными, по меньшей мере, приблизительно на 95% идентичными, по меньшей мере, приблизительно на 96% идентичными, по меньшей мере, приблизительно на 97% идентичными, по меньшей мере, приблизительно на 98% идентичными, по меньшей мере, приблизительно на 99% идентичными, по меньшей мере, приблизительно на 99,5% идентичными, или, по меньшей мере, приблизительно на 99,9% идентичными любому исходному белку RT, включая RT дикого типа, или мутантную RT, или фрагмент RT, или другой вариант RT, описанный или охватываемый здесь или известный в данной области.

[0486] В некоторых вариантах осуществления вариант RT может иметь 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, или до 100, или до 200, или до 300,

или до 400, или до 500 или больше аминокислотных изменений по сравнению с исходной RT. В некоторых вариантах осуществления вариант RT включает фрагмент исходной RT, так что фрагмент является, по меньшей мере, приблизительно на 70% идентичным, по меньшей мере, приблизительно на 80% идентичным, по меньшей мере, приблизительно на 90% идентичным, по меньшей мере, приблизительно на 95% идентичным, по меньшей мере, приблизительно на 96% идентичным, по меньшей мере, приблизительно на 97% идентичным, по меньшей мере, приблизительно на 98% идентичным, по меньшей мере, приблизительно на 99% идентичным, по меньшей мере, приблизительно на 99,5% идентичным, по меньшей мере, приблизительно на 99,9% идентичным соответствующему фрагменту исходной RT. В некоторых вариантах осуществления фрагмент имеет длину, соответствующую, по меньшей мере, на 30%, по меньшей мере, на 35%, по меньшей мере, на 40%, по меньшей мере, на 45%, по меньшей мере, на 50%, по меньшей мере, на 55%, по меньшей мере, на 60%, по меньшей мере, на 65%, по меньшей мере, на 70%, по меньшей мере, на 75%, по меньшей мере, на 80%, по меньшей мере, на 85%, по меньшей мере, на 90%, по меньшей мере, на 95%, по меньшей мере, на 96%, по меньшей мере, на 97%, по меньшей мере, на 98%, по меньшей мере, на 99%, или, по меньшей мере, на 99,5% длине аминокислоты соответствующей RT дикого типа (например SEQ ID NO: 1361485).

[0487] В некоторых вариантах осуществления в изобретении также могут использоваться фрагменты RT, которые сохраняют свою функциональность и которые представляют собой фрагменты любых описанных здесь белков RT. В некоторых вариантах осуществления фрагмент RT имеет длину, по меньшей мере, 100 аминокислот. В некоторых вариантах осуществления фрагмент имеет длину в виде, по меньшей мере 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, или до 600 или более аминокислот.

[0488] В других вариантах осуществления в изобретении также могут использоваться варианты RT, которые являются усеченными на N-конце или C-конце или на них обоих на определенное количество аминокислот, что приводит к образованию усеченного варианта, который все же сохраняет достаточную полимеразную функцию. В некоторых вариантах осуществления усеченный вариант RT имеет усечение длиной, по меньшей мере, 1, по меньшей мере, 2, по меньшей мере, 3, по меньшей мере, 4, по меньшей мере, 5, по меньшей мере, 6, по меньшей мере, 7, по меньшей мере, 8, по меньшей мере, 9, по меньшей мере, 10, по меньшей мере, 11, по меньшей мере, 12, по меньшей мере, 13, по

меньшей мере, 14, по меньшей мере, 15, по меньшей мере, 16, по меньшей мере, 17, по меньшей мере, 18, по меньшей мере, 19, по меньшей мере, 20, по меньшей мере, 21, по меньшей мере, 22, по меньшей мере, 23, по меньшей мере, 24, по меньшей мере, 25, по меньшей мере, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, или 250 аминокислот на N-конце белка. В других вариантах осуществления усеченный вариант RT имеет усечение длиной, по меньшей мере, 1, по меньшей мере, 2, по меньшей мере, 3, по меньшей мере, 4, по меньшей мере, 5, по меньшей мере, 6, по меньшей мере, 7, по меньшей мере, 8, по меньшей мере, 9, по меньшей мере, 10, по меньшей мере, 11, по меньшей мере, 12, по меньшей мере, 13, по меньшей мере, 14, по меньшей мере, 15, по меньшей мере, 16, по меньшей мере, 17, по меньшей мере, 18, по меньшей мере, 19, по меньшей мере, 20, по меньшей мере, 21, по меньшей мере, 22, по меньшей мере, 23, по меньшей мере, 24, по меньшей мере, 25, по меньшей мере, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, или 250 аминокислот на C-конце белка. В других вариантах осуществления усеченный вариант RT имеет усечения на N-конце и C-конце, которые имеют одинаковую или различающуюся длину.

[0489] Например, описанные здесь праймерные редакторы могут включать усеченную версию обратной транскриптазы M-MLV. В данном варианте осуществления обратная транскриптаза содержит 4 мутации (D200N, T306K, W313F, T330P; следует отметить, что мутация L603W, присутствующая в PE2, является отсутствующей вследствие усечения). Последовательность ДНК, кодирующая данный усеченный редактор, на 522 bp короче, чем PE2, и, таким образом, является потенциально полезной для вариантов использования, в рамках которых доставка последовательности ДНК является сложной вследствие ее размера (то есть при доставке на основе адено-ассоциированного вируса и лентивируса). Данный вариант осуществления обозначается как MMLV-RT(trunc) и имеет следующую аминокислотную последовательность:

mmlv-rt(TRUNC)	SGGSSGGSSGSETPGTSESATPESSGGSSGGSS*TLNIEDEYRLHETSKEPDVSLGST WLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRL LDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSG LPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNS PTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGF CRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPF ELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDA GKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNP*

	ATLLPLPEEGLQHNCLDNSRLIN (SEQ ID NO: 1361598) Обозначения Линкер: (SEQ ID NO: 1361528) RT (TRUNC): (SEQ ID NO: 1361597)

[0490] В различных вариантах осуществления описанные здесь праймерные редакторы могут включать один из вариантов Cas9, описанных далее, или его вариант в виде RT, который является, по меньшей мере, приблизительно на 70% идентичным, по меньшей мере, приблизительно на 80% идентичным, по меньшей мере, приблизительно на 90% идентичным, по меньшей мере, приблизительно на 95% идентичным, по меньшей мере, приблизительно на 96% идентичным, по меньшей мере, приблизительно на 97% идентичным, по меньшей мере, приблизительно на 98% идентичным, по меньшей мере, приблизительно на 99% идентичным, по меньшей мере, приблизительно на 99,5% идентичным, или, по меньшей мере, приблизительно на 99,9% идентичным любым исходным вариантам RT.

[0491] В литературе были описаны другие подверженные ошибкам обратные транскриптазы, каждая из которых предусмотрена для применения в рамках описанных здесь способов и композиций. Например, подверженные ошибкам обратные транскриптазы были описаны в Bebenek et al., "Error-prone Polymerization by HIV-1 Reverse Transcriptase," *J Biol Chem*, 1993, Vol. 268: 10324-10334 и Sebastian-Martin et al., "Transcriptional inaccuracy threshold attenuates differences in RNA-dependent DNA synthesis fidelity between retroviral reverse transcriptases," *Scientific Reports*, 2018, Vol. 8: 627, каждый из которых приведен здесь в качестве ссылки. Кроме того, обратные транскриптазы, включая подверженные ошибкам обратные транскриптазы, могут быть получены от коммерческого поставщика, включая обратную транскриптазу ProtoScript® (II), обратную транскриптазу AMV, обратную транскриптазу WarmStart®, и обратную транскриптазу M-MuLV, все от NEW ENGLAND BIOLABS®, или обратную транскриптазу AMV XL, обратную транскриптазу SMARTScribe, ультра-чистую обратную транскриптазу MMLV GPR, все от TAKARA BIO USA, INC. (бывш. CLONTECH).

[0492] В других вариантах осуществления в настоящих способах и композициях может применяться ДНК-полимераза, которая была подвергнута эволюции в обратную транскриптазу, как описано в Effefson *et al.*, "Synthetic evolutionary origin of a proofreading

reverse transcriptase," *Science*, June 24, 2016, Vol. 352: 1590-1593, содержание которого приведено здесь в качестве ссылки.

[0493] В определенных других вариантах осуществления обратная транскриптаза предусмотрена в виде компонента слитого белка, который также включает napDNAbp. Другими словами, в некоторых вариантах осуществления обратная транскриптаза слита с napDNAbp в виде слитого белка.

[0494] Некоторые примеры обратных транскриптаз, которые могут быть слиты с белками napDNAbp или поступать в виде отдельных белков в соответствии с различными вариантами осуществления данного изобретения, приведены ниже. Примеры обратных транскриптаз включают варианты, по меньшей мере, с 80%, по меньшей мере, с 85%, по меньшей мере, с 90%, по меньшей мере, с 95%, по меньшей мере, с 99% идентичностью последовательности относительно следующих ферментов дикого типа или частичных ферментов:

Описание	Последовательность (вариантные замещения относительно дикого типа)
Обратная транскриптаза (RT M-MLV) дикого типа вирус лейкоза мышей Молони Используется в PE1 (описанный здесь слитый белок праймерного редактора 1)	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIP GFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAK GVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVE ALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHG TRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALT QALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRL SIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 1361497)
RT M-MLV D200N	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIP GFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAK GVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVE ALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHG TRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALT QALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRL SIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 1361498)
RT M-MLV D200N T330P	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGY

Описание	Последовательность (вариантные замещения относительно дикого типа)
	RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIP GFAEMAAPLYPLTK**P**GTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAK GVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVE ALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHG TRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALT QALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRL SIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 1361499)
RT M-MLV D200N T330P L603W	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIP GFAEMAAPLYPLTK**P**GTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAK GVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVE ALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHG TRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALT QALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKRL SIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 1361500)
RT M-MLV D200N T330P L603W E69K	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQ**K**ARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIP GFAEMAAPLYPLTK**P**GTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAK GVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVE ALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHG TRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALT QALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKRL SIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 1361501)
RT M-MLV D200N T330P L603W E302R	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLR**R**FLGTAGFCRLWIP GFAEMAAPLYPLTK**P**GTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAK GVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVE ALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHG TRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALT QALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKRL SIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP(SEQ ID NO: 1361502)
RT M-MLV D200N T330P L603W E607K	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIP GFAEMAAPLYPLTK**P**GTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAK GVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVE ALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHG TRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALT

Описание	Последовательность (вариантные замещения относительно дикого типа)
	QALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTS**K**GKEIKNKDEILALLKALFLPKRL SIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 1361503)
RT M-MLV D200N T330P L603W L139P	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSG**P**PPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIP GFAEMAAPLYPLTK**P**GTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAK GVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVE ALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHG TRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALT QALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKRL SIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 1361504)
RT M-MLV D200N T330P L603W L435G	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIP GFAEMAAPLYPLTK**P**GTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAK GVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVI**G**APHAVE ALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHG TRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALT QALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKRL SIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 1361505)
RT M-MLV D200N T330P L603W N454K	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIP GFAEMAAPLYPLTK**P**GTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAK GVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVE ALVKQPPDRWLS**K**ARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHG TRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALT QALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKRL SIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 1361506)
RT M-MLV D200N T330P L603W T306K	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLG**K**AGFCRLWIP GFAEMAAPLYPLTK**P**GTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAK GVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVE ALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHG TRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALT QALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKRL SIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 1361507)
RT M-MLV D200N T330P L603W	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP

Описание	Последовательность (вариантные замещения относительно дикого типа)
W313F	QGFKNSPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRL**F**IP GFAEMAAPLYPLTK**P**GTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAK GVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVE ALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHG TRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALT QALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKRL SIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 1361508)
RT M-MLV D200N T330P L603W D524G E562Q D583N	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIP GFAEMAAPLYPLTK**P**GTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAK GVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVE ALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHG TRPDLTDQPLPDADHTWYT**G**GSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRA**Q**LIALT QALKMAEGKKLNVYT**N**SRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKRL SIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 1361509)
RT M-MLV D200N T330P L603W E302R W313F	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLR**R**FLGTAGFCRL**F**IP GFAEMAAPLYPLTK**P**GTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAK GVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVE ALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHG TRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALT QALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKRL SIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 1361510)
RT M-MLV D200N T330P L603W E607K L139P	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSG**P**PPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIP GFAEMAAPLYPLTK**P**GTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAK GVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVE ALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHG TRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALT QALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTS**K**GKEIKNKDEILALLKALFLPKRL SIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 1361511)
RT M-MLV P51L S67K T197A H204R E302K F309N W313F T330P L435G N454K D524G D583N H594Q D653N	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLII**L**LKATSTPVSIKQ YPM**K**QEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSP**A**LFDEAL**R**RDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLR**K**FLGTAG**N**CRL**F**IP GFAEMAAPLYPLTK**P**GTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAK GVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVI**G**APHAVE ALVKQPPDRWLS**K**ARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHG TRPDLTDQPLPDADHTWYT**G**GSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALT

Описание	Последовательность (вариантные замещения относительно дикого типа)
	QALKMAEGKKLNVYT**N**SRYAFATAHI**Q**GEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRL SIIHCPGHQKGHSAEARGNRMA**N**QAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 1361512)
RT M-MLV D200N P51L S67K T197A H204R E302K F309N W313F T330P L345G N454K D524G D583N H594Q D653N	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLII**L**LKATSTPVSIKQ YPM**K**QEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSP**A**LF**N**EAL**R**RDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLR**K**FLGTAG**N**CRL**F**IP GFAEMAAPLYPLTK**P**GTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAK GVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVI**G**APHAVE ALVKQPPDRWLS**K**ARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHG TRPDLTDQPLPDADHTWYT**G**GSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALT QALKMAEGKKLNVYT**N**SRYAFATAHI**Q**GEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRL SIIHCPGHQKGHSAEARGNRMA**N**QAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 1361513)
RT M-MLV D200N T330P L603W T306K W313F в PE2	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLG**K**AGFCRL**F**IP GFAEMAAPLYPLTK**P**GTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAK GVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVE ALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHG TRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALT QALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKRL SIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 1361514)

[0495] В различных вариантах осуществления описанные здесь праймерные редакторы (с RT, поступающей в виде слитого партнера или *in trans*) могут включать вариант RT, включающий одну или несколько из следующих мутаций: P P51L, S67K, E69K, L139P, T197A, D200N, H204R, F209N, E302K, E302R, T306K, F309N, W313F, T330P, L345G, L435G, N454K, D524G, E562Q, D583N, H594Q, L603W, E607K, или D653N в RT M-MLV дикого типа согласно SEQ ID NO: 1361485 или в соответствующем аминокислотном положении в другой последовательности полипептида RT дикого типа.

[0496] В различных других вариантах осуществления описанные здесь праймерные редакторы (с RT, поступающей в виде слитого партнера или *in trans*) могут включать вариант RT, включающий одну или несколько из следующих мутаций: P51X, S67X, E69X, L139X, T197X, D200X, H204X, F209X, E302X, T306X, F309X, W313X, T330X, L345X, L435X, N454X, D524X, E562X, D583X, H594X, L603X, E607X, или D653X в RT M-MLV дикого типа согласно SEQ ID NO: 1361485 или в соответствующем аминокислотном положении в другой последовательности полипептида RT дикого типа, где "X" может представлять собой любую аминокислоту.

[0497] В различных других вариантах осуществления описанные здесь праймерные редакторы (с RT, поступающей в виде слитого партнера или *in trans*) могут включать вариант RT, включающий мутацию P51X в RT M-MLV дикого типа согласно SEQ ID NO: 1361485 или в соответствующем аминокислотном положении в другой последовательности полипептида RT дикого типа, где "X" может представлять собой любую аминокислоту. В определенных вариантах осуществления X представляет собой L.

[0498] В различных других вариантах осуществления описанные здесь праймерные редакторы (с RT, поступающей в виде слитого партнера или *in trans*) могут включать вариант RT, включающий мутацию S67X в RT M-MLV дикого типа согласно SEQ ID NO: 1361485 или в соответствующем аминокислотном положении в другой последовательности полипептида RT дикого типа, где "X" может представлять собой любую аминокислоту. В некоторых вариантах осуществления X представляет собой K.

[0499] В различных других вариантах осуществления описанные здесь праймерные редакторы (с RT, поступающей в виде слитого партнера или *in trans*) могут включать вариант RT, включающий мутацию E69X в RT M-MLV дикого типа согласно SEQ ID NO: 1361485 или в соответствующем аминокислотном положении в другой последовательности полипептида RT дикого типа, где "X" может представлять собой любую аминокислоту. В некоторых вариантах осуществления X представляет собой K.

[0500] В различных других вариантах осуществления описанные здесь праймерные редакторы (с RT, поступающей в виде слитого партнера или *in trans*) могут включать вариант RT, включающий мутацию L139X в RT M-MLV дикого типа согласно SEQ ID NO: 1361485 или в соответствующем аминокислотном положении в другой последовательности полипептида RT дикого типа, где "X" может представлять собой любую аминокислоту. В некоторых вариантах осуществления X представляет собой P.

[0501] В различных других вариантах осуществления описанные здесь праймерные редакторы (с RT, поступающей в виде слитого партнера или *in trans*) могут включать вариант RT, включающий мутацию T197X в RT M-MLV дикого типа согласно SEQ ID NO: 1361485 или в соответствующем аминокислотном положении в другой последовательности полипептида RT дикого типа, где "X" может представлять собой любую аминокислоту. В некоторых вариантах осуществления X представляет собой A.

[0502] В различных других вариантах осуществления описанные здесь праймерные редакторы (с RT, поступающей в виде слитого партнера или *in trans*) могут включать вариант RT, включающий мутацию D200X в RT M-MLV дикого типа согласно SEQ ID NO: 1361485 или в соответствующем аминокислотном положении в другой последовательности полипептида RT дикого типа, где "X" может представлять собой любую аминокислоту. В некоторых вариантах осуществления X представляет собой N.

[0503] В различных других вариантах осуществления описанные здесь праймерные редакторы (с RT, поступающей в виде слитого партнера или *in trans*) могут включать вариант RT, включающий мутацию H204X в RT M-MLV дикого типа согласно SEQ ID NO: 1361485 или в соответствующем аминокислотном положении в другой последовательности полипептида RT дикого типа, где "X" может представлять собой любую аминокислоту. В некоторых вариантах осуществления X представляет собой R.

[0504] В различных других вариантах осуществления описанные здесь праймерные редакторы (с RT, поступающей в виде слитого партнера или *in trans*) могут включать вариант RT, включающий мутацию F209X в RT M-MLV дикого типа согласно SEQ ID NO: 1361485 или в соответствующем аминокислотном положении в другой последовательности полипептида RT дикого типа, где "X" может представлять собой любую аминокислоту. В некоторых вариантах осуществления X представляет собой N.

[0505] В различных других вариантах осуществления описанные здесь праймерные редакторы (с RT, поступающей в виде слитого партнера или *in trans*) могут включать вариант RT, включающий мутацию E302X в RT M-MLV дикого типа согласно SEQ ID NO: 1361485 или в соответствующем аминокислотном положении в другой последовательности полипептида RT дикого типа, где "X" может представлять собой любую аминокислоту. В некоторых вариантах осуществления X представляет собой K.

[0506] В различных других вариантах осуществления описанные здесь праймерные редакторы (с RT, поступающей в виде слитого партнера или *in trans*) могут включать вариант RT, включающий мутацию E302X в RT M-MLV дикого типа согласно SEQ ID NO: 1361485 или в соответствующем аминокислотном положении в другой последовательности полипептида RT дикого типа, где "X" может представлять собой любую аминокислоту. В некоторых вариантах осуществления X представляет собой R.

[0507] В различных других вариантах осуществления описанные здесь праймерные редакторы (с RT, поступающей в виде слитого партнера или *in trans*) могут включать вариант RT, включающий мутацию T306X в RT M-MLV дикого типа согласно SEQ ID NO: 1361485 или в соответствующем аминокислотном положении в другой последовательности полипептида RT дикого типа, где "X" может представлять собой любую аминокислоту. В некоторых вариантах осуществления X представляет собой K.

[0508] В различных других вариантах осуществления описанные здесь праймерные редакторы (с RT, поступающей в виде слитого партнера или *in trans*) могут включать вариант RT, включающий мутацию F309X в RT M-MLV дикого типа согласно SEQ ID NO: 1361485 или в соответствующем аминокислотном положении в другой последовательности полипептида RT дикого типа, где "X" может представлять собой любую аминокислоту. В некоторых вариантах осуществления X представляет собой N.

[0509] В различных других вариантах осуществления описанные здесь праймерные редакторы (с RT, поступающей в виде слитого партнера или *in trans*) могут включать вариант RT, включающий мутацию W313X в RT M-MLV дикого типа согласно SEQ ID NO: 1361485 или в соответствующем аминокислотном положении в другой последовательности полипептида RT дикого типа, где "X" может представлять собой любую аминокислоту. В некоторых вариантах осуществления X представляет собой F.

[0510] В различных других вариантах осуществления описанные здесь праймерные редакторы (с RT, поступающей в виде слитого партнера или *in trans*) могут включать вариант RT, включающий мутацию T330X в RT M-MLV дикого типа согласно SEQ ID NO: 1361485 или в соответствующем аминокислотном положении в другой последовательности полипептида RT дикого типа, где "X" может представлять собой любую аминокислоту. В некоторых вариантах осуществления X представляет собой P.

[0511] В различных других вариантах осуществления описанные здесь праймерные редакторы (с RT, поступающей в виде слитого партнера или *in trans*) могут включать вариант RT, включающий мутацию L345X в RT M-MLV дикого типа согласно SEQ ID NO: 1361485 или в соответствующем аминокислотном положении в другой последовательности полипептида RT дикого типа, где "X" может представлять собой любую аминокислоту. В некоторых вариантах осуществления X представляет собой G.

[0512] В различных других вариантах осуществления описанные здесь праймерные редакторы (с RT, поступающей в виде слитого партнера или *in trans*) могут включать вариант RT, включающий мутацию L435X в RT M-MLV дикого типа согласно SEQ ID NO: 1361485 или в соответствующем аминокислотном положении в другой последовательности полипептида RT дикого типа, где "X" может представлять собой любую аминокислоту. В некоторых вариантах осуществления X представляет собой G.

[0513] В различных других вариантах осуществления описанные здесь праймерные редакторы (с RT, поступающей в виде слитого партнера или *in trans*) могут включать вариант RT, включающий мутацию N454X в RT M-MLV дикого типа согласно SEQ ID NO: 1361485 или в соответствующем аминокислотном положении в другой последовательности полипептида RT дикого типа, где "X" может представлять собой любую аминокислоту. В некоторых вариантах осуществления X представляет собой K.

[0514] В различных других вариантах осуществления описанные здесь праймерные редакторы (с RT, поступающей в виде слитого партнера или *in trans*) могут включать вариант RT, включающий мутацию D524X в RT M-MLV дикого типа согласно SEQ ID NO: 1361485 или в соответствующем аминокислотном положении в другой последовательности полипептида RT дикого типа, где "X" может представлять собой любую аминокислоту. В некоторых вариантах осуществления X представляет собой G.

[0515] В различных других вариантах осуществления описанные здесь праймерные редакторы (с RT, поступающей в виде слитого партнера или *in trans*) могут включать вариант RT, включающий мутацию E562X в RT M-MLV дикого типа согласно SEQ ID NO: 1361485 или в соответствующем аминокислотном положении в другой последовательности полипептида RT дикого типа, где "X" может представлять собой любую аминокислоту. В некоторых вариантах осуществления X представляет собой Q.

[0516] В различных других вариантах осуществления описанные здесь праймерные редакторы (с RT, поступающей в виде слитого партнера или *in trans*) могут включать вариант RT, включающий мутацию D583X в RT M-MLV дикого типа согласно SEQ ID NO: 1361485 или в соответствующем аминокислотном положении в другой последовательности полипептида RT дикого типа, где "X" может представлять собой любую аминокислоту. В некоторых вариантах осуществления X представляет собой N.

[0517] В различных других вариантах осуществления описанные здесь праймерные редакторы (с RT, поступающей в виде слитого партнера или *in trans*) могут включать вариант RT, включающий мутацию H594X в RT M-MLV дикого типа согласно SEQ ID NO: 1361485 или в соответствующем аминокислотном положении в другой последовательности полипептида RT дикого типа, где "X" может представлять собой любую аминокислоту. В некоторых вариантах осуществления X представляет собой Q.
[0518] В различных других вариантах осуществления описанные здесь праймерные редакторы (с RT, поступающей в виде слитого партнера или *in trans*) могут включать вариант RT, включающий мутацию L603X в RT M-MLV дикого типа согласно SEQ ID NO: 1361485 или в соответствующем аминокислотном положении в другой последовательности полипептида RT дикого типа, где "X" может представлять собой любую аминокислоту. В некоторых вариантах осуществления X представляет собой W.
[0519] В различных других вариантах осуществления описанные здесь праймерные редакторы (с RT, поступающей в виде слитого партнера или *in trans*) могут включать вариант RT, включающий мутацию E607X в RT M-MLV дикого типа согласно SEQ ID NO: 1361485 или в соответствующем аминокислотном положении в другой последовательности полипептида RT дикого типа, где "X" может представлять собой любую аминокислоту. В некоторых вариантах осуществления X представляет собой K.
[0520] В различных других вариантах осуществления описанные здесь праймерные редакторы (с RT, поступающей в виде слитого партнера или *in trans*) могут включать вариант RT, включающий мутацию D653X в RT M-MLV дикого типа согласно SEQ ID NO: 1361485 или в соответствующем аминокислотном положении в другой последовательности полипептида RT дикого типа, где "X" может представлять собой любую аминокислоту. В некоторых вариантах осуществления X представляет собой N.
[0521] Описанная здесь система праймерного редактора (PE) охватывает любую публично доступную обратную транскриптазу, описанную или раскрытую в любом из следующих патентов США (каждый из которых полностью приведен здесь в качестве ссылки): 10202658; 10189831; 10150955; 9932567; 9783791; 9580698; 9534201; и 9458484, и любой ее вариант, который может быть получен с помощью известных способов внесения мутаций или известных способов эволюции белков. В следующих ссылках описаны

обратные транскриптазы в рамках данной области. Содержание каждой из них полностью приведено здесь в качестве ссылки.

[0522] Herzig, E., Voronin, N., Kucherenko, N. & Hizi, A. A Novel Leu92 Mutant of HIV-1 Reverse Transcriptase with a Selective Deficiency in Strand Transfer Causes a Loss of Viral Replication. *J. Virol.* **89**, 8119–8129 (2015).

[0523] Mohr, G. *et al.* A Reverse Transcriptase-Cas1 Fusion Protein Contains a Cas6 Domain Required for Both CRISPR RNA Biogenesis and RNA Spacer Acquisition. *Mol. Cell* **72**, 700-714.e8 (2018).

[0524] Zhao, C., Liu, F. & Pyle, A. M. An ultraprocessive, accurate reverse transcriptase encoded by a metazoan group II intron. *RNA* **24**, 183–195 (2018).

[0525] Zimmerly, S. & Wu, L. An Unexplored Diversity of Reverse Transcriptases in Bacteria. *Microbiol Spectr* **3**, MDNA3-0058–2014 (2015).

[0526] Ostertag, E. M. & Kazazian Jr, H. H. Biology of Mammalian L1 Retrotransposons. *Annual Review of Genetics* **35**, 501–538 (2001).

[0527] Perach, M. & Hizi, A. Catalytic Features of the Recombinant Reverse Transcriptase of Bovine Leukemia Virus Expressed in Bacteria. *Virology* **259**, 176–189 (1999).

[0528] Lim, D. *et al.* Crystal structure of the moloney murine leukemia virus RNase H domain. *J. Virol.* **80**, 8379–8389 (2006).

[0529] Zhao, C. & Pyle, A. M. Crystal structures of a group II intron maturase reveal a missing link in spliceosome evolution. *Nature Structural & Molecular Biology* **23**, 558–565 (2016).

[0530] Griffiths, D. J. Endogenous retroviruses in the human genome sequence. *Genome Biol.* **2**, REVIEWS1017 (2001).

[0531] Baranauskas, A. *et al.* Generation and characterization of new highly thermostable and processive M-MuLV reverse transcriptase variants. *Protein Eng Des Sel* **25**, 657–668 (2012).

[0532] Zimmerly, S., Guo, H., Perlman, P. S. & Lambowltz, A. M. Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell* **82**, 545–554 (1995).

[0533] Feng, Q., Moran, J. V., Kazazian, H. H. & Boeke, J. D. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87**, 905–916 (1996).

[0534] Berkhout, B., Jebbink, M. & Zsíros, J. Identification of an Active Reverse Transcriptase Enzyme Encoded by a Human Endogenous HERV-K Retrovirus. *Journal of Virology* **73**, 2365–2375 (1999).

[0535] Kotewicz, M. L., Sampson, C. M., D'Alessio, J. M. & Gerard, G. F. Isolation of cloned Moloney murine leukemia virus reverse transcriptase lacking ribonuclease H activity. *Nucleic Acids Res* **16**, 265–277 (1988).

[0536] Arezi, B. & Hogrefe, H. Novel mutations in Moloney Murine Leukemia Virus reverse transcriptase increase thermostability through tighter binding to template-primer. *Nucleic Acids Res* **37**, 473–481 (2009).

[0537] Blain, S. W. & Goff, S. P. Nuclease activities of Moloney murine leukemia virus reverse transcriptase. Mutants with altered substrate specificities. *J. Biol. Chem.* **268**, 23585–23592 (1993).

[0538] Xiong, Y. & Eickbush, T. H. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J* **9**, 3353–3362 (1990).

[0539] Herschhorn, A. & Hizi, A. Retroviral reverse transcriptases. *Cell. Mol. Life Sci.* **67**, 2717–2747 (2010).

[0540] Taube, R., Loya, S., Avidan, O., Perach, M. & Hizi, A. Reverse transcriptase of mouse mammary tumour virus: expression in bacteria, purification and biochemical characterization. *Biochem. J.* **329 (Pt 3)**, 579–587 (1998).

[0541] Liu, M. *et al.* Reverse Transcriptase-Mediated Tropism Switching in Bordetella Bacteriophage. *Science* **295**, 2091–2094 (2002).

[0542] Luan, D. D., Korman, M. H., Jakubczak, J. L. & Eickbush, T. H. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* **72**, 595–605 (1993).

[0543] Nottingham, R. M. *et al.* RNA-seq of human reference RNA samples using a thermostable group II intron reverse transcriptase. *RNA* **22**, 597–613 (2016).

[0544] Telesnitsky, A. & Goff, S. P. RNase H domain mutations affect the interaction between Moloney murine leukemia virus reverse transcriptase and its primer-template. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 1276–1280 (1993).

[0545] Halvas, E. K., Svarovskaia, E. S. & Pathak, V. K. Role of Murine Leukemia Virus Reverse Transcriptase Deoxyribonucleoside Triphosphate-Binding Site in Retroviral Replication and In Vivo Fidelity. *Journal of Virology* **74**, 10349–10358 (2000).

[0546] Nowak, E. *et al.* Structural analysis of monomeric retroviral reverse transcriptase in complex with an RNA/DNA hybrid. *Nucleic Acids Res* **41**, 3874–3887 (2013).

[0547] Stamos, J. L., Lentzsch, A. M. & Lambowitz, A. M. Structure of a Thermostable Group II Intron Reverse Transcriptase with Template-Primer and Its Functional and Evolutionary Implications. *Molecular Cell* **68**, 926-939.e4 (2017).

[0548] Das, D. & Georgiadis, M. M. The Crystal Structure of the Monomeric Reverse Transcriptase from Moloney Murine Leukemia Virus. *Structure* **12**, 819–829 (2004).

[0549] Avidan, O., Meer, M. E., Oz, I. & Hizi, A. The processivity and fidelity of DNA synthesis exhibited by the reverse transcriptase of bovine leukemia virus. *European Journal of Biochemistry* **269**, 859–867 (2002).

[0550] Gerard, G. F. *et al.* The role of template-primer in protection of reverse transcriptase from thermal inactivation. *Nucleic Acids Res* **30**, 3118–3129 (2002).

[0551] Monot, C. *et al.* The Specificity and Flexibility of L1 Reverse Transcription Priming at Imperfect T-Tracts. *PLOS Genetics* **9**, e1003499 (2013).

[0552] Mohr, S. *et al.* Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing. *RNA* **19**, 958–970 (2013).

[0553] Любая из приведенных выше ссылок, связанная с обратными транскриптазами, полностью приведена здесь в качестве ссылки, если иное еще не было указано.

D. <u>Слитые белки PE</u>

[0554] Описанная здесь система праймерного редактора (PE) охватывает слитые белки, включающие napDNAbp и полимеразу (например ДНК-зависимую ДНК-полимеразу или РНК-зависимую ДНК-полимеразу, например обратную транскриптазу), и необязательно соединенные линкером. Изобретение охватывает любой подходящий napDNAbp и полимеразу (например ДНК-зависимую ДНК-полимеразу или РНК-зависимую ДНК-полимеразу, например обратную транскриптазу) для комбинирования в виде единого слитого белка. Примеры napDNAbp и полимераз (например ДНК-зависимой ДНК-полимеразы или РНК-зависимой ДНК-полимеразы, например обратной транскриптазы) приведены здесь. Поскольку полимеразы хорошо известны в данной области и аминокислотные последовательности являются легкодоступными, данное изобретение не является каким-либо образом ограниченным данными конкретными указанными здесь полимеразами.

[0555] В различных вариантах осуществления слитые белки могут включать любую подходящую структурную конфигурацию. Например, слитый белок может включать в

направлении от N-конца к С-концу napDNAbp, слитый с полимеразой (например ДНК-зависимой ДНК-полимеразой или РНК-зависимой ДНК-полимеразой, например обратной транскриптазой). В других вариантах осуществления слитый белок может включать по направлению от N-конца к С-концу полимеразу (например обратную транскриптазу), слитую с napDNAbp. Слитый домен может быть необязательно соединен линкером, например аминокислотной последовательностью. В других вариантах осуществления слитые белки могут включать структуру NH_2-[napDNAbp]-[полимераза]-COOH; или NH_2-[полимераза]-[napDNAbp]-COOH, где каждый экземпляр "]-[" указывает на наличие необязательной последовательности линкера. В вариантах осуществления, где полимераза представляет собой обратную транскриптазу, слитые белки могут включать структуру NH_2-[napDNAbp]-[RT]-COOH; или NH_2-[RT]-[napDNAbp]-COOH, где каждый экземпляр "]-[" указывает на наличие необязательной последовательности линкера.

[0556] Пример слитого белка изображен на Фиг.14, на которой показан слитый белок, включающий обратную транскриптазу MLV ("MLV-RT"), слитую с никазой Cas9 ("Cas9(H840A)") с помощью последовательности линкера. Данный пример не является ограничивающим объем слитых белков, которые могут использоваться для описанной здесь системы праймерного редактора (PE).

[0557] В различных вариантах осуществления слитый белок праймерного редактора может иметь следующую аминокислотную последовательность (обозначенную здесь как "PE1"), которая включает вариант Cas9, включающий мутацию H840A (то есть никазу Cas9) и RT M-MLV дикого типа, а также N-концевую последовательность NLS (19 аминокислот) и аминокислотный линкер (32 аминокислоты), который присоединяет С-конец домена никазы Cas9 к N-концу домена RT. Слитый белок PE1 имеет следующую структуру: [NLS]-[Cas9(H840A)]-[линкер]-[MMLV_RT(дикого типа)]. Аминокислотная последовательность PE1 и его отдельные компоненты являются следующими:

Описание	Последовательность
Слитый белок PE1 Cas9(H840A)-MMLV_RT (дикого типа)	MKRTADGSEFESPKKKRKV**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKK NLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIE GDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGA SAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIV DLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN**

Описание	Последовательность
	EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS **GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGIL** **QTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPV** **ENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNR** **GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI** **TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN** **AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLA** **NGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS** **DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI** **DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY** **EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ** **AENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD** **SGGSSGGSSGSETPGTSESATPESSGGSSGGSS***TLNIEDEYRLHETSKEPDVSLGSTWLSDFP* *QAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSP* *WNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFF* *CLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLIL* *LQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTE* *ARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQ* *EIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPP* *CLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQ* *FGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQR* *KAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEI* *YRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAA* *ITETPDTSTLLIENSSP*SGGSKRTADGSEFEPKKKRKV (SEQ ID NO: 1361515) Обозначения: Последовательность ядерной локализации (NLS) Сверху:(SEQ ID NO: 1361532); Снизу: (SEQ ID NO: 1361541) **Cas9(H840A)** (SEQ ID NO: 1361454) **Линкер длиной 33 аминокислоты** (SEQ ID NO: 1361528) *Обратная транскриптаза M-MLV* (SEQ ID NO: 1361485)

[0558] В другом варианте осуществления слитый белок праймерного редактора может иметь следующую аминокислотную последовательность (обозначенную здесь как "PE2"), которая включает вариант Cas9, включающий мутацию H840A (то есть никазу Cas9) и RT M-MLV, включающую мутации D200N, T330P, L603W, T306K, и W313F, а также N-концевую последовательность NLS (19 аминокислот) и аминокислотный линкер (33 аминокислоты), который присоединяет C-конец домена никазы Cas9 к N-концу домена RT. Слитый белок PE2 имеет следующую структуру: [NLS]-[Cas9(H840A)]-[линкер]-[MMLV_RT(D200N)(T330P)(L603W)(T306K)(W313F)]. Аминокислотная последовательность PE2 является следующей:

Слитый белок PE2 Cas9(H840A)-MMLV_RT(D200N, T330P, L603W,	MKRTADGSEFESPKKKRKV**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKK** **NLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE** **EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIE** **GDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK** **NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS** **DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA** **GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL** **RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGA**

T306K, W313F)	**SAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIV DLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGIL QTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPV ENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNR GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLA NGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ AENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD SGGSSGGSSGSETPGTSESATPESSGGSSGGSS*TLNIEDEYRLHETSKEPDVSLGSTWLSDFP QAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSP WNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFF CLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLIL LQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTE ARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQ EIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPP CLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQ FGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQR KAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEI YRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAA ITETPDTSTLLIENSSP*SGGSKRTADGSEFEPKKKRKV (SEQ ID NO: 1361516) Обозначения: Последовательность ядерной локализации (NLS) Сверху:(SEQ ID NO: 1361532); Снизу: (SEQ ID NO: 1361541) **Cas9(H840A)** (SEQ ID NO: 1361454) **Линкер длиной 33 аминокислоты** (SEQ ID NO: 1361528) *Обратная транскриптаза M-MLV* (SEQ ID NO: 1361514)

[0559] В других вариантах осуществления слитый белок праймерного редактора может иметь следующие аминокислотные последовательности:

Слитый белок PE MMLV_RT(дикого типа)-32aa-Cas9(H840A)	MKRTADGSEFESPKKKRKV*TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVR QAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTN DYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAF EWRDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSE LDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKT PRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGL PDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKD AGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLP LPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIW AKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEI KNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIEN SSP*SGGSSGGSSGSETPGTSESATPESSGGSSGGSSDKKYSIGLDIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMA KVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLI YLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLS KSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFD NGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE

	GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH **DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTG** **WGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLH** **EHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRI** **EEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFL** **KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE** **LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY** **KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKY** **FFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTE** **VQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKE** **LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE** **LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLD** **KVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQS** **ITGLYETRIDLSQLGGD**SGGSKRTADGSEFEPKKKRKV (SEQ ID NO: 1361517) Обозначения: Последовательность ядерной локализации (NLS) Сверху:(SEQ ID NO: 1361532), Снизу: (SEQ ID NO: 1361541) **Cas9(H840A)** (SEQ ID NO: 1361454) **Линкер длиной 33 аминокислоты** (SEQ ID NO: 1361528?) *Обратная транскриптаза M-MLV* (SEQ ID NO: 1361497)
Слитый белок PE MMLV_RT(дикого типа)-60aa-Cas9(H840A)	MKRTADGSEFESPKKKRKV*TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVR* *QAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTN* *DYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAF* *EWRDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSE* *LDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKT* *PRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGL* *PDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKD* *AGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLP* *LPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIW* *AKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEI* *KNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIEN* *SSPS***GGSSGGSSGSETPGTSESATPESAGSYPYDVPDYAGSAAPAAKKKKLDGSGSGGSSGGS** **DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL** **KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA** **YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQ** **TYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFK** **SNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAP** **LSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPI** **LEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK** **ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNE** **KVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDY** **FKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREM** **IEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFM** **QLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPE** **NIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGR** **DMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYW** **RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEN** **DKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFV** **YGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI** **VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD** **SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKL** **PKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ** **HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF** **KYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD**SGGSKRTADGSEFEPKKKR KV (SEQ ID NO: 1361518)

	Обозначения: Последовательность ядерной локализации (NLS) Сверху:(SEQ ID NO: 1361532), Снизу: (SEQ ID NO: 1361541) **Cas9(H840A)**(SEQ ID NO: 1361454) **Линкер длиной 33 аминокислоты** (SEQ ID NO: 1361585) *Обратная транскриптаза M-MLV* (SEQ ID NO: 1361497)
Слитый белок PE Cas9(H840A)-FEN1-MMLV_RT(D200N, T330P, L603W, T306K, W313F)	MKRTADGSEFESPKKKRKV**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKK NLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIE GDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGA SAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIV DLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGIL QTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPV ENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNR GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLA NGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ AENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD SGGSSGGSSGSETPGTSESATPESSGGSSGGSS**GIQGLAKLIADVAPSAIRENDIKSYFGRKV AIDASMSIYQFLIAVRQGGDVLQNEEGETTSHLMGMFYRTIRMMENGIKPVYVFDGKPPQLKS GELAKRSERRAEAEKQLQQAQAAGAEQEVEKFTKRLVKVTKQHNDECKHLLSLMGIPYLDAPS EAEASCAALVKAGKVYAAATEDMDCLTFGSPVLMRHLTASEAKKLPIQEFHLSRILQELGLNQ EQFVDLCILLGSDYCESIRGIGPKRAVDLIQKHKSIEEIVRRLDPNKYPVPENWLHKEAHQLF LEPEVLDPESVELKWSEPNEEELIKFMCGEKQFSEERIRSGVKRLSKSRQGSTQGRLDDFFKV TGSLSSAKRKEPEPKGSTKKKAKTGAAGKFKRGK**SGGSSGGSSGSETPGTSESATPESSGGSS GGSS***TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPV SIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRV EDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTW TRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLG NLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCR LFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQ GYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAP HAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILA EAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAEL IALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFL PKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP**SGGSKRTADGSE FEPKKKRKV (SEQ ID NO: 1361519) Обозначения: Последовательность ядерной локализации (NLS) Сверху:(SEQ ID NO: 1361532), Снизу: (SEQ ID NO: 1361541) **Cas9(H840A)** (SEQ ID NO: 1361454) **Линкер длиной 32 аминокислоты** (SEQ ID NO: 1361528?) *Обратная транскриптаза M-MLV* (SEQ ID NO: 1361514) FEN1: SEQ ID NO: 1361542

[0560] В различных вариантах осуществления указанные здесь слитые белки праймерного редактора также могут включать любые варианты описанных выше последовательностей, имеющие аминокислотную последовательность, которая, по меньшей мере, приблизительно на 70% идентична, по меньшей мере, приблизительно на 80% идентична, по меньшей мере, приблизительно на 90% идентична, по меньшей мере, приблизительно на 95% идентична, по меньшей мере, приблизительно на 96% идентична, по меньшей мере, приблизительно на 97% идентична, по меньшей мере, приблизительно на 98% идентична, по меньшей мере, приблизительно на 99% идентична, по меньшей мере, приблизительно на 99,5% идентична, или, по меньшей мере, приблизительно на 99,9% идентична PE1, PE2 или любой из указанных выше последовательностей слитого праймерного редактора.

[0561] В некоторых вариантах осуществления могут использоваться линкеры для связывания любых пептидов или пептидных доменов или фрагментов по изобретению (например napDNAbp, связанный или слитый с обратной транскриптазой).

[0562] В других вариантах осуществления слитые белки праймерного редактора могут быть основаны на никазах SaCas9 или SpCas9 с измененной специфичностью в отношении PAM, таких как следующие примеры последовательностей:

Праймерный редактор SaCas9-RTM-MLV	MKRTADGSEFESPKKKRKVGKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKE ANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKG LSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAE LQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRR TYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLV ITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEF TNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEI EQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIP TTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEM QKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPF NYEVDHIIPRSVSFDNSFNNKVLVKQEEASKKGNRTPFQYLSSSDSKISYETFKKHI LNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFR VNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLD KAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKP NRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQT YQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDI TDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEA KKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLEN MNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKGSGGSSGGSSGS ETPGTSESATPESSGGSSGGSSTLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAE TGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQS PWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVL DLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRD LADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQ KQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEM AAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYA

	KGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVI LAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGL QHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIW AKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLT SEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAIT ETPDTSTLLIENSSPSGGSKRTADGSEFEPKKKRKV (SEQ ID NO: 1361599)
Праймерный редактор SpCas9(H840A)-VRQR-обратная транскриптаза вируса лейкоза мышей Молони	MKRTADGSEFESPKKKRKVDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDS FFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRL IYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKA ILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKR YDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE KMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNRE KIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLL FKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDN EENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKL INGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEH IANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQ RKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIAR KKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFL YLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSA YNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQ SITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSTLNIED EYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNK RVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEM GISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSEL DCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIK QALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAG WPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQA LLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADH TWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGK KLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSII HCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSPSGGSKRTADGSEF EPKKKRKV) (SEQ ID NO: 1361600)
Праймерный редактор SpCas9(H840A)-VRER-обратная транскриптаза вируса лейкоза мышей Молони	MKRTADGSEFESPKKKRKVDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDS FFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRL IYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKA ILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKR YDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE KMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNRE KIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLL FKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDN EENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKL INGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEH

	IANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQ RKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIAR KKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFL YLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSA YNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKEYRSTKEVLDATLIHQ SITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSTLNIED EYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNK RVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEM GISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSEL DCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIK QALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAG WPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQA LLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADH TWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGK KLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSII HCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSPSGGSKRTADGSEF EPKKKRKV (SEQ ID NO: 1361601)

[0563] В других вариантах осуществления предусмотренные здесь слитые белки праймерного редактора могут включать никазу Cas9 (например Cas9 (H840A)), слитую с усеченной версией обратной транскриптазы M-MLV. В данном варианте осуществления обратная транскриптаза также содержит 4 мутации (D200N, T306K, W313F, T330P; следует отметить, что мутация L603W, присутствующая в PE2, является отсутствующей вследствие усечения). Последовательность ДНК, кодирующая данный усеченный редактор, на 522 bp короче, чем PE2, и, таким образом, является потенциально полезной для вариантов использования, в рамках которых доставка последовательности ДНК является сложной вследствие ее размера (то есть при доставке на основе адено-ассоциированного вируса и лентивируса). Данный вариант осуществления обозначается как Cas9(H840A)-MMLV-RT(trunc) или "PE2-short" или "PE2-trunc" и имеет следующую аминокислотную последовательность:

CAS9(H840a)- mmlv-rt(TRUNC) ИЛИ PE2-SHORT	MKRTADGSEFESPKKKRKV**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD** **RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDS** **FFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRL** **IYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKA** **ILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS** **KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKR** **YDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE** **KMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNRE** **KIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM** **TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLL**

	FKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDN **EENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKL** **INGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEH** **IANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER** **MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD** **VDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQ** **RKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR** **EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF** **VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET** **NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIAR** **KKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI** **DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFL** **YLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSA** **YNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ** **SITGLYETRIDLSQLGGD**SGGSSGGSSGSETPGTSESATPESSGGSSGGSS*TLNIED* *EYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ* *YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNK* *RVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEM* *GISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSEL* *DCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ* *PTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIK* *QALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAG* *WPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQA* *LLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDNSRLIN*SGGSKRTADGSEFEPK KKRKV (SEQ ID NO: 1361602) Обозначения: Последовательность ядерной локализации (NLS) Сверху: (SEQ ID NO: 1361532) Снизу: (SEQ ID NO: 1361541) Cas9(H840A) (SEQ ID NO: 1361454) Линкер 1 длиной 33 аминокислоты (SEQ ID NO: 1361528) Обратная транскриптаза M-MLV TRUNCATED:(SEQ ID NO: 1361597) Линкер 2 длиной 33 аминокислоты (SEQ ID NO: 1361541) FEN1 (SEQ ID NO: 1361542)

[0564] См. Фиг.36, на которой приведена гистограмма для сравнения эффективности (то есть "% общих считываний секвенирования с указанным изменением или инделями") PE2, PE2-trunc, PE3, и PE3-trunc в различных участках-мишенях в различных линиях клеток. Данные показывают, что праймерные редакторы, включающие усеченные варианты RT, были примерно настолько же эффективны, что и праймерные редакторы, включающие неусеченные RT белки.

[0565] В различных вариантах осуществления указанные здесь слитые белки праймерного редактора также могут включать любые варианты описанных выше последовательностей, имеющие аминокислотную последовательность, которая, по меньшей мере, приблизительно на 70% идентична, по меньшей мере, приблизительно на 80% идентична, по меньшей мере, приблизительно на 90% идентична, по меньшей мере, приблизительно на 95% идентична, по меньшей мере, приблизительно на 96% идентична, по меньшей

мере, приблизительно на 97% идентична, по меньшей мере, приблизительно на 98% идентична, по меньшей мере, приблизительно на 99% идентична, по меньшей мере, приблизительно на 99,5% идентична, или, по меньшей мере, приблизительно на 99,9% идентична PE1, PE2 или любой из указанных выше последовательностей слитого праймерного редактора.

[0566] В определенных вариантах осуществления могут использоваться линкеры для связывания любых пептидов или пептидных доменов или фрагментов по изобретению (например napDNAbp, связанный или слитый с обратной транскриптазой).

E. Линкеры и другие домены слитого белка

[0567] Слитые белки PE могут включать различные другие домены помимо napDNAbp (например домен Cas9) и полимеразный домен (например домен RT). Например, в случае, если napDNAbp представляет собой Cas9, а полимераза представляет собой RT, слитые белки PE могут включать один или несколько линкеров, которые соединяют домен Cas9 с доменом RT. Линкеры могут также соединять другие функциональные домены, например последовательности ядерной локализации (NLS) или FEN1 (или другую флэп-эндонуклеазу) со слитыми белками PE или его доменом.

(i) Линкеры

[0568] В соответствии с используемым здесь значением термин "линкер" обозначает химическую группу или молекулу, связывающую две молекулы или фрагмента, например домен связывания и домен расщепления нуклеазы. В некоторых вариантах осуществления линкер соединяет домен связывания нРНК РНК-программируемой нуклеазы и каталитический домен полимеразы (например обратной транскриптазы). В некоторых вариантах осуществления линкер соединяет dCas9 и обратную транскриптазу. Как правило, линкер располагается между или фланкируется двумя группами, молекулами, или иными фрагментами и присоединяется к каждому из них через ковалентную связь, тем самым связывая их. В некоторых вариантах осуществления линкер представляет собой аминокислоту или несколько аминокислот (например пептид или белок). В некоторых вариантах осуществления линкер представляет собой органическую молекулу, группу, полимер, или химический фрагмент. В некоторых вариантах осуществления линкер имеет длину 5-100 аминокислот, например длину 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 30-35, 35-40,

40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, или 150-200 аминокислот. Также охватываются более длинные или короткие линкеры.

[0569] Линкер может быть представлять собой просто ковалентную цепь или он может представлять собой полимерный линкер, имеющий длину, соответствующую нескольким атомам. В определенных вариантах осуществления линкер представляет собой полипептид или основан на аминокислотах. В других вариантах осуществления линкер не является подобным пептиду. В некоторых вариантах осуществления линкер представляет собой ковалентную связь (например связь углерод-углерод, дисульфидную связь, связь углерод-гетероатом, и так далее). В некоторых вариантах осуществления линкер представляет собой связь углерод-азот в рамках амидной связи. В некоторых вариантах осуществления линкер является циклическим или ациклическим, замещенным или незамещенным, разветвленным или неразветвленным, алифатическим или гетероалифатическим линкером. В некоторых вариантах осуществления линкер является полимерным (например полиэтиленовым, полиэтиленгликолевым, полиамидным, полиэфирным, и так далее). В некоторых вариантах осуществления линкер включает мономер, димер, или полимер аминоалкановой кислоты. В некоторых вариантах осуществления линкер включает аминоалкановую кислоту (например глицин, этановую кислоту, аланин, бета-аланин, 3-аминопропановую кислоту, 4-аминобутановую кислоту, 5-пентановую кислоту, и так далее). В некоторых вариантах осуществления линкер включает мономер, димер, или полимер аминогексановой кислоты (Ahx). В некоторых вариантах осуществления линкер основан на карбоциклическом фрагменте (например циклопентане, циклогексане). В других вариантах осуществления линкер включает полиэтиленгликолевый фрагмент (ПЭГ). В других вариантах осуществления линкер включает аминокислоты. В некоторых вариантах осуществления линкер включает пептид. В некоторых вариантах осуществления линкер включает ариловый или гетероариловый фрагмент. В некоторых вариантах осуществления линкер основан на фенильном кольце. Линкер может включать функционализированные фрагменты для обеспечения присоединения нуклеофила (например тиола, амино) из пептида к линкеру. В качестве части линкера может использоваться любой электрофил. Примеры электрофилов включают, но без ограничения, активированные сложные эфиры,

активированные амиды, акцепторы Майкла, алкильные галиды, арильные галиды, ацильные галиды, и изотиоцианаты.

[0570] В некоторых других вариантах осуществления линкер включает аминокислотную последовательность (GGGGS)n (SEQ ID NO: 1361520), (G)n (SEQ ID NO: 1361521), (EAAAK)n (SEQ ID NO: 1361522), (GGS)n (SEQ ID NO: 1361523), (SGGS)n (SEQ ID NO: 1361524), (XP)n (SEQ ID NO: 1361525), или любую их комбинацию, при этом n независимо представляет собой целое число от 1 до 30, и при этом X представляет собой любую аминокислоту. В некоторых вариантах осуществления линкер включает аминокислотную последовательность (GGS)n (SEQ ID NO: 1361526), при этом n представляет собой 1, 3, или 7. В некоторых вариантах осуществления линкер включает аминокислотную последовательность SGSETPGTSESATPES (SEQ ID NO: 1361527). В некоторых вариантах осуществления линкер включает аминокислотную последовательность SGGSSGGSSGSETPGTSESATPESSGGSSGGS (SEQ ID NO: 1361528). В некоторых вариантах осуществления линкер включает аминокислотную последовательность SGGSGGSGGS (SEQ ID NO: 1361529). В некоторых вариантах осуществления линкер включает аминокислотную последовательность SGGS (SEQ ID NO: 1361530).

[0571] В частности, в различных вариантах осуществления могут использоваться следующие линкеры для соединения доменов праймерного редактора друг с другом:

[0572] GGS;

[0573] GGSGGS(SEQ ID NO: 1361523);

[0574] GGSGGSGGS(SEQ ID NO: 1361523);

[0575] SGGSSGGSSGSETPGTSESATPESSGGSSGGSS(SEQ ID NO: 1361528);

[0576] SGSETPGTSESATPES(SEQ ID NO: 1361527);

[0577] SGGSSGGSSGSETPGTSESATPESAGSYPYDVPDYAGSAAPAAKKKKLDGSGSGGSSGGS (SEQ ID NO: 1361585).

(ii) Последовательность ядерной локализации (NLS)

[0578] В различных вариантах осуществления слитые белки PE могут включать одну или несколько последовательностей ядерной локализации (NLS), которые помогают промотировать транслокацию белка в ядро клетки. Такие последовательности хорошо известны в данной области и могут включать следующие примеры:

Описание	Последовательность	SEQ ID NO:
NLS большого Т-антигена SV40	PKKKRKV	SEQ ID NO: 1361531
NLS	MKRTADGSEFESPKKKRKV	SEQ ID NO: 1361532
NLS	MDSLLMNRRKFLYQFKNVRWAKGRRETYLC	SEQ ID NO: 1361533
NLS нуклеоплазмина	AVKRPAATKKAGQAKKKKLD	SEQ ID NO: 1361534
NLS EGL-13	MSRRRKANPTKLSENAKKLAKEVEN	SEQ ID NO: 1361535
NLS c-MYC	PAAKRVKLD	SEQ ID NO: 1361536
NLS TUS-белка	KLKIKRPVK	SEQ ID NO: 1361537
NLS большого Т-антигена полиомавируса	VSRKRPRP	SEQ ID NO: 1361538
NLS антигена вируса гепатита	EGAPPAKRAR	SEQ ID NO: 1361539
NLS p53 мыши	PPQPKKKPLDGE	SEQ ID NO: 1361540
	SGGSKRTADGSEFEPKKKRKV	SEQ ID NO: 1361541

[0579] Приведенные выше примеры NLS не являются ограничивающими. Слитые белки PE могут включать любую последовательность NLS, включая любую из описанных в Cokol et al., "Finding nuclear localization signals," *EMBO Rep.*, 2000, 1(5): 411-415 и Freitas et al., "Mechanisms and Signals for the Nuclear Import of Proteins," *Current Genomics*, 2009, 10(8): 550-7, каждая из которых приведена здесь в качестве ссылки.

[0580] В различных вариантах осуществления описанные здесь праймерные редакторы и конструкты, кодирующие праймерные редакторы, дополнительно включают один или несколько, предпочтительно, по меньшей мере, два сигнала ядерной локализации. В определенных вариантах осуществления праймерные редакторы включают, по меньшей мере, две NLS. В вариантах осуществления, включающих, по меньшей мере, две NLS, они могут являться одинаковыми NLS или они могут являться различающимися NLS. Кроме того, NLS могут быть экспрессированы в виде части слитого белка с остальными частями праймерных редакторов. В некоторых вариантах осуществления одна или несколько NLS представляют собой состоящие из двух частей NLS ("bpNLS"). В определенных вариантах осуществления описанные слитые белки включают две состоящие из двух частей NLS. В некоторых вариантах осуществления описанные слитые белки включают более чем две состоящие из двух частей NLS.

[0581] Слитая NLS может располагаться на N-конце, C-конце, или внутри последовательности праймерного редактора (например будучи вставленной между кодируемым компонентом в виде napDNAbp (например Cas9) и полимеразным доменом (например доменом обратной транскпритазы).

[0582] NLS могут представлять собой любую известную в данной области последовательность NLS. NLS также могут представлять собой любые открытые в будущем NLS для ядерной локализации. NLS также могут представлять собой любую встречающуюся в природе NLS или любую не встречающуюся в природе NLS (например NLS с одной или несколькими требуемыми мутациями).

[0583] Термин "последовательность ядерной локализации" или "NLS" обозначает аминокислотную последовательность, которая промотирует импорт белка в ядро клетки, например за счет ядерного транспорта. Последовательности ядерной локализации известны в данной области и будут очевидны специалисту в данной области. Например, последовательности NLS описаны у Plank et al., в международной PCT заявке PCT/EP2000/011690, поданной 23 ноября 23 2000 г., опубликованной как WO/2001/038547 31 мая 2001 г., содержание которой приведено здесь в качестве ссылки. В некоторых вариантах осуществления NLS включает аминокислотную последовательность PKKKRKV (SEQ ID NO: 1361531), MDSLLMNRRKFLYQFKNVRWAKGRRETYLC (SEQ ID NO: 1361533), KRTADGSEFESPKKKRKV (SEQ ID NO: 1361659), или KRTADGSEFEPKKKRKV (SEQ ID NO: 1361660). В других вариантах осуществления NLS включает аминокислотные последовательности NLSKRPAAIKKAGQAKKKK (SEQ ID NO: 1361661), PAAKRVKLD (SEQ ID NO: 1361536), RQRRNELKRSF (SEQ ID NO: 1361662), NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY (SEQ ID NO: 1361663).

[0584] В одном аспекте изобретения праймерный редактор может быть модифицирован с использованием одного или нескольких сигналов ядерной локализации (NLS), предпочтительно, по меньшей мере, двух NLS. В определенных вариантах осуществления праймерные редакторы модифицированы с использованием двух или более NLS. Изобретение охватывает использование любого сигнала ядерной локализации, известного в данной области на момент изобретения, или любого сигнала ядерной локализации, который идентифицирован или иным образом сделан доступным в данной области после

подачи настоящей заявки. Соответствующий сигнал ядерной локализации представляет собой пептидную последовательность, которая направляет белок к ядру клетки, в которой экспрессируется последовательность. Сигнал ядерной локализации является преимущественно оснóвным, может быть расположен практически где угодно в рамках аминокислотной последовательности белка, в целом включает короткую последовательность от четырех аминокислот до восьми аминокислот (Autieri & Agrawal, (1998) *J. Biol. Chem*. 273: 14731-37, incorporated herein by reference), и обычно имеет высокое содержание остатков лизина и аргинина (Magin *et al*., (2000) *Virology* 274: 11-16, приведен здесь в качестве ссылки). Сигналы ядерной локализации зачастую включают остатки пролина. Было идентифицировано множество сигналов ядерной локализации, при этом они применялись для обеспечения транспорта биологических молекул из цитоплазмы в ядро клетки. См., например, Tinland et al., (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89:7442-46; Moede et al., (1999) *FEBS Lett.* 461:229-34, который приведен здесь в качестве ссылки. В настоящее время считается, что в транслокации принимают участие белки ядерных пор.

[0585] Большинство NLS могут быть отнесены к трем основным группам: (i) состоящий из одной части NLS, примером которого является NLS большого Т-антигена SV40 PKKKRKV (SEQ ID NO: 1361531)); (ii) состоящий из двух частей мотив, состоящий из двух основных доменов, разделенных переменным количеством спейсерных аминокислот, примером которого является NLS *Xenopus nucleoplasmin* (KRXXXXXXXXXXKKKL (SEQ ID NO: 1361664)); и (iii) нестандартные последовательности, такие как M9 белка hnRNP Al, NLS нуклеопротеина вируса гриппа, и NLS белка Gal4 дрожжей (Dingwall and Laskey 1991).

[0586] Сигналы ядерной локализации располагаются в различных точках аминокислотных последовательностей белков. NLS были идентифицированы на N-конце, С-конце, и в центральной части белков. Таким образом, изобретение предусматривает праймерные редакторы, которые могут быть модифицированы с использованием одного или нескольких NLS на С-конце, N-конце, а также во внутренней части праймерного редактора. Остатки более длинной последовательности, которые не функционирует в качестве остатков компонента в виде NLS, должны быть выбраны таким образом, чтобы не нарушать, например тонически или стерически, сам сигнал ядерной локализации.

Таким образом, несмотря на отсутствие строгих ограничений в плане композиции, включающей NLS последовательности, на практике такая последовательность может быть функционально ограничена в плане длины и композиции.

[0587] Настоящее изобретение охватывает любые подходящие средства для модификации праймерного редактора для включения одного или нескольких NLS. В одном аспекте праймерные редакторы могут быть сконструированы для экспрессии белка праймерного редактора, который трансляционным образом слит на его N-конце или C-конце (или их обоих) с одним или несколькими NLS, то есть для образования слитого конструкта праймерный редактор-NLS. В других вариантах осуществления кодирующая праймерный редактор нуклеотидная последовательность может быть генетически модифицирована для включения рамки считывания, которая кодирует один или несколько NLS во внутреннем участке кодируемого праймерного редактора. Кроме того, NLS могут включать различные аминокислотные линкеры или спейсерные участки, кодируемые между праймерным редактором и присоединенной на N-конце, C-конце, или в середине аминокислотной последовательностью NLS, например в центральной части белков. Таким образом, настоящее изобретение также предусматривает нуклеотидные конструкты, векторы и клетки-хозяева для экспрессии слитых белков, которые включают праймерный редактор и один или несколько NLS.

[0588] Описанные здесь праймерные редакторы также могут включать сигналы ядерной локализации, которые связаны с праймерным редактором с помощью одного или нескольких линкеров, например элемента в виде полимерного, аминокислотного, основанного на нуклеиновой кислоте, полисахаридного, или химического линкера. В рамках охватываемого объема изобретения линкеры не имеют каких-либо ограничений и могут представлять собой любой подходящий тип молекулы (например полимер, аминокислоту, полисахарид, нуклеиновую кислоту, липид, или любой синтетический химический линкерный домен) и быть присоединенными к праймерному редактору с помощью любой подходящей стратегии, которая обеспечивает образование связи (например ковалентной связи, водородной связи) между праймерным редактором и одним или несколькими NLS.

(iii)Флэп-эндонуклеазы (например FEN1)

[0589] В различных вариантах осуществления слитые белки PE могут включать одну или несколько флэп-эндонуклеаз (например FEN1), которые обозначают фермент, который катализирует удаление 5′ одноцепочечных ДНК-флэпов. Они представляют собой встречающиеся в природе ферменты, которые процессируют удаление 5′ флэпов, образованных в рамках клеточных процессов, включая репликацию ДНК. В рамках описанных здесь способов праймерного редактирования могут применяться эндогенно или *in trans* поступающие флэп-эндонуклеазы для удаления 5′ флэпа эндогенной ДНК, образованного в участке-мишени при праймерном редактировании. Флэп-эндонуклеазы известны в данной области и описаны в Patel et al., "Flap endonucleases pass 5′-flaps through a flexible arch using a disorder-thread-order mechanism to confer specificity for free 5′-ends," *Nucleic Acids Research*, 2012, 40(10): 4507-4519 и Tsutakawa et al., "Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily," *Cell*, 2011, 145(2): 198-211 (каждый из которых приведен здесь в качестве ссылки). Пример флэп-эндонуклеазы представляет собой FEN1, которая может быть описана в виде следующей аминокислотной последовательности:

Описание	Последовательность	SEQ ID NO:
FEN1 дикого типа (wt)	MGIQGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSIYQFLIAVRQGGDVLQN EEGETTSHLMGMFYRTIRMMENGIKPVYVFDGKPPQLKSGELAKRSERRAEAEKQ LQQAQAAGAEQEVEKFTKRLVKVTKQHNDECKHLLSLMGIPYLDAPSEAEASCAA LVKAGKVYAAATEDMDCLTFGSPVLMRHLTASEAKKLPIQEFHLSRILQELGLNQ EQFVDLCILLGSDYCESIRGIGPKRAVDLIQKHKSIEEIVRRLDPNKYPVPENWL HKEAHQLFLEPEVLDPESVELKWSEPNEEELIKFMCGEKQFSEERIRSGVKRLSK SRQGSTQGRLDDFFKVTGSLSSAKRKEPEPKGSTKKKAKTGAAGKFKRGK	SEQ ID NO: 1361542

[0590] Флэп-эндонуклеазы также могут включать любой вариант FEN1, мутант, или другой ортолог, гомолог, или вариант флэп-эндонуклеазы. Неограничивающие примеры вариантов FEN1 являются следующими:

Описание	Последовательность	SEQ ID NO:
FEN1 K168R (относительно FEN1 дикого типа)	MGIQGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSIYQFLIAVRQGGDVLQN EEGETTSHLMGMFYRTIRMMENGIKPVYVFDGKPPQLKSGELAKRSERRAEAEKQ LQQAQAAGAEQEVEKFTKRLVKVTKQHNDECKHLLSLMGIPYLDAPSEAEASCAA LV**R**AGKVYAAATEDMDCLTFGSPVLMRHLTASEAKKLPIQEFHLSRILQELGLNQ EQFVDLCILLGSDYCESIRGIGPKRAVDLIQKHKSIEEIVRRLDPNKYPVPENWL HKEAHQLFLEPEVLDPESVELKWSEPNEEELIKFMCGEKQFSEERIRSGVKRLSK SRQGSTQGRLDDFFKVTGSLSSAKRKEPEPKGSTKKKAKTGAAGKFKRGK	SEQ ID NO: 1361543
FEN1 S187A (относительно FEN1 дикого типа)	MGIQGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSIYQFLIAVRQGGDVLQN EEGETTSHLMGMFYRTIRMMENGIKPVYVFDGKPPQLKSGELAKRSERRAEAEKQ LQQAQAAGAEQEVEKFTKRLVKVTKQHNDECKHLLSLMGIPYLDAPSEAEASCAA LVKAGKVYAAATEDMDCLTFG**A**PVLMRHLTASEAKKLPIQEFHLSRILQELGLNQ EQFVDLCILLGSDYCESIRGIGPKRAVDLIQKHKSIEEIVRRLDPNKYPVPENWL HKEAHQLFLEPEVLDPESVELKWSEPNEEELIKFMCGEKQFSEERIRSGVKRLSK SRQGSTQGRLDDFFKVTGSLSSAKRKEPEPKGSTKKKAKTGAAGKFKRGK	SEQ ID NO: 1361544
FEN1 K354R (относитель	MGIQGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSIYQFLIAVRQGGDVLQN EEGETTSHLMGMFYRTIRMMENGIKPVYVFDGKPPQLKSGELAKRSERRAEAEKQ LQQAQAAGAEQEVEKFTKRLVKVTKQHNDECKHLLSLMGIPYLDAPSEAEASCAA	SEQ ID NO: 1361545

но FEN1 дикого типа)	LVKAGKVYAAATEDMDCLTFGSPVLMRHLTASEAKKLPIQEFHLSRILQELGLNQ EQFVDLCILLGSDYCESIRGIGPKRAVDLIQKHKSIEEIVRRLDPNKYPVPENWL HKEAHQLFLEPEVLDPESVELKWSEPNEEELIKFMCGEKQFSEERIRSGVKRLSK SRQGSTQGRLDDFFKVTGSLSSA**R**RKEPEPKGSTKKKAKTGAAGKFKRGK	
GEN1	MGVNDLWQILEPVKQHIPLRNLGGKTIAVDLSLWVCEAQTVKKMMGSVMKPHLRN LFFRISYLTQMDVKLVFVMEGEPPKLKADVISKRNQSRYGSSGKSWSQKTGRSHF KSVLRECLHMLECLGIPWVQAAGEAEAMCAYLNAGGHVDGCLTNDGDTFLYGAQT VYRNFTMNTKDPHVDCYTMSSIKSKLGLDRDALVGLAILLGCDYLPKGVPGVGKE QALKLIQILKGQSLLQRFNRWNETSCNSSPQLLVTKKLAHCSVCSHPGSPKDHER NGCRLCKSDKYCEPHDYEYCCPCEWHRTEHDRQLSEVENNIKKKACCCEGFPFHE VIQEFLLNKDKLVKVIRYQRPDLLLFQRFTLEKMEWPNHYACEKLLVLLTHYDMI ERKLGSRNSNQLQPIRIVKTRIRNGVHCFEIEWEKPEHYAMEDKQHGEFALLTIE EESLFEAAYPEIVAVYQKQKLEIKGKKQKRIKPKENNLPEPDEVMSFQSHMTLKP TCEIFHKQNSKLNSGISPDPTLPQESISASLNSLLLPKNTPCLNAQEQFMSSLRP LAIQQIKAVSKSLISESSQPNTSSHNISVIADLHLSTIDWEGTSFSNSPAIQRNT FSHDLKSEVESELSAIPDGFENIPEQLSCESERYTANIKKVLDEDSDGISPEEHL LSGITDLCLQDLPLKERIFTKLSYPQDNLQPDVNLKTLSILSVKESCIANSGSDC TSHLSKDLPGIPLQNESRDSKILKGDQLLQEDYKVNTSVPYSVSNTVVKTCNVRP PNTALDHSRKVDMQTTRKILMKKSVCLDRHSSDEQSAPVFGKAKYTTQRMKHSSQ KHNSSHFKESGHNKLSSPKIHIKETEQCVRSYETAENEESCFPDSTKSSLSSLQC HKKENNSGTCLDSPLPLRQRLKLRFQST	SEQ ID NO: 1361546
ERCC5	MGVQGLWKLLECSGRQVSPEALEGKILAVDISIWLNQALKGVRDRHGNSIENPHL LTLFHRLCKLLFFRIRPIFVFDGDAPLLKKQTLVKRRQRKDLASSDSRKTTEKLL KTFLKRQAIKTAFRSKRDEALPSLTQVRRENDLYVLPPLQEEEKHSSEEEDEKEW QERMNQKQALQEEFFHNPQAIDIESEDFSSLPPEVKHEILTDMKEFTKRRRTLFE AMPEESDDFSQYQLKGLLKKNYLNQHIEHVQKEMNQQHSGHIRRQYEDEGGFLKE VESRRVVSEDTSHYILIKGIQAKTVAEVDSESLPSSSKMHGMSFDVKSSPCEKLK TEKEPDATPPSPRTLLAMQAALLGSSSEEELESENRRQARGRNAPAAVDEGSISP RTLSAIKRALDDDEDVKVCAGDDVQTGGPGAEEMRINSSTENSDEGLKVRDGKGI PFTATLASSSVNSAEEHVASTNEGREPTDSVPKEQMSLVHVGTEAFPISDESMIK DRKDRLPLESAVVRHSDAPGLPNGRELTPASPTCTNSVSKNETHAEVLEQQNELC PYESKFDSSLLSSDDETKCKPNSASEVIGPVSLQETSSIVSVPSEAVDNVENVVS FNAKEHENFLETIQEQQTTESAGQDLISIPKAVEPMEIDSEESESDGSFIEVQSV ISDEELQAEFPETSKPPSEQGEEELVGTREGEAPAESESLLRDNSERDDVDGEPQ EAEKDAEDSLHEWQDINLEELETLESNLLAQQNSLKAQKQQQERIAATVTGQMFL ESQELLRLFGIPYIQAPMEAEAQCAILDLTDQTSGTITDDSDIWLFGARHVYRNF FNKNKFVEYYQYVDFHNQLGLDRNKLINLAYLLGSDYTEGIPTVGCVTAMEILNE FPGHGLEPLLKFSEWWHEAQKNPKIRPNPHDTKVKKKLRTLQLTPGFPNPAVAEA YLKPVVDDSKGSFLWGKPDLDKIREFCQRYFGWNRTKTDESLFPVLKQLDAQQTQ LRIDSFFRLAQQEKEDAKRIKSQRLNRAVTCMLRKEKEAAASEIEAVSVAMEKEF ELLDKAKRKTQKRGITNTLEESSSLKRKRLSDSKRKNTCGGFLGETCLSESSDGS SSEDAESSSLMNVQRRTAAKEPKTSASDSQNSVKEAPVKNGGATTSSSSDSDDDG GKEKMVLVTARSVFGKKRRKLRRARGRKRKT	SEQ ID NO: 1361547

[0591] В различных вариантах осуществления указанные здесь слитые белки праймерного редактора могут включать любой вариант флэп-эндонуклеазы описанных выше последовательностей, имеющий аминокислотную последовательность, которая, по меньшей мере, приблизительно на 70% идентична, по меньшей мере, приблизительно на 80% идентична, по меньшей мере, приблизительно на 90% идентична, по меньшей мере, приблизительно на 95% идентична, по меньшей мере, приблизительно на 96% идентична, по меньшей мере, приблизительно на 97% идентична, по меньшей мере, приблизительно на 98% идентична, по меньшей мере, приблизительно на 99% идентична, по меньшей

мере, приблизительно на 99,5% идентична, или, по меньшей мере, приблизительно на 99,9% идентична любой из указанных выше последовательностей.

[0592] Другие эндонуклеазы, которые могут использоваться в настоящих способах для обеспечения удаления 5'-концевого одноцепочечного ДНК-флэпа, включают, но без ограничения, эндонуклеазу (1) trex 2, (2) exo1 (например Keijzers et al., Biosci Rep. 2015, 35(3): e00206)

[0593] Trex 2

[0594] 3’ праймерная репарационная экзонуклеаза 2 (TREX2) - человек

[0595] Accession No. NM_080701

[0596] MSEAPRAETFVFLDLEATGLPSVEPEIAELSLFAVHRSSLENPEHDESGALVLPRV LDKLTLCMCPERPFTAKASEITGLSSEGLARCRKAGFDGAVVRTLQAFLSRQAGPICLVA HNGFDYDFPLLCAELRRLGARLPRDTVCLDTLPALRGLDRAHSHGTRARGRQGYSLGS LFHRYFRAEPSAAHSAEGDVHTLLLIFLHRAAELLAWADEQARGWAHIEPMYLPPDDPS LEA (SEQ ID NO: 1361665).

[0597] 3’ праймерная репарационная экзонуклеаза 2 (TREX2) - мышь

[0598] Кат. # NM_011907

[0599] MSEPPRAETFVFLDLEATGLPNMDPEIAEISLFAVHRSSLENPERDDSGSLVLPRV LDKLTLCMCPERPFTAKASEITGLSSESLMHCGKAGFNGAVVRTLQGFLSRQEGPICLVA HNGFDYDFPLLCTELQRLGAHLPQDTVCLDTLPALRGLDRAHSHGTRAQGRKSYSLASL FHRYFQAEPSAAHSAEGDVHTLLLIFLHRAPELLAWADEQARSWAHIEPMYVPPDGPSL EA (SEQ ID NO: 1361666)

[0600] 3’ праймерная репарационная экзонуклеаза 2 (TREX2) - крыса

[0601] Кат. # NM_001107580

[0602] MSEPLRAETFVFLDLEATGLPNMDPEIAEISLFAVHRSSLENPERDDSGSLVLPRV LDKLTLCMCPERPFTAKASEITGLSSEGLMNCRKAAFNDAVVRTLQGFLSRQEGPICLV AHNGFDYDFPLLCTELQRLGAHLPRDTVCLDTLPALRGLDRVHSHGTRAQGRKSYSLA SLFHRYFQAEPSAAHSAEGDVNTLLLIFLHRAPELLAWADEQARSWAHIEPMYVPPDGP SLEA (SEQ ID NO: 1361667)

[0603] ExoI

[0604] Экзонуклеаза 1 человека (EXO1) принимает участие во многих различных метаболических процессах ДНК, включая репарацию несовпадений ДНК (MMR),

опосредованное микрогомологией соединение концов, гомологичную рекомбинацию (HR), и репликацию. EXO1 человека принадлежит к семейству эукариотических нуклеаз, Rad2/XPG, которые также включают FEN1 и GEN1. Семейство Rad2/XPG является консервативным в рамках нуклеазного домена у видов от фага до человека. Продукт гена EXO1 демонстрирует активность как 5' экзонуклеазы, так и 5' флэпа. Кроме того, EXO1 содержит собственную активность 5' РНКазы H. EXO1 человека имеет высокую аффинность в плане процессирования двуцепочечной ДНК (дцДНК), разрывов, гэпов, псевдо-Y структур, и может корректировать Соединения Холлидея за счет собственной флэп-активности. EXO1 человека принимает участие в MMR и содержит консервативные домены связывания, взаимодействующие напрямую с MLH1 и MSH2. Нуклеолитическая активность положительно стимулируется PCNA, MutSα (комплексом MSH2/MSH6), 14-3-3, MRN и комплексом 9-1-1.

[0605] экзонуклеаза 1 (EXO1) Кат. # NM_003686 (Homo sapiens exonuclease 1 (EXO1), транскриптный вариант 3) – isoform A

[0606] MGIQGLLQFIKEASEPIHVRKYKGQVVAVDTYCWLHKGAIACAEKLAKGEPTDR YVGFCMKFVNMLLSHGIKPILVFDGCTLPSKKEVERSRRERRQANLLKGKQLLREGKVS EARECFTRSINITHAMAHKVIKAARSQGVDCLVAPYEADAQLAYLNKAGIVQAIITEDSD LLAFGCKKVILKMDQFGNGLEIDQARLGMCRQLGDVFTEEKFRYMCILSGCDYLSSLRG IGLAKACKVLRLANNPDIVKVIKKIGHYLKMNITVPEDYINGFIRANNTFLYQLVFDPIKR KLIPLNAYEDDVDPETLSYAGQYVDDSIALQIALGNKDINTFEQIDDYNPDTAMPAHSRS HSWDDKTCQKSANVSSIWHRNYSPRPESGTVSDAPQLKENPSTVGVERVISTKGLNLPR KSSIVKRPRSAELSEDDLLSQYSLSFTKKTKKNSSEGNKSLSFSEVFVPDLVNGPTNKKSV STPPRTRNKFATFLQRKNEESGAVVVPGTRSRFFCSSDSTDCVSNKVSIQPLDETAVTDK ENNLHESEYGDQEGKRLVDTDVARNSSDDIPNNHIPGDHIPDKATVFTDEESYSFESSKF TRTISPPTLGTLRSCFSWSGGLGDFSRTPSPSPSTALQQFRRKSDSPTSLPENNMSDVSQL KSEESSDDESHPLREEACSSQSQESGEFSLQSSNASKLSQCSSKDSDSEESDCNIKLLDSQS DQTSKLRLSHFSKKDTPLRNKVPGLYKSSSADSLSTTKIKPLGPARASGLSKKPASIQKR KHHNAENKPGLQIKLNELWKNFGFKKF (SEQ ID NO: 1361668)

[0607] экзонуклеаза 1 (EXO1) Кат. # NM_006027 (экзонуклеаза 1 Homo sapiens (EXO1), транскриптный вариант 3) – изоформа B

[0608] MGIQGLLQFIKEASEPIHVRKYKGQVVAVDTYCWLHKGAIACAEKLAKGEPTDR YVGFCMKFVNMLLSHGIKPILVFDGCTLPSKKEVERSRRERRQANLLKGKQLLREGKVS EARECFTRSINITHAMAHKVIKAARSQGVDCLVAPYEADAQLAYLNKAGIVQAIITEDSD LLAFGCKKVILKMDQFGNGLEIDQARLGMCRQLGDVFTEEKFRYMCILSGCDYLSSLRG IGLAKACKVLRLANNPDIVKVIKKIGHYLKMNITVPEDYINGFIRANNTFLYQLVFDPIKR KLIPLNAYEDDVDPETLSYAGQYVDDSIALQIALGNKDINTFEQIDDYNPDTAMPAHSRS HSWDDKTCQKSANVSSIWHRNYSPRPESGTVSDAPQLKENPSTVGVERVISTKGLNLPR KSSIVKRPRSAELSEDDLLSQYSLSFTKKTKKNSSEGNKSLSFSEVFVPDLVNGPTNKKSV STPPRTRNKFATFLQRKNEESGAVVVPGTRSRFFCSSDSTDCVSNKVSIQPLDETAVTDK ENNLHESEYGDQEGKRLVDTDVARNSSDDIPNNHIPGDHIPDKATVFTDEESYSFESSKF TRTISPPTLGTLRSCFSWSGGLGDFSRTPSPSPSTALQQFRRKSDSPTSLPENNMSDVSQL KSEESSDDESHPLREEACSSQSQESGEFSLQSSNASKLSQCSSKDSDSEESDCNIKLLDSQS DQTSKLRLSHFSKKDTPLRNKVPGLYKSSSADSLSTTKIKPLGPARASGLSKKPASIQKR KHHNAENKPGLQIKLNELWKNFGFKKDSEKLPPCKKPLSPVRDNIQLTPEAEEDIFNKPE CGRVQRAIFQ (SEQ ID NO: 1361669)

[0609] экзонуклеаза 1 (EXO1) Кат. # NM_001319224 (экзонуклеаза 1 Homo sapiens (EXO1), транскриптный вариант 4) – изоформа C

[0610] MGIQGLLQFIKEASEPIHVRKYKGQVVAVDTYCWLHKGAIACAEKLAKGEPTDR YVGFCMKFVNMLLSHGIKPILVFDGCTLPSKKEVERSRRERRQANLLKGKQLLREGKVS EARECFTRSINITHAMAHKVIKAARSQGVDCLVAPYEADAQLAYLNKAGIVQAIITEDSD LLAFGCKKVILKMDQFGNGLEIDQARLGMCRQLGDVFTEEKFRYMCILSGCDYLSSLRG IGLAKACKVLRLANNPDIVKVIKKIGHYLKMNITVPEDYINGFIRANNTFLYQLVFDPIKR KLIPLNAYEDDVDPETLSYAGQYVDDSIALQIALGNKDINTFEQIDDYNPDTAMPAHSRS HSWDDKTCQKSANVSSIWHRNYSPRPESGTVSDAPQLKENPSTVGVERVISTKGLNLPR KSSIVKRPRSELSEDDLLSQYSLSFTKKTKKNSSEGNKSLSFSEVFVPDLVNGPTNKKSVS TPPRTRNKFATFLQRKNEESGAVVVPGTRSRFFCSSDSTDCVSNKVSIQPLDETAVTDKE NNLHESEYGDQEGKRLVDTDVARNSSDDIPNNHIPGDHIPDKATVFTDEESYSFESSKFT RTISPPTLGTLRSCFSWSGGLGDFSRTPSPSPSTALQQFRRKSDSPTSLPENNMSDVSQLK SEESSDDESHPLREEACSSQSQESGEFSLQSSNASKLSQCSSKDSDSEESDCNIKLLDSQSD QTSKLRLSHFSKKDTPLRNKVPGLYKSSSADSLSTTKIKPLGPARASGLSKKPASIQKRK

HHNAENKPGLQIKLNELWKNFGFKKDSEKLPPCKKPLSPVRDNIQLTPEAEEDIFNKPEC GRVQRAIFQ (SEQ ID NO: 1361670)

(iv) Интеины и разделенные интеины

[0611] Будет являться очевидным, что в некоторых вариантах осуществления (например при доставке праймерного редактора in vivo с использованием частиц AAV) может являться полезным разделять полипептид (например дезаминазу или napDNAbp) или слитый белок (например праймерный редактор) на N-концевую половину и С-концевую половину и доставлять их по отдельности с последующим восстановлением полного белка (или слитого белка в соответствующем случае) за счет их солокализации в клетке. Каждая из отдельных половин белка или слитого белка может включать метку в виде разделенного интеина для осуществления восстановления полного белка или слитого белка за счет механизма транс-сплайсинга белка.

[0612] Транс-сплайсинг белка, катализируемый разделенными интеинами, представляет собой полностью ферментативный способ лигирования белка. Разделенный интеин представляет собой по существу непрерывный интеин (например мини-интеин), разделенный на две части, называемые N-интеином и С-интеином, соответственно. N-интеин и С-интеин разделенного интеина могут нековалентно связываться для образования активного интеина и катализировать реакцию сплайсинга по существу так же, как и непрерывный интеин. Разделенные интеины были обнаружены в природе, а также сконструированы в лабораториях. В соответствии с используемым здесь значением термин "разделенный интеин" обозначает любой интеин, в котором присутствует один или несколько разрывов пептидных связей между N-концевыми и С-концевыми последовательностями аминокислот, так что N-концевые и С-концевые последовательности становятся отдельными молекулами, которые могут нековалентно связываться или восстанавливаться с образованием интеина, который является функциональным для реакций транс-сплайсинга белка. Для получения разделенного интеина для использования в рамках способов по изобретению может применяться любой каталитически активный интеин или его фрагмент. Например, в одном аспекте разделенный интеин может быть получен из эукариотического интеина. В другом аспекте разделенный интеин может быть получен из бактериального интеина. В другом аспекте разделенный интеин может быть получен из архейного интеина. Предпочтительно,

полученный таким образом разделенный интеин будет иметь только аминокислотные последовательности, необходимые для катализирования реакций транс-сплайсинга.

[0613] В соответствии с используемым здесь значением термин "N-концевой разделенный интеин (In)" обозначает любую последовательность интеина, которая включает N-концевую аминокислотную последовательность, которая является функциональной для реакций транс-сплайсинга. Таким образом, In также включает последовательность, которая устраняется при осуществлении транс-сплайсинга. In может включать последовательность, которая представляет собой модификацию N-концевой части встречающейся в природе последовательности интеина. Например, In может включать дополнительные аминокислотные остатки и/или мутированные остатки, если включение подобных дополнительных и/или мутированных остатков не делает In нефункциональным в плане транс-сплайсинга. Предпочтительно, включение дополнительных и/или мутированных остатков повышает или стимулирует активность In, связанную с транс-сплайсингом.

[0614] В соответствии с используемым здесь значением термин "C-концевой разделенный интеин (Ic)" обозначает любую последовательность интеина, которая включает C-концевую аминокислотную последовательность, которая является функциональной для реакций транс-сплайсинга. В одном аспекте Ic включает от 4 до 7 последовательных аминокислотных остатков, по меньшей мере, 4 аминокислоты из которых происходят из последней β-цепи интеина, из которого он был получен. Таким образом, Ic также включает последовательность, которая устраняется при осуществлении транс-сплайсинга. Ic может включать последовательность, которая представляет собой модификацию C-концевой части встречающейся в природе последовательности интеина. Например, Ic может включать дополнительные аминокислотные остатки и/или мутированные остатки, если включение подобных дополнительных и/или мутированных остатков не делает Ic нефункциональным в плане транс-сплайсинга. Предпочтительно, включение дополнительных и/или мутированных остатков повышает или стимулирует активность Ic, связанную с транс-сплайсингом.

[0615] В некоторых вариантах осуществления изобретения пептид, связанный с Ic или In, может включать дополнительный химический фрагмент, включая, помимо прочего, флуоресцентные группы, биотин, полиэтиленгликоль (ПЭГ), аналоги аминокислот,

искусственные аминокислоты, фосфатные группы, гликозильные группы, радиоизотопные метки, и фармацевтические молекулы. В других вариантах осуществления пептид, связанный с Ic, может включать одну или несколько химически реактивных групп, включая, среди прочих, кетон, альдегид, остатки Cys и остатки Lys. N-интеин и C-интеин разделенного интеина могут нековалентно связываться для образования активного интеина и катализировать реакцию сплайсинга при присутствии "полипептида сплайсинга интеина (ISP)". В соответствии с используемым здесь значением термин "полипептид сплайсинга интеина (ISP)" обозначает часть аминокислотной последовательности разделенного интеина, которая сохраняется, когда Ic, In, или они оба удаляются из разделенного интеина. В определенных вариантах осуществления In включает ISP. В другом варианте осуществления Ic включает ISP. В другом варианте осуществления ISP представляет собой отдельный пептид, который не является ковалентно связанным ни с In, ни с Ic.

[0616] Разделенные интеины могут быть созданы из непрерывных интеинов путем конструирования одного или нескольких участков разделения в неструктурированной петле или внесения аминокислотной последовательности между 12 консервативными бета-цепями, находящимися в структуре мини-интеинов. Может присутствовать некоторая гибкость в плане расположения участка разрыва в участках между бета-цепями при условии, что создание разрыва не будет нарушать структуру интеина, в частности структурированных бета-цепей, до значительной степени, при которой теряется активность, связанная со сплайсингом белка.

[0617] При транс-сплайсинге белка один белок-предшественник состоит из части в виде N-экстеина и далее N-интеина, другой белок-предшественник состоит из C-интеина и далее части в виде C-экстеина, при этом реакция транс-сплайсинга (катализируемая совместно N- и C-интеинами) обеспечивает эксцизию данных двух последовательностей интеина и соединение данных двух последовательностей экстеина с помощью пептидной связи. Транс-сплайсинг белка, являясь ферментативной реакцией, может работать при очень низких (например микромолярных) концентрациях белков и может осуществляться при физиологических условиях.

[0618] Примеры последовательностей являются следующими:

[0619] НАЗВАНИЕ ПОСЛЕДОВАТЕЛЬНОСТИ ЛИГАНД-ЗАВИСИМОГО ИНТЕИНА

[0620] ИНТЕИН 2-4:

[0621] CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVVSWFD QGTRDVIGLRIAGGAIVWATPDHKVLTEYGWRAAGELRKGDRVAGPGGSGNSLALSLT ADQMVSALLDAEPPILYSEYDPTSPFSEASMMGLLTNLADRELVHMINWAKRVPGFVD LTLHDQAHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDM LLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRALDKITDTLIHL MAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKYKNVVPLYDLLLEMLDA HRLHAGGSGASRVQAFADALDDKFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHT LVAEGVVVHNC (SEQ ID NO: 1361671)

[0622] ИНТЕИН 3-2:

CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAVAKDGTLLARPVVSWFDQGTR DVIGLRIAGGAIVWATPDHKVLTEYGWRAAGELRKGDRVAGPGGSGNSLALSLTADQ MVSALLDAEPPILYSEYDPTSPFSEASMMGLLTNLADRELVHMINWAKRVPGFVDLTLH DQAHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLAT SSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRALDKITDTLIHLMAK AGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKYTNVVPLYDLLLEMLDAHRLH AGGSGASRVQAFADALDDKFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAE GVVVHNC (SEQ ID NO: 1361672)

[0623] ИНТЕИН 30R3-1:

CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVVSWFDQGTR DVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKGDRVAGPGGSGNSLALSLTADQ MVSALLDAEPPIPYSEYDPTSPFSEASMMGLLTNLADRELVHMINWAKRVPGFVDLTLH DQAHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLAT SSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRALDKITDTLIHLMAK AGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKYKNVVPLYDLLLEMLDAHRL HAGGSGASRVQAFADALDDKFLHDMLAEGLRYSVIREVLPTRRARTFDLEVEELHTLV AEGVVVHNC (SEQ ID NO: 1361673)

[0624] ИНТЕИН 30R3-2:

CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVVSWFDQGTR DVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKGDRVAGPGGSGNSLALSLTADQ MVSALLDAEPPILYSEYDPTSPFSEASMMGLLTNLADRELVHMINWAKRVPGFVDLTLH

DQAHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLAT
SSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRALDKITDTLIHLMAK
AGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKYKNVVPLYDLLLEMLDAHRL
HAGGSGASRVQAFADALDDKFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVA
EGVVVHNC (SEQ ID NO: 1361674)

[0625] ИНТЕИН 30R3-3:

CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVVSWFDQGTR
DVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKGDRVAGPGGSGNSLALSLTADQ
MVSALLDAEPPIPYSEYDPTSPFSEASMMGLLTNLADRELVHMINWAKRVPGFVDLTLH
DQAHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLAT
SSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRALDKITDTLIHLMAK
AGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKYKNVVPLYDLLLEMLDAHRL
HAGGSGASRVQAFADALDDKFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVA
EGVVVHNC (SEQ ID NO: 1361675)

[0626] ИНТЕИН 37R3-1:

CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVVSWFDQGTR
DVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKGDRVAGPGGSGNSLALSLTADQ
MVSALLDAEPPILYSEYNPTSPFSEASMMGLLTNLADRELVHMINWAKRVPGFVDLTLH
DQAHLLERAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLAT
SSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRALDKITDTLIHLMAK
AGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKYKNVVPLYDLLLEMLDAHRL
HAGGSGASRVQAFADALDDKFLHDMLAEGLRYSVIREVLPTRRARTFDLEVEELHTLV
AEGVVVHNC (SEQ ID NO: 1361676)

[0627] ИНТЕИН 37R3-2:

CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVVSWFDQGTR
DVIGLRIAGGAIVWATPDHKVLTEYGWRAAGELRKGDRVAGPGGSGNSLALSLTADQ
MVSALLDAEPPILYSEYDPTSPFSEASMMGLLTNLADRELVHMINWAKRVPGFVDLTLH
DQAHLLERAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLAT
SSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRALDKITDTLIHLMAK
AGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKYKNVVPLYDLLLEMLDAHRL

HAGGSGASRVQAFADALDDKFLHDMLAEGLRYSVIREVLPTRRARTFDLEVEELHTLV AEGVVVHNC (SEQ ID NO: 1361677)

[0628] ИНТЕИН 37R3-3:

CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAVAKDGTLLARPVVSWFDQGTR DVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKGDRVAGPGGSGNSLALSLTADQ MVSALLDAEPPILYSEYDPTSPFSEASMMGLLTNLADRELVHMINWAKRVPGFVDLTLH DQAHLLERAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLAT SSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRALDKITDTLIHLMAK AGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKYKNVVPLYDLLLEMLDAHRL HAGGSGASRVQAFADALDDKFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVA EGVVVHNC (SEQ ID NO: 13616718)

[0629] Несмотря на то, что интеины наиболее часто представлены в виде последовательного домена, некоторые из них существуют в естественной разделенной форме. В этом случае два фрагмента экспрессируются в виде отдельных полипептидов и должны быть связаны перед осуществлением сплайсинга, так называемого транс-сплайсинга белка.

[0630] Примером разделенного интеина является интеин Ssp DnaE, который включает две субъединицы, а именно DnaE-N и DnaE-C. Две различных субъединицы кодируются отдельными генами, а именно dnaE-n и dnaE-c, которые кодируют субъединицы DnaE-N и DnaE-C, соответственно. DnaE представляет собой встречающийся в природе разделенный интеин в Synechocytis sp. PCC6803 и способен направлять транс-сплайсинг двух отдельных белков, каждый из которых включает продукт слияния с DnaE-N или DnaE-C.

[0631] Дополнительные встречающиеся в природе или сконструированные последовательности разделенного интеина известны или могут быть реализованы на основе описанных здесь или доступных в данной области последовательностей цельного интеина. Примеры последовательностей разделенного интеина можно найти в Stevens et al., "A promiscuous split intein with expanded protein engineering applications," PNAS, 2017, Vol.114: 8538-8543; Iwai et al., "Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostc punctiforme, FEBS Lett, 580: 1853-1858, каждый из которых приведен здесь в качестве ссылки. Дополнительные последовательности разделенного интеина

можно найти, например, в WO 2013/045632, WO 2014/055782, WO 2016/069774, и EP2877490, содержание каждой из которых приведено здесь в качестве ссылки.

[0632] Кроме того, сплайсинг белка in trans был описан in vivo и in vitro (Shingledecker, et al., [0076] Gene 207:187 (1998), Southworth, et al., EMBO J. 17:918 (1998); Mills, et al., Proc. Natl. Acad. Sci. USA, 95:3543-3548 (1998); Lew, et al., J. Biol. Chem., 273:15887-15890 (1998); Wu, et al., Biochim. Biophys. Acta 35732:1 (1998b), Yamazaki, et al., J. Am. Chem. Soc. 120:5591 (1998), Evans, et al., J. Biol. Chem. 275:9091 (2000); Otomo, et al., Biochemistry 38:16040-16044 (1999); Otomo, et al., J. Biolmol. NMR 14:105-114 (1999); Scott, et al., Proc. Natl. Acad. Sci. USA 96:13638-13643 (1999)), при этом он обеспечивает возможность экспрессировать белок в виде двух неактивных фрагментов, которые далее подвергаются лигирования для образования функционального продукта, например как показано на Фиг.66 и 67 в плане образования полного слитого белка PE из двух отдельно экспрессируемых половин.

(v) Система рекрутирования РНК-белок

[0633] В различных вариантах осуществления два отдельных домена белка (например домен Cas9 и домен полимеразы) могут быть солокализованы друг с другом для образования функционального комплекса (аналогично функции слитого белка, включающего два данных отдельных домена белка) путем использования "системы рекрутирования РНК-белок", например "способа мечения MS2". Такие системы в целом обеспечивают мечение одного домена белка "доменом взаимодействия РНК-белок" (также известным как "домен рекрутирования РНК-белка"), а другого — "РНК-связывающим белком", который специфично распознает и связывается с доменом взаимодействия РНК-белок, например определенной структурой в виде шпильки. Данные типы систем могут быть использованы для солокализации доменов праймерного редактора, а также рекрутирования дополнительных видов функциональности к праймерному редактору, например домена UGI. В одном варианте осуществления система мечения MS2 основана на естественном взаимодействии белка оболочки бактериофага MS2 ("MCP" или "MS2cp") со структурой в виде шпильки, присутствующей в геноме фага, то есть "шпилькой MS2". В случае шпильки MS2 она распознается и связывается белком оболочки бактериофага MS2 (MCP). Таким образом, в одном примере сценария продукт слияния в виде дезаминазы-MS2 может рекрутировать продукт слияния Cas9-MCP.

[0634] Другие модулярные домены взаимодействия РНК-белок описаны в данной области, например в Johansson et al. , "RNA recognition by the MS2 phage coat protein," Sem Virol., 1997, Vol. 8(3): 176-185; Delebecque et al., "Organization of intracellular reactions with rationally designed RNA assemblies," Science, 2011, Vol. 333: 470-474; Mali et al., "Cas9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering," Nat. Biotechnol., 2013, Vol.31: 833-838; и Zalatan et al., "Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds," Cell, 2015, Vol.160: 339-350, каждый из которых полностью приведен здесь в качестве ссылки. Другие системы включают шпильку PP7, которая специфически рекрутирует белок PCP, и шпильку "com", которая специфически рекрутирует белок Com. См. Zalatan et al.

[0635] Нуклеотидная последовательность шпильки MS2 (или также обозначаемой как "аптамер MS2") представляет собой: GCCAACATGAGGATCACCCATGTCTGCAGGGCC (SEQ ID NO: 1361679).

[0636] Аминокислотная последовательность MCP или MS2cp является следующей: GSASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQNR KYTIKVEVPKVATQTVGGEELPVAGWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDG NPIPSAIAANSGIY (SEQ ID NO: 1361680).

(vi) Домен UGI

[0637] В других вариантах осуществления описанные здесь праймерные редакторы могут включать один или несколько доменов ингибитора урацилгликозилазы. Термин "ингибитор урацилгликозилазы (UGI)" или "домен UGI" в соответствии с используемым здесь значением обозначает белок, который способен ингибировать фермент эксцизионной репарации оснований урацил-ДНК-гликозилазы. В некоторых вариантах осуществления домен UGI включает UGI дикого типа или UGI согласно SEQ ID NO: 1361681. В некоторых вариантах осуществления предусмотренные здесь белки UGI включают фрагменты UGI и белки, являющиеся гомологичными UGI или фрагменту UGI. Например, в некоторых вариантах осуществления домен UGI включает фрагмент аминокислотной последовательности, указанной в SEQ ID NO: 1361681. В некоторых вариантах осуществления фрагмент UGI включает аминокислотную последовательность, которая включает, по меньшей мере, 60%, по меньшей мере, 65%, по меньшей мере, 70%, по меньшей мере, 75%, по меньшей мере, 80%, по меньшей мере, 85%, по меньшей мере,

90%, по меньшей мере, 95%, по меньшей мере, 96%, по меньшей мере, 97%, по меньшей мере, 98%, по меньшей мере, 99%, или, по меньшей мере, 99,5% аминокислотной последовательности, указанной в SEQ ID NO: 1361681. В некоторых вариантах осуществления UGI включает аминокислотную последовательность, являющуюся гомологичной аминокислотной последовательности, указанной в SEQ ID NO: 1361681, или аминокислотную последовательность, являющуюся гомологичной фрагменту аминокислотной последовательности, указанной в SEQ ID NO: 1361681. В некоторых вариантах осуществления белки, включающие UGI, или фрагменты UGI, или гомологи UGI, или фрагменты UGI, обозначаются как "варианты UGI". Вариант UGI имеет такую же гомологию, что и UGI или его фрагмент. Например, вариант UGI является, по меньшей мере, на 70% идентичным, по меньшей мере, на 75% идентичным, по меньшей мере, на 80% идентичным, по меньшей мере, на 85% идентичным, по меньшей мере, на 90% идентичным, по меньшей мере, на 95% идентичным, по меньшей мере, на 96% идентичным, по меньшей мере, на 97% идентичным, по меньшей мере, на 98% идентичным, по меньшей мере, на 99% идентичным, по меньшей мере, на 99,5% идентичным, или, по меньшей мере, на 99,9% идентичным UGI дикого типа или UGI, указанному в SEQ ID NO: 1361681. В некоторых вариантах осуществления вариант UGI включает фрагмент UGI, так что фрагмент является, по меньшей мере, на 70% идентичным, по меньшей мере, на 75% идентичным, по меньшей мере, на 80% идентичным, по меньшей мере, на 85% идентичным, по меньшей мере, на 90% идентичным, по меньшей мере, на 95% идентичным, по меньшей мере, на 96% идентичным, по меньшей мере, на 97% идентичным, по меньшей мере, на 98% идентичным, по меньшей мере, на 99% идентичным, по меньшей мере, на 99,5% идентичным, или, по меньшей мере, на 99,9% идентичными соответствующему фрагменту UGI дикого типа или UGI согласно SEQ ID NO: 1361681. В некоторых вариантах осуществления UGI включает следующую аминокислотную последовательность:

[0638] Ингибитор урацил-ДНК-гликозилазы:

[0639] >sp|P14739|UNGI_BPPB2 MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSD APEYKPWALVIQDSNGENKIKML (SEQ ID NO: 1361681).

[0640] Описанные здесь праймерные редакторы могут включать более одного домена UGI, которые могут быть разделены одним или несколькими линкерами, как описано здесь.

(vii) Дополнительные элементы PE

[0641] В определенных вариантах осуществления описанные здесь праймерные редакторы могут включать ингибитор репарации оснований. Термин "ингибитор репарации оснований" или "IBR" обозначает белок, который способен ингибировать активность фермента репарации нуклеиновой кислоты, например фермента эксцизионной репарации оснований. В некоторых вариантах осуществления IBR представляет собой ингибитор эксцизионной репарации основания OGG. В некоторых вариантах осуществления IBR представляет собой ингибитор эксцизионной репарации основания ("iBER"). Примеры ингибиторов эксцизионной репарации оснований включают ингибиторы APE1, Endo III, Endo IV, Endo V, Endo VIII, Fpg, hOGG1, hNEIL1, T7 EndoI, T4PDG, UDG, hSMUG1, и hAAG. В некоторых вариантах осуществления IBR представляет собой ингибитор Endo V или hAAG. В некоторых вариантах осуществления IBR представляет собой iBER, который может представлять собой каталитическую неактивную гликозилазу или каталитически неактивную деоксигеназу или низкомолекулярный или пептидный ингибитор оксидазы или его варианты. В некоторых вариантах осуществления IBR представляет собой iBER, который может представлять собой ингибитор TDG, ингибитор MBD4, или ингибитор фермента AlkBH. В некоторых вариантах осуществления IBR представляет собой iBER, который включает каталитически неактивный TDG или каталитически неактивный MBD4. Пример каталитически неактивного TDG представляет собой мутант N140A согласно SEQ ID NO: 3872 (TDG человека).

[0642] Некоторые примеры гликозилаз приведены ниже. Каталитически неактивные варианты любого из данных доменов гликозилазы представляют собой iBER, которые могут быть слиты с napDNAbp или доменом полимеразы праймерных редакторов, предусмотренных в рамках данного изобретения.

[0643] OGG (человек)

[0644] MPARALLPRRMGHRTLASTPALWASIPCPRSELRLDLVLPSGQSFRWREQSPAH
WSGVLADQVWTLTQTEEQLHCTVYRGDKSQASRPTPDELEAVRKYFQLDVTLAQLYH
HWGSVDSHFQEVAQKFQGVRLLRQDPIECLFSFICSSNNNIARITGMVERLCQAFGPRLI

QLDDVTYHGFPSLQALAGPEVEAHLRKLGLGYRARYVSASARAILEEQGGLAWLQQLR
ESSYEEAHKALCILPGVGTKVADCICLMALDKPQAVPVDVHMWHIAQRDYSWHPTTSQ
AKGPSPQTNKELGNFFRSLWGPYAGWAQAVLFSADLRQSRHAQEPPAKRRKGSKGPEG
(SEQ ID NO: 1361682)

[0645] MPG (человек)

[0646] MVTPALQMKKPKQFCRRMGQKKQRPARAGQPHSSSDAAQAPAEQPHSSSDAA
QAPCPRERCLGPPTTPGPYRSIYFSSPKGHLTRLGLEFFDQPAVPLARAFLGQVLVRRLPN
GTELRGRIVETEAYLGPEDEAAHSRGGRQTPRNRGMFMKPGTLYVYIIYGMYFCMNISS
QGDGACVLLRALEPLEGLETMRQLRSTLRKGTASRVLKDRELCSGPSKLCQALAINKSF
DQRDLAQDEAVWLERGPLEPSEPAVVAAARVGVGHAGEWARKPLRFYVRGSPWVSV
VDRVAEQDTQA (SEQ ID NO: 1361683)

[0647] MBD4 (человек)

[0648] MGTTGLESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQM
MIKRSSECNPLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISP
QGLKFRSKSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSN
NSNWNLRTRSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKP
KGKVTILKGIPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCIS
DAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDT
FLESEEIGTKVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERR
KTSLYFSSKYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGK
MAIPVLWKFLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKY
PIELHGIGKYGNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENHEKLSLS (SEQ ID
NO: 1361684)

[0649] TDG (человек)

[0650] MEAENAGSYSLQQAQAFYTFPFQQLMAEAPNMAVVNEQQMPEEVPAPAPAQEP
VQEAPKGRKRKPRTTEPKQPVEPKKPVESKKSGKSAKSKEKQEKITDTFKVKRKVDRFN
GVSEAELLTKTLPDILTFNLDIVIIGINPGLMAAYKGHHYPGPGNHFWKCLFMSGLSEVQ
LNHMDDHTLPGKYGIGFTNMVERTTPGSKDLSSKEFREGGRILVQKLQKYQPRIAVFNG
KCIYEIFSKEVFGVKVKNLEFGLQPHKIPDTETLCYVMPSSSARCAQFPRAQDKVHYYIK
LKDLRDQLKGIERNMDVQEVQYTFDLQLAQEDAKKMAVKEEKYDPGYEAAYGGAYG

ENPCSSEPCGFSSNGLIESVELRGESAFSGIPNGQWMTQSFTDQIPSFSNHCGTQEQEEES HA (SEQ ID NO: 1361685)

[0651] В некоторых вариантах осуществления описанные здесь слитые белки могут включать один или несколько гетерологичных доменов белка (например приблизительно или более чем 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, или более доменов помимо других компонентов праймерного редактора). Слитый белок может включать любую дополнительную последовательность белка и, необязательно, последовательность линкера между любыми двумя доменами. Другие примеры элементов, которые могут присутствовать, представляют собой последовательности локализации, например последовательности цитоплазматической локализации, последовательности экспорта, например последовательности ядерного экспорта, или другие последовательности локализации, а также метки последовательностей, которые могут использоваться для солюбилизации, очистки, или детектирования слитых белков.

[0652] Примеры доменов белка, которые могут быть слиты с праймерным редактором или его компонентом (например доменом napDNAbp, полимеразным доменом, или доменом NLS) включают, без ограничения, эпитопные метки и последовательности репортерных генов. Неограничивающие примеры эпитопных меток включают гистидиновые (His) метки, метки V5, метки FLAG, метки гемагглютинина гриппа (HA), метки Myc, метки VSV-G, и тиоредоксиновые (Trx) метки. Примеры репортерных генов включают, но без ограничения, глутатион-5-трансферазу (GST), пероксидазу хрена (HRP), хлорамфениколацетилтрансферазу (CAT), бета-галактозидазу, бета-глюкуронидазу, люциферазу, зеленый флуоресцентный белок (GFP), HcRed, DsRed, голубой флуоресцентный белок (CFP), желтый флуоресцентный белок (YFP), и аутофлуоресцентные белки, включая синий флуоресцентный белок (BFP). Праймерный редактор может быть слит с генной последовательностью, кодирующей белок или фрагмент белка, который связывается с молекулами ДНК или связывается с другими клеточными молекулами, включая, но без ограничения, белок связывания мальтозы (MBP), S-метку, продукты слияния ДНК-связывающего домена Lex A (DBD), продукты слияния ДНК-связывающего домена GAL4, и продукты слияния белка BP16 вируса простого герпеса (HSV). Дополнительные домены, которые могут образовывать часть

праймерного редактора, описаны в публикации патента США 2011/0059502, опубликованной 10 марта 2011 и полностью приведенной здесь в качестве ссылки.

[0653] В одном аспекте изобретения репортерный ген, который включает, но без ограничения, глутатион-5-трансферазу (GST), пероксидазу хрена (HRP), хлорамфениколацетилтрансферазу (CAT), бета-галактозидазу, бета-глюкуронидазу, люциферазу, зеленый флуоресцентный белок (GFP), HcRed, DsRed, голубой флуоресцентный белок (CFP), желтый флуоресцентный белок (YFP), и аутофлуоресцентные белки, включая синий флуоресцентный белок (BFP), может быть внесен в клетку для кодирования генного продукта, который служит в качества маркера для измерения изменения или модифицирования экспрессии генного продукта. В определенных вариантах осуществления изобретения генный продукт представляет собой люциферазу. В дополнительном варианте осуществления изобретения экспрессия генного продукта является пониженной.

[0654] Подходящие предусмотренные здесь белковые метки включают, но без ограничения, метки на основе белка-носителя карбоксила биотина (BCCP), myc-метки, метки на основе кальмодулина, FLAG-метки, гемагглютининовые (HA) метки, полигистидиновые метки, также называемые гистидиновыми метками или His-метками, метки на основе связывающего мальтозу белка (MBP), nus-метки, метки на основе глутатион-S-трансферазы (GST), метки на основе зеленого флуоресцентного белка (GFP), тиоредоксиновые метки, S-метки, Softag (например Softag 1, Softag 3), strep-метки, метки на основе биотинлигазы, метки FlAsH, метки V5 и SBP-метки. Дополнительные подходящие последовательности будут очевидны специалистам в данной области. В некоторых вариантах осуществления слитый белок включает одну или несколько His меток.

[0655] В некоторых вариантах осуществления настоящего изобретения активность системы праймерного редактирования может регулироваться во времени за счет изменения времени нахождения, количества, и/или активности экспрессируемых компонентов системы PE. Например, как описано здесь, PE может быть слит с доменом белка, который способен модифицировать внутриклеточное время полужизни PE. В определенных вариантах осуществления с использованием двух или более векторов (например векторной системы, в рамках которой описанные здесь компоненты

кодируются на двух или более отдельных векторах), активность системы PE может регулироваться во времени путем контролирования времени, за которое доставляются векторы. Например, в некоторых вариантах осуществления вектор, кодирующий нуклеазную систему, может доставлять PE перед вектором, кодирующим матрицу. В других вариантах осуществления вектор, кодирующий PEgRNA, может доставлять направляющую перед вектором, кодирующим систему PE. В некоторых вариантах осуществления векторы, кодирующие систему PE и PEgRNA, доставляются одновременно. В определенных вариантах осуществления одновременно доставляемые векторы обеспечивают доставку в определенное время, например, PE, PEgRNA, и/или компонентов направляющей нРНК второй цепи. В дополнительных вариантах осуществления РНК (такая как, например, нуклеазный транскрипт), транскрибированная из кодирующей последовательности на векторах, может дополнительно включать, по меньшей мере, один элемент, который способен модифицировать внутриклеточное время полужизни РНК и/или модулировать управление трансляцией. В некоторых вариантах осуществления время полужизни РНК может быть повышено. В некоторых вариантах осуществления время полужизни РНК может быть снижено. В некоторых вариантах осуществления элемент может быть способен повышать стабильность РНК. В некоторых вариантах осуществления элемент может быть способен понижать стабильность РНК. В некоторых вариантах осуществления элемент может располагаться в 3' UTR РНК. В некоторых вариантах осуществления элемент может включать сигнал полиаденилирования (PA). В некоторых вариантах осуществления элемент может включать кэп, например конец в виде расположенной выше иРНК или PEgRNA. В некоторых вариантах осуществления РНК может не включать PA, так что она подвергается более быстрому расщеплению в клетке после транскрипции. В некоторых вариантах осуществления элемент может включать, по меньшей мере, один элемент с высоким содержанием AU (ARE). ARE могут быть связаны связывающими ARE белками (ARE-BP) таким образом, который зависит от типа ткани, типа клетки, времени, клеточной локализации, и внешней среды. В некоторых вариантах осуществления дестабилизирующий элемент может промотировать распад РНК, влиять на стабильность РНК, или активировать трансляцию. В некоторых вариантах осуществления ARE может иметь длину в виде от 50 до 150 нуклеотидов. В некоторых вариантах осуществления ARE

может включать, по меньшей мере, одну копию последовательности AUUUA. В некоторых вариантах осуществления, по меньшей мере, один ARE может быть добавлен к 3' UTR РНК. В некоторых вариантах осуществления элемент может представлять собой вирус гепатита Вудчака (WHP).

[0656] Посттрансляционный регуляторный элемент (WPRE), который создает третичную структуру для стимулирования экспрессии из транскрипта. В дополнительных вариантах осуществления элемент представляет собой модифицированную и/или усеченную последовательность WPRE, которая способна стимулировать экспрессию из транскрипта, как описано, например, в Zufferey et al., J Virol, 73(4): 2886-92 (1999) and Flajolet et al., J Virol, 72(7): 6175-80 (1998). В некоторых вариантах осуществления WPRE или его эквивалент может быть добавлен к 3' UTR РНК. В некоторых вариантах осуществления элемент может быть выбран из других мотивов последовательности РНК с высоким содержанием быстро или медленно разлагающихся транскриптов.

[0657] В некоторых вариантах осуществления вектор, кодирующий PE или PEgRNA, может подвергаться саморазрушению за счет расщепления последовательности-мишени, присутствующей на векторе, за счет системы PE. Расщепление может устранять непрерывную транскрипцию PE или PEgRNA из вектора. Несмотря на то, что транскрипция может осуществляться на линеаризованном векторе в течение некоторого количества времени, экспрессируемые транскрипты или белки, подвергаемые внутриклеточному расщеплению, будут иметь меньше времени для продуцирования нецелевых эффектов без непрерывной поддержки на основе экспрессии кодирующих векторов.

F. Комплексы PE

[0658] В различных вариантах осуществления описанные здесь слитые белки праймерного редактора могут образовывать комплексы с описанной здесь PEgRNA. Таким образом, далее здесь приведены комплексы, включающие (i) любой из указанных здесь слитых белков PE и (ii) PEgRNA (например терапевтическую PEgRNA согласно Перечню последовательностей), связанную с napDNAbp (например доменом Cas9) слитого белка PE. Без связи с какой-либо конкретной теорией, слитые белки PE могут быть направлены к требуемому участку-мишени программируемым образом путем конструирования подходящей PEgRNA для специфичного и эффективного нацеливания слитого белка PE

на требуемый геномный участок-мишень для внесения требуемой мутации или иной генетической модификации. Однако в некоторых случаях пригодность участка-мишени для праймерного редактирования может зависеть от присутствия подходящим образом расположенного PAM. Широкая совместимость указанных здесь доменов Cas9 в плане PAM имеет потенциал в плане расширения области нацеливания редакторов оснований для включения участков-мишеней, которые не лежат на расстоянии приблизительно 15 нуклеотидов от стандартной последовательности PAM 5'- NGG-3'. Специалист в данной области будет способен сконструировать подходящую последовательность PEgRNA для нацеливания на требуемую геномную последовательность на основе данного изобретения и знаний в данной области.

[0659] В некоторых вариантах осуществления PE1 образует комплекс с PEgRNA и имеет следующую конструкцию: [NLS]-[Cas9(H840A)]-[линкер]-[MMLV_RT(дикого типа)]. Структура является следующей:

Описание	Последовательность
Комплекс PE1 Cas9(H840A) - MMLV_RT(дикого типа) + PEgRNA	MKRTADGSEFESPKKKRKV**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKK** **NLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE** **EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIE** **GDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK** **NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS** **DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA** **GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL** **RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGA** **SAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIV** **DLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN** **EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS** **GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGIL** **QTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPV** **ENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNR** **GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI** **TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN** **AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLA** **NGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS** **DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI** **DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY** **EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ** **AENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD** **SGGSSGGSSGSETPGTSESATPESSGGSSGGSS***TLNIEDEYRLHETSKEPDVSLGSTWLSDFP* *QAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSP* *WNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFF* *CLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLIL* *LQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTE* *ARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQ* *EIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPP* *CLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQ* *FGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQR* *KAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEI*

Описание	Последовательность
	*YRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAA ITETPDTSTLLIENSSP*SGGSKRTADGSEFEPKKKRKV (SEQ ID NO: 1361582) + PEgRNA . Обозначения: Последовательность ядерной локализации (NLS) Сверху (SEQ ID NO: 1361532); Снизу (SEQ ID NO: 1361541) **Cas9(H840A)** (SEQ ID NO: 1361454) **Линкер длиной 33 аминокислоты** (SEQ ID NO: 1361528) *Обратная транскриптаза M-MLV* (SEQ ID NO: 1361497)

[0660] В некоторых вариантах осуществления PE2 образует комплекс с PEgRNA. Данный комплекс может быть обозначен как "PE3" и имеет следующую структуру: [NLS]-[Cas9(H840A)]-[линкер]-[MMLV_RT(D200N)(T330P)(L603W)(T306K)(W313F)] + PEgRNA . PE3 имеет следующую структуру:

PE3 PE2 + PEgRNA PE2 представляет собой Cas9(H840A)-MMLV_RT D200N T330P L603W T306K W313F	MKRTADGSEFESPKKKRKV**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKK NLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIE GDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGA SAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIV DLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGIL QTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPV ENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNR GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLA NGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ AENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD SGGSSGGSSGSETPGTSESATPESSGGSSGGSS***TLNIEDEYRLHETSKEPDVSLGSTWLSDFP QAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSP WNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFF CLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLIL LQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTE ARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQ EIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPP CLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQ FGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQR KAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEI YRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAA ITETPDTSTLLIENSSP*SGGSKRTADGSEFEPKKKRKV (SEQ ID NO: 1361583)

	+ PEgRNA . Обозначения: Последовательность ядерной локализации (NLS) Сверху (SEQ ID NO: 1361532); Снизу (SEQ ID NO: 1361541) **Cas9(H840A)** (SEQ ID NO: 1361454) **Линкер длиной 33 аминокислоты** (SEQ ID NO: 1361528) *Обратная транскриптаза M-MLV* (SEQ ID NO: 1361514)

[0661] В некоторых вариантах осуществления PEgRNA имеет длину около 15-100 нуклеотидов и включает последовательность из, по меньшей мере, 10 последовательных нуклеотидов, которые комплементарны последовательности-мишени. В некоторых вариантах осуществления направляющая РНК имеет длину 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, или 50 нуклеотидов. В некоторых вариантах осуществления направляющая РНК включает последовательность из 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, или 40 последовательных нуклеотидов, которые комплементарны последовательности-мишени. В некоторых вариантах осуществления последовательность-мишень представляет собой последовательность ДНК. В некоторых вариантах осуществления последовательность-мишень располагается в геноме организма. В некоторых вариантах осуществления организм представляет собой прокариота. В некоторых вариантах осуществления прокариот представляет собой бактерию. В некоторых вариантах осуществления бактерия представляет собой *E. coli*. В некоторых вариантах осуществления организм представляет собой эукариота. В некоторых вариантах осуществления организм представляет собой растение или грибок. В некоторых вариантах осуществления организм представляет собой позвоночное. В некоторых вариантах осуществления позвоночное представляет собой млекопитающее. В некоторых вариантах осуществления млекопитающее представляет собой человека. В некоторых вариантах осуществления организм представляет собой клетку. В некоторых вариантах осуществления клетка представляет собой клетку человека. В некоторых вариантах осуществления клетка представляет собой клетку HEK293T или U2OS.

[0662] В некоторых вариантах осуществления последовательность-мишень включает последовательность, ассоциированную с заболеванием или расстройством. В некоторых вариантах осуществления последовательность-мишень включает точечную мутацию, ассоциированную с заболеванием или расстройством. В некоторых вариантах

осуществления последовательность-мишень включает точечную мутацию T→C. В некоторых вариантах осуществления комплекс обеспечивает дезаминирование точечной мутации-мишени C, при этом дезаминирование приводит к получению последовательности, которая не ассоциирована с заболеванием или расстройством. В некоторых вариантах осуществления точечная мутация-мишень C присутствует в цепи ДНК, которая не является комплементарной направляющей РНК. В некоторых вариантах осуществления последовательность-мишень включает точечную мутацию T→A. В некоторых вариантах осуществления комплекс обеспечивает дезаминирование точечной мутации-мишени A, при этом дезаминирование приводит к получению последовательности, которая не ассоциирована с заболеванием или расстройством. В некоторых вариантах осуществления точечная мутация-мишень A присутствует в цепи ДНК, которая не является комплементарной направляющей РНК.

[0663] В некоторых вариантах осуществления комплекс PE дополнительно включает направляющую РНК для осуществления разрыва второй цепи, что обозначает внесение второго разрыва в месте ниже первого разрыва на не подвергаемой редактированию цепи (то есть первоначального участка разрыва, который обеспечивает получение свободного 3'-конца для использования в рамках праймирования обратной транскриптазы на удлиненной части направляющей РНК). В некоторых вариантах осуществления первый разрыв и второй разрыв располагаются на противоположных цепях. В других вариантах осуществления первый разрыв и второй разрыв располагаются на противоположных цепях. В другом варианте осуществления первый разрыв располагается на не подвергаемой редактированию цепи (то есть цепи, которая образует одноцепочечную часть R-петли), а второй разрыв располагается на цепи-мишени. Второй разрыв расположен, по меньшей мере, на 5 нуклеотидов ниже первого разрыва, или, по меньшей мере, на 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, или 30 или более нуклеотидов ниже первого разрыва. Без связи с теорией, второй разрыв индуцирует клеточные процессы репарации и репликации эндогенной ДНК с замещением не подвергаемой редактированию цепи вместо замещения требуемой подвергаемой редактированию цепи. В некоторых вариантах осуществления подвергаемая редактированию цепь представляет собой не являющуюся мишенью цепь, а не подвергаемая редактированию цепь представляет собой цепь-мишень. В других

вариантах осуществления подвергаемая редактированию цепь представляет собой цепь-мишень, а не подвергаемая редактированию цепь представляет собой не являющуюся мишенью цепь.

[0664] Разрыв второй цепи может быть осуществлен с помощью второй направляющей РНК, которая образует комплекс со слитым белком PE на второй, но расположенной поблизости последовательности протоспейсера, и обеспечивает осуществление разрыва.

[0665] В некоторых вариантах осуществления осуществление разрыва второй цепи может быть реализовано после внесения требуемого изменения. Данная концепция обозначается как "временно́й разрыв второй цепи". Это обеспечивает возможность избежать одновременных разрывов на обеих цепях, которые могут приводить к двуцепочечным разрывам ДНК. Временно́е никирование второй цепи может быть осуществлено различными способами, включая внесение второй направляющей РНК после реализации требуемого изменения.

[0666] В одном варианте осуществления настоящее изобретение предусматривает комплекс "PE3b", что обозначает комплекс PE3 плюс направляющую РНК для никирования второй цепи. Данный комплекс имеет следующую структуру: [NLS]-[Cas9(H840A)]-[линкер]-[MMLV_RT(D200N)(T330P)(L603W)(T306K)(W313F)] + PEgRNA + направляющая РНК для никирования второй цепи. PE3b имеет следующую структуру:

PE3b PE2 + PEgRNA + Направляющая РНК для никирования второй цепи PE2 представляет собой Cas9(H840A)-MMLV_RT D200N T330P L603W T306K W313F	MKRTADGSEFESPKKKRKV**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKK** **NLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE** **EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIE** **GDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK** **NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS** **DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA** **GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL** **RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGA** **SAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIV** **DLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN** **EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS** **GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGIL** **QTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPV** **ENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNR** **GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI** **TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN** **AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLA** **NGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS** **DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI** **DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY** **EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ** **AENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD** **SGGSSGGSSGSETPGTSESATPESSGGSSGGSS***TLNIEDEYRLHETSKEPDVSLGSTWLSDFP* *QAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSP*

	WNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFF *CLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLIL* *LQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTE* *ARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQ* *EIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPP* *CLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQ* *FGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQR* *KAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEI* *YRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAA* *ITETPDTSTLLIENSSP*SGGSKRTADGSEFEPKKKRKV (SEQ ID NO: 1361583) + PEgRNA + направляющая РНК для никирования второй цепи. Обозначения: Последовательность ядерной локализации (NLS) Сверху (SEQ ID NO: 1361532); Снизу (SEQ ID NO: 1361541) **Cas9(H840A)** (SEQ ID NO: 1361454) **Линкер длиной 33 аминокислоты** (SEQ ID NO: 1361528) *Обратная транскриптаза M-MLV* (SEQ ID NO: 1361514)

[0667] Комплексы PE могут доставляться в клетки в виде интактных слитых комплексов. Комплексы PE могут также доставляться в клетки с помощью одного или нескольких векторов экспрессии (например векторов на основе лентивируса или векторов на основе адено-ассоциированного вируса). Например, доставка требуемого комплекса PE может включать один вектор экспрессии, который кодирует слитый белок PE и PEgRNA и, необязательно, направляющую РНК для никирования второй цепи из одного промотора или различных промоторов. В другом примере доставка требуемого комплекса PE может включать два или более векторов экспрессии, которые кодируют слитый белок PE из первого вектора экспрессии и PEgRNA и/или необязательную направляющую РНК для никирования второй цепи из второго вектора экспрессии.

[0668] PEgRNA, которая может быть включена в качестве части данного комплекса, включает любую из терапевтических PEgRNA, включенных в Перечень последовательностей, например последовательности целой PEgRNA согласно SEQ ID NO: 1-135514.

IV. Способы использования и лечение на основе PE

[0669] В другом аспекте изобретение относится к способам редактирования последовательности-мишени ДНК с использованием "праймерного редактора". В соответствии с используемым здесь значением термин "праймерное редактирование" обозначает новый подход к редактированию генов с использованием napDNAbp, полимераз, и специализированных направляющих РНК, как описано в настоящем

описании, примеры которого приведены в рамках вариантов осуществления на Фиг.1A-1H. Праймерное редактирования также может быть описано как "мишень-инициированная обратная транскрипция" (TPRT), поскольку молекула-мишень ДНК используется для праймирования синтеза цепи ДНК полимеразой (например обратной транскриптазой). Использование термина "обратная транскрипция" для обозначения "мишень-инициированной обратной транскрипции" предназначено не для ограничения праймерного редактирования использованием обратных транскриптаз, а для обозначения того, что TPRT или праймерные редакторы могут включать любую полимеразу (например ДНК-зависимую ДНК-полимеразу или РНК-зависимую ДНК-полимеразу). В различных вариантах осуществления праймерное редактирование осуществляется путем контактирования молекулы-мишени ДНК (для которое требуется внести изменение в нуклеотидной последовательности) с программируемым нуклеиновой кислотой ДНК-связывающим белком (napDNAbp) в виде комплекса с направляющей РНК праймерного редактора. Со ссылкой на Фиг.1E, направляющая РНК праймерного редактора включает удлинение на 3′- или 5′-конце направляющей РНК или во внутримолекулярном положении в направляющей РНК и кодирует требуемое нуклеотидное изменение (например изменение одного нуклеотида, вставку, или делецию). На этапе (a) комплекс в виде napDNAbp/удлиненной нРНК контактирует с молекулой ДНК, при этом удлиненная нРНК направляет napDNAbp для связывания с локусом-мишенью. На этапе (b) осуществляется разрыв в одной из цепей ДНК локуса-мишени (например с помощью нуклеазы или химического агента), тем самым обеспечивая создание свободного 3′-конца в одной из цепей локуса-мишени. В определенных вариантах осуществления разрыв создается в цепи ДНК, которая соответствует цепи R-петли, то есть цепи, которая не гибридизируется с последовательностью направляющей РНК, то есть "не являющейся мишенью цепи". Однако разрыв может быть осуществлен в любой одной из цепей. То есть разрыв должен быть осуществлен в "цепи-мишени" R-петли (то есть цепи, гибридизированной с протоспейсером удлиненной нРНК) или "не являющейся мишенью цепи" (то есть цепи, образующей одноцепочечную часть R-петли, которая является комплементарной цепи-мишени). На этапе (c) 3′-конец цепи ДНК (образованный разрывом) взаимодействует с удлиненной частью направляющей РНК для праймирования обратной транскрипции (то есть "праймированной мишенью RT"). В некоторых

вариантах осуществления 3′-конец цепи ДНК гибридизируется с определенным участком связывания праймера на удлиненной части направляющей РНК, то есть "праймирующей последовательностью обратной транскриптазы". На этапе (d) вносится обратная транскриптаза, которая синтезирует одиночную цепь ДНК от 3′-конца праймированного участка к 3′-концу направляющей РНК праймерного редактора. Это образует одноцепочечный ДНК-флэп, включающий требуемое нуклеотидное изменение (например изменение одного основания, вставку, или делецию, или их комбинацию) и являющийся иным образом гомологичным эндогенной ДНК в участке разрыва или являющейся смежной с ним. На этапе (e) napDNAbp и направляющая РНК высвобождаются. Этапы (f) и (g) связаны с коррекцией одноцепочечного ДНК-флэпа, так что требуемое нуклеотидное изменение становится встроенным в локус-мишень. Данный процесс может быть смещен в сторону образования требуемого продукта путем удаления соответствующего 5′ эндогенного ДНК-флэпа, который образуется когда 3′ одноцепочечный ДНК-флэп вносится и гибридизируется с последовательностью эндогенной ДНК. Без связи с теорией, процессы репарации и репликации клеточной эндогенной ДНК обеспечивают коррекцию несовпадающей ДНК для встраивания нуклеотидного изменения (изменений) с образованием требуемого измененного продукта. Процесс также может быть смещен в сторону образования продукта с "никированием второй цепи", как показано в качестве примера на Фиг.1D. Данный процесс может обеспечивать внесение, по меньшей мере, одного или нескольких из следующих генетических изменений: трансверсий, транзиций, делеций, и вставок.

[0670] Термин "праймерный редактор (PE)" или "праймерный редактор" обозначает композиции, используемые в рамках способа геномного редактирования с применением описанной здесь мишень-инициированной обратной транскрипции (TPRT), включая, но без ограничения, napDNAbp, обратные транскриптазы, слитые белки (например включающие napDNAbp и обратные транскриптазы), направляющие РНК праймерного редактора, и комплексы, включающие слитые белки и направляющие РНК праймерного редактора, а также вспомогательные элементы, например компоненты для никирования второй цепи и эндонуклеазы для удаления 5′ эндогенного ДНК-флэпа для обеспечения сдвига процесса праймерного редактирования в сторону образования подвергнутого редактированию продукта.

[0671] В различных вариантах осуществления система праймерного редактора (PE) может включать применение подверженной ошибкам обратной транскриптазы для реализации нацеленного мутагенеза, то есть мутирования только хорошо определенной последовательности ДНК в геноме или другого ДНК-элемента в клетке. На Фиг.22 показана схема примера процесса осуществления направленного мутагенеза с использованием подверженной ошибкам обратной транскриптазы в локусе-мишени с помощью программируемого нуклеиновой кислотой ДНК-связывающего белка (napDNAbp) в виде комплекса с направляющей РНК праймерного редактора. Данный процесс может быть обозначен как вариант осуществления праймерного редактирования для направленного мутагенеза. Направляющая РНК праймерного редактора включает удлинение на 3′- или 5′-конце направляющей РНК или во внутримолекулярном положении в направляющей РНК. На этапе (a) комплекс napDNAbp/нРНК контактирует с молекулой ДНК, при этом нРНК направляет napDNAbp для связывания с подвергаемым мутагенезу локусом-мишенью. На этапе (b) осуществляется разрыв в одной из цепей ДНК локуса-мишени (например с помощью нуклеазы или химического агента), тем самым обеспечивая создание свободного 3′-конца в одной из цепей локуса-мишени. В определенных вариантах осуществления разрыв создается в цепи ДНК, которая соответствует цепи R-петли, то есть цепи, которая не гибридизируется с последовательностью направляющей РНК. На этапе (c) 3′-концевая цепь ДНК взаимодействует с удлиненной частью направляющей РНК для праймирования обратной транскрипции. В некоторых вариантах осуществления 3′-концевая цепь ДНК гибридизируется с определенным участком связывания праймера на удлиненной части направляющей РНК. На этапе (d) вносится подверженная ошибкам обратная транскриптаза, которая синтезирует подвергнутую мутагенезу одиночную цепь ДНК от 3′-конца праймированного участка к 3′-концу направляющей РНК. Примеры мутаций указаны с помощью звездочки "*". Это обеспечивает образование одноцепочечного ДНК-флэпа, включающего требуемый подвергнутый мутагенезу участок. На этапе (e) napDNAbp и направляющая РНК высвобождаются. Этапы (f) и (g) связаны с коррекцией одноцепочечного ДНК-флэпа (включающего подвергнутый мутагенезу участок), так что требуемый подвергнутый мутагенезу участок становится встроенным в локус-мишень. Данный процесс может быть смещен в сторону образования требуемого продукта путем

удаления соответствующего 5′ эндогенного ДНК-флэпа, который образуется когда 3′ одноцепочечный ДНК-флэп вносится и гибридизируется с комплементарной последовательностью на другой цепи. Процесс также может быть смещен в сторону образования продукта с никированием второй цепи, как показано в качестве примера на Фиг.1D. После процессов репарации и/или репликации эндогенной ДНК мутагенизированный участок становится встроенным в обе цепи ДНК локуса ДНК.

[0672] В других вариантах осуществления PEgRNA может использоваться в рамках описанных здесь систем праймерного редактирования для внесения генетического изменения в последовательность-мишень ДНК. Пример такого процесса изображен на Фиг.29. На фигурах приведена схема взаимодействия стандартной PEgRNA с участком-мишенью двуцепочечной ДНК. Двуцепочечная ДНК показана с верхней цепью в направлении от 3′ к 5′ и нижней цепью в направлении от 5′ к 3′. Верхняя цепь включает "протоспейсер" и последовательность PAM и обозначена как "цепь-мишень". Комплементарная нижняя цепь обозначена как "не являющаяся мишенью цепь". Хотя это и не показано, изображенная PEgRNA будет образовывать комплекс с Cas9 или его эквивалентом. Как показано на схеме, спейсер PEgRNA отжигается на комплементарном участке на цепи-мишени, который обозначен как протоспейсер, который расположен непосредственно ниже последовательности PAM и имеет длину приблизительно 20 нуклеотидов. Это взаимодействие образует ДНК/РНК гибрид между спейсерной РНК и ДНК протоспейсера и индуцирует образование R-петли в участке напротив протоспейсера. Как описано в других разделах, затем белок Cas9 (не показан) индуцирует разрыв в не являющейся мишенью цепи, как показано. Это затем ведет к образованию участка в виде 3′ оцДНК-флэпа, который, в соответствии с *z*, взаимодействует с 3′-концом PEgRNA в участке связывания праймера. 3′-конец оцДНК-флэпа (то есть последовательность праймера в виде обратной транскриптазы) отжигается на участке связывания праймера (A) на PEgRNA, тем самым обеспечивая праймирование обратной транскриптазы. Далее обратная транскриптаза (например поступающая *in trans* или поступающая *cis* в виде слитого белка, присоединенного к конструкту Cas9) затем обеспечивает полимеризацию одиночной цепи ДНК, которая кодируется матрицей для редактирования (B) и плечом гомологии (C). Полимеризация продолжается по направлению к 5′-концу плеча удлинения. Полимеризованная цепь оцДНК образует 3′-

концевой оцДНК-флэп, который, как описано в других разделах (например как показано на Фиг.1E) вносится в эндогенную ДНК, вытесняя соответствующую эндогенную цепь (которая удаляется в виде 5′-концевого ДНК-флэпа эндогенной ДНК) и обеспечивая внесение требуемого нуклеотидного изменения (изменения одной пары азотистых оснований нуклеотида, делеций, вставок (включая целые гены) с помощью раундов естественной репарации/репликации ДНК).

[0673] Настоящее изобретение относится к способам лечения субъекта, у которого диагностировано заболевание, ассоциированное или вызванное точечной мутацией, которая может быть скорректирована с помощью предусмотренной здесь системы праймерного редактора (PE). Например, в некоторых вариантах осуществления предусмотрен способ, который включает введение субъекту, имеющему такое заболевание, например рак, ассоциированный с точечной мутацией, как описано выше, эффективного количества описанной здесь системы праймерного редактора(PE), которая корректирует точечную мутацию или вносит деактивирующую мутацию в ассоциированный с заболеванием ген на основе определяемой гомологией репарации в присутствии донорской молекулы ДНК, включающей требуемое генетическое изменение. В некоторых вариантах осуществления предусмотрен способ, который включает введение субъекту, имеющему такое заболевание, например рак, ассоциированный с точечной мутацией, как описано выше, эффективного количества описанной здесь системы праймерного редактора (PE), которая корректирует точечную мутацию или вносит деактивирующую мутацию в ассоциированный с заболеванием ген. В некоторых вариантах осуществления заболевание представляет собой пролиферативное заболевание. В некоторых вариантах осуществления заболевание представляет собой генетическое заболевание. В некоторых вариантах осуществления заболевание представляет собой неопластическое заболевание. В некоторых вариантах осуществления заболевание представляет собой метаболическое заболевание. В некоторых вариантах осуществления заболевание представляет собой лизосомное заболевание накопления. Другие заболевания, которые могут подвергаться лечению путем корректирования точечной мутации или введения деактивирующей мутации в ассоциированный с заболеванием ген, будут известны специалистам в данной области, при этом изобретение не является ограниченным в этом отношении.

[0674] Настоящее изобретение предусматривает способы лечения дополнительных заболеваний или расстройств, например заболеваний или расстройств, которые ассоциированы или вызваны точечной мутацией, которая может быть скорректирована с помощью TPRT-опосредованного редактирования генов. Некоторые подобные заболевания описаны здесь, при этом дополнительные подходящие заболевания, которые могут подвергаться лечению с использованием описанных здесь стратегий и слитых белков, будут очевидны специалистам в данной области на основе настоящего описания. Примеры подходящих заболеваний и расстройств перечислены ниже. Очевидно, что нумерация конкретных положений или остатков в соответствующих последовательностях зависит от конкретного используемого белка и схемы нумерации. Нумерация может различаться, например в предшественниках зрелого белка и самом зрелом белке, при этом различия в последовательностях от одного вида к другому могут влиять на нумерацию. Специалист в данной области будет способен идентифицировать соответствующий остаток в любом гомологичном белке и в соответствующей кодирующей нуклеиновой кислоте с помощью способов, которые хорошо известны в данной области, например путем выравнивания последовательностей и определения гомологичных остатков. Примеры подходящих заболеваний и расстройств включают, без ограничения, такие заболевания и расстройства как: 2-метил-3-гидроксимасляная ацидурия; дефицит 3 бета-гидроксистероиддегидрогеназы; 3-метилглутаконовая ацидурия; дефицит 3-оксо-5-альфа-стероид-дельта-4-дегидрогеназы; 46, реверсия пола XY, тип 1, 3 и 5; дефицит 5-оксопролиназы; дефицит 6-пирувоил-тетрагидроптеринсинтазы; синдром Аарског; синдром Аазе; ахондрогенез 2 типа; ахроматопсия 2 и 7; приобретенный синдром удлиненного интервала QT; акрокаллозальный синдром по типу Шинцеля; акрокапитофеморальная дисплазия; акродизостоз 2 с гормональной резистентностью или без нее; акроэритрокератодермия; акромикрическая дисплазия; Acth-независимая макронодулярная гиперплазия надпочечников 2; синдром активированного PI3K-дельта; острая перемежающаяся порфирия; дефицит семейства ацил-КоА дегидрогеназ, член 9; синдром Адамса-Оливера 5 и 6; дефицит аденинфосфорибозилтрансферазы; дефицит аденилаткиназы; гемолитическая анемия вследствие дефицита аденилосукцинатлиазы; подростковый нефронофтиз; почечно-печеночно-панкреатическая дисплазия; синдром Меккеля тип 7; адренолейкодистрофия; буллезный соединительный эпидермолиз у

взрослых; буллезный эпидермолиз, соединительный, локализованный вариант; цероидный липофусциноз нейронов взрослых; цероидный липофусциноз нейронов взрослых; атаксия у взрослых с глазодвигательной апраксией; синдром ADULT; афибриногенемия и врожденная афибриногенемия; аутосомно-рецессивная агаммаглобулинемия 2; возрастная дегенерация желтого пятна 3, 6, 11 и 12; синдромы Айкарди-Гутьереса 1, 4 и 5; хилбеновая волчанка 1; синдромы Алажилля 1 и 2; болезнь Александра; алькаптонурия; синдром Аллана-Херндона-Дадли; врожденная общая алопеция; энцефалопатия Альперса; дефицит альфа-1-антитрипсина; аутосомно-доминантный, аутосомно-рецессивный и Х-сцепленный рецессивный синдром Альпорта; болезнь Альцгеймера, семейная, 3, со спастическим парапарезом и апраксией; болезнь Альцгеймера, типы 1, 3 и 4; IIA1 несовершенный амелогенез, тип гипокальцификации и тип гипоматурации; дефицит аминоацилазы 1; синдром детской эпилепсии амишей; амилоидогенный транстиретиновый амилоидоз; амилоидная кардиомиопатия, связанная с транстиретином; кардиомиопатия; боковой амиотрофический склероз типов 1, 6, 15 (с лобно-височной деменцией или без нее), 22 (с лобно-височной деменцией или без нее) и 10; лобно-височная деменция с включениями TDP43, связанная с TARDBP; синдром Андерманна; синдром Андерсена-Тавиля; врожденный синдром удлиненного интервала QT; анемия, несфероцитарная гемолитическая, вследствие дефицита G6PD; синдром Ангельмана; тяжелая неонатальная энцефалопатия с микроцефалией; предрасположенность к аутизму, Х-сцепленный 3; ангиопатия, наследственная, с нефропатией, аневризмами и мышечными спазмами; ангиотензин-превращающий фермент, доброкачественное повышение в сыворотке крови; аниридия, мозжечковая атаксия и умственная отсталость; анонихия; дефицит антитромбина III; синдром Антли-Бикслера с генитальными аномалиями и нарушенным стероидогенезом; аневризм аорты, семейная торакальная 4, 6 и 9; аневризмы грудной аорты и расслоение аорты; синдром мультисистемной дисфункции гладких мышц; болезнь моямоя 5; апластическая анемия; очевидный избыток минералокортикоидов; дефицит аргиназы; дефицит аргининосукцинатлиазы; дефицит ароматазы; аритмогенная кардиомиопатия правого желудочка 5, 8 и 10 типов; первичная семейная гипертрофическая кардиомиопатия; врожденный множественный артрогрипоз, дистальный, Х-сцепленный; холестазный синдром на фоне артрогрипоза и почечной дисфункции ; артрогрипоз, нарушение функции почек и холестаз 2; дефицит

аспарагинсинтетазы; нарушение миграции нейронов; атаксия при дефиците витамина E; атаксия сенсорная, аутосомно-доминантная; синдром атаксии-телеангиэктазии; наследственный предрасполагающий к раку синдром; атрансферринемия; фибрилляция предсердий, семейная, 11, 12, 13 и 16; дефекты межпредсердной перегородки 2, 4 и 7 (с дефектами атриовентрикулярной проводимости или без них); остановка предсердий 2; дефект атриовентрикулярной перегородки 4; наследственная луковичная атрофия; синдром ATR-X; аурикуло-мыщелковый синдром 2; аутоиммунное заболевание, мультисистемное, с младенческим началом ; аутоиммунный лимфопролиферативный синдром, тип 1a; аутосомно-доминантная гипогидротическая эктодермальная дисплазия; аутосомно-доминантная прогрессирующая наружная офтальмоплегия с делециями 1 и 3 митохондриальной ДНК; аутосомно-доминантная торсионная дистония 4; аутосомно-рецессивная центроядерная миопатия; аутосомно-рецессивный врожденный ихтиоз 1, 2, 3, 4A и 4B; аутосомно-рецессивная кутислакса типа IA и 1B; синдром аутосомно-рецессивной гипогидротической эктодермальной дисплазии; эктодермальная дисплазия 11b; гипогидротическая / волос / зубов, аутосомно-рецессивная; аутосомно-рецессивная гипофосфатемическая болезнь костей; синдром Аксенфельда-Ригера тип 3; синдром Бейнбриджа-Роперса; синдром Баннаяна-Райли-Рувалькабы; синдром опухоли гамартомы PTEN; синдромы Барайцера-Винтера 1 и 2; синдром Бараката; синдромы Барде-Бидля 1, 11, 16 и 19; синдром голых лимфоцитов тип 2, группа комплементации E; синдром Барттера антенатальный тип 2; синдром Барттера 3, 3 типа с гипокальциурией и 4 типа; кальцификация базальных ганглиев, идиопатическая, 4; монилетрикс; доброкачественная семейная гематурия; доброкачественные семейные неонатальные судороги 1 и 2; судороги, доброкачественные семейные неонатальные 1, и/или миокимия; судороги, ранняя детская эпилептическая энцефалопатия 7; доброкачественные семейные неонатально-младенческие судороги; доброкачественная наследственная хорея; доброкачественная лопаточно-перонеальная мышечная дистрофия с кардиомиопатией; синдром Бернара-Сулье, типы A1 и A2 (аутосомно-доминантный); бестрофинопатия, аутосомно-рецессивная; бета-талассемия; миопатия Бетлема и миопатия Бетлема 2; кристаллическая корнеоретинальная дистрофия Бьетти; нарушение синтеза желчной кислоты, врожденный, 2; дефицит биотинидазы; синдром дисморфизма и умственной отсталости Бирка-Барела; блефарофимоз, птоз и обратный эпикантус; синдром Блума;

синдром Борхесона-Форссмана-Лемана; синдром Буше-Нейхаузера; брахидактилия, типы A1 и A2; брахидактилия при гипертонии; заболевание мелких сосудов головного мозга с кровотечением; дефицит киназы кетоациддегидрогеназы с разветвленной цепью; жаберные синдромы 2 и 3; рак груди с ранним началом; рак груди-яичников, семейный 1, 2 и 4; синдром хрупкой роговицы 2; миопатия Броди ; бронхоэктазия с повышенным содержанием хлорида пота или без него 3; синдром Брауна-Виалетто-Ван Лаэра и синдром Брауна-Виалетто-Ван Лаэра 2; синдром Бругада; синдром Бругада 1; фибрилляция желудочков; пароксизмальная семейная фибрилляция желудочков; синдром Бругада и синдром Бругада 4; синдром удлиненного интервала QT; внезапная сердечная смерть; макулярная дистрофия ("бычий глаз"); болезнь Штаргардта 4; конус-стержневая дистрофия 12; буллезная ихтиозиформная эритродермия; синдром Бёрна-Маккеуна; кандидоз, семейный, 2, 5, 6 и 8; гликопротеиновый синдром дефицита углеводов I и II типа; дефицит карбоангидразы VA, гипераммониемия вследствие нее; рак толстой кишки; аритмия сердца; синдром удлиненного интервала QT, подтип LQT1; кардиоэнцефаломиопатия со смертельным исходом из-за недостаточности цитохром-с-оксидазы; кардиофациально-кожный синдром; кардиомиопатия; болезнь Данона; гипертрофическая кардиомиопатия; кардиомиопатия без компактации левого желудочка; синдром Карневале; комплекс Карни, тип 1; дефицит карнитин-ацилкарнитинтранслоказы; дефицит карнитин-пальмитоилтрансферазы I, II, II (позднее начало) и II (младенческий); катаракта 1, 4, аутосомно-доминантная, аутосомно-доминантная, множественных типов, с микрокорнеей, коппокоподобная, ювенильная, с микрокорнеей и глюкозурией, и ядерная диффузная непрогрессирующая; катехоламинергическая полиморфная желудочковая тахикардия; синдром каудальной регрессии; дефицит Cd8, семейный; заболевание центрального ядра; центромерная нестабильность хромосом 1,9 и 16 и иммунодефицит; мозжечковая атаксия в младенчестве с прогрессирующей внешней офтальмоплегией и мозжечковой атаксией, умственной отсталостью и синдромом дисбаланса 2; церебральная амилоидная ангиопатия, связанная с APP; церебральная аутосомно-доминантная и рецессивная артериопатия с подкорковыми инфарктами и лейкоэнцефалопатией; кавернозные мальформации головного мозга 2; цереброокулофациоскелетный синдром 2; церебро-окуло-фациально-скелетный синдром; цереброретинальная микроангиопатия с

кальцификациями и кистами; цероидный липофусциноз нейронов 2, 6, 7 и 10; синдром Чедиака-Хигаши, взрослый тип; болезнь Шарко-Мари-Тута типов 1B, 2B2, 2C, 2F, 2I, 2U (аксональная), 1C (демиелинизирующая), доминантная промежуточная C, рецессивная промежуточная A, 2A2, 4C, 4D, 4H, IF, IVF и X; спинальная мышечная атрофия лопатки и надкостницы; дистальная спинномозговая мышечная атрофия, врожденная непрогрессирующая; спинальная мышечная атрофия дистальная, аутосомно-рецессивная, 5 ; CHARGE-ассоциация; гипофосфатазия в детстве; гипофосфатазия у взрослых; холецистит; прогрессирующий семейный внутрипеченочный холестаз 3; внутрипеченочный холестаз при беременности 3; болезнь накопления холестанола; дефицит холестеринмонооксигеназы (расщепление боковой цепи); хондродисплазия типа Бломстранда; хондродисплазия точечная 1, X-сцепленная рецессивная и 2 X-сцепленная доминантная; синдром CHOPS; хроническая гранулематозная болезнь, аутосомно-рецессивный цитохром b-положительный, типы 1 и 2; синдром Чадли-Маккалоу; цилиарная дискинезия, первичная, 7, 11, 15, 20 и 22; цитруллинемия I типа; цитруллинемия I и II типа; клейдокраниальный дизостоз; c-подобный синдром; синдром Кокейна типа A,; дефицит коэнзима Q10, первичный 1, 4 и 7; синдром Коффина-Сириса/умственная отсталость; синдром Коффина-Лоури; синдром Коэна; синдром холодового потоотделения 1; синдром Коула-Карпентера 2; комбинированные клеточные и гуморальные иммунные дефекты с гранулемами; комбинированная d-2- и l-2-гидроксиглутаровая ацидурия; комбинированная малоновая и метилмалоновая ацидурия; недостатки комбинированного окислительного фосфорилирования 1, 3, 4, 12, 15 и 25; комбинированный частичный и полный дефицит 17-альфа-гидроксилазы/17,20-лиазы; общий вариабельный иммунодефицит 9; компонент 4 комплемента, частичный дефицит из-за дисфункционального ингибитора c1; дефицит фактора комплемента B; конусный монохроматизм; дистрофия палочек и колбочек 2 и 6; дистрофия палочек и колбочек, несовершенный амелогенез; врожденная гиперплазия надпочечников и Врожденная гипоплазия надпочечников, X-сцепленная; врожденная амегакариоцитарная тромбоцитопения; врожденная аниридия; врожденная центральная гиповентиляция; болезнь Гиршпрунга 3; врожденная контрактурная арахнодактилия; врожденные контрактуры конечностей и лица, гипотония и задержка развития; врожденное нарушение гликозилирования типов 1B, 1D, 1G, 1H, 1J, 1K, 1N, 1P, 2C, 2J, 2K, IIm; врожденная

дизеритропоэтическая анемия I и II типа; врожденная эктодермальная дисплазия лица; врожденная эритропоэтическая порфирия; врожденная генерализованная липодистрофия 2 типа; врожденный порок сердца, несколько типов, 2; врожденный порок сердца; прерванная дуга аорты; врожденный избыточный липоматозный рост, сосудистые аномалии и эпидермальные невусы; немелкоклеточный рак легкого; новообразование яичника; нарушение сердечной проводимости неспецифическое; врожденная микроворсинчатая атрофия; врожденная мышечная дистрофия; врожденная мышечная дистрофия из-за частичного дефицита LAMA2; врожденная мышечная дистрофия-дистрогликанопатия с аномалиями мозга и глаз, типы A2, A7, A8, A11 и A14; врожденная мышечная дистрофия-дистрогликанопатия с умственной отсталостью, типы B2, B3, B5 и B15; врожденная мышечная дистрофия-дистрогликанопатия без умственной отсталости, тип B5; врожденный мышечно-гипертрофический церебральный синдром; врожденный миастенический синдром, чувствительный к ацетазоламиду; врожденная миопатия с диспропорцией типов волокон; врожденная колобома глаза; врожденная стационарная куриная слепота типа 1A, 1B, 1C, 1E, 1F и 2A; копропорфирия; плоская роговица 2; дистрофия роговицы, эндотелий Фукса, 4; эндотелиальная дистрофия роговицы 2 типа; хрупкость роговицы, кератоглобус, синие склеры и гипермобильность суставов; синдромы Корнелии де Ланге 1 и 5; ишемическая болезнь сердца аутосомно-доминантная 2; ишемическая болезнь сердца; гиперальфалипопротеинемия 2; корковая дисплазия, сложная, с другими пороками развития головного мозга 5 и 6; корковые аномалии, затылочные; дефицит кортикостероид-связывающего глобулина; дефицит кортикостеронметилоксидазы 2 типа; синдром Костелло; синдром Каудена 1; болезнь Пертеса; краниодиафизарная дисплазия, аутосомно-доминантная; краниосиностоз 1 и 4; краниосиностоз и стоматологические аномалии; дефицит креатина, X-связанный; синдром Крузона; синдром криптофтальма; крипторхизм, односторонний или двусторонний; симфалангизм Кушинга; кожная злокачественная меланома 1; дряблая кожа при остеодистрофии и тяжелых патологиях легких, желудочно-кишечного тракта и мочевыводящих путей; цианоз, преходящий неонатальный и атипичный нефропатический; муковисцидоз; цистинурия; дефицит цитохром-c-оксидазы I; дефицит цитохром-c оксидазы; D-2-гидроксиглутаровая ацидурия 2; болезнь Дарье, сегментарная; глухота с лабиринтной аплазией, микротией и микродонтией (LAMM) ; глухота,

аутосомно-доминантная 3a, 4, 12, 13, 15, аутосомно-доминантная несиндромная нейросенсорная 17, 20 и 65; глухота, аутосомно-рецессивная 1A, 2, 3, 6, 8, 9, 12, 15, 16, 18b, 22, 28, 31, 44, 49, 63, 77, 86 и 89; глухота, кохлеарная, с миопией и интеллектуальными нарушениями, без вестибулярного поражения, аутосомно-доминантная, X-сцепленная 2; дефицит 2-метилбутирил-КоА дегидрогеназы; дефицит 3-гидроксиацил-КоА дегидрогеназы; дефицит альфа-маннозидазы; дефицит декарбоксилазы ароматических L-аминокислот; дефицит бисфосфоглицератмутазы; дефицит бутирил-КоА дегидрогеназы; дефицит ферроксидазы; дефицит галактокиназы; дефицит гуанидиноацетатметилтрансферазы; дефицит гиалуроноглюкозаминидазы; дефицит рибозо-5-фосфатизомеразы; дефицит стероид-11-бета-монооксигеназы; дефицит УДФ-глюкозо-гексозо-1-фосфатуридилилтрансферазы; дефицит ксантиноксидазы; болезнь Дежерина-Соттаса; болезнь Шарко-Мари-Тута, типы ID и IVF; синдром Дежерина-Соттаса, аутосомно-доминантный; дефицит дендритных клеток, моноцитов, B-лимфоцитов и естественных лимфоцитов-киллеров; дисплазия Десбуки 2; синдром Десбуки; DFNA 2 несиндромная потеря слуха; сахарный и несахарный диабет с атрофией зрительного нерва и глухотой; сахарный диабет 2 типа и инсулинозависимый, 20; анемия Даймонда-Блэкфана 1, 5, 8 и 10; диарея 3 (секреторная, натриевая, врожденная, синдромальная) и 5 (с тафтинговой энтеропатией, врожденная); дикарбоновая аминоацидурия; диффузная ладонно-подошвенная кератодермия ботнического типа; дигитореноцеребральный синдром; дефицит дигидроптеридинредуктазы; дилатационная кардиомиопатия 1A, 1AA, 1C, 1G, 1BB, 1DD, 1FF, 1HH, 1I, 1KK, 1N, 1S, 1Y и 3B; некомпакция левого желудочка 3; нарушение стероидогенеза из-за дефицита оксидоредуктазы цитохрома p450; дистальный артрогрипоз типа 2B; дистальная наследственная мотонейронопатия типа 2B; дистальная миопатия по типу Маркесбери-Григгса; дистальная спинномозговая мышечная атрофия, X-сцепленная 3; дистихиазно-лимфедемный синдром; доминирующий дистрофический буллезный эпидермолиз с отсутствием кожи; доминирующая наследственная атрофия зрительного нерва; синдром Доннаи-Барроу; дефицит дофамин-бета-гидроксилазы; рецептор дофамина d2, снижение плотности в мозге; болезнь Даулинга-Дего 4; сотовая дистрофия сетчатки Дойна; Malattia leventinese; синдром Дуэйна 2 типа; синдром Дубина-Джонсона; мышечная дистрофия Дюшенна; мышечная дистрофия Беккера; дисфибриногенемия; врожденный дискератоз

аутосомно-доминантный и аутосомно-доминантный, 3; врожденный дискератоз, аутосомно-рецессивный, 1, 3, 4 и 5; врожденный дискератоз X-сцепленный; дискинезия, семейная, с миокимией лица; дисплазминогенемия; дистония 2 (скрученность, аутосомно-рецессивная), 3 (скрученность, X-сцепленная), 5 (Dopa-чувствительный тип), 10, 12, 16, 25, 26 (миоклоническая); судороги, доброкачественные семейные детские, 2; ранняя детская эпилептическая энцефалопатия 2, 4, 7, 9, 10, 11, 13 и 14; атипичный синдром Ретта; ранний Т-клеточный предшественник острого лимфобластного лейкоза; синдром хрупкости кожи при эктодермальной дисплазии; синдром эктодермальной дисплазии-синдактилии 1; Ectopia lentis, изолированная аутосомно-рецессивная и доминантная; эктродактилия, эктодермальная дисплазия и синдром расщелины губы/неба 3; синдром Элерса-Данлоса тип 7 (аутосомно-рецессивный), классический тип, тип 2 (прогероидный), с дефицитом гидроксилизина, тип 4, вариант типа 4, и вследствие дефицита тенасцина-X; врожденная мышечная дистрофия типа Эйхсфельда; эндокринная цереброостеодисплазия; синдром повышенных s-колбочек; синдром расширенного вестибулярного акведука; недостаточность энтерокиназы; бородавчатая эпидермодисплазия; простая буллезная эпидермолиза и мышечная дистрофия пояса конечностей, симплекс с пятнистой пигментацией, симплекс с атрезией привратника, симплекс, аутосомно-рецессивный и с атрезией привратника; эпидермолитическая ладонно-подошвенная кератодермия; семейные фебрильные судороги 8; эпилепсия, отсутствие в детстве 2, 12 (идиопатическая генерализованная, предрасположенность) 5 (ночная лобная доля), ночная лобная доля 1 типа, частичная, с переменными очагами, прогрессирующая миоклоническая 3 и X-сцепленная, с различными нарушениями обучаемости и поведенческими расстройствами; эпилептическая энцефалопатия, в детстве, в раннем младенчестве, 1, 19, 23, 25, 30 и 32; эпифизарная дисплазия, множественная, с миопией и кондуктивной глухотой; эпизодическая атаксия 2 типа; эпизодический болевой синдром, семейный, 3; синдром Эпштейна; синдром Фехтнера; эритропоэтическая протопорфирия; устойчивость к эстрогенам; экссудативная витреоретинопатия 6; болезнь Фабри и болезнь Фабри, кардиальный вариант; фактор H, VII, X, v и фактор viii, комбинированный дефицит 2, xiii, субъединица, дефицит; семейный аденоматозный полипоз 1 и 3; семейная амилоидная нефропатия с крапивницей и глухотой; семейная холодовая крапивница; семейная аплазия дермы; семейная доброкачественная пузырчатка; семейный рак груди; восприимчивость к

раку груди; остеосаркома; рак поджелудочной железы 3; семейная кардиомиопатия; семейный аутовоспалительный холодовой синдром 2 ; семейный колоректальный рак; семейная экссудативная витреоретинопатия, Х-сцепленная; семейная гемиплегическая мигрень 1 и 2 типа; семейная гиперхолестеринемия; семейная гипертрофическая кардиомиопатия 1, 2, 3, 4, 7, 10, 23 и 24; семейная гипокалиемия-гипомагниемия; семейная гипоплазия, гломеруло-кистозная почка; семейная инфантильная миастения; семейная юношеская подагра; семейная средиземноморская лихорадка и Семейная средиземноморская лихорадка, аутосомно-доминантная; семейная порэнцефалия; семейная поздняя кожная порфирия; семейный гемангиоматоз легочных капилляров; семейная почечная глюкозурия; семейная почечная гипоурикемия; семейная рестриктивная кардиомиопатия 1; семейная гиперлипопротеинемия 1 и 3 типа; анемия Фанкони, группа комплементации E, I, N и O; синдром Фанкони-Бикеля; восприимчивость к фавизму; фебрильные судороги, семейные, 11; синдром Фейнгольда 1; количественный признак гемоглобина плода, локус 1; синдром FG и синдром FG 4; фиброз экстраокулярных мышц, врожденный, 1, 2, 3a (с поражением глазного яблока или без него), 3b; болезнь рыбьего глаза; дистрофия роговицы Флека; синдром плавающей гавани; фокальная эпилепсия с нарушением речи с умственной отсталостью или без нее; фокальный сегментарный гломерулосклероз 5; дефекты фронтальной части мозга; синдром Франка-Тер-Хаара; синдром Борроне ди Рокко Кровато; синдром Фрейзера; опухоль Вильмса 1; синдром Фримена-Шелдона; лобно-метафизарная дисплазия 1 и 3; лобно-височная деменция; лобно-височная деменция и/или боковой амиотрофический склероз 3 и 4; лобно-височная деменция, связанная с хромосомой 3, и лобно-височная деменция, убиквитин-положительная; дефицит фруктозобифосфатазы; синдром Фурмана; дефицит трансаминазы гамма-аминомасляной кислоты; синдром Гамсторпа-Вольфарта; болезнь Гоше 1 типа и подострая нейронопатия; паралич взгляда, семейный горизонтальный, с прогрессирующим сколиозом; генерализованный доминантный дистрофический буллезный эпидермолиз; генерализованная эпилепсия с фебрильными судорогами плюс 3, тип 1, тип 2; эпилептическая энцефалопатия типа Леннокса-Гасто; гигантская аксональная нейропатия; тромбастения Гланцмана; глаукома 1, открытый угол, e, F и G; глаукома 3, первичная врожденная, d; глаукома, врожденная и глаукома, врожденная, колобома; глаукома, первично открытоугольная, ювенильное начало;

восприимчивость к глиоме 1; синдром дефицита переносчика глюкозы 1 типа; нарушение транспорта глюкозо-6-фосфата; синдром дефицита GLUT1 2; эпилепсия, генерализованная идиопатическая, подверженность, 12; дефицит глутаматформиминотрансферазы; глутаровая ацидемия IIA и IIB; глутаровая ацидурия, тип 1; дефицит глутатионсинтетазы; болезнь накопления гликогена 0 (мышцы), II (взрослая форма), IXa2, IXc, тип 1A; тип II, тип IV, IV (комбинированный печеночный и миопатический), тип V и тип VI; синдром Гольдмана-Фавра; синдром Гордона; синдром Горлина; последовательность голопрозэнцефалии; голопрозэнцефалия 7; гранулематозная болезнь, хроническая, X-сцепленная, вариант; гранулезно-клеточная опухоль яичника; синдром серых тромбоцитов; синдром Грищелли тип 3; дистрофия роговицы Греноува типа I; задержка роста и умственного развития, нижнечелюстной дизостоз, микроцефалия и волчья пасть; дефицит гормона роста при аномалиях гипофиза; нечувствительность к гормону роста при иммунодефиците; дефицит GTP циклогидролазы I; синдром Хайду-Чейни; синдром матки, кисти, и стопы; нарушение слуха; гемангиома, инфантильная капиллярная; гематологическое новообразование; гемохроматоз типа 1, 2B и 3; микрососудистые осложнения диабета 7; количественный признак уровня трансферрина в сыворотке, локус 2; болезнь гемоглобина H неделеционная; гемолитическая анемия, несфероцитарная, вследствие дефицита глюкозо-фосфат-изомеразы; гемофагоцитарный лимфогистиоцитоз, семейный, 2; гемофагоцитарный лимфогистиоцитоз, семейный, 3; дефицит кофактора гепарина II; наследственный энтеропатический акродерматит; синдром наследственного рака груди и яичников; расстройство, подобное атаксии-телеангиэктазии; наследственный диффузный рак желудка; наследственная диффузная лейкоэнцефалопатия со сфероидами; наследственные факторы дефицита II, IX, VIII.; наследственная геморрагическая телеангиэктазия 2 типа ; наследственная нечувствительность к боли при ангидрозе; наследственная лимфедема I типа; наследственная моторная и сенсорная нейропатия с атрофией зрительного нерва; наследственная миопатия с ранней дыхательной недостаточностью; наследственная невралгическая амиотрофия; наследственные неполипозные колоректальные новообразования; синдром Линча I и II; наследственный панкреатит; предрасположенность к хроническому панкреатиту; наследственная сенсорная и вегетативная нейропатия типа IIB и IIA; наследственная сидеробластная анемия; синдром

Германского-Пудлака 1, 3, 4 и 6; гетеротаксия, висцеральная, 2, 4 и 6, аутосомная; гетеротаксия, висцеральная, X-сцепленная; гетеротопия; гистиоцитарный медуллярный ретикулез; синдром гистиоцитоза-лимфаденопатии плюс; дефицит синтетазы холокарбоксилазы; голопрозэнцефалия 2, 3,7 и 9; синдром Холта-Орама; гомоцистеинемия из-за дефицита MTHFR, дефицита CBS и гомоцистинурии, чувствительная к пиридоксину; гомоцистинурия-мегалобластная анемия из-за нарушения метаболизма кобаламина, тип комплементации cblE; синдром Хауэла-Эванса; синдром Гурлера; синдром Хатчинсона-Гилфорда; гидроцефалия; гипераммонемия, тип III; гиперхолестеринемия и гиперхолестеринемия, аутосомно-рецессивные; гиперэкплексия 2 и наследственная гиперэкплексия; синдром гиперферритинемии катаракты; гиперглицинурия; гипериммуноглобулин D при периодической лихорадке; мевалоновая ацидурия; синдром гипериммуноглобулина E; гиперинсулинемическая гипогликемия семейная 3, 4 и 5; синдром гиперинсулинизма-гипераммониемии; гиперлизинемия; гиперманганеземия с дистонией, полицитемией и циррозом печени; синдром гиперорнитинемии-гипераммониемии-гомоцитруллинурии; гиперпаратиреоз 1 и 2; гиперпаратиреоз тяжелой степени у новорожденных; гиперфенилаланинемия, с дефицитом bh4, a, из-за частичного дефицита pts, с дефицитом BH4, D и без pku; гиперфосфатазия с синдромом умственной отсталости 2, 3 и 4; гипертрихотическая остеохондродисплазия; гипобеталипопротеинемия, семейная, связанная с apob32; гипокальциемия аутосомно-доминантная 1; гипокальциурическая гиперкальциемия, семейная, 1 и 3 типы; гипохондрогенез; гипохромная микроцитарная анемия с перегрузкой железом; гипогликемия с дефицитом гликогенсинтетазы в печени; гипогонадотропный гипогонадизм 11 с аносмией или без нее; гипогидротическая эктодермальная дисплазия с иммунодефицитом; гипогидротическая X-сцепленная эктодермальная дисплазия; гипокалиемический периодический паралич 1 и 2; гипомагниемия 1, кишечная; гипомагниемия, судороги и умственная отсталость; гипомиелинизирующая лейкодистрофия 7; синдром гипоплазии левых отделов сердца; дефект атриовентрикулярной перегородки и общий атриовентрикулярный переход; гипоспадия 1 и 2, X-сцепленная; гипотироидизм, врожденный, незобный, 1; гипотрихоз 8 и 12; синдром гипотрихоз-лимфедема-телеангиэктазия; I группа крови; буллезный ихтиоз Сименса; ихтиоз отшелушивающий; ихтиозный синдром при недоношенности;

идиопатическая кальцификация базальных ганглиев 5; идиопатический фиброзирующий альвеолит, хроническая форма; врожденный дискератоз, аутосомно-доминантный, 2 и 5; идиопатическая гиперкальциемия у младенцев; иммунная дисфункция с инактивацией Т-клеток из-за дефекта поступления кальция 2; иммунодефицит 15, 16, 19, 30, 31C, 38, 40, 8 из-за дефекта в cd3-zeta, с гипер-IgM типа 1 и 2, и X-сцепленный с дефектом магния, инфекцией вируса Эпштейна-Барра и неоплазией; иммунодефицит-центромерная нестабильность-синдром лицевых аномалий 2; миопатия с тельцами включения 2 и 3; миопатия Нонака; детские судороги и пароксизмальный хореоатетоз, семейный; детский кортикальный гиперостоз; детский ганглиозидоз GM1; детская гипофосфатазия; детский нефронофтиз; детский нистагм, X-сцепленный; детский паркинсонизм-дистония; бесплодие, связанное с многохвостыми сперматозоидами и избытком ДНК; резистентность к инсулину; инсулинорезистентный сахарный диабет и черный акантоз; синдром секреторной диареи инсулинозависимого сахарного диабета; интерстициальный нефрит, кариомегалия; задержка внутриутробного развития, метафизарная дисплазия, врожденная гипоплазия надпочечников и генитальные аномалии; дефект йодтирозильной связи; дефицит IRAK4; доминантный тип иридогониодисгенеза и тип 1; накопление железа в головном мозге; ишиопателлярная дисплазия; гиперплазия островковых клеток; изолированный дефицит 17,20-лиазы; изолированный дефицит лутропина; дефицит изовалерил-КоА дегидрогеназы; синдром Янковича-Риверы; синдром Джервелла и Ланге-Нильсена 2 ; синдром Жубера 1, 6, 7, 9/15 (дигенный), 14, 16 и 17 и орофациальный синдром xiv; узловой буллезный эпидермолиз Херлитца; юношеский GM>1<ганглиозидоз; синдром юношеского полипоза; юношеский полипоз/синдром наследственной геморрагической телеангиэктазии; юношеский ретиношизис; синдром Кабуки; синдром Каллмана 1, 2 и 6; задержка полового созревания; болезнь Канзаки; синдром Карак; синдром Картагенера; синдром Кенни-Каффи 2 типа; синдром Кеппена-Любинского; кератоконус 1; фолликулярный кератоз; кератоз ладонно-подошвенный полосатый 1; синдром Киндлера; L-2-гидроксиглутаровая ацидурия; синдром Ларсена, доминантный тип; решеточная дистрофия роговицы III типа; амавроз Лебера; синдром Зеллвегера; нарушения биогенеза пероксисом; спектр синдрома Зеллвегера; врожденный амавроз Лебера 11, 12, 13, 16, 4, 7 и 9; атрофия зрительного нерва Лебера; глухота, вызванная аминогликозидами; глухота, несиндромная нейросенсорная,

митохондриальная; некомпакция левого желудочка 5; пороки развития левой-правой оси; болезнь Ли; дефицит митохондриальной короткоцепочечной эноил-КоА гидратазы 1; синдром Ли из-за дефицита митохондриального комплекса I; болезнь Лейнера; дисхондростеоз Лери-Вейля; синдром летальной врожденной контрактуры 6; нарушение адгезии лейкоцитов I и III типа; лейкодистрофия, гипомиелинизирующая, 11 и 6; лейкоэнцефалопатия с атаксией, поражением ствола и спинного мозга и повышением уровня лактата, с исчезновением белого вещества, и прогрессирующая, с недостаточностью яичников; тотальная лейконихия; деменция с тельцами Леви; синдром Лихтенштейна-Норра; синдром Ли-Фраумени 1; синдром Lig4; конечностно-поясная мышечная дистрофия, тип 1B, 2A, 2B, 2D, C1, C5, C9, C14; врожденная мышечная дистрофия-дистрогликанопатия с аномалиями мозга и глаз, тип A14 и B14; комбинированная недостаточность липазы; липидный протеиноз; липодистрофия, семейная парциальная, тип 2 и 3; лиссэнцефалия 1, 2 (X-сцепленная), 3, 6 (с микроцефалией), X-сцепленная; подкорковая ламинарная гетеротопия, X-сцепленная; печеночная недостаточность острая младенческая; синдром Лойса-Дитца 1, 2, 3; синдром удлиненного интервала QT 1, 2, 2/9, 2/5, (дигенный), 3, 5 и 5, приобретенный, восприимчивость; рак легкого; лимфедема, наследственная, id; лимфедема первичная с миелодисплазией; лимфопролиферативный синдром 1, 1 (X-сцепленный) и 2; дефицит липазы лизосомальной кислоты; макроцефалия, макросомия, синдром лицевого дисморфизма; макулярная дистрофия, желточная форма, у взрослых; восприимчивость к злокачественной гипертермии 1 типа; злокачественная неходжкинская лимфома; злокачественная меланома; злокачественная опухоль простаты; мандибулоакральный дизостоз; мандибулоакральная дисплазия с липодистрофией типа A или B, атипичная; мандибуло-лицевой дизостоз по типу Тричера-Коллинза, аутосомно-рецессивный; дефицит маннозо-связывающего белка; заболевание мочи типа "кленовый сироп" 1A и 3 типа; синдром Мардена-Уокера; синдром Марфана; синдром Маринеско-Съёргена; синдром Марцольфа; диабет зрелого возраста у молодых людей типа 1, типа 2, типа 11, типа 3 и типа 9; аномалия Мэй-Хегглина; расстройства, связанные с MYH9; синдром Себастьяна; синдром МакКьюна-Олбрайта; соматотропная аденома; опухоль полового канатика и стромы гонад; синдром Кушинга; синдром МакКьюсика-Кауфмана; синдром нейроакантоцитоза Маклеода; синдром Меккеля-Грубера; дефицит ацил-кофермента A-

дегидрогеназы со средней длиной цепи; медуллобластома; мегаленцефальная лейкоэнцефалопатия с подкорковыми кистами 1 и 2a; мегалэнцефалия мраморной кожи телеангиэктатическая врожденная; связанный с PIK3CA спектр чрезмерного роста; синдром мегалэнцефалии-полимикрогирии-полидактилии-гидроцефалии 2; мегалобластная анемия, тиамин-чувствительная, с сахарным диабетом и нейросенсорной глухотой; синдромы Мейера-Горлина 1 и 4; синдром Мельника-Нидлза; менингиома; умственная отсталость, Х-сцепленная, 3, 21, 30 и 72; умственная отсталость и микроцефалия с гипоплазией моста и мозжечка; умственная отсталость Х-сцепленная синдромная 5; умственная отсталость, протрузия верхней челюсти и косоглазие ; умственная отсталость, аутосомно-доминантный тип 12, 13, 15, 24, 3, 30, 4, 5, 6 и 9; умственная отсталость, аутосомно-рецессивный 15, 44, 46 и 5; умственная отсталость, стереотипные движения, эпилепсия и/или пороки развития головного мозга; умственная отсталость, синдромальная, типа Класа-Йенсена, Х-сцепленная; умственная отсталость, Х-сцепленная, неспецифическая, синдромная, по типу Хедера и синдромальная, тип wu; врожденная мышечная дистрофия с недостаточностью мерозина; метахроматическая лейкодистрофия ювенильного, позднего инфантильного и взрослого типов; метахроматическая лейкодистрофия; метатрофическая дисплазия; метгемоглобинемия I и 2 типа; дефицит метионинаденозилтрансферазы, аутосомно-доминантный; метилмалоновая ацидемия с гомоцистинурией,; метилмалоновая ацидурия типа cblB ; метилмалоновая ацидурия из-за дефицита метилмалонил-КоА мутазы; метилмалоновая ацидурия, тип mut (0); микроцефальная остеодиспластическая примордиальная карликовость 2 типа; микроцефалия с хориоретинопатией, лимфедемой или умственной отсталостью или без них; микроцефалия, грыжа пищеводного отверстия диафрагмы и нефротический синдром; микроцефалия; гипоплазия мозолистого тела; спастическая параплегия 50, аутосомно-рецессивная; глобальная задержка развития; гипомиелинизация ЦНС; атрофия головного мозга; микроцефалия, нормальный интеллект и иммунодефицит; синдром микроцефалии-капиллярной мальформации; микроцитарная анемия; синдромный микрофтальм 5, 7 и 9; микрофтальм, изолированный 3, 5, 6, 8 и с колобомой 6; микросферофакия; мигрень, семейная базилярная; синдром Миллера; миопатия миопатия с наружной офтальмоплегией; миопатия врожденная с стержнями; синдром Митчелла-Райли; митохондриальная недостаточность 3-гидрокси-3-метилглутарил-КоА-синтазы;

дефицит митохондриального комплекса I, II, III, III (ядерный тип 2, 4 или 8); синдром истощения митохондриальной ДНК 11, 12 (кардиомиопатический тип), 2, 4B (тип MNGIE), 8B (тип MNGIE); синдром истощения митохондриальной ДНК 3 и 7, гепатоцеребральные типы и 13 (энцефаломиопатический тип); дефицит митохондриального переносчика фосфата и переносчика пирувата; дефицит митохондриального трифункционального белка; дефицит длинноцепочечной 3-гидроксиацил-КоА дегидрогеназы; мышечная дистрофия Миёси 1; миопатия, дистальная, с началом в передней большеберцовой кости; синдром Мора-Транебьерга; дефицит кофактора молибдена, группа комплементации A; синдром Мовата-Вильсона; муколипидоз III гамма; мукополисахаридоз типа VI, типа VI (тяжелый) и типа VII; мукополисахаридоз, MPS-I-H / S, MPS-II, MPS-III-A, MPS-III-B, MPS-III-C, MPS-IV-A, MPS-IV-B; пигментный ретинит 73; ганглиозидоз GM1 типа 1 (с поражением сердца) 3; многоцентровая остеолизная нефропатия; мультицентрический остеолиз, нодулез и артропатия; множественные врожденные аномалии; дефект межпредсердной перегородки 2; синдром множественных врожденных аномалий-гипотонии-судорог 3; множественные кожные и венозные пороки слизистых оболочек; множественная эндокринная неоплазия 1 и 4 типа; множественная эпифизарная дисплазия 5 или доминантная; множественные атрезии желудочно-кишечного тракта; синдром множественного птеригиума типа Эскобара; множественный дефицит сульфатазы; синдром множественных синостозов 3; дефицит гуанинoксидазы AMP в мышцах; заболевание мышц глаз и головного мозга; мышечная дистрофия, врожденная, мегакониального типа; миастения, семейная инфантильная, 1; миастенический синдром, врожденный, 11, связанный с дефицитом рецепторов ацетилхолина; миастенический синдром, врожденный, 17, 2A (медленный канал), 4B (быстрый канал) и без тубулярных агрегатов; дефицит миелопероксидазы; MYH-ассоциированный полипоз; карцинома эндометрия; инфаркт миокарда 1; миоклоническая дистония; миоклонико-атоническая эпилепсия; миоклонус с эпилепсией с рваными красными волокнами; миофибриллярная миопатия 1 и связанная с ZASP; миоглобинурия, острая рецидивирующая, аутосомно-рецессивная; синдром мионевральной желудочно-кишечной энцефалопатии; мозжечковая атаксия младенческого возраста с прогрессирующей внешней офтальмоплегией; синдром истощения митохондриальной ДНК 4B, тип MNGIE; миопатия, центральноядерная, 1,

врожденная, с избытком мышечных веретен, дистальная, 1, лактоацидоз и сидеробластная анемия 1, митохондриально-прогрессирующая анемия с врожденной катарактой, потерей слуха и задержкой развития и тубулярным агрегатом, 2; близорукость 6; миосклероз аутосомно-рецессивный; миотония врожденная; врожденная миотония, аутосомно-доминантная и рецессивная формы; синдром ногтей-надколенника; синдром Нэнси-Хорана; нанофтальм 2; нейрогепатопатия Навахо; немалиновая миопатия 3 и 9; неонатальная гипотония; интеллектуальная инвалидность; судороги; задержка речи и языкового развития; умственная отсталость, аутосомно-доминантная 31; неонатальный внутрипеченочный холестаз, вызванный дефицитом цитрина; нефрогенный несахарный диабет, Нефрогенный несахарный диабет, Х-сцепленный; нефролитиаз/остеопороз, гипофосфатемический, 2 ; нефронофтис 13, 15 и 4; бесплодие; церебелло-окуло-почечный синдром (нефронофтиз, глазодвигательная апраксия и аномалии мозжечка); нефротический синдром, тип 3, тип 5, с глазными аномалиями или без них, тип 7 и тип 9; синдром прогерии Нестора-Гильермо; синдром Ней-Лаксова 1; нейродегенерация с накоплением железа в головном мозге 4 и 6; нейроферритинопатия; нейрофиброматоз 1 и 2 типа; нейрофибросаркома; нейрогипофизарный несахарный диабет; невропатия, наследственная сенсорная, тип IC; дефект транспорта нейтральной 1 аминокислоты; болезнь накопления нейтральных липидов с миопатией; синдром нейтрофильного иммунодефицита; синдром Николаидеса-Барайцера; болезнь Ниманна-Пика типа C1, C2, типа A и типа C1, взрослая форма; некетотическая гиперглицинемия; синдром Нунана 1 и 4, синдром LEOPARD 1; заболевание, подобное синдрому Нунана, с ювенильным миеломоноцитарным лейкозом или без него; нормокалиемический периодический паралич, чувствительный к калию; болезнь Норума; эпилепсия, потеря слуха и синдром задержки умственного развития; умственная отсталость, Х-сцепленная 102 и синдромальная 13; ожирение; глазной альбинизм, тип I; глазокожный альбинизм типа 1B, типа 3 и типа 4; окулодентодигитальная дисплазия; одонтогипофосфатазия; одонтотрихомелический синдром; болезнь Огучи; синдром олигодонтии-колоректального рака; синдром Опица G/BBB; атрофия зрительного нерва 9; орально-лицевой-дигитальный синдром; дефицит орнитинаминотрансферазы; орофациальная расщелина 11 и 7, синдром расщелины губы / неба и эктодермальной дисплазии; синдром Орставика-Линдеманна-Сольберга; остеоартроз с легкой хондродисплазией; рассекающий остеохондрит;

несовершенный остеогенез тип 12, тип 5, тип 7, тип 8, тип I, тип III, с нормальными склерами, доминантная форма, рецессивный перинатальный летальный; полосчатая остеопатия с краниальный склерозом; остеопетроз аутосомно-доминантный тип 1 и 2, рецессивный 4, рецессивный 1, рецессивный 6; остеопороз с псевдоглиомой; ото-небно-дигитальный синдром, типы I и II; дисгенезия яичников 1; овариолейкодистрофия; врожденная пахионихия 4 и 2 типа; болезнь Педжета костей, семейная; синдром Паллистера-Холла; ладонно-подошвенная кератодермия, неэпидермолитическая, очаговая или диффузная; агенезия поджелудочной железы и врожденный порок сердца; синдром Папийона-Лефевре; параганглиомы 3; врожденная парамиотония фон Эйленбурга; карцинома паращитовидной железы; болезнь Паркинсона 14, 15, 19 (ювенильное начало), 2, 20 (раннее начало), 6 (аутосомно-рецессивное раннее начало) и 9; частичный альбинизм; частичный дефицит гипоксантин-гуанинфосфорибозилтрансферазы; узорчатая дистрофия пигментного эпителия сетчатки; PC-K6a; болезнь Пелицея-Мерцбахера; синдром Пендреда; периферическая демиелинизирующая нейропатия, центральная дисмиелинизация; болезнь Гиршпрунга; постоянный неонатальный сахарный диабет; сахарный диабет, перманентный неонатальный, с неврологическими особенностями; неонатальный инсулинозависимый сахарный диабет; диабет зрелости у молодых, тип 2; нарушение биогенеза пероксисом 14B, 2A, 4A, 5B, 6A, 7A и 7B; синдром Перро 4; синдром Перри; стойкая гиперинсулинемическая гипогликемия младенческого возраста; семейный гиперинсулинизм; фенотипы; фенилкетонурия; феохромоцитома; наследственные синдромы параганглиомы-феохромоцитомы; параганглиомы 1; карциноидная опухоль кишечника; синдром Каудена 3; дефицит фосфоглицератдегидрогеназы; дефицит фосфоглицераткиназы 1; фоточувствительная трихотиодистрофия; болезнь накопления фитановой кислоты; болезнь Пика; синдром Пирсона; пигментная дистрофия сетчатки; пигментная узловая болезнь коры надпочечников, первичная, 1; пиломатриксома; синдром Питта-Хопкинса; гипофизарный гиперкортизолизм; дефицит гормона гипофиза, комбинированный 1, 2, 3 и 4; дефицит ингибитора активатора плазминогена типа 1; дефицит плазминогена, тип I; расстройство свертываемости тромбоцитов 15 и 8; пойкилодермия, наследственное фиброзирование с контрактурами сухожилий, миопатией и фиброзом легких; поликистоз почек 2, взрослый тип и детский тип; поликистозная липомембранозная остеодисплазия со склерозирующей

лейкоэнцефалопатией; миопатия с полиглюкозановыми тельцами 1 с иммунодефицитом или без него ; полимикрогирия, асимметричная, двусторонняя лобно-теменная; полинейропатия, потеря слуха, атаксия, пигментный ретинит и катаракта; понтоцеребеллярная гипоплазия 4 типа; синдром подколенного птеригиума; порэнцефалия 2; порокератоз 8, диссеминированный поверхностный актинический тип; дефицит порфобилиногенсинтазы; поздняя кожная порфирия; атаксия задней колонны с пигментным ретинитом; задняя полярная катаракта 2 типа; синдром Прадера-Вилли; преждевременная недостаточность яичников 4, 5, 7 и 9; первичная аутосомно-рецессивная микроцефалия 10, 2, 3 и 5; первичная цилиарная дискинезия 24; первичная дилатационная кардиомиопатия; некомпакция левого желудочка 6; 4, некомпакция левого желудочка 10; пароксизмальная фибрилляция предсердий; первичная гипероксалурия, тип I, тип и тип III; первичная гипертрофическая остеоартропатия, аутосомно-рецессивная 2; первичная гипомагниемия; первичная открытоугольная глаукома с юношеским дебютом 1; первичная легочная гипертензия; синдром первоцвета; прогрессирующая семейная блокада сердца типа 1B; прогрессирующий семейный внутрипеченочный холестаз 2 и 3; прогрессирующий внутрипеченочный холестаз; прогрессирующая миоклоническая эпилепсия с атаксией; прогрессирующая ревматоидная дисплазия; прогрессирующая склерозирующая полиодистрофия; дефицит пролидазы; дефицит пролиндегидрогеназы; шизофрения 4; дефицит пропердина, X-сцепленный; пропионовая академия; дефицит пропротеинконвертазы 1/3; рак простаты, наследственный, 2; дефект протана; протеинурия; финский врожденный нефротический синдром; синдром Протея; аденокарцинома груди; синдром псевдоахондропластической спондилоэпифизарной дисплазии; псевдогипоальдостеронизм, аутосомно-доминантный и рецессивный тип 1 и тип 2; псевдогипопаратиреоз типа 1A, псевдопсевдогипопаратиреоз; псевдонеонатальная адренолейкодистрофия; псевдопервичный гиперальдостеронизм; эластическая псевдоксантома; обобщенный кальциноз артерий младенческого возраста 2; псевдоксантомное эластикоподобное расстройство с недостаточностью множественных факторов свертывания крови; восприимчивость к псориазу 2; синдром опухоли гамартомы PTEN; легочная артериальная гипертензия, связанная с наследственной геморрагической телеангиэктазией; фиброз легких и/или недостаточность костного мозга, связанные с теломерами, 1 и 3; легочная гипертензия, первичная, 1, с наследственной геморрагической

телеангиэктазией; дефицит пуриннуклеозидфосфорилазы; дефицит пируваткарбоксилазы; дефицит пируватдегидрогеназы E1-альфа; дефицит пируваткиназы эритроцитов; синдром Рейна; разопатия; рецессивный дистрофический буллезный эпидермолиз; поражение ногтей несиндромное врожденное, 8; синдром Рейфенштейна; почечная адисплазия; почечный дефект транспорта карнитина; синдром почечной колобомы; почечная дисплазия; дисплазия почек, пигментная дистрофия сетчатки, мозжечковая атаксия и дисплазия скелета; почечный канальцевый ацидоз, дистальный, аутосомно-рецессивный, с поздней нейросенсорной тугоухостью или с гемолитической анемией; почечный канальцевый ацидоз, проксимальный, с глазными аномалиями и умственной отсталостью; дистрофия конуса сетчатки 3B; пигментный ретинит; пигментный ретинит 10, 11, 12, 14, 15, 17 и 19; пигментный ретинит 2, 20, 25, 35, 36, 38, 39, 4, 40, 43, 45, 48, 66, 7, 70, 72; ретинобластома; расстройство Ретта; синдром предрасположенности к рабдоидным опухолям 2; регматогенная отслойка сетчатки по аутосомно-доминантному типу; точечная ризомелическая хондродисплазия 2 и 3 типа; синдром фокомелии Робертса-SC; синдром Робиноу-Сорауфа; синдром Робиноу, аутосомно-рецессивный, аутосомно-рецессивный, с брахи-син-полидактилией; синдром Ротмунда-Томсона; синдром Рападилино; митохондриальное заболевание, связанное с RRM2B; синдром Рубинштейна-Тайби; болезнь Салла; болезнь Сандхоффа, взрослые и младенческие типы; саркоидоз, раннее начало; синдром Блау; болезнь Шиндлера, тип 1; шизэнцефалия; шизофрения 15; дисплазия Шнекенбекена; шванноматоз 2; синдром Шварца-Джампеля тип 1; склерорнеа аутосомно-рецессивная; склеростеоз; вторичный гипотиреоз ; синдром Сегавы, аутосомно-рецессивный; синдром Сеньора-Локена 4 и 5,; сенсорная атаксическая нейропатия, дизартрия и офтальмопарез; дефицит сепиаптеринредуктазы; синдром SeSAME; тяжелый комбинированный иммунодефицит из-за дефицита ADA, с микроцефалией, задержкой роста и чувствительностью к ионизирующему излучению, атипичный, аутосомно-рецессивный, T-клеточно-отрицательный, B-клеточный, NK-отрицательный или NK-положительный; тяжелая врожденная нейтропения; тяжелая врожденная нейтропения 3, аутосомно-рецессивная или доминантная; тяжелая врожденная нейтропения и 6, аутосомно-рецессивная; тяжелая миоклоническая эпилепсия в младенчестве; генерализованная эпилепсия с фебрильными судорогами плюс, типы 1 и 2; тяжелая X-сцепленная миотубулярная миопатия; синдром короткого интервала QT 3;

низкий рост с неспецифическими аномалиями скелета; низкий рост, атрезия слухового прохода, гипоплазия нижней челюсти, аномалии скелета; низкий рост, ониходисплазия, дисморфизм лица и гипотрихоз; врожденная карликовость; дисплазия грудного отдела коротких ребер 11 или 3 с полидактилией или без нее; сиалидоз I и II типа; синдром серебряной спастической параплегии; замедленная скорость нервной проводимости, аутосомно-доминантная; синдром Смита-Лемли-Опица; синдром Снайдера-Робинсона; соматотропная аденома; пролактинома; семейная предрасположенность к аденоме гипофиза; синдром Сотоса 1 или 2; спастическая атаксия 5, аутосомно-рецессивная, тип Шарлевуа-Сагеней, 1,10 или 11, аутосомно-рецессивная; боковой амиотрофический склероз 5 типа; спастическая параплегия 15, 2, 3, 35, 39, 4, аутосомно-доминантная, 55, аутосомно-рецессивная и 5A; нарушение синтеза желчных кислот, врожденное, 3; сперматогенная недостаточность 11, 3 и 8; сфероцитоз 4 и 5 типов; миопатия сфероидного тела; спинальная мышечная атрофия, с преобладанием нижних конечностей 2, аутосомно-доминантная; спинальная мышечная атрофия II типа; спиноцеребеллярная атаксия 14, 21, 35, 40 и 6; аутосомно-рецессивная спиноцеребеллярная атаксия 1 и 16; гипоплазия селезенки; синдром спондилокарпотарзального синостоза; спондилохейродисплазия, подобная синдрому Элерса-Данлоса, с нарушением иммунной регуляции, тип Аггрекана, с врожденными вывихами суставов, тип короткой конечности-руки, тип Седагата, с дистрофией палочек и колбочек, тип Козловски; парастремматическая карликовость; болезнь Штаргардта 1; дистрофия колбочек и палочек 3; синдром Стиклера тип 1; дисплазия Кнеста; синдром Стиклера, типы 1 (несиндромный глазной) и 4; васкулопатия, связанная с укусом, младенческое начало; синдром Сторморкена; синдром Стерджа-Вебера, пороки развития капилляров, врожденные, 1; дефицит сукцинил-КоА-ацетоацетаттрансферазы; дефицит сахаразы-изомальтазы; синдром внезапной детской смерти; дефицит сульфитоксидазы, изолированный; суправальварный стеноз аорты; дисфункция метаболизма ПАВ, легочная, 2 и 3; симфалангизм проксимальный, 1b; синдактилия типа Ченани-Ленца; синдактилия 3-го типа; синдромная X-сцепленная умственная отсталость 16; эквиноварусная косолапость; болезнь Танжера; TARP-синдром; болезнь Тея-Сакса, вариант B1, Gm2-ганглиозидоз (у взрослых), Gm2-ганглиозидоз (у взрослых); синдром Темтами; синдром Тенорио; терминальная костная дисплазия; недостаток тестостерон-17-бета-дегидрогеназы; тетраамелия, аутосомно-рецессивная;

тетралогия Фалло; синдром гипоплазии левых отделов сердца 2; артериальный ствол; пороки развития сердца и магистральных сосудов; дефект межжелудочковой перегородки 1; дистрофия роговицы Тиля-Бенке; аневризмы грудной аорты и расслоение аорты; марфановский синдром; синдром 3M 2; тромбоцитопения, дисфункция тромбоцитов, гемолиз и несбалансированный синтез глобина; Х-сцепленная тромбоцитопения; тромбофилия, наследственная, из-за дефицита протеина С, аутосомно-доминантная и рецессивная; агенезия щитовидной железы; рак щитовидной железы, фолликулярный; метаболизм гормонов щитовидной железы, анормальный; резистентность к гормонам щитовидной железы, генерализованная, аутосомно-доминантная; тиреотоксический периодический паралич и Тиреотоксический периодический паралич 2; общая резистентность к высвобождающему тиротропин гормону; синдром Тимоти; синдром периодической лихорадки, связанный с рецепторами TNF (TRAPS); агенезия зубов, выборочная, 3 и 4; пируэтная тахикардия; синдром Таунса-Брокса брангиоретенального типа; транзиторный буллезный дермолиз новорожденного; синдром Тричера-Коллинза 1; трихомегалия с умственной отсталостью, карликовостью и пигментной дегенерацией сетчатки; трихоринофаланговая дисплазия I типа; трихоринофалангеальный синдром 3 типа; триметиламинурия; синдром туберозного склероза; лимфангиомиоматоз ; туберозный склероз 1 и 2; тирозиназа-отрицательный окулокожный альбинизм; тирозиназа-положительный окулокожный альбинизм; тирозинемия I типа; дефицит UDP-глюкозо-4-эпимеразы; врожденная мышечная дистрофия Ульриха; отсутствие локтевой и малоберцовой костей при тяжелой недостаточности конечностей; синдром Апшоу-Шульман; дефицит уроканатгидратазы; синдром Ашера, типы 1, 1B, 1D, 1G, 2A, 2C и 2D; пигментный ретинит 39; уФ-чувствительный синдром; синдром Ван дер Вуде; синдром Ван Мальдергема 2; синдром лимфангиэктазии-лимфедемы Хеннекама 2; вариегатная порфирия; вентрикуломегалия при кистозной болезни почек; синдром Верхей; дефицит ацил-КоА-дегидрогеназы с очень длинной цепью; пузырно-мочеточниковый рефлюкс 8; висцеральная гетеротаксия 5, аутосомная; висцеральная миопатия; витамин D-зависимый рахит 1 и 2 типа; желточная дистрофия; болезнь фон Виллебранда типа 2M и типа 3; синдром Ваарденбурга типа 1, 4C и 2E (с неврологическим поражением); синдром Клейна-Ваарденберга; врожденная мышечная дистрофия Уокера-Варбурга; микросиндром Варбурга 2 и 4; бородавки, гипогаммаглобулинемия, инфекции и миелокатексис; синдром

Уивера; синдром Вейля-Марчезани 1 и 3; синдром Вейля-Марчезани; синдром Вайссенбахера-Цвеймюллера; болезнь Верднига-Гофмана; болезнь Шарко-Мари-Тута; синдром Вернера; заболевания, связанные с WFS1; синдром Видеманна-Штайнера; болезнь Вильсона; вольфрамоподобный синдром, аутосомно-доминантный; болезнь Ворта; болезнь Ван Бухема 2 типа; Xeroderma pigmentosum, группа комплементации b, группа D, группа E и группа G; X-сцепленная агаммаглобулинемия; X-сцепленная наследственная моторная и сенсорная нейропатия; X-сцепленный ихтиоз с дефицитом стерилсульфатазы; X-сцепленная перивентрикулярная гетеротопия; ото-небно-дигитальный синдром, тип I; X-сцепленный тяжелый комбинированный иммунодефицит; синдром Циммермана-Лабанда и синдром Циммермана-Лабанда 2; и зонулярная пылевидная катаракта 3.

[0675] В определенном аспекте настоящее изобретение предусматривает способы на основе TPRT для лечения субъекта, у которого диагностировано заболевание, связанное экспансией повторов (также известное как связанное с экспансией повторов заболевание или связанное с тринуклеотидными повторами заболевание). Связанные с экспансией повторов заболевания возникают когда микросателлитные повторы подвергаются экспансии сверх пороговой длины. В настоящее время считается, что, по меньшей мере, 30 генетических заболеваний вызываются экспансиями повторов. Научное понимание данной разнообразной группы заболеваний возникло в ранних 1990-х годах с открытием того, что тринуклеотидные повторы связаны с несколькими серьезными наследственными состояниями, включая синдром ломкой X-хромосомы, спинальную и бульбарную мышечную атрофию, миотоническую дистрофию и болезнь Хантингтона (Nelson et al, "The unstable repeats - three evolving faces of neurological disease," Neuron, March 6, 2013, Vol.77; 825-843 [HYPERLINK: https://www.ncbi.nlm.nih.gov/pubmed/23473314], который полностью приведен здесь в качестве ссылки), а также синдром Хау-Ривера, синдром Якобсена, дентато-рубро-паллидо-льюисову атрофию (DRPLA), болезнь Мачадо-Джозефа, синполидактилию (SPD II), генитальный синдром кисти-стопы (HFGS), клидокраниальную дисплазию (CCD), голопрозэнцефалическое расстройство (HPE), врожденный синдром центральной гиповентиляции (CCHS), ARX-несиндромную X-сцепленную умственную отсталость (XLMR) и окулофарингеальную мышечную дистрофию (OPMD) (см.). Было обнаружено, что нестабильность микросателлитных

повторов является признаком данных состояний, как и предполагалось, то есть явлением, при котором экспансия повторов может возникать с каждым последующим поколением, что приводит к более тяжелому фенотипу и более раннему возрасту возникновения заболевания у потомства. Считается, что экспансии повторов вызывают заболевания за счет нескольких различных механизмов. Более конкретно, экспансии могут нарушать функцию клеток на уровне гена, иРНК транскрипта, и/или кодируемого белка. При некоторых состояниях мутации действуют за счет механизма потери функции путем сайленсинга содержащих повторы генов. При других из них заболевание возникает вследствие механизмов приобретения функции, при которых иРНК транскрипт или белок приобретает новые аберрантные функции.

[0676] В рамках одного варианта осуществления на Фиг.23 изображен способ лечения заболевания, связанного с тринуклеотидными повторами. В целом, данный подход включает применение геномного редактирования на основе TPRT в комбинации с удлиненной нРНК, которая включает участок, который кодирует требуемую и здоровую замещающую последовательность тринуклеотидного повтора, которая должна заместить эндогенную связанную с заболеванием последовательность тринуклеотидного повтора за счет механизма процесса праймерного редактирования. Схема примера конструкции нРНК для сокращения последовательностей тринуклеотидных повторов и сокращение тринуклеотидных повторов с помощью редактирования генома на основе TPRT показаны на Фиг.23.

[0677] Экспансия тринуклеотидных повторов связана с рядом заболеваний человека, включая болезнь Хантингтона, синдром ломкой Х-хромосомы и атаксию Фридрейха. Наиболее часто встречающийся тринуклеотидный повтор содержит триплеты CAG, хотя также встречаются триплеты GAA (атаксия Фридрейха), и триплеты CGG (синдром ломкой Х-хромосомы). Унаследование предрасположенности к экспансии или приобретение уже экспансированного родительского аллеля повышает вероятность возникновения заболевания. Патогенные экспансии тринуклеотидных повторов гипотетически могут быть скорректированы с помощью праймерного редактирования.

[0678] Участок выше участка повтора может быть никирован с помощью РНК-направляемой нуклеазы и затем использован для праймирования синтеза новой цепи ДНК, которая содержит здоровое количество повторов (которое зависит от конкретного гена и

заболевания) в соответствии с общим механизмом, показанным на Фиг.1E или Фиг.22. После последовательности повтора добавляется короткий фрагмент гомологии, который соответствует последовательности, примыкающей к другому концу повтора (красная цепь). Внесение новой синтезированной цепи с помощью TPRT системы и последующее замещение эндогенной ДНК новым синтезированным флэпом ведет к сокращению повторяющегося аллеля. Термин "сокращенный" обозначает укорачивание длины участка нуклеотидного повтора, что тем самым приводит к репарации участка тринуклеотидного повтора.

[0679] В различных вариантах осуществления требуемое нуклеотидное изменение может представлять собой замещение отдельного нуклеотида (например изменение в виде транзиции или трансверсии), делецию, или вставку. Например, требуемое нуклеотидное изменение может представлять собой (1) замещение G на T, (2) замещение G на A, (3) замещение G на C, (4) замещение T на G, (5) замещение T на A, (6) замещение T на C, (7) замещение C на G, (8) замещение C на T, (9) замещение C на A, (10) замещение A на T, (11) замещение A на G, или (12) замещение A на C.

[0680] В других вариантах осуществления требуемое нуклеотидное изменение может обеспечивать конвертирование (1) пары оснований G:C на пару оснований T:A, (2) пары оснований G:C на пару оснований A:T, (3) пары оснований G:C на пару оснований C:G, (4) пары оснований T:A на пару оснований G:C, (5) пары оснований T:A на пару оснований A:T, (6) пары оснований T:A на пару оснований C:G, (7) пары оснований C:G на пару оснований G:C, (8) пары оснований C:G на пару оснований T:A, (9) пары оснований C:G на пару оснований A:T, (10) пары оснований A:T на пару оснований T:A, (11) пары оснований A:T на пару оснований G:C, или (12) пары оснований A:T на пару оснований C:G.

[0681] В других вариантах осуществления способ обеспечивает внесение требуемого нуклеотидного изменения, которое представляет собой вставку. В определенных случаях вставка имеет длину, по меньшей мере, 1, по меньшей мере, 2, по меньшей мере, 3, по меньшей мере, 4, по меньшей мере, 5, по меньшей мере, 6, по меньшей мере, 7, по меньшей мере, 8, по меньшей мере, 9, по меньшей мере, 10, по меньшей мере, 11, по меньшей мере, 12, по меньшей мере, 13, по меньшей мере, 14, по меньшей мере, 15, по меньшей мере, 16, по меньшей мере, 17, по меньшей мере, 18, по меньшей мере, 19, по

меньшей мере, 20, по меньшей мере, 30, по меньшей мере, 40, по меньшей мере, 50, по меньшей мере, 60, по меньшей мере, 70, по меньшей мере, 80, по меньшей мере, 90, по меньшей мере, 100, по меньшей мере, 200, по меньшей мере, 300, по меньшей мере, 400, или, по меньшей мере, 500 нуклеотидов.

[0682] В других вариантах осуществления способ обеспечивает внесение требуемого нуклеотидного изменения, которое представляет собой делецию. В определенных случаях делеция имеет длину, по меньшей мере, 1, по меньшей мере, 2, по меньшей мере, 3, по меньшей мере, 4, по меньшей мере, 5, по меньшей мере, 6, по меньшей мере, 7, по меньшей мере, 8, по меньшей мере, 9, по меньшей мере, 10, по меньшей мере, 11, по меньшей мере, 12, по меньшей мере, 13, по меньшей мере, 14, по меньшей мере, 15, по меньшей мере, 16, по меньшей мере, 17, по меньшей мере, 18, по меньшей мере, 19, по меньшей мере, 20, по меньшей мере, 30, по меньшей мере, 40, по меньшей мере, 50, по меньшей мере, 60, по меньшей мере, 70, по меньшей мере, 80, по меньшей мере, 90, по меньшей мере, 100, по меньшей мере, 200, по меньшей мере, 300, по меньшей мере, 400, или, по меньшей мере, 500 нуклеотидов.

[0683] В различных вариантах осуществления требуемое нуклеотидное изменение обеспечивает корректирование связанного с заболеванием гена. Связанный с заболеванием ген может быть связан с моногенетическим заболеванием, выбранным из группы, состоящей из: дефицита аденозиндезаминазы (ADA); дефицита альфа-1 антитрипсина; муковисцидоза; мышечной дистрофии Дюшенна; галактоземии; гемохроматоза; болезни Хантингтона; болезни кленового сиропа; синдрома Марфана; нейрофиброматоза 1 типа; врожденной пахионихии; фенилкеотнурии; тяжелого комбинированного иммунодефицита; серповидноклеточной анемии; синдрома Смита-Лемли-Опица; и болезни Тея-Сакса. В других вариантах осуществления связанный с заболеванием ген может быть связан с полигенным заболеванием, выбранным из группы, состоящей из: болезни сердца, высокого кровяного давления, болезни Альцгеймера, артрита, диабета, рака, и ожирения.

[0684] Нуклеотидная последовательность-мишень может включать последовательность-мишень (например точечную мутацию), ассоциированную с заболеванием, расстройством, или состоянием. Последовательность-мишень может включать точечную мутацию T на C (или A на G), ассоциированную с заболеванием, расстройством, или состоянием, и при

этом дезаминирование мутантного основания C приводит к опосредованной репарацией несовпадения коррекции в последовательности, которая не ассоциирована с заболеванием, расстройством, или состоянием. Последовательность-мишень может включать точечную мутацию G на A (или C на T), ассоциированную с заболеванием, расстройством, или состоянием, и при этом дезаминирование мутантного основания A приводит к опосредованной репарацией несовпадения коррекции в последовательности, которая не ассоциирована с заболеванием, расстройством, или состоянием. Последовательность-мишень может кодировать белок, и при этом точечная мутация располагается в кодоне и приводит к изменению аминокислоты, кодируемой мутантным кодоном, по сравнению с кодоном дикого типа. Последовательность-мишень также может быть расположена в участке сплайсинга, при этом точечная мутация приводит к изменению в сплайсинге иРНК транскрипта по сравнению с транскриптом дикого типа. Кроме того, мишень может быть расположена в некодирующей последовательности гена, например в промоторе, при этом точечная мутация приводит к повышенной или пониженной экспрессии гена.

[0685] Таким образом, в некоторых аспектах дезаминирование мутантного C приводит к изменению аминокислоты, кодируемой мутантным кодоном, что в некоторых случаях может приводить к экспрессии аминокислоты дикого типа. В других аспектах дезаминирование мутантного A приводит к изменению аминокислоты, кодируемой мутантным кодоном, что в некоторых случаях может приводить к экспрессии аминокислоты дикого типа.

[0686] Описанные здесь способы, включающие контактирование клетки с композицией или rAAV частицей, могут осуществляться *in vitro, ex vivo,* или in vivo. В определенных вариантах осуществления этап контактирования осуществляется в субъекте. В определенных вариантах осуществления у субъекта было диагностировано заболевание, расстройство, или состояние.

[0687] В некоторых вариантах осуществления описанные здесь способы включают контактирование клетки млекопитающего с композицией или rAAV частицей. В определенных вариантах осуществления способы включают контактирования клетки сетчатки, кортикальной клетки или церебеллярной клетки.

[0688] Разделенный белок Cas9 или разделенный праймерный редактор, доставляемый с помощью описанных здесь способов, предпочтительно имеет сравнимую активность по

сравнению с исходным белком Cas9 или праймерным редактором (то есть неразделенным белком, доставляемым в клетку или экспрессируемым в клетке в целом). Например, разделенный белок Cas9 или разделенный праймерный редактор сохраняет, по меньшей мере, 50% (например, по меньшей мере, 50%, по меньшей мере, 60%, по меньшей мере, 70%, по меньшей мере, 80%, по меньшей мере, 90%, по меньшей мере, 95%, по меньшей мере, 98%, по меньшей мере, 99%, или 100%) активности исходного белка Cas9 или праймерного редактора. В некоторых вариантах осуществления разделенный белок Cas9 или разделенный праймерный редактор является более активным (например, 2-кратно, 5-кратно, 10- кратно, 100- кратно, 1000- кратно, или более), чем исходный белок Cas9 или праймерный редактор.

[0689] Описанные здесь композиции могут быть введены нуждающемуся в этом субъекту в терапевтически эффективном количестве для лечения и/или предотвращения заболевания или расстройства, которым страдает субъект. Любое заболевание или расстройство, которое может быть подвергнуто лечению и/или предотвращено с помощью технологии геномного редактирования на основе CRISPR/Cas9, может быть подвергнуто лечению с помощью описанного здесь разделенного белка Cas9 или разделенного праймерного редактора. Следует понимать, что если нуклеотидные последовательности, кодирующие разделенный белок Cas9 или праймерный редактор, далее не кодируют нРНК, то может быть введен отдельный вектор нуклеиновой кислоты, кодирующей нРНК, наряду с описанными здесь композициями.

[0690] Примеры подходящих заболеваний, расстройств или состояний включают, без ограничения, заболевание или расстройство, выбранное из группы, состоящей из: муковисцидоза, фенилкетонурии, эпидермолитического гиперкератоза (ЕНК), хронической обструктивной болезни легких (ХОБЛ), болезни Шарко-Мари-Тута типа 4J, нейробластомы (NB), болезни фон Виллебранда (vWD), врожденной миотонии, наследственного амилоидоза почек, дилатационной кардиомиопатии, наследственной лимфедемы, семейной болезни Альцгеймера, прионной болезни, хронического детского неврологического кожно-суставного синдрома (CINCA), врожденной глухоты, болезни Ниманна-Пика типа С (NPC), и миопатии, связанной с десмином (DRM). В определенных вариантах осуществления заболевание или состояние представляет собой болезнь Ниманна-Пика типа С (NPC).

[0691] В некоторых вариантах осуществления заболевание, расстройство, или состояние ассоциировано с точечной мутацией в гене NPC, гене *DNMT1*, гене *PCSK9*, или гене *TMC1*. В определенных вариантах осуществления точечная мутация представляет собой мутацию T3182C в *NPC*, которая приводит к аминокислотному замещению I1061T.
[0692] В определенных вариантах осуществления точечная мутация представляет собой мутацию A545G в *TMC1*, которая приводит к аминокислотному замещению Y182C. *TMC1* кодирует белок, который образует механочувствительные ионные каналы в сенсорных волосковых клетках внутреннего уха и необходим для нормального функционирования слуха. Аминокислотное замещение Y182C ассоциировано с врожденной глухотой.
[0693] В некоторых вариантах осуществления заболевание, расстройство или состояние ассоциировано с точечной мутацией, которая приводит к образованию стоп-кодона, например преждевременного стоп-кодона в кодирующем участке гена.
[0694] Дополнительные примеры заболеваний, расстройств и состояний включают муковисцидоз (см., например, Schwank *et al.*, Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. Cell stem cell. 2013; 13: 653-658; и Wu et. al. *Cell stem cell.* **2013**; 13: 653-658; и Wu et. al., Correction of a genetic disease in mouse via use of CRISPR-Cas9. *Cell stem cell.* **2013;** 13: 659-662, при этом ни в одном из них не используется дезаминазный слитый белок для корректирования генетического дефекта); фенилкетонурию – например мутацию фенилаланин на серин в положении 835 (мышь) или 240 (человек) или гомологичном остатке в гене фенилаланингидроксилазы (мутация T>C) – см., например, McDonald *et al.*, *Genomics.* **1997;** 39:402-405; синдром Бернара-Сулье (BSS) – например мутацию фенилаланин на серин в положении 55 или гомологичном остатке, или цистеин на аргинин в положении 24 или гомологичном остатке в гликопротеине мембраны тромбоцитов IX (мутация T>C) – см., например, Noris *et al.*, *British Journal of Haematology.* **1997;** 97: 312-320, и Ali *et al.*, *Hematol.* **2014;** 93: 381-384; эпидермолитический гиперкератоз (ЕНК) – например мутацию лейцин на пролин в положении 160 или 161 (если считать начальный метионин) или гомологичном остатке в кератине 1 (мутация T>C) – см., например, Chipev *et al.*, *Cell.* **1992;** 70: 821-828, также см. кат. # P04264 в базе данных UNIPROT на www[dot]uniprot[dot]org; хроническую обструктивную болезнь легких (ХОБЛ) – например мутацию лейцин на пролин в положении 54 или 55 (если считать начальный метионин) или гомологичном остатке в

процессированной форме α_1-антитрипсина или остатке 78 в непроцессированной форме или гомологичном остатке (мутация T>C) – см., например, Poller *et al.*, *Genomics.* **1993;** 17: 740-743, см. кат. # P01011 в базе данных UNIPROT; болезнь Шарко-Мари-Тута типа 4J – например мутацию изолейцин на треонин в положении 41 или гомологичном остатке в FIG4 (мутация T>C) – см., например, Lenk *et al.*, PLoS Genetics. 2011; 7: e1002104; нейробластому (NB) – например мутацию в положении 197 или гомологичном остатке в каспазе-9 (мутация T>C) – см., например, Kundu *et al.*, *3 Biotech.* **2013, 3:225-234**; болезнь фон Виллебранда (vWD) – например мутацию цистеин на аргинин в положении 509 или гомологичном остатке в процессированной форме фактора Виллебранда, или в положении 1272 или гомологичном остатке в непроцессированной форме фактора Виллебранда (мутация T>C) – см., например, Lavergne *et al.*, *Br. J. Haematol.* **1992**, также см. кат. # P04275 в базе данных UNIPROT**;** 82: 66-72; врожденную миотонию – например мутацию цистеин на аргинин в положении 277 или гомологичном остатке в гене мышечного хлоридного канала CLCN1 (мутация T>C mutation) – см., например, Weinberger *et al.*, *The J. of Physiology.* **2012;** 590: 3449-3464; наследственный амилоидоз почек – например мутацию стоп-кодон на аргинин в положении 78 или гомологичном остатке в процессированной форме аполипопротеина AII или в положении 101 или гомологичном остатке в непроцессированной форме (мутация T>C) – см., например, Yazaki *et al.*, *Kidney Int.* **2003;** 64: 11-16; дилатационную кардиомиопатию (DCM) – например мутацию триптофан на аргинин в положении 148 или гомологичном остатке в гене FOXD4 (мутация T>C), см., например, Minoretti et. al., *Int. J. of Mol. Med.* **2007;** 19: 369-372; наследственную лимфедему – например мутацию гистидин на аргинин в положении 1035 или гомологичном остатке в VEGFR3 тирозинкиназе (мутация A>G), см., например, Irrthum *et al.*, *Am. J. Hum. Genet.* **2000;** 67: 295-301; фамильную болезнь Альцгеймера – например мутацию изолейцин на валин в положении 143 или гомологичном остатке в презенилине 1 (мутация A>G), см., например, Gallo et. al., *J. Alzheimer's disease.* **2011;** 25: 425-431; прионную болезнь – например мутацию метионин на валин в положении 129 или гомологичном остатке в прионном белке (мутация A>G) – см., например, Lewis et. al., *J. of General Virology.* **2006;** 87: 2443-2449; хронический детский неврологический кожно-суставной синдром (CINCA) – например мутацию тирозин на цистеин в положении 570 или гомологичном остатке в криопирине (мутация A>G) – см., например, Fujisawa et. al.

Blood. **2007;** 109: 2903-2911; и связанную с десмином миопатию (DRM) – например мутацию аргинин на глицин в положении 120 или гомологичном остатке в кристаллине αβ (мутация A>G mutation) – см., например, Kumar *et al.*, *J. Biol. Chem.* **1999;** 274: 24137-24141. Полное содержание всех ссылок и записей в базах данных приведено здесь в качестве ссылки.

V. Фармацевтические композиции

[0695] Другие аспекты настоящего изобретения относятся к фармацевтическим композициям, включающим любые из различных компонентов описанной здесь системы праймерного редактора (PE) (например включая, но без ограничения, napDNAbp, обратные транскриптазы, слитые белки (например включающие napDNAbp и обратные транскриптазы), направляющие РНК праймерного редактора, направляющие РНК для никирования второй цепи и комплексы, включающие слитые белки и направляющие РНК праймерного редактора, а также вспомогательные элементы, такие как компоненты для никирования второй цепи и нуклеазы для удаления 5' эндогенного ДНК-флэпа для обеспечения процесса праймерного редактирования с образованием подвергнутого редактированию продукта). Такие композиции могут включать PE1, PE2, PE3, или PE3b.

[0696] Термин "фармацевтическая композиция" в соответствии с используемым здесь значением обозначает композицию, составленную для фармацевтического применения. В некоторых вариантах осуществления фармацевтическая композиция дополнительно включает фармацевтически приемлемый носитель. В некоторых вариантах осуществления фармацевтическая композиция включает дополнительные агенты (например для специфичной доставки, увеличения периода полужизни, или другие терапевтические соединения).

[0697] В соответствии с используемым здесь значением термин "фармацевтически приемлемый носитель" означает фармацевтически приемлемый материал, композицию или носитель, такой как жидкий или твердый наполнитель, разбавитель, инертный наполнитель, вспомогательное средство для производства (например лубрикант, магниевый тальк, стеарат кальция или цинка, или стеариновую кислоту), или материал для инкапсуляции разбавителя, принимающий участие в переносе или транспортировке соединения из одного участка организма (например участка доставки) в другой участок (например орган, ткань, или часть тела). Фармацевтически приемлемый носитель

является "приемлемым" в плане того, что он является совместимым с другими ингредиентами состава и не наносит вреда ткани субъекта (например является физиологически совместимым, стерильным, имеет физиологический pH и так далее). Некоторые примеры материалов, которые могут служить в качестве фармацевтически приемлемых носителей, включают: (1) сахара, такие как лактоза, глюкоза и сахароза; (2) крахмалы, такие как кукурузный крахмал и картофельный крахмал; (3) целлюлоза и ее производные, такие как карбоксиметилцеллюлоза натрия, метилцеллюлоза, этилцеллюлоза, микрокристаллическая целлюлоза и ацетат целлюлозы; (4) порошкообразный трагакант; (5) солод; (6) желатин; (7) смазывающие агенты, такие как стеарат магния, лаурилсульфат натрия и тальк; (8) наполнители, такие как масло какао и воски для суппозиториев; (9) масла, такие как арахисовое масло, хлопковое масло, сафлоровое масло, кунжутное масло, оливковое масло, кукурузное масло и соевое масло; (10) гликоли, такие как пропиленгликоль; (11) полиолы, такие как глицерин, сорбит, маннит и полиэтиленгликоль (ПЭГ); (12) сложные эфиры, такие как этилолеат и этиллаурат; (13) агар; (14) буферные агенты, такие как гидроксид магния и гидроксид алюминия; (15) альгиновую кислоту; (16) апирогенную воду; (17) изотонический физиологический раствор; (18) раствор Рингера; (19) этиловый спирт; (20) pH-буферные растворы; (21) сложные полиэфиры, поликарбонаты и/или полиангидриды; (22) наполнители, такие как полипептиды и аминокислоты; (23) компонент сыворотки, такой как сывороточный альбумин, HDL и LDL; (22) C2-C12 спирты, такие как этанол; и (23) другие нетоксичные совместимые вещества, используемые в фармацевтических составах. В составе также могут присутствовать смачивающие агенты, красители, агенты для высвобождения, агенты для покрытия, подсластители, ароматизаторы, отдушки, консерванты и антиоксиданты. Такие термины как "инертный наполнитель", "носитель", "фармацевтически приемлемый носитель" или им подобные используются здесь взаимозаменяемым образом.

[0698] В некоторых вариантах осуществления фармацевтическая композиция составлена для доставки субъекту, например для редактирования генов. Подходящие способы введения описанной здесь фармацевтической композиции включают, без ограничения: местное, подкожное, трансдермальное, внутрикожное, внутриочаговое, внутрисуставное, внутрибрюшинное, внутрипузырное, чрезслизистое, десневое, внутризубное,

интракохлеарное, транстимпаническое, внутриорганное, эпидуральное, интратекальное, внутримышечное, внутривенное, внутрисосудистое, внутрикостное, периокулярное, внутриопухолевое, внутримозговое и интрацеребровентрикулярное введение.

[0699] В некоторых вариантах осуществления описанную здесь фармацевтическую композицию вводят местно в участок заболевания (например участок с опухолью). В некоторых вариантах осуществления описанную здесь фармацевтическую композицию вводят субъекту с помощью инъекции, с помощью катетера, с помощью суппозитория, или с помощью имплантата, при этом имплантат является пористым, непористым, или изготовленным из мягкого материала, включая мембрану, например силиконовую мембрану или волокно.

[0700] В других вариантах осуществления описанная здесь фармацевтическая композиция доставляется в рамках системы с контролируемым высвобождением. В одном варианте осуществления может использоваться помпа (см., например, Langer, 1990, Science 249:1527-1533; Sefton, 1989, CRC Crit. Ref. Biomed. Eng. 14:201; Buchwald et al., 1980, Surgery 88:507; Saudek et al., 1989, N. Engl. J. Med. 321:574). В другом варианте осуществления могут использоваться полимерные материалы. (См., например, Medical Applications of Controlled Release (Langer and Wise eds., CRC Press, Boca Raton, Fla., 1974); Controlled Drug Bioavailability, Drug Product Design and Performance (Smolen and Ball eds., Wiley, New York, 1984); Ranger and Peppas, 1983, Macromol. Sci. Rev. Macromol. Chem. 23:61. Также см. Levy *et al.*, 1985, Science 228:190; During *et al.*, 1989, Ann. Neurol. 25:351; Howard *et al.*, 1989, J. Neurosurg. 71:105.) Также описаны другие системы с контролируемым высвобождением, например см. Langer выше.

[0701] В некоторых вариантах осуществления фармацевтическая композиция составлена в соответствии со стандартными процедурами в виде композиции, адаптированной для внутривенного или подкожного введения субъекту, например человеку. В некоторых вариантах осуществления фармацевтическая композиция для введения с помощью инъекции представляет собой растворы в стерильном изотоническом водном буфере. При необходимости фармацевтическая композиция также может включать солюбилизирующий агент и местный анестетик, например лигнокаин, для облегчения боли в месте инъекции. В целом ингредиенты присутствуют либо по отдельности, либо смешиваются вместе в виде единичной лекарственной формы, например в виде сухого

лиофилизированного порошка или безводного концентрата в герметично запечатанном контейнере, например ампуле или саше, с указанием количества активного агента. При введении фармацевтической композиции с помощью инфузии она может быть разведена с использованием бутылки для инфузий, содержащей стерильную подходящую для фармацевтического применения воду или солевой раствор. При введении фармацевтической композиции с помощью инъекции может быть добавлена ампула со стерильной водой или солевым раствором для инъекций, так что ингредиенты могут быть смешаны перед введением.

[0702] Фармацевтическая композиция для системного введения может представлять собой жидкость, например стерильный солевой раствор, лактатный раствор Рингера или Хэнка. Кроме того, фармацевтическая композиция может быть представлена в твердых формах и повторно растворена или суспендирована непосредственно перед применением. Также охватываются лиофилизированные формы.

[0703] Фармацевтическая композиция может содержаться в липидной частице или пузырьке, например липосоме или микрокристалле, который также является подходящим для парентерального введения. Частицы могут иметь любую подходящую структуру, например однослойную или многослойную, при условии, что в них содержатся композиции. Соединения могут быть заключены в "стабилизированных плазмид-липидных частицах" (SPLP), содержащих фузогенный липид диолеоилфосфатидилэтаноламин (DOPE), низкие уровни (5-10 мол.%) катионного липида, и стабилизированых за счет полиэтиленгликолевого (ПЭГ) покрытия (Zhang Y. P. *et al.*, *Gene Ther.* 1999, 6:1438-47). Положительно заряженные липиды, такие как N-[1-(2,3-диолеоилокси)пропил]-N,N,N-триметил-аммониймутилсульфат, или "DOTAP", являются особенно предпочтительными для таких частиц и пузырьков. Получение подобных липидных частиц является хорошо известным. См., например, патенты США 4880635; 4906477; 4911928; 4917951; 4920016; и 4921757; каждый из которых приведен здесь в качестве ссылки.

[0704] Описанная здесь фармацевтическая композиция может быть, например, введена или упакована в виде единичной дозы. Термин "единичная доза" при использовании в отношении фармацевтической композиции по настоящему изобретению обозначает физически отдельные единицы, являющиеся подходящими в качестве единичной дозы для

субъекта, при этом каждая единица содержит предварительно определенное количество активного материала, рассчитанное для продуцирования требуемого терапевтического эффекта, совместно с требуемым разбавителем; то есть носителем или индифферентным веществом.

[0705] Кроме того, фармацевтическая композиция может быть представлена в виде фармацевтического набора, включающего (a) контейнер, содержащий соединение по изобретению в лиофилизированной форме, и (b) второй контейнер, содержащий фармацевтически приемлемый разбавитель (например стерильную воду) для инъекции. Фармацевтически приемлемый разбавитель может применяться для восстановления или разбавления лиофилизированного соединения по изобретению. С таким контейнером (контейнерами) может быть необязательно связана этикетка в форме, соответствующей требованиям государственного учреждения, регулирующего производство, применение или продажу лекарственный средств или биологических продуктов, на которой отражено одобрение данного учреждения, регулирующего производство, применение, или продажу, для введения человеку.

[0706] В другом аспекте включено изделие, содержащее материалы, которые могут применяться для лечения описанных выше заболеваний. В некоторых вариантах осуществления изделие включает контейнер и этикетку. Подходящие контейнеры включают, например, бутылки, флаконы, шприцы и пробирки для тестов. Контейнеры могут быть изготовлены из различных материалов, таких как стекло или пластик. В некоторых вариантах осуществления контейнер содержит композицию, которая является эффективной для лечения описанного здесь заболевания, и может иметь стерильное входное отверстие. Например, контейнер может представлять собой пакет с внутривенным раствором или флакон, имеющий пробку, протыкаемую иглой для подкожных инъекций. Активный агент в композиции представляет собой соединение по изобретению. В некоторых вариантах осуществления этикетка, которая расположена на контейнере или связана с контейнером, указывает, что композиция применяется для лечения выбранного заболевания. Изделие может дополнительно включать второй контейнер, включающий фармацевтически приемлемый буфер, например фосфатно-солевой буфер, раствор Рингера, или раствор декстрозы. Он может дополнительно включать другие материалы, являющиеся желательными с коммерческой и

пользовательской точки зрения, включая другие буферы, разбавители, фильтры, иглы, шприцы, и вкладыши в упаковку с инструкциями по применению.

VI. Вирусные способы доставки

[0707] В некоторых аспектах изобретение предусматривает способы, включающие доставку одного или нескольких полинуклеотидов, например одного или нескольких векторов, как описано здесь, кодирующих один или несколько компонентов системы праймерного редактора (PE), описанной здесь, одного или нескольких ее транскриптов, и/или одного или нескольких транскрибированных из нее белков в клетку-хозяин. В некоторых аспектах изобретение дополнительно предусматривает клетки, полученные с помощью таких способов, и организмы (такие как животные, растения, или грибы), состоящие или полученные из таких клеток. В некоторых вариантах осуществления редактор оснований, как описано здесь, в комбинации (и, необязательно, в виде комплекса) с направляющей последовательностью доставляется в клетку. Для внесения нуклеиновых кислот в клетки млекопитающих или ткани-мишени могут применяться стандартные способы переноса генов, которые основаны или не основаны на вирусах. Такие способы могут применяться для введения нуклеиновых кислот, кодирующих компоненты редактора оснований, в клетки в культуре или в организм-хозяин. Невирусные векторные системы доставки включают ДНК-плазмиды, РНК (например описанный здесь транскрипт вектора), "голую" нуклеиновую кислоту, и нуклеиновую кислоту в виде комплекса со средством для доставки, например липосомой. Вирусные векторные системы доставки включают ДНК и РНК вирусы, которые имеют либо эписомальный, или встроенный геном после доставки в клетку. См. обзор процедур генной терапии в Anderson, Science 256:808-813 (1992); Nabel & Felgner, TIBTECH 11:211-217 (1993); Mitani & Caskey, TIBTECH 11:162-166 (1993); Dillon, TIBTECH 11:167-175 (1993); Miller, Nature 357:455-460 (1992); Van Brunt, Biotechnology 6(10):1149-1154 (1988); Vigne, Restorative Neurology and Neuroscience 8:35-36 (1995); Kremer & Perricaudet, British Medical Bulletin 51(1):31-44 (1995); Haddada et al., в Current Topics in Microbiology and Immunology Doerfler and Bihm (eds) (1995); и Yu et al., Gene Therapy 1:13-26 (1994).

[0708] Способы невирусной доставки нуклеиновых кислот включают липофекцию, нуклеофекцию, микроинъекцию, биолистику, вирусомы, липосомы, иммунолипосомы, конъюгаты поликатион или липид:нуклеиновая кислота, голую ДНК, искусственные

вирионы, и агент-энхансированный захват ДНК. Липофекция описана, например, в патентах США 5049386, 4946787, и 4897355), а реагенты для липофекции являются коммерчески доступными (например Transfectam™ и Lipofectin™). Катионные и нейтральные липиды, которые являются подходящими для эффективной липофекции полинуклеотидов на основе распознавания рецептора, включают указанные у Feigner, WO 91/17424; WO 91/16024. Доставка может осуществляться в клетки (например при введении *in vitro* или *ex vivo*) или ткани-мишени (например при введении *in vivo*).

[0709] Получение комплексов липид:нуклеиновая кислота, включая липосомы-мишени, такие как иммунолипидные комплексы, хорошо известно специалисту в данной области (см., например g., Crystal, Science 270:404-410 (1995); Blaese et al., Cancer Gene Ther. 2:291-297 (1995); Behr et al., Bioconjugate Chem. 5:382-389 (1994); Remy et al., Bioconjugate Chem. 5:647-654 (1994); Gao et al., Gene Therapy 2:710-722 (1995); Ahmad et al., Cancer Res. 52:4817-4820 (1992); патенты США 4186183, 4217344, 4235871, 4261975, 4485054, 4501728, 4774085, 4837028, и 4946787).

[0710] Применение систем на основе вирусной РНК или ДНК для доставки нуклеиновых кислот основано на продвинутых процессах нацеливания вируса на определенные клетки в организме и направлении полезной нагрузки вируса в ядро. Вирусные векторы могут быть введена напрямую пациентам (*in vivo*) или они могут быть использованы для лечения клеток *in vitro*, при этом модифицированные клетки могут необязательно быть введены пациентам (*ex vivo*). Стандартные системы на основе вирусов могут включать ретровирусные, лентивирусные, аденовирусные, адено-ассоциированные и основанные на вирусе простого герпеса векторы для переноса генов. Встраивание в геном хозяина возможно в случае способов на основе переноса генов с помощью ретровируса, лентивируса и адено-ассоциированного вируса, что зачастую приводит к долгосрочной экспрессии вставленного трансгена. Кроме того, высокие показатели эффективности трансдукции наблюдались в случае многих различных типов клеток и тканей-мишеней.

[0711] Тропизм вирусов может быть изменен за счет встраивания белков чужеродной оболочки, что обеспечивает увеличение потенциальной популяции клеток-мишеней. Лентивирусные векторы представляют собой ретровирусные векторы, которые способны трансдуцировать или инфицировать неделящиеся клетки и обычно продуцируют высокие вирусные титры. Таким образом, выбор ретровирусной системы для трансфера генов

будет зависеть от ткани-мишени. Ретровирусные векторы состоят из cis-действующих длинных концевых повторов с показателем упаковывания до 6-10 kb чужеродной последовательности. Минимальные cis-действующие LTR являются достаточными для репликации и упаковывания векторов, которые затем используются для встраивания терапевтического гена в клетку-мишень для обеспечения перманентной экспрессии трансгена. Широко применяемые ретровирусные векторы включают основанные на вирусе лейкоза мышей (MuLV), вирусе лейкоза гиббонов (GaLV), вирусе иммунодефицита обезьян (SIV), вирусе иммунодефицита человека (ВИЧ), и их комбинациях (см., например, Buchscher et al., J. Virol. 66:2731-2739 (1992); Johann et al., J. Virol. 66:1635-1640 (1992); Sommnerfelt et al., Virol. 176:58-59 (1990); Wilson et al., J. Virol. 63:2374-2378 (1989); Miller et al., J. Virol. 65:2220-2224 (1991); PCT/US94/05700). В рамках вариантах использования, где является предпочтительной транзиентная экспрессия, могут использоваться системы на основе аденовирусов. Векторы на основе аденовирусов способны обеспечивать очень высокую эффективность трансдукции во многих типах клеток и не требуют разделения клеток. С помощью подобных векторов был достигнут высокий титр и уровни экспрессии. Данный вектор может быть получен в больших количествах в рамках относительно простой системы. Векторы на основе адено-ассоциированного вируса ("AAV") также могут применяться для трансдукции клеток с использованием нуклеиновых кислот-мишеней, например при *in vitro* получении нуклеиновых кислот и пептидов и для *in vivo* и *ex vivo* процедур генной терапии (см., например West *et al.*, Virology 160:38-47 (1987); патент США 4797368; WO 93/24641; Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J. Clin. Invest. 94:1351 (1994). Конструирование рекомбинантных AAV векторов описано в ряде публикаций, включая патент США 5173414; Tratschin et al., Mol. Cell. Biol. 5:3251-3260 (1985); Tratschin, et al., Mol. Cell. Biol. 4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); и Samulski et al., J. Virol. 63:03822-3828 (1989).

[0712] Упаковывающие клетки обычно используются для образования вирусных частиц, которые способны инфицировать клетку-хозяина. Такие клетки включают клетки 293, которые упаковывают аденовирус, и клетки ψ2 или клетки PA317, которые упаковывают ретровирус. Вирусные векторы, используемые в рамках генной терапии, обычно получают путем продуцирования линии клеток, которые упаковывают вектор нуклеиновой кислоты

в вирусную частицу. Векторы обычно содержат минимальные вирусные последовательности, требуемые для упаковывания и последующего встраивания в хозяина, при этом другие вирусные последовательности замещаются кассетой экспрессии для подвергаемого экспрессии полинуклеотида (полинуклеотидов). Отсутствующие вирусные функции обычно обеспечиваются *in trans* за счет линии упаковывающих клеток. Например, AAV векторы, используемые в рамках генной терапии, обычно обладают только ITR последовательностями из генома AAV, которые необходимы для упаковывания и встраивания в геном хозяина. Вирусная ДНК упаковывается в линии клеток, которая содержит хелперную плазмиду, кодирующую другие гены AAV, а именно rep и cap, но не имеет ITR последовательностей. Линия клеток также может быть инфицирована аденовирусом в качестве хелпера. Хелперный вирус промотирует репликацию AAV вектора и экспрессию генов AAV из хелперной плазмиды. Хелперная плазмида не упаковывается в значительных количествах вследствие отсутствия ITR последовательностей. Контаминация аденовирусом может быть снижена, например, с помощью тепловой обработки, к которой аденовирус является более чувствительным, чем AAV. Дополнительные способы доставки нуклеиновых кислот в клетки известны специалистам в данной области. См., например, публикацию US 20030087817, приведенную здесь в качестве ссылки.

VII. Наборы, векторы, клетки, и доставка

[0713] Некоторые аспекты настоящего изобретения относятся к наборам, включающим конструкт нуклеиновой кислоты, включающий нуклеотидную последовательность, кодирующую различные компоненты описанной здесь системы праймерного редактора (PE) (например включая, но без ограничения, napDNAbp, обратные транскриптазы, слитые белки (например включающие napDNAbp и обратные транскриптазы), направляющие РНК праймерного редактора, и комплексы, включающие слитые белки и направляющие РНК праймерного редактора, а также вспомогательные элементы, такие как компоненты для никирования второй цепи и нуклеазы для удаления 5' эндогенного ДНК-флэпа для обеспечения приведения процесса праймерного редактирования к образованию подвергнутого редактированию продукта). В некоторых вариантах осуществления нуклеотидная последовательность включает гетерологичный промотор, который обеспечивает экспрессию компонентов системы праймерного редактора (PE).

[0714] Некоторые аспекты данного изобретения предусматривают наборы, включающие один или несколько конструктов нуклеиновой кислоты, кодирующей различные компоненты описанной здесь системы праймерного редактора (PE), например включающие нуклеотидную последовательность, кодирующую компоненты системы праймерного редактора (PE), способные модифицировать последовательность-мишень ДНК. В некоторых вариантах осуществления нуклеотидная последовательность включает гетерологичный промотор, который обеспечивает экспрессию компонентов системы праймерного редактора (PE).

[0715] Некоторые аспекты данного изобретения предусматривают наборы, включающие конструкт нуклеиновой кислоты, включающий (a) нуклеотидную последовательность, кодирующую napDNAbp (например домен Cas9), слитый с обратной транскриптазой, и (b) гетерологичный промотор, который обеспечивает экспрессию последовательности (a).

[0716]

Наборы

[0717] Композиции по настоящему изобретению могут быть собраны в наборы. В некоторых вариантах осуществления набор включает векторы нуклеиновой кислоты для экспрессии описанных здесь праймерных редакторов. В других вариантах осуществления набор дополнительно включает подходящие направляющие нуклеотидные последовательности (например PEgRNA и нРНК для второго участка) или векторы нуклеиновой кислоты для экспрессии таких направляющих нуклеотидных последовательностей для нацеливания белка Cas9 или праймерного редактора на требуемую последовательность-мишень.

[0718] Описанный здесь набор может включать один или несколько контейнеров, включающих компоненты для осуществления описанных здесь способов и необязательно инструкции по применению. Любой из описанных здесь наборов может дополнительно включать компоненты, требуемые для осуществления способов проведения анализов. Каждый компонент наборов где это возможно может быть представлен в жидкой форме (например в виде раствора) или в твердой форме (например в виде сухого порошка). В определенных случаях некоторые компоненты могут подвергаться восстановлению или быть обработаны иным образом (например до активной формы), например путем

добавления подходящего растворителя или другого компонента (например воды), который может входить или не входить в состав набора.

[0719] В некоторых вариантах осуществления наборы могут необязательно включать инструкции и/или материалы относительно применения представленных компонентов. В соответствии с используемым здесь значением термин "инструкции" может определять компонент в виде инструкции и/или материала и обычно включает письменные инструкции, расположенные на упаковке или связанные с упаковкой. Инструкции также могут включать любые устные или электронные инструкции, предоставленные любым образом, так что пользователю станет быстро очевидно, что инструкции следует ассоциировать с набором, например аудиовизуальные (например в виде видеокассеты, DVD и так далее), основанные на интернете и/или получаемые через веб, и так далее. Письменные инструкции могут быть в любой форме, соответствующей требованиям государственного учреждения, регулирующего производство, применение или продажу лекарственный средств или биологических продуктов, которые также могут отражать одобрение данного учреждения, регулирующего производство, применение, или продажу, для введения человеку. В соответствии с используемым здесь значением термин "материалы" включает все способы ведения бизнеса, включая способы образовательного, больничного, и другого клинического инструктирования, научные запросы, разработку и открытие лекарств, академические исследования, активность в рамках фармацевтической промышленности, включая продажи лекарственных средств, и любые виды рекламы и другую связанную с продвижением активность, включая письменные, устные и электронные коммуникации в любой форме, связанные с данным изобретением. Кроме того, наборы могут включать другие компоненты в зависимости от определенного вида использования, как описано здесь.

[0720] Наборы могут содержать любой один или несколько описанных здесь компонентов в одном или нескольких контейнерах. Компоненты могут быть получены в стерильном виде, упакованы в шприц и доставлены в охлажденном виде. В качестве альтернативы, они могут быть заключены во флаконе или другом контейнере для хранения. Второй контейнер может включать другие компоненты, полученные в стерильном виде. В качестве альтернативы, наборы могут включать активные агенты, предварительно смешанные и доставленные во флаконе, пробирке, или другом контейнере.

[0721] Наборы могут иметь различные формы, такие как блистерный пакет, пакет, упакованный в термоусадочную пленку, пакет с вакуумным запечатыванием, запечатываемый термоформованный лоток или аналогичный пакет или лоток, с аксессуарами, свободно упакованными внутри пакета, одной или несколькими пробирками, контейнерами, коробкой или пакетом. Наборы могут быть стерилизованы после добавления аксессуаров, что позволяет не запаковывать отдельные аксессуары в контейнере. Наборы могут быть простерилизованы с помощью любых подходящих способов стерилизации, таких как стерилизация радиацией, тепловая стерилизация, или другие способы стерилизации, известные в данной области. Наборы также могут включать другие компоненты в зависимости от определенного вида использования, например контейнеры, питательную среду для клеток, соли, буферы, реагенты, шприцы, иглы ткань, такую как марля, для нанесения или удаления дезинфицирующего средства, одноразовые перчатки, подставку для агентов перед введением и так далее. Некоторые аспекты настоящего изобретения относятся к наборам, включающим конструкт нуклеиновой кислоты, включающий нуклеотидную последовательность, кодирующую различные компоненты описанной здесь системы праймерного редактирования (например включая, но без ограничения, napDNAbp, обратные транскриптазы, полимеразы, слитые белки (например включающие napDNAbp и обратные транскриптазы (или, более широко, полимеразы), удлиненные направляющие РНК, и комплексы, включающие слитые белки и удлиненные направляющие РНК, а также вспомогательные элементы, такие как компоненты для никирования второй цепи (например нРНК для никирования второй цепи) и нуклеазы для удаления 5' эндогенного ДНК-флэпа для реализации процесса праймерного редактирования с образованием подвергнутого редактированию продукта). В некоторых вариантах осуществления нуклеотидная последовательность (последовательности) включает гетерологичный промотор (или более одного промотора), который обеспечивает экспрессию компонентов системы праймерного редактирования.

[0722] Другие аспекты данного изобретения предусматривают наборы, включающие один или несколько конструктов нуклеиновой кислоты, кодирующей различные компоненты описанной здесь системы праймерного редактирования, например включающие нуклеотидную последовательность, кодирующую компоненты системы праймерного редактирования, способные модифицировать последовательность-мишень ДНК. В

некоторых вариантах осуществления нуклеотидная последовательность включает гетерологичный промотор, который обеспечивает экспрессию компонентов системы праймерного редактирования.

[0723] Некоторые аспекты данного изобретения предусматривают наборы, включающие конструкт нуклеиновой кислоты, включающий (a) нуклеотидную последовательность, кодирующую napDNAbp (например домен Cas9), слитый с обратной транскриптазой, и (b) гетерологичный промотор, который обеспечивает экспрессию последовательности (a).

Клетки

[0724] Клетки, которые могут содержать любую из описанных здесь композиций, включают прокариотические клетки и эукариотические клетки. Описанные здесь способы применяются для доставки белка Cas9 или праймерного редактора в эукариотическую клетку (например клетку млекопитающего, такую как клетка человека). В некоторых вариантах осуществления клетка представлена in vitro (например, является культивируемой клеткой). В некоторых вариантах осуществления клетка представлена in vivo (например располагается в субъекте, таком как представляющий собой человека субъект). В некоторых вариантах осуществления клетка представлена ex vivo (например выделена из субъекта и может быть введена обратно тому же или другому субъекту).

[0725] Клетки млекопитающих по настоящему изобретению включают клетки человека, клетки приматов (например клетки Vero), клетки крыс (например клетки GH3, клетки OC23) или клетки мышей (например клетки MC3T3). Существует множество линий клеток человека, включая, без ограничения, эмбриональные клетки почки человека (HEK), клетки HeLa, раковые клетки из 60 линий раковых клеток Национального института рака (National Cancer Institute) (NCI60), клетки DU145 (рак простаты), клетки Lncap (рак простаты), клетки MCF-7 (рак груди), клетки MDA-MB-438 (рак груди), клетки PC3 (рак простаты), клетки T47D (рак груди), клетки THP-1 (острый миелоидный лейкоз), клетки U87 (глиобластома), клетки нейробластомы человека SHSY5Y (клонированные из миеломы) и клетки Saos-2 (рак кости). В некоторых вариантах осуществления rAAV векторы доставляют в эмбриональные клетки почки человека (HEK) (например клетки HEK 293 или HEK 293T). В некоторых вариантах осуществления векторы rAAV доставляются в стволовые клетки (например стволовые клетки человека), такие как, например, плюрипотентные стволовые клетки (например плюрипотентные стволовые

клетки человека, включая индуцированные плюрипотентные стволовые клетки человека (hiPSC)). Термин "стволовая клетка" обозначает клетку со способностью делиться в течение бесконечного времени в культуре и порождать специализированные клетки. Термин "плюрипотентная стволовая клетка" обозначает тип стволовой клетки, которая способна дифференцироваться во все ткани организма, но не способна поддерживать полное развитие организма самостоятельным образом. Термин "индуцированная плюрипотентная стволовая клетка человека" обозначает соматическую (например зрелую или взрослую) клетку, которая была перепрограммирована в соответствующее эмбриональной стволовой клетке состояние за счет индуцирования экспрессии генов и факторов, являющихся важными для поддержания основных свойств эмбриональных стволовых клеток (см., например, Takahashi and Yamanaka, Cell 126 (4): 663 - 76, 2006, который приведен здесь в качестве ссылки). Индуцированные плюрипотентные стволовые клетки человека экспрессируют маркеры стволовых клеток и способны продуцировать клетки, характерные для всех трех зародышевых листков (эктодермы, энтодермы, мезодермы).

[0726] Дополнительные неограничивающие примеры линий клеток, которые могут применяться в соответствии с настоящим изобретением, включают 293-T, 293-T, 3T3, 4T1, 721, 9L, A-549, A172, A20, A253, A2780, A2780ADR, A2780cis, A431, ALC, B16, B35, BCP-1, BEAS-2B, bEnd.3, BHK-21, BR 293, BxPC3, C2C12, C3H-10T1/2, C6, C6/36, Cal-27, CGR8, CHO, CML T1, CMT, COR-L23, COR-L23/5010, COR-L23/CPR, COR-L23/R23, COS-7, COV-434, CT26, D17, DH82, DU145, DuCaP, E14Tg2a, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, Hepa1c1c7, клетки High Five, HL-60, HMEC, HT-29, HUVEC, клетки J558L, клетки Jurkat, JY, клетки K562, KCL22, KG1, Ku812, KYO1, LNCap, Ma-Mel 1, 2, 3....48, MC-38, MCF-10A, MCF-7, MDA-MB-231, MDA-MB-435, MDA-MB-468, MDCK II, MG63, MONO-MAC 6, MOR/0.2R, MRC5, MTD-1A, MyEnd, NALM-1, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NW-145, OPCN/OPCT Peer, PNT-1A/PNT 2, PTK2, Raji, клетки RBL, RenCa, RIN-5F, RMA/RMAS, S2, клетки Saos-2, Sf21, Sf9, SiHa, SKBR3, SKOV-3, T-47D, T2, T84, THP1, U373, U87, U937, VCaP, WM39, WT-49, X63, YAC-1 и клетки YAR.

[0727] Некоторые аспекты данного изобретения предусматривают клетки, включающие любой из описанных здесь конструктов. В некоторых вариантах осуществления клетка-

хозяин транзиентно или нетранзиентно трансфицируется одним или несколькими описанными здесь векторами. В некоторых вариантах осуществления клетка трансфицируется в естественном виде, в котором она находится в субъекте. В некоторых вариантах осуществления клетку, которая подвергается трансфекции, забирают от субъекта. В некоторых вариантах осуществления клетку получают из клеток, взятых от субъекта, например из линии клеток. В данной области известно большое количество линий клеток для культивирования тканей. Примеры линий клеток включают, но без ограничения: C8161, CCRF-CEM, MOLT, mIMCD-3, NHDF, HeLa-S3, Huh1, Huh4, Huh7, HUVEC, HASMC, HEKn, HEKa, MiaPaCell, Panc1, PC-3, TF1, CTLL-2, C1R, Rat6, CV1, RPTE, A10, T24, J82, A375, ARH-77, Calu1, SW480, SW620, SKOV3, SK-UT, CaCo2, P388D1, SEM-K2, WEHI-231, HB56, TIB55, Jurkat, J45.01, LRMB, Bcl-1, BC-3, IC21, DLD2, Raw264.7, NRK, NRK-52E, MRC5, MEF, Hep G2, HeLa B, HeLa T4, COS, COS-1, COS-6, COS-M6A, эпителиальные клетки BS-C-1 почки обезьяны, эмбриональный фибробласт мыши BALB/3T3, 3T3 Swiss, 3T3-L1, фибробласты зародышей человека 132-d5; фибробласты мыши 10.1, 293-T, 3T3, 721, 9L, A2780, A2780ADR, A2780cis, A 172, A20, A253, A431, A-549, ALC, B16, B35, клетки BCP-1, BEAS-2B, bEnd.3, BHK-21, BR 293. BxPC3. C3H-10T1/2, C6/36, Cal-27, CHO, CHO-7, CHO-IR, CHO-K1, CHO-K2, CHO-T, CHO Dhfr −/−, COR-L23, COR-L23/CPR, COR-L23/5010, COR-L23/R23, COS-7, COV-434, CML T1, CMT, CT26, D17, DH82, DU145, DuCaP, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, HEK-293, HeLa, Hepa1c1c7, HL-60, HMEC, HT-29, Jurkat, клетки JY, клетки K562, Ku812, KCL22, KG1, KYO1, LNCap, Ma-Mel 1-48, MC-38, MCF-7, MCF-10A, MDA-MB-231, MDA-MB-468, MDA-MB-435, MDCK II, MDCK 11, MOR/0.2R, MONO-MAC 6, MTD-1A, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NALM-1, NW-145, линии клеток OPCN/OPCT, Peer, PNT-1A/PNT 2, RenCa, RIN-5F, RMA/RMAS, клетки Saos-2, Sf-9, SkBr3, T2, T-47D, T84, линию клеток THP1, U373, U87, U937, VCaP, клетки Vero, WM39, WT-49, X63, YAC-1, YAR, и их трансгенные варианты.

[0728] Линии клеток доступны из ряда источников, известных специалистам в данной области (см., например, Американскую коллекцию типовых культур (American Type Culture Collection, ATCC) (Manassus, Va.)). В некоторых вариантах осуществления клетка, трансфицированная одним или несколькими описанными здесь векторами, применяется

для получения новой линии клеток, включающих одну или несколько последовательностей из вектора. В некоторых вариантах осуществления клетка, транзиентно трансфицированная компонентами системы CRISPR, как описано здесь, (например в виде транзиентной трансфекции одним или несколькими векторами, или трансфекции РНК) и модифицированная за счет активности комплекса CRISPR, применяется для получения новой линии клеток, включающих модификацию, но не имеющих какой-либо другой экзогенной последовательности. В некоторых вариантах осуществления клетки транзиентно или нетранзиентно трансфицируют одним или несколькими описанными здесь векторами или линии клеток, полученные из подобных клеток, применяют при анализе одного или нескольких тестируемых соединений.

Векторы

[0729] Некоторые аспекты настоящего изобретения относятся к применению рекомбинантных вирусных векторов (например адено-ассоциированных вирусных векторов, аденовирусных векторов, или векторов на основе вируса простого герпеса) для доставки описанных здесь праймерных редакторов или их компонентов, например разделенного белка Cas9 или разделенных праймерных редакторов, в клетку. В случае подхода на основе разделенного PE N-концевая часть слитого белка PE и C-концевая часть продукта слияния PE доставляются с помощью отдельных рекомбинантных вирусных векторов (например адено-ассоциированных вирусных векторов, аденовирусных векторов, или векторов на основе вируса простого герпеса) в одну клетку, поскольку полноразмерный белок Cas9 или праймерные редакторы превышают лимит упаковывания различных вирусных векторов, например rAAV (~4,9 kb).

[0730] Таким образом, в одном варианте осуществления изобретение охватывает векторы, способные обеспечивать доставку слитых белков разделенного праймерного редактора или его разделенных компонентов. В некоторых вариантах осуществления предусмотрена композиция для доставки разделенного белка Cas9 или разделенного праймерного редактора в клетку (например клетку млекопитающего, клетку человека). В некоторых вариантах осуществления композиция по настоящему изобретению включает: (i) первую частицу рекомбинантного адено-ассоциированного вируса (rAAV), включающую первую нуклеотидную последовательность, кодирующую N-концевую часть белка Cas9 или праймерный редактор, слитый на его C-конце с интеином-N; и (ii) вторую частицу

рекомбинантного адено-ассоциированного вируса (rAAV), включающую вторую нуклеотидную последовательность, кодирующую интеин-C, слитый с N-концом C-концевой части белка Cas9 или праймерного редактора.. rAAV частицы по настоящему изобретению включают rAAV вектор (то есть рекомбинантный геном rAAV), инкапсулированный в белках капсида вируса.

[0731] В некоторых вариантах осуществления rAAV вектор включает: (1) гетерологичный участок нуклеиновой кислоты, включающий первую или вторую нуклеотидную последовательность, кодирующую N-концевую часть или C-концевую часть разделенного белка Cas9 или разделенного праймерного редактора в любой форме, как описано здесь, (2) одну или несколько нуклеотидных последовательностей, включающих последовательность, которая обеспечивает экспрессию гетерологичного участка нуклеиновой кислоты (например промотор), и (3) один или несколько участков нуклеиновой кислоты, включающих последовательность, которая обеспечивает встраивание гетерологичного участка нуклеиновой кислоты (необязательно с одним или несколькими участками нуклеиновой кислоты, включающими последовательность, которая обеспечивает экспрессию), в геном клетки. В некоторых вариантах осуществления вирусные последовательности, которые обеспечивают встраивание, включают последовательности инвертированных концевых повторов (ITR). В некоторых вариантах осуществления первая или вторая нуклеотидная последовательность, кодирующая N-концевую часть или C-концевую часть разделенного белка Cas9 или разделенного праймерного редактора, фланкируется на каждой стороне последовательностью ITR. В некоторых вариантах осуществления вектор нуклеиновой кислоты дополнительно включает участок, кодирующий белок AAV Rep, как описано здесь, который содержится либо в участке, фланкируемом ITR, либо вне данного участка. Последовательности ITR могут быть получены из любого серотипа AAV (например 1, 2, 3, 4, 5, 6, 7, 8, 9, или 10) или могут быть получены из более чем одного серотипа. В некоторых вариантах осуществления последовательности ITR получены из AAV2 или AAV6.

[0732] Таким образом, в некоторых вариантах осуществления описанные здесь rAAV частицы включают, по меньшей мере, одну rAAV2 частицу, rAAV6 частицу, rAAV8 частицу, rPHP.B частицу, rPHP.eB частицу, или rAAV9 частицу, или их вариант. В

определенных вариантах осуществления описанные rAAV частицы представляют собой rPHP.B частицы, rPHP.eB частицы, rAAV9 частицы.

[0733] Последовательности ITR и плазмиды, содержащие последовательности ITR, хорошо известны в данной области и являются коммерчески доступными (см., например, продукты и услуги, предоставляемые Vector Biolabs, Philadelphia, PA; Cellbiolabs, San Diego, CA; Agilent Technologies, Santa Clara, Ca; и Addgene, Cambridge, MA; и Gene delivery to skeletal muscle results in sustained expression and systemic delivery of a therapeutic protein. Kessler PD, Podsakoff GM, Chen X, McQuiston SA, Colosi PC, Matelis LA, Kurtzman GJ, Byrne BJ. Proc Natl Acad Sci USA. 1996 Nov 26;93(24):14082-7; и Curtis A. Machida. Methods in Molecular Medicine™. Viral Vectors for Gene Therapy Methods and Protocols. 10.1385/1-59259-304-6:201 © Humana Press Inc. 2003. Chapter 10. Targeted Integration by Adeno-Associated Virus. Matthew D. Weitzman, Samuel M. Young Jr., Toni Cathomen and Richard Jude Samulski; патенты США 5139941 и 5962313, все из которых приведены здесь в качестве ссылки).

[0734] В некоторых вариантах осуществления rAAV вектор по настоящему изобретению включает один или несколько регуляторных элементов для контролирования экспрессии гетерологичного участка нуклеиновой кислоты (например промоторы, терминаторы транскрипции, и/или другие регуляторные элементы). В некоторых вариантах осуществления первая и/или вторая нуклеотидная последовательность функционально связана с одной или несколькими (например 1, 2, 3, 4, 5 или более) терминаторами транскрипции. Неограничивающие примеры терминаторов транскрипции, которые могут применяться в соответствии с настоящим изобретением, включают терминаторы транскрипции гена гормона роста коровы (bGH), гена гормона роста человека (hGH), SV40, CW3, ϕ, или их комбинации. Была протестирована эффективность нескольких терминаторов транскрипции для определения их соответствующих эффектов на уровень экспрессии разделенного белка Cas9 или разделенного праймерного редактора. В некоторых вариантах осуществления терминатор транскрипции, используемый в настоящем изобретении, представляет собой терминатор транскрипции bGH. В некоторых вариантах осуществления rAAV вектор дополнительно включает посттранскрипционный регуляторный элемент вируса гепатита сурка (WPRE). В определенных вариантах осуществления WPRE представляет собой усеченную последовательность WPRE,

например "W3". В некоторых вариантах осуществления WPRE вставлен 5' относительно терминатора транскрипции. При транскрипции такие последовательности создают третичную структуру, которая повышает экспрессию, особенно из вирусных векторов.

[0735] В некоторых вариантах осуществления используемые здесь векторы кодируют белки слитого PE или любые из его компонентов (например napDNAbp, линкеры, или полимеразы). Кроме того, применяемые здесь векторы могут кодировать PEgRNA и/или вспомогательную нРНК для никирования второй цепи. Векторы могут быть способны обеспечивать экспрессию одной или нескольких кодирующих последовательностей в клетке. В некоторых вариантах осуществления клетка может представлять собой прокариотическую клетку, например бактериальную клетку. В некоторых вариантах осуществления клетка может представлять собой эукариотическую клетку, например клетку дрожжей, растения, насекомого или млекопитающего. В некоторых вариантах осуществления эукариотическая клетка может представлять собой клетку млекопитающего. В некоторых вариантах осуществления эукариотическая клетка может представлять собой клетку грызуна. В некоторых вариантах осуществления эукариотическая клетка может представлять собой клетку человека. Подходящие промоторы для обеспечения экспрессии в различных типах клеток известны в данной области. В некоторых вариантах осуществления промотор может относиться к дикому типу. В других вариантах осуществления промотор может быть модифицирован для более эффективной экспрессии. В других вариантах осуществления промотор может быть усечен, но все же сохранять свою функцию. Например, промотор может иметь нормальный размер или уменьшенный размер, являющийся подходящим для правильного упаковывания вектора в вирус.

[0736] В некоторых вариантах осуществления промоторы, которые могут применяться в рамках векторов праймерного редактора, могут являться конститутивными, индуцируемыми, или специфичными в отношении ткани. В некоторых вариантах осуществления промоторы могут представлять собой конститутивные промоторы. Неограничивающие примеры конститутивных промоторов включают немедленный ранний промотор цитомегаловируса (CMV), промотор вируса обезьяны (SV40), большой поздний промотор аденовируса (MLP), промотор вируса саркомы Рауса (RSV), промотор вируса опухоли молочной железы мыши (MMTV), промотор фосфоглицераткиназы (

PGK), промотор фактора-альфа элонгации (EFla), промоторы убиквитина, промоторы актина, промоторы тубулина, промоторы иммуноглобулина, их функциональный фрагмент или комбинацию любых из них. В некоторых вариантах осуществления промотор может представлять собой промотор CMV. В некоторых вариантах осуществления промотор может представлять собой усеченный промотор CMV. В других вариантах осуществления промотор может представлять собой промотор EFla. В некоторых вариантах осуществления промотор может представлять собой индуцируемый промотор. Неограничивающие примеры индуцируемых промоторов включают промоторы, которые индуцируются температурным шоком, светом, химическими веществами, пептидами, металлами, стероидами, антибиотиками, или спиртом. В некоторых вариантах осуществления индуцируемый промотор может представлять собой промотор, имеющий низкий базальный (неиндуцированный) уровень экспрессии, например промотор Tet-On® (Clontech). В некоторых вариантах осуществления промотор может представлять собой специфический в отношении ткани промотор. В некоторых вариантах осуществления специфический в отношении ткани промотор исключительно или преимущественно экспрессируется в ткани печени. Неограничивающие примеры специфических в отношении ткани промоторов включают промотор B29, промотор CD14, промотор CD43, промотор CD45, промотор CD68, промотор десмина, промотор эластазы-1, промотор эндоглина, промотор фибронектина, промотор Flt-1, промотор GFAP, промотор GPIIb, промотор ICAM- 2, промотор INF-β, промотор Mb, промотор Nphsl, промотор OG-2, промотор SP-B, промотор SYN1, и промотор WASP.

[0737] В некоторых вариантах осуществления векторы праймерного редактора (например включая любые векторы, кодирующие слитый белок праймерного редактора и/или PEgRNA и/или вспомогательные нРНК для никирования второй цепи) могут включать индуцируемые промоторы для начала экспрессии только после его доставки в клетку-мишень. Неограничивающие примеры индуцируемых промоторов включают промоторы, которые индуцируются температурным шоком, светом, химическими веществами, пептидами, металлами, стероидами, антибиотиками, или спиртом. В некоторых вариантах осуществления индуцируемый промотор может представлять собой промотор, имеющий низкий базальный (неиндуцированный) уровень экспрессии, например промотор Tet-On® (Clontech).

[0738] В дополнительных вариантах осуществления векторы праймерного редактора (например включая любые векторы, кодирующие слитый белок праймерного редактора и/или PEgRNA и/или вспомогательные нРНК для никирования второй цепи) могут включать специфические в отношении ткани промоторы для начала экспрессии только после его доставки в определенную ткань. Неограничивающие примеры специфических в отношении ткани промоторов включают промотор B29, промотор CD14, промотор CD43, промотор CD45, промотор CD68, промотор десмина, промотор эластазы-1, промотор эндоглина, промотор фибронектина, промотор Flt-1, промотор GFAP, промотор GPIIb, промотор ICAM- 2, промотор INF-β, промотор Mb, промотор Nphsl, промотор OG-2, промотор SP-B, промотор SYN1, и промотор WASP.

[0739] В некоторых вариантах осуществления нуклеотидная последовательность, кодирующая PEgRNA (или любые направляющие РНК, используемые в рамках праймерного редактирования) может быть функционально связана, по меньшей мере, с одной последовательностью контроля транскрипции или трансляции. В некоторых вариантах осуществления нуклеотидная последовательность, кодирующая направляющую РНК, может быть функционально связана, по меньшей мере, с одним промотором. В некоторых вариантах осуществления промотор может распознаваться РНК-полимеразой III (Pol III). Неограничивающие примеры Pol III промоторов включают промоторы U6, HI, и тРНК. В некоторых вариантах осуществления нуклеотидная последовательность, кодирующая направляющую РНК, может быть функционально связана с промотором U6 мыши или человека. В других вариантах осуществления нуклеотидная последовательность, кодирующая направляющую РНК, может быть функционально связана с промотором HI мыши или человека. В некоторых вариантах осуществления нуклеотидная последовательность, кодирующая направляющую РНК, может быть функционально связана с тРНК промотором мыши или человека. В вариантах осуществления с более чем одной направляющей РНК промоторы, применяемые для обеспечения экспрессии, могут являться одинаковыми или различающимися. В некоторых вариантах осуществления нуклеотид, кодирующий crРНК направляющей РНК, и нуклеотид, кодирующей tracrРНК направляющей РНК, могут быть внесены на одном векторе. В некоторых вариантах осуществления нуклеотид, кодирующий crРНК, и нуклеотид, кодирующей tracrРНК, могут быть внесены с помощью одного промотора. В

некоторых вариантах осуществления crРНК и tracrРНК могут быть транскрибированы в один транскрипт. Например, crРНК и tracrРНК могут быть процессированы из одного транскрипта для образования направляющей РНК в виде двойной молекулы. В качестве альтернативы, crРНК и tracrРНК могут быть транскрибированы в направляющую РНК в виде одной молекулы.

[0740] В некоторых вариантах осуществления нуклеотидная последовательность, кодирующая направляющую РНК, может быть расположена на одном векторе, включающем нуклеотидную последовательность, кодирующую слитый белок PE. В некоторых вариантах осуществления экспрессия направляющей РНК и слитого белка PE может быть обеспечена за счет их соответствующих промоторов. В некоторых вариантах осуществления экспрессия направляющей РНК может быть обеспечена за счет того же промотора, который обеспечивает экспрессию слитого белка PE. В некоторых вариантах осуществления направляющая РНК и транскрипт слитого белка PE могут содержаться в одном транскрипте. Например, направляющая РНК может располагаться в нетранслируемом участке (UTR) транскрипта белка Cas9. В некоторых вариантах осуществления направляющая РНК может располагаться в 5' UTR транскрипта слитого белка PE. В других вариантах осуществления направляющая РНК может располагаться в 3' UTR транскрипта слитого белка PE. В некоторых вариантах осуществления внутриклеточный период полужизни транскрипта слитого белка PE может быть снижен за счет включения направляющей РНК в его 3' UTR с укорачиванием, тем самым, длины его 3' UTR. В дополнительных вариантах осуществления направляющая РНК может располагаться в интроне транскрипта слитого белка PE. В некоторых вариантах осуществления в интрон, в котором располагается направляющая РНК, могут быть добавлены подходящие участки сплайсинга, так что направляющая РНК подвергается сплайсингу из транскрипта правильным образом. В некоторых вариантах осуществления экспрессия белка Cas9 и направляющей РНК в непосредственной близости на одном векторе может обеспечивать более эффективное образование комплекса CRISPR.

[0741] Система праймерного редактора на основе векторов может включать один вектор, или два вектора, или три вектора, или четыре вектора, или пять векторов, или более. В некоторых вариантах осуществления векторная система может включать один единственный вектор, который кодирует как слитый белок PE, так и PEgRNA. В других

вариантах осуществления векторная система может включать два вектора, при этом один вектор кодирует слитый белок PE, а другой кодирует PEgRNA. В дополнительных вариантах осуществления векторная система может включать три вектора, при этом третий вектор кодирует нРНК для никирования второй цепи, применяемую в рамках описанных здесь способов.

[0742] В некоторых вариантах осуществления композиция, включающая rAAV частицу (в любой предусмотренной здесь форме), дополнительно включает фармацевтически приемлемый носитель. В некоторых вариантах осуществления композиция включает подходящие фармацевтические носители для введения субъектам, представляющим собой человека или животных.

[0743] Некоторые примеры материалов, которые могут служить в качестве фармацевтически приемлемых носителей, включают: (1) сахара, такие как лактоза, глюкоза и сахароза; (2) крахмалы, такие как кукурузный крахмал и картофельный крахмал; (3) целлюлоза и ее производные, такие как карбоксиметилцеллюлоза натрия, метилцеллюлоза, этилцеллюлоза, микрокристаллическая целлюлоза и ацетат целлюлозы; (4) порошкообразный трагакант; (5) солод; (6) желатин; (7) смазывающие агенты, такие как стеарат магния, лаурилсульфат натрия и тальк; (8) наполнители, такие как масло какао и воски для суппозиториев; (9) масла, такие как арахисовое масло, хлопковое масло, сафлоровое масло, кунжутное масло, оливковое масло, кукурузное масло и соевое масло; (10) гликоли, такие как пропиленгликоль; (11) полиолы, такие как глицерин, сорбит, маннит и полиэтиленгликоль (ПЭГ); (12) сложные эфиры, такие как этилолеат и этиллаурат; (13) агар; (14) буферные агенты, такие как гидроксид магния и гидроксид алюминия; (15) альгиновую кислоту; (16) апирогенную воду; (17) изотонический физиологический раствор; (18) раствор Рингера; (19) этиловый спирт; (20) pH-буферные растворы; (21) сложные полиэфиры, поликарбонаты и/или полиангидриды; (22) наполнители, такие как полипептиды и аминокислоты; (23) компонент сыворотки, такой как сывороточный альбумин, HDL и LDL; (22) C2-C12 спирты, такие как этанол; и (23) другие нетоксичные совместимые вещества, используемые в фармацевтических составах. В составе также могут присутствовать смачивающие агенты, красители, агенты для высвобождения, агенты для покрытия, подсластители, ароматизаторы, отдушки, консерванты и антиоксиданты. Такие термины как "инертный наполнитель", "носитель",

"фармацевтически приемлемый носитель" или им подобные используются здесь взаимозаменяемым образом.

Способы доставки

[0744] В некоторых аспектах изобретение предусматривает способы, включающие доставку одного или нескольких полинуклеотидов, например одного или нескольких векторов, как описано здесь, одного или нескольких их транскриптов, и/или одного или нескольких транскрибированных из них белков в клетку-хозяин. В некоторых аспектах изобретение дополнительно предусматривает клетки, полученные с помощью таких способов, и организмы (такие как животные, растения, или грибы), состоящие или полученные из таких клеток. В некоторых вариантах осуществления редактор оснований, как описано здесь, в комбинации (и, необязательно, в виде комплекса) с направляющей последовательностью доставляется в клетку.

[0745] Примеры стратегий доставки описаны в других разделах здесь и включают стратегии на основе векторов, доставку на основе рибонуклеопротеинового комплекса РЕ, и доставку РЕ с помощью способов на основе иРНК.

[0746] В некоторых вариантах осуществления предусмотренный способ доставки включает нуклеофекцию, микроинъекцию, биолистику, вирусомы, липосомы, иммунолипосомы, конъюгаты поликатион или липид:нуклеиновая кислота, голую ДНК, искусственные вирионы, и агент-энхансированный захват ДНК.

[0747] Примеры способов доставки нуклеиновых кислот включают липофекцию, нуклеофекцию, электропорацию, стабильную геномную интеграцию (например piggyback), микроинъекцию, биолистику, вирусомы, липосомы, иммунолипосомы, конъюгаты поликатион или липид:нуклеиновая кислота, голую ДНК, искусственные вирионы, и агент-энхансированный захват ДНК. Липофекция описана, например, в патентах США 5049386, 4946787, и 4897355), а реагенты для липофекции являются коммерчески доступными (например Transfectam™, Lipofectin™ и SF Cell Line 4D-Nucleofector X Kit™ (Lonza)). Катионные и нейтральные липиды, которые являются подходящими для эффективной липофекции полинуклеотидов на основе распознавания рецептора, включают указанные у Feigner, WO 91/17424; WO 91/16024. Доставка может осуществляться в клетки (например при введении in vitro или ex vivo) или ткани-мишени

(например при введении in vivo). Доставка может быть достигнута за счет использования RNP комплексов.

[0748] Получение комплексов липид:нуклеиновая кислота, включая липосомы-мишени, такие как иммунолипидные комплексы, хорошо известно специалисту в данной области (см., например, Crystal, Science 270:404-410 (1995); Blaese et al., Cancer Gene Ther. 2:291-297 (1995); Behr et al., Bioconjugate Chem. 5:382-389 (1994); Remy et al., Bioconjugate Chem. 5:647-654 (1994); Gao et al., Gene Therapy 2:710-722 (1995); Ahmad et al., Cancer Res. 52:4817-4820 (1992); патенты США 4186183, 4217344, 4235871, 4261975, 4485054, 4501728, 4774085, 4837028, и 4946787).

[0749] В других вариантах осуществления способ доставки и предусмотренный здесь вектор представляет собой RNP комплекс. RNP доставка слитых белков заметно повышает специфичность редактирования оснований в плане ДНК. RNP доставка слитых белков ведет к устранению одновременного целевого и нецелевого редактирования ДНК. RNP доставка обеспечивает устранение нецелевого редактирования в участках без повторов, в то же время обеспечивая целевое редактирование, сравнимое с плазмидной доставкой, и значительно снижает нецелевое редактирование ДНК даже в участке 2 VEGFA с высоким содержанием повторов. См. Rees, H.A. et al., Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery, Nat. Commun. 8, 15790 (2017), патент США 9526784, выданный 27 декабря 27 2016 г., и патент США 9737604, выданный 22 августа 2017 г., каждый из которых приведен здесь в качестве ссылки.

[0750] Дополнительные способы доставки нуклеиновых кислот в клетки известны специалистам в данной области. См., например, публикацию US 2003/0087817, приведенную здесь в качестве ссылки.

[0751] Другие способы настоящего изобретения предусматривают способы доставки конструктов праймерного редактора в клетку для образования полного и функционального праймерного редактора в клетке. Например, в некоторых вариантах осуществления клетка контактирует с описанной здесь композицией (например композициями, включающими нуклеотидные последовательности, кодирующие разделенный Cas9 или разделенный праймерный редактор или AAV частицы, содержащие векторы нуклеиновой кислоты, включающие такие нуклеотидные последовательности). В

некоторых вариантах осуществления контактирование приводит к доставке таких нуклеотидных последовательностей в клетку, при этом N-концевая часть белка Cas9 или праймерного редактора и С-концевая часть белка Cas9 или праймерного редактора экспрессируются в клетке и соединяются для образования полного белка Cas9 или полного праймерного редактора.

[0752] Является очевидным, что любая предусмотренная здесь rAAV частица, молекула нуклеиновой кислоты или композиция может быть внесена в клетку любым подходящим способом, либо стабильно, или транзиентно. В некоторых вариантах осуществления описанные белки могут быть трансфецированы в клетку. В некоторых вариантах осуществления клетка может быть трансдуцирована или трансфицирована молекулой нуклеиновой кислоты. Например, клетка может быть трансдуцирована (например вирусом, кодирующим разделенный белок) или трансфицирована (например плазмидой, кодирующей разделенный белок) молекулой нуклеиновой кислоты, которая кодирует разделенный белок, или rAAV частицей, содержащей вирусный геном, кодирующий одну или несколько молекул нуклеиновой кислоты. Такая трансдукция может представлять собой стабильную или транзиентную трансдукцию. В некоторых вариантах осуществления клетки, экспрессирующие разделенный белок или содержащие разделенный белок, могут быть трансдуцированы или трансфицированы одной или несколькими последовательностями направляющей РНК, например при доставке разделенного белка Cas9 (например nCas9). В некоторых вариантах осуществления плазмида, экспрессирующая разделенный белок, может быть внесена в клетки с помощью электропорации, транзиентной (например с помощью липофекции) и стабильной геномной интеграции (например с помощью piggybac), и вирусной трансдукции, или других способов, известных специалистам в данной области.

[0753] В определенных вариантах осуществления предусмотренные здесь композиции включают липид и/или полимер. В определенных вариантах осуществления липид и/или полимер является катионным. Получение подобных липидных частиц является хорошо известным. См., например, патенты США 4880635; 4906477; 4911928; 4917951; 4920016; 4921757; и 9737604, каждый из которых приведен здесь в качестве ссылки.

[0754] Последовательность направляющей РНК может иметь длину 15-100 нуклеотидов и включать последовательность из, по меньшей мере, 10, по меньшей мере, 15, или, по

меньшей мере, 20 последовательных нуклеотидов, которая комплементарна нуклеотидной последовательности-мишени. Направляющая РНК может включать последовательность из 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, или 40 последовательных нуклеотидов, которые комплементарны нуклеотидной последовательности-мишени. Направляющая РНК может иметь длину 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, или 50 нуклеотидов.

[0755] В некоторых вариантах осуществления нуклеотидная последовательность-мишень представляет собой последовательность ДНК в геноме, например эукариотическом геноме. В определенных вариантах осуществления нуклеотидная последовательность-мишень расположена в геноме млекопитающего (например человека).

[0756] Композиции по данному изобретению могут быть введены или упакованы в виде, например, единичной дозы. Термин "единичная доза" при использовании в отношении фармацевтической композиции по настоящему изобретению обозначает физически отдельные единицы, являющиеся подходящими в качестве единичной дозы для субъекта, при этом каждая единица содержит предварительно определенное количество активного материала, рассчитанное для продуцирования требуемого терапевтического эффекта, совместно с требуемым разбавителем; то есть носителем или индифферентным веществом.

[0757] Лечение заболевания или расстройства включает замедление развития или прогрессирования заболевания, или снижение тяжести заболевания. Лечение заболевания не обязательно требует результатов в виде излечения.

[0758] В соответствии с используемым здесь значением, "замедление" развития заболевания означает задерживание, сдерживание, замедление, притормаживание, стабилизацию, и/или отсрочку прогрессирования заболевания. Данное замедление может иметь различную длительность в зависимости от истории заболевания и/или подвергаемых лечению субъектов. Способ, который обеспечивает "замедление" или облегчение развития заболевания или задержку развития заболевания, представляет собой способ, который снижает вероятность развития одного или нескольких симптомов заболевания в рамках определенного промежутка времени и/или снижение тяжести симптомов в рамках определенного промежутка времени при сравнении с отсутствием

применения данного способа. Такие сравнения обычно основаны на клинических исследованиях, использовании ряда субъектов, являющихся достаточными для получения статистически значимого результата.

[0759] "Развитие" или "прогрессирование" заболевания означает начальные проявления и/или наступающее прогрессирование заболевания. Развитие заболевания может являться детектируемым и может быть проанализировано с помощью стандартных клинических способов, которые хорошо известны в данной области. Однако термин "развитие" также обозначает прогрессирование, которое может являться недетектируемым. Для целей данного описания термины "развитие" или "прогрессирование" обозначают биологическое течение симптомов. Термин "развитие" включает возникновение, повторное возникновение, и развитие.

[0760] В соответствии с используемым здесь значением "развитие" или "возникновение" заболевания включают первоначальное развитие и/или повторное возникновение. Стандартные способы, известные средним специалистам в области медицины, могут применяться для введения выделенного полипептида или фармацевтической композиции субъекту в зависимости от типа подвергаемого лечению заболевания или участка заболевания.

[0761] Предполагается, что специалист в данной области может полностью реализовать настоящее изобретение на основе приведенного выше описания без дополнительных уточнений. Приведенные ниже конкретные варианты осуществления, таким образом, являются лишь иллюстративными и не ограничивают остальной объем описания каким-либо образом. Все упомянутые здесь публикации приведены в качестве ссылок для указанных здесь целей или описываемого предмета.

ПОСЛЕДОВАТЕЛЬНОСТИ

[0762] В рамках данной заявке подробно описано множество аминокислотных и нуклеотидных последовательностей, связанных с различными аспектами настоящего изобретения, включая примеры последовательностей Cas9, последовательностей обратной транскриптазы, последовательностей слитых белков, линкеров, направляющих РНК, и другие последовательности. Кроме того, в Примере 2 (и других разделах) описан процесс и алгоритм для конструирования и/или определения последовательности тысяч примеров

направляющих РНК праймерного редактора (PEgRNA) для репарации примеров последовательностей из базы данных ClinVar.

[0763] Данная заявка подается с Перечнем последовательностей. Перечень последовательностей включает описание каждой PEgRNA, определенной в соответствии с Примером 2. В рамках Примера 2 осуществляется определение последовательности всего 133515 примеров полной последовательности PEgRNA. Каждая из данных последовательностей присутствует / включена в Перечень последовательностей и идентифицирована в виде SEQ ID NO: 1-135514 и 813085-880462. Кроме того, и как описано в других разделах, каждая из данных PEgRNA состоит из спейсера (SEQ ID NO: 135515 – 271028 и 880463-947840) и плеча удлинения (SEQ ID NO: 271029 – 406542 и 947841-1015218). Кроме того, каждая PEgRNA включает ядро нРНК, например как определено согласно SEQ ID NO: 1361579-1361580. Каждое плечо удлинения согласно SEQ ID NO: 271029 – 406542 и 947841-1015218 дополнительно состоит из участка связывания праймера (SEQ ID NO: 406543 – 542056 и 1015219-1082596), матрицы для редактирования (SEQ ID NO.: 542057 – 677570 и 1082597-1149974), и плеча гомологии (SEQ ID NO.: 677571 – 813084 и 1149975-1217352). PEgRNA может необязательно включать 5′-концевой модификаторный участок и/или 3′-концевой модификаторный участок. PEgRNA может также включать сигнал терминации обратной транскрипции (например SEQ ID NO: 1361560-1361566) на 3'-конце PEgRNA.

[0764] Для каждой полноразмерной последовательности PEgRNA, приведенной в Перечне последовательностей (SEQ ID NO: 1 – 135514), Перечень последовательностей включает набор из пяти (5) соответствующих подпоследовательностей: а именно: (1) спейсер, (2) плечо удлинения, (3) участок связывания праймера, (4) матрица для редактирования, и (5) плечо гомологии. Набор подпоследовательностей для любой полноразмерной последовательности PEgRNA может быть определен с помощью следующей математической операции.

Определение набора подпоследовательностей для каждой PEgRNA

[0765] Для каждой последовательности PEgRNA (например SEQ ID NO: 1) в Перечне последовательностей следующие последовательности в Перечне последовательностей представляют собой набор соответствующих подпоследовательностей:

[0766] Для SEQ ID NO: 1-813084: (1) спейсер, (2) плечо удлинения, (3) участок связывания праймера, (4) матрица для редактирования, и (5) плечо гомологии являются следующими:

[0767] Спейсер: Для каждой последовательности PEgRNA соответствующая последовательность спейсера идентифицирована в виде номера идентификатора последовательности PEgRNA (например номера "1" для SEQ ID NO: 1), прибавленного к числу 135514. Например, спейсер, соответствующий PEgRNA согласно SEQ ID NO: 1, представляет собой SEQ ID NO: 135515.

[0768] Плечо удлинения: Для каждой последовательности PEgRNA соответствующее плеча удлинения идентифицирована в виде номера идентификатора последовательности PEgRNA (например номера "1" для SEQ ID NO: 1), прибавленного к числу 271028 (135514 X 2). Например, плечо удлинения, соответствующее PEgRNA согласно SEQ ID NO: 1, представляет собой SEQ ID NO: 271029.

[0769] Участок связывания праймера: Для каждой последовательности PEgRNA соответствующий участок связывания праймера идентифицирован в виде номера идентификатора последовательности PEgRNA (например номера "1" для SEQ ID NO: 1), прибавленного к числу 406542 (135514 X 3). Например, участок связывания праймера, соответствующий PEgRNA согласно SEQ ID NO: 1, представляет собой SEQ ID NO: 406542.

[0770] Матрица для редактирования: Для каждой последовательности PEgRNA соответствующая матрица для редактирования идентифицирована в виде номера идентификатора последовательности PEgRNA (например номера "1" для SEQ ID NO: 1), прибавленного к числу 542056 (135514 X 4). Например, матрица для редактирования, соответствующий PEgRNA согласно SEQ ID NO: 1, представляет собой SEQ ID NO: 542057.

[0771] Плечо гомологии: Для каждой последовательности PEgRNA соответствующее плечо гомологии идентифицировано в виде номера идентификатора последовательности PEgRNA (например номера "1" для SEQ ID NO: 1), прибавленного к числу 677570 (135514 X 5). Например, плечо гомологии, соответствующее PEgRNA согласно SEQ ID NO: 1, представляет собой SEQ ID NO: 677571.

[0772] Примеры других наборов последовательностей PEgRNA (то есть включающих любую PEgRNA из последовательностей 1-813084, каждая из которых была

сконструирована для SpCas9 (NG) (SEQ ID NO: 1361594) или SpCas9 (NGG) (SEQ ID NO: 1361593), и соответствующий спейсер, плечо удлинения, участок связывания праймера, матрицу для редактирования и плечо гомологии) приведены в следующей таблице:

Группа A	Полная PEgRNA	Спейсер	Плечо удлинения	Участок связывания праймера	Матрица для редактирования	Плечо гомологии
SEQ ID NO:	1	135515	271029	406543	542057	677571
SEQ ID NO:	2	135516	271030	406544	542058	677572
SEQ ID NO:	3	135517	271031	406545	542059	677573
SEQ ID NO:	4	135518	271032	406546	542060	677574
SEQ ID NO:	5	135519	271033	406547	542061	677575
SEQ ID NO:	6	135520	271034	406548	542062	677576

...

SEQ ID NO:	135509	271023	406537	542051	677565	813079
SEQ ID NO:	135510	271024	406538	542052	677566	813080
SEQ ID NO:	135511	271025	406539	542053	677567	813081
SEQ ID NO:	135512	271026	406540	542054	677568	813082
SEQ ID NO:	135513	271027	406541	542055	677569	813083
SEQ ID NO:	135514	271028	406542	542056	677570	813084

[0773] Со ссылкой на столбец с последовательностью "Полной PEgRNA", для SpCas9 (NG) были сконструированы следующие последовательности: SEQ ID NO: 1 - 5647, 11805 - 16732, 22103 - 25050, 28363 - 29187, 30093 - 32319, 35189 - 36933, 38922 - 39997, 41226 - 42469, 43878 - 44208, 44586 - 46456, 48645 - 49697, 50844 - 52070, 53532 - 54670, 55949 - 57576, 59335 - 60913, 62672 - 64332, 66233 - 67299, 68520 - 69273, 70195 - 72171, 74385 - 74390, 74398 - 77256, 80717 - 81275, 81899 - 81962, 82033 - 82033, 82036 - 82044, 82057 - 82063, 82072 - 82075, 82080 - 82084, 82090 - 82092, 82096 - 82100, 82106 - 82110, 82117 - 82122, 82129 - 82405, 82715 - 84431, 86323 - 86687, 87092 - 87715, 88417 - 88800, 89256 - 89791, 90405 - 92752, 95411 - 98661, 102329 - 103777, 105393 - 107009, 108826 - 109348, 109932 - 110356, 110863 - 111265, 111744 - 112224, 112822 - 113854, 115060 - 115952, 116995 - 117667, 118418 - 118426, 118436 - 119980, 121698 - 121921, 122175 - 122445, 122774 - 124123, 125657 - 126486, 127395 - 127872, 128428 - 128931, 129509 - 130164, 130892 - 131784, 132784 – 134059.

[0774] Со ссылкой на столбец с последовательностью "Полной PEgRNA", для SpCas9 (NGG) были сконструированы следующие последовательности: SEQ ID NO: 5648 - 11804, 16733 - 22102, 25051 - 28362, 29188 - 30092, 32320 - 35188, 36934 - 38921, 39998 - 41225, 42470 - 43877, 44209 - 44585, 46457 - 48644, 49698 - 50843, 52071 - 53531, 54671 - 55948, 57577 - 59334, 60914 - 62671, 64333 - 66232, 67300 - 68519, 69274 - 70194, 72172 - 74384, 74391 - 74397, 77257 - 80716, 81276 - 81898, 81963 - 82032, 82034 - 82035, 82045 - 82056, 82064 - 82071, 82076 - 82079, 82085 - 82089, 82093 - 82095, 82101 - 82105, 82111 - 82116, 82123 - 82128, 82406 - 82714, 84432 - 86322, 86688 - 87091, 87716 - 88416, 88801 - 89255, 89792 - 90404, 92753 - 95410, 98662 - 102328, 103778 - 105392, 107010 - 108825, 109349 - 109931, 110357 - 110862, 111266 - 111743, 112225 - 112821, 113855 - 115059, 115953 - 116994, 117668 - 118417, 118427 - 118435, 119981 - 121697, 121922 - 122174, 122446 - 122773, 124124 - 125656, 126487 - 127394, 127873 - 128427, 128932 - 129508, 130165 - 130891, 131785 - 132783, 134060 – 135514.

[0775] Для SEQ ID NO: 813085-1217352: (1) спейсер, (2) плечо удлинения, (3) участок связывания праймера, (4) матрица для редактирования, и (5) плечо гомологии являются следующими:

[0776] Спейсер: Для каждой последовательности PEgRNA соответствующая последовательность спейсера идентифицирована в виде номера идентификатора последовательности PEgRNA (например номера "813085" для SEQ ID NO: 813085), прибавленного к числу 67378. Например, спейсер, соответствующий PEgRNA согласно SEQ ID NO: 813085, представляет собой SEQ ID NO: 880463.

[0777] Плечо удлинения: Для каждой последовательности PEgRNA соответствующее плечо удлинения идентифицировано в виде номера идентификатора последовательности PEgRNA (например номера "813085" для SEQ ID NO: 813085), прибавленного к числу 134756 (67378 X 2). Например, плечо удлинения, соответствующее PEgRNA согласно SEQ ID NO: 813085, представляет собой SEQ ID NO: 947841.

[0778] Участок связывания праймера: Для каждой последовательности PEgRNA соответствующий участок связывания праймера идентифицирован в виде номера идентификатора последовательности PEgRNA (например номера "813085" для SEQ ID NO: 813085), прибавленного к числу 202134 (67378 X 3). Например, участок связывания

праймера, соответствующий PEgRNA согласно SEQ ID NO: 813085, представляет собой SEQ ID NO: 1015219.

[0779] Матрица для редактирования: Для каждой последовательности PEgRNA соответствующая матрица для редактирования идентифицирована в виде номера идентификатора последовательности PEgRNA (например номера "813085" для SEQ ID NO: 813085), прибавленного к числу 269512 (67378 X 4). Например, матрица для редактирования, соответствующее PEgRNA согласно SEQ ID NO: 813085, представляет собой SEQ ID NO: 1082597.

[0780] Плечо гомологии: Для каждой последовательности PEgRNA соответствующее плечо гомологии идентифицировано в виде номера идентификатора последовательности PEgRNA (например номера "813085" для SEQ ID NO: 813085), прибавленного к числу 336890 (67378 X 5). Например, матрица для редактирования, соответствующая PEgRNA согласно SEQ ID NO: 813085, представляет собой SEQ ID NO: 1149975.

[0781] Общее количество последовательностей, представленных в Перечне последовательностей, составляет 1217352. Представлены всего 202892 полных последовательностей PEgRNA (каждая из которых включает, по меньшей мере, спейсер, ядро нРНК, и плечо удлинения). Представлено одинаковое количество (1) спейсеров, (2) плечей удлинения, (3) участков связывания праймера, (4) матриц для редактирования, и (5) плечей гомологии, при этом наборы каждых из них определены как указано выше.

[0782] Примеры других наборов последовательностей PEgRNA (то есть включающих любую PEgRNA из последовательностей 813085-1217352 и соответствующий спейсер, плечо удлинения, участок связывания праймера, матрицу для редактирования и плечо гомологии) приведены в следующей таблице:

Группа B	Полная PEgRNA	Спейсер	Плечо удлинения	Участок связывания праймера	Матрица для редактирования	Плечо гомологии
SEQ ID NO:	813085	880463	947841	1015219	1082597	1149975
SEQ ID NO:	813086	880464	947842	1015220	1082598	1149976
SEQ ID NO:	813087	880465	947843	1015221	1082599	1149977
SEQ ID NO:	813088	880466	947844	1015222	1082600	1149978
SEQ ID NO:	813089	880467	947845	1015223	1082601	1149979
SEQ ID NO:	813090	880468	947846	1015224	1082602	1149980

...

SEQ ID NO:	880457	947835	1015213	1082591	1149969	1217347
SEQ ID NO:	880458	947836	1015214	1082592	1149970	1217348
SEQ ID NO:	880459	947837	1015215	1082593	1149971	1217349
SEQ ID NO:	880460	947838	1015216	1082594	1149972	1217350
SEQ ID NO:	880461	947839	1015217	1082595	1149973	1217351
SEQ ID NO:	880462	947840	1015218	1082596	1149974	1217352

[0783] Каждая из последовательностей SEQ ID NO: 813085-1217352 была сконструирована для SaCas9-KKH (SEQ ID NO: 1361596).

[0784] Подаваемый Перечень последовательностей предназначен и фактически образует часть настоящего описания при подаче.

[0785] Описание содержания Перечня последовательностей (то есть списка) соответствует Таблице XY:

SEQ ID NO:*	**Описание молекулы**	**Классификация**
1-135514	PEgRNA - полная последовательность	PEgRNA или ее компонент
135515-271028	Спейсер	"
271029-406542	Плечо удлинения	"
406543-542056	Участок связывания праймера	"
542057-677570	Матрица для редактирования	"
677571-813084	Плечо гомологии	"
813085- 880462	PEgRNA - полная последовательность	"
880463- 947840	Спейсер	"
947841- 1015218	Плечо удлинения	"
1015219- 1082596	Участок связывания праймера	"
1082597- 1149974	Матрица для редактирования	"
1149975- 1217352	Плечо гомологии	"
1217353-1289387	Связанные с заболеванием варианты (входн.) - вариант + соотв. последовательность	Приоритетный связанный с заболеванием вариант из базы данных ClinVar**, включая 200 bp выше и ниже от связанного с заболеванием варианта***
1289388-1361420	Здоровые аллели (выходн.) - здоровый аллель + соотв. последовательность	Соответствующий здоровый аллель, включая 200 bp

		выше и ниже от здорового аллеля***
1361421-1361428	Стандартный SpCas9 дикого типа	napDNAbp/ Cas9
1361429-1361442	Ортологи Cas9 дикого типа	napDNAbp/ Cas9
1361443-1361444	Вариант в виде мертвого Cas9	napDNAbp/ Cas9
1361445-1361456, and 1361593-1361596	Другие варианты Cas9	napDNAbp/ Cas9
1361457-1361462	Варианты Cas9 малого размера	napDNAbp/ Cas9
1361463-1361471	Эквиваленты Cas9	napDNAbp/ Cas9
1361472-1361474	Cas9 с подвергнутым экспансии PAM	napDNAbp/ Cas9
1361475-1361484	Циркулярные пермутанты Cas9	napDNAbp/ Cas9
1361485-1361496	Обратные транскриптазы дикого типа	RT
1361497-1361514, and 1361597-1361598	Варианты обратных транскриптаз	RT
1361515-1361519, and 1361602	Слитые белки PE	Продукт слияния PE
1361520-1361530, 1361585, and 1361603	Линкеры	Линкеры
1361531-1361541, and 1361659-1361664	Последовательность ядерной локализации	NLS
1361542-1361547	Флэп-эндонуклеазы (FEN1/GEN1/ERCC5, и так далее)	Флэп-эндонуклеазы
1361548-1361559	Участки-мишени и последовательности-мишени	PEgRNA
1361560-1361565	Терминаторы транскрипции	PEgRNA
1361566-1361578	Различные PEgRNA	PEgRNA
1361579-1361581	Конструирование PEgRNA на основе алгоритма	PEgRNA
1361582-1361584	Комплексы PE	Комплексы PE
1361586-1361592	Последовательности меток	МЕТКА
1361665-1361667		TREX
1361668-1361670		EXO1
1361671-1361678		ИНТЕИН
1361679-1361680		Шпилька
1361681		Урацил
1361682-1361685		IBR
1361604-1361658	Последовательности, которые указаны только на фигурах	SEQ на фигуре

Участки распознавания PAM

Для каждой PEgRNA Cas9 может иметь ассоциированные с ней участки распознавания PAM. В некоторых вариантах осуществления участок распознавания PAM представляет собой NGG. В некоторых вариантах осуществления участок распознавания PAM представляет собой NG. В некоторых вариантах осуществления участок распознавания PAM представляет собой KKH. В следующей таблице указано соответствие участков PAM, на которые нацелена PEgRNA по изобретению:

Таблица: XX: соответствие PAM

	NG	**NGG**	**KKH**
SEQ ID NO:	1 - 5467	5648 - 11804	813085 - 1217352
SEQ ID NO:	11805 - 16732	16733 - 22102	
SEQ ID NO:	22103 - 25050	25051 - 28362	
SEQ ID NO:	28363 - 29187	29188 - 30092	
SEQ ID NO:	30093 - 32319	32320 - 35188	
SEQ ID NO:	35189 - 36933	36934 - 38921	
SEQ ID NO:	38922 - 39997	39998 - 41225	
SEQ ID NO:	41226 - 42469	42470 - 43877	
SEQ ID NO:	43878 - 44208	44209 - 44585	
SEQ ID NO:	44586 - 46456	46457 - 48644	
SEQ ID NO:	48645 - 49697	49698 - 50843	
SEQ ID NO:	50844 - 52070	52071 - 53531	
SEQ ID NO:	53532 - 54670	54671 - 55948	
SEQ ID NO:	55949 - 57576	57577 - 59334	
SEQ ID NO:	59335 - 60913	60914 - 62671	
SEQ ID NO:	62672 - 64332	64333 - 66232	
SEQ ID NO:	66233 - 67299	67300 - 68519	
SEQ ID NO:	68520 - 69273	69274 - 70194	
SEQ ID NO:	70195 - 72171	72172 - 74384	

	NG	NGG	KKH
SEQ ID NO:	74385 - 74390	74391 - 74397	
SEQ ID NO:	74398 - 77256	77257 - 80716	
SEQ ID NO:	80717 - 81275	81276 - 81898	
SEQ ID NO:	81899 - 81962	81963 - 82032	
SEQ ID NO:	82033 - 82033	82034 - 82035	
SEQ ID NO:	82036 - 82044	82045 - 82056	
SEQ ID NO:	82057 - 82063	82064 - 82071	
SEQ ID NO:	82072 - 82075	82076 - 82079	
SEQ ID NO:	82080 - 82084	82085 - 82089	
SEQ ID NO:	82090 - 82092	82093 - 82095	
SEQ ID NO:	82096 - 82100	82101 - 82105	
SEQ ID NO:	82106 - 82110	82111 - 82116	
SEQ ID NO:	82117 - 82122	82123 - 82128	
SEQ ID NO:	82129 - 82405	82406 - 82714	
SEQ ID NO:	82715 - 84431	84432 - 86322	
SEQ ID NO:	86323 - 86687	86688 - 87091	
SEQ ID NO:	87092 - 87715	87716 - 88416	
SEQ ID NO:	88417 - 88800	88801 - 89255	
SEQ ID NO:	89256 - 89791	89792 - 90404	
SEQ ID NO:	90405 - 92752	92753 - 95410	
SEQ ID NO:	95411 - 98661	98662 - 102328	
SEQ ID NO:	102329 - 103777	103778 - 105392	
SEQ ID NO:	105393 - 107009	107010 - 108825	
SEQ ID NO:	108826 - 109348	109349 - 109931	
SEQ ID NO:	109932 - 110356	110357 - 110862	
SEQ ID NO:	110863 - 111265	111266 - 111743	
SEQ ID NO:	111744 - 112224	112225 - 112821	

	NG	NGG	KKH
SEQ ID NO:	112822 - 113854	113855 - 115059	
SEQ ID NO:	115060 - 115952	115953 - 116994	
SEQ ID NO:	116995 - 117667	117668 - 118417	
SEQ ID NO:	118418 - 118426	118427 - 118435	
SEQ ID NO:	118436 - 119980	119981 - 121697	
SEQ ID NO:	121698 - 121921	121922 - 122174	
SEQ ID NO:	122175 - 122445	122446 - 122773	
SEQ ID NO:	122774 - 124123	124124 - 125656	
SEQ ID NO:	125657 - 126486	126487 - 127394	
SEQ ID NO:	127395 - 127872	127873 - 128427	
SEQ ID NO:	128428 - 128931	128932 - 129508	
SEQ ID NO:	129509 - 130164	130165 - 130891	
SEQ ID NO:	130892 - 131784	131785 - 132783	
SEQ ID NO:	132784 - 134059	134060 - 135514	

ПРИМЕРЫ

ПРИМЕР 1. ПРАЙМЕРНОЕ РЕДАКТИРОВАНИЕ (PE) ДЛЯ ВНЕСЕНИЯ ТОЧНЫХ НУКЛЕОТИДНЫХ ИЗМЕНЕНИЙ В ГЕНОМ

[0786] Целью является разработать трансформативную технологию геномного редактирования для точного и универсального внесения изменений отдельных нуклеотидов в геномах млекопитающих. Данная технология позволила бы ученым исследовать эффекты вариаций отдельных нуклеотидов практически в любом гене млекопитающих и потенциально обеспечила бы возможность терапевтического вмешательства для корректирования патогенных точечных мутаций у представляющих собой людей пациентов.

[0787] Использование системы на основе коротких регулярно расположенных группами палиндромных повторов (CRISPR) для геномного редактирования произвело революцию в биологических науках[1-3]. Несмотря на то, что манипулирование генами с помощью

CRISPR теперь является обыденным, точное внесение изменений отдельных нуклеотидов остается трудной задачей, несмотря на то, что оно является необходимым для исследования или корректирования большого количества вызывающих заболевания мутаций. Направленная гомологичная репарация (HDR) способна обеспечивать реализацию подобных изменений, но связана с низкой эффективностью (зачастую <5%), потребностью в донорских матрицах для репарации ДНК, и негативными эффектами образования двуцепочечных разрывов ДНК (DSB). Недавно в лаборатории профессора Liu было разработано редактирование оснований, которое обеспечивает достижение эффективного редактирования отдельных нуклеотидов без DSB. В рамках редакторов оснований (BE) система CRISPR комбинируется с ферментами в виде модифицирующей основания дезаминазы для конвертации пар-мишеней оснований C•G или A•T в A•T или G•C, соответственно[4-6]. Несмотря на то, что они уже широко используются исследователями по всему миру (в виде > 5000 BE конструктов лаборатории профессора Liu, поставленных Addgene), текущие BE обеспечивают конверсию лишь четырех из двенадцати возможных пар оснований и не способны корректировать малые вставки или делеции. Более того, объем нацеливания при редактировании оснований ограничен редактированием не являющихся мишенями оснований C или A, являющихся смежными с основанием-мишенью ("побочное редактирование"), а также необходимостью наличия последовательности PAM на расстоянии 15±2 bp от основания-мишени. Таким образом, преодоление данных ограничений значительно расширит варианты исследовательского и терапевтического использования редактирования оснований для геномного редактирования.

[0788] Здесь предлагается разработать новый подход точного редактирования, который обеспечивает множество преимуществ редактирования оснований, а именно устранение двуцепочечных разрывов и потребности в донорных матрицах для репарации ДНК, в то же время обеспечивая преодоление его основных ограничений. Для достижения данной амбициозной цели предлагается напрямую вносить подвергнутые редактированию цепи ДНК в геномные участки-мишени с помощью мишень-инициированной обратной транскрипции (TPRT). В рамках описываемой здесь конструкции направляющая РНК CRISPR (нРНК) будет сконструирована таким образом, чтобы нести матрицу, кодирующую синтез мутагенной цепи ДНК, для выполнения ассоциированным

ферментом в виде обратной транскриптазы (RT). Никированный нуклеазой (Cas9) CRISPR участок-мишень ДНК будет служить в качестве праймера для обратной транскрипции, обеспечивая прямое встраивание любого требуемого нуклеотидного изменения.

Эксперимент 1

[0789] Реализация обратной транскрипции мутагенных цепей ДНК с направляющей РНК в качестве матрицы. Предшествующие исследования показали, что после расщепления ДНК, но до диссоциации комплекса Cas9 высвобождает не являющуюся мишенью цепь ДНК для высвобождения свободного 3′-конца. Предполагается, что данная цепь ДНК является доступной для удлинения ферментами в виде полимеразы и что нРНК могут быть сконструированы путем удлинения их 5′ или 3′-конца для использования в качестве матриц для синтеза ДНК. В предварительных *in vitro* исследованиях было определено, что никированные цепи ДНК в комплексах в виде связанных Cas9:нРНК действительно могут праймировать обратную транскрипцию с использованием связанной нРНК в качестве матрицы (фермент RT *in trans*). Далее будут определены различные линкеры нРНК , участки связывания праймера, и матрицы для редактирования для определения оптимальных правил конструирования *in vitro*. Затем будут оценены *in vitro* различные ферменты RT, действующие *in trans* или в виде продуктов слияния с Cas9. Наконец будут идентифицированы конструкции модифицированной нРНК, которые сохраняют эффективную активность в виде связывания и осуществления разрезов в клетках. Успешное достижение данной цели заложит основу для реализации синтеза мутагенных цепей в клетках.

Эксперимент 2

[0790] Реализация праймерного редактирования в клетках человека. На основе механизмов процессирования и репарации ДНК предполагается, что мутагенные цепи ДНК (одноцепочечные флэпы) могут применяться для обеспечения специфичного и эффективного редактирования нуклеотидов-мишеней. В рамках внушающих оптимизм предварительных исследования была определена реализуемость данной стратегии на основе демонстрации редактирования с использованием модели в виде плазмидных субстратов, содержащих мутагенные флэпы. Одновременно с Экспериментом 1 будут дополнительно проанализированы результаты репарации на основе систематически

варьирующейся длины мутагенного флэпа, композиции последовательностей, идентичности целевых нуклеотидов, и 3′-конца. Также будут протестированы малые вставки и делеции длиной от 1 до 3 нуклеотидов. Параллельно и отдельно от Эксперимента 1 будет проанализирована архитектура Cas9-RT, включая слитые белки и стратегии нековалентного рекрутирования. Архитектура Cas9-RT и удлиненные нРНК будут проанализированы на предмет клеточного редактирования в нескольких участках-мишенях в геноме человека и затем будут оптимизированы для высокой эффективности. В случае успеха данная цель немедленно сделает возможным геномное редактирование на основе TPRT для использования в рамках фундаментальных наук.

Эксперимент 3

[0791] Реализация специфического в отношении участка редактирования патогенных мутаций в культивируемых клетках человека. Потенциальная универсальность данной технологии может обеспечивать редактирование трансверсивных мутаций и инделей, которые в настоящее время не могут быть скорректированы с помощью BE. На основе результатов, полученных в рамках Эксперимента 1 и Эксперимента 2, патогенные трансверсивные мутации будут являться мишенями в рамках культивируемых клеток человека, включая вызывающую серповидноклеточную анемию мутацию в бета-глобине (для корректирования требуется трансверсия A•T на T•A) и наиболее частую мутацию при болезни Уилсона в ATP7B (для корректирования требуется трансверсия G•C на T•A). Также будет проанализирована коррекция мутаций в виде малых вставок и делеций , включая 3-нуклеотидную делецию ΔF508 в CFTR, которая вызывает муковисцидоз. В случае успеха это заложит основу для разработки мощных терапевтических подходов, которые направлены на лечение данных важных заболеваний человека.

Подход

[0792] Целью является разработать стратегию геномного редактирования, которая обеспечивает прямое внесение точечных мутаций в геномных участках-мишенях. В рамках фазы разработки технологии усилия сосредоточатся на конструировании белка и РНК для встраивания функциональности TPRT в систему CRISPR/Cas. Будут применяться in vitro анализы для осторожной оценки функции каждого этапа TPRT начиная с самого начала (Эксперимент 1). На втором этапе будут оценены результаты редактирования в клетках млекопитающих с применением комбинации модели в виде субстратов и

модифицированных CRISPR/Cas систем (Эксперимент 2). Наконец, на этапе использования технология будет применяться для корректирования мутаций, которые являлись недостижимыми при геномном редактировании с использованием других способов (Эксперимент 3).

[0793] Общая схема редактирования показана на Фиг.1A-1B. Никазы Cas9 содержат деактивирующие мутации в домене нуклеазы HNH (Spy Cas9 H840A или N863A), рестриктирующие расщепление ДНК до PAM-содержащей цепи (не являющейся мишенью цепи). Направляющие РНК (нРНК) сконструированы для включения матрицы для обратной транскрипции. Показано 5' удлинение нРНК, но также могут быть реализованы и 3' удлинения. Никаза Cas9 слита с ферментом в виде обратной транскриптазы (RT) через C-конец или N-конец. Комплекс нРНК:Cas9-RT нацелен на представляющий интерес участок ДНК и образует R-петлю после вытеснения не являющейся мишенью цепи. Cas9 разрывает не являющуюся мишенью цепь ДНК. Высвобождение никированной цепи освобождает свободный 3′-OH конец, который может праймировать обратную транскрипцию с использованием удлиненной нРНК в качестве матрицы. Данная реакция синтеза ДНК осуществляется с помощью слитого фермента в виде RT. нРНК матрица кодирует последовательность ДНК, которая гомологична исходному ДНК дуплексу за исключением нуклеотида, который является мишенью для редактирования. Продукт обратной транскрипции представляет собой одноцепочечный ДНК-флэп, который кодирует требуемое изменение. Данный флэп, который содержит свободный 3′-конец, может уравновешиваться со смежной цепью ДНК, что приводит к образованию 5′ флэпа. Предполагается, что данный вид флэпов служит в качестве эффективного субстрата для FEN1 (флэп-эндонуклеазы 1), представляющей собой фермент, который естественным образом обеспечивает эксцизию 5′ флэпов из фрагментов Оказаки при синтезе запаздывающей цепи ДНК и удаление 5′ флэпов после синтеза с замещением цепи, который осуществляется при репарации эксцизии длинных последовательностей оснований. Лигирование никированной ДНК приводит к образованию несовпадающей пары оснований. Данный промежуточный продукт может либо подвергнуться возврату к исходной паре оснований или конверсии в требуемую подвергнутую редактированию пару за счет процессов репарации несовпадений (MMR). В

качестве альтернативы, полуконсервативная репликация ДНК может обеспечивать получение по одной копии каждого из продукта с возвратом и изменением.

1. Реализация обратной транскрипции мутагенных цепей ДНК с направляющей РНК в качестве матрицы.

Предпосылки и обоснование

[0794] В рамках предложенной стратегии геномного редактирования никированная Cas9 не являющаяся мишенью цепь ДНК (PAM-содержащая цепь, которая образует R-петлю) действует в качестве праймера для синтеза ДНК. Предполагается, что это возможно на основе нескольких фрагментов биохимических и структурных данных. Эксперименты по защите на основе нуклеазы[32], кристаллографические исследования[33], и окна редактирования оснований[4,24] продемонстрировали высокую степень гибкости и разнообразия нуклеотидов не являющейся мишенью цепи от -20 до -10 в рамках так называемой R-петли Cas9-связанного комплекса (нумерация обозначает расстояние 5′ от первого нуклеотида PAM). Более того, удаленная от PAM часть расщепленной не являющейся мишенью цепи может быть вытеснена из тесно связанных третичных комплексов при добавлении комплементарной оцДНК *in trans*[20]. Данные исследования поддерживают вывод о том, что не являющаяся мишенью цепь имеет высокую гибкость, доступна для ферментов, и что после никирования 3′-конец удаленного от PAM фрагмента высвобождается до диссоциации Cas9. Более того, предполагается, что нРНК могут быть удлинены для функционирования в качестве матриц для синтеза ДНК. Предшествующие исследования показали, что нРНК для SpCas9, SaCas9, и LbCas12a (ранее Cpf1) переносят удлинение нРНК РНК с использованием аптамеров[34], лиганд-индуцируемых саморасщепляемых рибозимов[35], и длинных некодирующих РНК36. Данное описание образует прецедент в плане двух основных свойств, которые будут использоваться. При оценке данной стратегии будут проанализированы некоторые системы CRISPR-Cas наряду с конструкциями 5′ и 3′ удлиненной нРНК с использованием комбинации *in vitro* и клеточных анализов (Фиг.2A-2C).

[0795] Конструкции модифицированных нРНК для праймерного редактирования показаны на Фиг.3A-3B. Синтез ДНК осуществляется от 5′ к 3′ с копированием, таким образом, РНК матрицы в направлении от 3′ к 5′. Конструкция 5′ удлинения содержит линкерный участок, участок связывания праймера, где отжигается никированная цепь

ДНК, и матрицу для синтеза ДНК с помощью обратной транскрипции. 3′ удлиненная нРНК содержит участок связывания праймера и матрицу обратной транскрипции. В некоторых случаях 3′ РНК шпилька ядра нРНК модифицирована для соответствия последовательности-мишени ДНК, поскольку *in vitro* эксперименты показали, что обратная транскрипция приводит к ~3-нуклеотидному удлинению в ядре нРНК для конструктов 3′ удлиненной нРНК (похоже, что модификация последовательности шпильки хорошо переносится при условии реализации компенсаторных изменений, которые сохраняют структуру РНК шпильки). Синтез ДНК осуществляется от 5′ к 3′ с добавлением нуклеотидов к 3′ ОН растущей цепи ДНК.

Предварительные результаты

[0796] Никированная Cas9 ДНК праймирует обратную транскрипцию нРНК матриц. Для оценки доступности никированной не являющейся мишенью цепи ДНК могут быть проведены *in vitro* биохимические анализы с использованием нуклеазы Cas9 из *S. pyogenes* (SpCas9) и флуоресцентно меченных Cy5 субстратов в виде ДНК дуплексов (51 пара оснований). Сначала была получена группа нРНК, содержащих 5′ удлинения с различной длиной матрицы для синтеза с помощью *in vitro* транскрипции (общая конструкция показана на Фиг.2B). Анализы сдвига электрофоретической подвижности (EMSA) с не имеющей нуклеазной активности Cas9 (dCas9) показали, что 5′ удлиненные нРНК сохраняют аффинность связывания мишени (данные не показаны). Далее была протестирована активность TPRT на подвергнутых предварительному никированию меченных Cy5 субстратах в виде ДНК дуплексов с использованием dCas9, 5'-удлиненных нРНК, и обратной транскриптазы вируса лейкоза мышей Молони (M-MLV) (Superscript III). После 1 часа инкубирования при 37 °C продукты анализировали с помощью денатурирующего электрофореза на основе полиакриламидного геля (PAGE) и подвергали визуализации на основе флуоресценции Cy5 (Фиг.4A). Каждый из вариантов 5′-удлиненной нРНК приводил к значительному образованию продукта, при этом наблюдаемые размеры продукта в виде ДНК соответствовали длине матрицы в виде удлинения (Фиг.4B). Является важным то, что при отсутствии dCas9 подвергнутые предварительному никированию субстраты удлинялись до полной соответствующей 51 bp длины ДНК субстрата, что является сильным свидетельством того, что комплементарная цепь ДНК, а не нРНК, была использована в качестве матрицы для синтеза ДНК, когда

dCas9 отсутствовал (Фиг.4C). Следует отметить, что система была сконструирована таким образом, что новая синтезированная цепь ДНК отражает продукт, который будет необходим для редактирования участка-мишени (гомологичной цепи с изменением одного нуклеотида). Данный результат приводит к тому, что связывание Cas9:нРНК высвобождает 3′-конец никированной не являющейся мишенью цепи, а также к тому, что не являющаяся мишенью цепь является доступной для обратной транскрипции.

[0797] Далее оценивали не подвергнутые никированию дцДНК субстраты с использованием мутанта Cas9(H840A), который осуществляет разрыв не являющейся мишенью цепи ДНК. Сначала для тестирования никазной активности Cas9(H840A) с 5′-удлиненными нРНК проводили *in vitro* анализы расщепления, как описано ранее[37]. Несмотря на то, что никирование было нарушено по сравнению со стандартной нРНК, были образованы значительные продукты расщепления (Фиг.4D). Является важным то, что продукты RT также наблюдались при осуществлении TPRT реакций с 5′-удлиненными нРНК и Cas9(H840A), хотя и с более низким выходом, который, вероятно, может быть объяснен сниженной никирующей активностью (Фиг.4D). Данный результат показывает, что комплексы в виде 5′-удлиненная нРНК:Cas9(H840A) могут никировать ДНК и служить матрицей для обратной транскрипции.

[0798] Наконец 3′ удлинения нРНК оценивали на предмет никирования Cas9(H840A) и TPRT. В отличие от 5′-удлиненных нРНК, расщепление ДНК за счет 3′-удлиненных нРНК не ухудшалось до какой-либо детектируемой степени по сравнению со стандартной нРНК. Является значимым то, что матрицы в виде 3′-удлиненных нРНК также поддерживают эффективную обратную транскрипцию как с подвергнутыми предварительному никированию, так и интактными субстратами в виде ДНК дуплексов при поступлении RT M-MLV *in trans* (Фиг.4E). Удивительным является то, что только один продукт наблюдался для 3′-удлиненных матриц, что свидетельствует о том, что обратная транскрипция терминируется в определенном участке в рамках каркаса нРНК. Добавление гомополимерного хвоста продукта с использованием концевой трансферазы с последующим удлинением фрагментом Klenow и секвенированием по Сэнгеру показало, что копировалась полная нРНК матрица для синтеза наряду с 3 концевыми нуклеотидами ядра нРНК. В будущем флэп-конец будет перепрограммирован для модифицирования концевой последовательности нРНК[38,39]. Этот результат показывает, что 3′-удлиненные

нРНК могут служить в качестве эффективных направляющих для нацеливания нуклеаз и могут служить матрицами для обратной транскрипции.

[0799] Cas9-TPRT использует никированную ДНК и нРНК in cis. Эксперименты на основе двойного цвета использовали для определения того, осуществляется ли предпочтительным образом RT реакция с нРНК in cis (связывание с тем же комплексом) (см. Фиг.8). Два отдельных эксперимента проводили для 5′-удлиненных и 3′-удлиненных нРНК. Для указанного эксперимента в отдельных пробирках получали третичные комплексы dCas9, нРНК, и ДНК субстрата. В одной пробирке нРНК кодирует длинный RT продукт, а ДНК субстрат помечен Cy3 (красный); в другой пробирке нРНК кодирует короткий RT продукт, а ДНК субстрат помечен Cy5 (синий). После короткого инкубирования комплексы смешивали и затем обрабатывали ферментом RT и dNTP. Продукты разделяли с помощью денатурирующего PAGE с применением мочевины и визуализировали с помощью флуоресценции в Cy3 и Cy5 каналах. Определяли, что продукты реакции предпочтительно образуются с использованием нРНК матрицы, которая предварительно образовывала комплексы с ДНК субстратом, что свидетельствует о том, что RT реакция вероятно происходит in cis. Эти результаты поддерживают вывод о том. что один комплекс Cas9:нРНК может нацеливаться на участок ДНК и служить матрицей для обратной транскрипции мутагенной цепи ДНК.

(viii) Тестирование TPRT с другими Cas системами

[0800] Эксперименты, аналогичные описанным в предыдущих разделах, будут проведены с использованием других Cas систем, включая Cas9 из *S. aureus* и Cas12a из *L. bacterium* (см. Фиг.2A-2C). Если TPRT также будет продемонстрирована для данных вариантов Cas, то потенциальный объем редактирования и вероятность общего успеха в клетках будут повышены.

(ix) Тестирование TPRT со слитыми белками RT-Cas9

[0801] Сначала будет проанализирована группа коммерчески доступных или способных быть подвергнутыми очистке ферментов RT *in trans* на предмет TPRT активности. Помимо уже протестированной RT из M-MLV, будет протестирована RT из вируса миобластоза птиц (AMV), интрон Группы II *Geobacillus stearothermophilus* (GsI-IIC)[41,42], и интрон Группы II *Eubacterium rectale* (Eu.re.I2)[43,44]. Является значимым то, что последние две RT реализуют TPRT в рамках их природных биологических контекстов. Где это

возможно будут протестированы деактивирующие РНКазу мутации и другие потенциально полезные модификации ферментов RT. После идентификации функциональных RT при их поступлении *in trans* каждая из них будет проанализирована в виде слитого белка с вариантами Cas9. Будут протестированы как N-концевая, так и N-концевая ориентация продукта слияния наряду с различной длиной и архитектурой линкера. Будут использованы эксперименты на основе кинетического времени для определения того, может ли TPRT осуществляться при использовании ферментов RT in cis. Если может быть сконструирована архитектура продукта слияния RT-Cas9, обеспечивающая эффективные химические процессы в рамках TPRT, то это значительно повысит вероятность функционального редактирования в контексте клетки.

(x) Нацеливание Cas9 с использованием модифицированных нРНК в клетках

[0802] Модифицированные нРНК-кандидаты, разработанные в рамках предыдущих подзадач, будут проанализированы в рамках эксперимента на основе культуры клеток человека (HEK293) для подтверждения эффективности нацеливания Cas9. Путем использования стандартных анализов образования инделей на основе SpCas9 дикого типа[45], модифицированные нРНК будут подвергнуты всестороннему сравнению со стандартными нРНК в рамках 5 или более участков в геноме человека. Эффективность геномного редактирования будет описана на основе мультиплексного секвенирования ампликонов с помощью расположенной в лаборатории платформы Illumina MiSeq. Предполагается, что результаты, полученные в рамках этого и предыдущих разделов, обеспечат возможность получения информации для последующих итераций цикла разработки-получения-тестирования, в рамках которых нРНК могут быть оптимизированы как в плане обратной транскрипции на основе матриц, так и эффективного нацеливания Cas9 в клетках.

[0803] Результаты *in vitro* анализов показаны на Фиг.5-7. In vitro эксперименты продемонстрировали, что никированная не являющаяся мишенью цепь ДНК является гибкой и доступной для праймирования синтеза ДНК и что удлинение нРНК может служить матрицей для обратной транскрипции (см. Фиг.5). В рамках данного набора экспериментов использовались 5′-удлиненные нРНК (сконструированные как показано на Фиг.3А-3В) с матрицами для синтеза различной длины (перечислены слева). В данном наборе экспериментов флуоресцентно меченные (Cy5) ДНК-мишени использовали в

качестве субстратов и подвергали предварительному никированию. Использованный в этих экспериментах Cas9 представляет собой каталитически мертвый Cas9 (dCas9), поэтому он не может осуществлять разрывы ДНК, но все же может эффективно связываться. Superscript III, являющийся коммерческой RT, получаемой на основе вируса лейкоза мышей Молони (M-MLV), поступал in trans. Сначала из очищенных компонентов были образованы комплексы dCas9:нРНК. Затем добавляли флуоресцентно меченный субстрат ДНК совместно с dNTP и ферментом RT. Спустя 1 час инкубирования при 37 °C продукты реакции анализировали с помощью денатурирующего электрофореза на основе полиакриламидного геля с применением мочевины (PAGE). Изображение на геле показывает удлинение исходной цепи ДНК до показателей длины, которые соответствуют длине матрицы обратной транскрипции. Следует отметить, что реакции осуществляли при отсутствии полученных с помощью dCas9 продуктов ДНК с длиной в виде 51 нуклеотида независимо от использованной нРНК. Данный продукт соответствует использованию комплементарной цепи ДНК в качестве матрицы для синтеза ДНК, а не РНК (данные не показаны). Таким образом, для обеспечения синтеза ДНК на РНК матрице необходимо связывание с Cas9. Данный набор *in vitro* экспериментов похож на эксперименты, показанные на Фиг.5, за исключением того, что ДНК субстрат не подвергнут предварительному никированию и используется никаза Cas9 (мутант SpyCas9 H840A). Как показано на геле, никаза эффективно расщепляет цепь ДНК при использовании стандартной нРНК (нРНК_0, дорожка 3). Наблюдается множество продуктов расщепления, что соответствует предыдущим биохимическим исследованиям SpyCas9. 5′ удлинение нарушает никирующую активность (дорожки 4-8), однако некоторый продукт RT все же наблюдается. На Фиг.7 показано, что 3′ удлинения поддерживают синтез ДНК и не оказывают значительного влияния на активность никазы Cas9. Подвергнутые предварительному никированию субстраты (черная стрелка) практически полностью конвертируются в продукты RT при использовании никазы dCas9 или Cas9 (дорожки 4 и 5). Более чем 50% конверсия в продукт RT (красная стрелка) наблюдается с полными субстратами (дорожка 3). Для определения длины и последовательности продукта RT полосу продукта выделяли из геля, отделяли и секвенировали. Это показывало, что RT обеспечивало 3-нуклеотидное удлинение в 3′-концевой шпильке ядра нРНК. Последующие эксперименты демонстрировали, что данные три нуклеотида могут быть

изменены для соответствия последовательностям-мишеням ДНК при условии, что были реализованы комплементарные изменения для сохранения структуры шпильки РНК.

(xi) Потенциальные сложности и альтернативы

[0804] (1) RT не функционирует в качестве продукта слияния: молекулярная скученность и/или нежелательные геометрические свойства могут препятствовать полимеразному удлинению за счет слитых с Cas9 ферментов RT. Сначала может быть протестирована оптимизация линкера. Будут проанализированы циркулярно пермутированные варианты Cas9, которые могут обеспечивать переориентацию пространственной связи между ДНК праймером, нРНК и ферментом RT. Могут быть протестированы стратегии нековалентного рекрутирования RT, как подробно описано в Эксперименте 2. (2) Сниженная эффективность нацеливания Cas за счет вариантов удлиненной нРНК: это скорее всего является проблемой для 5′-удлиненных нРНК. На основе структурных данных[24] могут быть сконструированы и подвергнуты скринингу мутанты Cas9 для идентификации вариантов с большей устойчивостью к удлинению нРНК. Кроме того, библиотеки нРНК могут быть подвергнуты скринингу в клетках на предмет линкеров, которые повышают связанную с нацеливанием активность.

Значимость

[0805] Данные предварительные результаты показывают, что никазы Cas9 и удлиненные нРНК могут инициировать мишень-инициированную обратную транскрипцию на связанных ДНК-мишенях с использованием обратной транскриптазы, поступающей *in trans*. Важно, что было определено, что связывание Cas9 является критически важным для образования продукта. Хотя это вероятно не является абсолютным требованием для геномного редактирования в клетках, дальнейшее развитие системы, включающей функцию фермента RT in cis, будет значительно повышать вероятность успеха в рамках вариантов использования на основе клеток. Реализация остальных аспектов в рамках этой Задачи заложит молекулярную основу для осуществления точного геномного редактирования в контексте генома человека.

2. Реализация праймерного редактирования в клетках человека.

Предпосылки и обоснование

[0806] В рамках предлагаемой стратегии комплекс в виде модифицированной RT-Cas9:нРНК обеспечит внесение мутагенных 3′ ДНК-флэпов в геномные участки-мишени.

Предполагается, что мутагенные 3′ флэпы, содержащие одно несовпадение, будут встроены с помощью механизма репарации ДНК за счет энергетически доступного уравнивания со смежными 5′ флэпами, которые будут предпочтительно удалены (Фиг.1B). Механизмы репликации и репарации ДНК встречают 5′ оцДНК-флэпы при процессировании фрагментов Оказаки[46] и при эксцизионной репарации длинных последовательностей оснований (LP-BER)[47]. 5′ флэпы являются предпочтительными субстратами для широко экспрессируемой флэп-эндонуклеазы FEN1, которая рекрутируется в участки репарации ДНК за счет гомотримерного скользящего зажимного комплекса PCNA[48]. PCNA также служит каркасом для одновременного рекрутирования других факторов репарации, включая ДНК-лигазу Lig1[49]. Действуя в качестве "пояса с инструментами", PCNA ускоряет последовательное расщепление и лигирование флэпов, что является критически важным для процессирования миллионов фрагментов Оказаки, образованных при каждом делении клеток[50,51]. На основе аналогии с данными природными промежуточными продуктами ДНК, предполагается, что мутагенные цепи будут встроены с помощью уравнивания с 5′ флэпами с последующей скоординированной эксцизией 5′ флэпов и лигированием. Затем должна произойти репарация несовпадений (MMR) на любой цепи с равной вероятностью, что ведет к редактированию или возврату (Фиг.1B). В качестве альтернативы сначала может произойти репликация ДНК, которая ведет напрямую к встраиванию изменения в новую синтезированную дочернюю цепь. В то время как наиболее высокий ожидаемый выход данного процесса составляет 50%, множественные попытки редактирования субстратов могут сдвигать реакцию к завершению вследствие необратимости репарации изменения.

Предварительные результаты

[0807] ДНК-флэпы индуцируют специфический в отношении участка мутагенез в модели в виде плазмидных субстратов в дрожжах и клетках HEK. Для тестирования предложенной стратегии редактирования были инициированы исследования с использованием модели в виде плазмидных субстратов, содержащих мутагенные 3′ флэпы, которые похожи на продукт TPRT. Был создан двойной флуоресцентный репортерный белок, который кодирует стоп-кодон между GFP и mCherry. Мутагенные флэпы кодируют корректирование стоп-кодона (Фиг.9A), что обеспечивает синтез mCherry. Таким образом, эффективность мутагенеза может быть количественно

проанализирована на основе соотношений GFP:mCherry. Плазмидные субстраты были получены *in vitro* и внесены в дрожжи (*S. cerevisiae*) или клетки человека (HEK293). В обеих системах наблюдался высокочастотный мутагенез (Фиг.9B), при этом выделенные колонии дрожжей содержали либо подвергнутое возврату основание, мутированное основание, или смесь обоих продуктов (Фиг.9C). Детектирование последнего свидетельствует о том, что плазмидная репликации произошла до MMR в данных случаях, и дополнительно свидетельствует о том, что эксцизия и лигирование флэпа предшествует MMR. Данный результат свидетельствует об осуществимости редактирования ДНК с использованием 3′ мутагенных цепей.

(i) Системные исследования с использованием модели в виде флэп-субстратов

[0808] На основе описанных выше предварительных результатов будет проанализирован более широкий спектр флэп-субстратов в клетках HEK для определения принципов эффективного редактирования. 3′ оцДНК-флэпы будут систематически варьироваться для определения влияния образований пар с несовпадениями, расположения мутагенного нуклеотида в рамках флэпа, и идентичности концевого нуклеотида (Фиг.9D). Также будут протестированы вставки и делеции отдельных нуклеотидов. Для анализа точности редактирования будет использоваться ампликонное секвенирование. Данные результаты помогут определить конструкцию нРНК матриц для обратной транскрипции.

[0809] *In vitro* TPRT на плазмидных субстратах ведет к эффективным результатам редактирования. TPRT реакции, разработанные в рамках Эксперимента 1, использовали для индуцирования мутагенеза в плазмидном субстрате. Реакцию осуществляли на плазмидных субстратах циркулярной ДНК (см. Фиг.10). Это исключает вероятность диссоциации цепи ДНК в качестве механизма удлинения RT в рамках предыдущих *in vitro* экспериментов. Это также обеспечивало возможность тестирования репарации ДНК флэп-субстратов в клетках. Репортерную плазмиду с двойной флуоресценцией конструировали для экспрессии в дрожжах (S. cerevisiae). Эта плазмида кодирует GFP (зеленый флуоресцентный белок) и mCherry (красный флуоресцентный белок) с промежуточным стоп-кодоном (TGA). Экспрессия данного конструкта в дрожжах продуцирует только GFP. Плазмида была использована в качестве субстрата для *in vitro* TRT [Cas9(H840A) никаза, модифицированная нРНК, фермент RT MLV, dNTP]. Удлинение нРНК кодирует мутацию в стоп-кодоне. Цепь флэпа используется для репарации стоп-кодона и, как

ожидается, будет продуцировать плазмиду, которая экспрессирует как GFP, так и mCherry в виде слитого белка. На Фиг.10 показаны дрожжевые плазмидные трансформанты с двойным флуоресцентным белком. Трансформирование исходной плазмидой или *in vitro* Cas9(H840A) никированной плазмидой приводит к получению колоний, экспрессирующих только зеленый GFP. Реакция TRT с 5′-удлиненными или 3′-удлиненными нРНК продуцирует смесь зеленых и желтых колоний. Последние экспрессируют как GFP, так и mCherry. Еще больше желтых колоний наблюдается при использовании 3′-удлиненной нРНК. Также показан положительный контроль, который не содержит стоп-кодон.

[0810] Этот результат показывает, что длинные двуцепочечные субстраты могут подвергаться TPRT и что продукты TPRT индуцируют редактирование в эукариотических клетках.

[0811] Был осуществлен другой эксперимент, аналогичный указанному выше эксперименту на основе праймерного редактирования, но вместо внесения точечной мутации в стоп-кодон праймерное редактирование обеспечивает внесение вставки (слева) или делеции (справа) в виде одного нуклеотида, которая корректирует мутацию сдвига рамки считывания и обеспечивает возможность синтеза mCherry ниже (см. Фиг.11). В обоих экспериментах использовались 3′-удлиненные нРНК. Отдельные колонии из TRT трансформантов отбирали и анализировали с помощью секвенирования по Сэнгеру (см.Фиг.12). Зеленые колонии содержали плазмиды с исходной последовательность ДНК, в то время как желтые колонии содержали точную мутацию, осуществленную с помощью нРНК для праймерного редактирования. Другие точечные мутации или индели не наблюдались.

(ii) Реализация праймерного редактирования в клетках HEK с использованием архитектур RT-Cas9

[0812] Оптимизированные конструкты из предшествующих задач будут адаптированы для экспрессии в млекопитающих и редактирования в участках-мишенях в геноме человека. Будет протестировано множество ферментов RT и архитектур слияния, помимо смежного нацеливания с использованием вторичных нРНК (являющихся усеченными для предотвращения никирования). Также будет проанализировано нековалентное рекрутирование RT с использованием системы Sun-Tag[52] и системы на основе аптамера

MS2[53]. Будут использоваться анализы образования инделей для оценки эффективности нацеливания с использованием стандартных нРНК и продуктов слияния RT-Cas9 (как описано выше). Затем для каждого участка генома будут проанализированы удлиненные нРНК и пары RT-Cas9 на предмет редактирования одного нуклеотида. Результаты редактирования будут проанализированы с помощью MiSeq.

[0813] Исходные эксперименты на основе клеток HEK были проведены с использованием продуктов слияния Cas9-RT. Редактирование с помощью компонентов, экспрессируемых в клетках, требует никазы Cas9(H840A), обратной транскриптазы (экспрессируемой в виде продукта слияния или поступающей *in trans*), и модифицированной нРНК с 3′ удлинением (см. Фиг.14). Предварительные исследования показали, что длина участка связывания праймера в удлинении нРНК являлась важной в плане повышения эффективности редактирования в клетках человека (см. Фиг.15).

(iii)Оптимизация параметров праймерного редактирования в клетках HEK

[0814] После идентификации архитектур Cas9-RT, которые могут обеспечивать реализацию праймерного редактирования в клетках, компоненты и конструкция будут оптимизированы для достижения высокоэффективного редактирования. Расположение и идентичность нуклеотидов кодируемой точечной мутации, а также общая длина новой синтезированной цепи ДНК будут варьироваться для оценки объема редактирования и потенциальных ограничений. Также будут проанализированы мутации в виде коротких вставок и делеций. Конструкты для экспрессии белков будут являться кодон-оптимизированными. В случае успеха будет обеспечено эффективное праймерное редактирование в клетках млекопитающих.

[0815] Предварительные результаты. Были сконструированы дополнительные нРНК для доведения фермента RT до более высокой местной концентрации в локусе редактирования для случая, если внутримолекулярная обратная транскрипция за счет слитого фермента RT была бы невозможной. Эти вспомогательные направляющие подвергались усечению по 5′-концу (14-15 nt спейсер), что, как было показано ранее, предотвращает образование разрезов за счет Cas9, но сохраняет связывание (см. Фиг.16). Для анализа данной стратегии был выбран локус HEK3.

(iv) Потенциальные сложности и альтернативы

[0816] 1) Расщепление нРНК в клетках: если концы удлиненной нРНК являются усеченными в клетках, то могут быть внесены стабилизирующие вторичные структуры, или могут быть протестированы синтетические нРНК со стабилизирующими модификациями. (2) Отсутствие наблюдаемого редактирования в клетках человека: будут проанализированы дополнительные стратегии, включая вторичное нацеливание продуктов слияния RT-Cas9 на смежные участки генома[54]. Кроме того, могут быть проанализированы потенциальные стратегии на основе направленной эволюции в E. coli или S. cerevisiae.

Значимость

[0817] Если праймерное редактирование может быть осуществлено в экспериментальных линиях клеток, то это окажет немедленное влияние на основные биомедицинские исследования за счет обеспечения быстрого получения и описания большого количества точечных мутаций в генах человека. Универсальность данного способа и его ортогональное окно редактирования относительно редакторов оснований обеспечит подход для внесения многих мутаций, которые недоступны в настоящее время. Более того, если праймерное редактирование может быть оптимизировано для достижения высокой эффективности и чистоты продукта, то его потенциальная применимость для корректирования связанных с заболеваниями мутаций в других типах клеток человека будет значительной.

3. Реализация специфического в отношении участка редактирования патогенных мутаций в культивируемых клетках человека.

Предпосылки и обоснование.

[0818] Большое количество патогенных мутаций не может быть скорректировано с помощью доступных в настоящее время редакторов оснований вследствие связанных с PAM ограничений или потребности в корректировании трансверсивных или индельных мутаций. При праймерном редактировании все транзиции и трансверсии являются теоретически возможными, как и малые вставки и делеции. Более того, в плане PAM окно праймерного редактирования (предполагается от -3 до +4) отличается от окна редакторов оснований (от -18 до -12) (Фиг.13). Менделевские условия, которые в настоящее время не могут быть скорректированы с помощью редакторов оснований, включают: (1) фаундер-мутацию Glu6Val в гемоглобине бета при серповидноклеточной анемии (для

корректирования требуется трансверсия A•T на T•A); (2) наиболее частый вариант His1069Gln в ATP7B при болезни Уилсона (для корректирования требуется трансверсия G•C на T•A); и (3) мутацию ΔPhe508 в CFTR, которая вызывает муковисцидоз (требуется 3-нуклеотидная вставка). Каждая из данных мишеней содержит подходящим образом расположенный PAM для нацеливания SpCas9 и праймерного редактирования.

Предварительные результаты.

(i) Редактирование T на A в клетках HEK3 не является реализуемым с помощью доступного в настоящее время редактирования оснований, но реализуемо с помощью TPRT редактирования (см. Фиг.17A-17C).

[0819] На Фиг.17A показан график, демонстрирующий % конверсии T в A в нуклеотиде-мишени после трансфекции компонентов в эмбриональные клетки почки человека (HEK). Эти данные показывают результаты с использованием N-концевого слияния обратной транскриптазы MLV дикого типа и никазы Cas9(H840A) (линкер длиной 32 аминокислоты). Эффективность редактирования была значительно повышена, когда длина участка связывания праймера удлиняется от 7 нуклеотидов до 11 или 12 нуклеотидов. Кроме того, вспомогательная направляющая A, которая располагается непосредственно выше локуса редактирования (см. Фиг.16), значительно повышает активность в виде редактирования, в частности для участков связывания праймера более короткой длины. Эффективность редактирования подвергали количественному анализу с помощью ампликонного секвенирования с использованием платформы Illumina MiSeq. На Фиг.17B также показан % конверсии T на A в нуклеотиде-мишени после трансфекции компонентов в эмбриональные клетки почки человека (HEK), однако эти данные показывают результаты с применением C-концевого слияния фермента RT. Здесь вспомогательная направляющая A не имеет такого эффекта, при этом эффективность редактирования в целом является более высокой. На Фиг.17C показаны данные, которые демонстрируют результаты с использованием N-концевого слияния обратной транскриптазы MLV дикого типа и никазы Cas9(H840A), аналогичных использованным на Фиг.17A; однако линкер между RT MLV и Cas9 имеет длину 60 аминокислот вместо 32 аминокислот.

(ii) Результаты редактирования T на A в участке HEK3 с помощью TPRT редактирования демонстрируют высокую чистоту.

[0820] На Фиг.18 показан результат анализа на основе секвенирования с использованием высокопроизводительного ампликонного секвенирования. Результаты демонстрируют наиболее часто встречающиеся генотипы подвергнутых редактированию клеток. Следует отметить, что не было получено заметных продуктов в виде инделей, при этом требуемая точечная мутация (T на A) была точно внесена без побочных изменений. Первая последовательность показывает исходный генотип. Два верхних продукта представляют собой начальный генотип, содержащий эндогенный полиморфизм (G или A). Два нижних продукта соответствуют правильным образом отредактированным генотипам.

(iii)Мутанты RT MLV обеспечивают улучшение редактирования.

[0821] Мутантные обратные транскриптазы, описанные в Baranauskas, et al (doi:10.1093/protein/gzs034), тестировали в виде C-концевых продуктов слияния с никазой Cas9(H840A) для нацеленного нуклеотидного редактирования в эмбриональных клетках почки человека (HEK). Плазмида редактора Cas9-RT была котрансфицирована с плазмидой, кодирующей 3′ направляющую РНК праймерного редактора, которая служит матрицей для обратной транскрипции. На Фиг.19 показа эффективность редактирования в нуклеотиде-мишени (синие столбцы) наряду с показателями инделей (оранжевые столбцы). WT обозначает фермент RT MLV дикого типа. Мутантные ферменты (от M1 до M4) содержат перечисленные справа мутации. Показатели редактирования количественно анализировали с помощью высокопроизводительного секвенирования ампликонов геномной ДНК.

(iv) Никирование комплементарной цепи с использованием второй нНРК обеспечивает улучшение редактирования.

[0822] В рамках данного эксперимента оценивается эффективность редактирования нуклеотида-мишени при внесении одноцепочечного разрыва в комплементарную цепь ДНК рядом с нуклеотидом-мишенью, при этом предполагается, что это направит репарацию несовпадений к предпочтительному удалению исходного нуклеотида и конвертации пары оснований с реализацией требуемого изменения. Конструкт Cas9(H840A)-RT для редактирования был котрансфицирован с двумя плазмидами, кодирующими направляющие РНК, одна из которых служит матрицей для реакции обратной транскрипции, а другая нацелена на комплементарную цепь ДНК для никирования. Было протестировано никирование на различном расстоянии от нуклеотида-

мишени (оранжевые треугольники) (см. Фиг.20). Показана эффективность редактирования в паре азотистых оснований-мишеней (синие столбцы) наряду с показателями образования инделей (оранжевые столбцы). Пример "нет" не содержит никирующей комплементарную цепь направляющей РНК. Показатели редактирования количественно анализировали с помощью высокопроизводительного секвенирования ампликонов геномной ДНК.

[0823] На Фиг.21 демонстрируются обработанные данные высокопроизводительного секвенирования, показывающие требуемую трансверсивную мутацию T на A и общее отсутствие других значительных побочных продуктов редактирования генома.

[0824] Объем. На Фиг.13 показан потенциальный объем новой технологии редактирования и сравнен с технологиями редактирования оснований на основе дезаминазы. Ранее разработанные редакторы основания нацелены на участок ~15 ± 2 bp выше PAM. Путем конвертации нуклеотидов-мишеней C или A в T или G, соответственно, ранее разработанные редакторы оснований обеспечивают все переходные мутации (конверсии A:T в G:C). Однако ранее разработанные редакторы оснований не способны вносить трансверсивные мутации (A на T, A на C, G на T, G на C, T на A , T на G, C на A, C на G). Более того, при наличии нескольких нуклеотидов-мишеней в окне редактирования могут возникнуть дополнительные нежелательные изменения.

[0825] Новая технология праймерного редактирования теоретически может обеспечивать внесение конверсии любого нуклеотида и пары оснований, а также потенциально изменений в виде малой вставки и делеции. В плане PAM окна праймерного редактирования начинаются в участке никирования ДНК (на 3 основания выше от PAM) и заканчиваются в пока ещё не определенном положении ниже PAM. Следует отметить, что данное окно редактирования отличается от окна редактирования редакторов оснований на основе дезаминазы. Поскольку TPRT системы осуществляют редактирование с помощью ферментов в виде ДНК-полимеразы, они потенциально имеют все их полезные свойства, включая универсальность и точность.

(v) Корректирование патогенных мутаций в полученных от пациента линиях клеток.

[0826] Линии клеток, несущих соответствующие мутации (серповидноклеточная анемия: CD34+ гемопоэтические стволовые клетки; болезнь Уилсона: культивированные фибробласты; муковисцидоз: культивированные клетки бронхиального эпителия) будут получены от ATCC, Coriell Biobank, или сотрудничающих аффилированных лабораторий

Harvard/Broad. Эффективность редактирования будет оценена с помощью высокопроизводительного секвенирования, при этом эффективность подвергнутого корректированию генотипа будет протестирована с помощью фенотипических анализов (ВЭЖХ на основе гемоглобина, иммуноокрашивания ATP7B, и анализа мембранного потенциала CFTR).

(vi) Описание активности в виде нецелевого редактирования.

[0827] Потенциальное нецелевое редактирование будет подвергнуто скринингу с использованием стандартных способов, таких как GUIDE-seq[55] и CIRCLE-seq[56] с применением нРНК-мишеней, соединенных с Cas9 дикого типа. При идентификации побочных мишеней данные локусы будут проанализированы в рамках подвергнутых TPRT редактированию клетках для идентификации случаев нецелевого редактирования.

(vii) Потенциальные сложности и альтернативы.

[0828] (1) Низкая эффективность редактирования: праймерные редакторы могут требовать оптимизации для каждой мишени. В этом случае могут быть протестированы библиотеки нРНК для идентификации вариантов для конкретных видов использования с наивысшими показателями функционирования. Экспрессия продукта слияния RT-Cas и ядерная локализация могут быть оптимизированы. Для ограничения нецелевого редактирования может использоваться липосомная доставка на основе RNP.

(viii) Последующие эксперименты.

[0829] Оптимизация конструкций нРНК может быть достигнута путем дополнительного анализа длины участка связывания праймера и удлинения матрицы для редактирования. Объем тестирования и универсальность будут включать различные нуклеотидные конверсии, малые вставки и делеции, а также различные положения редактирования относительно PAM и несколько участков в геноме человека. Оптимизация компонента в виде RT будет включать анализ мутаций в RT MLV для повышения активности (деактивация РНКазы H, повышение аффинности связывания праймера-матрицы, модифицирование процессирования) и новые ферменты RT (RT интронов группы II, другие ретровирусные RT).

Значимость.

[0830] Изменения в виде одного нуклеотида в отдельных генах вызывают множество генетических заболеваний. Разработка описанной здесь технологи геномного

редактирования и ее применение в рамках связанных с заболеванием типов клеток заложит основу для перехода к клиническому использованию. Для некоторых заболеваний, таких как серповидноклеточная анемия, одна точечная мутация соответствует доминантному генотипу во всей популяции. Однако для многих других генетических заболеваний наблюдается значительная гетерогенность различных точечных мутаций в одном гене у популяции пациентов, каждая из которых приводит к аналогичному фенотипу заболевания. Таким образом, в качестве универсального способа геномного редактирования, который теоретически может быть нацелен на большое количество подобных мутаций, данная технология может обеспечивать огромную потенциальную пользу для многих из таких пациентов, а также их семей. Если доказательство принципа для данных вариантов использования может быть достигнуто в клетках, то это заложит основу для исследований в рамках моделей заболеваний с использованием животных.

Преимущества

[0831] Точность: требуемое изменение напрямую кодируется в последовательности нуклеиновой кислоты. Универсальность: в теории может являться возможным осуществлять любую конверсию пар оснований, включая изменения в виде трансверсий, а также малые вставки или делеции. Существует окно редактирование, отличное от окна редактирования редакторов оснований в плане последовательности смежного с протоспейсером мотива (PAM) Cas9. Этот способ обеспечивает достижение многих возможностей редактирования на основе направленной гомологичной репарации (HDR), но без значительных ограничений HDR (которая является неэффективной в большинстве типов клеток и обычно сопровождающейся избытком нежелательных побочных продуктов, таких как индели). Кроме того, при нем не осуществляются двуцепочечные разрывы ДНК (DSB, то есть малое количество инделей, транслокаций, крупных делеций, активации p53 и так далее).

ПРИМЕР 2. ПРОЦЕСС КОНСТРУИРОВАНИЯ ТЕРАПЕВТИЧЕСКИХ НАПРАВЛЯЮЩИХ РНК ПРАЙМЕРНОГО РЕДАКТОРА (PEGRNA) И ПОТЕНЦИАЛЬНЫЕ МИШЕНИ И PEGRNA ДЛЯ КОРРЕКТИРОВАНИЯ

ПАТОГЕННЫХ ВАРИАНТОВ ГЕНОВ ЧЕЛОВЕКА С ПОМОЩЬЮ ПРАЙМЕРНОГО РЕДАКТИРОВАНИЯ

Введение

[0832] Праймерное редактирование представляет собой трансформативный инструмент для геномного редактирования. Среди многих его возможных вариантов использования праймерное редактирование представляет собой новую стратегию корректирования патогенных мутаций с потенциальной терапевтической пользой. Clinvar представляет собой публично доступную базу данных описанных мутаций человека, которые могут быть ассоциированы с заболеваниями.

[0833] В данном Примере демонстрируется составление компьютерной программы/алгоритма для определения последовательностей PEgRNA с использованием простой процедуры и его применение для 436042 уникальных мутаций из Clinvar. В данном Примере сначала описана структура и функциональные элементы PEgRNA и затем описана процедура составления, включая определения для описания компьютерной программы/алгоритма, в рамках которого используется простая процедура для конструирования потенциально терапевтической PEgRNA, а также варианты конструкции PEgRNA.

[0834] Кроме того, в данном Примере приведен подаваемый Перечень последовательностей, включающий примеры последовательностей PEgRNA для корректирования потенциально патогенных мутаций из Clinvar. Описание Перечня последовательностей приведено выше.

Структура PEgRNA

[0835] На Фиг.27 и 28 приведена схема двух возможных конфигураций PEgRNA, которые могут быть сконструированы в соответствии с этим Примером.

[0836] На Фиг.27 приведена структура варианта осуществления предусмотренной здесь PEgRNA, которая может быть сконструирована в соответствии с методологией, определенной в рамках Примера 2. PEgRNA включает три основных компонента в направлении от 5′ к 3′, а именно: спейсер, ядро нРНК, и плечо удлинения на 3′-конце. Плечо удлинения может быть дополнительно разделено на следующие структурные элементы в направлении от 5′ к 3′, а именно: участок связывания праймера (A), матрицу

для редактирования (B), и плечо гомологии (C). Кроме того, PEgRNA может включать необязательный 3′-концевой модификаторный участок (e1) и необязательный 5′-концевой модификаторный участок (e2). Кроме того, PEgRNA может включать сигнал терминации транскрипции на 3′-конце PEgRNA (не показан). Данные структурные элементы дополнительно определены здесь. Изображение структуры PEgRNA не является ограничивающим и охватывает вариации в плане расположения элементов. Например, необязательные модификаторы последовательности (e1) и (e2) могут быть расположены внутри или между любыми из показанных участков без ограничения в виде расположения на 3′- или 5′-концах.

[0837] На Фиг.28 приведена структура другого варианта осуществления предусмотренной здесь PEgRNA, которая может быть сконструирована в соответствии с методологией, определенной в рамках Примера 2. PEgRNA включает три основных компонента в направлении от 5′ к 3′, а именно: спейсер, ядро нРНК, и плечо удлинения на 3′-конце. Плечо удлинения может быть дополнительно разделено на следующие структурные элементы в направлении от 5′ к 3′, а именно: участок связывания праймера (A), матрицу для редактирования (B), и плечо гомологии (C). Кроме того, PEgRNA может включать необязательный 3′-концевой модификаторный участок (e1) и необязательный 5′-концевой модификаторный участок (e2). Кроме того, PEgRNA может включать сигнал терминации транскрипции на 3′-конце PEgRNA (не показан). Данные структурные элементы дополнительно определены здесь. Изображение структуры PEgRNA не является ограничивающим и охватывает вариации в плане расположения элементов. Например, необязательные модификаторы последовательности (e1) и (e2) могут быть расположены внутри или между любыми из показанных участков без ограничения в виде расположения на 3′- или 5′-концах.

[0838] Структурные элементы PEgRNA могут быть описаны следующим образом:

[0839] Спейсер: Часть нРНК, обычно длиной 20 nt, которая является комплементарной последовательности-мишени ДНК в геноме. Спейсер комплементарен протоспейсеру в ДНК. Спейсер может быть модифицирован различными неограничивающими способами. Например, спейсер может быть модифицирован путем добавления 5′-G для образования спейсера длиной 21 nt, когда крайний 5' нуклеотид не является G. Данная модификация улучшает транскрипцию из промотора U6.

[0840] Ядро нРНК или "каркас нРНК": Данная часть нРНК в целом располагается на 3' стороне спейсера и включает несколько шпилек, которые взаимодействуют с белком Cas9. Примеры ядер нРНК могут включать, например:

[0841] Последовательность каркаса нРНК 1: GTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC (86 nt) (SEQ ID NO: 1361579).

[0842] Последовательность каркаса 2: GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC (76 nt) (SEQ ID NO: 1361580).

[0843] Предусмотренная здесь PEgRNA может включать другие последовательности ядра нРНК, которые сохраняют такую же или аналогичную способность нРНК связывать Cas9.

[0844] Плечо удлинения: Предусмотренная здесь PEgRNA включает плечо удлинения, которое включает различные функциональные элементы, включая участок связывания праймера, матрицу для редактирования, и плечо гомологии. Плечо удлинения может быть расположено на 3´-конце ядра нРНК. В других вариантах осуществления плечо удлинения может быть расположено на 5´-конце спейсера. Плечо удлинения практически комплементарно окружающей геномной последовательности, при этом требуемое изменение соответствует некомплементарному участку. Последовательность плеча удлинения может варьироваться для включения следующих вариаций: вариаций в плане длины участка связывания праймера (см. ниже); вариаций в плане длины кодируемых участков.

[0845] Участок связывания праймера: Участок связывания праймера располагается на 3'-конце плеча удлинения PEgRNA и комплементарен оцДНК-флэпу, вытесненному спейсером после осуществления разрыва, опосредованного Cas. Гибрид оцДНК:РНК служит в качестве праймера для обратной транскрипции остатка удлинения PEgRNA в направлении полимеризации от 5' к 3'. Для *Sp*Cas9 разрыв RuvC располагается между положениями протоспейсера -4 и -3, где -1 соответствует крайнему 3' нуклеотиду, а -20 соответствует крайнему 5' нуклеотиду спейсера длиной 20 nt. Участок связывания праймера длиной X сконструирован в виде X нуклеотидов начиная с 3' положения спейсера -4 до положения спейсера -4 включительно. Стандартные вариации включают различную длину в диапазоне от около 8 nt до около 20 nt (например от 8 до 17 nt).

[0846] Матрица для редактирования: Матрица для редактирования является смежной с 5'-концом участка связывания праймера и кодирует требуемую подвергнутую редактированию последовательность (например изменение, делецию или вставку одного нуклеотида).

[0847] Плечо гомологии: Плечо гомологии располагается на 5'-конце матрицы для редактирования и является комплементарным нативному геномному окружению.

[0848] Матрица для обратной транскрипции или участок, кодирующий оцДНК-флэп, который становится встроенным в эндогенную ДНК: Удлинение PEgRNA за исключением участка связывания праймера (поскольку участок связывания праймера не становится матрицей для полимеризации оцДНК за счет обратной транскриптазы). Матрица в виде обратной транскриптазы содержит матрицу для редактирования и плечо гомологии. После отсоединения спейсера от протоспейсера праймерная оцДНК-последовательность в виде обратной транскриптазы заново связывается со своим нативным геномным окружением, оставляя матрицу в виде обратной транскриптазы в виде 3' оцДНК-флэпа. Стандартные вариации включают: может варьироваться длина флэпа, например от 7 nt до 34 nt (однако охватывается широкий диапазон длины флэпа, как описано здесь); может варьироваться положение редактирования в рамках флэпа. Успешное редактирование наблюдалось при использовании плеча гомологии длиной до 2 nt. Изменения, как правило, но не обязательно, располагаются как можно ближе к разрыву.

[0849] Последовательность терминатора транскрипции: Последовательность на 3'-конце PEgRNA, которая терминирует транскрипцию при получении PEgRNA, например при экспрессии из промотора U6 или другого промотора. Пример последовательности терминатора представляет собой TTTTTTGTTTT (SEQ ID NO: 1361581).

[0850] Дополнительные вариации конструкций PEgRNA являются возможными и охватываются здесь. Например, предполагается, что PEgRNA может быть сконструирована с плечами удлинения PEgRNA, содержащими некомплементарные последовательности 5' плеча удлинения или 3' участка связывания праймера или с ними обоими, например для осуществления взаимодействия в виде целующихся петель или для функционирования в качестве защитной шпильки для стабильности РНК. Данные последовательности показаны на Фиг.27 и 28 в виде элементов последовательности e1 и

e2 и обозначены как "необязательные 3'- или 5'-концевые модификаторные участки". Кроме того, предполагается, что PEgRNA может быть сконструирована с использованием стратегий и способов на основе приоритизации между несколькими конструкциями-кандидатами. Примеры включают избежание удлинений PEgRNA, где крайний 5' нуклеотид представляет собой цитозин вследствие нарушения нативных взаимодействий нуклеотид-белок в комплексе sgРНК:Cas9 или использование инструментов для предсказания вторичной структуры РНК для выбора предпочтительной длины PBS и длины флэпа на основе спейсера и требуемого изменения.

Алгоритм конструирования PEgRNA

[0851] С учетом входного аллеля, выходного аллеля и системы CRISPR для праймерного редактирования (являющегося важным мотива PAM и соответствующего положения разрыва праймерного редактора) алгоритм обеспечивает возможность конструирования PEgRNA или перечня PEgRNA, которые способны обеспечивать редактирование входного аллеля с получением выходного аллеля. Разница в плане последовательности между выходным и входным аллелями обозначается как требуемое изменение. Соответствующий вариант осуществления может представлять собой входной аллель, соответствующий патогенной мутации, и выходной аллель, соответствующей скорректированной последовательности дикого типа.

[0852] Классы изменений могут быть индуцированы одиночной PEgRNA, включая замещения одного нуклеотида, вставки от 1 nt до приблизительно 40 nt, делеции от 1 nt до приблизительно 30 nt, и комбинацию или смесь всех из указанных выше. Известно, что праймерное редактирование поддерживает изменения данных типов от положения спейсера -3 (то есть непосредственно 3' разрыва) до положения спейсера +27 (30 nt 3' разрыва во входном аллеле). Каждое из данных указанных чисел соответствует параметру алгоритма, который может меняться с течением времени при увеличении знаний о праймерном редактировании. С другой стороны, изменения в положении протоспейсера -4 с использованием системы SpCas9 наблюдались в рамках праймерного редактирования, которые скорее всего связаны с произвольным расщеплением RuvC между положениями протоспейсера -5 и -4.

[0853] В рамках алгоритма осуществляется нумерация всех спейсеров, совместимых с мотивом PAM выбранной системы CRISPR во входном аллеле на обеих цепях, затем

удаление протоспейсеров, ассоциированные положения разрыва которых не совместимы с праймерным редактированием с получением выходного аллеля вследствие того, что разрыв расположен на 3' стороне требуемого изменения на данной цепи, или поскольку расстояние между разрывом и требуемым изменением является слишком большим (более чем определенное пользователем пороговое значение, например 30 nt).

[0854] Для каждого спейсера в рамках алгоритма осуществляется конструирование последовательности спейсера и матрицы для редактирования с использованием последовательности входного аллеля, положения разрыва, и последовательности требуемого изменения. Затем алгоритм обеспечивает выбор одного или нескольких значений длины участка связывания праймера, которая может варьироваться от 8 nt до 17 nt; длины плеча гомологии, которая может варьироваться от 2 nt до 33 nt; и последовательности каркаса нРНК, которая может представлять собой GTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATC AACTTGAAAAAGTGGCACCGAGTCGGTGC (SEQ ID NO: 1361579) или GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAA AGTGGCACCGAGTCGGTGC (SEQ ID NO: 1361580), или другой последовательности каркаса нРНК, которая сохраняет вторичную структуру РНК дикого типа. Для каждой выбранной комбинации параметров алгоритм обеспечивает конструирование последовательности плеча гомологии, участка связывания праймера, и каркаса нРНК с использованием заданных параметров. Затем алгоритм обеспечивает получение последовательности PEgRNA за счет конкатенации последовательности протоспейсера, каркаса нРНК, удлинения PEgRNA, и терминатора.

[0855] После простой фильтрации на основе мутаций зародышевых линий, обозначенных в качестве патогенных или вероятно патогенных, были идентифицированы 72020 уникальных мутаций из Clinvar, которые являются совместимыми с праймерным редактированием с использованием Cas9-NG, и 63496 уникальных мутаций из Clinvar, которые являются совместимыми с праймерным редактированием с использованием *Sp*Cas9 и NGG PAM. Следует обратить внимание, что дополнительные мутации будут являться корректируемыми при использовании праймерного редактора, содержащего другой вариант Cas9 с другой совместимостью в плане PAM.

[0856] Данные мутации были разделены на четыре класса согласно клинической значимости на основе малой частоты аллелей, количества отправителей, наличия противоречий в интерпретации отправителей, и наличия или отсутствия обзора мутации, выполненного группой экспертов.

[0857] Среди 63496 *Sp*Cas9-совместимых мутаций:

- 4627 мутаций идентифицированы на наиболее значительном уровне (четыре)
- 13943 мутации идентифицированы на уровне значимости три или четыре
- 44385 мутаций идентифицированы на уровне значимости два, три или четыре.

[0858] В приведенном Перечне последовательностей указана одна PEgRNA на каждую уникальную мутацию, выбранная в виде PEgRNA с наиболее коротким расстоянием между разрывом и изменением. PEgRNA были сконструированы с длиной плеча гомологии 13 и длиной участка связывания праймера 13. Спейсеры с участками разрыва на расстоянии более 20 nt от изменения исключались. Использованная последовательность каркаса нРНК являлась следующей GTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATC AACTTGAAAAAGTGGCACCGAGTCGGTGC (SEQ ID NO: 1361579). Использованная терминаторная последовательность являлась следующей: TTTTTTGTTTT (SEQ ID NO: 1361581).

[0859] Перечень последовательностей включает описание каждой PEgRNA, определенной в соответствии с Примером 2. В рамках Примера 2 осуществляется определение последовательности всего 133515 примеров полной последовательности PEgRNA. Каждая из данных последовательностей присутствует / включена в Перечень последовательностей и идентифицирована в виде SEQ ID NO: 1-135514. Кроме того, и как описано в других разделах, каждая из данных PEgRNA включает спейсер (SEQ ID NO: 135515 – 271028) и плечо удлинения (SEQ ID NO: 271029 – 406542). Кроме того, каждая PEgRNA включает ядро нРНК, например как определено согласно SEQ ID NO: 1361579-1361580. Каждое плечо удлинения согласно SEQ ID NO: 271029 – 406542 дополнительно состоит из участка связывания праймера (SEQ ID NO: 406543 – 542056), матрицы для редактирования (SEQ ID NO.: 542057 – 677570), и плеча гомологии (SEQ ID NO.: 677571 – 813084). PEgRNA может необязательно включать 5′-концевой модификаторный участок (SEQ ID NO: EE-EE) и/или 3′-концевой модификаторный участок (SEQ ID NO: FF-FF).

PEgRNA может также включать сигнал терминации обратной транскрипции (например SEQ ID NO: 1361560-1361565) на 3'-конце PEgRNA.

[0860] Для каждой полноразмерной последовательности PEgRNA, приведенной в Перечне последовательностей (SEQ ID NO: 1 – 135514), Перечень последовательностей включает набор из пяти (5) соответствующих подпоследовательностей: а именно: (1) спейсер, (2) плечо удлинения, (3) участок связывания праймера, (4) матрицу для редактирования, и (5) плечо гомологии. Набор подпоследовательностей для любой полноразмерной последовательности PEgRNA может быть определен с помощью следующей математической операции.

[0861] Для каждой последовательности PEgRNA (например SEQ ID NO: 1) в Перечне последовательностей следующие последовательности в Перечне последовательностей представляют собой набор соответствующих подпоследовательностей: (1) спейсер, (2) плечо удлинения, (3) участок связывания праймера, (4) матрица для редактирования, и (5) плечо гомологии являются следующими:

[0862] Спейсер: Для каждой последовательности PEgRNA соответствующая последовательность спейсера идентифицирована в виде номера идентификатора последовательности PEgRNA (например номера "1" для SEQ ID NO: 1), прибавленного к числу 135514. Например, спейсер, соответствующий PEgRNA согласно SEQ ID NO: 1, представляет собой SEQ ID NO: 135515.

[0863] Плечо удлинения: Для каждой последовательности PEgRNA соответствующее плечо удлинения идентифицировано в виде номера идентификатора последовательности PEgRNA (например номера "1" для SEQ ID NO: 1), прибавленного к числу 271028 (135514 X 2). Например, плечо удлинения, соответствующее PEgRNA согласно SEQ ID NO: 1, представляет собой SEQ ID NO: 271029.

[0864] Участок связывания праймера: Для каждой последовательности PEgRNA соответствующий участок связывания праймера идентифицирован в виде номера идентификатора последовательности PEgRNA (например номера "1" для SEQ ID NO: 1), прибавленного к числу 406542 (135514 X 3). Например, участок связывания праймера, соответствующий PEgRNA согласно SEQ ID NO: 1, представляет собой SEQ ID NO: 406542.

[0865] Матрица для редактирования: Для каждой последовательности PEgRNA соответствующая матрица для редактирования идентифицирована в виде номера идентификатора последовательности PEgRNA (например номера "1" для SEQ ID NO: 1), прибавленного к числу 542056 (135514 X 4). Например, матрица для редактирования, соответствующая PEgRNA согласно SEQ ID NO: 1, представляет собой SEQ ID NO: 542057.

[0866] Плечо гомологии: Для каждой последовательности PEgRNA соответствующее плечо гомологии идентифицировано в виде номера идентификатора последовательности PEgRNA (например номера "1" для SEQ ID NO: 1), прибавленного к числу 677570 (135514 X 5). Например, матрица для редактирования, соответствующая PEgRNA согласно SEQ ID NO: 1, представляет собой SEQ ID NO: 677571.

[0867] Общее количество последовательностей, представленных в Перечне последовательностей, составляет 813084. Представлены всего 135514 полных последовательностей PEgRNA (каждая из которых включает, по меньшей мере, спейсер, ядро нРНК, и плечо удлинения). Представлено одинаковое количество (1) спейсеров, (2) плечей удлинения, (3) участков связывания праймера, (4) матриц для редактирования, и (5) плечей гомологии, при этом наборы каждых из них определены как указано выше.

[0868] Примеры других наборов последовательностей PEgRNA (то есть включающих любую PEgRNA из и соответствующий спейсер, плечо удлинения, участок связывания праймера, матрицу для редактирования и плечо гомологии) приведены в следующей таблице:

	Полная PEgRNA	Спейсер	Плечо удлинения	Участок связывания праймера	Матрица для редактирования	Плечо гомологии
SEQ ID NO:	1	135515	271029	406543	542057	677571
SEQ ID NO:	2	135516	271030	406544	542058	677572
SEQ ID NO:	3	135517	271031	406545	542059	677573
SEQ ID NO:	4	135518	271032	406546	542060	677574
SEQ ID NO:	5	135519	271033	406547	542061	677575
SEQ ID NO:	6	135520	271034	406548	542062	677576

…

SEQ ID NO:	135509	271023	406537	542051	677565	813079

SEQ ID NO:	135510	271024	406538	542052	677566	813080
SEQ ID NO:	135511	271025	406539	542053	677567	813081
SEQ ID NO:	135512	271026	406540	542054	677568	813082
SEQ ID NO:	135513	271027	406541	542055	677569	813083
SEQ ID NO:	135514	271028	406542	542056	677570	813084

[0869] Вариации указанных конструкций PEgRNA включают все ранее указанные вариации, включая варьирующуюся последовательность каркаса нРНК, длину участка связывания праймера, длину флэпа, и так далее.

Ссылки (для Примера 2)

[0870] Landrum, M. J., Lee, J. M., Riley, G. R., Jang, W., Rubinstein, W. S., Church, D. M., & Maglott, D. R. (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic acids research, 42(Database issue), D980–D985. doi:10.1093/nar/gkt1113

[0871] Stenson, P. D., Mort, M., Ball, E. V., Evans, K., Hayden, M., Heywood, S., … Cooper, D. N. (2017). The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. Human genetics, 136(6), 665–677. doi:10.1007/s00439-017-1779-6

[0872] Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., … Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell, 156(5), 935–949. doi:10.1016/j.cell.2014.02.001

[0873] Lorenz, R., Bernhart, S. H., Höner Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., & Hofacker, I. L. (2011). ViennaRNA Package 2.0. Algorithms for molecular biology : AMB, 6, 26. doi:10.1186/1748-7188-6-26

ПРИМЕР 3. КОНСТРУИРОВАНИЕ И МОДИФИЦИРОВАНИЕ PEGRNA ДЛЯ ПРАЙМЕРНОГО РЕДАКТИРОВАНИЯ

Краткое содержание

[0874] Здесь описана группа конструкций PEgRNA и стратегий, которые могут повышать эффективность праймерного редактирования (PE).

Предпосылки

[0875] Праймерное редактирование (PE) представляет собой технологию геномного редактирования, которая может обеспечивать замену, вставку, или удаление определенных последовательностей ДНК в генетическом локусе-мишени с использованием информации, кодируемой в направляющей РНК праймерного редактора (PEgRNA). Праймерные редакторы (PE) состоят из программируемого на основе последовательности связывающего ДНК белка с нуклеазной активностью (Cas9), слитого с ферментом в виде обратной транскриптазы (RT). PE образуют комплексы с PEgRNA, которые содержат информацию для нацеливания определенных локусов ДНК в рамках их спейсерных последовательностей, а также информацию с указанием требуемого изменения в сконструированном удлинении, встроенном в стандартный каркас sgРНК. Комплексы PE:PEgRNA связывают и никируют программируемый локус-мишень ДНК, обеспечивая гибридизацию никированной цепи ДНК со сконструированной последовательностью связывания праймера (PBS) PEgRNA. Затем домен обратной транскриптазы копирует кодирующую изменение информацию в часть в виде RT матрицы PEgRNA с использованием никированной геномной ДНК в качестве праймера для полимеризации ДНК. Последующие процессы репарации ДНК обеспечивают встраивание новой синтезированной подвергнутой редактированию цепи ДНК в геномный локус. В то время как универсальность праймерного редактирования является многообещающей в качестве исследовательского инструмента и потенциального способа лечения, существует несколько ограничений в плане эффективности и объема вследствие многоэтапного процесса, требуемого для редактирования. Например, нежелательные структуры РНК, которые могут образовываться в PEgRNA, могут приводить к ингибированию копирования изменений ДНК из PEgRNA в геномный локус. Один потенциальный способ усовершенствования технологии PE заключается в изменении конструкции и модифицировании критического компонента PEgRNA. Усовершенствования конструкции данных PEgRNA вероятно будут необходимы для повышения эффективности PE, а также обеспечения внесения более длинных вставляемых последовательностей в геном.

Описание

[0876] Здесь описана группа конструкций PEgRNA, которые, как предполагается, будут повышать эффективность PE. Данные конструкции основаны на ряде ранее описанных подходов для повышения эффективности и/или стабильности sgРНК, также на

использовании ряда новых стратегий. Данные усовершенствования могут относится к одной или нескольким из ряда различных категорий: (1) конструкциям для обеспечения эффективной экспрессии функциональных PEgRNA из не относящихся к полимеразе III (pol III) промоторов, которые будут обеспечивать экспрессию более длинных PEgRNA без обременительных требований в плане последовательности; (2) усовершенствованиям ядра, Cas9-связывающему каркасу PEgRNA, которые могут повысить эффективность; (3) модификациям PEgRNA для повышения процессирования RT, обеспечивающим вставку более длинных последовательностей в геномные локусы-мишени; и (4) добавлению мотивов РНК на 5'- или 3'-концы PEgRNA, которые повышают стабильность PEgRNA, стимулируют процессирование RT, предотвращают неправильный фолдинг PEgRNA, или рекрутируют дополнительные факторы, являющиеся важными для геномного редактирования. Здесь описан ряд подобных потенциальных конструкций PEgRNA в рамках каждой категории. Некоторые из данных конструкций были ранее описаны для повышения активности sgРНК с использованием Cas9 и отмечены соответствующим образом. Здесь также описана платформа для эволюции PEgRNA для мишеней в виде определенной последовательности, которая будет обеспечивать совершенствование каркаса PEgRNA и повышение активности PE (v). Следует отметить, что данные конструкции также могут легко применяться для усовершенствования PEgRNA, распознаваемой любым Cas9 или его подвергнутым эволюции вариантом.

(i) Экспрессия PEgRNA из не относящихся к pol III промоторов

[0877] sgРНК обычно экспрессируются из промотора мяРНК U6. Данный промотор рекрутирует pol III для экспрессии ассоциированной РНК и может применяться для экспрессии коротких РНК, которые остаются в ядре. Однако pol III не является высокопроцессивной и не способна экспрессировать РНК с длиной более чем несколько сотен нуклеотидов на уровнях, требуемых для эффективного геномного редактирования[183]. Кроме того, pol III может вызывать остановку или терминирование на длинных участках в виде U, что потенциально ограничивает разнообразие цепей, которые могут быть вставлены с помощью PEgRNA. Были проанализированы другие промоторы, которые рекрутируют полимеразу II (например pCMV) или полимеразу I (например промотор мяРНК U1) на их способность экспрессировать более длинные sgРНК[183]. Однако

данные промоторы обычно частично транскрибируются, что будет приводить к дополнительной последовательности 5' спейсера в экспрессируемой PEgRNA, что, как было показано, приводит к заметному снижению активности Cas9:sgРНК в зависимости от участка. Кроме того, в то время как транскрибированные из pol III PEgRNA могут просто приводить к терминации на участке из 6-7 U, PEgRNA, транскрибированные из pol II или pol I, будут требовать другого сигнала терминации. Зачастую такие сигналы также приводят к полиаденилированию, которое приведет к нежелательному транспорту PEgRNA из ядра. Аналогичным образом, РНК, экспрессируемые из pol II промоторов, например pCMV, обычно являются 5'-кэпированными, что также приводит к их экспорту из ядра.

[0878] Ранее Rinn и коллеги провели скрининг ряда платформ экспрессии на предмет продуцирования sgРНК, меченных длинными некодирующими РНК (днкРНК)[183]. Данные платформы включают РНК, экспрессированные из pCMV, которые приводят к терминации в элементе ENE из нкРНК MALAT1 человека[184], элементе PAN ENE из KSHV[185], или 3' боксе из sgРНК U1[186]. Следует отметить, что нкРНК MALAT1 и PAN ENE образуют тройные спирали, защищающие polyA-хвост [184, 187]. Ожидается, что помимо обеспечения экспрессии РНК, данные конструкты также могут повышать стабильность РНК (см. раздел iv). Также анализировали использование промотора из sgРНК U1 для обеспечения экспрессии данных более длинных sgРНК[183]. Предполагается, что данные системы экспрессии также обеспечат возможность экспрессии более длинных PEgRNA. Кроме того, был разработан ряд способов для расщепления части pol II промотора, которая будет транскрибирована в виде части PEgRNA, обеспечивая добавление либо саморасщепляемого рибозима, например рибозимов типа "головка молотка" (hammerhead)[188], "пистолет" (pistol)[189], "топор" (hatchet)[189], "шпилька" (hairpin)[190], VS[191], "твистер"[192], или "твистер-систер" (twister sister)[192], или других саморасщепляемых элементов для процессирования транскрибированной направляющей, или шпильки, которая распознается Csy4[193] и также ведет к процессированию направляющей. Кроме того, существует гипотеза, что встраивание нескольких мотивов ENE может приводить к повышенной экспрессии PEgRNA и стабильности, как ранее было продемонстрировано для РНК и элемента KSHV PAN[185]. Также ожидается, что циркуляризация PEgRNA в

форме циркулярной интронной РНК (циРНК) также может приводить к повышенной экспрессии РНК и стабильности, а также ядерной локализации[194].

Последовательности:

[0879] Платформа экспрессии PEgRNA, состоящая из pCMV, шпильки Csy4, PEgRNA, и MALAT1 ENE

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTT
ACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAA
TAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTA
TTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATT
GACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTC
CTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTA
CATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTC
AATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCC
CATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGT
GAACCGTCAGATCGTTCACTGCCGTATAGGCAGGGCCCAGACTGAGCACGTGAGTTTTAGAGCT
AGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTC
CTCTGCCATCAAAGCGTGCTCAGTCTGTTTTAGGGTCATGAAGGTTTTTCTTTTCCTGAGAAAA
CAACACGTATTGTTTTCTCAGGTTTTGCTTTTTGGCCTTTTTCTAGCTTAAAAAAAAAAAAAGC
AAAAGATGCTGGTGGTTGGCACTCCTGGTTTCCAGGACGGGGTTCAAATCCCTGCGGCGTCTTT
GCTTTGACT (SEQ ID NO: 1361567)

[0880] Платформа экспрессии PEgRNA, состоящая из pCMV, шпильки Csy4, PEgRNA, и PAN ENE

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTT
ACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAA
TAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTA
TTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATT
GACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTC
CTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTA
CATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTC
AATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCC
CATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGT
GAACCGTCAGATCGTTCACTGCCGTATAGGCAGGGCCCAGACTGAGCACGTGAGTTTTAGAGCT
AGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTC
CTCTGCCATCAAAGCGTGCTCAGTCTGTTTTGTTTTGGCTGGGTTTTTCCTTGTTCGCACCGGA
CACCTCCAGTGACCAGACGGCAAGGTTTTTATCCCAGTGTATATTGGAAAAACATGTTATACTT
TTGACAATTTAACGTGCCTAGAGCTCAAATTAAACTAATACCATAACGTAATGCAACTTACAAC
ATAAATAAAGGTCAATGTTTAATCCATAAAAAAAAAAAAAAAAAA (SEQ ID NO: 1361568)

[0881] Платформа экспрессии PEgRNA, состоящая из pCMV, шпильки Csy4, PEgRNA, и 3xPAN ENE

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTT
ACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAA
TAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTA
TTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATT
GACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTC
CTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTA
CATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTC
AATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCC
CATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGT
GAACCGTCAGATCGTTCACTGCCGTATAGGCAGGGCCCAGACTGAGCACGTGAGTTTTAGAGCT
AGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTC
CTCTGCCATCAAAGCGTGCTCAGTCTGTTTTGTTTTGGCTGGGTTTTTCCTTGTTCGCACCGGA
CACCTCCAGTGACCAGACGGCAAGGTTTTTATCCCAGTGTATATTGGAAAAACATGTTATACTT
TTGACAATTTAACGTGCCTAGAGCTCAAATTAAACTAATACCATAACGTAATGCAACTTACAAC
ATAAATAAAGGTCAATGTTTAATCCATAAAAAAAAAAAAAAAAAAAAACACACTGTTTTGGCTGG
GTTTTTCCTTGTTCGCACCGGACACCTCCAGTGACCAGACGGCAAGGTTTTTATCCCAGTGTAT
ATTGGAAAAACATGTTATACTTTTGACAATTTAACGTGCCTAGAGCTCAAATTAAACTAATACC
ATAACGTAATGCAACTTACAACATAAATAAAGGTCAATGTTTAATCCATAAAAAAAAAAAAAAA
AAAATCTCTCTGTTTTGGCTGGGTTTTTCCTTGTTCGCACCGGACACCTCCAGTGACCAGACGG
CAAGGTTTTTATCCCAGTGTATATTGGAAAAACATGTTATACTTTTGACAATTTAACGTGCCTA
GAGCTCAAATTAAACTAATACCATAACGTAATGCAACTTACAACATAAATAAAGGTCAATGTTT
AATCCATAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 1361569)

[0882] Платформа экспрессии PEgRNA, состоящая из pCMV, шпильки Csy4, PEgRNA, и 3′ бокса

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTT
ACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAA
TAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTA
TTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATT
GACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTC
CTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTA
CATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTC
AATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCC
CATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGT
GAACCGTCAGATCGTTCACTGCCGTATAGGCAGGGCCCAGACTGAGCACGTGAGTTTTAGAGCT
AGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTC
CTCTGCCATCAAAGCGTGCTCAGTCTGTTTGTTTCAAAAGTAGACTGTACGCTAAGGGTCATAT
CTTTTTTTGTTTGGTTTGTGTCTTGGTTGGCGTCTTAAA (SEQ ID NO: 1361570)

[0883] Платформа экспрессии PEgRNA, состоящая из pU1, шпильки Csy4, PEgRNA, и 3′ бокса

CTAAGGACCAGCTTCTTTGGGAGAGAACAGACGCAGGGGCGGGAGGGAAAAAGGGAGAGGCAGA
CGTCACTTCCCCTTGGCGGCTCTGGCAGCAGATTGGTCGGTTGAGTGGCAGAAAGGCAGACGGG
GACTGGGCAAGGCACTGTCGGTGACATCACGGACAGGGCGACTTCTATGTAGATGAGGCAGCGC
AGAGGCTGCTGCTTCGCCACTTGCTGCTTCACCACGAAGGAGTTCCCGTGCCCTGGGAGCGGGT
TCAGGACCGCTGATCGGAAGTGAGAATCCCAGCTGTGTGTCAGGGCTGGAAAGGGCTCGGGAGT
GCGCGGGGCAAGTGACCGTGTGTGTAAAGAGTGAGGCGTATGAGGCTGTGTCGGGGCAGAGGCC
CAAGATCTCAGTTCACTGCCGTATAGGCAGGGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGA
AATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTC
TGCCATCAAAGCGTGCTCAGTCTGTTTCAGCAAGTTCAGAGAAATCTGAACTTGCTGGATTTTT
GGAGCAGGGAGATGGAATAGGAGCTTGCTCCGTCCACTCCACGCATCGACCTGGTATTGCAGTA
CCTCCAGGAACGGTGCACCCACTTTCTGGAGTTTCAAAAGTAGACTGTACGCTAAGGGTCATAT
CTTTTTTTGTTTGGTTTGTGTCTTGGTTGGCGTCTTAAA (SEQ ID NO: 1361571)

(ii) Усовершенствования каркаса PEgRNA

[0884] Ядро, представляющее собой Cas9-связывающий каркас PEgRNA, может быть вероятно усовершенствовано для повышения активности PE. Некоторые подобные подходы уже были продемонстрированы. Например, первый парный элемент каркаса (P1) содержит парный элемент GTTTT-AAAAC. Было показано, что такие повторы T приводят к остановке pol III и преждевременной терминации РНК транскрипта. Было показано, что разумная мутация одной из пар T-A на пару G-C в этой части P1 повышает активность sgРНК, что свидетельствует о том, что данный подход также может применяться для PEgRNA[195]. Кроме того, также было показано, что повышение длины P1 стимулирует фолдинг sgРНК и ведет к повышенной активности[195], что свидетельствует о том, что оно является другой возможностью повысить активность PEgRNA. Наконец, усовершенствование каркаса PEgRNA за счет направленной эволюции PEgRNA на определенной ДНК-мишени вероятно также приведет к повышенной активности. Это описано в разделе (v).

Последовательности:

[0885] PEgRNA, содержащую удлинение P1 длиной 6 nt

GGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGCTCATGAAAATGAGCTAGCAAGTTAAAATAA
GGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAGCGTGCTC
AGTCTGTTTTTTT (SEQ ID NO: 1361572)

[0886] PEgRNA, содержащую мутацию T-A на G-C в P1

GGCCCAGACTGAGCACGTGAGTTTGAGAGCTAGAAATAGCAAGTTTAAATAAGGCTAGTCCGTT
ATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAGCGTGCTCAGTCTGTTTTTT
T (SEQ ID NO: 1361573)

(iii) Усовершенствования процессирования RT за счет модификаций участка PEgRNA в виде матрицы

[0887] С увеличением размера вставки, матрицей которого служит PEgRNA, она с большей вероятностью будет расщепляться эндонуклеазами, подвергаться спонтанному гидролизу или фолдингу во вторичные структуры, не способные подвергаться обратной транскрипции за счет RT или нарушающие фолдинг каркаса PEgRNA и последующее связывание Cas9-RT. Соответственно, является вероятным, что для реализации крупных вставок может являться необходимым модифицирование матрицы PEgRNA, например для вставок целых генов. Некоторые стратегии для этого включают встраивание модифицированных нуклеотидов в синтетическую или полусинтетическую PEgRNA, которые делают РНК более устойчивой к расщеплению или гидролизу или менее подверженной включению ингибиторных вторичных структур[196]. Подобные модификации могут включать 8-аза-7-деазагуанозин, который будет снижать содержание вторичной структуры РНК в последовательностях с высоким содержанием G; "заблокированные" нуклеиновые кислоты (LNA), которые снижают расщепление и повышают определенные виды вторичной структуры РНК; модификации в виде 2'-O-метил, 2'-фтор, или 2'-O-метоксиэтокси, которые повышают стабильность РНК. Такие модификации также могут быть включены в другие части PEgRNA для повышения стабильности и активности. В качестве альтернативы или дополнительно, матрица PEgRNA может быть сконструирована таким образом, что она как кодирует требуемый продукт в виде белка, так и с большей вероятностью включает простые вторичные структуры, которые способны подвергаться обратному фолдингу за счет RT. Такие простые структуры будут действовать в качестве термодинамического стока, делая менее вероятным появление более сложных структур, предотвращающих обратную транскрипцию. Наконец, также является возможным разделить матрицу на две отдельные PEgRNA. При такой конструкции PE будет использоваться для инициирования транскрипции, а также рекрутирования отдельной матричной РНК к являющемуся мишенью участку с помощью РНК-связывающего белка, слитого с Cas9, или элемента распознавания РНК на самой PEgRNA, например аптамера MS2. RT может либо напрямую связываться с данной отдельной матричной РНК, или инициировать обратную транскрипцию на исходной

PEgRNA перед заменой на вторую матрицу. Такой подход может обеспечивать возможность осуществления длинных вставок за счет как предотвращения неправильного фолдинга PEgRNA при добавлении длинной матрицы, так и отсутствия необходимости в диссоциации Cas9 из генома для реализации длинных вставок, что вероятно может ингибировать длинные вставки на основе PE.

(iv) Внесение дополнительных мотивов РНК на 5′ или 3′-концах

[0888] Конструкции PEgRNA также могут быть улучшены за счет внесения дополнительных мотивов на любом конце РНК. Некоторые подобные мотивы, например PEN ENE из KSHV и ENE из MALAT1, были описаны выше в части (i)[184,185] в качестве возможных способов терминации экспрессии длинных PEgRNA из не относящихся к pol III промоторов. Эти элементы образуют тройные спирали РНК, которые окружают polyA-хвост, что приводит к их удержанию в ядре[184,187]. Однако за счет образования сложных структур на 3′-конце PEgRNA, которые закрывают концевой нуклеотид, данные структуры также вероятно будут помогать предотвращать опосредованное экзонуклеазой расщепление PEgRNA. Другие структурные элементы, вставленные на 3′-конце, также могут повышать стабильность РНК, хотя и без обеспечения терминации из не относящихся к pol III промоторов. Такие мотивы могут включать шпильки или квадруплексы РНК, которые будут закрывать 3′-конец[197], или саморасщепляющиеся рибозимы, например HDV, которые будут приводить к образованию 2'-3′-циклического фосфата на 3′-конце, а также потенциально делать PEgRNA менее подверженной расщеплению экзонуклеазами[198]. Индуцирование циклизации PEgRNA за счет неполного сплайсинга для образования циРНК также может повышать стабильность PEgRNA и приводить к удержанию PEgRNA в ядре[194].

[0889] Дополнительные мотивы РНК также могут повышать процессирование RT или повышать активность PEgRNA за счет повышения связывания RT с дуплексом ДНК-РНК. Добавление нативной последовательности, связанной RT в ее родственном ретровирусном геноме, может повышать активность RT[199]. Это может включать нативный участок связывания праймера (PBS), полипуриновый тракт (PPT), или целующиеся петли, принимающие участие в димеризациии ретровирусного генома и инициировании транскрипции[199]. Добавление мотивов димеризации, таких как целующиеся петли или

пары тетрапетля GNRA/рецептор тетрапетли[200], на 5′- и 3′-концах PEgRNA также может приводить к эффективной циркуляризации PEgRNA, что повышает стабильность. Кроме того, предполагается, что добавление данных мотивов может обеспечивать физическое разделение спейсера и праймера PEgRNA, предотвращать закрытие спейсера, ухудшающее активность PE. Короткие 5′ удлинения PEgRNA, которые образуют малую опорную шпильку в спейсерном участке, также могут благоприятным образом конкурировать против участка отжига PEgRNA, связывающего спейсер. Наконец, целующиеся петли также могут использоваться для рекрутирования других матричных РНК в участок генома и обеспечения обмена активностью RT от одной РНК к другой (раздел iii).

Последовательности

[0890] Продукт слияния PEgRNA-HDV

GGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTT
ATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAGCGTGCTCAGTCTGGGCCGG
CATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTTCGGCATGGCGAATGGGACTT
TTTTT (SEQ ID NO: 1361574)

[0891] Целующаяся петля PEgRNA-MMLV

GGTGGGAGACGTCCCACCGGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTA
AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAGC
TTCGACCGTGCTCAGTCTGGTGGGAGACGTCCCACCTTTTTTT (SEQ ID NO: 1361575)

[0892] Целующаяся петля PEgRNA-VS рибозим

GAGCAGCATGGCGTCGCTGCTCACGGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGC
AAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCAT
CAAAGCTTCGACCGTGCTCAGTCTCCATCAGTTGACACCCTGAGGTTTTTTT (SEQ ID NO:
1361576)

[0893] PEgRNA-GNRA тетрапетля/рецептор тетрапетли

GCAGACCTAAGTGGUGACATATGGTCTGGGCCCAGACTGAGCACGTGAGTTTTAGAGCTAUACG
TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTUACGAAGTGGGACCGAGTCGGTCCTCTG
CCATCAAAGCTTCGACCGTGCTCAGTCTGCATGCGATTAGAAATAATCGCATGTTTTTTT
(SEQ ID NO: 1361577)

[0894] PEgRNA матрица с заменой вторичного продукта слияния в виде РНК-HDV

TCTGCCATCAAAGCTGCGACCGTGCTCAGTCTGGTGGGAGACGTCCCACCGGCCGGCATGGTCC
CAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTTCGGCATGGCGAATGGGACTTTTTTT
(SEQ ID NO: 1361578)

(v) Эволюция PEgRNA

[0895] Каркас PEgRNA вероятно может быть дополнительно усовершенствован за счет направленной эволюции аналогично тому как были усовершенствованы SpCas9 и редакторы оснований[201]. Направленная эволюция может повышать степень распознавания PEgRNA Cas9 или подвергнутыми эволюции вариантами Cas9. Кроме того, является возможным, что различные последовательности каркаса PEgRNA будут оптимальными в рамках различных геномных локусов, обеспечивая повышение активности PE в рассматриваемом участке, снижая нецелевые виды активности, или осуществляя оба из данных действий. Наконец, эволюция каркасов PEgRNA, к которым были добавлены другие мотивы РНК, будет практически наверняка повышать активность слитой PEgRNA по сравнению с не подвергавшейся эволюции слитой РНК. Например, эволюция аллостерических рибозимов, состоящих из аптамеров c-di-GMP-I и рибозимов типа "головка молотка" (hammerhead), приводила к значительному повышению активности[202], что свидетельствует о том, что эволюция также будет повышать активность продуктов слияния PEgRNA и рибозимов типа "головка молотка" (hammerhead). Кроме того, в то время как Cas9 в настоящее время в целом не совместим с 5′ удлинением sgРНК, направленная эволюция вероятно обеспечит реализацию мутаций, которые устранят данную несовместимость, обеспечивая возможность использования дополнительных мотивов РНК.

Альтернативные подходы

[0896] Как описано здесь, ряд данных подходов уже был описан для применения с комплексами Cas9:sgРНК, однако конструкции для повышения активности PEgRNA не были описаны. Другие стратегии для внесения программируемых мутаций в геном включают редактирование оснований, направленную гомологичную репарацию (HDR), точное опосредуемое микрогомологией соединение концов (MMEJ), или опосредованное транспозазой редактирование. Однако все эти подходы имеют значительные недостатки по сравнению с PE. В то время как современные редакторы оснований являются более

эффективными, чем существующие PE, они могут обеспечивать внесение лишь определенных классов геномных мутаций и могут приводить к дополнительным нежелательным нуклеотидным конверсиям в представляющем интерес участке. HDR является реализуемой только в очень малом количестве типов клеток и приводит к сравнительно высоким показателям случайных мутаций в виде вставок и делеций (инделям). Точное MMEJ может приводить к предсказуемой репарации двуцепочечных разрывов, но в целом ограничено внесением делеций, имеет сильную зависимость от участка, и также может иметь сравнительно высокие показатели нежелательных инделей. Было показано, что опосредованное транспозазой редактирование на сегодняшний день функционирует только в бактериях. Таким образом, усовершенствования PE являются вероятно наилучшим путем для терапевтического корректирования большого количества геномных мутаций.

ПРИМЕР 4. ВСТРАИВАНИЕ 3′ ОПОРНОЙ ПЕТЛИ В УЧАСТОК СВЯЗЫВАНИЯ ПРАЙМЕРА (PBS) ПОВЫШАЕТ АКТИВНОСТЬ PEgRNA

[0897] Для дальнейшего повышения активности PE авторы предусмотрели добавление последовательности опорной петли на 3′-конце PEgRNA, имеющей 3′ плечо удлинения. На Фиг.37A приведен пример стандартной PEgRNA SpCas9, имеющей 3′ плечо удлинения (верхняя молекула). 3′ плечо удлинения, в свою очередь, включает RT матрицу (которая включает требуемое изменение) и участок связывания праймера (PBS) на 3′-конце молекулы. Молекула оканчивается последовательностью poly(U), включающей три азотистых основания U (то есть 5′-UUU-3′).

[0898] С другой стороны, в нижней части Фиг.37A показана та же молекула PEgRNA, что и в верхней части Фиг.37A, но где последовательность из 9 азотистых оснований в виде 5′-GAAANNNNN-3′ была вставлена между 3′-концом участка связывания праймера и 5′-концом концевой последовательности poly(U). Данная структура сворачивается обратно на саму себя на 180° с образованием структуры РНК в виде "опорной петли", где последовательности 5′-NNNNN-3′ вставки в виде 9 азотистых оснований отжигаются с комплементарной последовательностью в участке связывания праймера и где часть в виде 5′-GAAA-3′ образует поворот на 180°. Свойства последовательности опорной петли, показанные на Фиг.37A, не ограничивают и не сужают объем возможных опорных петель, которые могут быть использованы вместо нее. Кроме того, последовательность опорной

петли будет зависеть от комплементарной последовательности участка связывания праймера. Однако главным является то, что в рамах различных вариантов осуществления последовательность опорной петли может иметь первую часть в виде последовательности, которая образует поворот на 180°, и вторую часть в виде последовательности, которая имеет последовательность, являющуюся комплементарной части участка связывания праймера.

[0899] Без связи с теорией, предполагается, что последовательность опорной петли обеспечивает возможность использования PEgRNA со значительно более длинными участками связывания праймера, чем те, которые были бы возможны в ином случае. В свою очередь, предполагается, что более длинные последовательности PBS повышают активность PEgRNA PE. Более конкретно, вероятная функция опорной петли заключает в закрытии PBS или, по меньшей мере, минимизации взаимодействия PBS со спейсером. Образование стабильной шпильки между PBS и спейсером может приводить к деактивации PEgRNA. Без опорной петли данное взаимодействие может требовать ограничения длины PBS. Блокирование или минимизация взаимодействия между спейсером и PBS с помощью 3'-концевой опорной петли может приводить к повышению активности PE.

[0900] На Фиг.37B показаны результаты Примера 4, который демонстрирует, что эффективность праймерного редактирования в клетках HEK или клетках EMX повышается с применением PEgRNA, содержащей элементы в виде опорных петель, в то время как процентная величина образования инделей в значительной степени не меняется.

ВАРИАНТЫ ОСУЩЕСТВЛЕНИЯ

[0901] Приведенные ниже варианты осуществления находятся в рамках объема настоящего изобретения. Более того, изобретение охватывает все вариации, комбинации, и пермутации данных вариантов осуществления, в которых одно или несколько ограничений, элементов, пунктов, и описательных элементов из одного или нескольких перечисленных вариантов осуществления внесены в другой перечисленный вариант осуществления в рамках данного раздела. Например, любой перечисленный вариант осуществления, который зависит от другого варианта осуществления, может быть

модифицирован для включения одного или нескольких ограничений, указанных в любом другом перечисленном варианте осуществления в рамках данного раздела, который зависит от того же первоначального варианта осуществления. Если элементы представлены в виде перечней, например в формате группы Маркуша, каждая подгруппа данных элементов также включается, при этом любой элемент (элементы) могут быть удалены из группы. Следует понимать, что, в целом, при обозначении изобретения или аспектов изобретения в качестве включающих определенные элементы и/или свойства, определенные варианты осуществления изобретения или аспекты изобретения состоят или состоят по существу из подобных элементов и/или свойств. Следует также отметить, что термины "включающий" и "содержащий" являются открытыми и допускают включение дополнительных элементов или этапов. При указании диапазонов их границы включаются. Более того, если иное не указано или не следует из контекста и понимания среднего специалиста в данной области, значения, которые выражены в виде диапазонов, могут соответствовать любому конкретному значению или поддиапазону в рамках указанных диапазонов в различных вариантах осуществления изобретения до десятой доли единицы нижней границы диапазона, если из контекста явно не следует иное.

1. Направляющая РНК, включающая спейсер, ядро нРНК, и плечо удлинения, где направляющая РНК включает последовательность, выбранную из группы, состоящей из SEQ ID NO: 1-135514, или последовательность, имеющую, по меньшей мере, 90% идентичность последовательности относительно любой из SEQ ID NO: 1-135514.

2. Направляющая РНК, включающая спейсер, ядро нРНК, и плечо удлинения, где спейсер включает нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 135515 – 271028, или спейсер, имеющий нуклеотидную последовательность, имеющую, по меньшей мере, 90% идентичность последовательности относительно любой из SEQ ID NO: 135515 – 271028.

3. Направляющая РНК, включающая спейсер, ядро нРНК, и плечо удлинения, где плечо удлинения имеет нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 271029 – 406542, или плечо удлинения, имеющее нуклеотидную

последовательность, имеющую, по меньшей мере, 90% идентичность последовательности относительно любой из SEQ ID NO: 271029 – 406542.

4. Направляющая РНК, включающая спейсер, ядро нРНК, и плечо удлинения, где плечо удлинения включает (i) участок связывания праймера, (ii) матрицу для редактирования, и (iii) плечо гомологии.

5. Направляющая РНК, включающая спейсер, ядро нРНК, и плечо удлинения, где плечо удлинения включает участок связывания праймера, имеющий нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 406543 – 542056, или участок связывания праймера, имеющий нуклеотидную последовательность, которая, по меньшей мере, на 90% идентична последовательности любой из SEQ ID NO: 406543 – 542056.

6. Направляющая РНК, включающая спейсер, ядро нРНК, и плечо удлинения, где плечо удлинения включает матрицу для редактирования, включающую нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 542057 – 677570, или матрицу для редактирования, имеющую нуклеотидную последовательность, которая, по меньшей мере, на 90% идентична любой из SEQ ID NO: 542057 – 677570.

7. Направляющая РНК, включающая спейсер, ядро нРНК, и плечо удлинения, где плечо удлинения включает плечо гомологии, имеющее нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 677571 – 813084, или плечо гомологии, имеющее нуклеотидную последовательность, которая, по меньшей мере, на 90% идентична любой из SEQ ID NO: 677571 – 813084.

8. Направляющая РНК, включающая:

(i) спейсер, имеющий нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 135515 – 271028, или спейсер, имеющий нуклеотидную последовательность, имеющую, по меньшей мере, 90% идентичность последовательности относительно любой из SEQ ID NO: 135515 – 271028, и

(ii) плечо удлинения, выбранное из группы, состоящей из SEQ ID NO: 271029 – 406542, или плечо удлинения, имеющее нуклеотидную последовательность, имеющую, по меньшей мере, 90% идентичность последовательности относительно SEQ ID NO: 271029 – 406542.

9. Направляющая РНК, включающая:

(i) спейсер, имеющий нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 135515 – 271028, или спейсер, имеющий нуклеотидную последовательность, которая, по меньшей мере, на 90% идентична любой из SEQ ID NO: 135515 – 271028, и

(ii) участок связывания праймера, выбранный из группы, состоящей из SEQ ID NO: 406543 – 542056, или участок связывания праймера, имеющий нуклеотидную последовательность, которая, по меньшей мере, на 90% идентична любой из SEQ ID NO: 406543 – 542056.

10. Направляющая РНК, включающая:

(i) спейсер, имеющий нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 135515 – 271028, или спейсер, имеющий нуклеотидную последовательность, имеющую, по меньшей мере, 90% идентичность последовательности относительно любой из SEQ ID NO: 135515 – 271028, и

(ii) матрицу для редактирования, имеющую нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 542057 – 677570, или матрицу для редактирования, имеющую нуклеотидную последовательность, которая, по меньшей мере, на 90% идентична любой из SEQ ID NO: 542057 – 677570.

11. Направляющая РНК, включающая:

(i) спейсер, имеющий нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 135515 – 271028, или спейсер, имеющий нуклеотидную последовательность, которая, по меньшей мере, на 90% идентична любой из SEQ ID NO: 135515 – 271028, и

(ii) плечо гомологии, имеющее нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 677571 – 813084, или спейсер, имеющий нуклеотидную последовательность, которая, по меньшей мере, на 90% идентична любой из SEQ ID NO: 677571 – 813084.

12. Направляющая РНК по любому из вариантов осуществления 1-11, дополнительно включающая сигнал терминации SEQ ID NO: 813086 или сигнал терминации, имеющий, по меньшей мере, 90% идентичность последовательности относительно SEQ ID NO: 813086.

13. Направляющая РНК по любому из вариантов осуществления 1-12, дополнительно включающая 5'-концевой модификаторный участок, включающий последовательность шпильки, последовательность стебля/петли, или последовательность опорной петли.

14. Направляющая РНК по любому из вариантов осуществления 1-13, дополнительно включающая 3'-концевой модификаторный участок, включающий последовательность шпильки, последовательность стебля/петли, или последовательность опорной петли.

15. Направляющая РНК по любому из вариантов осуществления 1-14, дополнительно включающая ядро нРНК, включающее SEQ ID NO: 813085, или ядро нРНК, имеющее, по меньшей мере, 90% идентичность последовательности относительно SEQ ID NO: 813085.

16. Направляющая РНК по любому из вариантов осуществления 1-15, где направляющая РНК способна связываться с napDNAbp, подходящим для праймерного редактирования, и направлять napDNAbp к последовательности-мишени ДНК.

17. Направляющая РНК по варианту осуществления 16, где последовательность-мишень нуклеиновой кислоты включает цепь-мишень (или включающую PAM цепь) и комплементарную не являющуюся мишенью цепь (или не включающую PAM цепь), где спейсер направляющей РНК гибридизируется с комплементарной не являющейся мишенью цепью (не включающей PAM цепью) для образования гибрида РНК-ДНК и R-петли.

18. Направляющая РНК по любому из вариантов осуществления 1-17, где участок связывания праймера имеет длину от приблизительно 8 до приблизительно 20 нуклеотидов.

19. Направляющая РНК по любому из вариантов осуществления 1-18, где участок связывания праймера имеет длину 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, или 20 нуклеотидов.

20. Направляющая РНК по любому из указанных выше вариантов осуществления, где участок связывания праймера имеет длину, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, или, по меньшей мере, 20 нуклеотидов.

21. Направляющая РНК по любому из вариантов осуществления 1-19, где плечо гомологии комплементарно цепи ДНК-мишени.

22. Направляющая РНК по любому из вариантов осуществления 1-10, где плечо удлинения имеет длину от приблизительно 7 до приблизительно 500 нуклеотидов.

23. Направляющая РНК по любому из указанных выше вариантов осуществления, где плечо удлинения имеет длину, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, по меньшей мере, 20 нуклеотидов, по меньшей мере, 21 нуклеотид, по меньшей мере, 22 нуклеотида, по меньшей мере, 23 нуклеотида, по меньшей мере, 24 нуклеотида, по меньшей мере, 25 нуклеотидов, по меньшей мере, 26 нуклеотидов, по меньшей мере, 27 нуклеотидов, по меньшей мере, 28 нуклеотидов, по меньшей мере, 29 нуклеотидов, по меньшей мере, 30 нуклеотидов, по

меньшей мере, 31 нуклеотид, по меньшей мере, 32 нуклеотида, по меньшей мере, 33 нуклеотида, по меньшей мере, 34 нуклеотида, по меньшей мере, 35 нуклеотидов, по меньшей мере, 36 нуклеотидов, по меньшей мере, 37 нуклеотидов, по меньшей мере, 38 нуклеотидов, по меньшей мере, 39 нуклеотидов, по меньшей мере, 40 нуклеотидов, или, по меньшей мере, 100 нуклеотидов.

24. Направляющая РНК по любому из указанных выше вариантов осуществления, где плечо удлинения имеет длину, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, по меньшей мере, 20 нуклеотидов, по меньшей мере, 21 нуклеотид, по меньшей мере, 22 нуклеотида, по меньшей мере, 23 нуклеотида, по меньшей мере, 24 нуклеотида, по меньшей мере, 25 нуклеотидов, по меньшей мере, 26 нуклеотидов, по меньшей мере, 27 нуклеотидов, по меньшей мере, 28 нуклеотидов, по меньшей мере, 29 нуклеотидов, по меньшей мере, 30 нуклеотидов, по меньшей мере, 31 нуклеотид, по меньшей мере, 32 нуклеотида, по меньшей мере, 33 нуклеотида, по меньшей мере, 34 нуклеотида, по меньшей мере, 35 нуклеотидов, по меньшей мере, 36 нуклеотидов, по меньшей мере, 37 нуклеотидов, по меньшей мере, 38 нуклеотидов, по меньшей мере, 39 нуклеотидов, по меньшей мере, 40 нуклеотидов, или, по меньшей мере, 100 нуклеотидов.

25. Направляющая РНК по любому из указанных выше вариантов осуществления, где плечо гомологии имеет длину, по меньшей мере, 1 нуклеотид, по меньшей мере, 2 нуклеотида, по меньшей мере, 3 нуклеотида, по меньшей мере, 4 нуклеотида, по меньшей мере, 5 нуклеотидов, по меньшей мере, 6 нуклеотидов, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, по меньшей мере, 20

нуклеотидов, по меньшей мере, 21 нуклеотид, по меньшей мере, 22 нуклеотида, по меньшей мере, 23 нуклеотида, по меньшей мере, 24 нуклеотида, по меньшей мере, 25 нуклеотидов, по меньшей мере, 26 нуклеотидов, по меньшей мере, 27 нуклеотидов, по меньшей мере, 28 нуклеотидов, по меньшей мере, 29 нуклеотидов, или, по меньшей мере, 30 нуклеотидов.

26. Направляющая РНК по любому из указанных выше вариантов осуществления, где матрица для редактирования и плечо гомологии могут использоваться обратной транскриптазой в качестве последовательности матрицы для синтеза соответствующего одноцепочечного ДНК-флэпа, имеющего 3'-конец, где ДНК-флэп комплементарен цепи последовательности эндогенной ДНК-мишени, примыкающей к участку разрыва, и где одноцепочечный ДНК-флэп включает нуклеотидное изменение, кодируемое матрицей для редактирования.

27. Направляющая РНК по варианту осуществления 25, где одноцепочечный ДНК-флэп вытесняет эндогенную одноцепочечную ДНК, имеющую 5′-конец, в последовательности-мишени ДНК, которая была никирована.

28. Направляющая РНК по варианту осуществления 26, где эндогенная одноцепочечная ДНК, имеющая свободный 5′-конец, подвергается эксцизии клеткой.

29. Направляющая РНК по варианту осуществления 27, где клеточная репарация одноцепочечного ДНК-флэпа приводит к внесению нуклеотидного изменения, тем самым обеспечивая образование требуемого продукта.

30. Направляющая РНК по варианту осуществления 28, где требуемое нуклеотидное изменение представляет собой вставку.

31. Направляющая РНК по варианту осуществления 29, где плечо вставка имеет длину, по меньшей мере, 1 нуклеотид, по меньшей мере, 2 нуклеотида, по меньшей мере, 3 нуклеотида, по меньшей мере, 4 нуклеотида, по меньшей мере, 5 нуклеотидов, по меньшей мере, 6 нуклеотидов, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13

нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, по меньшей мере, 20 нуклеотидов, по меньшей мере, 21 нуклеотид, по меньшей мере, 22 нуклеотида, по меньшей мере, 23 нуклеотида, по меньшей мере, 24 нуклеотида, по меньшей мере, 25 нуклеотидов, по меньшей мере, 26 нуклеотидов, по меньшей мере, 27 нуклеотидов, по меньшей мере, 28 нуклеотидов, по меньшей мере, 29 нуклеотидов, по меньшей мере, 30 нуклеотидов, по меньшей мере, 31 нуклеотид, по меньшей мере, 32 нуклеотида, по меньшей мере, 33 нуклеотида, по меньшей мере, 34 нуклеотида, по меньшей мере, 35 нуклеотидов, по меньшей мере, 36 нуклеотидов, по меньшей мере, 37 нуклеотидов, по меньшей мере, 38 нуклеотидов, по меньшей мере, 39 нуклеотидов, по меньшей мере, 40 нуклеотидов, или, по меньшей мере, 100 нуклеотидов.

32. Направляющая РНК по варианту осуществления 29, где вставка представляет собой последовательность, кодирующую полипептид.

33. Комплекс для праймерного редактирования, включающий napDNAbp, обратную транскриптазу, и любую одну из направляющих РНК по вариантам осуществления 1-32.

34. Комплекс для праймерного редактирования по варианту осуществления 32, где napDNAbp и обратная транскриптаза образуют слитый белок.

35. Комплекс для праймерного редактирования по варианту осуществления 32, где napDNAbp представляет собой Cas9.

36. Комплекс для праймерного редактирования по варианту осуществления 34, где Cas9 выбран из группы, состоящей из никаз Cas9 или их вариантов.

37. Комплекс для праймерного редактирования по варианту осуществления 34, где Cas9 имеет аминокислотную последовательность, выбранную из группы, состоящей из SEQ ID NO: 1-135514.

38. Комплекс для праймерного редактирования по варианту осуществления 33, где слитый белок имеет аминокислотную последовательность, выбранную из группы, состоящей из SEQ ID NO: 1-135514.

39. Комплекс для праймерного редактирования по варианту осуществления 33, где слитый белок включает линкер, соединяющий napDNAbp и обратную транскриптазу.

40. Комплекс для праймерного редактирования по варианту осуществления 38, где линкер имеет аминокислотную последовательность, выбранную из группы, состоящей из SEQ ID NO: 1-135514.

41. Один или несколько полинуклеотидов, кодирующих комплекс для праймерного редактирования по любому из вариантов осуществления 32-39.

42. Вектор, включающий полинуклеотид по варианту осуществления 41 и один или несколько промоторов, которые обеспечивают экспрессию направляющей РНК и слитого белка комплекса для праймерного редактирования.

43. Клетка, включающая вектор по варианту осуществления 41.

44. Клетка, включающая комплекс для праймерного редактирования по любому из вариантов осуществления 32-39.

45. Фармацевтическая композиция, включающая: (i) направляющую РНК по любому из вариантов осуществления 1-31, комплекс для праймерного редактирования по вариантам осуществления 32-39, полинуклеотид по варианту осуществления 40, или вектор по варианту осуществления 41; и (ii) фармацевтически приемлемый наполнитель.

46. Способ внесения нуклеотидного изменения в последовательность нуклеиновой кислоты, при этом способ включает: контактирование последовательности нуклеиновой кислоты с комплексом, включающим слитый белок и направляющую РНК по любому из вариантов осуществления 1-31 или любому из вариантов осуществления 83-85, где слитый белок включает napDNAbp и полимеразу, и где направляющая РНК включает спейсер, ядро нРНК, и плечо удлинения, которое включает матрицу для редактирования, кодирующую нуклеотидное изменение, тем самым обеспечивая

(i) никирование последовательности двуцепочечной ДНК в цепи-мишени (или включающей PAM цепи) и создание свободной одноцепочечной ДНК, имеющей 3´-конец;

(ii) гибридизирование 3'-конца свободной одноцепочечной ДНК с направляющей РНК в участке связывания праймера, тем самым обеспечивая праймирование полимеразы;

(iii) полимеризацию цепи ДНК с 3´-конца, тем самым обеспечивая создание одноцепочечного ДНК-флэпа, включающего нуклеотидное изменение; и

(iv) замещение цепи эндогенной ДНК, непосредственно снизу примыкающей к участку разрыва на цепи-мишени (или включающей PAM цепи), одноцепочечным ДНК-флэпом, тем самым обеспечивая внесение требуемого нуклеотидного изменения в последовательность двуцепочечной ДНК.

47. Способ по варианту осуществления 46, где нуклеотидное изменение представляет собой замещение одного нуклеотида, делецию, вставку, или их комбинацию.

48. Способ по варианту осуществления 46, где замещение одного нуклеотида представляет собой транзицию или трансверсию.

49. Способ по варианту осуществления 46, где нуклеотидное изменение представляет собой (1) замещение G на T, (2) замещение G на A, (3) замещение G на C, (4) замещение T на G, (5) замещение T на A, (6) замещение T на C, (7) замещение C на G, (8) замещение C на T, (9) замещение C на A, (10) замещение A на T, (11) замещение A на G, или (12) замещение A на C.

50. Способ по варианту осуществления 46, где нуклеотидное изменение обеспечивает конвертирование (1) пары оснований G:C на пару оснований T:A, (2) пары оснований G:C на пару оснований A:T, (3) пары оснований G:C на пару оснований C:G, (4) пары оснований T:A на пару оснований G:C, (5) пары оснований T:A на пару оснований A:T, (6) пары оснований T:A на пару оснований C:G, (7) пары оснований C:G на пару оснований G:C, (8) пары оснований C:G на пару оснований T:A, (9) пары оснований C:G на пару

оснований A:T, (10) пары оснований A:T на пару оснований T:A, (11) пары оснований A:T на пару оснований G:C, или (12) пары оснований A:T на пару оснований C:G.

51. Способ по варианту осуществления 46, где нуклеотидное изменение представляет собой вставку или делецию 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, или 25 нуклеотидов.

52. Способ по варианту осуществления 46, где нуклеотидное изменение представляет собой вставку кодирующей полипептид последовательности.

53. Способ по варианту осуществления 46, где нуклеотидное изменение обеспечивает корректирование связанного с заболеванием гена.

54. Способ по варианту осуществления 46, где связанный с заболеванием ген связан с моногенетическим заболеванием, выбранным из группы, состоящей из: дефицита аденозиндезаминазы (ADA); дефицита альфа-1 антитрипсина; муковисцидоза; мышечной дистрофии Дюшенна; галактоземии; гемохроматоза; болезни Хантингтона; болезни кленового сиропа; синдрома Марфана; нейрофиброматоза 1 типа; врожденной пахионихии; фенилкеотнурии; тяжелого комбинированного иммунодефицита; серповидноклеточной анемии; синдрома Смита-Лемли-Опица; и болезни Тея-Сакса.

55. Способ по варианту осуществления 46, где связанный с заболеванием ген связан с полигенным заболеванием, выбранным из группы, состоящей из: болезни сердца, высокого кровяного давления, болезни Альцгеймера, артрита, диабета, рака, и ожирения.

56. Компьютеризированный способ определения структуры направляющей РНК праймерного редактора (PEgRNA), при этом способ включает использование, по меньшей мере, одного компьютерного аппаратного процессора для осуществления:

анализа данных, связанных с:

входным аллелем;

выходным аллелем; и

слитым белком, включающим программируемый нуклеиновой кислотой ДНК-связывающий белок и полимеразу (например обратную транскриптазу); и

определение структуры PEgRNA на основе входного аллеля, выходного аллеля, и слитого белка, где структура PEgRNA конструируется для ассоциирования со слитым белком для изменения входного аллеля с получением выходного аллеля, включая определение для структуры PEgRNA одного или нескольких из следующих признаков:

спейсера, комплементарного нуклеотидной последовательности-мишени во входном аллеле;

каркаса нРНК для взаимодействия со слитым белком; и

удлинения, включающего одну или несколько из:

последовательности матрицы для синтеза ДНК, включающей требуемое нуклеотидное изменение для изменения входного аллеля с получением выходного аллеля;

участка связывания праймера;

необязательно сигнала терминации, примыкающего к матрице для синтеза ДНК;

необязательно первого модификатора, примыкающего к сигналу терминации; и

необязательно второго модификатора, примыкающего к участку связывания праймера.

57. Способ по варианту осуществления 56, дополнительно включающий определение спейсера и удлинения, а также определение того, что спейсер расположен на 5'-конце структуры PEgRNA, а удлинение расположено на 3'-конце структуры PEgRNA.

58. Способ по варианту осуществления 56, дополнительно включающий определение спейсера и удлинения, где спейсер расположен на 5'-конце структуры PEgRNA, а удлинение расположено 3' относительно спейсера.

59. Способ по варианту осуществления 56, где доступ к данным, связанным с входным аллелем и выходным аллелем, включает доступ к базе данных, включающей набор входных аллелей и ассоциированных выходных аллелей.

60. Способ по варианту осуществления 59, где доступ к базе данных включает доступ к базе данных ClinVar, включающей множество записей, где каждая запись включает

входной аллель из набора входных аллелей и выходной аллель из набора выходных аллелей.

61. Способ по варианту осуществления 59, где определение структуры PEgRNA включает определение одной или нескольких структур PEgRNA для каждого входного аллеля и ассоциированного выходного аллеля в наборе.

62. Способ по варианту осуществления 56, где доступ к данным, связанным со слитым белком, включает определение слитого белка из множества слитых белков.

63. Способ по варианту осуществления 56, где слитый белок включает белок Cas9.

64. Способ по варианту осуществления 63, где слитый белок включает белок Cas9-NG или белок SpCas9.

65. Способ по варианту осуществления 56, где изменение входного аллеля на выходной аллель включает изменение одного нуклеотида, вставку одного или нескольких нуклеотидов, делецию одного или нескольких нуклеотидов, или их комбинацию.

66. Способ по варианту осуществления 56, дополнительно включающий определение спейсера, при этом спейсер включает нуклеотидную последовательности длиной приблизительно 20 нуклеотидов.

67. Способ по варианту осуществления 66, дополнительно включающий определение спейсера на основе положения изменения в соответствующей нуклеотидной последовательности протоспейсера.

68. Способ по варианту осуществления 67, где изменение внесено в окне редактирования, которое располагается приблизительно от положения -3 протоспейсера до положения +27 протоспейсера.

69. Способ по варианту осуществления 67, дополнительно включающий:

определение набора исходных протоспейсеров-кандидатов на основе входного аллеля и слитого белка, где каждый исходных протоспейсер-кандидат включает PAM слитого белка во входном аллеле;

определение одного или нескольких исходных протоспейсеров-кандидатов из набора исходных протоспейсеров-кандидатов, где каждый из них включает несовместимое положение разрыва;

удаление определенного одного или нескольких исходных протоспейсеров-кандидатов из набора для получения набора оставшихся протоспейсеров-кандидатов; и

где определение структуры PEgRNA включает определение множества структур PEgRNA, где каждая структура PEgRNA включает различный спейсер, определенный на основе соответствующего протоспейсера из набора оставшихся протоспейсеров-кандидатов.

70. Способ по варианту осуществления 55, дополнительно включающий определение удлинения и матрицы для синтеза ДНК (например последовательности RT матрицы), где матрица для синтеза ДНК (например последовательность RT матрицы) включает приблизительно от 7 нуклеотидов до приблизительно 34 нуклеотидов.

71. Способ по варианту осуществления 56, где определение PEgRNA включает:

определение спейсера на основе входного аллеля и/или слитого белка; и

определение матрицы для синтеза ДНК (например последовательности RT матрицы) на основе спейсера.

72. Способ по варианту осуществления 56, где матрица для синтеза ДНК (например последовательность RT матрицы) кодирует одноцепочечный ДНК-флэп, который комплементарен последовательности эндогенной ДНК, являющейся смежной относительно участка разрыва, при этом одноцепочечный ДНК-флэп включает требуемое нуклеотидное изменение.

73. Способ по варианту осуществления 72, где одноцепочечный ДНК-флэп способен гибридизироваться с последовательностью эндогенной ДНК, являющейся смежной относительно участка разрыва, тем самым обеспечивая внесение требуемого нуклеотидного изменения.

74. Способ по варианту осуществления 72, где одноцепочечный ДНК-флэп способ вытеснять последовательность эндогенной ДНК, являющуюся смежной относительно участка разрыва.

75. Способ по варианту осуществления 72, где клеточная репарация одноцепочечного ДНК-флэпа приводит к внесению нуклеотидного изменения, тем самым обеспечивая образование требуемого продукта.

76. Способ по варианту осуществления 56, где слитый белок при образовании комплекса с PEgRNA способен связываться с последовательностью-мишенью ДНК.

77. Способ по варианту осуществления 76, где последовательность-мишень ДНК включает цепь-мишень, в которой осуществляется изменение, и комплементарную не являющуюся мишенью цепь.

78. Способ по варианту осуществления 56, где входной аллель включает патогенную мутацию ДНК, а выходной аллель включает скорректированную последовательность ДНК.

79. Способ по варианту осуществления 56, где входной аллель представляет собой любой один из связанных с заболеванием аллелей SEQ ID NO: 1217353-1289387.

80. Способ по варианту осуществления 56, где выходной аллель представляет собой любой один из здоровых аллелей SEQ ID NO: 1289388-1361420.

81. Система, включающая:

по меньшей мере, один процессор; и

по меньшей мере, один машиночитаемый носитель данных, на котором закодированы инструкции, которые при выполнении побуждают, по меньшей мере, один процессор выполнять способ по любому из вариантов осуществления 56-81.

82. По меньшей мере, один машиночитаемый носитель данных, на котором закодированы инструкции, которые при выполнении побуждают, по меньшей мере, один процессор выполнять способ по любому из вариантов осуществления 56-81.

83. Способ редактирования оснований с использованием структуры PEgRNA, определенной в соответствии со способом по любому из вариантов осуществления 56-81.

84. PEgRNA, определенная в соответствии со способом по любому одному из вариантов осуществления 56-81.

85. Направляющая РНК для применения в рамках праймерного редактирования для корректирования связанного с заболеванием аллеля в последовательности-мишени ДНК для получения здорового аллеля, при этом указанная направляющая включает спейсер, ядро нРНК, и плечо удлинения, где спейсер способен связываться с участком длиной ~20 нуклеотидов в SEQ ID NO: 1217353-1289387 или его комплементарной цепью.

86. Направляющая РНК, включающая спейсер, ядро нРНК, и плечо удлинения, где плечо удлинения включает матрицу для синтеза ДНК и участок связывания праймера, являющиеся эффективными при осуществлении праймерного редактирования.

87. Направляющая РНК по варианту осуществления 85, где участок изменения в любой из нуклеотидных последовательностей SEQ ID NO: 1217353-1289387 начинается в положении 201 в направлении от 5' к 3'.

88. Направляющая РНК для праймерного редактирования, включающая спейсер, ядро нРНК, и плечо удлинения, где плечо удлинения включает участок связывания праймера и матрицу для синтеза ДНК.

89. Направляющая РНК по варианту осуществления 88, где участок связывания праймера имеет нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 406543 – 542056 (участок связывания праймера), или нуклеотидную последовательность, которая имеет, по меньшей мере, 90% идентичность последовательности относительно любой из SEQ ID NO: 406543 – 542056.

90. Направляющая РНК по варианту осуществления 88, где матрица для синтеза ДНК включает нуклеотидную последовательность SEQ ID NO: 542057 – 677570 (матрица для редактирования) или нуклеотидную последовательность, которая имеет, по меньшей мере,

90% идентичность последовательности относительно любой из SEQ ID NO: 542057 – 677570.

91. Направляющая РНК по варианту осуществления 88, где матрица для синтеза ДНК включает нуклеотидную последовательность SEQ ID NO: 677571 – 813084 (плечо гомологии) или нуклеотидную последовательность, которая имеет, по меньшей мере, 90% идентичность последовательности относительно любой из SEQ ID NO: 677571 – 813084.

92. Направляющая РНК по варианту осуществления 88, где матрица для синтеза ДНК включает матрицу для редактирования и плечо гомологии, где матрица для редактирования включает нуклеотидную последовательность SEQ ID NO: 542057 – 677570, а плечо гомологии включает нуклеотидную последовательность SEQ ID NO: 677571 – 813084.

93. Направляющая РНК по любому из вариантов осуществления 86-92, дополнительно включающая сигнал терминации SEQ ID NO: 813086 или сигнал терминации, имеющий, по меньшей мере, 90% идентичность последовательности относительно SEQ ID NO: 813086.

94. Направляющая РНК по любому из вариантов осуществления 86-93, дополнительно включающая 5'-концевой модификаторный участок, включающий последовательность шпильки, последовательность стебля/петли, или последовательность опорной петли.

95. Направляющая РНК по любому из вариантов осуществления 86-94, дополнительно включающая 3'-концевой модификаторный участок, включающий последовательность шпильки, последовательность стебля/петли, или последовательность опорной петли.

96. Направляющая РНК по любому из вариантов осуществления 86-95, дополнительно включающая ядро нРНК, включающее SEQ ID NO: 813085, или ядро нРНК, имеющее, по меньшей мере, 90% идентичность последовательности относительно SEQ ID NO: 813085.

97. Направляющая РНК по любому из вариантов осуществления 86-96, где направляющая РНК способна связываться с napDNAbp, подходящим для праймерного редактирования, и направлять napDNAbp к последовательности-мишени ДНК.

98. Направляющая РНК по варианту осуществления 97, где последовательность-мишень нуклеиновой кислоты включает цепь-мишень (или включающую PAM или подвергаемую редактированию цепь) и комплементарную не являющуюся мишенью цепь (или не включающую PAM или не подвергаемую редактированию цепь), где спейсер направляющей РНК гибридизируется с не включающей PAM цепью для образования гибрида РНК-ДНК и R-петли.

99. Направляющая РНК по любому из вариантов осуществления 86-98, где участок связывания праймера имеет длину от приблизительно 8 до приблизительно 20 нуклеотидов.

100. Направляющая РНК по любому из вариантов осуществления 86-99, где участок связывания праймера имеет длину 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, или 20 нуклеотидов.

101. Направляющая РНК по любому из вариантов осуществления 86-100, где плечо удлинения имеет длину, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, по меньшей мере, 20 нуклеотидов, по меньшей мере, 21 нуклеотид, по меньшей мере, 22 нуклеотида, по меньшей мере, 23 нуклеотида, по меньшей мере, 24 нуклеотида, по меньшей мере, 25 нуклеотидов, по меньшей мере, 26 нуклеотидов, по меньшей мере, 27 нуклеотидов, по меньшей мере, 28 нуклеотидов, по меньшей мере, 29 нуклеотидов, по меньшей мере, 30 нуклеотидов, по меньшей мере, 31 нуклеотид, по меньшей мере, 32 нуклеотида, по меньшей мере, 33 нуклеотида, по меньшей мере, 34 нуклеотида, по меньшей мере, 35 нуклеотидов, по меньшей мере, 36 нуклеотидов, по меньшей мере, 37 нуклеотидов, по меньшей мере, 38 нуклеотидов, по меньшей мере, 39 нуклеотидов, по меньшей мере, 40 нуклеотидов, или, по меньшей мере, 100 нуклеотидов.

102. Направляющая РНК по любому из вариантов осуществления 86-101, где участок связывания праймера имеет длину, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, или, по меньшей мере, 20 нуклеотидов.

103. Направляющая РНК по любому из вариантов осуществления 86-102, где матрица для синтеза ДНК имеет длину, по меньшей мере, 1 нуклеотид, по меньшей мере, 2 нуклеотида, по меньшей мере, 3 нуклеотида, по меньшей мере, 4 нуклеотида, по меньшей мере, 5 нуклеотидов, по меньшей мере, 6 нуклеотидов, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, по меньшей мере, 20 нуклеотидов, по меньшей мере, 21 нуклеотид, по меньшей мере, 22 нуклеотида, по меньшей мере, 23 нуклеотида, по меньшей мере, 24 нуклеотида, по меньшей мере, 25 нуклеотидов, по меньшей мере, 26 нуклеотидов, по меньшей мере, 27 нуклеотидов, по меньшей мере, 28 нуклеотидов, по меньшей мере, 29 нуклеотидов, по меньшей мере, 30 нуклеотидов, по меньшей мере, 31 нуклеотид, по меньшей мере, 32 нуклеотида, по меньшей мере, 33 нуклеотида, по меньшей мере, 34 нуклеотида, по меньшей мере, 35 нуклеотидов, по меньшей мере, 36 нуклеотидов, по меньшей мере, 37 нуклеотидов, по меньшей мере, 38 нуклеотидов, по меньшей мере, 39 нуклеотидов, по меньшей мере, 40 нуклеотидов, или, по меньшей мере, 100 нуклеотидов.

104. Направляющая РНК по любому из вариантов осуществления 86-103, где матрица для синтеза ДНК может использоваться РНК-зависимой ДНК-полимеразой (например обратной транскриптазой) в качестве матрицы для синтеза соответствующего одноцепочечного ДНК-флэпа, имеющего 3'-конец, где ДНК-флэп комплементарен цепи последовательности эндогенной ДНК-мишени, примыкающей к участку разрыва, и где

одноцепочечный ДНК-флэп включает требуемое нуклеотидное изменение, кодируемое матрицей для синтеза ДНК.

105. Направляющая РНК по варианту осуществления 104, где одноцепочечный ДНК-флэп вытесняет эндогенную одноцепочечную ДНК, имеющую 5′-конец, в последовательности-мишени ДНК, которая была никирована.

106. Направляющая РНК по варианту осуществления 105, где эндогенная одноцепочечная ДНК, имеющая свободный 5′-конец, подвергается эксцизии клеткой.

107. Направляющая РНК по варианту осуществления 105, где клеточная репарация одноцепочечного ДНК-флэпа приводит к внесению нуклеотидного изменения, тем самым обеспечивая образование подвергнутого редактированию ДНК-продукта.

108. Направляющая РНК по варианту осуществления 107, где нуклеотидное изменение представляет собой вставку.

109. Направляющая РНК по варианту осуществления 108, где вставка имеет длину, по меньшей мере, 1 нуклеотид, по меньшей мере, 2 нуклеотида, по меньшей мере, 3 нуклеотида, по меньшей мере, 4 нуклеотида, по меньшей мере, 5 нуклеотидов, по меньшей мере, 6 нуклеотидов, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, по меньшей мере, 20 нуклеотидов, по меньшей мере, 21 нуклеотид, по меньшей мере, 22 нуклеотида, по меньшей мере, 23 нуклеотида, по меньшей мере, 24 нуклеотида, по меньшей мере, 25 нуклеотидов, по меньшей мере, 26 нуклеотидов, по меньшей мере, 27 нуклеотидов, по меньшей мере, 28 нуклеотидов, по меньшей мере, 29 нуклеотидов, по меньшей мере, 30 нуклеотидов, по меньшей мере, 31 нуклеотид, по меньшей мере, 32 нуклеотида, по меньшей мере, 33 нуклеотида, по меньшей мере, 34 нуклеотида, по меньшей мере, 35 нуклеотидов, по меньшей мере, 36 нуклеотидов, по меньшей мере, 37 нуклеотидов, по меньшей мере, 38

нуклеотидов, по меньшей мере, 39 нуклеотидов, по меньшей мере, 40 нуклеотидов, или, по меньшей мере, 100 нуклеотидов.

110. Направляющая РНК по варианту осуществления 108, где вставка представляет собой последовательность, кодирующую полипептид.

111. Комплекс для праймерного редактирования, включающий napDNAbp, РНК-зависимую ДНК-полимеразу, и любую одну из направляющих РНК по вариантам осуществления 86-110.

112. Комплекс для праймерного редактирования по варианту осуществления 111, где napDNAbp и РНК-зависимая ДНК-полимераза образуют слитый белок.

113. Комплекс для праймерного редактирования по варианту осуществления 111, где napDNAbp представляет собой Cas9.

114. Комплекс для праймерного редактирования по варианту осуществления 113, где Cas9 представляет собой никазу Cas9 или ее вариант.

115. Комплекс для праймерного редактирования по варианту осуществления 113, где Cas9 имеет аминокислотную последовательность, выбранную из группы, состоящей из SEQ ID NO: 1-135514.

116. Комплекс для праймерного редактирования по варианту осуществления 112, где слитый белок имеет аминокислотную последовательность, выбранную из группы, состоящей из SEQ ID NO: 1-135514.

117. Комплекс для праймерного редактирования по варианту осуществления 112, где слитый белок включает линкер, соединяющий napDNAbp и РНК-зависимую ДНК-полимеразу.

118. Комплекс для праймерного редактирования по варианту осуществления 117, где линкер имеет аминокислотную последовательность, выбранную из группы, состоящей из SEQ ID NO: 1-135514.

119. Один или несколько полинуклеотидов, кодирующих комплекс для праймерного редактирования по любому из вариантов осуществления 111-118.

120. Вектор, включающий полинуклеотид по варианту осуществления 119 и один или несколько промоторов, которые обеспечивают экспрессию направляющей РНК и слитого белка комплекса для праймерного редактирования.

121. Клетка, включающая вектор по варианту осуществления 120.

122. Клетка, включающая комплекс для праймерного редактирования по любому из вариантов осуществления 111-118.

123. Фармацевтическая композиция, включающая: (i) направляющую РНК по любому из вариантов осуществления 84-110, комплекс для праймерного редактирования по вариантам осуществления 111-118, полинуклеотид по варианту осуществления 119, или вектор по варианту осуществления 120; и (ii) фармацевтически приемлемый наполнитель.

124. Способ внесения нуклеотидного изменения в последовательность нуклеиновой кислоты, при этом способ включает: контактирование последовательности нуклеиновой кислоты с комплексом, включающим слитый белок и направляющую РНК по любому из вариантов осуществления 109-116, где слитый белок включает napDNAbp и РНК-зависимую ДНК-полимеразу, где направляющая РНК включает спейсер, ядро нРНК, и плечо удлинения, которое включает матрицу для синтеза ДНК и участок связывания праймера, при этом указанная матрица для синтеза ДНК кодирует нуклеотидное изменение, и где спейсер способен отжигаться на не включающей PAM цепи, являющейся проксимальной относительно доступного PAM и протоспейсера, тем самым обеспечивая

(i) никирование последовательности двуцепочечной ДНК на включающей PAM цепи, тем самым обеспечивая создание свободной одноцепочечной ДНК, имеющей 3´-конец;

(ii) гибридизирование 3'-конца свободной одноцепочечной ДНК с направляющей РНК в участке связывания праймера, тем самым обеспечивая праймирование РНК-зависимой ДНК-полимеразы;

(iii) полимеризацию цепи ДНК от 3'-конца ДНК с кодированием из матрицы для синтеза ДНК, тем самым обеспечивая получения одноцепочечного ДНК-флэпа, удлиненного от 3'-конца ДНК, где флэп включает нуклеотидное изменение;

(iv) замещение цепи эндогенной ДНК, непосредственно снизу примыкающей к участку разрыва на включающей PAM цепи, одноцепочечным ДНК-флэпом, тем самым обеспечивая внесение нуклеотидного изменения в последовательность двуцепочечной ДНК.

125. Способ по варианту осуществления 124, где при завершении этапа (v) в клетке клетка осуществляет репарацию не подвергаемой редактированию цепи за счет репарации и/или репликации клеточной ДНК.

126. Способ по варианту осуществления 124, где нуклеотидное изменение представляет собой замещение одного нуклеотида, делецию, вставку, или их комбинацию

127. Способ по варианту осуществления 124, где замещение одного нуклеотида представляет собой транзицию или трансверсию.

128. Способ по варианту осуществления 124, где замещение одного нуклеотида представляет собой (1) замещение G на T, (2) замещение G на A, (3) замещение G на C, (4) замещение T на G, (5) замещение T на A, (6) замещение T на C, (7) замещение C на G, (8) замещение C на T, (9) замещение C на A, (10) замещение A на T, (11) замещение A на G, или (12) замещение A на C.

129. Способ по варианту осуществления 124, где замещение одного нуклеотида обеспечивает конвертирование (1) пары оснований G:C на пару оснований T:A, (2) пары оснований G:C на пару оснований A:T, (3) пары оснований G:C на пару оснований C:G, (4) пары оснований T:A на пару оснований G:C, (5) пары оснований T:A на пару оснований A:T, (6) пары оснований T:A на пару оснований C:G, (7) пары оснований C:G на пару оснований G:C, (8) пары оснований C:G на пару оснований T:A, (9) пары оснований C:G на пару оснований A:T, (10) пары оснований A:T на пару оснований T:A,

(11) пары оснований A:T на пару оснований G:C, или (12) пары оснований A:T на пару оснований C:G.

130. Способ по варианту осуществления 124, где нуклеотидное изменение представляет собой вставку или делецию 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, или 25 нуклеотидов.

131. Способ по варианту осуществления 124, где нуклеотидное изменение представляет собой вставку кодирующей полипептид последовательности.

132. Способ по варианту осуществления 124, где нуклеотидное изменение обеспечивает корректирование связанного с заболеванием гена.

133. Способ по варианту осуществления 132, где связанный с заболеванием ген связан с моногенетическим заболеванием, выбранным из группы, состоящей из: дефицита аденозиндезаминазы (ADA); дефицита альфа-1 антитрипсина; муковисцидоза; мышечной дистрофии Дюшенна; галактоземии; гемохроматоза; болезни Хантингтона; болезни кленового сиропа; синдрома Марфана; нейрофиброматоза 1 типа; врожденной пахионихии; фенилкеотнурии; тяжелого комбинированного иммунодефицита; серповидноклеточной анемии; синдрома Смита-Лемли-Опица; и болезни Тея-Сакса.

134. Способ по варианту осуществления 132, где связанный с заболеванием ген связан с полигенным заболеванием, выбранным из группы, состоящей из: болезни сердца, высокого кровяного давления, болезни Альцгеймера, артрита, диабета, рака, и ожирения.

135. Направляющая РНК для применения в рамках праймерного редактирования для изменения нуклеотидной последовательности молекулы-мишени ДНК с помощью вставки, делеции, инверсии, замещения, или их комбинации для получения соответствующей подвергнутой редактированию молекулы ДНК, где:

(i) направляющая РНК способна образовывать комплекс со слитым белком, включающим napDNAbp и домен, включающий активность РНК-зависимой ДНК-полимеразы;

(ii) направляющая РНК включает (a) спейсер, который способен отжигаться на не включающей PAM цепи, являющейся проксимальной относительно доступного PAM и протоспейсера на включающей PAM цепи в молекуле-мишени ДНК, и (b) ядро нРНК;

(iii) направляющая РНК дополнительно включает плечо удлинения на 5'- или 3'-конце направляющей РНК;

(iv) плечо удлинения включает (a) участок связывания праймера и (b) матрицу для синтеза ДНК, где матрица для синтеза ДНК кодирует одноцепочечный ДНК-флэп, который включает изменение для внесения вместо эндогенной цепи непосредственно ниже участка разрыва на включающей PAM цепи;

(v) молекула-мишень ДНК выбрана из группы, состоящей из SEQ ID NO: 1217353-1289387; a

(vi) соответствующая подвергнутая редактированию молекула ДНК выбрана из группы, состоящей из SEQ ID NO: 1289388-1361420.

136. Направляющая РНК по варианту осуществления 135, где молекула-мишень ДНК представляет собой последовательность варианта Clinvar.

137. Направляющая РНК по варианту осуществления 135, где napDNAbp представляет собой Cas9, Cas12e, Cas12d, Cas12a, Cas12b1, Cas13a, Cas12c, или Argonaute, или вариант Cas9, Cas12e, Cas12d, Cas12a, Cas12b1, Cas13a, Cas12c, или Argonaute.

138. Направляющая РНК по варианту осуществления 135, где домен napDNAbp включает никазную активность.

139. Направляющая РНК по варианту осуществления 135, где napDNAbp представляет собой Cas9 или его вариант.

140. Направляющая РНК по варианту осуществления 135, где napDNAbp представляет собой имеющий нуклеазную активность Cas9, не имеющий нуклеазной активности Cas9 (dCas9), или никазу Cas9 (nCas9).

141. Направляющая РНК по варианту осуществления 135, где napDNAbp представляет собой никазу Cas9 (nCas9).

142. Направляющая РНК по варианту осуществления 135, где napDNAbp включает аминокислотную

143. Направляющая РНК по варианту осуществления 135, где napDNAbp представляет собой SpCas9 дикого типа или его вариант согласно любой одной из аминокислотных последовательностей 1361421-1361428 или аминокислотную последовательность, имеющую, по меньшей мере, 80% идентичность последовательности относительно любой из SEQ ID NO: 1361421-1361428.

144. Направляющая РНК по варианту осуществления 135, где napDNAbp представляет собой ортолог SpCas9 согласно любой одной из аминокислотных последовательностей 1361429-1361442 или аминокислотную последовательность, имеющую, по меньшей мере, 80% идентичность последовательности относительно любой из SEQ ID NO: 1361429-1361442.

145. Направляющая РНК по варианту осуществления 135, где napDNAbp представляет собой любую одну из аминокислотных последовательностей 1361421-1361484 или аминокислотную последовательность, имеющую, по меньшей мере, 80% идентичность последовательности относительно любой из SEQ ID NO: 1361421-1361484.

146. Направляющая РНК по варианту осуществления 135, где домен, включающий активность РНК-зависимой ДНК-полимеразы представляет собой обратную транскриптазу.

147. Направляющая РНК по варианту осуществления 146, где обратная транскриптаза представляет собой встречающуюся в природе обратную транскриптазу дикого типа, имеющую аминокислотную последовательность любой одной из SEQ ID NO: 1361485-1361496 или аминокислотную последовательность, имеющую, по меньшей мере, 80% идентичность последовательности относительно любой из SEQ ID NO: 1361485-1361496.

148. Направляющая РНК по варианту осуществления 146, где обратная транскриптаза представляет собой вариант обратной транскриптазы, имеющий аминокислотную последовательность любой одной из SEQ ID NO: 1361497-1361514 или аминокислотную последовательность, имеющую, по меньшей мере, 80% идентичность последовательности относительно любой из SEQ ID NO: 1361497-1361514.

149. Направляющая РНК по варианту осуществления 135, где слитый белок включает аминокислотную последовательность любой одной из SEQ ID NO: 1361515-1361519 или аминокислотную последовательность, имеющую, по меньшей мере, 80% идентичность последовательности относительно любой из SEQ ID NO: 1361515-1361519.

150. Направляющая РНК по варианту осуществления 135, где слитый белок включает аминокислотную последовательность SEQ ID NO: 1361515 (PE1) или 1361516 (PE2), или аминокислотную последовательность, имеющую, по меньшей мере, 80% идентичность последовательности относительно любой из SEQ ID NO: 1361515 или 1361516.

151. Направляющая РНК по варианту осуществления 135, где доступная последовательность PAM зависит от napDNAbp, использованного на этапе (i).

152. Направляющая РНК по варианту осуществления 135, где доступная последовательность PAM выбрана из группы, состоящей из: (a) 5′-NGG-3′ (стандартная последовательность PAM), (b) 5′-NNG-3′, (c) 5′-NNA-3′, (d) 5′-NNC-3′, (e) 5′-NNT-3′, (f) 5′-NGT-3′, (g) 5′-NGA-3′, (h) 5′-NGC-3′, (i) 5′-NAA-3′, (j) 5′-NAC-3′, (k) 5′-NAG-3′, и (l) 5′-NAT-3′, выбор которой зависит от выбора napDNAbp.

153. Направляющая РНК по варианту осуществления 135, где участок изменения в любой из нуклеотидных последовательностей SEQ ID NO: 1217353-1289387 этапа (v) начинается в положении 201 в направлении от 5' к 3'.

154. Направляющая РНК по варианту осуществления 135, где нуклеотидное изменение представляет собой нуклеотидное замещение, делецию, вставку, или их комбинацию.

155. Направляющая РНК по варианту осуществления 135, где нуклеотидное замещение представляет собой транзицию или трансверсию.

156. Направляющая РНК по варианту осуществления 135, где замещение одного нуклеотида представляет собой (1) замещение G на T, (2) замещение G на A, (3) замещение G на C, (4) замещение T на G, (5) замещение T на A, (6) замещение T на C, (7) замещение C на G, (8) замещение C на T, (9) замещение C на A, (10) замещение A на T, (11) замещение A на G, или (12) замещение A на C.

157. Направляющая РНК по варианту осуществления 135, где замещение одного нуклеотида обеспечивает конвертирование (1) пары оснований G:C на пару оснований T:A, (2) пары оснований G:C на пару оснований A:T, (3) пары оснований G:C на пару оснований C:G, (4) пары оснований T:A на пару оснований G:C, (5) пары оснований T:A на пару оснований A:T, (6) пары оснований T:A на пару оснований C:G, (7) пары оснований C:G на пару оснований G:C, (8) пары оснований C:G на пару оснований T:A, (9) пары оснований C:G на пару оснований A:T, (10) пары оснований A:T на пару оснований T:A, (11) пары оснований A:T на пару оснований G:C, или (12) пары оснований A:T на пару оснований C:G.

158. Направляющая РНК по варианту осуществления 135, где требуемое нуклеотидное изменение представляет собой вставку или делецию 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, или 25 нуклеотидов.

159. Направляющая РНК по варианту осуществления 135, где нуклеотидное изменение представляет собой вставку кодирующей полипептид последовательности.

160. Направляющая РНК по варианту осуществления 135, где нуклеотидное изменение обеспечивает корректирование связанного с заболеванием гена.

161. Направляющая РКН по варианту осуществления 160, где связанный с заболеванием ген связан с моногенетическим заболеванием, выбранным из группы, состоящей из: дефицита аденозиндезаминазы (ADA); дефицита альфа-1 антитрипсина; муковисцидоза; мышечной дистрофии Дюшенна; галактоземии; гемохроматоза; болезни Хантингтона; болезни кленового сиропа; синдрома Марфана; нейрофиброматоза 1 типа; врожденной пахионихии; фенилкеотнурии; тяжелого комбинированного иммунодефицита; серповидноклеточной анемии; синдрома Смита-Лемли-Опица; и болезни Тея-Сакса.

162. Направляющая РНК по варианту осуществления 160, где связанный с заболеванием ген связан с полигенным заболеванием, выбранным из группы, состоящей из: болезни сердца, высокого кровяного давления, болезни Альцгеймера, артрита, диабета, рака, и ожирения.

163. Способ внесения нуклеотидного изменения в последовательность нуклеиновой кислоты, при этом способ включает: контактирование последовательности нуклеиновой кислоты с комплексом, включающим слитый белок и направляющую РНК по любому из вариантов осуществления 1-32 или 135-162.

164. Способ по варианту осуществления 163, где слитый белок включает napDNAbp и РНК-зависимую ДНК-полимеразу.

165. Способ по варианту осуществления 163, где направляющая РНК включает спейсер, ядро нРНК, и плечо удлинения, которое включает матрицу для синтеза ДНК и участок связывания праймера.

166. Способ по варианту осуществления 165, где матрица для синтеза ДНК кодирует нуклеотидное изменение.

167. Способ по любому из вариантов осуществления 163-166, где направляющая РНК способна связываться с napDNAbp, подходящим для праймерного редактирования, и направлять napDNAbp к последовательности-мишени ДНК.

168. Способ по варианту осуществления 167, где последовательность-мишень нуклеиновой кислоты включает цепь-мишень (или включающую PAM или подвергаемую редактированию цепь) и комплементарную не являющуюся мишенью цепь (или не включающую PAM или не подвергаемую редактированию цепь), где спейсер направляющей РНК гибридизируется с не включающей PAM цепью для образования гибрида РНК-ДНК и R-петли.

169. Способ по варианту осуществления 165, где участок связывания праймера имеет длину от приблизительно 8 до приблизительно 20 нуклеотидов.

170. Способ по варианту осуществления 165, где участок связывания праймера имеет длину 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, или 20 нуклеотидов.

171. Способ по варианту осуществления 165, где плечо удлинения имеет длину, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, по меньшей мере, 20 нуклеотидов, по меньшей мере, 21 нуклеотид, по меньшей мере, 22 нуклеотида, по меньшей мере, 23 нуклеотида, по меньшей мере, 24 нуклеотида, по меньшей мере, 25 нуклеотидов, по меньшей мере, 26 нуклеотидов, по меньшей мере, 27 нуклеотидов, по меньшей мере, 28 нуклеотидов, по меньшей мере, 29 нуклеотидов, по меньшей мере, 30 нуклеотидов, по меньшей мере, 31 нуклеотид, по меньшей мере, 32 нуклеотида, по меньшей мере, 33 нуклеотида, по меньшей мере, 34 нуклеотида, по меньшей мере, 35 нуклеотидов, по меньшей мере, 36 нуклеотидов, по меньшей мере, 37 нуклеотидов, по меньшей мере, 38 нуклеотидов, по меньшей мере, 39 нуклеотидов, по меньшей мере, 40 нуклеотидов, или, по меньшей мере, 100 нуклеотидов.

172. Способ по варианту осуществления 165, где участок связывания праймера имеет длину, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, или, по меньшей мере, 20 нуклеотидов.

173. Способ по варианту осуществления 165, где матрица для синтеза ДНК имеет длину, по меньшей мере, 1 нуклеотид, по меньшей мере, 2 нуклеотида, по меньшей мере, 3 нуклеотида, по меньшей мере, 4 нуклеотида, по меньшей мере, 5 нуклеотидов, по

меньшей мере, 6 нуклеотидов, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, по меньшей мере, 20 нуклеотидов, по меньшей мере, 21 нуклеотид, по меньшей мере, 22 нуклеотида, по меньшей мере, 23 нуклеотида, по меньшей мере, 24 нуклеотида, по меньшей мере, 25 нуклеотидов, по меньшей мере, 26 нуклеотидов, по меньшей мере, 27 нуклеотидов, по меньшей мере, 28 нуклеотидов, по меньшей мере, 29 нуклеотидов, по меньшей мере, 30 нуклеотидов, по меньшей мере, 31 нуклеотид, по меньшей мере, 32 нуклеотида, по меньшей мере, 33 нуклеотида, по меньшей мере, 34 нуклеотида, по меньшей мере, 35 нуклеотидов, по меньшей мере, 36 нуклеотидов, по меньшей мере, 37 нуклеотидов, по меньшей мере, 38 нуклеотидов, по меньшей мере, 39 нуклеотидов, по меньшей мере, 40 нуклеотидов, или, по меньшей мере, 100 нуклеотидов.

174. Способ по варианту осуществления 165, где матрица для синтеза ДНК может использоваться РНК-зависимой ДНК-полимеразой (например обратной транскриптазой) в качестве матрицы для синтеза соответствующего одноцепочечного ДНК-флэпа, имеющего 3'-конец, где ДНК-флэп комплементарен цепи последовательности эндогенной ДНК-мишени, примыкающей к участку разрыва, и где одноцепочечный ДНК-флэп включает требуемое нуклеотидное изменение, кодируемое матрицей для синтеза ДНК.

175. Способ по варианту осуществления 174, где одноцепочечный ДНК-флэп вытесняет эндогенную одноцепочечную ДНК, имеющую 5′-конец, в последовательности-мишени ДНК, которая была никирована.

176. Способ по варианту осуществления 175, где эндогенная одноцепочечная ДНК, имеющая свободный 5′-конец, подвергается эксцизии клеткой.

177. Способ по варианту осуществления 175, где клеточная репарация одноцепочечного ДНК-флэпа приводит к внесению нуклеотидного изменения, тем самым обеспечивая образование подвергнутого редактированию ДНК-продукта.

178. Способ по варианту осуществления 177, где нуклеотидное изменение представляет собой вставку.

179. Способ по варианту осуществления 178, где вставка имеет длину, по меньшей мере, 1 нуклеотид, по меньшей мере, 2 нуклеотида, по меньшей мере, 3 нуклеотида, по меньшей мере, 4 нуклеотида, по меньшей мере, 5 нуклеотидов, по меньшей мере, 6 нуклеотидов, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, по меньшей мере, 20 нуклеотидов, по меньшей мере, 21 нуклеотид, по меньшей мере, 22 нуклеотида, по меньшей мере, 23 нуклеотида, по меньшей мере, 24 нуклеотида, по меньшей мере, 25 нуклеотидов, по меньшей мере, 26 нуклеотидов, по меньшей мере, 27 нуклеотидов, по меньшей мере, 28 нуклеотидов, по меньшей мере, 29 нуклеотидов, по меньшей мере, 30 нуклеотидов, по меньшей мере, 31 нуклеотид, по меньшей мере, 32 нуклеотида, по меньшей мере, 33 нуклеотида, по меньшей мере, 34 нуклеотида, по меньшей мере, 35 нуклеотидов, по меньшей мере, 36 нуклеотидов, по меньшей мере, 37 нуклеотидов, по меньшей мере, 38 нуклеотидов, по меньшей мере, 39 нуклеотидов, по меньшей мере, 40 нуклеотидов, или, по меньшей мере, 100 нуклеотидов.

180. Способ по варианту осуществления 178, где вставка представляет собой последовательность, кодирующую полипептид.

ССЫЛКИ

[0902] Каждая из следующих ссылок полностью приведена здесь в качестве ссылки.

1. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–821 (2012).

2. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819–823 (2013).

3. Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell* **168**, 20–36 (2017).

4. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).

5. Nishida, K. *et al.* Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, aaf8729 (2016).

6. Gaudelli, N. M. *et al.* Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).

7. Stenson, P. D. *et al.* The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum. Genet.* **136**, 665–677 (2017).

8. Dunbar, C. E. *et al.* Gene therapy comes of age. *Science* **359**, eaan4672 (2018).

9. Cox, D. B. T., Platt, R. J. & Zhang, F. Therapeutic genome editing: prospects and challenges. *Nat. Med.* **21**, 121–131 (2015).

10. Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* **9**, 1911 (2018).

11. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–485 (2015).

12. Kleinstiver, B. P. *et al.* High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).

13. Hu, J. H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63 (2018).

14. Nishimasu, H. *et al.* Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* **361**, 1259–1262 (2018).

15. Jasin, M. & Rothstein, R. Repair of strand breaks by homologous recombination. *Cold Spring Harb. Perspect. Biol.* **5**, a012740 (2013).

16. Paquet, D. *et al.* Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* **533**, 125–129 (2016).

17. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018).

18. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.* **24**, 927–930 (2018).

19. Ihry, R. J. *et al.* p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. *Nat. Med.* **24**, 939–946 (2018).

20. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* **34**, 339–344 (2016).

21. Srivastava, M. *et al.* An Inhibitor of Nonhomologous End-Joining Abrogates Double-Strand Break Repair and Impedes Cancer Progression. *Cell* **151**, 1474–1487 (2012).

22. Chu, V. T. *et al.* Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.* **33**, 543–548 (2015).

23. Maruyama, T. *et al.* Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat. Biotechnol.* **33**, 538–542 (2015).

24. Kim, Y. B. *et al.* Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat. Biotechnol.* **35**, 371–376 (2017).

25. Li, X. *et al.* Base editing with a Cpf1–cytidine deaminase fusion. *Nat. Biotechnol.* **36**, 324–327 (2018).

26. Gehrke, J. M. *et al.* An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat. Biotechnol.* (2018). doi:10.1038/nbt.4199

27. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* 1 (2018). doi:10.1038/s41576-018-0059-1.

28. Ostertag, E. M. & Kazazian Jr, H. H. Biology of Mammalian L1 Retrotransposons. *Annu. Rev. Genet.* **35**, 501–538 (2001).

29. Zimmerly, S., Guo, H., Perlman, P. S. & Lambowltz, A. M. Group II intron mobility occurs by target DNA- primed reverse transcription. *Cell* **82**, 545–554 (1995).

30. Luan, D. D., Korman, M. H., Jakubczak, J. L. & Eickbush, T. H. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* **72**, 595–605 (1993).

31. Feng, Q., Moran, J. V., Kazazian, H. H. & Boeke, J. D. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87**, 905–916 (1996).

32. Jinek, M. *et al.* Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. *Science* **343**, 1247997 (2014).

33. Jiang, F. *et al.* Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* aad8282 (2016). doi:10.1126/science.aad8282

34. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152**, 1173–1183 (2013).

35. Tang, W., Hu, J. H. & Liu, D. R. Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation. *Nat. Commun.* **8**, 15939 (2017).

36. Shechner, D. M., Hacisuleyman, E., Younger, S. T. & Rinn, J. L. Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. *Nat. Methods* **12**, 664–670 (2015).

37. Anders, C. & Jinek, M. Chapter One - In Vitro Enzymology of Cas9. in *Methods in Enzymology* (eds. Doudna, J. A. & Sontheimer, E. J.) **546**, 1–20 (Academic Press, 2014).

38. Briner, A. E. *et al.* Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality. *Mol. Cell* **56**, 333–339 (2014).

39. Nowak, C. M., Lawson, S., Zerez, M. & Bleris, L. Guide RNA engineering for versatile Cas9 functionality. *Nucleic Acids Res.* **44**, 9555–9564 (2016).

40. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67 (2014).

41. Mohr, S. *et al.* Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing. *RNA* **19**, 958–970 (2013).

42. Stamos, J. L., Lentzsch, A. M. & Lambowitz, A. M. Structure of a Thermostable Group II Intron Reverse Transcriptase with Template-Primer and Its Functional and Evolutionary Implications. *Mol. Cell* **68**, 926- 939.e4 (2017).

43. Zhao, C. & Pyle, A. M. Crystal structures of a group II intron maturase reveal a missing link in spliceosome evolution. *Nat. Struct. Mol. Biol.* **23**, 558–565 (2016).

44. Zhao, C., Liu, F. & Pyle, A. M. An ultraprocessive, accurate reverse transcriptase encoded by a metazoan group II intron. *RNA* **24**, 183–195 (2018).

45. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).

46. Liu, Y., Kao, H.-I. & Bambara, R. A. Flap endonuclease 1: a central component of DNA metabolism. *Annu. Rev. Biochem.* **73**, 589–615 (2004).

47. Krokan, H. E. & Bjørås, M. Base Excision Repair. *Cold Spring Harb. Perspect. Biol.* **5**, (2013).

48. Kelman, Z. PCNA: structure, functions and interactions. *Oncogene* **14**, 629–640 (1997).

49. Choe, K. N. & Moldovan, G.-L. Forging Ahead through Darkness: PCNA, Still the Principal Conductor at the Replication Fork. *Mol. Cell* **65**, 380–392 (2017).

50. Li, X., Li, J., Harrington, J., Lieber, M. R. & Burgers, P. M. Lagging strand DNA synthesis at the eukaryotic replication fork involves binding and stimulation of FEN-1 by proliferating cell nuclear antigen. *J. Biol. Chem.* **270**, 22109–22112 (1995).

51. Tom, S., Henricksen, L. A. & Bambara, R. A. Mechanism whereby proliferating cell nuclear antigen stimulates flap endonuclease 1. *J. Biol. Chem.* **275**, 10498–10505 (2000).

52. Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**, 635–646 (2014).

53. Bertrand, E. *et al.* Localization of ASH1 mRNA particles in living yeast. *Mol. Cell* **2**, 437–445 (1998).

54. Dahlman, J. E. *et al.* Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease. *Nat. Biotechnol.* **33**, 1159–1161 (2015).

55. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).

56. Tsai, S. Q. *et al.* CIRCLE-seq: a highly sensitive *in vitro* screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat. Methods* **14**, 607–614 (2017).

ЭКВИВАЛЕНТЫ И ОХВАТЫВАЕМЫЙ ОБЪЕМ

[0903] В рамках пунктов формулы изобретения существительные могут обозначать один или более одного, если иное не указано или не следует из контекста. Пункты формулы изобретения или фрагменты описания, которые включают "или" между одним или несколькими элементами группы, являются действительными, если один, более чем один, или все из элементов группы присутствуют, используются, или иным образом связаны с определенным продуктом или процессом, если иное не указано или не следует из контекста. Изобретение включает варианты осуществления, в которых конкретно один элемент группы присутствует, используется, или иным образом связан с определенным продуктом или процессом. Изобретение включает варианты осуществления, в которых

более чем один или все из элементов группы присутствуют, используются, или иным образом являются связанными с определенным продуктом или процессом.

[0904] Более того, изобретение охватывает все вариации, комбинации, и пермутации, в рамках которых одно или несколько ограничений, элементов, пунктов, и описательных элементов из одного или нескольких перечисленных пунктов формулы изобретения внесены в другой пункт формулы изобретения. Например, любой пункт формулы изобретения, который зависит от другого пункта формулы изобретения, может быть модифицирован для включения одного или нескольких ограничений, указанных в любом другом пункте формулы изобретения, который зависит от того же первоначального пункта формулы изобретения. Если элементы представлены в виде перечней, например в формате группы Маркуша, каждая подгруппа данных элементов также включается, при этом любой элемент (элементы) могут быть удалены из группы. Следует понимать, что, в целом, при обозначении изобретения или аспектов изобретения в качестве включающих определенные элементы и/или свойства, определенные варианты осуществления изобретения или аспекты изобретения состоят или состоят по существу из подобных элементов и/или свойств. Для упрощения такие варианты осуществления не были явно приведены *in haec verba* здесь. Следует также отметить, что термины "включающий" и "содержащий" являются открытыми и допускают включение дополнительных элементов или этапов. При указании диапазонов их границы включаются. Более того, если иное не указано или не следует из контекста и понимания среднего специалиста в данной области, значения, которые выражены в виде диапазонов, могут соответствовать любому конкретному значению или поддиапазону в рамках указанных диапазонов в различных вариантах осуществления изобретения до десятой доли единицы нижней границы диапазона, если из контекста явно не следует иное.

[0905] Данная заявка включает ссылки на различные выданные патенты, опубликованные патентные заявки, журнальные статьи, и другие публикации, все из которых приведены здесь в качестве ссылки. В случае несоответствия между какими-либо включенными ссылками и настоящим описанием описание обладает приоритетом. Кроме того, любой отдельный вариант осуществления настоящего изобретения, который соответствует уровню техники, может быть явно исключен из любого одного или нескольких пунктов формулы изобретения. Поскольку такие варианты осуществления рассматриваются в

качестве известных среднему специалисту в данной области, они могут быть исключены, даже если исключение явно не указано здесь. Любой конкретный вариант осуществления изобретения может быть исключен из любого пункта формулы изобретения по любой причине, которая связана или не связана с наличием уровня техники.

[0906] Специалистам в данной области будут очевидны или они будут способны определить с помощью стандартных экспериментов многие эквиваленты определенных описанных здесь вариантов осуществления. Объем настоящих описанных здесь вариантов осуществления не ограничивается приведенным выше Описанием, но основан на прилагаемой формуле изобретения. Средним специалистам в данной области будет очевидно, что могут быть реализованы различные изменения и модификации данного изобретения без выхода за рамки духа или объема настоящего изобретения, как определено в рамках приведенной ниже формулы изобретения.

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Направляющая РНК, включающая спейсер, ядро нРНК, и плечо удлинения, где направляющая РНК включает последовательность, выбранную из группы, состоящей из SEQ ID NO: 1-135514, или последовательность, имеющую, по меньшей мере, 90% идентичность последовательности относительно любой из SEQ ID NO: 1-135514.

2. Направляющая РНК, включающая спейсер, ядро нРНК, и плечо удлинения, где спейсер включает нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 135515 – 271028, или спейсер, имеющий нуклеотидную последовательность, имеющую, по меньшей мере, 90% идентичность последовательности относительно любой из SEQ ID NO: 135515 – 271028.

3. Направляющая РНК, включающая спейсер, ядро нРНК, и плечо удлинения, где плечо удлинения имеет нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 271029 – 406542, или плечо удлинения, имеющее нуклеотидную последовательность, имеющую, по меньшей мере, 90% идентичность последовательности относительно любой из SEQ ID NO: 271029 – 406542.

4. Направляющая РНК, включающая спейсер, ядро нРНК, и плечо удлинения, где плечо удлинения включает (i) участок связывания праймера, (ii) матрицу для редактирования, и (iii) плечо гомологии.

5. Направляющая РНК, включающая спейсер, ядро нРНК, и плечо удлинения, где плечо удлинения включает участок связывания праймера, имеющий нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 406543 – 542056, или участок связывания праймера, имеющий нуклеотидную последовательность, которая, по меньшей мере, на 90% идентична последовательности любой из SEQ ID NO: 406543 – 542056.

6. Направляющая РНК, включающая спейсер, ядро нРНК, и плечо удлинения, где плечо удлинения включает матрицу для редактирования, включающую нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 542057 – 677570, или матрицу для редактирования, имеющую нуклеотидную последовательность, которая, по меньшей мере, на 90% идентична любой из SEQ ID NO: 542057 – 677570.

7. Направляющая РНК, включающая спейсер, ядро нРНК, и плечо удлинения, где плечо удлинения включает плечо гомологии, имеющее нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 677571 – 813084, или плечо гомологии, имеющее нуклеотидную последовательность, которая, по меньшей мере, на 90% идентична любой из SEQ ID NO: 677571 – 813084.

8. Направляющая РНК, включающая:
   (i) спейсер, имеющий нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 135515 – 271028, или спейсер, имеющий нуклеотидную последовательность, имеющую, по меньшей мере, 90% идентичность последовательности относительно любой из SEQ ID NO: 135515 – 271028, и
   (ii) плечо удлинения, выбранное из группы, состоящей из SEQ ID NO: 271029 – 406542, или плечо удлинения, имеющее нуклеотидную последовательность, имеющую, по меньшей мере, 90% идентичность последовательности относительно SEQ ID NO: 271029 – 406542.

9. Направляющая РНК, включающая:
   (i) спейсер, имеющий нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 135515 – 271028, или спейсер, имеющий нуклеотидную последовательность, которая, по меньшей мере, на 90% идентична любой из SEQ ID NO: 135515 – 271028, и
   (ii) участок связывания праймера, выбранный из группы, состоящей из SEQ ID NO: 406543 – 542056, или участок связывания праймера, имеющий

нуклеотидную последовательность, которая, по меньшей мере, на 90% идентична любой из SEQ ID NO: 406543 – 542056.

10. Направляющая РНК, включающая:

(i) спейсер, имеющий нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 135515 – 271028, или спейсер, имеющий нуклеотидную последовательность, имеющую, по меньшей мере, 90% идентичность последовательности относительно любой из SEQ ID NO: 135515 – 271028, и

(ii) матрицу для редактирования, имеющую нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 542057 – 677570, или матрицу для редактирования, имеющую нуклеотидную последовательность, которая, по меньшей мере, на 90% идентична любой из SEQ ID NO: 542057 – 677570.

11. Направляющая РНК, включающая:

(i) спейсер, имеющий нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 135515 – 271028, или спейсер, имеющий нуклеотидную последовательность, которая, по меньшей мере, на 90% идентична любой из SEQ ID NO: 135515 – 271028, и

(ii) плечо гомологии, имеющее нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 677571 – 813084, или спейсер, имеющий нуклеотидную последовательность, которая, по меньшей мере, на 90% идентична любой из SEQ ID NO: 677571 – 813084.

12. Направляющая РНК по любому из предыдущих пунктов, дополнительно включающая сигнал терминации SEQ ID NO: 813086 или сигнал терминации, имеющий, по меньшей мере, 90% идентичность последовательности относительно SEQ ID NO: 813086.

13. Направляющая РНК по любому из предыдущих пунктов, дополнительно включающая 5'-концевой модификаторный участок, включающий

последовательность шпильки, последовательность стебля/петли, или последовательность опорной петли.

14. Направляющая РНК по любому из предыдущих пунктов, дополнительно включающая 3'-концевой модификаторный участок, включающий последовательность шпильки, последовательность стебля/петли, или последовательность опорной петли.

15. Направляющая РНК по любому из предыдущих пунктов, дополнительно включающая ядро нРНК, включающее SEQ ID NO: 813085, или ядро нРНК, имеющее, по меньшей мере, 90% идентичность последовательности относительно SEQ ID NO: 813085.

16. Направляющая РНК по любому из предыдущих пунктов, отличающаяся тем, что направляющая РНК способна связываться с napDNAbp, подходящим для праймерного редактирования, и направлять napDNAbp к последовательности-мишени ДНК.

17. Направляющая РНК по п.16, отличающаяся тем, что последовательность-мишень нуклеиновой кислоты включает цепь-мишень (или включающую PAM цепь) и комплементарную не являющуюся мишенью цепь (или не включающую PAM цепь), где спейсер направляющей РНК гибридизируется с комплементарной не являющейся мишенью цепью (не включающей PAM цепью) для образования гибрида РНК-ДНК и R-петли.

18. Направляющая РНК по любому из предыдущих пунктов, отличающаяся тем, что участок связывания праймера имеет длину от приблизительно 8 до приблизительно 20 нуклеотидов.

19. Направляющая РНК по любому из предыдущих пунктов, отличающаяся тем, что участок связывания праймера имеет длину 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, или 20 нуклеотидов.

20. Направляющая РНК по любому из предыдущих пунктов, отличающаяся тем, что участок связывания праймера имеет длину, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, или, по меньшей мере, 20 нуклеотидов.

21. Направляющая РНК по любому из предыдущих пунктов, отличающаяся тем, что плечо гомологии комплементарно цепи ДНК-мишени.

22. Направляющая РНК по любому из предыдущих пунктов, отличающаяся тем, что плечо удлинения имеет длину от приблизительно 7 до приблизительно 500 нуклеотидов.

23. Направляющая РНК по любому из предыдущих пунктов, отличающаяся тем, что плечо удлинения имеет длину, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, по меньшей мере, 20 нуклеотидов, по меньшей мере, 21 нуклеотид, по меньшей мере, 22 нуклеотида, по меньшей мере, 23 нуклеотида, по меньшей мере, 24 нуклеотида, по меньшей мере, 25 нуклеотидов, по меньшей мере, 26 нуклеотидов, по меньшей мере, 27 нуклеотидов, по меньшей мере, 28 нуклеотидов, по меньшей мере, 29 нуклеотидов, по меньшей мере, 30 нуклеотидов, по меньшей мере, 31 нуклеотид, по меньшей мере, 32 нуклеотида, по меньшей мере, 33 нуклеотида, по меньшей мере, 34 нуклеотида, по меньшей мере, 35 нуклеотидов, по меньшей мере, 36 нуклеотидов, по меньшей мере, 37 нуклеотидов, по меньшей мере, 38 нуклеотидов,

по меньшей мере, 39 нуклеотидов, по меньшей мере, 40 нуклеотидов, или, по меньшей мере, 100 нуклеотидов.

24. Направляющая РНК по любому из предыдущих пунктов, отличающаяся тем, что матрица для редактирования имеет длину, по меньшей мере, 1 нуклеотид, по меньшей мере, 2 нуклеотида, по меньшей мере, 3 нуклеотида, по меньшей мере, 4 нуклеотида, по меньшей мере, 5 нуклеотидов, по меньшей мере, 6 нуклеотидов, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, по меньшей мере, 20 нуклеотидов, по меньшей мере, 21 нуклеотид, по меньшей мере, 22 нуклеотида, по меньшей мере, 23 нуклеотида, по меньшей мере, 24 нуклеотида, по меньшей мере, 25 нуклеотидов, по меньшей мере, 26 нуклеотидов, по меньшей мере, 27 нуклеотидов, по меньшей мере, 28 нуклеотидов, по меньшей мере, 29 нуклеотидов, по меньшей мере, 30 нуклеотидов, по меньшей мере, 31 нуклеотид, по меньшей мере, 32 нуклеотида, по меньшей мере, 33 нуклеотида, по меньшей мере, 34 нуклеотида, по меньшей мере, 35 нуклеотидов, по меньшей мере, 36 нуклеотидов, по меньшей мере, 37 нуклеотидов, по меньшей мере, 38 нуклеотидов, по меньшей мере, 39 нуклеотидов, по меньшей мере, 40 нуклеотидов, или, по меньшей мере, 100 нуклеотидов.

25. Направляющая РНК по любому из предыдущих пунктов, отличающаяся тем, что плечо гомологии имеет длину, по меньшей мере, 1 нуклеотид, по меньшей мере, 2 нуклеотида, по меньшей мере, 3 нуклеотида, по меньшей мере, 4 нуклеотида, по меньшей мере, 5 нуклеотидов, по меньшей мере, 6 нуклеотидов, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов,

по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, по меньшей мере, 20 нуклеотидов, по меньшей мере, 21 нуклеотид, по меньшей мере, 22 нуклеотида, по меньшей мере, 23 нуклеотида, по меньшей мере, 24 нуклеотида, по меньшей мере, 25 нуклеотидов, по меньшей мере, 26 нуклеотидов, по меньшей мере, 27 нуклеотидов, по меньшей мере, 28 нуклеотидов, по меньшей мере, 29 нуклеотидов, или, по меньшей мере, 30 нуклеотидов.

26. Направляющая РНК по любому из предыдущих пунктов, отличающаяся тем, что матрица для редактирования и плечо гомологии могут использоваться обратной транскриптазой в качестве последовательности матрицы для синтеза соответствующего одноцепочечного ДНК-флэпа, имеющего 3'-конец, где ДНК-флэп комплементарен цепи последовательности эндогенной ДНК-мишени, примыкающей к участку разрыва, и где одноцепочечный ДНК-флэп включает нуклеотидное изменение, кодируемое матрицей для редактирования.

27. Направляющая РНК по п. 26, отличающаяся тем, что одноцепочечный ДНК-флэп вытесняет эндогенную одноцепочечную ДНК, имеющую 5′-конец, в последовательности-мишени ДНК, которая была никирована.

28. Направляющая РНК по п.27, отличающаяся тем, что эндогенная одноцепочечная ДНК, имеющая свободный 5′-конец, подвергается эксцизии клеткой.

29. Направляющая РНК по п. 27, отличающаяся тем, что клеточная репарация одноцепочечного ДНК-флэпа приводит к внесению нуклеотидного изменения, тем самым обеспечивая образование требуемого продукта.

30. Направляющая РНК по п. 29, отличающаяся тем, что требуемое нуклеотидное изменение представляет собой вставку.

31. Направляющая РНК по п.30, отличающаяся тем, что вставка имеет длину, по меньшей мере, 1 нуклеотид, по меньшей мере, 2 нуклеотида, по меньшей мере, 3 нуклеотида, по меньшей мере, 4 нуклеотида, по меньшей мере, 5 нуклеотидов, по меньшей мере, 6 нуклеотидов, по меньшей мере, 7 нуклеотидов, по меньшей мере,

8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, по меньшей мере, 20 нуклеотидов, по меньшей мере, 21 нуклеотид, по меньшей мере, 22 нуклеотида, по меньшей мере, 23 нуклеотида, по меньшей мере, 24 нуклеотида, по меньшей мере, 25 нуклеотидов, по меньшей мере, 26 нуклеотидов, по меньшей мере, 27 нуклеотидов, по меньшей мере, 28 нуклеотидов, по меньшей мере, 29 нуклеотидов, по меньшей мере, 30 нуклеотидов, по меньшей мере, 31 нуклеотид, по меньшей мере, 32 нуклеотида, по меньшей мере, 33 нуклеотида, по меньшей мере, 34 нуклеотида, по меньшей мере, 35 нуклеотидов, по меньшей мере, 36 нуклеотидов, по меньшей мере, 37 нуклеотидов, по меньшей мере, 38 нуклеотидов, по меньшей мере, 39 нуклеотидов, по меньшей мере, 40 нуклеотидов, или, по меньшей мере, 100 нуклеотидов.

32. Направляющая РНК по п.30, отличающаяся тем, что вставка представляет собой последовательность, кодирующую полипептид.

33. Комплекс для праймерного редактирования, включающий napDNAbp, обратную транскриптазу, и любую одну из направляющих РНК по п.п.1-32.

34. Комплекс для праймерного редактирования по п.33, отличающийся тем, что napDNAbp и обратная транскриптаза образуют слитый белок.

35. Комплекс для праймерного редактирования по п.33, отличающийся тем, что napDNAbp представляет собой Cas9.

36. Комплекс для праймерного редактирования по п. 35, отличающийся тем, что Cas9 выбран из группы, состоящей из никаз Cas9 или их вариантов.

37. Комплекс для праймерного редактирования по п.35, отличающийся тем, что Cas9 имеет аминокислотную последовательность, выбранную из группы, состоящей из SEQ ID NO: 1-135514.

38. Комплекс для праймерного редактирования по п.34, отличающийся тем, что слитый белок имеет аминокислотную последовательность, выбранную из группы, состоящей из SEQ ID NO: 1-135514.

39. Комплекс для праймерного редактирования по п.34, отличающийся тем, что слитый белок включает линкер, соединяющий napDNAbp и обратную транскриптазу.

40. Комплекс для праймерного редактирования по п.39, отличающийся тем, что линкер имеет аминокислотную последовательность, выбранную из группы, состоящей из SEQ ID NO: 1-135514.

41. Один или несколько полинуклеотидов, кодирующих комплекс для праймерного редактирования по любому из п.п.33-40.

42. Вектор, включающий полинуклеотид по п.41 и один или несколько промоторов, которые обеспечивают экспрессию направляющей РНК и слитого белка комплекса для праймерного редактирования.

43. Клетка, включающая вектор по п.42.

44. Клетка, включающая комплекс для праймерного редактирования по любому из п.п.33-40.

45. Фармацевтическая композиция, включающая: (i) направляющую РНК по любому из п.п.1-32, комплекс для праймерного редактирования по п.п.33-40, полинуклеотид по п.41, или вектор по п.42; и (ii) фармацевтически приемлемый наполнитель.

46. Способ внесения нуклеотидного изменения в последовательность нуклеиновой кислоты, при этом способ включает: контактирование последовательности нуклеиновой кислоты с комплексом, включающим слитый белок и направляющую РНК по любому из п.п.1-32 или любому из п.п.56-81, где слитый белок включает napDNAbp и полимеразу, и где направляющая РНК включает спейсер, ядро нРНК,

и плечо удлинения, которое включает матрицу для редактирования, кодирующую нуклеотидное изменение, тем самым обеспечивая

(i) никирование последовательности двуцепочечной ДНК в цепи-мишени (или включающей PAM цепи) и создание свободной одноцепочечной ДНК, имеющей 3´-конец;

(ii) гибридизирование 3'-конца свободной одноцепочечной ДНК с направляющей РНК в участке связывания праймера, тем самым обеспечивая праймирование полимеразы;

(iii) полимеризацию цепи ДНК с 3´-конца, тем самым обеспечивая создание одноцепочечного ДНК-флэпа, включающего нуклеотидное изменение; и

(iv) замещение цепи эндогенной ДНК, непосредственно снизу примыкающей к участку разрыва на цепи-мишени (или включающей PAM цепи), одноцепочечным ДНК-флэпом, тем самым обеспечивая внесение требуемого нуклеотидного изменения в последовательность двуцепочечной ДНК.

47. Способ по п.46, отличающийся тем, что нуклеотидное изменение представляет собой замещение одного нуклеотида, делецию, вставку, или их комбинацию

48. Способ по п.46, отличающийся тем, что замещение одного нуклеотида представляет собой транзицию или трансверсию.

49. Способ по п.46, отличающийся тем, что нуклеотидное изменение представляет собой (1) замещение G на T, (2) замещение G на A, (3) замещение G на C, (4) замещение T на G, (5) замещение T на A, (6) замещение T на C, (7) замещение C на G, (8) замещение C на T, (9) замещение C на A, (10) замещение A на T, (11) замещение A на G, или (12) замещение A на C.

50. Способ по п.46, отличающийся тем, что нуклеотидное изменение обеспечивает конвертирование (1) пары оснований G:C на пару оснований T:A, (2) пары оснований G:C на пару оснований A:T, (3) пары оснований G:C на пару оснований C:G, (4) пары оснований T:A на пару оснований G:C, (5) пары оснований T:A на пару оснований A:T, (6) пары оснований T:A на пару оснований C:G, (7) пары

оснований C:G на пару оснований G:C, (8) пары оснований C:G на пару оснований T:A, (9) пары оснований C:G на пару оснований A:T, (10) пары оснований A:T на пару оснований T:A, (11) пары оснований A:T на пару оснований G:C, или (12) пары оснований A:T на пару оснований C:G.

51. Способ по п.46, отличающийся тем, что нуклеотидное изменение представляет собой вставку или делецию 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, или 25 нуклеотидов.

52. Способ по п.46, отличающийся тем, что нуклеотидное изменение представляет собой вставку кодирующей полипептид последовательности.

53. Способ по п.46, отличающийся тем, что нуклеотидное изменение обеспечивает корректирование связанного с заболеванием гена.

54. Способ по п.46, отличающийся тем, что связанный с заболеванием ген связан с моногенетическим заболеванием, выбранным из группы, состоящей из: дефицита аденозиндезаминазы (ADA); дефицита альфа-1 антитрипсина; муковисцидоза; мышечной дистрофии Дюшенна; галактоземии; гемохроматоза; болезни Хантингтона; болезни кленового сиропа; синдрома Марфана; нейрофиброматоза 1 типа; врожденной пахионихии; фенилкеотнурии; тяжелого комбинированного иммунодефицита; серповидноклеточной анемии; синдрома Смита-Лемли-Опица; и болезни Тея-Сакса.

55. Способ по п.46, отличающийся тем, что связанный с заболеванием ген связан с полигенным заболеванием, выбранным из группы, состоящей из: болезни сердца, высокого кровяного давления, болезни Альцгеймера, артрита, диабета, рака, и ожирения.

56. Направляющая РНК для применения в рамках праймерного редактирования для корректирования связанного с заболеванием аллеля в последовательности-мишени ДНК для получения здорового аллеля, при этом указанная направляющая включает спейсер, ядро нРНК, и плечо удлинения, где спейсер способен связываться с

участком длиной ~20 нуклеотидов в SEQ ID NO: 1217353-1289387 или его комплементарной цепью.

57. Направляющая РНК, включающая спейсер, ядро нРНК, и плечо удлинения, где плечо удлинения включает матрицу для синтеза ДНК и участок связывания праймера, являющиеся эффективными при осуществлении праймерного редактирования.

58. Направляющая РНК по п.56, отличающаяся тем, что участок изменения в любой из нуклеотидных последовательностей SEQ ID NO: 1217353-1289387 начинается в положении 201 в направлении от 5' к 3'.

59. Направляющая РНК для праймерного редактирования, включающая спейсер, ядро нРНК, и плечо удлинения, где плечо удлинения включает участок связывания праймера и матрицу для синтеза ДНК.

60. Направляющая РНК по п. 59, отличающаяся тем, что участок связывания праймера имеет нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 406543 – 542056 (участок связывания праймера), или нуклеотидную последовательность, которая имеет, по меньшей мере, 90% идентичность последовательности относительно любой из SEQ ID NO: 406543 – 542056.

61. Направляющая РНК по п.59, отличающаяся тем, что матрица для синтеза ДНК включает нуклеотидную последовательность SEQ ID NO: 542057 – 677570 (матрица для редактирования) или нуклеотидную последовательность, которая имеет, по меньшей мере, 90% идентичность последовательности относительно любой из SEQ ID NO: 542057 – 677570.

62. Направляющая РНК по п.59, отличающаяся тем, что матрица для синтеза ДНК включает нуклеотидную последовательность SEQ ID NO: 677571 – 813084 (плечо гомологии) или нуклеотидную последовательность, которая имеет, по меньшей мере, 90% идентичность последовательности относительно любой из SEQ ID NO: 677571 – 813084.

63. Направляющая РНК по п.59, отличающаяся тем, что матрица для синтеза ДНК включает матрицу для редактирования и плечо гомологии, где матрица для редактирования включает нуклеотидную последовательность SEQ ID NO: 542057 – 677570, а плечо гомологии включает нуклеотидную последовательность SEQ ID NO: 677571 – 813084.

64. Направляющая РНК по любому из п.п.56-63, дополнительно включающая сигнал терминации SEQ ID NO: 813086 или сигнал терминации, имеющий, по меньшей мере, 90% идентичность последовательности относительно SEQ ID NO: 813086.

65. Направляющая РНК по любому из п.п.56-64, дополнительно включающая 5'-концевой модификаторный участок, включающий последовательность шпильки, последовательность стебля/петли, или последовательность опорной петли.

66. Направляющая РНК по любому из п.п.56-65, дополнительно включающая 3'-концевой модификаторный участок, включающий последовательность шпильки, последовательность стебля/петли, или последовательность опорной петли.

67. Направляющая РНК по любому из п.п.56-66, дополнительно включающая ядро нРНК, включающее SEQ ID NO: 813085, или ядро нРНК, имеющее, по меньшей мере, 90% идентичность последовательности относительно SEQ ID NO: 813085.

68. Направляющая РНК по любому из п.п.56-67, отличающаяся тем, что направляющая РНК способна связываться с napDNAbp, подходящим для праймерного редактирования, и направлять napDNAbp к последовательности-мишени ДНК.

69. Направляющая РНК по п.68, отличающаяся тем, что последовательность-мишень нуклеиновой кислоты включает цепь-мишень (или включающую PAM или подвергаемую редактированию цепь) и комплементарную не являющуюся мишенью цепь (или не включающую PAM или не подвергаемую редактированию цепь), где спейсер направляющей РНК гибридизируется с не включающей PAM цепью для образования гибрида РНК-ДНК и R-петли.

70. Направляющая РНК по любому из п.п.56-69, отличающаяся тем, что участок связывания праймера имеет длину от приблизительно 8 до приблизительно 20 нуклеотидов.

71. Направляющая РНК по любому из п.п.56-70, где участок связывания праймера имеет длину 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, или 20 нуклеотидов.

72. Направляющая РНК по любому из п.п.56-71, отличающаяся тем, что плечо удлинения имеет длину, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, по меньшей мере, 20 нуклеотидов, по меньшей мере, 21 нуклеотид, по меньшей мере, 22 нуклеотида, по меньшей мере, 23 нуклеотида, по меньшей мере, 24 нуклеотида, по меньшей мере, 25 нуклеотидов, по меньшей мере, 26 нуклеотидов, по меньшей мере, 27 нуклеотидов, по меньшей мере, 28 нуклеотидов, по меньшей мере, 29 нуклеотидов, по меньшей мере, 30 нуклеотидов, по меньшей мере, 31 нуклеотид, по меньшей мере, 32 нуклеотида, по меньшей мере, 33 нуклеотида, по меньшей мере, 34 нуклеотида, по меньшей мере, 35 нуклеотидов, по меньшей мере, 36 нуклеотидов, по меньшей мере, 37 нуклеотидов, по меньшей мере, 38 нуклеотидов, по меньшей мере, 39 нуклеотидов, по меньшей мере, 40 нуклеотидов, или, по меньшей мере, 100 нуклеотидов.

73. Направляющая РНК по любому из п.п.56-72, где участок связывания праймера имеет длину, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей

мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, или, по меньшей мере, 20 нуклеотидов.

74. Направляющая РНК по любому из п.п.56-73, отличающаяся тем, что матрица для синтеза ДНК имеет длину, по меньшей мере, 1 нуклеотид, по меньшей мере, 2 нуклеотида, по меньшей мере, 3 нуклеотида, по меньшей мере, 4 нуклеотида, по меньшей мере, 5 нуклеотидов, по меньшей мере, 6 нуклеотидов, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, по меньшей мере, 20 нуклеотидов, по меньшей мере, 21 нуклеотид, по меньшей мере, 22 нуклеотида, по меньшей мере, 23 нуклеотида, по меньшей мере, 24 нуклеотида, по меньшей мере, 25 нуклеотидов, по меньшей мере, 26 нуклеотидов, по меньшей мере, 27 нуклеотидов, по меньшей мере, 28 нуклеотидов, по меньшей мере, 29 нуклеотидов, по меньшей мере, 30 нуклеотидов, по меньшей мере, 31 нуклеотид, по меньшей мере, 32 нуклеотида, по меньшей мере, 33 нуклеотида, по меньшей мере, 34 нуклеотида, по меньшей мере, 35 нуклеотидов, по меньшей мере, 36 нуклеотидов, по меньшей мере, 37 нуклеотидов, по меньшей мере, 38 нуклеотидов, по меньшей мере, 39 нуклеотидов, по меньшей мере, 40 нуклеотидов, или, по меньшей мере, 100 нуклеотидов.

75. Направляющая РНК по любому из п.п.56-74, отличающаяся тем, что матрица для синтеза ДНК может использоваться РНК-зависимой ДНК-полимеразой (например обратной транскриптазой) в качестве матрицы для синтеза соответствующего одноцепочечного ДНК-флэпа, имеющего 3'-конец, где ДНК-флэп комплементарен цепи последовательности эндогенной ДНК-мишени, примыкающей к участку разрыва, и где одноцепочечный ДНК-флэп включает требуемое нуклеотидное изменение, кодируемое матрицей для синтеза ДНК.

76. Направляющая РНК по п.75, отличающаяся тем, что одноцепочечный ДНК-флэп вытесняет эндогенную одноцепочечную ДНК, имеющую 5′-конец, в последовательности-мишени ДНК, которая была никирована.

77. Направляющая РНК по п.76, отличающаяся тем, что эндогенная одноцепочечная ДНК, имеющая свободный 5′-конец, подвергается эксцизии клеткой.

78. Направляющая РНК по п.77, отличающаяся тем, что клеточная репарация одноцепочечного ДНК-флэпа приводит к внесению нуклеотидного изменения, тем самым обеспечивая образование подвергнутого редактированию ДНК-продукта.

79. Направляющая РНК по п.78, отличающаяся тем, что нуклеотидное изменение представляет собой вставку.

80. Направляющая РНК по п.79, отличающаяся тем, что вставка имеет длину, по меньшей мере, 1 нуклеотид, по меньшей мере, 2 нуклеотида, по меньшей мере, 3 нуклеотида, по меньшей мере, 4 нуклеотида, по меньшей мере, 5 нуклеотидов, по меньшей мере, 6 нуклеотидов, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, по меньшей мере, 20 нуклеотидов, по меньшей мере, 21 нуклеотид, по меньшей мере, 22 нуклеотида, по меньшей мере, 23 нуклеотида, по меньшей мере, 24 нуклеотида, по меньшей мере, 25 нуклеотидов, по меньшей мере, 26 нуклеотидов, по меньшей мере, 27 нуклеотидов, по меньшей мере, 28 нуклеотидов, по меньшей мере, 29 нуклеотидов, по меньшей мере, 30 нуклеотидов, по меньшей мере, 31 нуклеотид, по меньшей мере, 32 нуклеотида, по меньшей мере, 33 нуклеотида, по меньшей мере, 34 нуклеотида, по меньшей мере, 35 нуклеотидов, по меньшей мере, 36 нуклеотидов, по меньшей мере, 37 нуклеотидов, по меньшей мере, 38 нуклеотидов, по меньшей мере, 39 нуклеотидов, по меньшей мере, 40 нуклеотидов, или, по меньшей мере, 100 нуклеотидов.

81. Направляющая РНК по п.79, отличающаяся тем, что вставка представляет собой последовательность, кодирующую полипептид.

82. Комплекс для праймерного редактирования, включающий napDNAbp, РНК-зависимую ДНК-полимеразу, и любую одну из направляющих РНК по п.п.56-81.

83. Комплекс для праймерного редактирования по п.82, отличающийся тем, что napDNAbp и РНК-зависимая ДНК-полимераза образуют слитый белок.

84. Комплекс для праймерного редактирования по п.82, отличающийся тем, что napDNAbp представляет собой Cas9.

85. Комплекс для праймерного редактирования по п.84, отличающийся тем, что Cas9 представляет собой никазу Cas9 или ее вариант.

86. Комплекс для праймерного редактирования по п.84, отличающийся тем, что Cas9 имеет аминокислотную последовательность, выбранную из группы, состоящей из SEQ ID NO: 1-135514.

87. Комплекс для праймерного редактирования по п.83, отличающийся тем, что слитый белок имеет аминокислотную последовательность, выбранную из группы, состоящей из SEQ ID NO: 1-135514.

88. Комплекс для праймерного редактирования по п.83, отличающийся тем, что слитый белок включает линкер, соединяющий napDNAbp и РНК-зависимую ДНК-полимеразу.

89. Комплекс для праймерного редактирования по п.88, отличающийся тем, что линкер имеет аминокислотную последовательность, выбранную из группы, состоящей из SEQ ID NO: 1-135514.

90. Один или несколько полинуклеотидов, кодирующих комплекс для праймерного редактирования по любому из п.п.82-89.

91. Вектор, включающий полинуклеотид по п.90 и один или несколько промоторов, которые обеспечивают экспрессию направляющей РНК и слитого белка комплекса для праймерного редактирования.

92. Клетка, включающая вектор по п.91.

93. Клетка, включающая комплекс для праймерного редактирования по любому из п.п.82-89.

94. Фармацевтическая композиция, включающая: (i) направляющую РНК по любому из п.п.56-81, комплекс для праймерного редактирования по п.п.82-89, полинуклеотид по п.90, или вектор по п.91; и (ii) фармацевтически приемлемый наполнитель.

95. Способ внесения нуклеотидного изменения в последовательность нуклеиновой кислоты, при этом способ включает: контактирование последовательности нуклеиновой кислоты с комплексом, включающим слитый белок и направляющую РНК по любому из п.п.56-81, где слитый белок включает napDNAbp и РНК-зависимую ДНК-полимеразу, где направляющая РНК включает спейсер, ядро нРНК, и плечо удлинения, которое включает матрицу для синтеза ДНК и участок связывания праймера, при этом указанная матрица для синтеза ДНК кодирует нуклеотидное изменение, и где спейсер способен отжигаться на не включающей PAM цепи, являющейся проксимальной относительно доступного PAM и протоспейсера, тем самым обеспечивая

(i) никирование последовательности двуцепочечной ДНК на включающей PAM цепи, тем самым обеспечивая создание свободной одноцепочечной ДНК, имеющей 3´-конец;

(ii) гибридизирование 3'-конца свободной одноцепочечной ДНК с направляющей РНК в участке связывания праймера, тем самым обеспечивая праймирование РНК-зависимой ДНК-полимеразы;

(iii) полимеризацию цепи ДНК от 3'-конца ДНК с кодированием из матрицы для синтеза ДНК, тем самым обеспечивая получения одноцепочечного ДНК-

флэпа, удлиненного от 3'-конца ДНК, где флэп включает нуклеотидное изменение;

(iv) замещение цепи эндогенной ДНК, непосредственно снизу примыкающей к участку разрыва на включающей PAM цепи, одноцепочечным ДНК-флэпом, тем самым обеспечивая внесение нуклеотидного изменения в последовательность двуцепочечной ДНК.

96. Способ по п.95, отличающийся тем, что при завершении этапа (v) в клетке клетка осуществляет репарацию не подвергаемой редактированию цепи за счет репарации и/или репликации клеточной ДНК.

97. Способ по п.95, отличающийся тем, что нуклеотидное изменение представляет собой замещение одного нуклеотида, делецию, вставку, или их комбинацию

98. Способ по п.97, отличающийся тем, что замещение одного нуклеотида представляет собой транзицию или трансверсию.

99. Способ по п.97, отличающийся тем, что замещение одного нуклеотида представляет собой (1) замещение G на T, (2) замещение G на A, (3) замещение G на C, (4) замещение T на G, (5) замещение T на A, (6) замещение T на C, (7) замещение C на G, (8) замещение C на T, (9) замещение C на A, (10) замещение A на T, (11) замещение A на G, или (12) замещение A на C.

100. Способ по п.97, отличающийся тем, что замещение одного нуклеотида обеспечивает конвертирование (1) пары оснований G:C на пару оснований T:A, (2) пары оснований G:C на пару оснований A:T, (3) пары оснований G:C на пару оснований C:G, (4) пары оснований T:A на пару оснований G:C, (5) пары оснований T:A на пару оснований A:T, (6) пары оснований T:A на пару оснований C:G, (7) пары оснований C:G на пару оснований G:C, (8) пары оснований C:G на пару оснований T:A, (9) пары оснований C:G на пару оснований A:T, (10) пары оснований A:T на пару оснований T:A, (11) пары оснований A:T на пару оснований G:C, или (12) пары оснований A:T на пару оснований C:G.

101. Способ по п.97, отличающийся тем, что нуклеотидное изменение представляет собой вставку или делецию 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, или 25 нуклеотидов.

102. Способ по п.97, отличающийся тем, что нуклеотидное изменение представляет собой вставку кодирующей полипептид последовательности.

103. Способ по п.97, отличающийся тем, что нуклеотидное изменение обеспечивает корректирование связанного с заболеванием гена.

104. Способ по п.103, отличающийся тем, что связанный с заболеванием ген связан с моногенетическим заболеванием, выбранным из группы, состоящей из: дефицита аденозиндезаминазы (ADA); дефицита альфа-1 антитрипсина; муковисцидоза; мышечной дистрофии Дюшенна; галактоземии; гемохроматоза; болезни Хантингтона; болезни кленового сиропа; синдрома Марфана; нейрофиброматоза 1 типа; врожденной пахионихии; фенилкеотнурии; тяжелого комбинированного иммунодефицита; серповидноклеточной анемии; синдрома Смита-Лемли-Опица; и болезни Тея-Сакса.

105. Способ по п.103, отличающийся тем, что связанный с заболеванием ген связан с полигенным заболеванием, выбранным из группы, состоящей из: болезни сердца, высокого кровяного давления, болезни Альцгеймера, артрита, диабета, рака, и ожирения.

106. Направляющая РНК для применения в рамках праймерного редактирования для изменения нуклеотидной последовательности молекулы-мишени ДНК с помощью вставки, делеции, инверсии, замещения, или их комбинации для получения соответствующей подвергнутой редактированию молекулы ДНК, где:

(i) направляющая РНК способна образовывать комплекс со слитым белком, включающим napDNAbp и домен, включающий активность РНК-зависимой ДНК-полимеразы;

(ii) направляющая РНК включает (a) спейсер, который способен отжигаться на не включающей PAM цепи, являющейся проксимальной относительно доступного PAM и протоспейсера на включающей PAM цепи в молекуле-мишени ДНК, и (b) ядро нРНК;

(iii) направляющая РНК дополнительно включает плечо удлинения на 5'- или 3'-конце направляющей РНК;

(iv) плечо удлинения включает (a) участок связывания праймера и (b) матрицу для синтеза ДНК, где матрица для синтеза ДНК кодирует одноцепочечный ДНК-флэп, который включает изменение для внесения вместо эндогенной цепи непосредственно ниже участка разрыва на включающей PAM цепи;

(v) молекула-мишень ДНК выбрана из группы, состоящей из SEQ ID NO: 1217353-1289387; и

(vi) соответствующая подвергнутая редактированию молекула ДНК выбрана из группы, состоящей из SEQ ID NO: 1289388-1361420.

107. Направляющая РНК по п.106, отличающаяся тем, что молекула-мишень ДНК представляет собой последовательность варианта Clinvar.

108. Направляющая РНК по п.106, отличающаяся тем, что napDNAbp представляет собой Cas9, Cas12e, Cas12d, Cas12a, Cas12b1, Cas13a, Cas12c, или Argonaute, или вариант Cas9, Cas12e, Cas12d, Cas12a, Cas12b1, Cas13a, Cas12c, или Argonaute.

109. Направляющая РНК по п.106, отличающаяся тем, что домен napDNAbp включает никазную активность.

110. Направляющая РНК по п.106, отличающаяся тем, что napDNAbp представляет собой Cas9 или его вариант.

111. Направляющая РНК по п.106, отличающаяся тем, что napDNAbp представляет собой имеющий нуклеазную активность Cas9, не имеющий нуклеазной активности Cas9 (dCas9), или никазу Cas9 (nCas9).

112. Направляющая РНК по п.106, отличающаяся тем, что napDNAbp представляет собой никазу Cas9 (nCas9).

113. Направляющая РНК по п.106, отличающаяся тем, что napDNAbp включает аминокислотную

114. Направляющая РНК по п.106, отличающаяся тем, что napDNAbp представляет собой SpCas9 дикого типа или его вариант согласно любой одной из аминокислотных последовательностей 1361421-1361428 или аминокислотную последовательность, имеющую, по меньшей мере, 80% идентичность последовательности относительно любой из SEQ ID NO: 1361421-1361428.

115. Направляющая РНК по п.106, отличающаяся тем, что napDNAbp представляет собой ортолог SpCas9 согласно любой одной из аминокислотных последовательностей 1361429-1361442 или аминокислотную последовательность, имеющую, по меньшей мере, 80% идентичность последовательности относительно любой из SEQ ID NO: 1361429-1361442.

116. Направляющая РНК по п.106, отличающаяся тем, что napDNAbp представляет собой любую одну из аминокислотных последовательностей 1361421-1361484 или аминокислотную последовательность, имеющую, по меньшей мере, 80% идентичность последовательности относительно любой из SEQ ID NO: 1361421-1361484.

117. Направляющая РНК по п. 106, отличающаяся тем, что домен, включающий активность РНК-зависимой ДНК-полимеразы, представляет собой обратную транскриптазу.

118. Направляющая РНК по п.117, отличающаяся тем, что обратная транскриптаза представляет собой встречающуюся в природе обратную транскриптазу дикого типа, имеющую аминокислотную последовательность любой одной из SEQ ID NO: 1361485-1361496 или аминокислотную последовательность, имеющую, по меньшей мере, 80% идентичность последовательности относительно любой из SEQ ID NO: 1361485-1361496.

119. Направляющая РНК по п.117, отличающаяся тем, что обратная транскриптаза представляет собой вариант обратной транскриптазы, имеющий аминокислотную последовательность любой одной из SEQ ID NO: 1361497-1361514 или аминокислотную последовательность, имеющую, по меньшей мере, 80% идентичность последовательности относительно любой из SEQ ID NO: 1361497-1361514.

120. Направляющая РНК по п.106, отличающаяся тем, что слитый белок включает аминокислотную последовательность любой одной из SEQ ID NO: 1361515-1361519 или аминокислотную последовательность, имеющую, по меньшей мере, 80% идентичность последовательности относительно любой из SEQ ID NO: 1361515-1361519.

121. Направляющая РНК по п.106, отличающаяся тем, что слитый белок включает аминокислотную последовательность SEQ ID NO: 1361515 (PE1) или 1361516 (PE2), или аминокислотную последовательность, имеющую, по меньшей мере, 80% идентичность последовательности относительно любой из SEQ ID NO: 1361515 или 1361516.

122. Направляющая РНК по п.106, отличающаяся тем, что доступная последовательность PAM зависит от napDNAbp, использованного на этапе (i).

123. Направляющая РНК по п.106, отличающаяся тем, что доступная последовательность PAM выбрана из группы, состоящей из: (a) 5′-NGG-3′ (стандартная последовательность PAM), (b) 5′-NNG-3′, (c) 5′-NNA-3′, (d) 5′-NNC-3′, (e) 5′-NNT-3′, (f) 5′-NGT-3′, (g) 5′-NGA-3′, (h) 5′-NGC-3′, (i) 5′-NAA-3′, (j) 5′-NAC-3′, (k) 5′-NAG-3′, и (l) 5′-NAT-3′, выбор которой зависит от выбора napDNAbp.

124. Направляющая РНК по п.106, отличающаяся тем, что участок изменения в любой из нуклеотидных последовательностей SEQ ID NO: 1217353-1289387 этапа (v) начинается в положении 201 в направлении от 5' к 3'.

125. Направляющая РНК по п.106, отличающаяся тем, что нуклеотидное изменение представляет собой нуклеотидное замещение, делецию, вставку, или их комбинацию.

126. Направляющая РНК по п.106, отличающаяся тем, что нуклеотидное замещение представляет собой транзицию или трансверсию.

127. Направляющая РНК по п.106, отличающаяся тем, что замещение одного нуклеотида представляет собой (1) замещение G на T, (2) замещение G на A, (3) замещение G на C, (4) замещение T на G, (5) замещение T на A, (6) замещение T на C, (7) замещение C на G, (8) замещение C на T, (9) замещение C на A, (10) замещение A на T, (11) замещение A на G, или (12) замещение A на C.

128. Направляющая РНК по п.106, отличающаяся тем, что замещение одного нуклеотида обеспечивает конвертирование (1) пары оснований G:C на пару оснований T:A, (2) пары оснований G:C на пару оснований A:T, (3) пары оснований G:C на пару оснований C:G, (4) пары оснований T:A на пару оснований G:C, (5) пары оснований T:A на пару оснований A:T, (6) пары оснований T:A на пару оснований C:G, (7) пары оснований C:G на пару оснований G:C, (8) пары оснований C:G на пару оснований T:A, (9) пары оснований C:G на пару оснований A:T, (10) пары оснований A:T на пару оснований T:A, (11) пары оснований A:T на пару оснований G:C, или (12) пары оснований A:T на пару оснований C:G.

129. Направляющая РНК по п.106, отличающаяся тем, что требуемое нуклеотидное изменение представляет собой вставку или делецию 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, или 25 нуклеотидов.

130. Направляющая РНК по п.106, отличающаяся тем, что нуклеотидное изменение представляет собой вставку кодирующей полипептид последовательности.

131. Направляющая РНК по п.106, отличающаяся тем, что нуклеотидное изменение обеспечивает корректирование связанного с заболеванием гена.

132. Направляющая РКН по п.131, отличающаяся тем, что связанный с заболеванием ген связан с моногенетическим заболеванием, выбранным из группы, состоящей из: дефицита аденозиндезаминазы (ADA); дефицита альфа-1 антитрипсина; муковисцидоза; мышечной дистрофии Дюшенна; галактоземии; гемохроматоза; болезни Хантингтона; болезни кленового сиропа; синдрома Марфана; нейрофиброматоза 1 типа; врожденной пахионихии; фенилкеотнурии; тяжелого комбинированного иммунодефицита; серповидноклеточной анемии; синдрома Смита-Лемли-Опица; и болезни Тея-Сакса.

133. Направляющая РНК по п.131, отличающаяся тем, что связанный с заболеванием ген связан с полигенным заболеванием, выбранным из группы, состоящей из: болезни сердца, высокого кровяного давления, болезни Альцгеймера, артрита, диабета, рака, и ожирения.

134. Способ внесения нуклеотидного изменения в последовательность нуклеиновой кислоты, при этом способ включает: контактирование последовательности нуклеиновой кислоты с комплексом, включающим слитый белок и направляющую РНК по любому из п.п.56-81.

135. Способ по п.134, отличающийся тем, что слитый белок включает napDNAbp и РНК-зависимую ДНК-полимеразу.

136. Способ по п.134, отличающийся тем, что направляющая РНК включает спейсер, ядро нРНК, и плечо удлинения, которое включает матрицу для синтеза ДНК и участок связывания праймера.

137. Способ по п.136, отличающийся тем, что матрица для синтеза ДНК кодирует нуклеотидное изменение.

138. Способ по любому из п.п.134-137, отличающийся тем, что направляющая РНК способна связываться с napDNAbp, подходящим для праймерного редактирования, и направлять napDNAbp к последовательности-мишени ДНК.

139. Способ по п.138, отличающийся тем, что последовательность-мишень нуклеиновой кислоты включает цепь-мишень (или включающую PAM или подвергаемую редактированию цепь) и комплементарную не являющуюся мишенью цепь (или не включающую PAM или не подвергаемую редактированию цепь), где спейсер направляющей РНК гибридизируется с не включающей PAM цепью для образования гибрида РНК-ДНК и R-петли.

140. Способ по п.136, отличающийся тем, что участок связывания праймера имеет длину от приблизительно 8 до приблизительно 20 нуклеотидов.

141. Способ по п.136, отличающийся тем, что участок связывания праймера имеет длину 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, или 20 нуклеотидов.

142. Способ по п.136, отличающийся тем, что плечо удлинения имеет длину, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, по меньшей мере, 20 нуклеотидов, по меньшей мере, 21 нуклеотид, по меньшей мере, 22 нуклеотида, по меньшей мере, 23 нуклеотида, по меньшей мере, 24 нуклеотида, по меньшей мере, 25 нуклеотидов, по меньшей мере, 26 нуклеотидов, по меньшей мере, 27 нуклеотидов, по меньшей мере, 28 нуклеотидов, по меньшей мере, 29 нуклеотидов, по меньшей мере, 30 нуклеотидов, по меньшей мере, 31 нуклеотид, по меньшей мере, 32 нуклеотида, по меньшей мере, 33 нуклеотида, по меньшей мере, 34 нуклеотида, по меньшей мере, 35 нуклеотидов, по меньшей мере, 36 нуклеотидов, по меньшей мере, 37 нуклеотидов, по меньшей мере, 38 нуклеотидов, по меньшей мере, 39 нуклеотидов, по меньшей мере, 40 нуклеотидов, или, по меньшей мере, 100 нуклеотидов.

143. Способ по п.136, отличающийся тем, что участок связывания праймера имеет длину, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, или, по меньшей мере, 20 нуклеотидов.

144. Способ по п.136, отличающийся тем, что матрица для синтеза ДНК имеет длину, по меньшей мере, 1 нуклеотид, по меньшей мере, 2 нуклеотида, по меньшей мере, 3 нуклеотида, по меньшей мере, 4 нуклеотида, по меньшей мере, 5 нуклеотидов, по

меньшей мере, 6 нуклеотидов, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, по меньшей мере, 20 нуклеотидов, по меньшей мере, 21 нуклеотид, по меньшей мере, 22 нуклеотида, по меньшей мере, 23 нуклеотида, по меньшей мере, 24 нуклеотида, по меньшей мере, 25 нуклеотидов, по меньшей мере, 26 нуклеотидов, по меньшей мере, 27 нуклеотидов, по меньшей мере, 28 нуклеотидов, по меньшей мере, 29 нуклеотидов, по меньшей мере, 30 нуклеотидов, по меньшей мере, 31 нуклеотид, по меньшей мере, 32 нуклеотида, по меньшей мере, 33 нуклеотида, по меньшей мере, 34 нуклеотида, по меньшей мере, 35 нуклеотидов, по меньшей мере, 36 нуклеотидов, по меньшей мере, 37 нуклеотидов, по меньшей мере, 38 нуклеотидов, по меньшей мере, 39 нуклеотидов, по меньшей мере, 40 нуклеотидов, или, по меньшей мере, 100 нуклеотидов.

145. Способ по п.136, отличающийся тем, что матрица для синтеза ДНК может использоваться РНК-зависимой ДНК-полимеразой (например обратной транскриптазой) в качестве матрицы для синтеза соответствующего одноцепочечного ДНК-флэпа, имеющего 3'-конец, где ДНК-флэп комплементарен цепи последовательности эндогенной ДНК-мишени, примыкающей к участку разрыва, и где одноцепочечный ДНК-флэп включает требуемое нуклеотидное изменение, кодируемое матрицей для синтеза ДНК.

146. Способ по п.145, отличающийся тем, что одноцепочечный ДНК-флэп вытесняет эндогенную одноцепочечную ДНК, имеющую 5′-конец, в последовательности-мишени ДНК, которая была никирована.

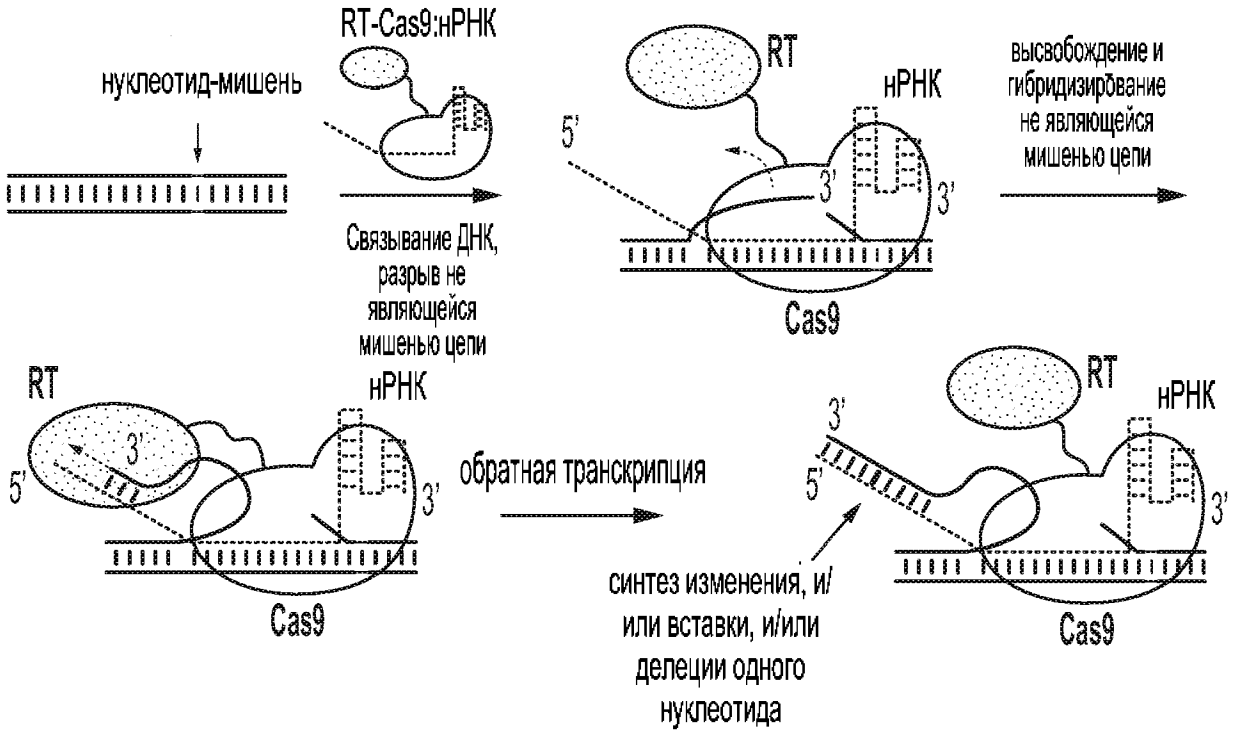
147. Способ по п.146, отличающийся тем, что эндогенная одноцепочечная ДНК, имеющая свободный 5′-конец, подвергается эксцизии клеткой.

148. Способ по п.145, отличающийся тем, что клеточная репарация одноцепочечного ДНК-флэпа приводит к внесению нуклеотидного изменения, тем самым обеспечивая образование подвергнутого редактированию ДНК-продукта.

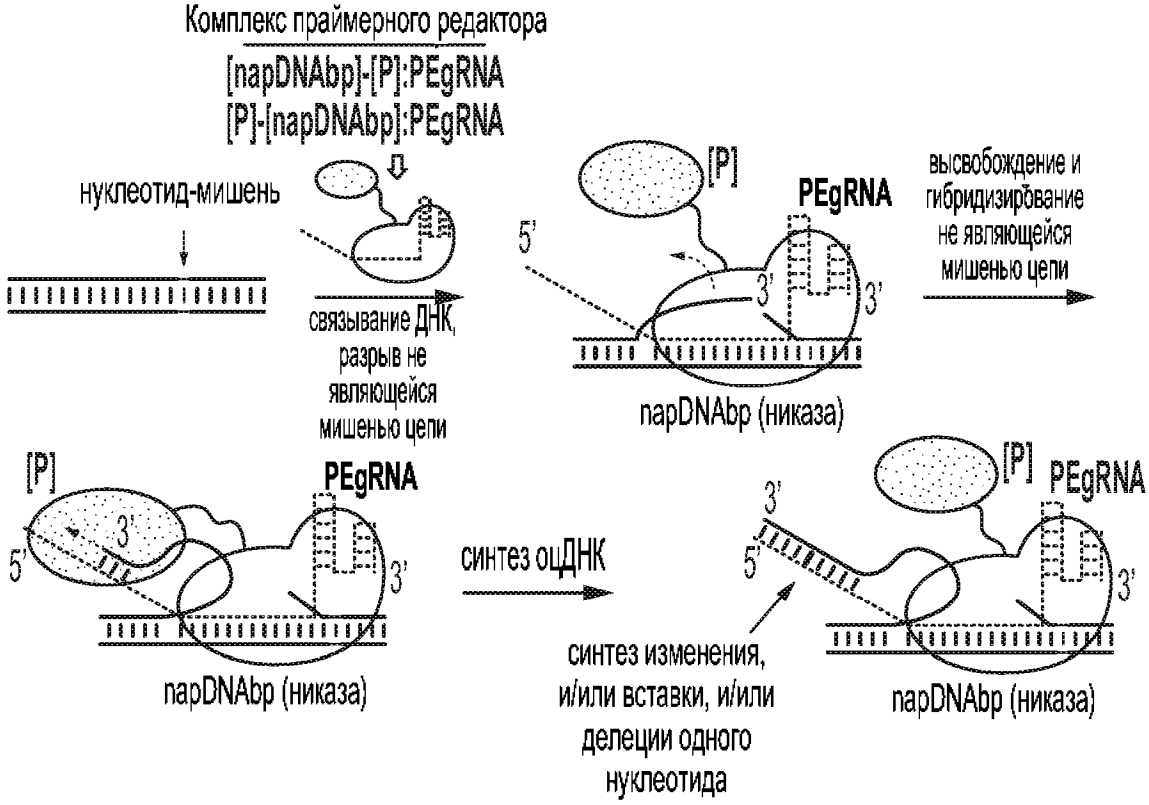
149. Способ по п.148, отличающийся тем, что нуклеотидное изменение представляет собой вставку.

150. Способ по п.149, отличающийся тем, что вставка имеет длину, по меньшей мере, 1 нуклеотид, по меньшей мере, 2 нуклеотида, по меньшей мере, 3 нуклеотида, по меньшей мере, 4 нуклеотида, по меньшей мере, 5 нуклеотидов, по меньшей мере, 6 нуклеотидов, по меньшей мере, 7 нуклеотидов, по меньшей мере, 8 нуклеотидов, по меньшей мере, 9 нуклеотидов, по меньшей мере, 10 нуклеотидов, по меньшей мере, 11 нуклеотидов, по меньшей мере, 12 нуклеотидов, по меньшей мере, 13 нуклеотидов, по меньшей мере, 14 нуклеотидов, по меньшей мере, 15 нуклеотидов, по меньшей мере, 16 нуклеотидов, по меньшей мере, 17 нуклеотидов, по меньшей мере, 18 нуклеотидов, по меньшей мере, 19 нуклеотидов, по меньшей мере, 20 нуклеотидов, по меньшей мере, 21 нуклеотид, по меньшей мере, 22 нуклеотида, по меньшей мере, 23 нуклеотида, по меньшей мере, 24 нуклеотида, по меньшей мере, 25 нуклеотидов, по меньшей мере, 26 нуклеотидов, по меньшей мере, 27 нуклеотидов, по меньшей мере, 28 нуклеотидов, по меньшей мере, 29 нуклеотидов, по меньшей мере, 30 нуклеотидов, по меньшей мере, 31 нуклеотид, по меньшей мере, 32 нуклеотида, по меньшей мере, 33 нуклеотида, по меньшей мере, 34 нуклеотида, по меньшей мере, 35 нуклеотидов, по меньшей мере, 36 нуклеотидов, по меньшей мере, 37 нуклеотидов, по меньшей мере, 38 нуклеотидов, по меньшей мере, 39 нуклеотидов, по меньшей мере, 40 нуклеотидов, или, по меньшей мере, 100 нуклеотидов.

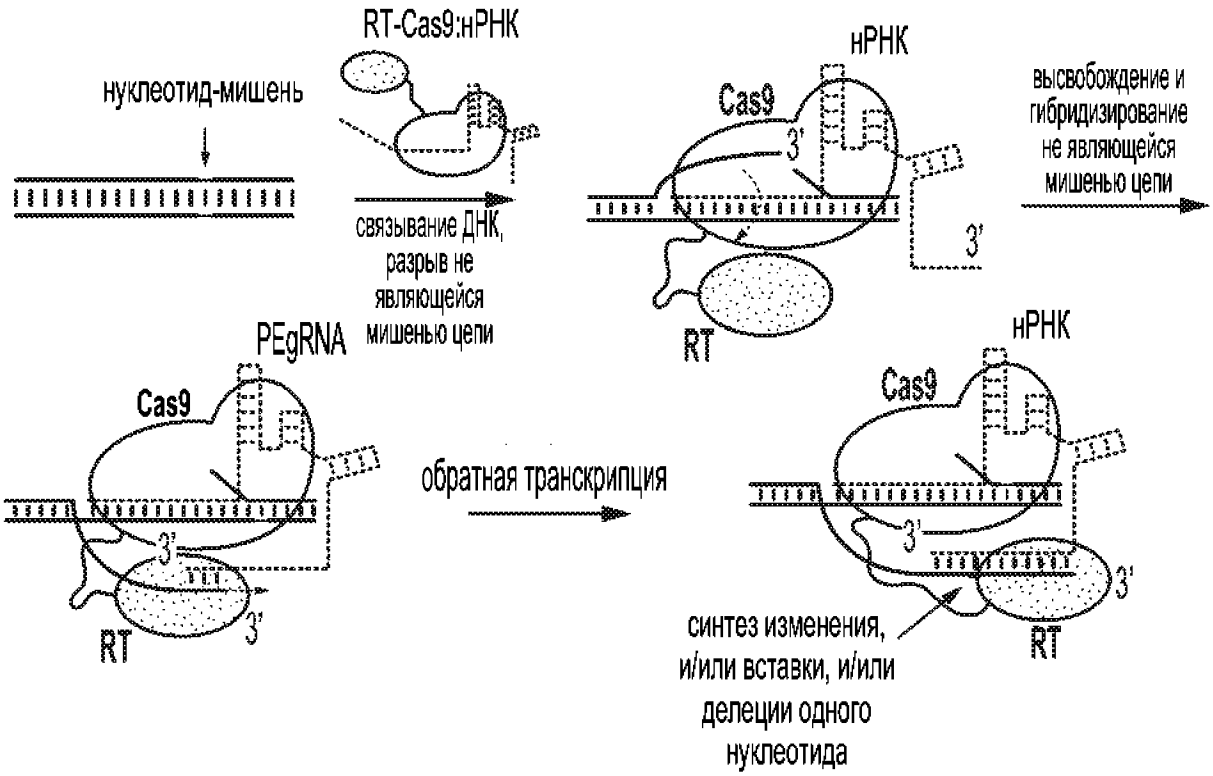
151. Способ по п.149, отличающийся тем, что вставка представляет собой последовательность, кодирующую полипептид.


нуклеотид-мишень
RT-Cas9:нРНК
Связывание ДНК, разрыв не являющейся мишенью цепи
RT
нРНК
5'
3'
3'
Cas9
высвобождение и гибридизирование не являющейся мишенью цепи
RT
нРНК
3'
5'
3'
Cas9
обратная транскрипция
RT
нРНК
3'
5'
3'
Cas9
синтез изменения, и/или вставки, и/или делеции одного нуклеотида

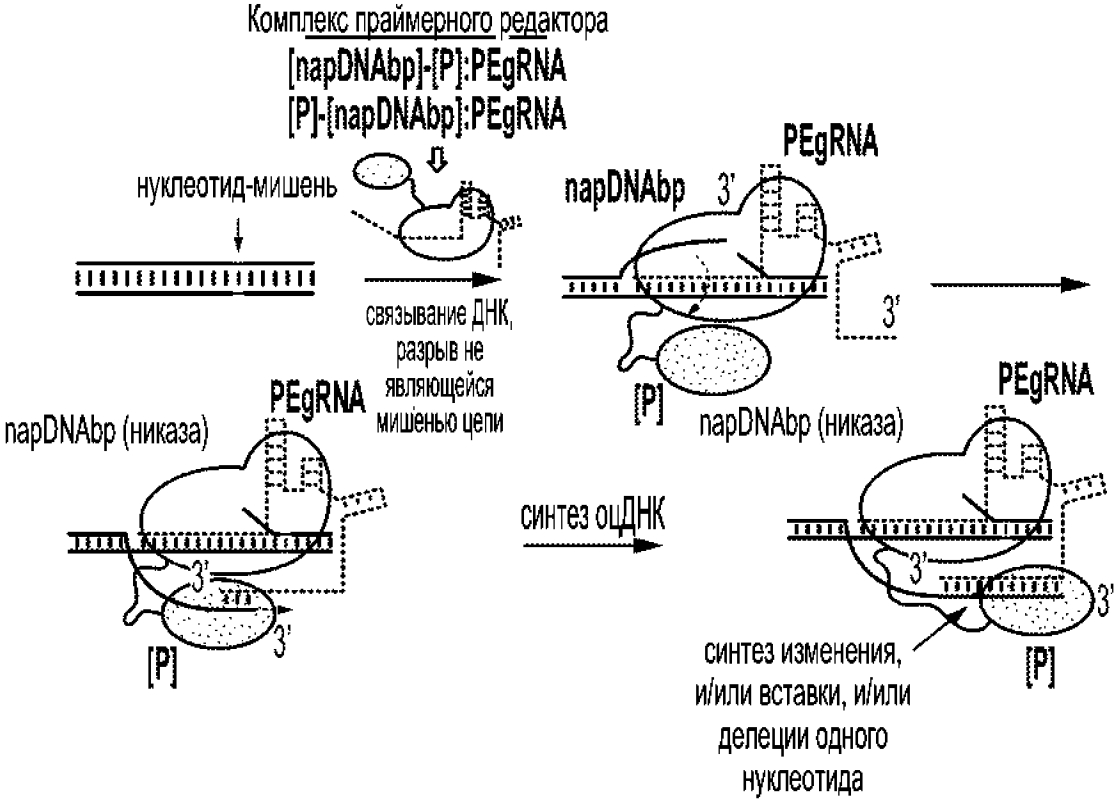
ФИГ.1А.1


Комплекс праймерного редактора
[napDNAbp]-[P]:PEgRNA
[P]-[napDNAbp]:PEgRNA
нуклеотид-мишень
связывание ДНК, разрыв не являющейся мишенью цепи
[P]
PEgRNA
5'
3'
3'
napDNAbp (никаза)
высвобождение и гибридизирбвание не являющейся мишенью цепи
[P]
PEgRNA
3'
5'
3'
napDNAbp (никаза)
синтез оцДНК
3'
5'
[P]
PEgRNA
3'
синтез изменения, и/или вставки, и/или делеции одного нуклеотида
napDNAbp (никаза)

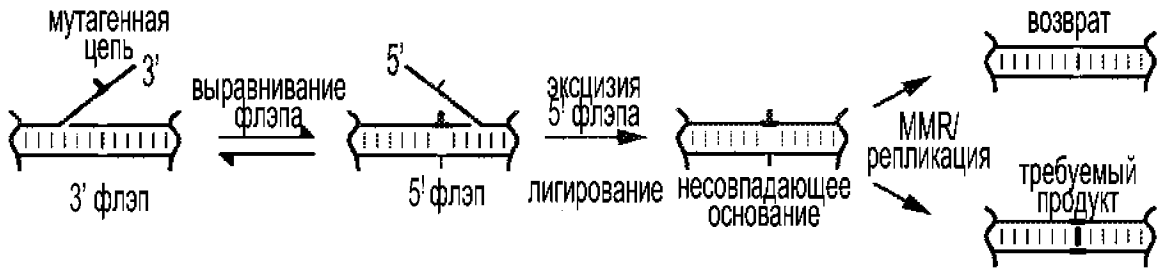
ФИГ.1А.2


нуклеотид-мишень
RT-Cas9:нРНК
связывание ДНК, разрыв не являющейся мишенью цепи
нРНК
Cas9
3'
3'
RT
высвобождение и гибридизирование не являющейся мишенью цепи
PEgRNA
Cas9
3'
3'
RT
обратная транскрипция
нРНК
Cas9
3'
3'
RT
синтез изменения, и/или вставки, и/или делеции одного нуклеотида

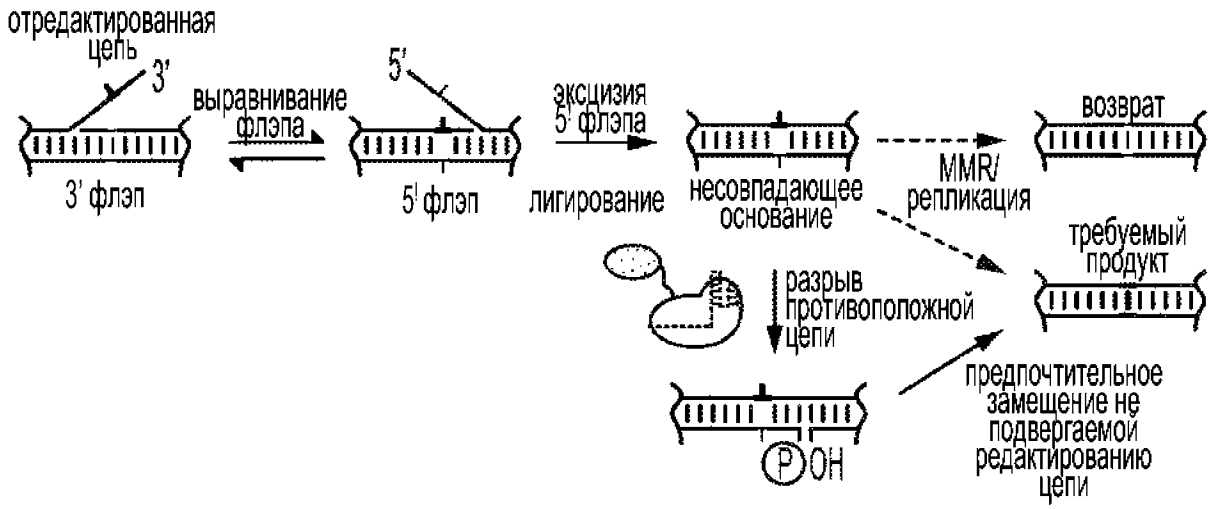
ФИГ.1В.1


Комплекс праймерного редактора
[napDNAbp]-[P]:PEgRNA
[P]-[napDNAbp]:PEgRNA
нуклеотид-мишень
связывание ДНК, разрыв не являющейся мишенью цепи
napDNAbp
PEgRNA
3'
[P]
napDNAbp (никаза)
синтез оцДНК
синтез изменения, и/или вставки, и/или делеции одного нуклеотида

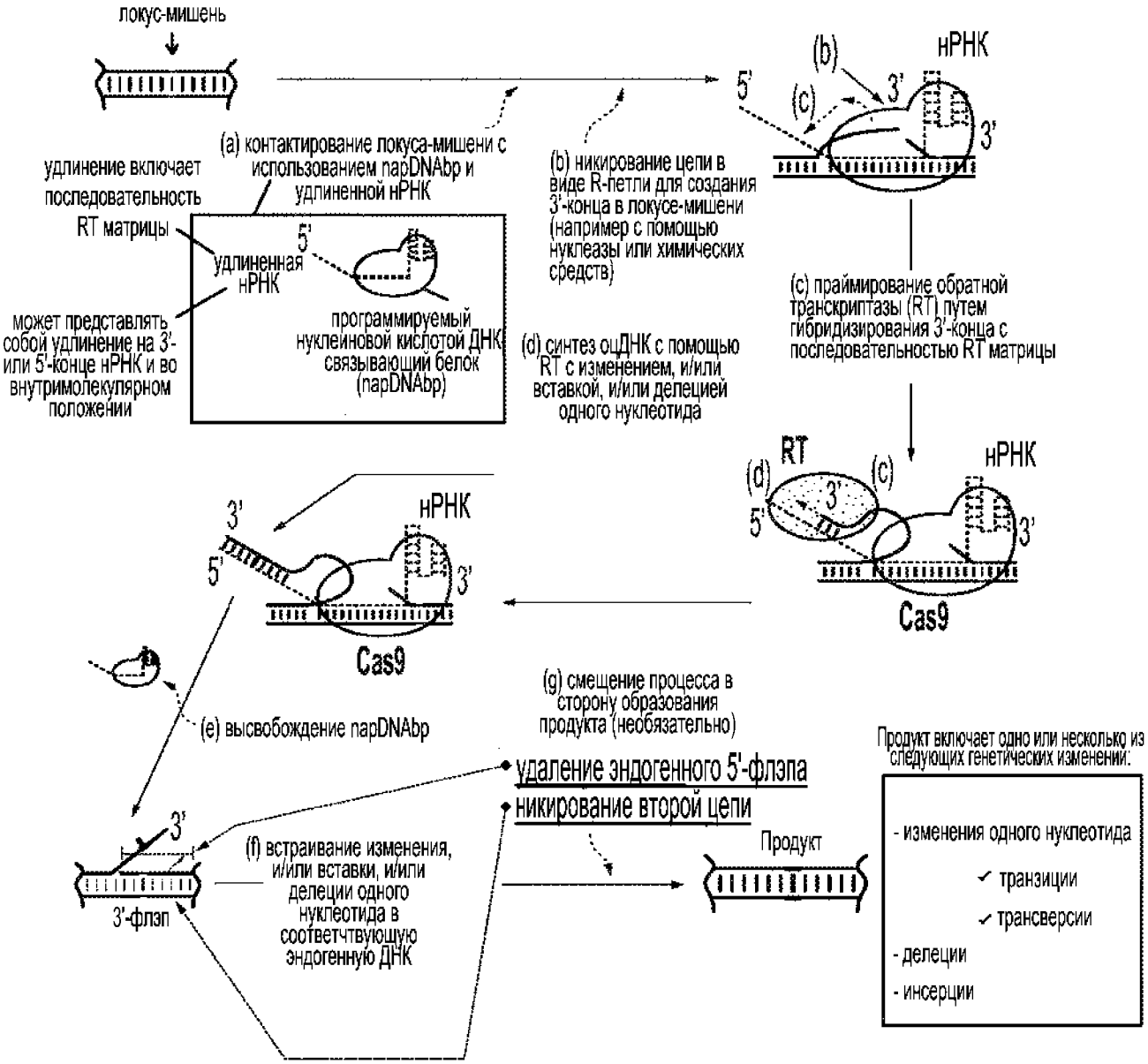
ФИГ.1В.2


мутагенная цепь
3'
выравнивание флэпа
5'
эксцизия 5' флэпа
возврат
MMR/ репликация
3' флэп
5' флэп
лигирование
несовпадающее основание
требуемый продукт

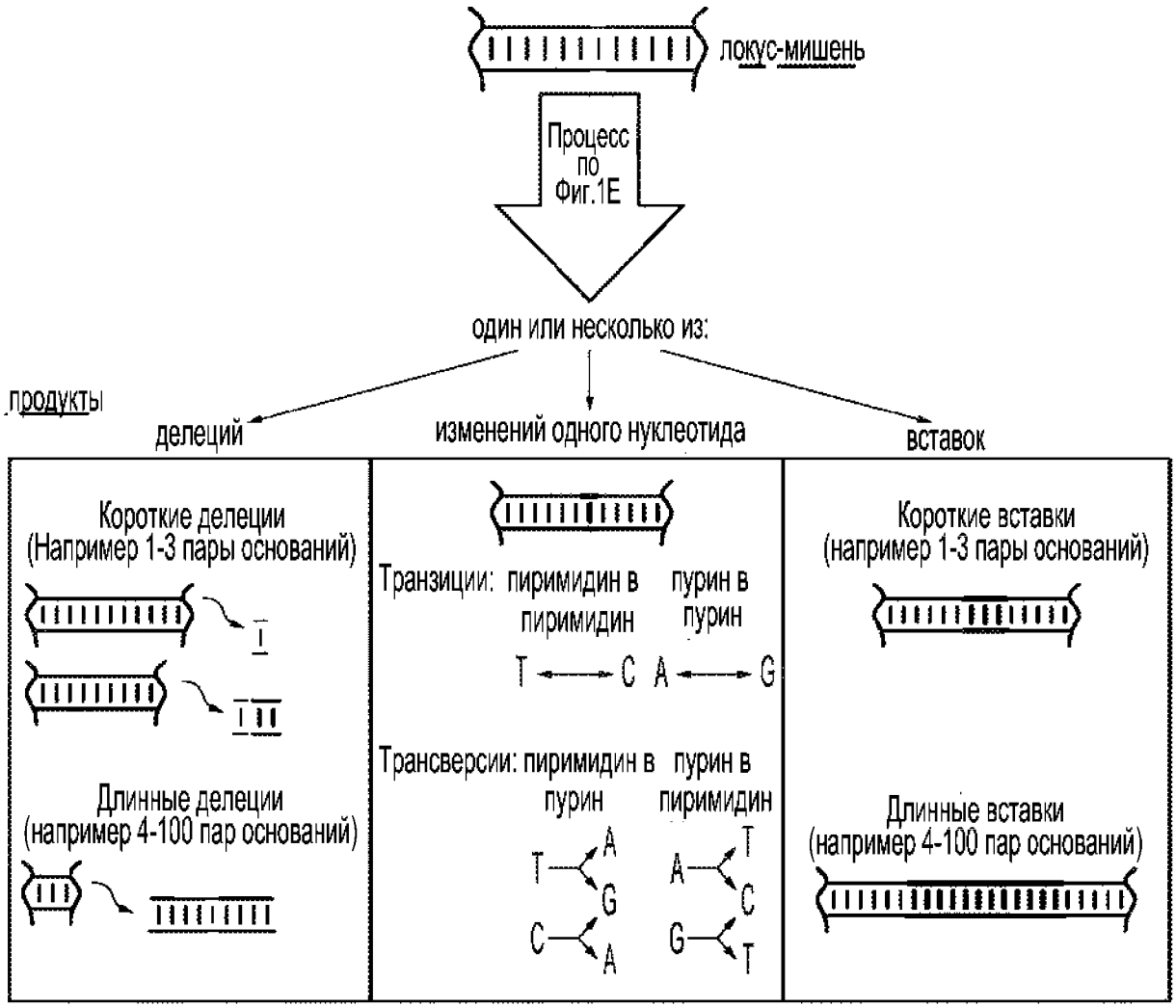
ФИГ.1C


отредактированная цепь
3'
выравнивание флэпа
5'
эксцизия 5' флэпа
возврат
MMR/ репликация
3' флэп
5' флэп
лигирование
несовпадающее основание
требуемый продукт
разрыв противоположной цепи
предпочтительное замещение не подвергаемой редактированию цепи
P
OH

ФИГ.1D

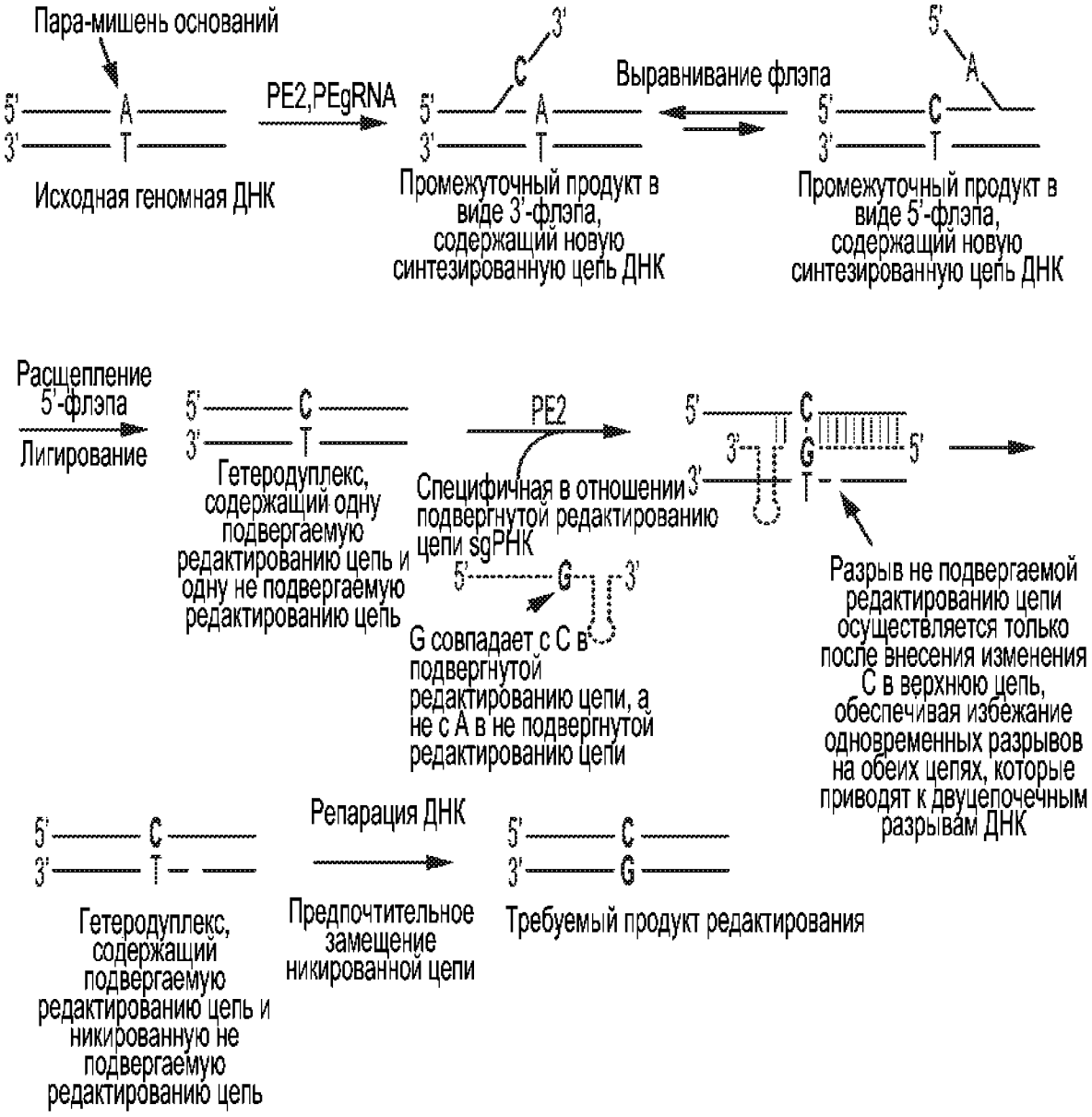

локус-мишень
(a) контактирование локуса-мишени с использованием napDNAbp и удлиненной нРНК
удлинение включает последовательность RT матрицы
может представлять собой удлинение на 3'- или 5'-конце нРНК и во внутримолекулярном положении
5'
удлиненная нРНК
программируемый нуклеиновой кислотой ДНК связывающий белок (napDNAbp)
(b) никирование цепи в виде R-петли для создания 3'-конца в локусе-мишени (например с помощью нуклеазы или химических средств)
(b)
нРНК
5'
(c)
3'
3'
(c) праймирование обратной транскриптазы (RT) путем гибридизирования 3'-конца с последовательностью RT матрицы
(d) синтез оцДНК с помощью RT с изменением, и/или вставкой, и/или делецией одного нуклеотида
RT
(d)
(c)
нРНК
3'
5'
3'
Cas9
3'
нРНК
5'
3'
Cas9
(e) высвобождение napDNAbp
(g) смещение процесса в сторону образования продукта (необязательно)
удаление эндогенного 5'-флэпа
никирование второй цепи
Продукт
Продукт включает одно или несколько из следующих генетических изменений:
- изменения одного нуклеотида
✓ транзиции
✓ трансверсии
- делеции
- инсерции
3'
3'-флэп
(f) встраивание изменения, и/или вставки, и/или делеции одного нуклеотида в соответствующую эндогенную ДНК

ФИГ.1E


локус-мишень
Процесс по Фиг.1E
один или несколько из:
продукты
делеций
изменений одного нуклеотида
вставок
Короткие делеции
(Например 1-3 пары оснований)
Длинные делеции
(например 4-100 пар оснований)
Транзиции: пиримидин в пиримидин
пурин в пурин
T ⟷ C A ⟷ G
Трансверсии: пиримидин в пурин
пурин в пиримидин
T A G C A T C G T
Короткие вставки
(например 1-3 пары оснований)
Длинные вставки
(например 4-100 пар оснований)

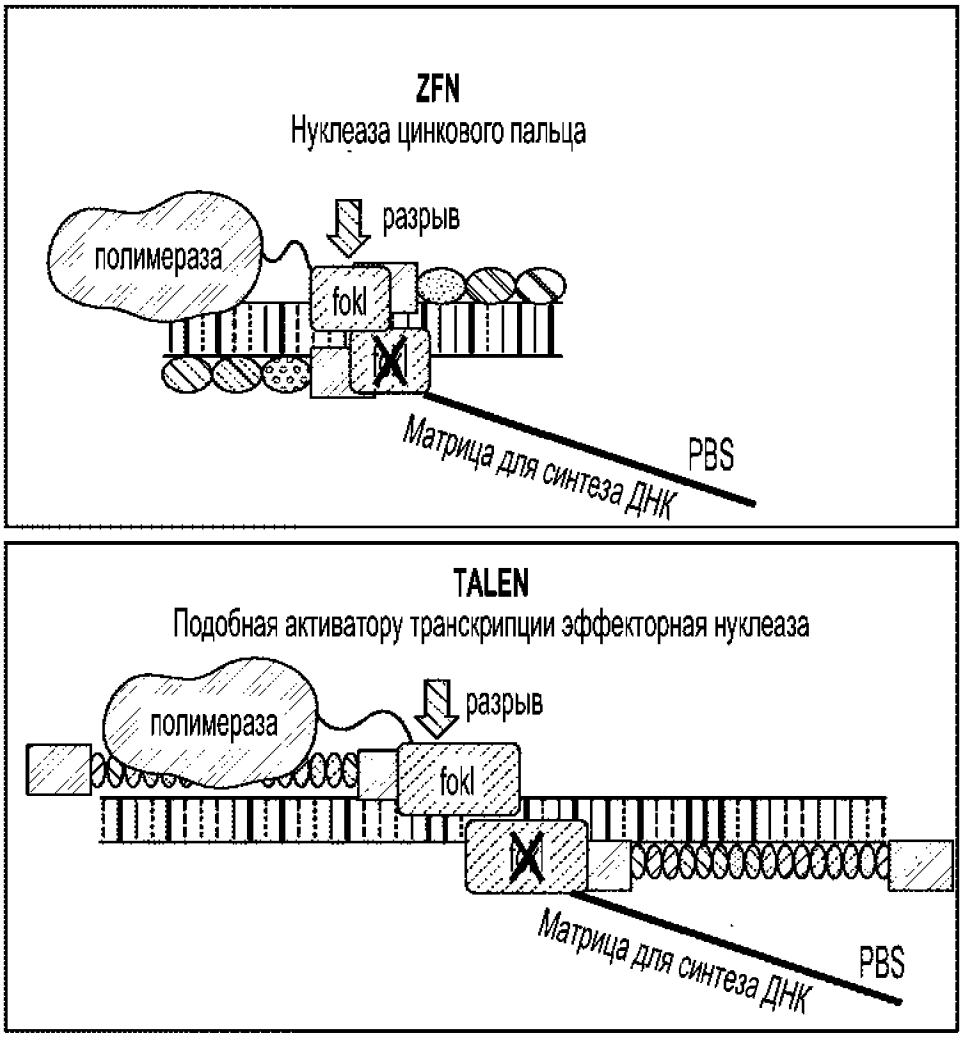
ФИГ.1F

Система редактирования PE3b: специфическое в плане редактирования никирование комплементарной (не подвергнутой редактированию) цепи ДНК



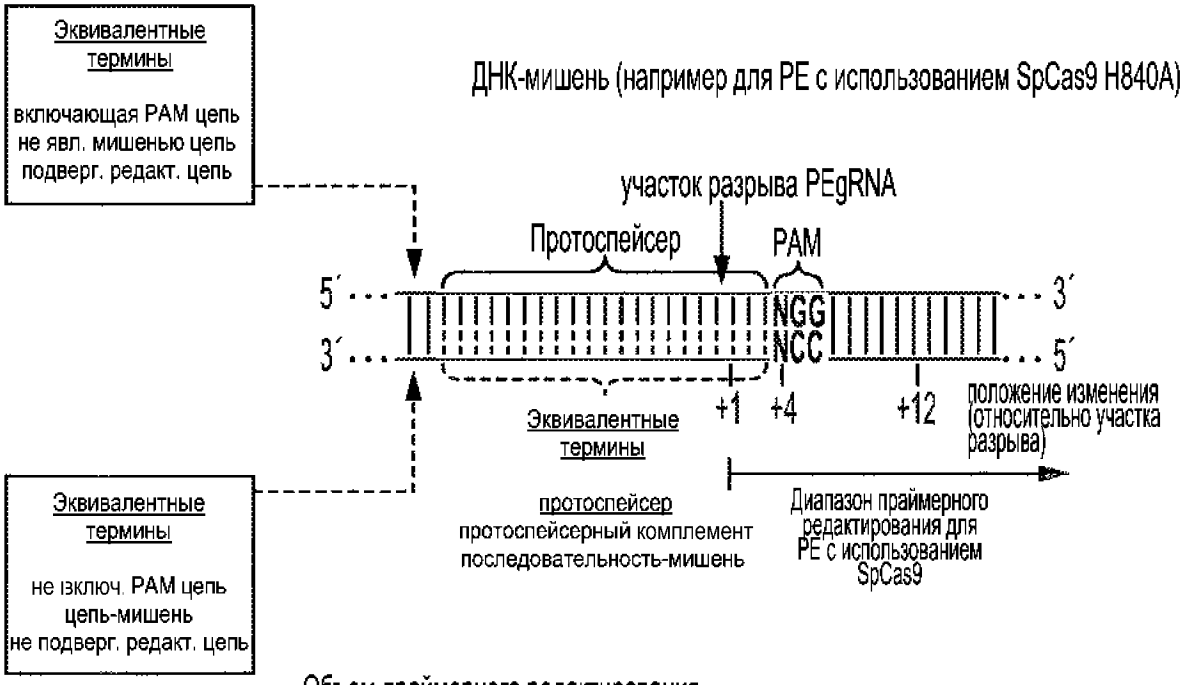
ФИГ.1G

Праймерное редактирование с использованием любой программируемой нуклеазы



Зависимый от полимеразы синтез замещающей цепи, кодируемой полимеразой с использованием матрицы для синтеза ДНК, которая содержит требуемое изменение. Предполагается, что механизм работает аналогично праймерному редактированию. Дополнительные факторы или эффекторы могут быть добавлены in trans или в виде продуктов слияния для обеспечения реакции в целом (например (a) хеликаза для раскручивания ДКН в участке разрыва для получения цепи с разрывом с 3'-концом, доступным в качестве праймера, (b) FEN1 для осуществления удаления эндогенной цепи на цепи с разрывом для направления реакции к замещению эндогенной цепи синтезированной цепью, или (c) комплекс nCas9:нРНК для создания второго участка разрыва, который может обеспечивать встраивание синтезированной коррекции за счет предпочтительной клеточной репарации не подвергаемой редактированию цепи)

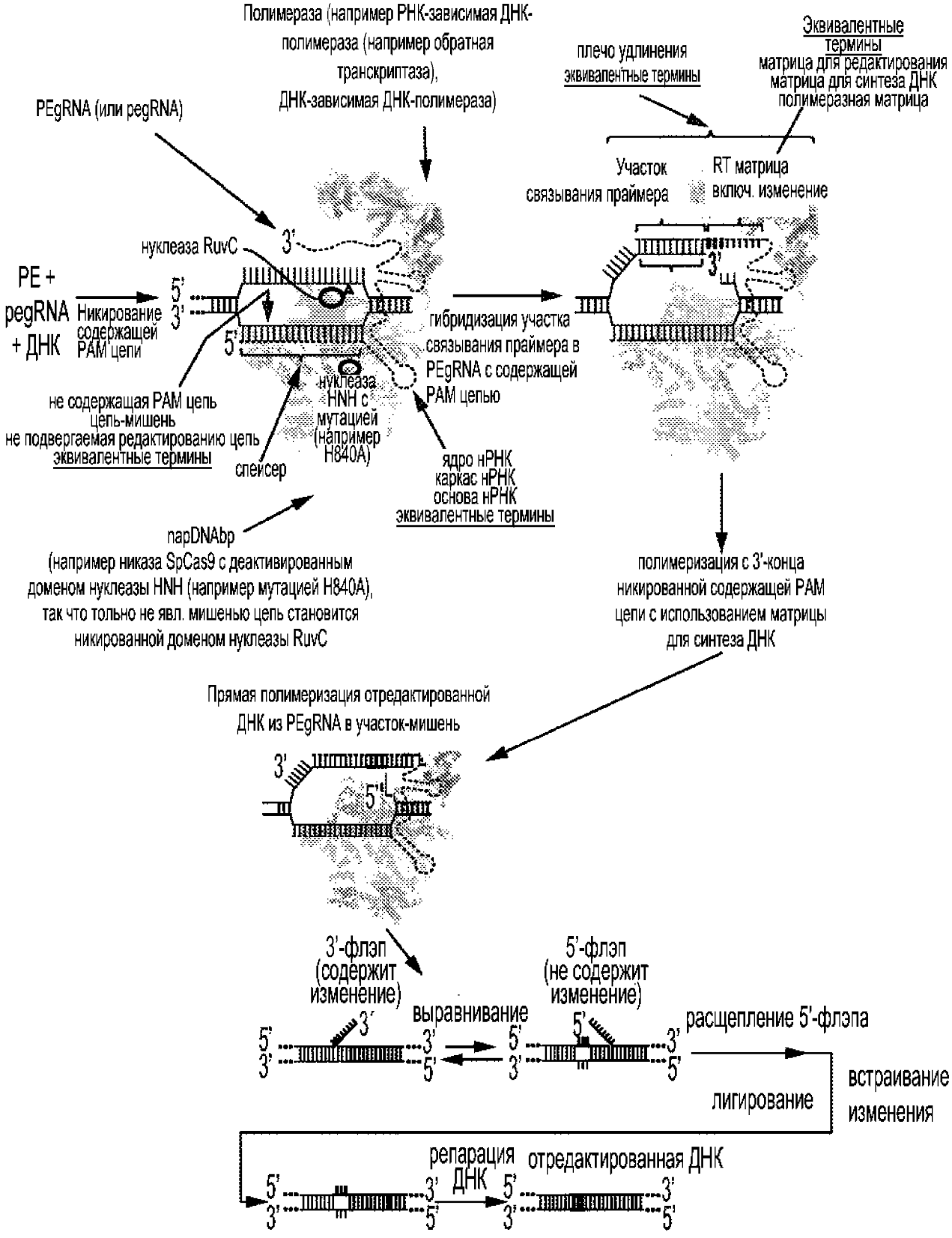
ФИГ.1Н



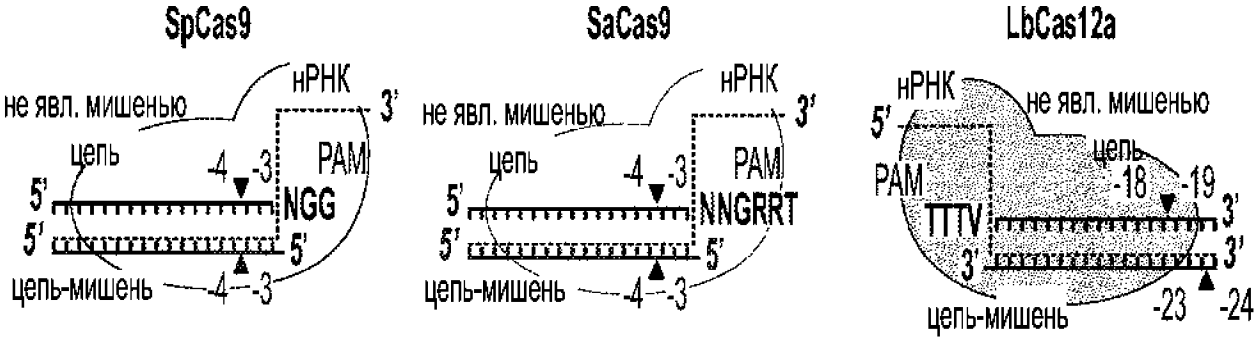
Объем праймерного редактирования

Все 4 переходные точечные мутации
Все 8 трансверсивных точечных мутаций
Вставки (от 1 bp до >44 bp)
Делеции (от 1 bp до >80 bp)
Комбинации указанных выше

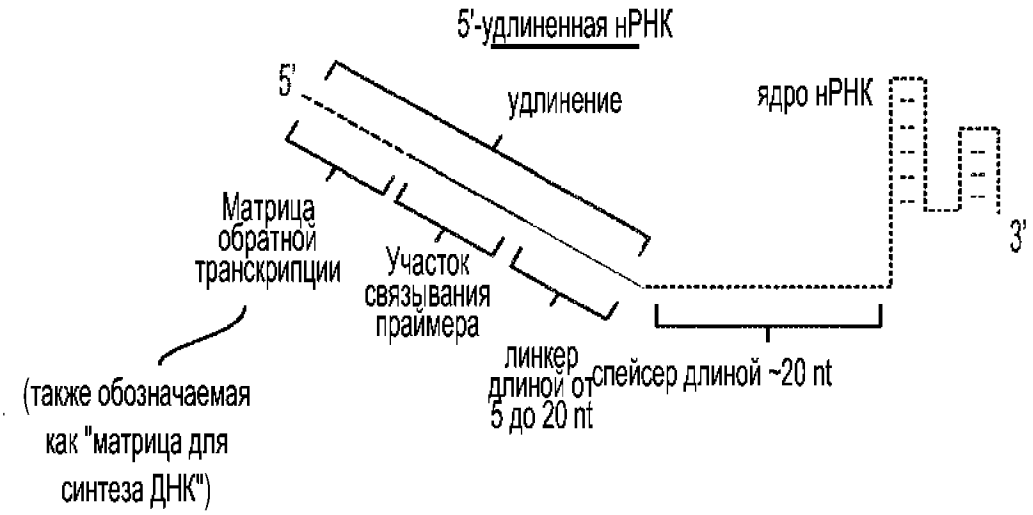
ФИГ.1I


Полимераза (например РНК-зависимая ДНК-полимераза (например обратная транскриптаза), ДНК-зависимая ДНК-полимераза)
плечо удлинения
эквивалентные термины
Эквивалентные термины
матрица для редактирования
матрица для синтеза ДНК
полимеразная матрица
PEgRNA (или pegRNA)
Участок связывания праймера
RT матрица включ. изменение
нуклеаза RuvC
3'
5'
3'
PE + pegRNA + ДНК
Никирование содержащей PAM цепи
5'
3'
гибридизация участка связывания праймера в PEgRNA с содержащей PAM цепью
не содержащая PAM цепь
цепь-мишень
не подвергаемая редактированию цепь
эквивалентные термины
спейсер
нуклеаза HNH с мутацией (например H840A)
ядро нРНК
каркас нРНК
основа нРНК
эквивалентные термины
napDNAbp
(например никаза SpCas9 с деактивированным доменом нуклеазы HNH (например мутацией H840A), так что только не явл. мишенью цепь становится никированной доменом нуклеазы RuvC
полимеризация с 3'-конца никированной содержащей PAM цепи с использованием матрицы для синтеза ДНК
Прямая полимеризация отредактированной ДНК из PEgRNA в участок-мишень
3'
5'
3'-флэп (содержит изменение)
5'-флэп (не содержит изменение)
выравнивание
расщепление 5'-флэпа
встраивание изменения
лигирование
репарация ДНК
отредактированная ДНК

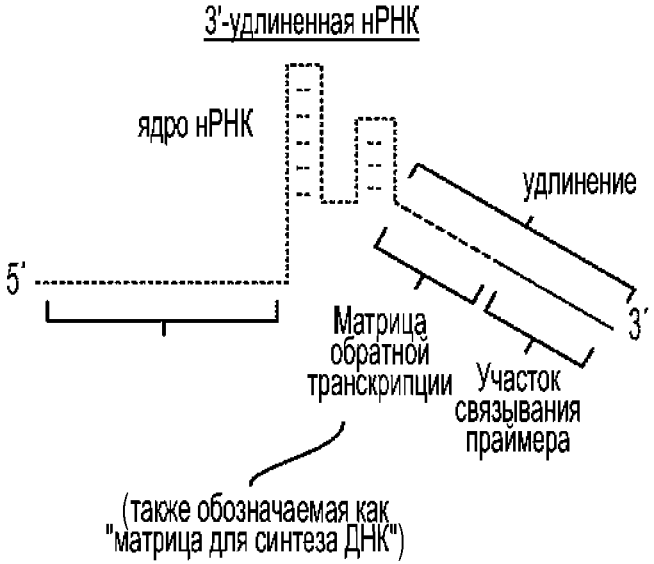
ФИГ.1J


SpCas9
не явл. мишенью
нРНК
3'
цепь
-4
-3
PAM
5'
NGG
5'
5'
цепь-мишень
-4
-3
SaCas9
не явл. мишенью
нРНК
3'
цепь
-4
-3
PAM
5'
NNGRRT
5'
5'
цепь-мишень
-4
-3
LbCas12a
нРНК
5'
не явл. мишенью
PAM
цепь
-18
-19
TTTV
3'
3'
3'
цепь-мишень
-23
-24

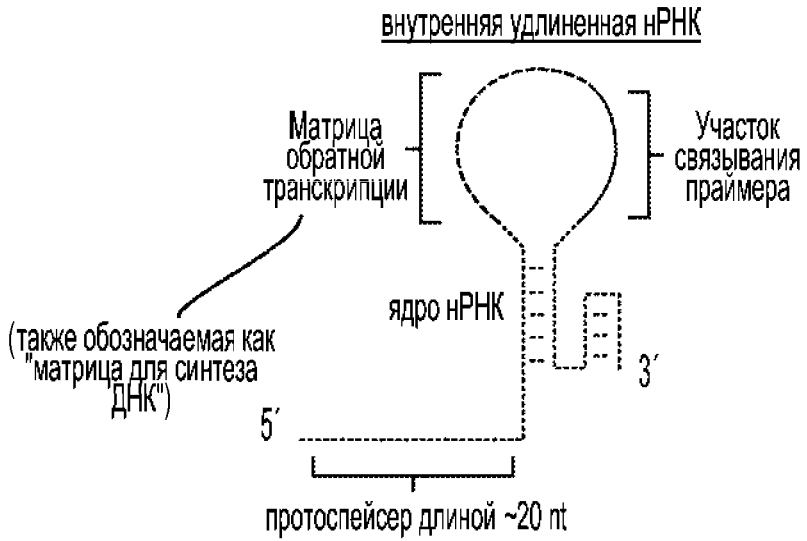
ФИГ.2


5'-удлиненная нРНК
5'
удлинение
ядро нРНК
3'
Матрица обратной транскрипции
Участок связывания праймера
линкер длиной от 5 до 20 nt
спейсер длиной ~20 nt
(также обозначаемая как "матрица для синтеза ДНК")

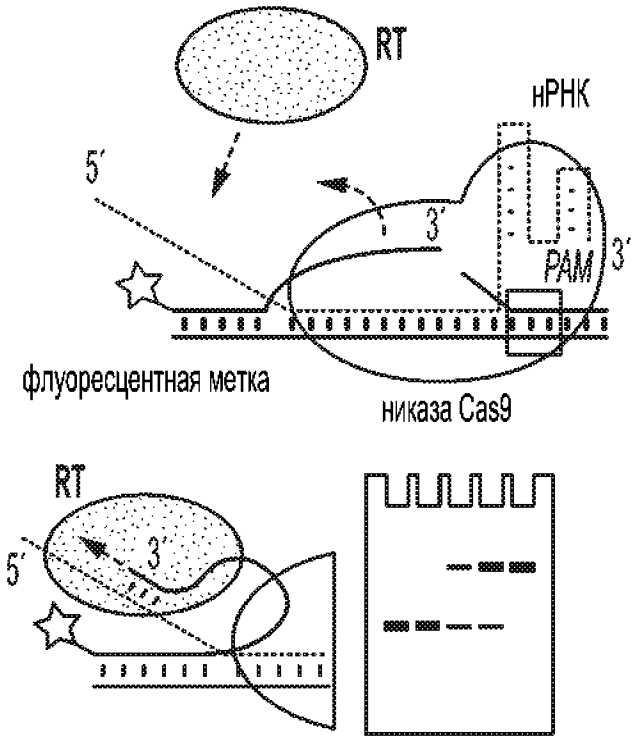
ФИГ.3А


3'-удлиненная нРНК
ядро нРНК
удлинение
5'
3'
Матрица обратной транскрипции
Участок связывания праймера
(также обозначаемая как "матрица для синтеза ДНК")

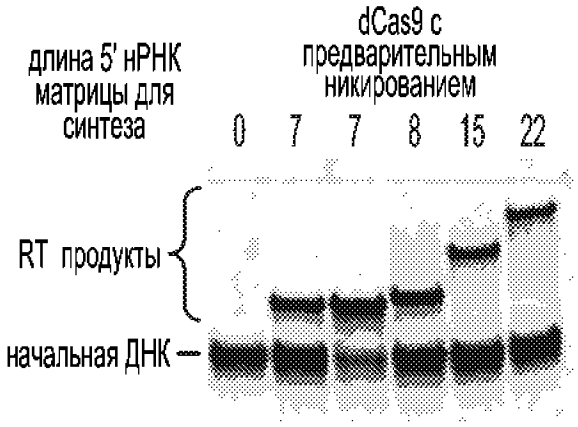
ФИГ.В


внутренняя удлиненная нРНК
Матрица обратной транскрипции
Участок связывания праймера
(также обозначаемая как "матрица для синтеза ДНК")
ядро нРНК
3'
5'
протоспейсер длиной ~20 nt

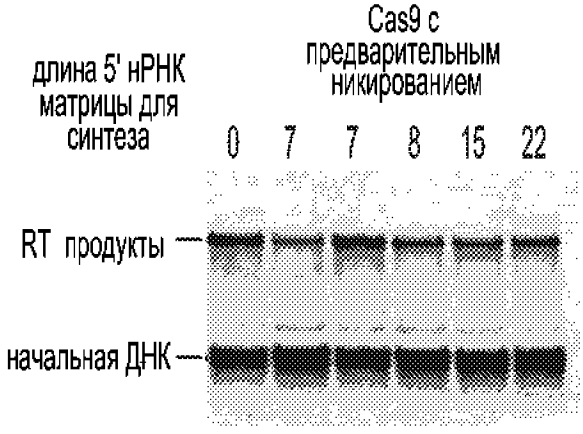
ФИГ.3С



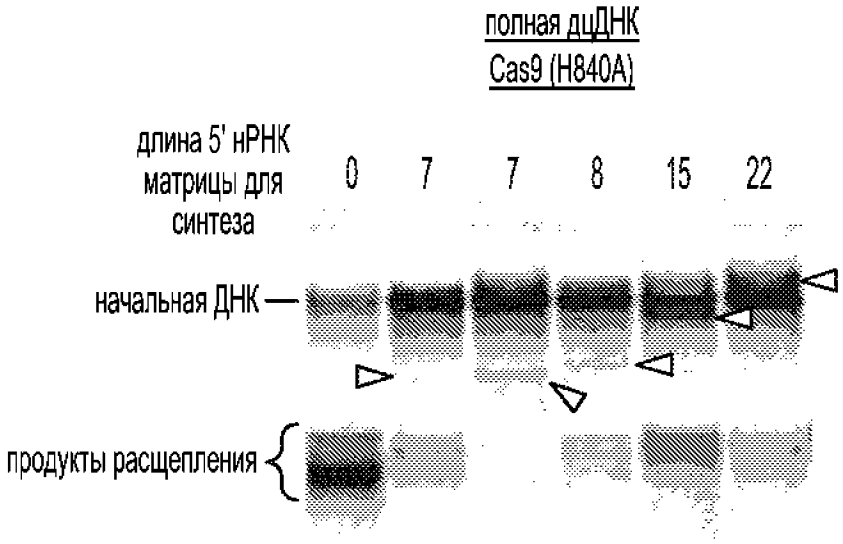
ФИГ.4А



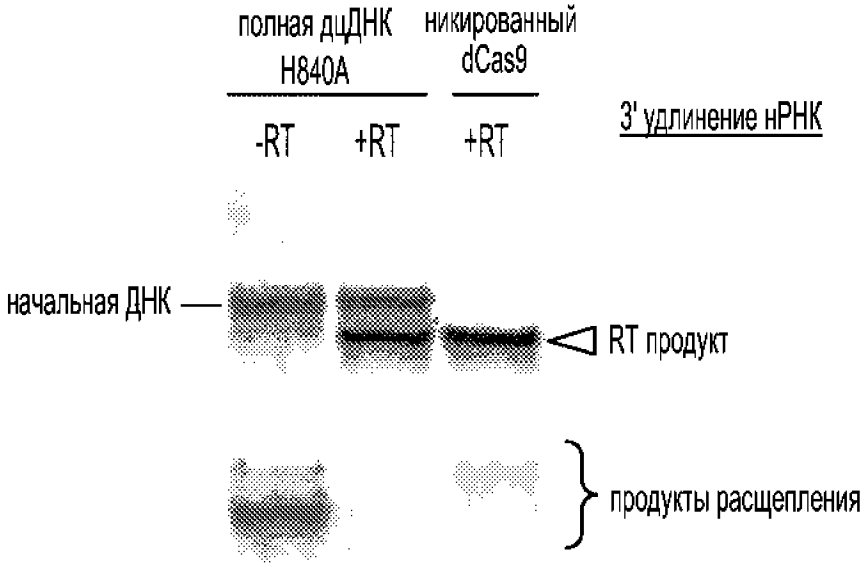
ФИГ.4В



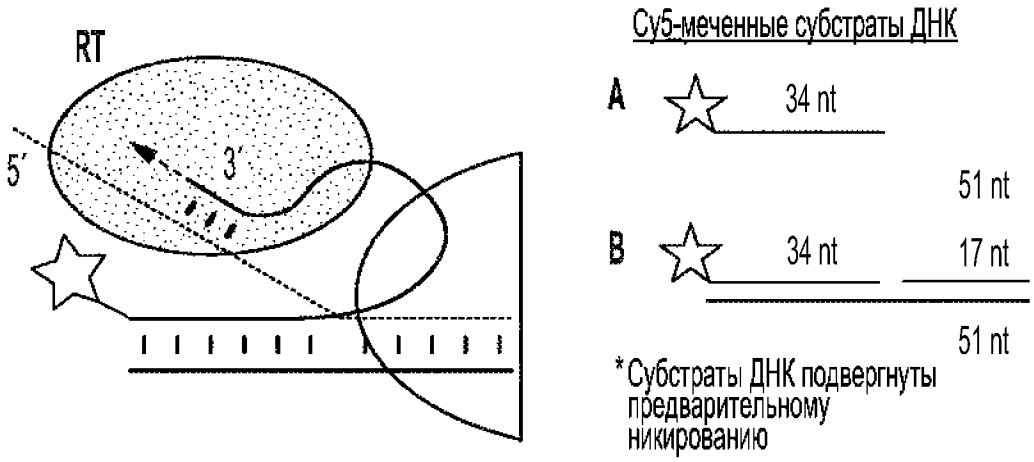
ФИГ.4С

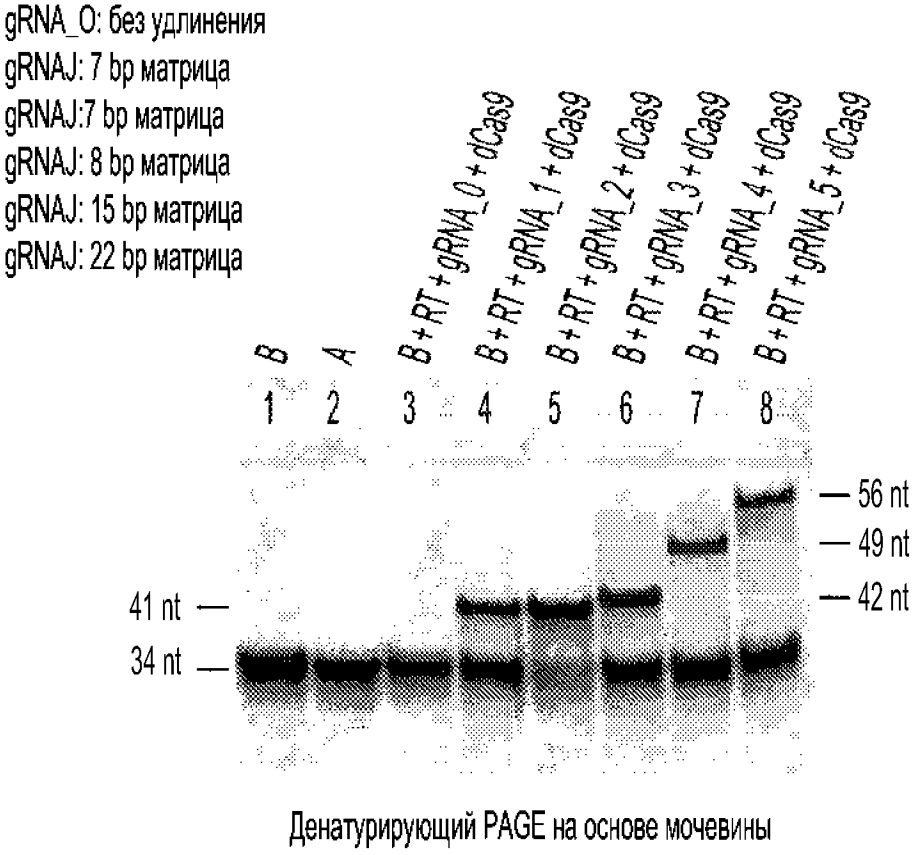

полная дцДНК
Cas9 (H840A)
длина 5' нРНК матрицы для синтеза
0
7
7
8
15
22
начальная ДНК
продукты расщепления

ФИГ.4D

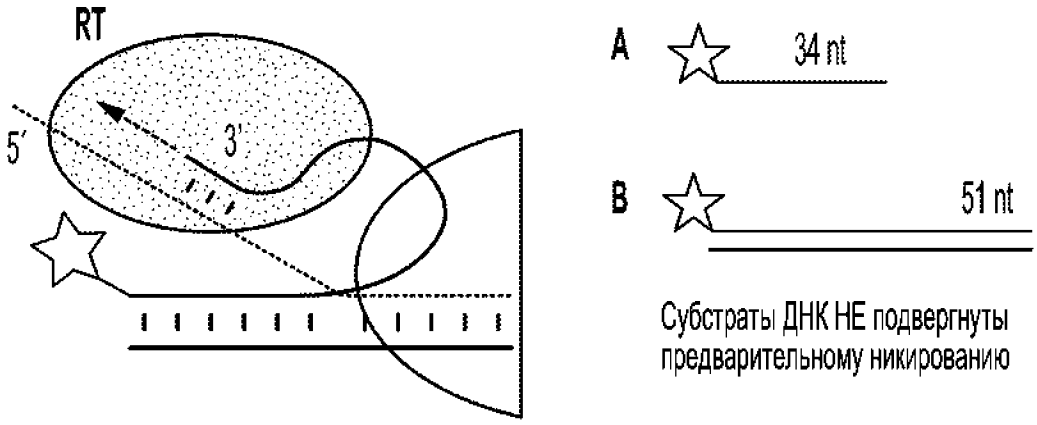

полная дцДНК H840A
никированный dCas9
-RT
+RT
+RT
3' удлинение нРНК
начальная ДНК
RT продукт
продукты расщепления

ФИГ.4E

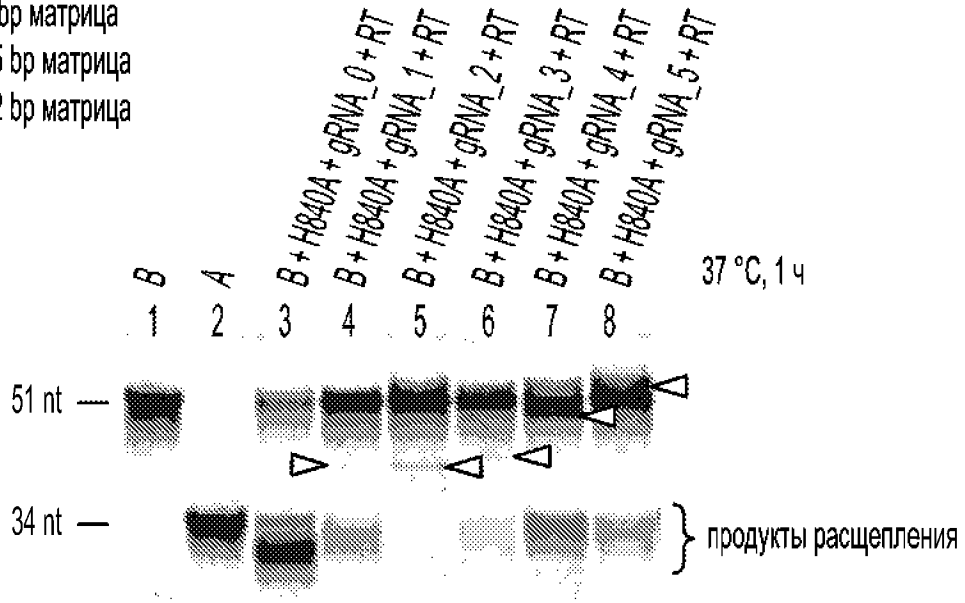

RT
5'
3'
Cy5-меченные субстраты ДНК
A
34 nt
51 nt
B
34 nt
17 nt
51 nt
*Субстраты ДНК подвергнуты предварительному никированию


gRNA_O: без удлинения
gRNAJ: 7 bp матрица
gRNAJ:7 bp матрица
gRNAJ: 8 bp матрица
gRNAJ: 15 bp матрица
gRNAJ: 22 bp матрица
B
A
B+RT+gRNA_0+dCas9
B+RT+gRNA_1+dCas9
B+RT+gRNA_2+dCas9
B+RT+gRNA_3+dCas9
B+RT+gRNA_4+dCas9
B+RT+gRNA_5+dCas9
1
2
3
4
5
6
7
8
56 nt
49 nt
42 nt
41 nt
34 nt
Денатурирующий PAGE на основе мочевины

ФИГ.5

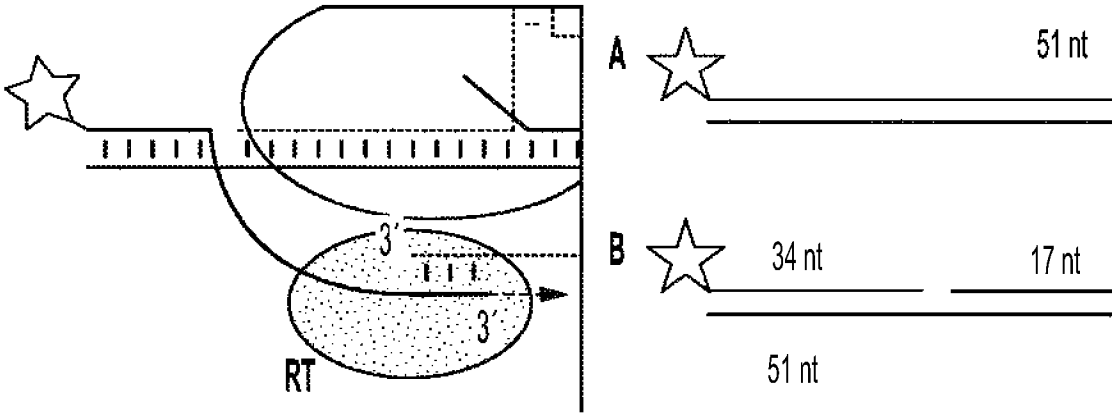


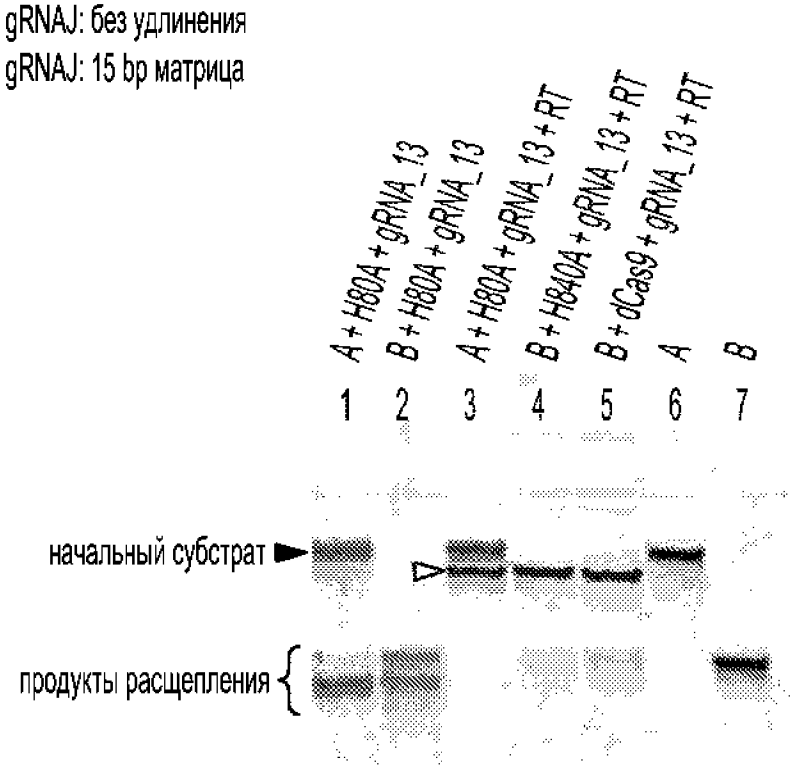
gRNA_0: без удлинения
gRNAJ: 7 bp матрица
gRNA_2:7 bp матрица
gRNAJ: 8 bp матрица
gRNAJ: 15 bp матрица
gRNAJ: 22 bp матрица



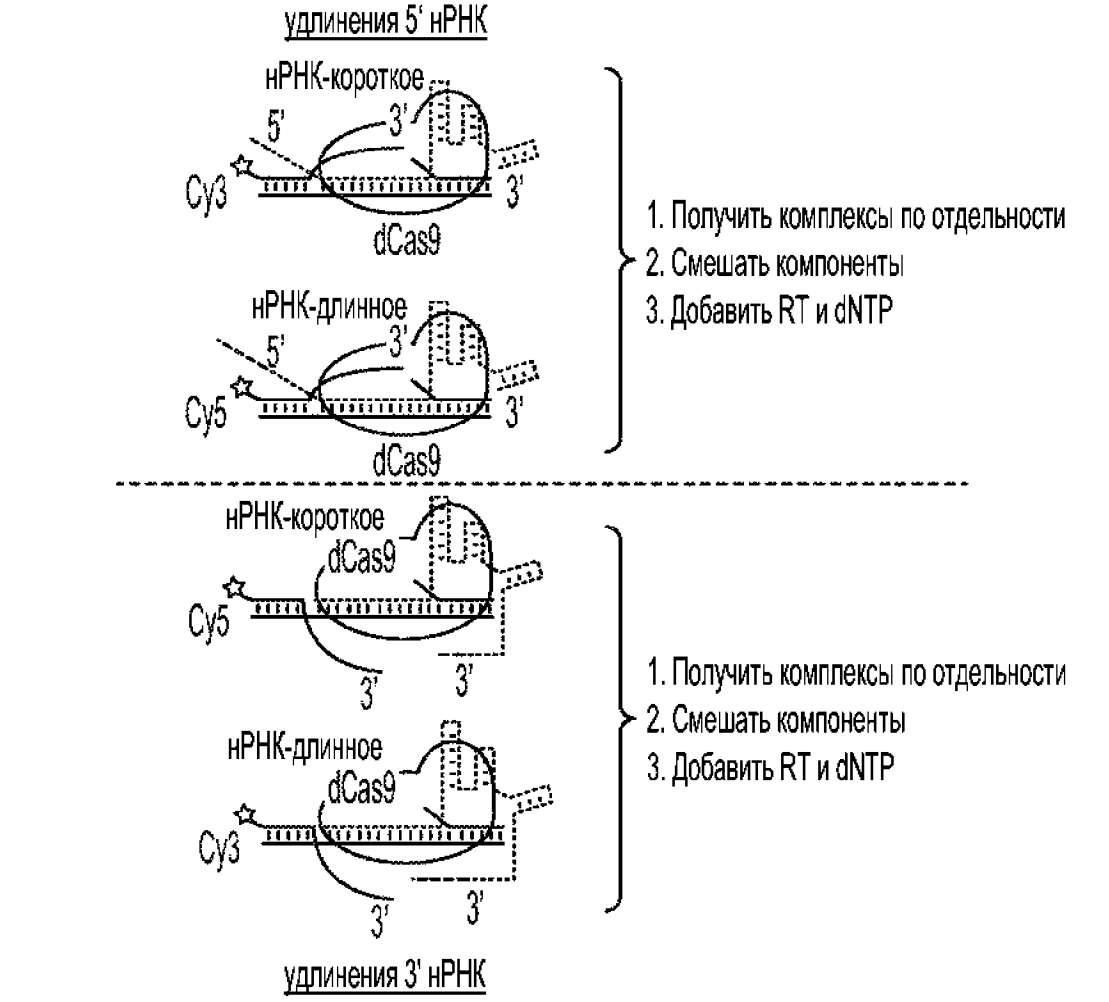
15% денатурирующий PAGE на основе мочевины

ФИГ.6


A
51 nt
B
34 nt
17 nt
51 nt
3´
3´
RT

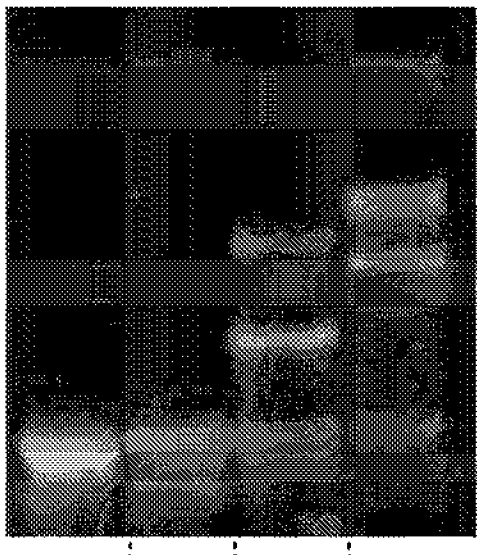

gRNAJ: без удлинения
gRNAJ: 15 bp матрица
A + H80A + gRNA_13
B + H80A + gRNA_13
A + H80A + gRNA_13 + RT
B + H840A + gRNA_13 + RT
B + dCas9 + gRNA_13 + RT
A
B
1 2 3 4 5 6 7
начальный субстрат
продукты расщепления

ФИГ.7



$$\frac{cis_{Cy3}/trans_{Cy3}}{trans_{Cy5}/cis_{Cy5}}$$

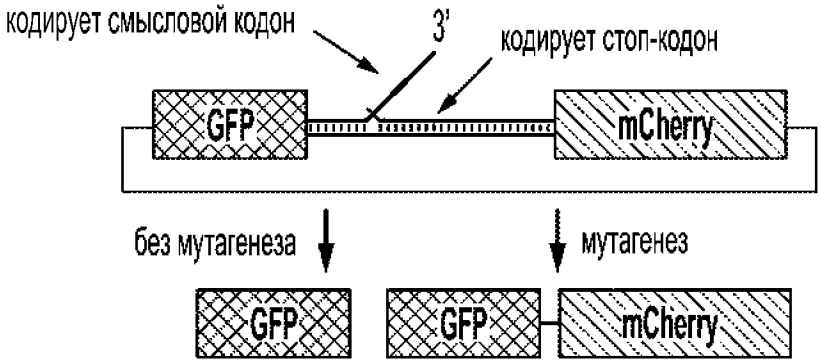
Соотношения продуктов: 7.4 4.7



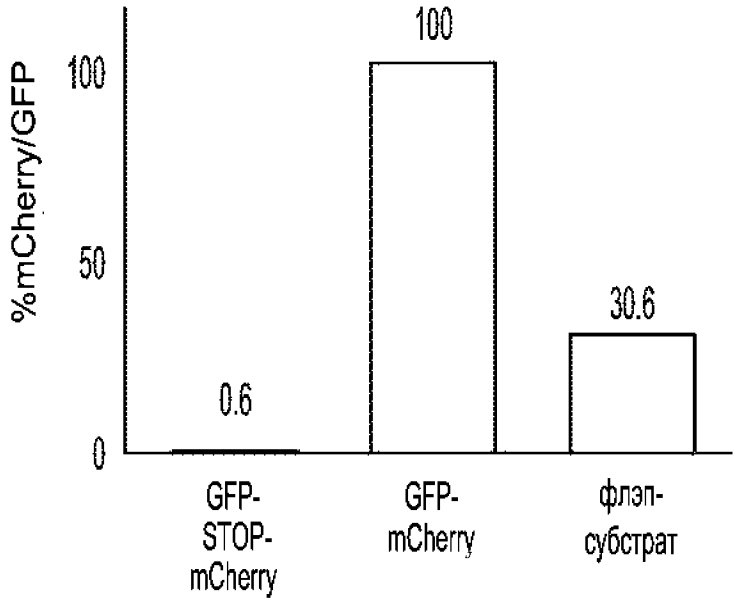
продукты RT

Cy3 | Cy5 | 5'-удл. | 3'-удл.

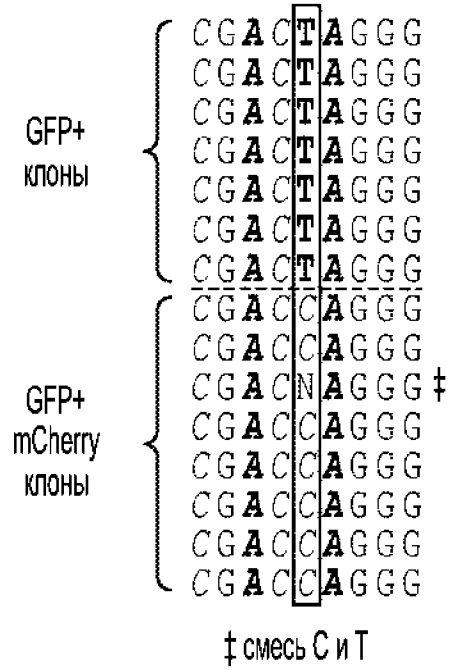
ФИГ.8


кодирует смысловой кодон
3'
кодирует стоп-кодон
GFP
mCherry
без мутагенеза
мутагенез
GFP
GFP
mCherry

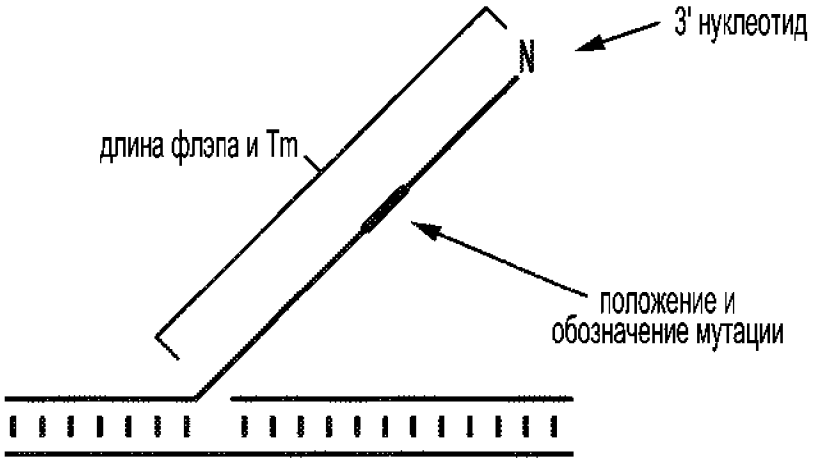
ФИГ.9А


%mCherry/GFP
100
50
0
100
30.6
0.6
GFP-STOP-mCherry
GFP-mCherry
флэп-субстрат

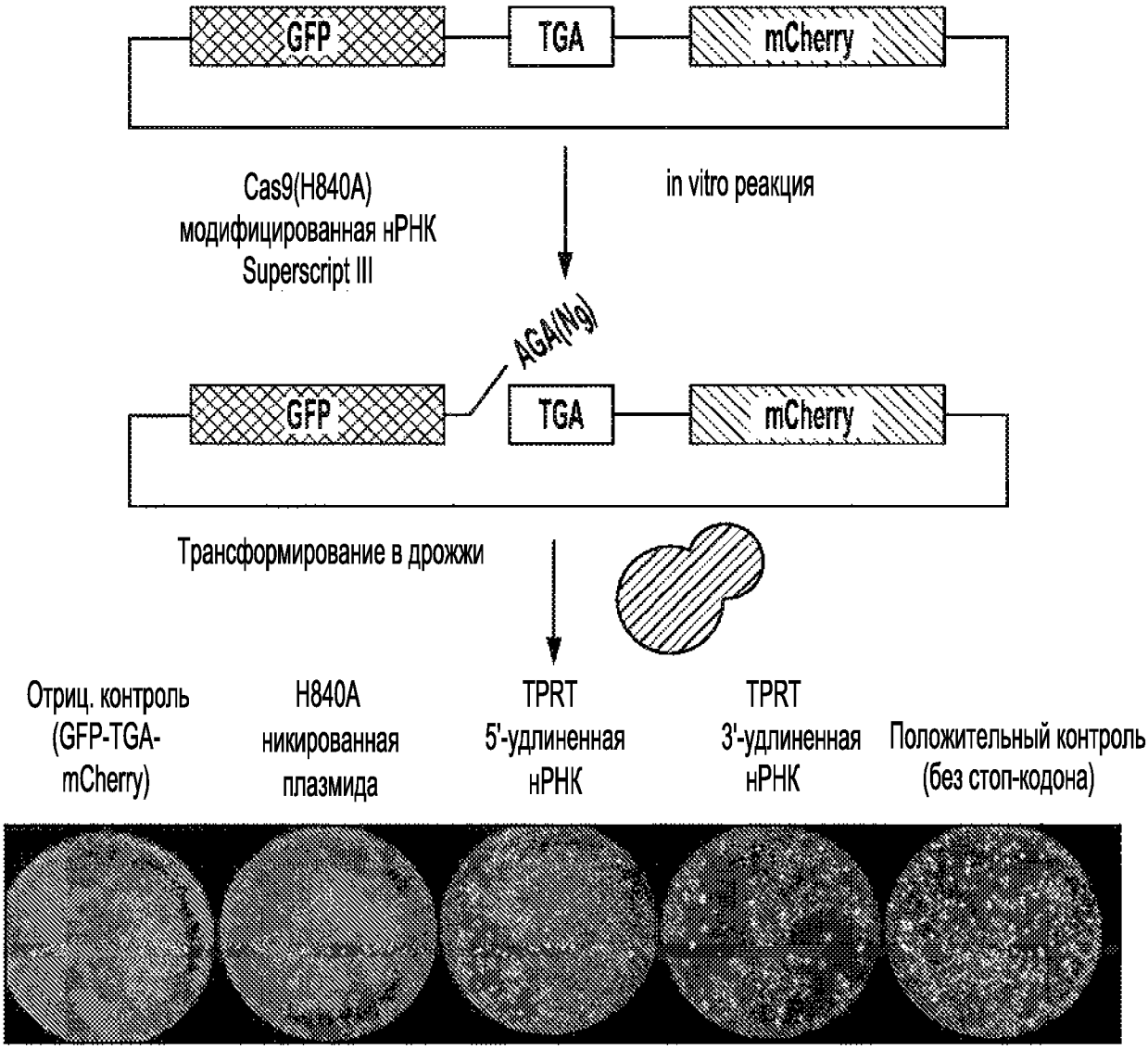
ФИГ.9В


GFP+
клоны
CGACTAGGG
CGACTAGGG
CGACTAGGG
CGACTAGGG
CGACTAGGG
CGACTAGGG
CGACTAGGG
GFP+
mCherry
клоны
CGACCAGGG
CGACCAGGG
CGACNAGGG ‡
CGACCAGGG
CGACCAGGG
CGACCAGGG
CGACCAGGG
CGACCAGGG
‡ смесь C и T

ФИГ.9C

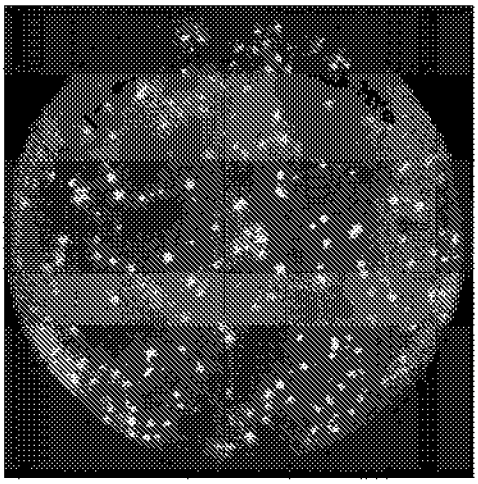

3' нуклеотид
N
длина флэпа и Tm
положение и
обозначение мутации

ФИГ.9D

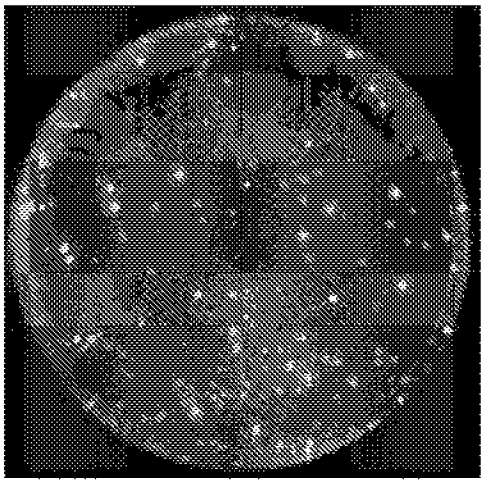

GFP
TGA
mCherry
Cas9(H840A)
модифицированная нРНК
Superscript III
in vitro реакция
AGA(Ng)
GFP
TGA
mCherry
Трансформирование в дрожжи
Отриц. контроль (GFP-TGA-mCherry)
H840A никированная плазмида
TPRT 5'-удлиненная нРНК
TPRT 3'-удлиненная нРНК
Положительный контроль (без стоп-кодона)

ФИГ.10

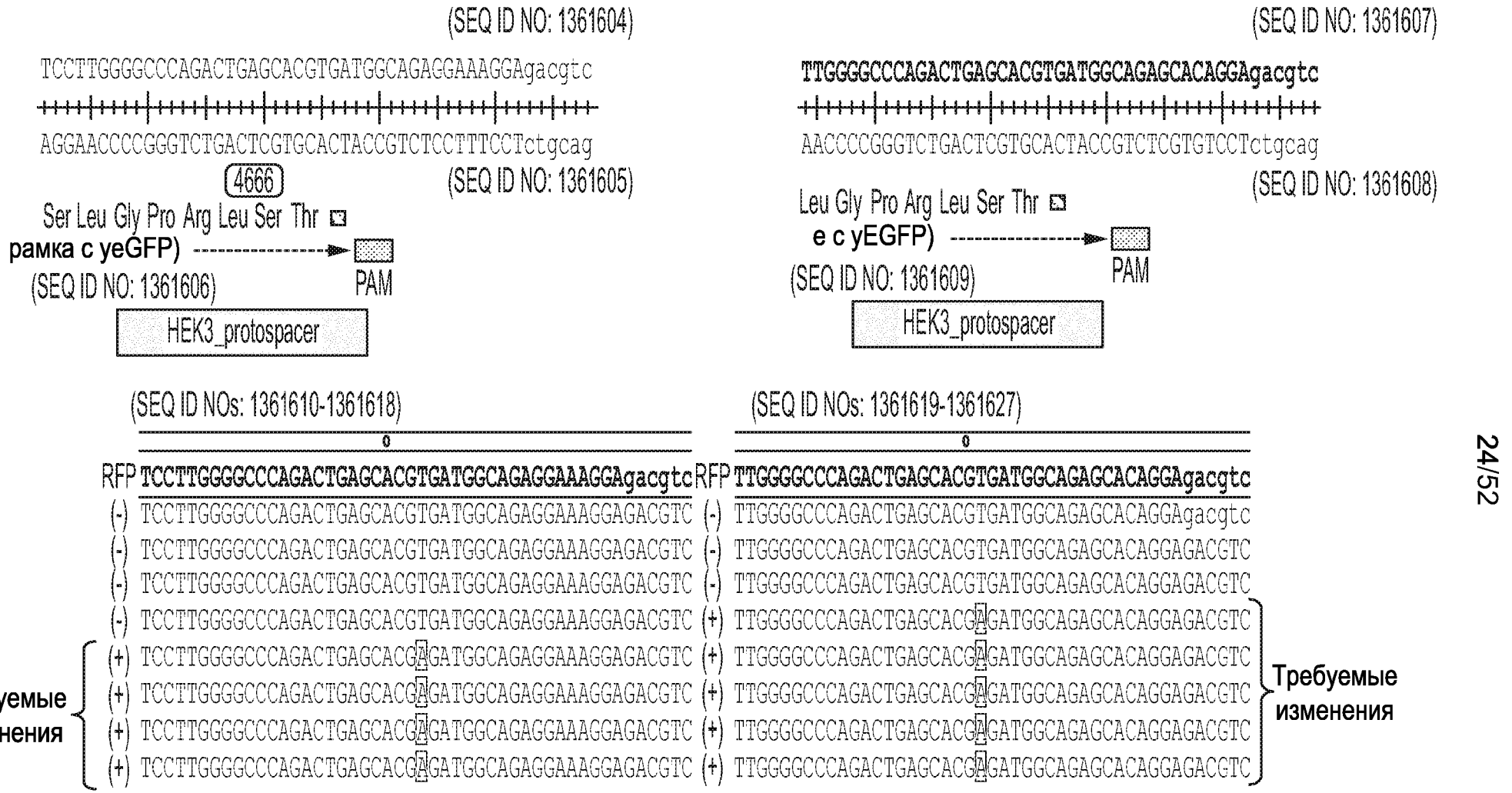
Коррекция вставки (TPRT обеспечивает делецию 1 нуклеотида



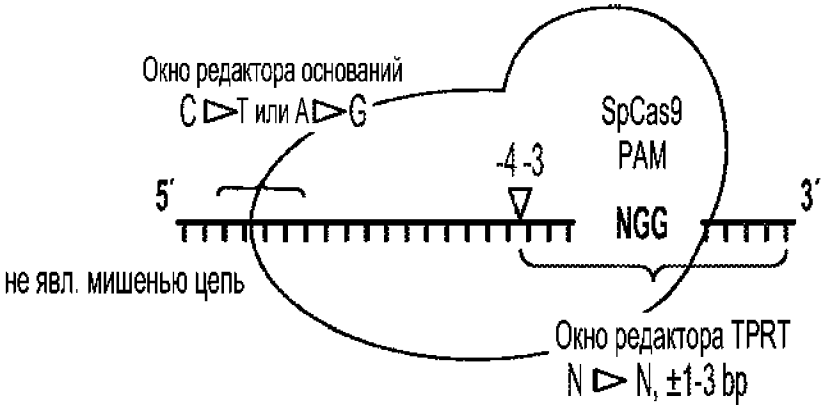
Коррекция делеции (TPRT обеспечивает вставку 1 нуклеотида



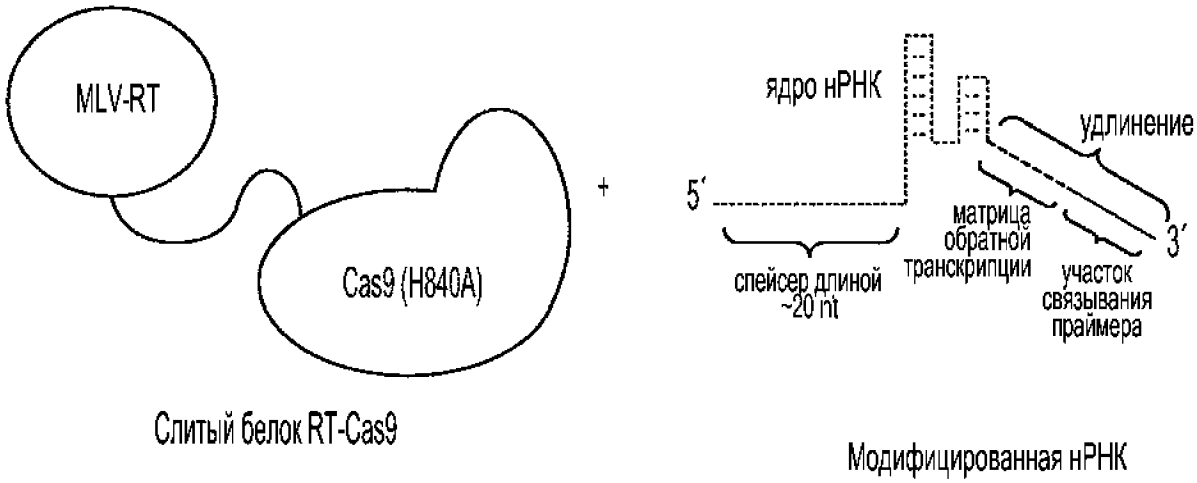
ФИГ.11


(SEQ ID NO: 1361604)
TCCTTGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAgacgtc
AGGAACCCCGGGTCTGACTCGTGCACTACCGTCTCCTTTCCTctgcag
4666
(SEQ ID NO: 1361605)
Ser Leu Gly Pro Arg Leu Ser Thr
рамка с yeGFP)
(SEQ ID NO: 1361606)
PAM
HEK3_protospacer
(SEQ ID NO: 1361607)
TTGGGGCCCAGACTGAGCACGTGATGGCAGAGCACAGGAgacgtc
AACCCCGGGTCTGACTCGTGCACTACCGTCTCGTGTCCTctgcag
(SEQ ID NO: 1361608)
Leu Gly Pro Arg Leu Ser Thr
e c yEGFP)
(SEQ ID NO: 1361609)
PAM
HEK3_protospacer
(SEQ ID NOs: 1361610-1361618)
(SEQ ID NOs: 1361619-1361627)
RFP TCCTTGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAgacgtc
RFP TTGGGGCCCAGACTGAGCACGTGATGGCAGAGCACAGGAgacgtc
Требуемые изменения
Требуемые изменения

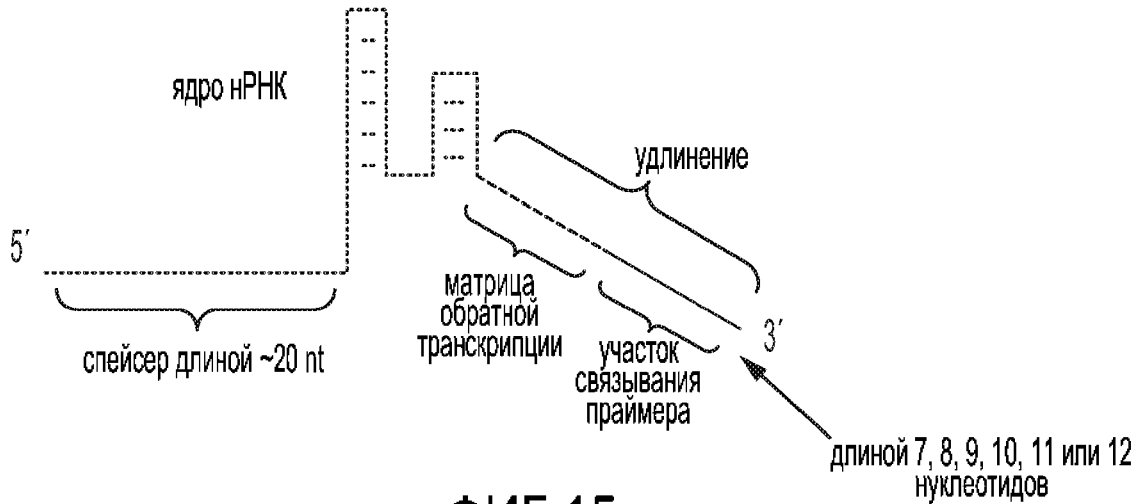
ФИГ.12


Окно редактора оснований
C ▷T или A ▷G
SpCas9
PAM
-4 -3
5′
3′
NGG
не явл. мишенью цепь
Окно редактора TPRT
N ▷ N, ±1-3 bp

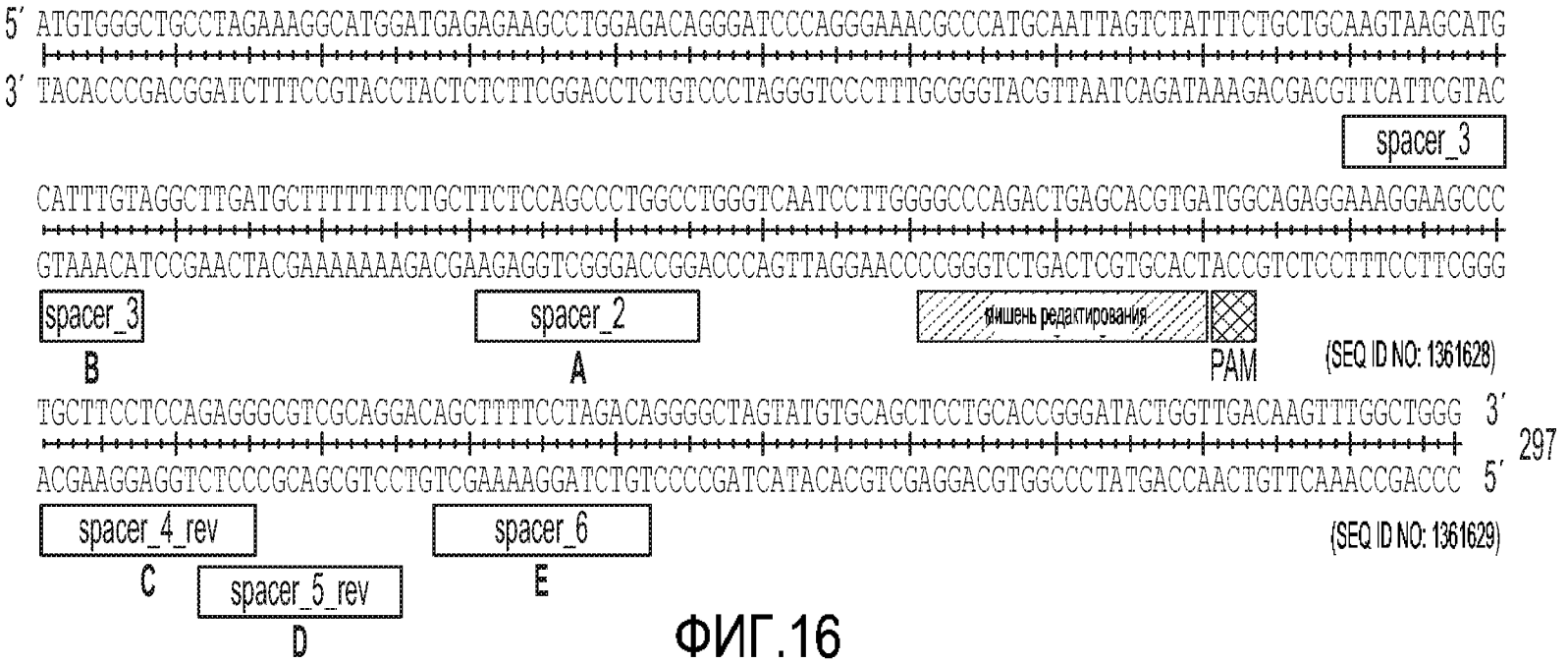
ФИГ.13


MLV-RT
+
Cas9 (H840A)
Слитый белок RT-Cas9
ядро нРНК
удлинение
5′
3′
матрица обратной транскрипции
спейсер длиной ~20 nt
участок связывания праймера
Модифицированная нРНК

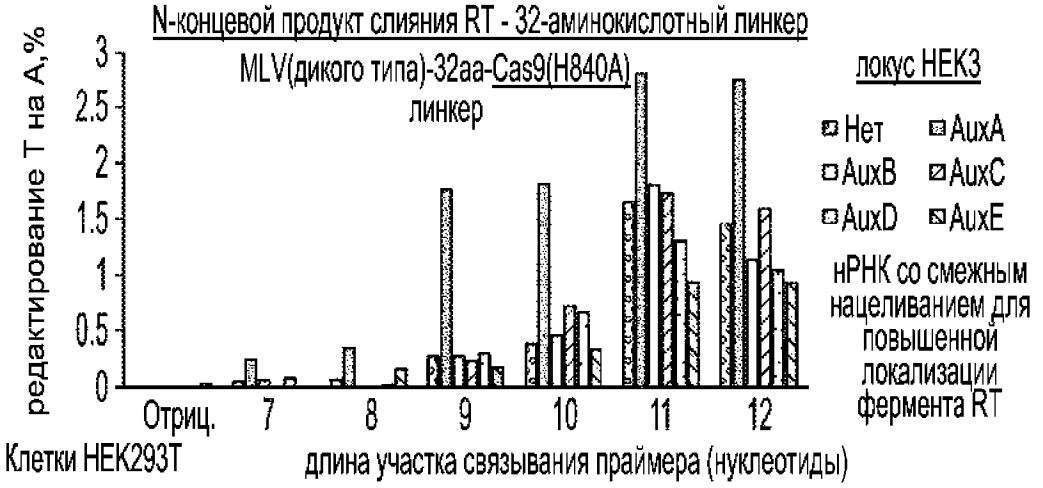
ФИГ.14


ядро нРНК
5′
спейсер длиной ~20 nt
матрица обратной транскрипции
участок связывания праймера
удлинение
3′
длиной 7, 8, 9, 10, 11 или 12 нуклеотидов

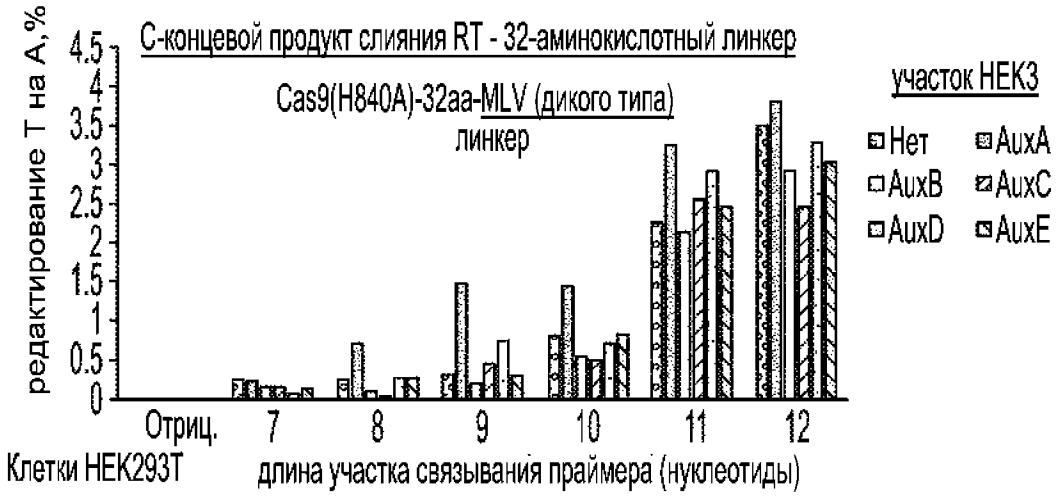
ФИГ.15


5′ ATGTGGGCTGCCTAGAAAGGCATGGATGAGAGAAGCCTGGAGACAGGGATCCCAGGGAAACGCCCATGCAATTAGTCTATTTCTGCTGCAAGTAAGCATG
3′ TACACCCGACGGATCTTTCCGTACCTACTCTCTTCGGACCTCTGTCCCTAGGGTCCCTTTGCGGGTACGTTAATCAGATAAAGACGACGTTCATTCGTAC
spacer_3
CATTTGTAGGCTTGATGCTTTTTTTCTGCTTCTCCAGCCCTGGCCTGGGTCAATCCTTGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAAGCCC
GTAAACATCCGAACTACGAAAAAAAGACGAAGAGGTCGGGACCGGACCCAGTTAGGAACCCCGGGTCTGACTCGTGCACTACCGTCTCCTTTCCTTCGGG
spacer_3
B
spacer_2
A
мишень редактирования
PAM
(SEQ ID NO: 1361628)
TGCTTCCTCCAGAGGGCGTCGCAGGACAGCTTTTCCTAGACAGGGGCTAGTATGTGCAGCTCCTGCACCGGGATACTGGTTGACAAGTTTGGCTGGG 3′
297
ACGAAGGAGGTCTCCCGCAGCGTCCTGTCGAAAAGGATCTGTCCCCGATCATACACGTCGAGGACGTGGCCCTATGACCAACTGTTCAAACCGACCC 5′
spacer_4_rev
C
spacer_5_rev
D
spacer_6
E
(SEQ ID NO: 1361629)

ФИГ.16

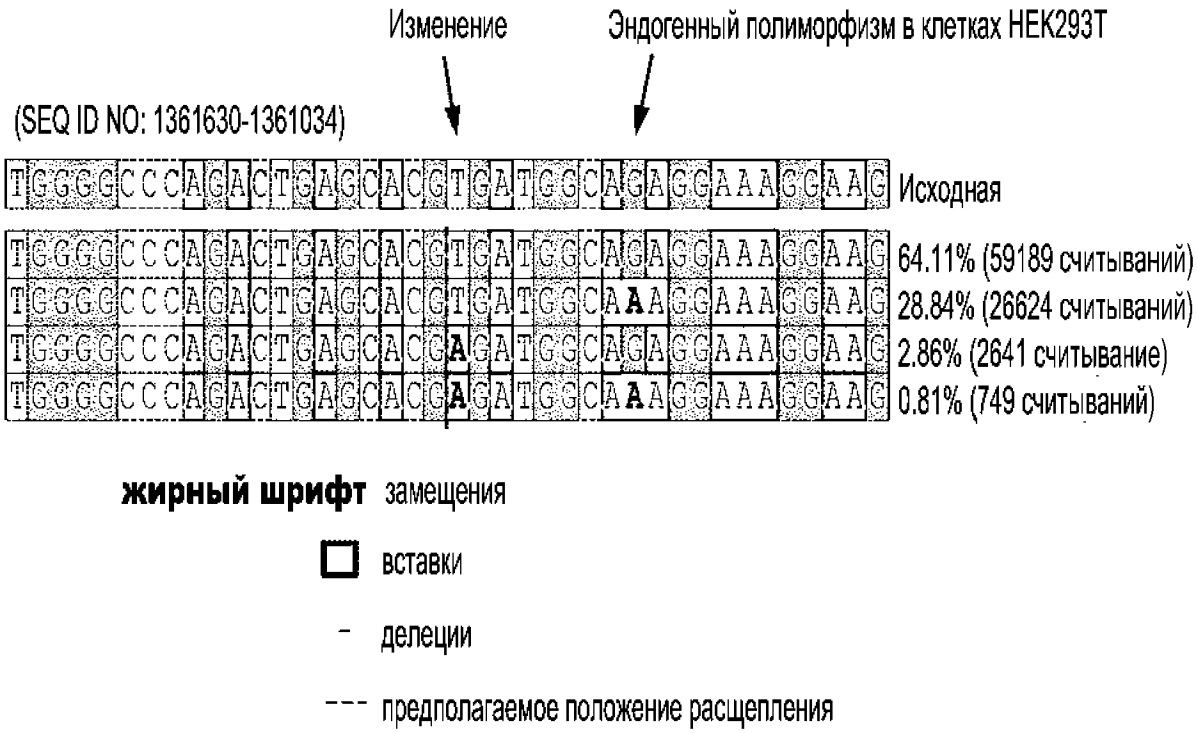


ФИГ.17A

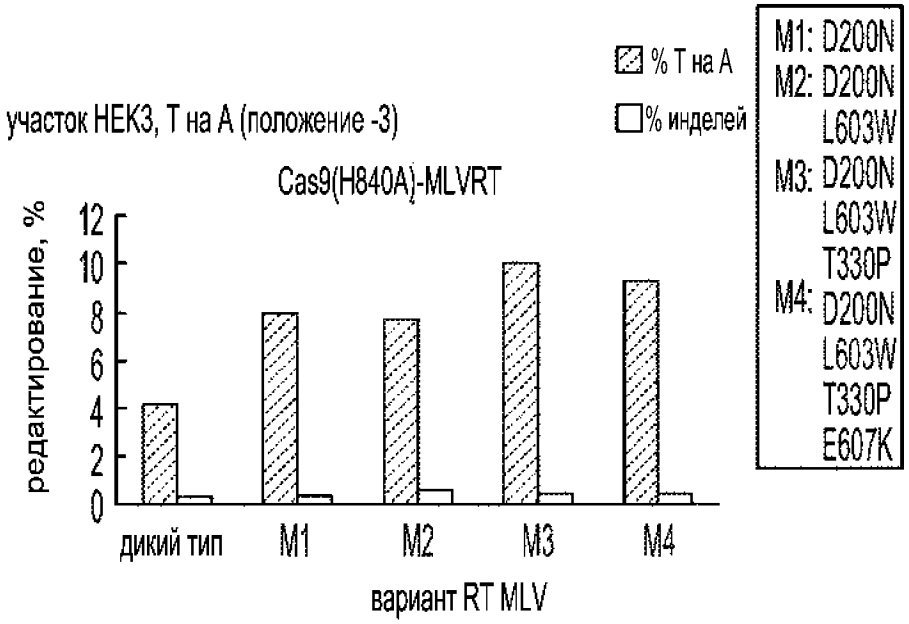


ФИГ.17B

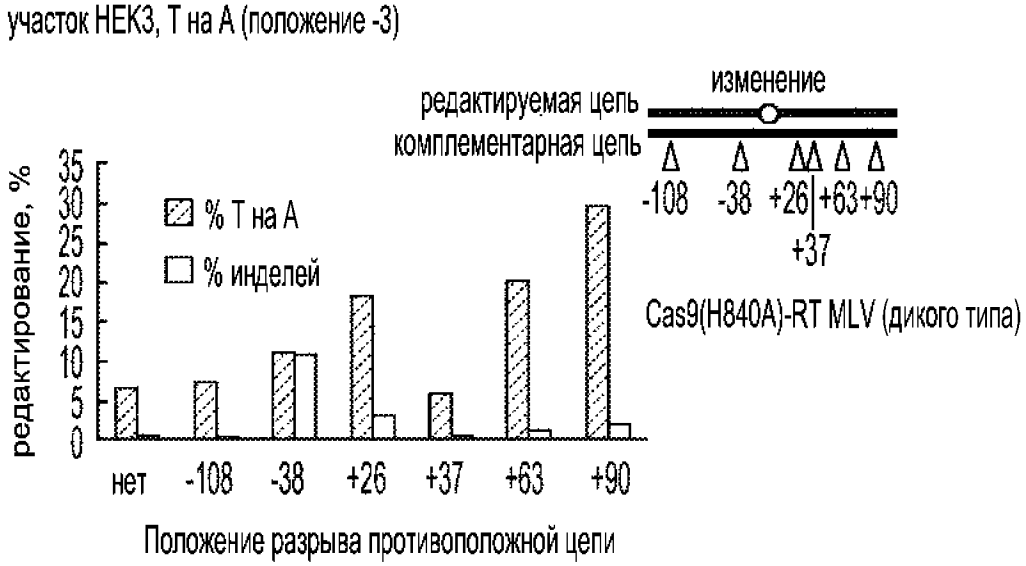
ФИГ.17C


Изменение
Эндогенный полиморфизм в клетках HEK293T
(SEQ ID NO: 1361630-1361034)
TGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAAG
Исходная
TGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAAG
64.11% (59189 считываний)
TGGGGCCCAGACTGAGCACGTGATGGCAAAGGAAAGGAAG
28.84% (26624 считываний)
TGGGGCCCAGACTGAGCACGAGATGGCAGAGGAAAGGAAG
2.86% (2641 считывание)
TGGGGCCCAGACTGAGCACGAGATGGCAAAGGAAAGGAAG
0.81% (749 считываний)
жирный шрифт замещения
вставки
- делеции
--- предполагаемое положение расщепления

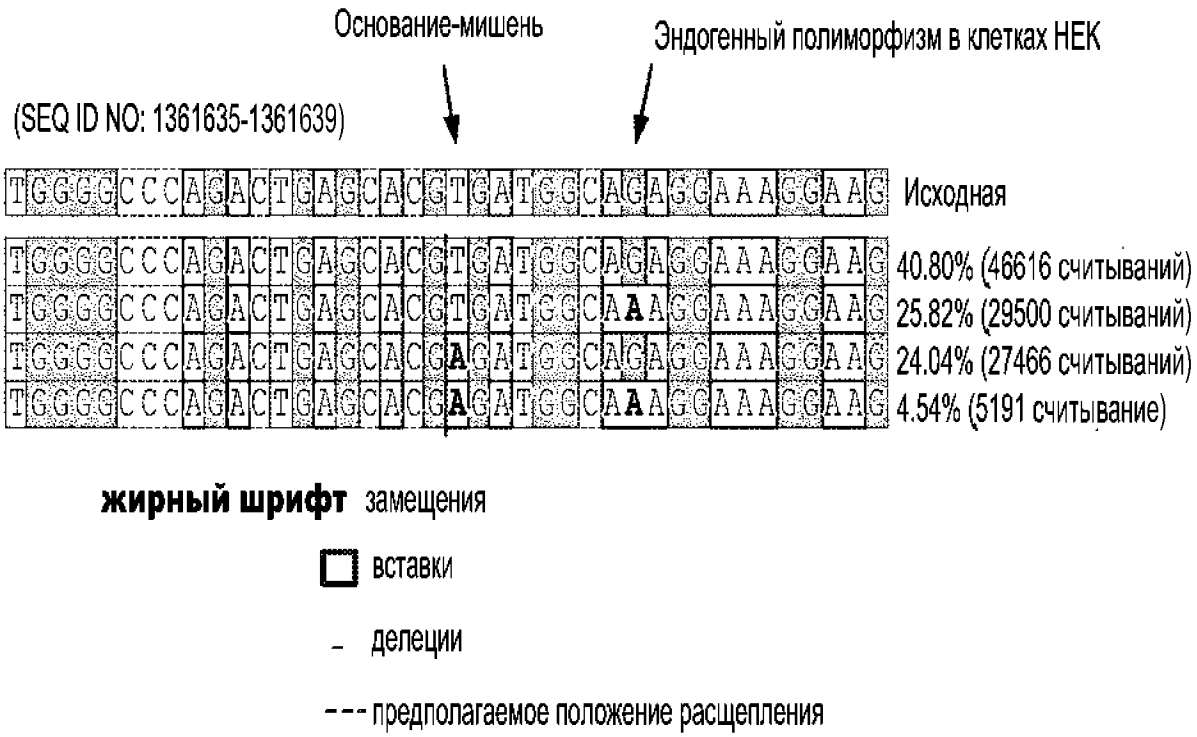
ФИГ.18


участок HEK3, T на A (положение -3)
Cas9(H840A)-MLVRT
% T на A
% инделей
редактирование, %
0
2
4
6
8
10
12
дикий тип
M1
M2
M3
M4
вариант RT MLV
M1: D200N
M2: D200N L603W
M3: D200N L603W T330P
M4: D200N L603W T330P E607K

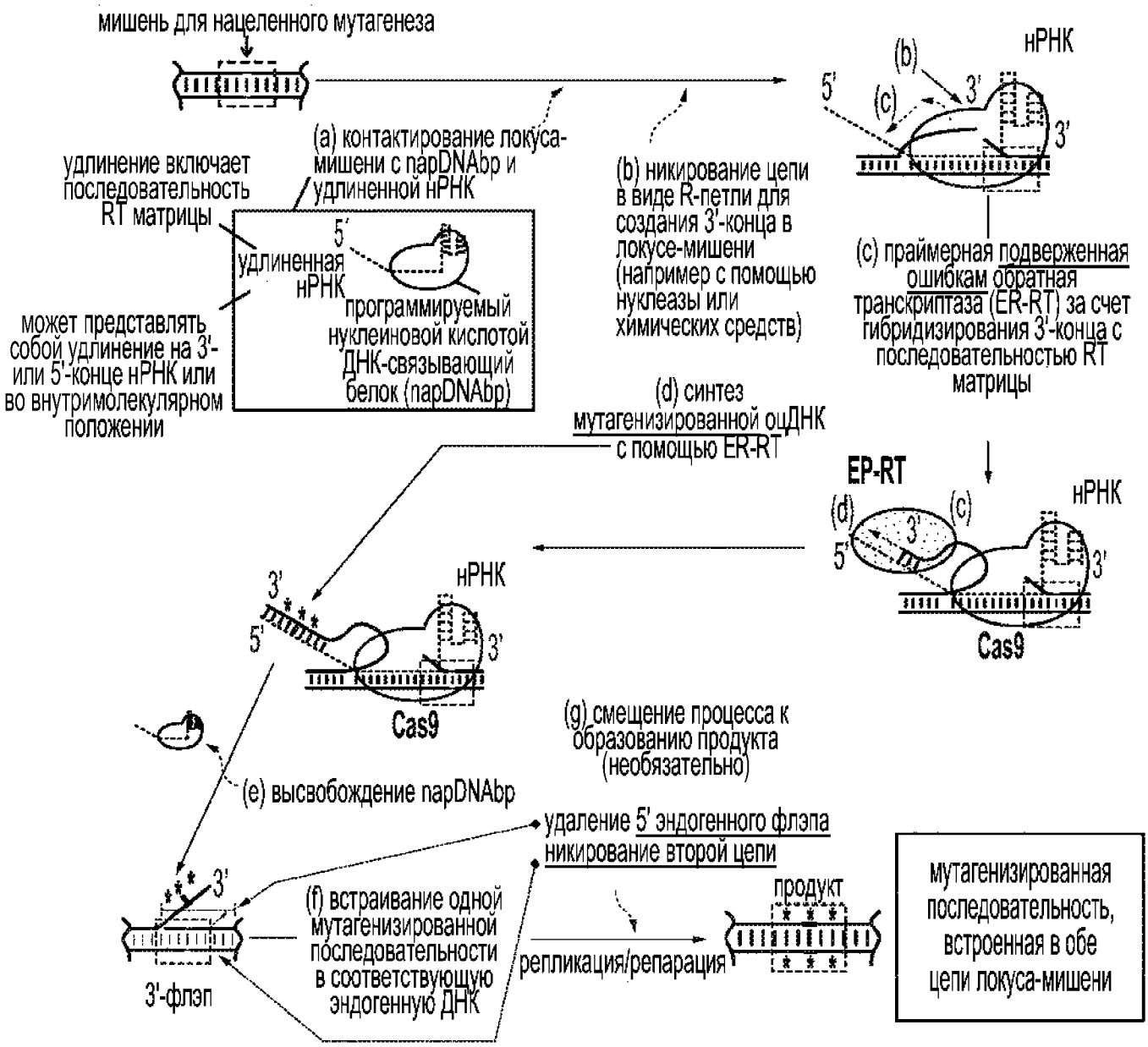
ФИГ.19


участок HEK3, T на A (положение -3)
редактируемая цепь
комплементарная цепь
изменение
-108
-38
+26
+63
+90
+37
Cas9(H840A)-RT MLV (дикого типа)
% T на A
% инделей
редактирование, %
0
5
10
15
20
25
30
35
нет
-108
-38
+26
+37
+63
+90
Положение разрыва противоположной цепи

ФИГ.20


Основание-мишень
Эндогенный полиморфизм в клетках HEK
(SEQ ID NO: 1361635-1361639)
TGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAAG
Исходная
TGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAAG
40.80% (46616 считываний)
TGGGGCCCAGACTGAGCACGTGATGGCAAAGGAAAGGAAG
25.82% (29500 считываний)
TGGGGCCCAGACTGAGCACGAGATGGCAGAGGAAAGGAAG
24.04% (27466 считываний)
TGGGGCCCAGACTGAGCACGAGATGGCAAAGGAAAGGAAG
4.54% (5191 считывание)
жирный шрифт замещения
вставки
делеции
предполагаемое положение расщепления

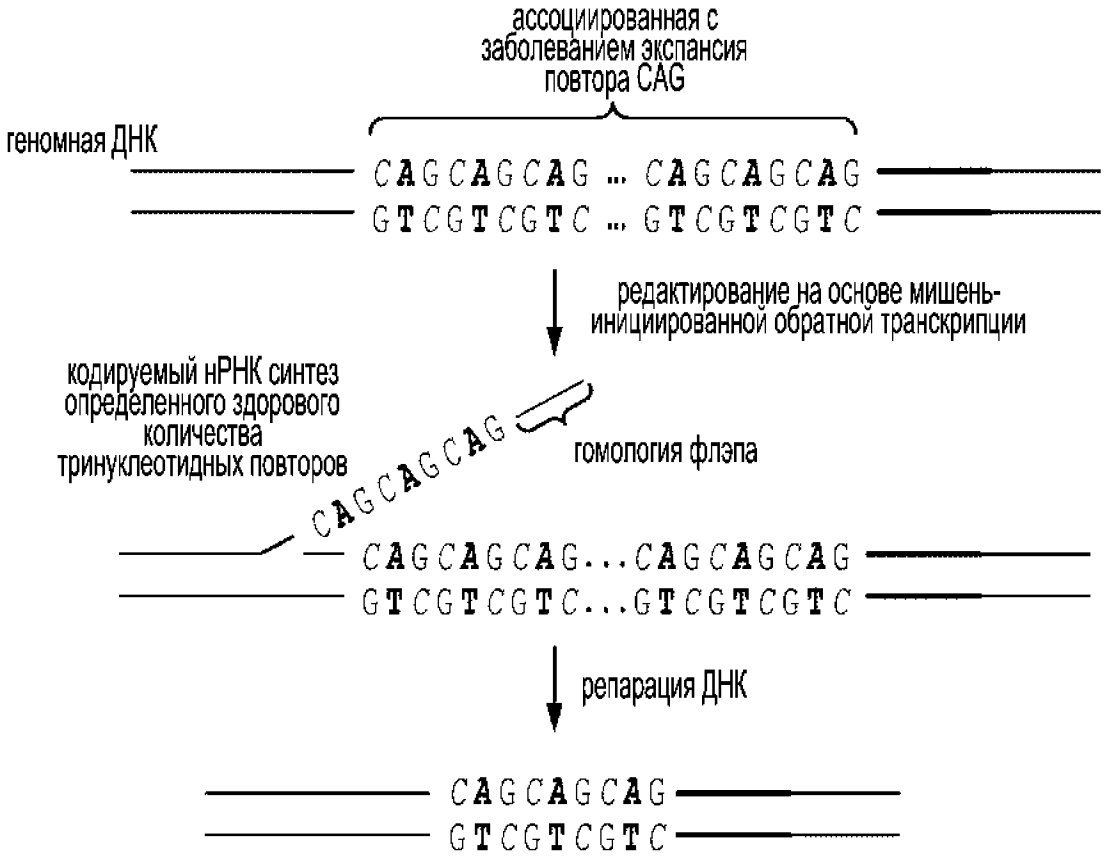
ФИГ.21


мишень для нацеленного мутагенеза
(a) контактирование локуса-мишени с napDNAbp и удлиненной нРНК
удлинение включает последовательность RT матрицы
удлиненная нРНК
5'
программируемый нуклеиновой кислотой ДНК-связывающий белок (napDNAbp)
может представлять собой удлинение на 3'- или 5'-конце нРНК или во внутримолекулярном положении
(b) никирование цепи в виде R-петли для создания 3'-конца в локусе-мишени (например с помощью нуклеазы или химических средств)
нРНК
(b)
(c)
5'
3'
3'
(c) праймерная подверженная ошибкам обратная транскриптаза (ER-RT) за счет гибридизирования 3'-конца с последовательностью RT матрицы
(d) синтез мутагенизированной оцДНК с помощью ER-RT
EP-RT
нРНК
(d)
(c)
3'
5'
RT
3'
Cas9
3'
5'
нРНК
3'
Cas9
(g) смещение процесса к образованию продукта (необязательно)
(e) высвобождение napDNAbp
удаление 5' эндогенного флэпа
никирование второй цепи
3'
(f) встраивание одной мутагенизированной последовательности в соответствующую эндогенную ДНК
продукт
репликация/репарация
3'-флэп
мутагенизированная последовательность, встроенная в обе цепи локуса-мишени

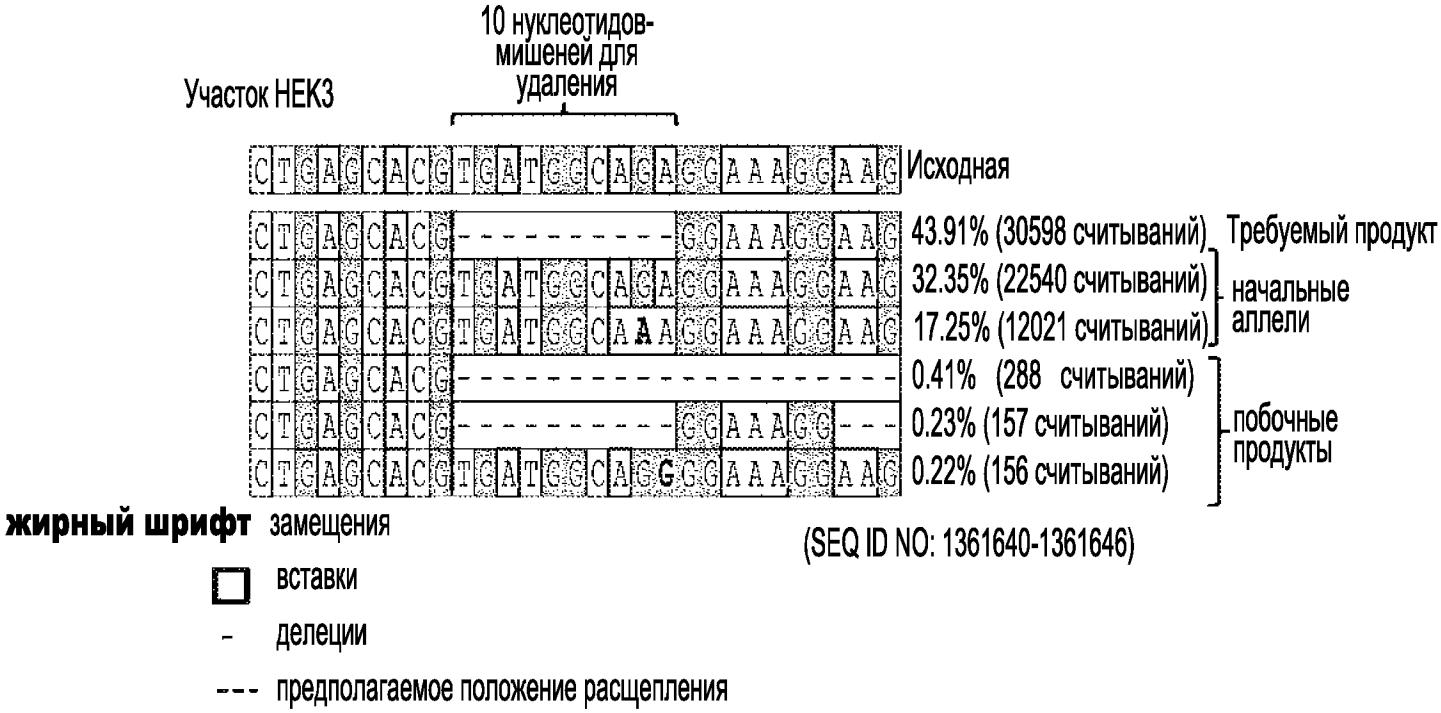
ФИГ.22

Конструирование нРНК для сокращения последовательностей тринуклеотидных повторов

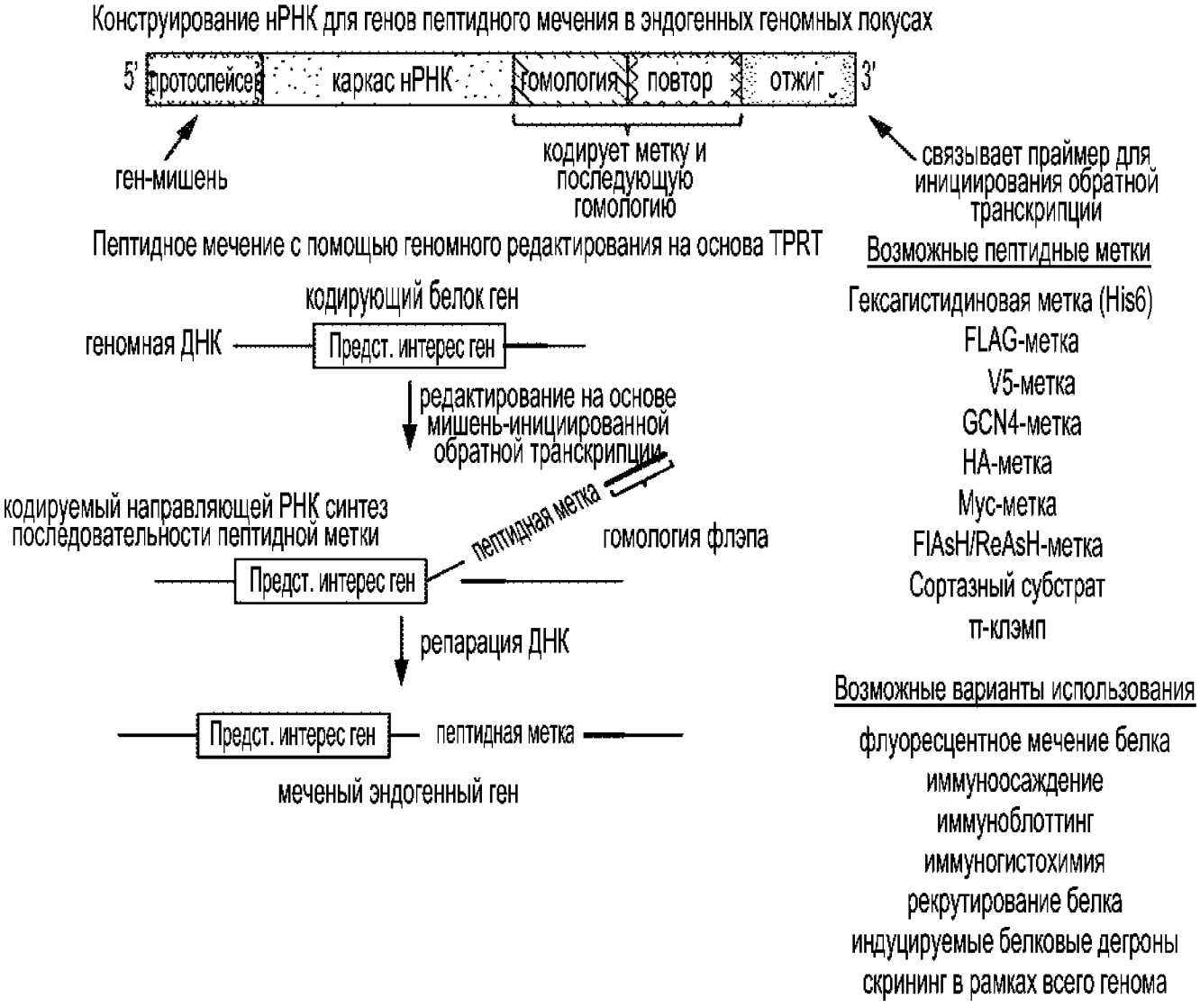
Сокращение тринуклеотидных повторов с помощью геномного редактирования на основе TPRT



ФИГ.23


Участок HEK3
10 нуклеотидов-мишеней для удаления
CTGAGCACGTGATGGCAGAGGAAAGGAAG Исходная
CTGAGCACG----------GGAAAGGAAG 43.91% (30598 считываний)
CTGAGCACGTGATGGCAGAGGAAAGGAAG 32.35% (22540 считываний)
CTGAGCACGTGATGGCAAAGGAAAGGAAG 17.25% (12021 считываний)
CTGAGCACG---------- 0.41% (288 считываний)
CTGAGCACG----------GGAAAGG--- 0.23% (157 считываний)
CTGAGCACGTGATGGCAGGGAAAGGAAG 0.22% (156 считываний)
Требуемый продукт
начальные аллели
побочные продукты
жирный шрифт замещения
вставки
- делеции
--- предполагаемое положение расщепления
(SEQ ID NO: 1361640-1361646)

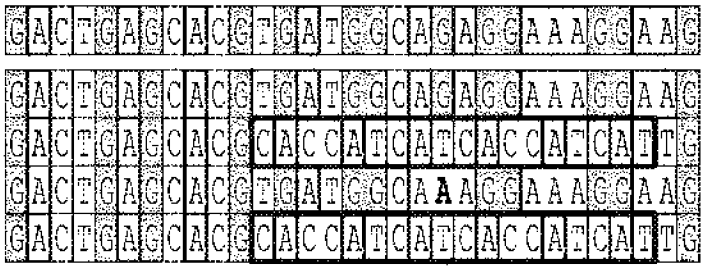
ФИГ.24


Конструирование нРНК для генов пептидного мечения в эндогенных геномных локусах
5'
протоспейсер
каркас нРНК
гомология
повтор
отжиг
3'
ген-мишень
кодирует метку и последующую гомологию
связывает праймер для инициирования обратной транскрипции
Пептидное мечение с помощью геномного редактирования на основа TPRT
кодирующий белок ген
геномная ДНК
Предст. интерес ген
редактирование на основе мишень-инициированной обратной транскрипции
кодируемый направляющей РНК синтез последовательности пептидной метки
пептидная метка
гомология флэпа
Предст. интерес ген
репарация ДНК
Предст. интерес ген
пептидная метка
меченый эндогенный ген
Возможные пептидные метки
Гексагистидиновая метка (His6)
FLAG-метка
V5-метка
GCN4-метка
HA-метка
Myc-метка
FlAsH/ReAsH-метка
Сортазный субстрат
π-клэмп
Возможные варианты использования
флуоресцентное мечение белка
иммуноосаждение
иммуноблоттинг
иммуногистохимия
рекрутирование белка
индуцируемые белковые дегроны
скрининг в рамках всего генома

ФИГ.25

(SEQ ID NO: 1361647-1361651)

Участок HEK3, Вставка His-метки



жирный шрифт замещения

вставки

\- делеции

--- предполагаемое положение расщепления

(SEQ ID NO: 1361652-1361655)

Участок HEK3, Вставка FLAG-метки

жирный шрифт замещения

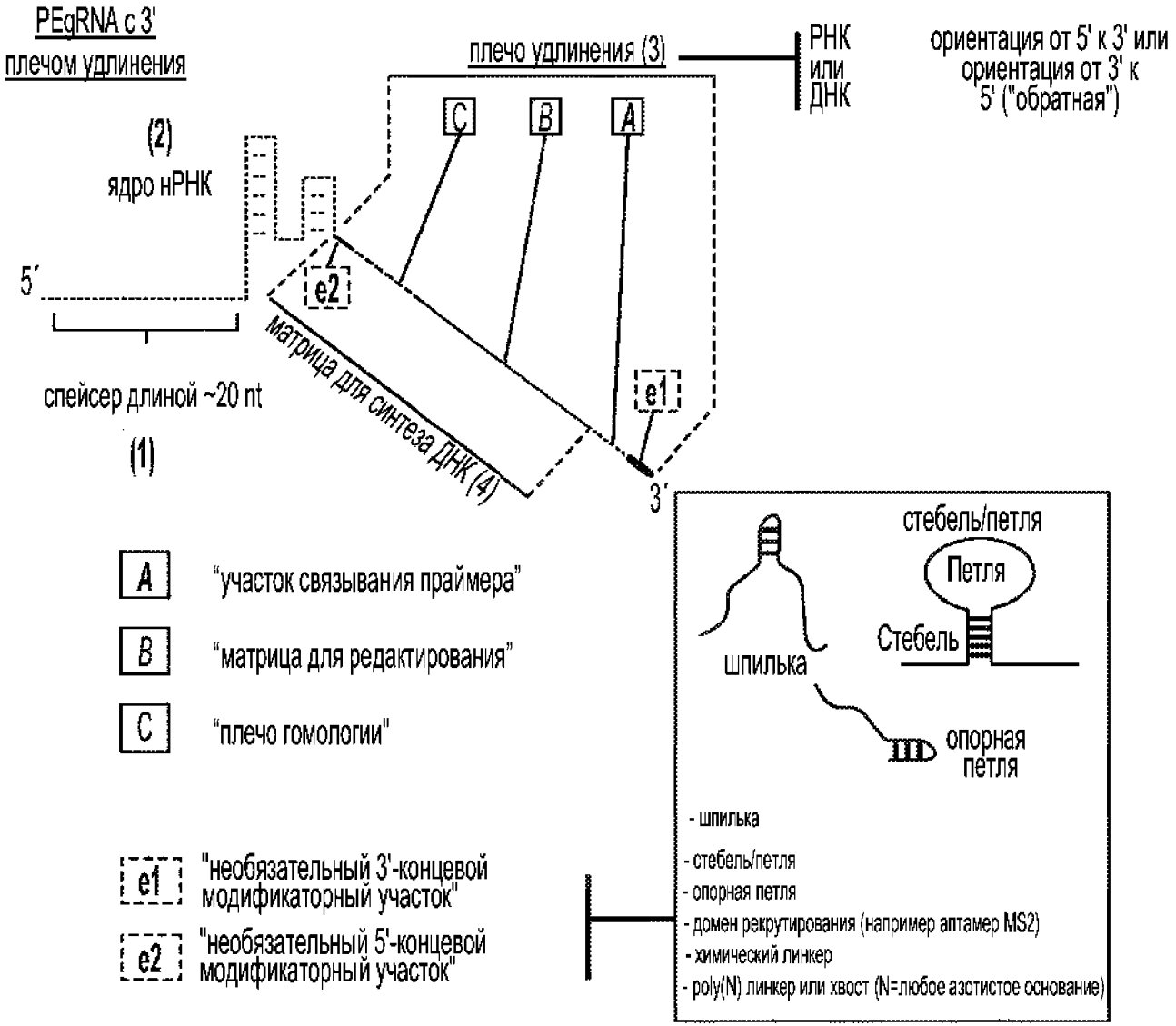
вставки

\- делеции

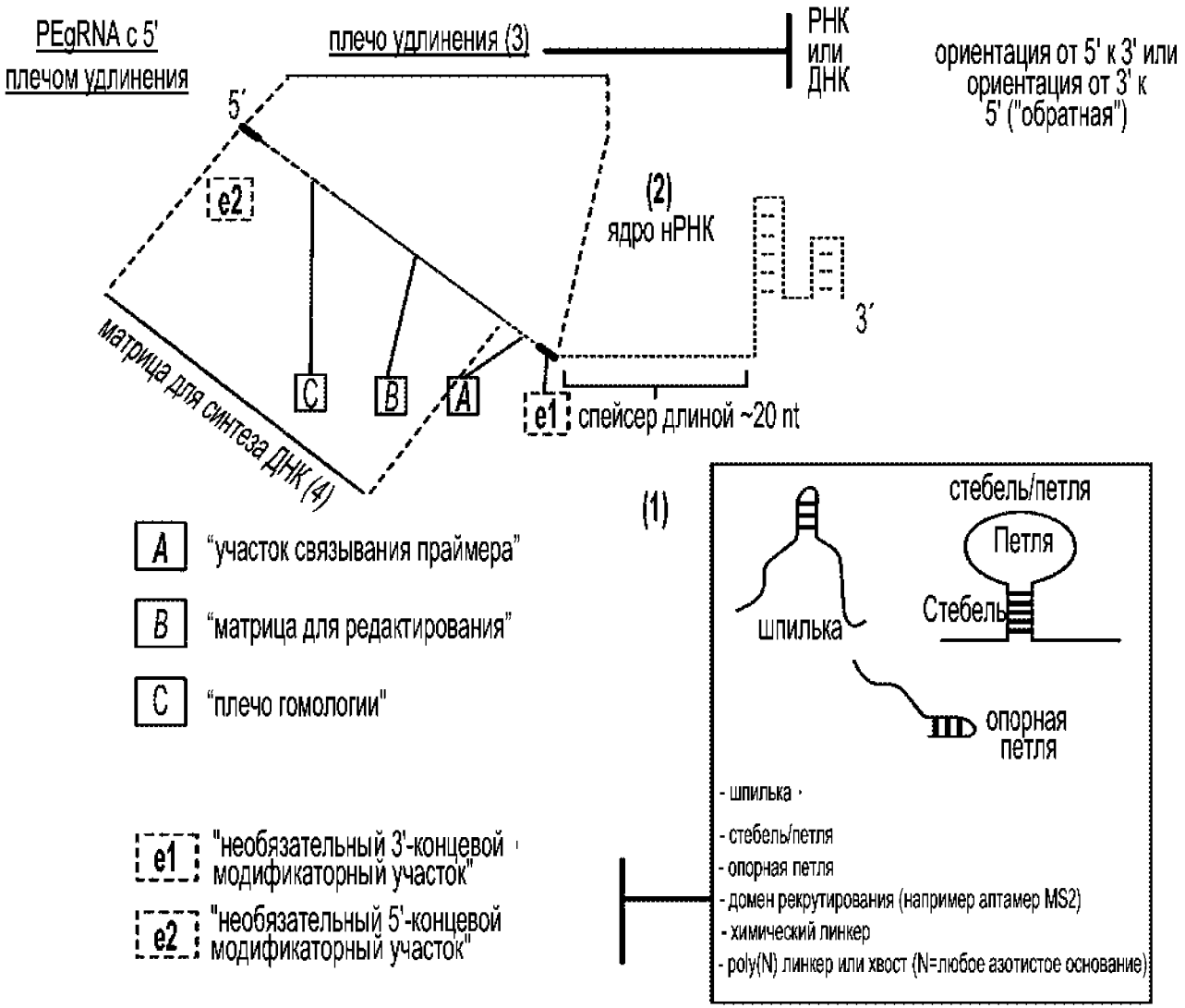
--- предполагаемое положение расщепления

след. обратить внимание, что вставка длиной 24 nt не показана

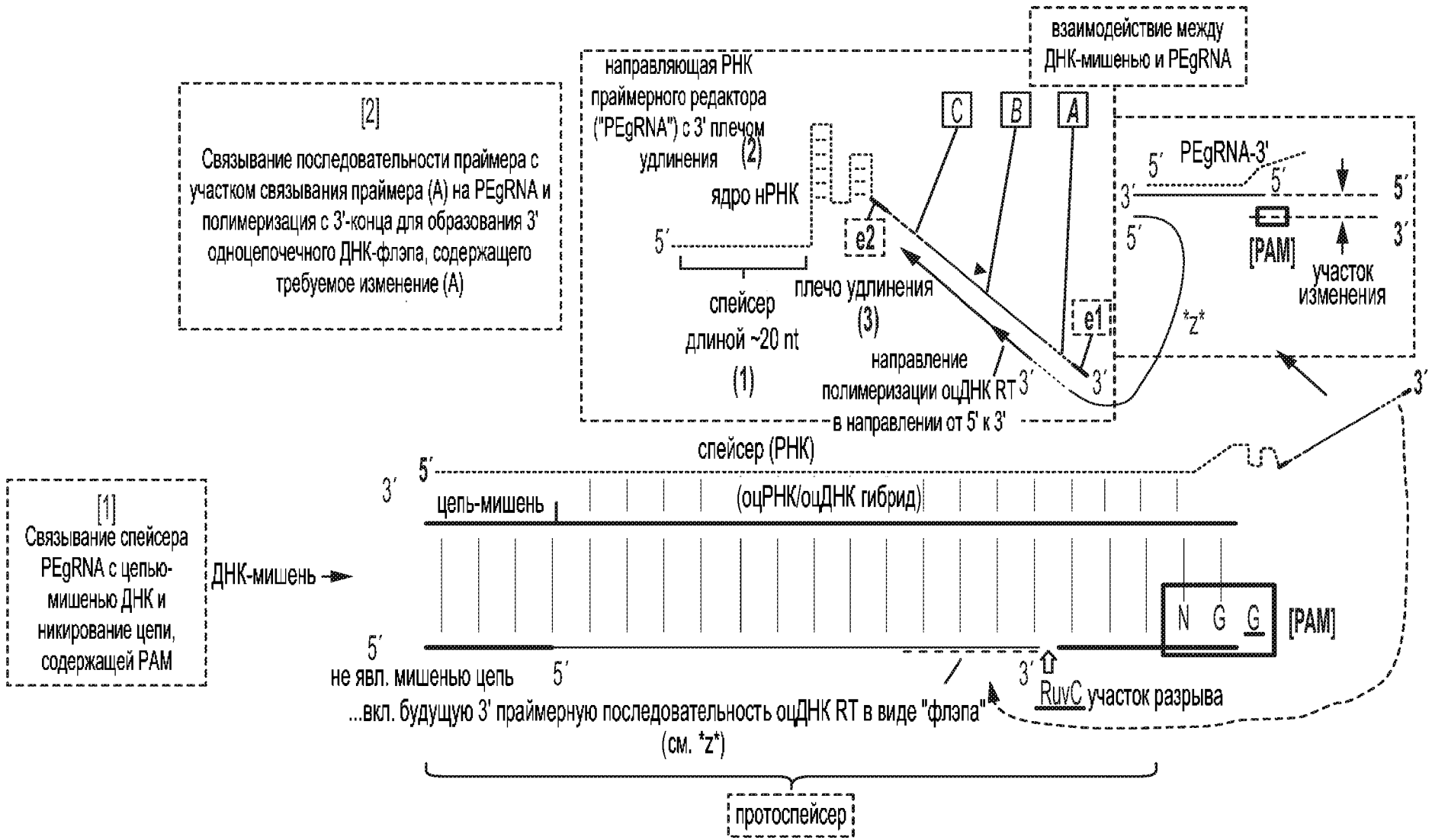
ФИГ.26


PEgRNA с 3'
плечом удлинения
(2)
ядро нРНК
5'
спейсер длиной ~20 nt
(1)
плечо удлинения (3)
РНК
или
ДНК
ориентация от 5' к 3' или
ориентация от 3' к
5' ("обратная")
C
B
A
e2
e1
матрица для синтеза ДНК (4)
3'
A
"участок связывания праймера"
B
"матрица для редактирования"
C
"плечо гомологии"
e1
"необязательный 3'-концевой
модификаторный участок"
e2
"необязательный 5'-концевой
модификаторный участок"
стебель/петля
Петля
Стебель
шпилька
опорная
петля
- шпилька
- стебель/петля
- опорная петля
- домен рекрутирования (например аптамер MS2)
- химический линкер
- poly(N) линкер или хвост (N=любое азотистое основание)

ФИГ.27


PEgRNA с 5' плечом удлинения
плечо удлинения (3)
РНК или ДНК
ориентация от 5' к 3' или ориентация от 3' к 5' ("обратная")
5'
e2
(2) ядро нРНК
3'
матрица для синтеза ДНК (4)
C
B
A
e1
спейсер длиной ~20 nt
A "участок связывания праймера"
B "матрица для редактирования"
C "плечо гомологии"
e1 "необязательный 3'-концевой модификаторный участок"
e2 "необязательный 5'-концевой модификаторный участок"
(1)
стебель/петля
Петля
Стебель
шпилька
опорная петля
- шпилька
- стебель/петля
- опорная петля
- домен рекрутирования (например аптамер MS2)
- химический линкер
- poly(N) линкер или хвост (N=любое азотистое основание)

ФИГ.28


[2]
Связывание последовательности праймера с участком связывания праймера (A) на PEgRNA и полимеризация с 3'-конца для образования 3' одноцепочечного ДНК-флэпа, содержащего требуемое изменение (A)
направляющая РНК праймерного редактора ("PEgRNA") с 3' плечом удлинения (2)
ядро нРНК
спейсер длиной ~20 nt (1)
e2
плечо удлинения (3)
C
B
A
e1
направление полимеризации оцДНК RT
в направлении от 5' к 3'
взаимодействие между ДНК-мишенью и PEgRNA
PEgRNA-3'
[PAM]
участок изменения
z
[1]
Связывание спейсера PEgRNA с цепью-мишенью ДНК и никирование цепи, содержащей PAM
ДНК-мишень
спейсер (РНК)
цепь-мишень
(оцРНК/оцДНК гибрид)
N G G
[PAM]
не явл. мишенью цепь
RuvC участок разрыва
...вкл. будущую 3' праймерную последовательность оцДНК RT в виде "флэпа"
(см. *z*)
протоспейсер

ФИГ.29

Пример 1 - SpCas9 (NGG)

CCGGGGCTGCCCTGTGAGGG*G*TTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTAT

Полная нРНК (спейсер + ядро + плечо удлинения) (SEQ ID NO: 135529)

CCGGGGCTGCCCTGTGAGGG

Спейсер (SEQ ID NO: 271043)

Пример 2 - SpCas9 (NGG)

GTGAAGAATTGGCTTATACA*GTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTAT*

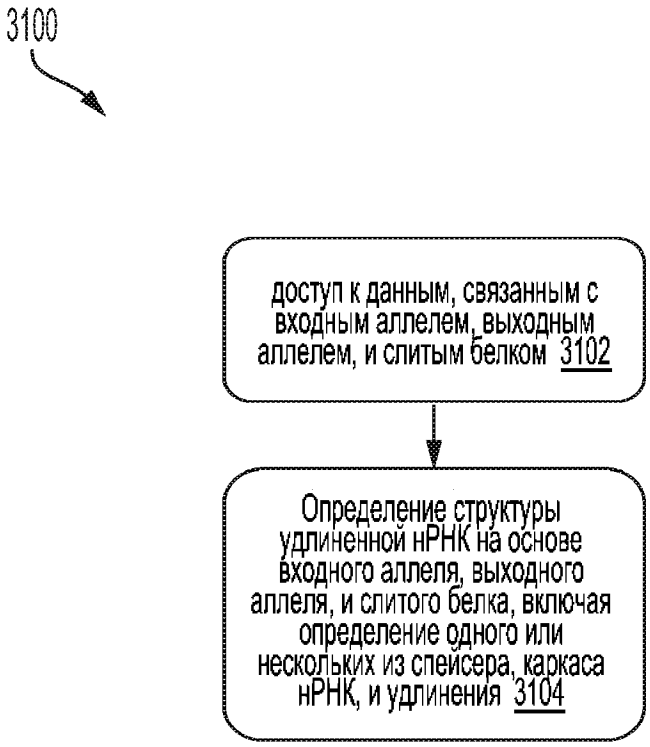
Полная нРНК (спейсер + ядро + плечо удлинения) (SEQ ID NO: 135880)

GTGAAGAATTGGCTTATACA

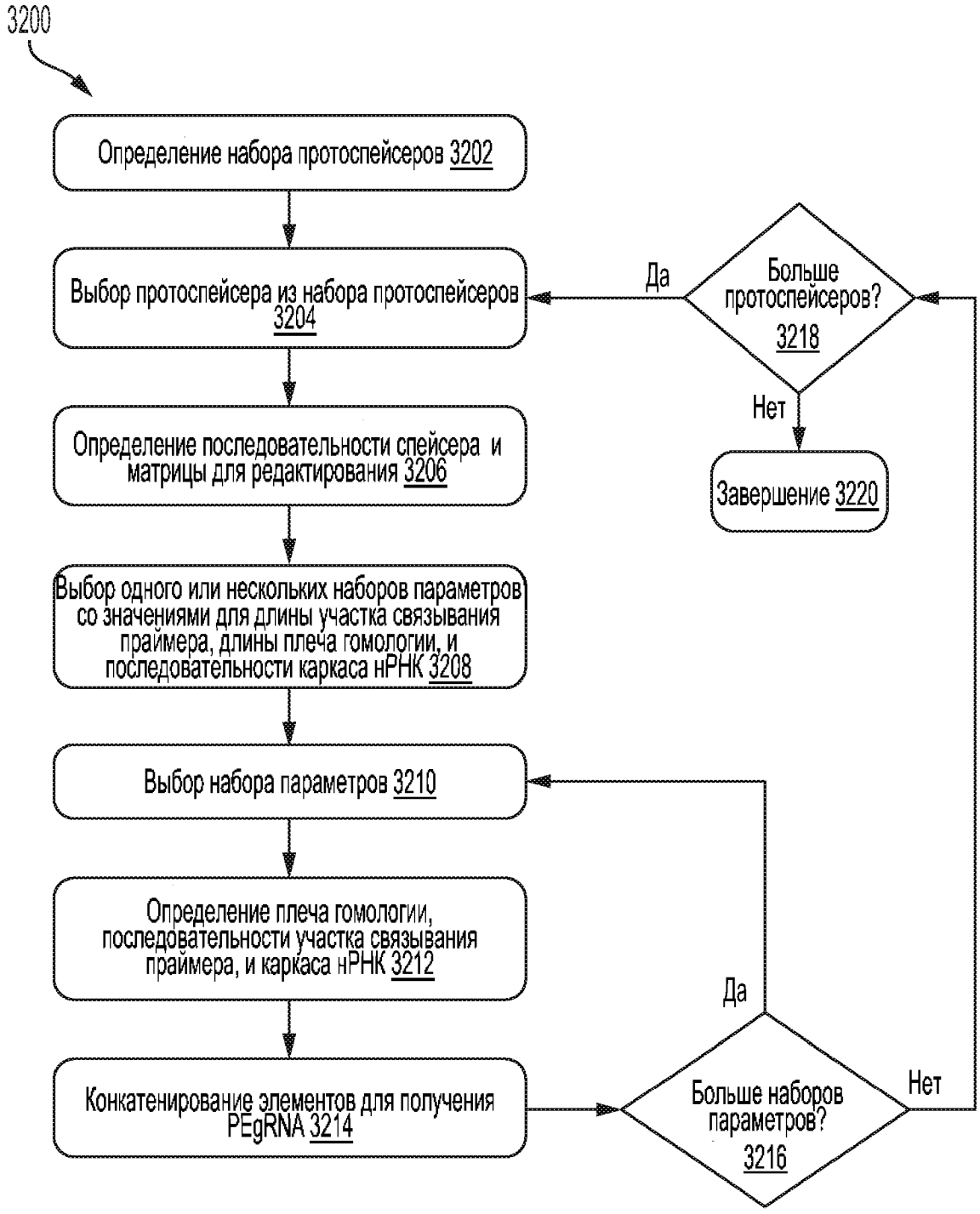
Спейсер (SEQ ID NO: 880463)

ФИГ.30

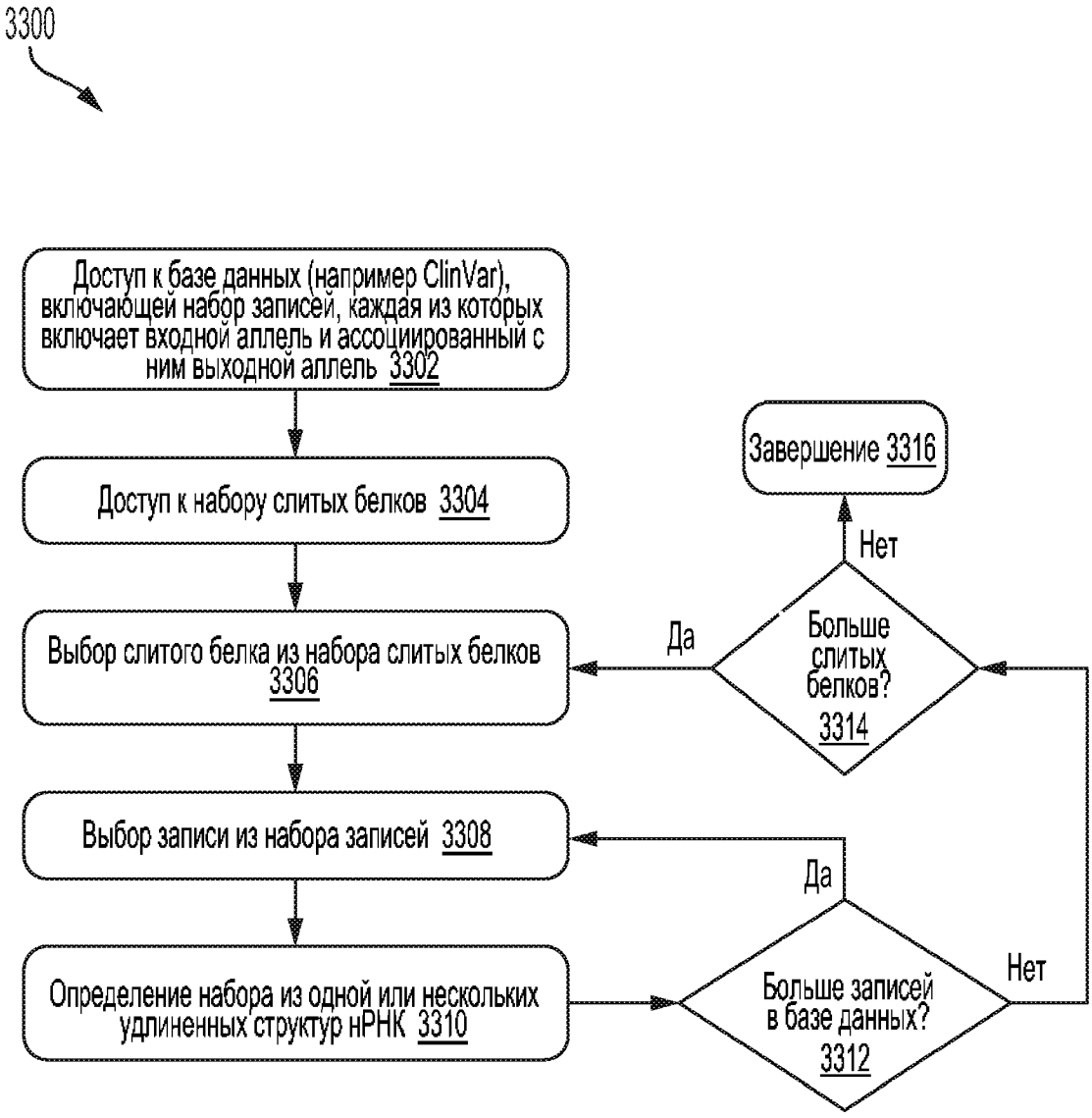
ФИГ.30
продолжение


3100
доступ к данным, связанным с входным аллелем, выходным аллелем, и слитым белком 3102
Определение структуры удлиненной нРНК на основе входного аллеля, выходного аллеля, и слитого белка, включая определение одного или нескольких из спейсера, каркаса нРНК, и удлинения 3104

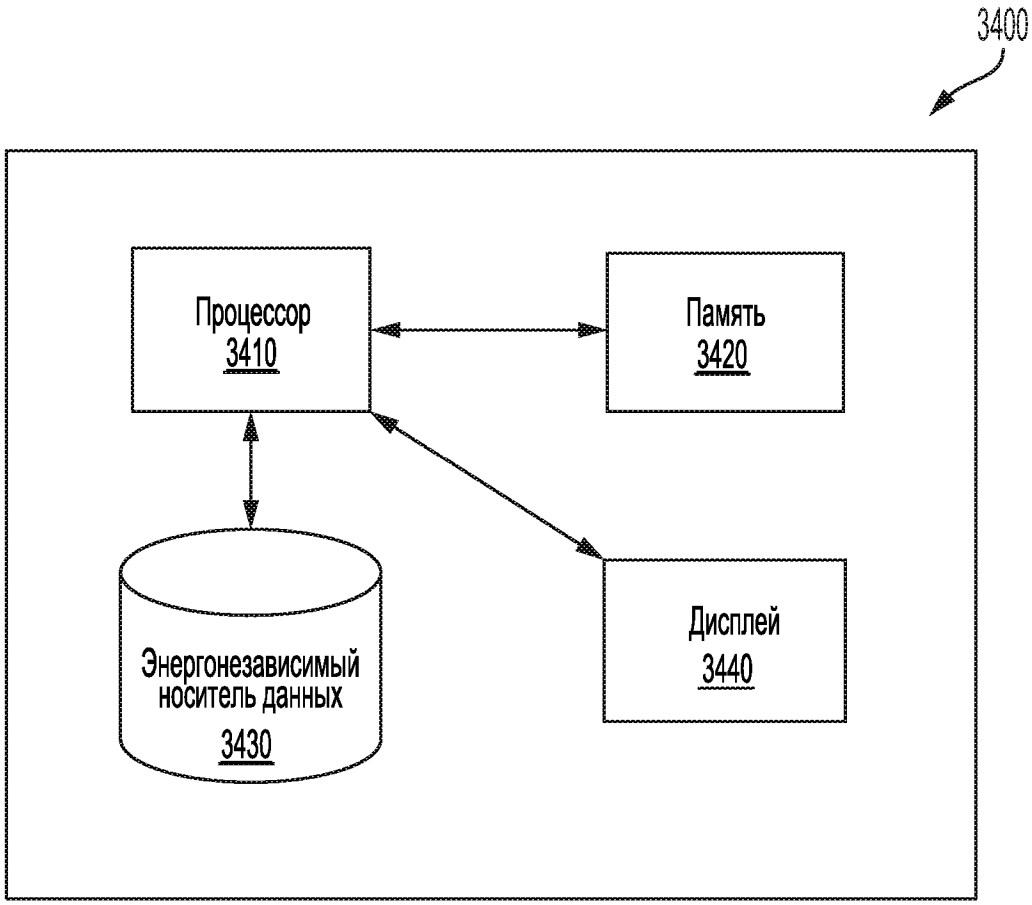
ФИГ.31


3200
Определение набора протоспейсеров 3202
Выбор протоспейсера из набора протоспейсеров 3204
Определение последовательности спейсера и матрицы для редактирования 3206
Выбор одного или нескольких наборов параметров со значениями для длины участка связывания праймера, длины плеча гомологии, и последовательности каркаса нРНК 3208
Выбор набора параметров 3210
Определение плеча гомологии, последовательности участка связывания праймера, и каркаса нРНК 3212
Конкатенирование элементов для получения PEgRNA 3214
Больше наборов параметров? 3216
Да
Нет
Больше протоспейсеров? 3218
Да
Нет
Завершение 3220

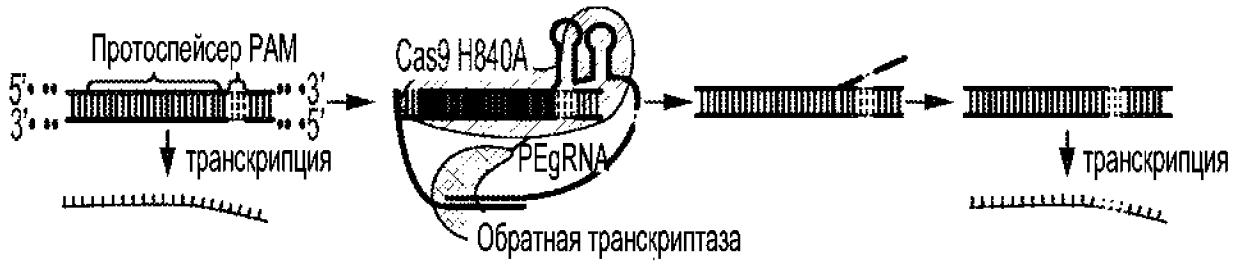
ФИГ.32


3300
Доступ к базе данных (например ClinVar), включающей набор записей, каждая из которых включает входной аллель и ассоциированный с ним выходной аллель 3302
Доступ к набору слитых белков 3304
Выбор слитого белка из набора слитых белков 3306
Выбор записи из набора записей 3308
Определение набора из одной или нескольких удлиненных структур нРНК 3310
Больше записей в базе данных? 3312
Да
Нет
Больше слитых белков? 3314
Да
Нет
Завершение 3316

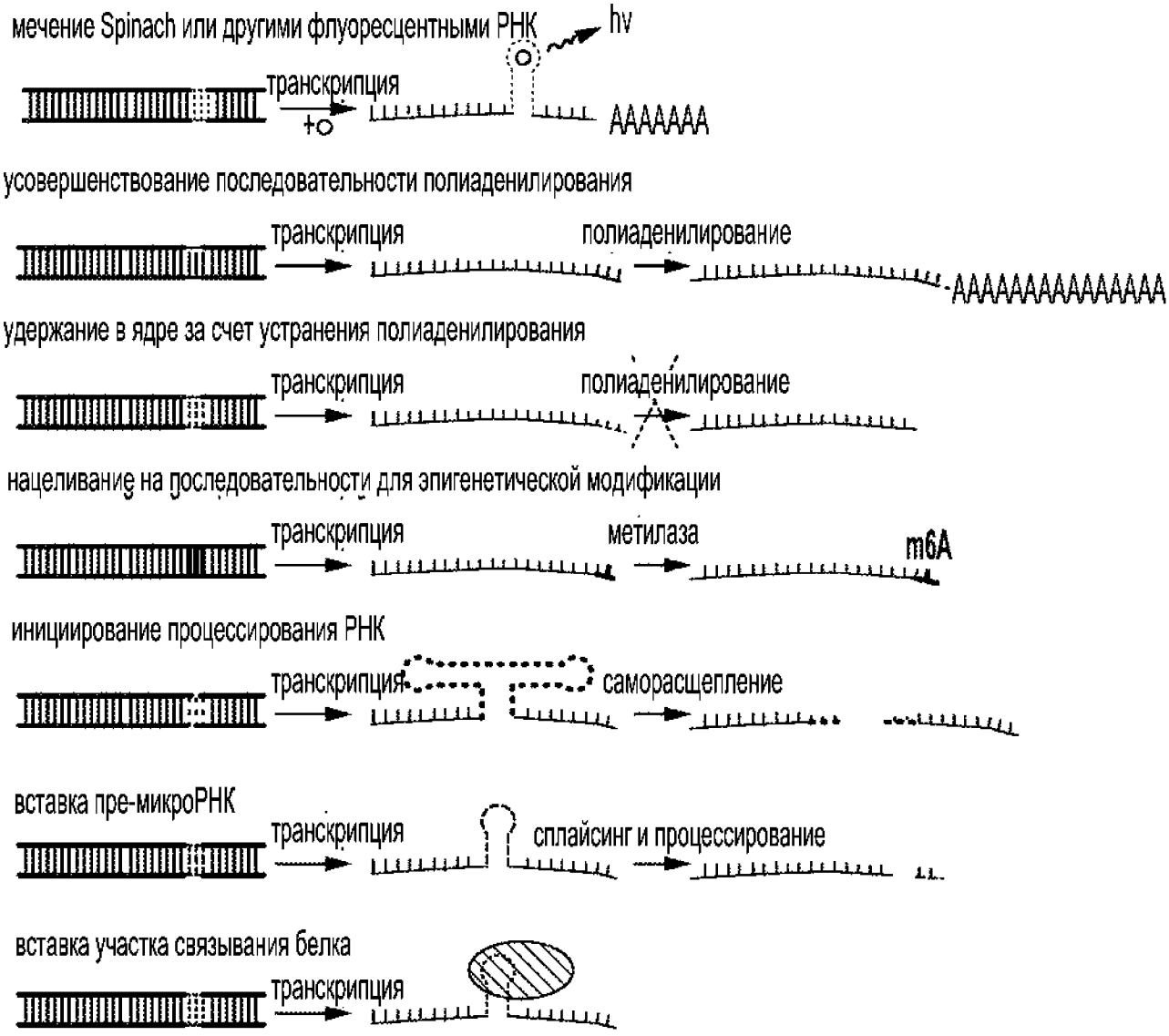
ФИГ.33


3400
Процессор
3410
Память
3420
Энергонезависимый
носитель данных
3430
Дисплей
3440

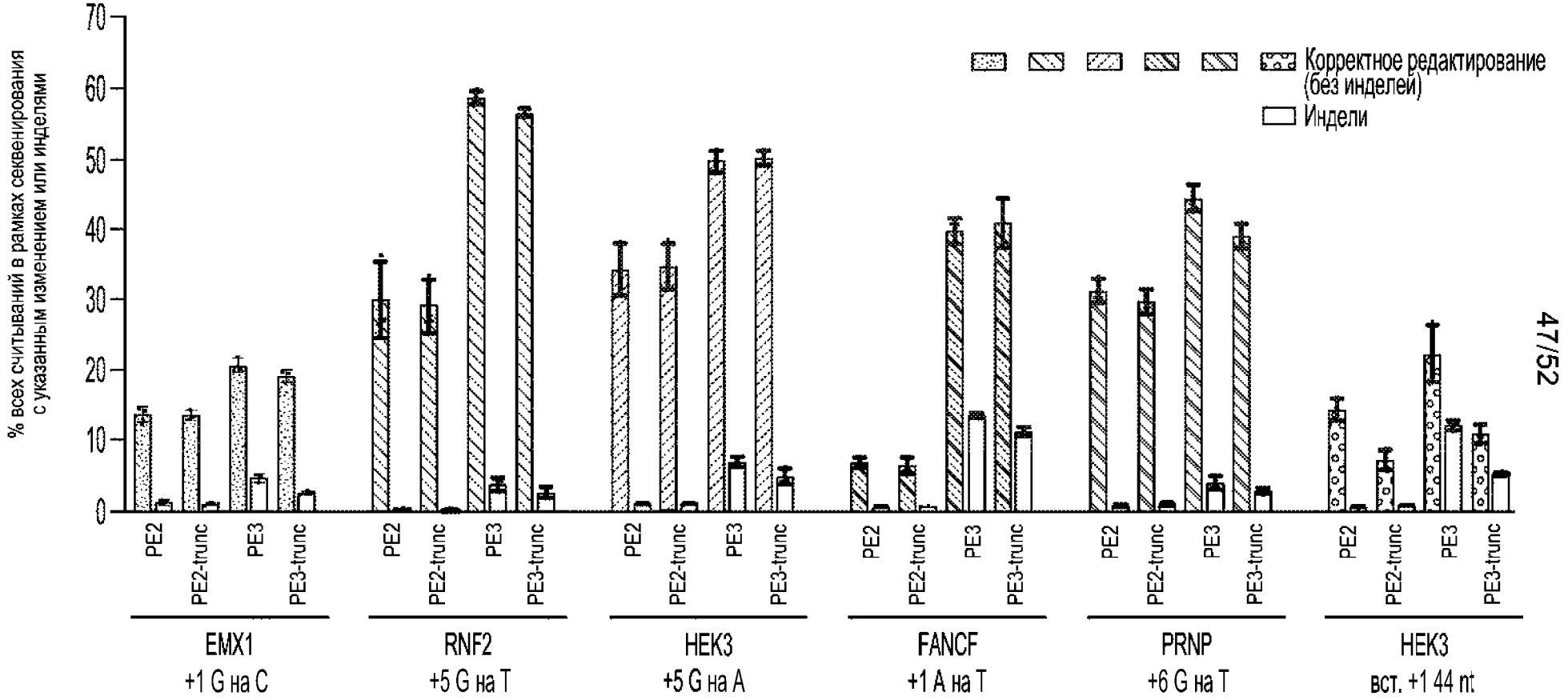
ФИГ.34



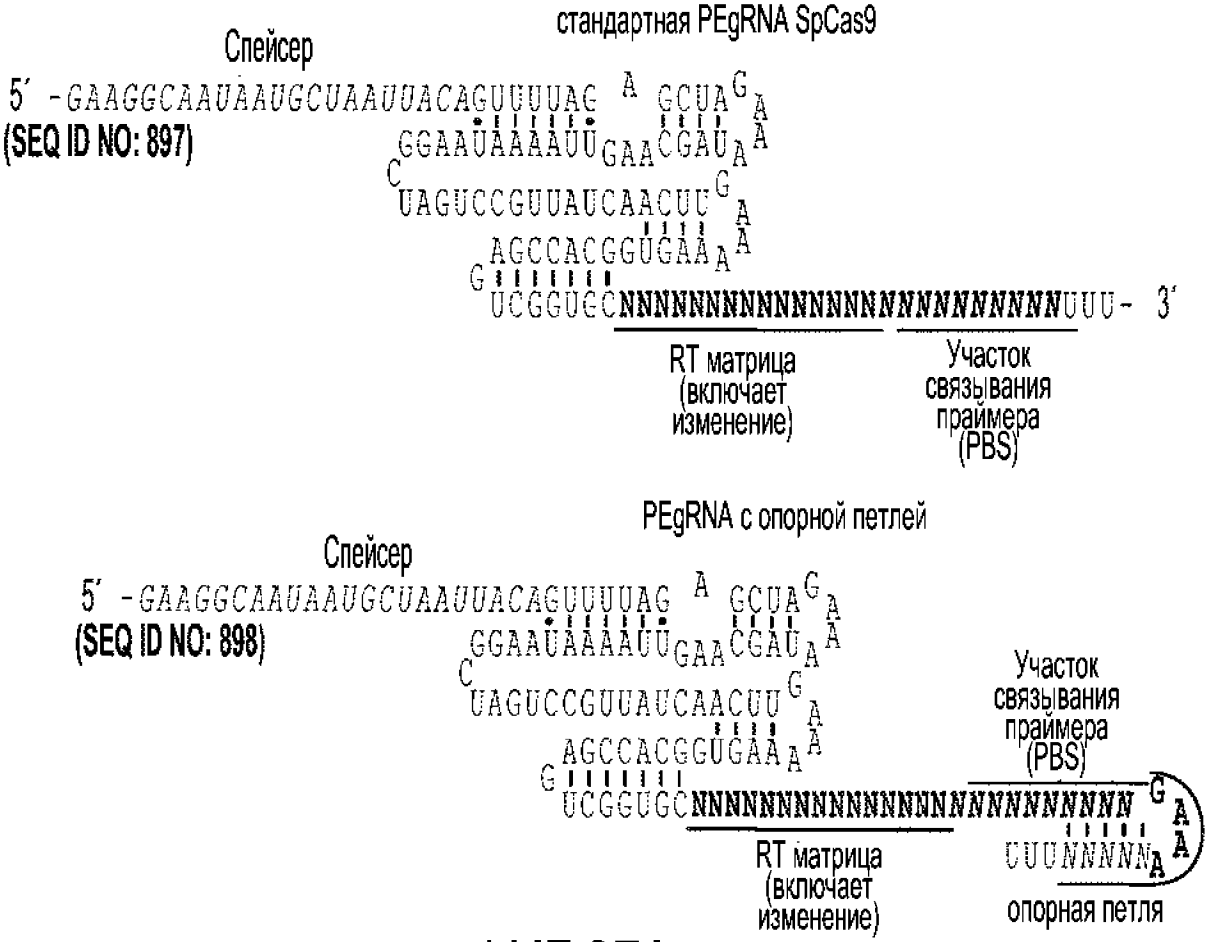
ФИГ.35A



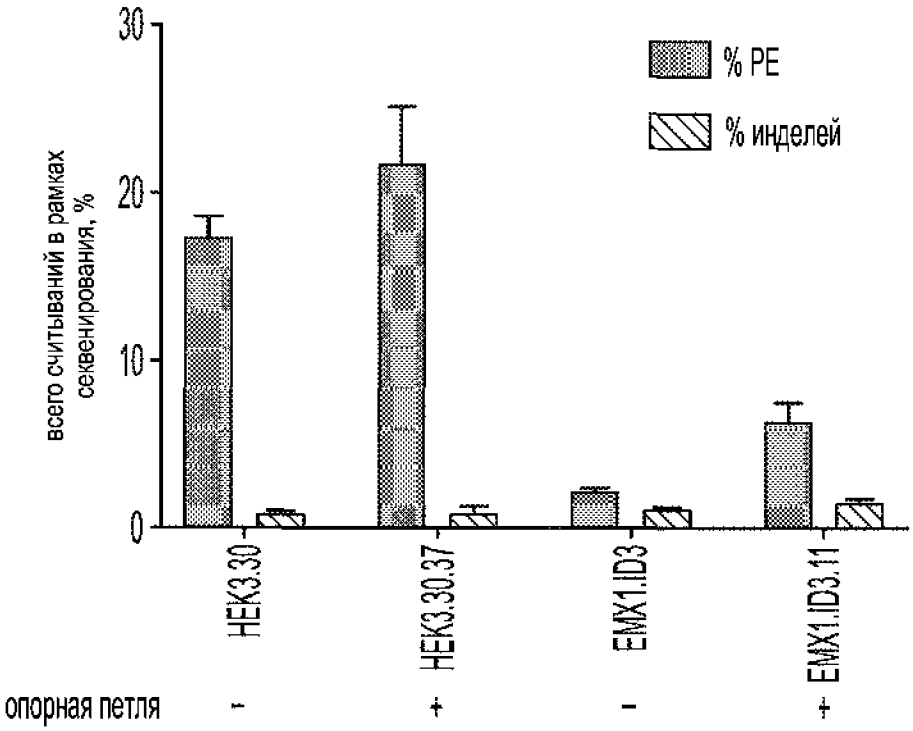
ФИГ.35B


% всех считываний в рамках секвенирования с указанным изменением или инделями
70
60
50
40
30
20
10
0
Корректное редактирование (без инделей)
Индели
PE2
PE2-trunc
PE3
PE3-trunc
EMX1
+1 G на C
RNF2
+5 G на T
HEK3
+5 G на A
FANCF
+1 A на T
PRNP
+6 G на T
HEK3
вст. +1 44 пт

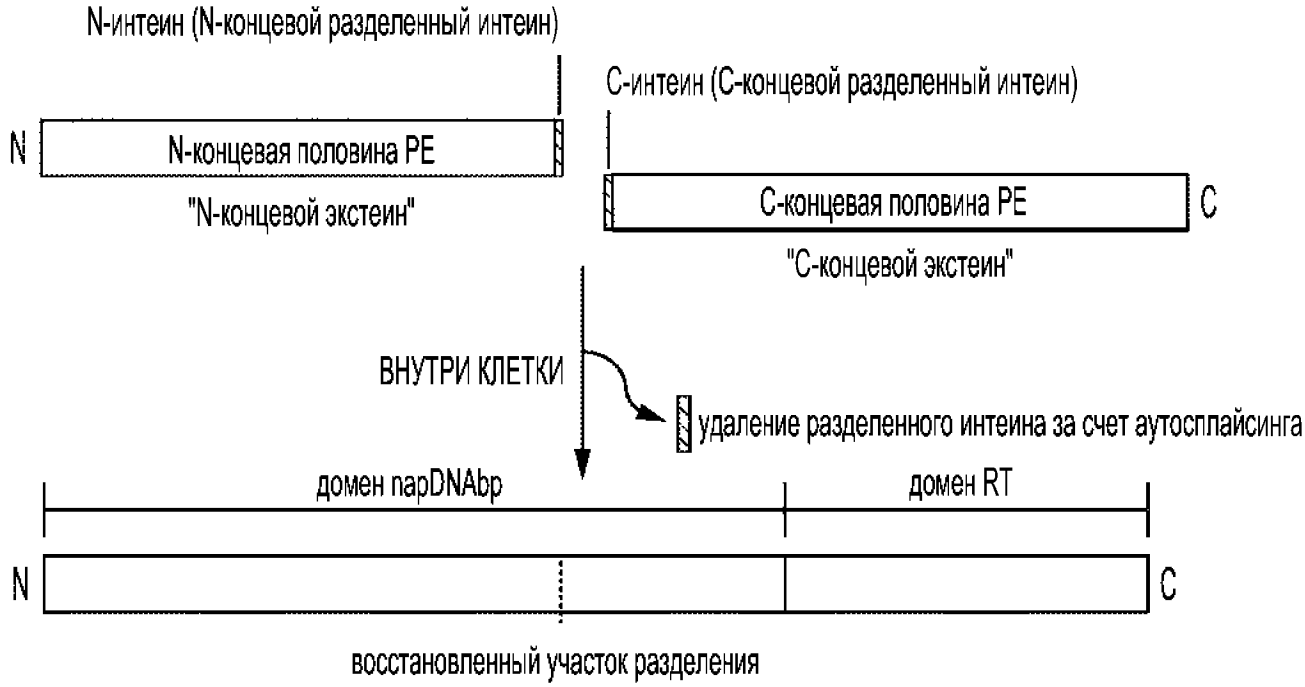
ФИГ.36


стандартная PEgRNA SpCas9
Спейсер
5′ -GAAGGCAAUAAUGCUAAUUACAGUUUUAG A GCUA G A
(SEQ ID NO: 897)
GGAAUAAAAUU GAA CGAU A A
C
UAGUCCGUUAUCAACUU G A
AGCCACGGUGAA A A
G
UCGGUGCNNNNNNNNNNNNNNNNNNNNNNUUU- 3′
RT матрица (включает изменение)
Участок связывания праймера (PBS)
PEgRNA с опорной петлей
Спейсер
5′ -GAAGGCAAUAAUGCUAAUUACAGUUUUAG A GCUA G A
(SEQ ID NO: 898)
GGAAUAAAAUU GAA CGAU A A
C
UAGUCCGUUAUCAACUU G A
AGCCACGGUGAA A A
G
UCGGUGCNNNNNNNNNNNNNNNNNNNN G A
UUUNNNNN A A
Участок связывания праймера (PBS)
RT матрица (включает изменение)
опорная петля

ФИГ.37А

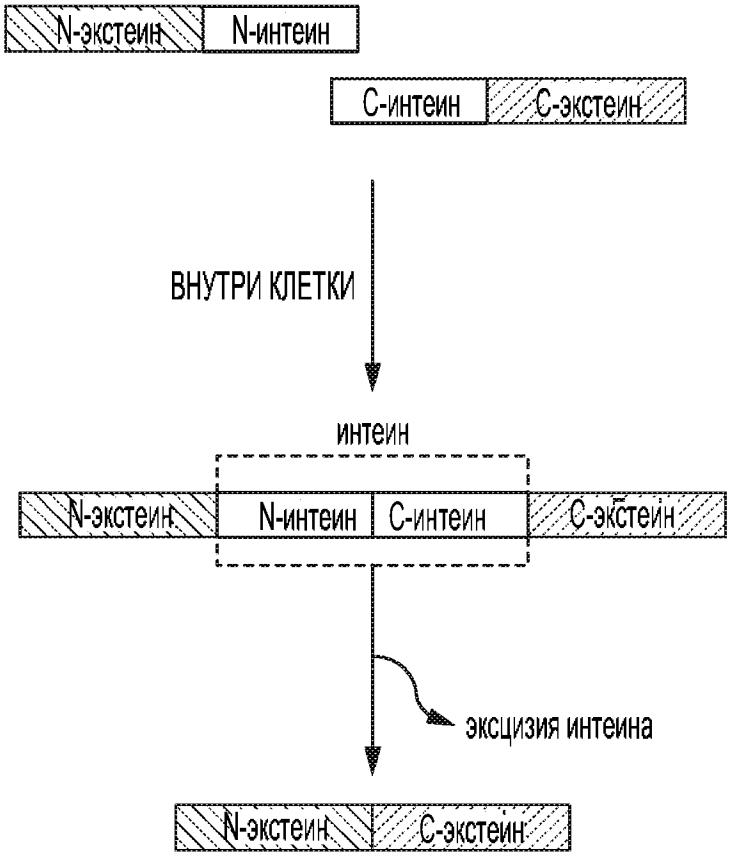

30
20
10
0
всего считываний в рамках секвенирования, %
% PE
% инделей
HEK3.30
HEK3.30.37
EMX1.ID3
EMX1.ID3.11
опорная петля
-
+
-
+

ФИГ.37B

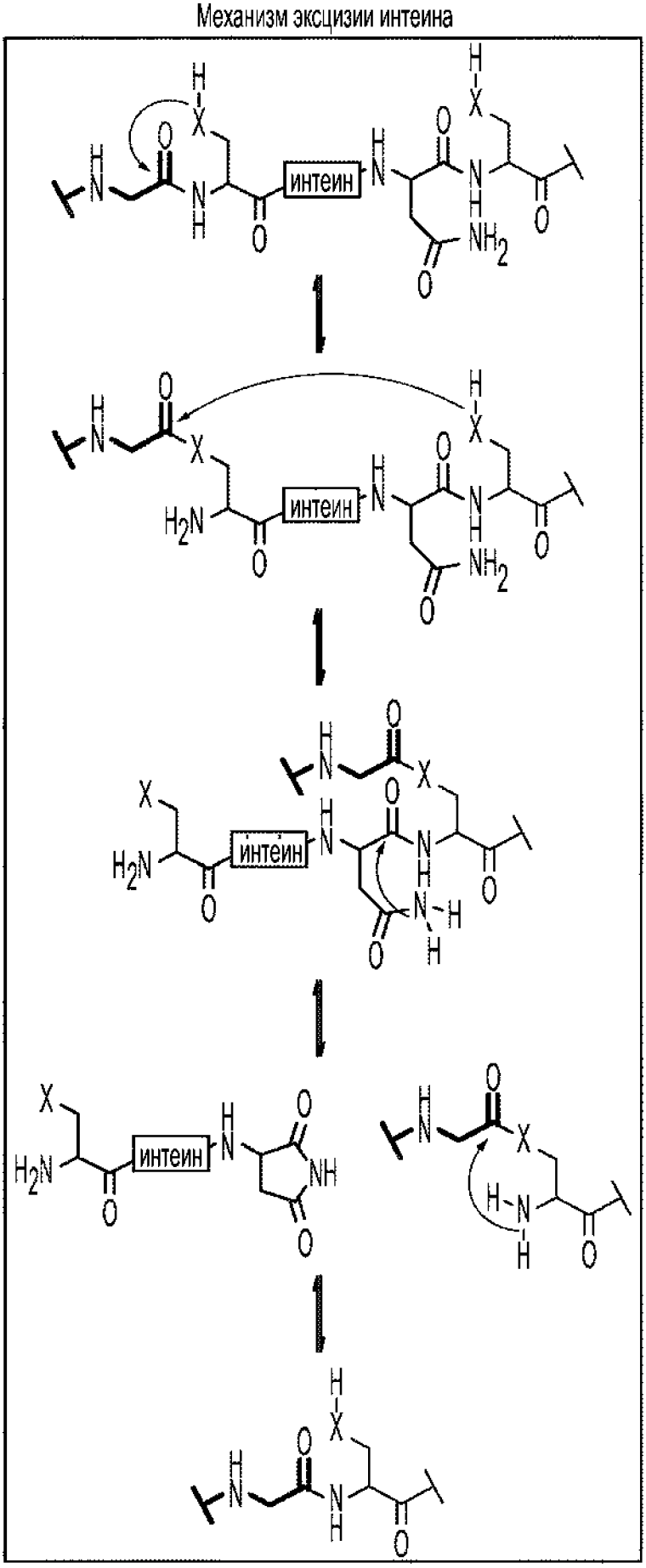


Праймерный редактор

ФИГ.38


N-экстеин
N-интеин
C-интеин
C-экстеин
ВНУТРИ КЛЕТКИ
интеин
N-экстеин
N-интеин
C-интеин
C-экстеин
эксцизия интеина
N-экстеин
C-экстеин

ФИГ.39А



ФИГ.39В