

(19)



**Евразийское
патентное
ведомство**

(11) **046214**

(13) **B1**

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ

(45) Дата публикации и выдачи патента
2024.02.16

(21) Номер заявки
201890565

(22) Дата подачи заявки
2016.08.25

(51) Int. Cl. **G01N 33/53** (2006.01)
G01N 33/543 (2006.01)
G01N 33/553 (2006.01)

(54) КОМПОЗИЦИИ И СПОСОБЫ УЛУЧШЕНИЯ СПЕЦИФИЧНОСТИ В ГЕНОМНОЙ ИНЖЕНЕРИИ С ПРИМЕНЕНИЕМ РНК-НАПРАВЛЯЕМЫХ ЭНДОНУКЛЕАЗ

(31) 62/209,466

(32) 2015.08.25

(33) US

(43) 2019.04.30

(86) PCT/US2016/048798

(87) WO 2017/035416 2017.03.02

(71)(73) Заявитель и патентовладелец:
ДБЮК ЮНИВЕРСИТИ (US)

(72) Изобретатель:
**Джозефс Эрик, Коцак Девран,
Маршалек Петр, Герсбах Чарльз А.
(US)**

(74) Представитель:
Медведев В.Н. (RU)

(56) WO-A2-2014197748
US-A1-20140068797
US-A1-20140295557
WO-A1-2014065596
JOSEPHS et al.: "Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage," Nucleic Acids Research, 17 September 2015 (17.09.2015), Vol. 43, No. 18, Pgs. 8924-8941, entire document

(57) В данном изобретении раскрываются оптимизированные направляющие РНК (gRNA) и способы конструирования и применения указанных оптимизированных gRNA, которые характеризуются повышенной специфичностью связывания с мишенью и уменьшенным нецелевым связыванием.

B1

046214

**046214
B1**

Перекрестная ссылка на родственные заявки

Данная заявка заявляет приоритет по предварительной заявке на патент США № 62/209466, поданной 25 августа 2015 г., которая включена в данный документ посредством ссылки во всей своей полноте.

Заявление о государственной заинтересованности

Настоящее изобретение было создано при государственной поддержке в соответствии с Федеральными грантами № MCB1244297 и CBET1151035, выданными Национальным научным фондом (National Science Foundation), и F32GM11250201, R01DA036865 и DP2OD008586, выданными Национальными институтами здравоохранения (National Institutes of Health). Правительство обладает определенными правами на настоящее изобретение.

Область техники

Настоящее изобретение направлено на оптимизированные направляющие РНК (gRNA) и способы конструирования и применения указанных gRNA, которые характеризуются повышенной специфичностью связывания с мишенью и уменьшенным нецелевым связыванием.

Предпосылки изобретения

РНК-направляемые эндонуклеазы, в частности белок Cas9, привлекли внимание в качестве потенциального "идеального инструмента для геномной инженерии", поскольку их можно направлять с помощью одиночной молекулы "направляющей РНК", чтобы вырезать ДНК практически с любой последовательностью. Недавно эту способность использовали для ряда перспективных биологических и медицинских применений, порождая беспрецедентное воодушевление и надежду на их будущее использование. Однако на практике геномная инженерия требует чрезвычайно точного контроля над способностью избирательно нацеливаться и вырезать точные последовательности ДНК, чтобы нецелевая ДНК не была случайно повреждена и не подверглась мутации.

Cas9 представляет собой эндонуклеазу прокариотического CRISPR (короткие палиндромные повторы, регулярно расположенные группами) типа II - CRISPR-ассоциированного (Cas) ответа на встраиваемую чужеродную ДНК. Во время этого ответа Cas9 сначала связывается дуплексом РНК CRISPR (crRNA): trans-активирующая crRNA (tracrRNA), а затем направляется для расщепления ДНК, которая содержит сайты "протоспейсера" размером 20 пар оснований (п.о.), комплементарные вариабельному сегменту размером 20 п.о. из crRNA (фиг. 1А). При связывании одиночной направляющей РНК (sgRNA) комплекс Cas9-sgRNA связывается с последовательностями "протоспейсера" размером 20 п.о. в целевой ДНК при условии, что за протоспейсером непосредственно следует прилегающий к протоспейсеру мотив (PAM, в данном документе "TGG"). После связывания эндонуклеаза Cas9 осуществляет двухнитевые разрывы (треугольники) в протоспейсере. По сути, единственным ограничением для последовательностей, на которые может целенаправленно воздействовать Cas9, является то, что короткий прилегающий к протоспейсеру мотив (PAM), такой как "NGG" в случае Cas9 *S. pyogenes*, должен следовать непосредственно за сайтами протоспейсера в молекуле чужеродной ДНК. Анализ в ходе проведения кристаллографических и биохимических экспериментов показывает, что специфичность связывания и расщепления протоспейсера достигается сначала путем распознавания сайтов PAM собственно белком Cas9 с последующим встраиванием нити связанным комплексом РНК и прямым спариванием оснований по Уотсону-Крику с протоспейсером (фиг. 1А).

Способность Cas9 модульно "программироваться" с помощью одиночной РНК-шпильки для целенаправленного воздействия на практически любой сайт ДНК породила необычайный энтузиазм после того, как системы CRISPR-Cas9 повторно адаптировали для ряда разнообразных биотехнологических применений. В частности, была сконструирована одиночная направляющая РНК (sgRNA)-шпилька, которая объединяет ключевые компоненты дуплексов crRNA: tracrRNA в одиночные функциональные молекулы. С помощью этой sgRNA Cas9 можно вводить в различные организмы с получением целенаправленных двухнитевых разрывов *in vivo* для проведения масштабной и в то же время простой геномной инженерии. Cas9 без нуклеазной активности (D10A/H840A, известную как "dCas9") и химерные производные dCas9 также использовали для изменения экспрессии генов путем целенаправленного связывания на промоторных сайтах или рядом с ними *in vivo*, а также для введения целенаправленных эпигенетических модификаций.

Нецелевое связывание и расщепление с помощью Cas9 является проблемой, поскольку оно может отрицательно повлиять на его потенциальное использование на практике. Значительные усилия были предприняты для улучшения специфичности активности Cas9/dCas9. Во-первых, наиболее масштабные усилия в значительной степени проводятся за счет интеллектуального выбора последовательностей-мишеней без аналогичных других последовательностей в геноме, хотя недавнее исследование показало, что эти способы показали неудовлетворительные результаты в их способности прогнозировать нецелевое расщепление. Кроме того, также были предприняты усилия по непосредственному конструированию самого белка путем введения точечных мутаций, которые, как было установлено, модулируют или увеличивают специфичность связывания PAM или протоспейсера. Производные Cas9, которые вызывают разрывы только одной нити ДНК, а не осуществляют расщепление двухнитевой ДНК, также используют парами ("парные никазы") с допущением, что вероятность нецелевого однострессового разрыва в нескольких сайтах, достаточно близко расположенных относительно друг к другу, для получения двунитевого

разрыва, будет чрезвычайно редкой. И наконец, был проведен ряд работ по созданию самих вариантов направляющей РНК в попытке достичь большей специфичности. Прежние усилия, когда к направляющим РНК добавляли 5'-удлинения, чтобы дополнить дополнительные нуклеотиды за протоспейсером, не показали повышенной специфичности расщепления Cas9 *in vivo*. Скорее всего, они расщеплялись приблизительно до своей стандартной длины в живых клетках (фиг. 1А). Для применений в геномной инженерии, особенно для применений с целью терапии, требуется исключительная специфичность при целенаправленном воздействии на ген, чтобы не повреждать нецелевую ДНК и не допускать запрещенных мутаций. Тем не менее, существует ряд сообщений о нецелевом связывании и расщеплении с помощью Cas9, что может отрицательно повлиять на ее потенциальное использование на практике.

Остается потребность в уменьшении нецелевого связывания и повышении специфичности нуклеазы при использовании системы CRISPR/Cas9.

Краткое описание изобретения

Настоящее изобретение направлено на способ получения оптимизированной направляющей РНК (gRNA). Способ предусматривает: а) идентификацию представляющей интерес целевой области, при этом представляющая интерес целевая область содержит последовательность протоспейсера; б) определение полинуклеотидной последовательности полноразмерной gRNA, которая нацеливается на представляющую интерес целевую область, при этом полноразмерная gRNA содержит нацеливающиеся на протоспейсер последовательность или сегмент; в) определение по меньшей мере одного или нескольких нецелевых сайтов для полноразмерной gRNA; г) получение полинуклеотидной последовательности первой gRNA, при этом первая gRNA содержит полинуклеотидную последовательность полноразмерной gRNA и сегмент РНК, при этом сегмент РНК содержит полинуклеотидную последовательность, имеющую М нуклеотидов в длину, которая комплементарна нуклеотидному сегменту нацеливающихся на протоспейсер последовательности или сегменту, при этом РНК расположена на 5'-конце полинуклеотидной последовательности полноразмерной gRNA, при этом первая gRNA необязательно содержит линкер между 5'-концом полинуклеотидной последовательности полноразмерной gRNA и сегментом РНК, при этом линкер содержит полинуклеотидную последовательность, имеющую N нуклеотидов в длину, при этом первая gRNA способна встраиваться в последовательность протоспейсера, и связываться с последовательностью ДНК, которая комплементарна последовательности протоспейсера, и образовывать дуплекс с протоспейсером, и при этом первая gRNA способна встраиваться в нецелевой сайт, и связываться с последовательностью ДНК, которая комплементарна нецелевому сайту, и образовывать нецелевой дуплекс; е) вычисление оценочного показателя или вычислительное моделирование кинетики встраивания и времени нахождения, в течение которого gRNA остается встроенной в дуплексы протоспейсера и нецелевого сайта, где динамику встраивания оценивают нуклеотид за нуклеотидом путем определения энергетических различий между дополнительным встраиванием отличающейся gRNA и повторным отжигом первой gRNA на последовательности ДНК, которая комплементарна последовательности протоспейсера; ф) сравнение значений предполагаемого времени нахождения в сайтах протоспейсера и/или нецелевых сайтах первой gRNA со значениями предполагаемого времени нахождения полноразмерной gRNA или усеченной gRNA (tru-gRNA) в сайтах протоспейсера и/или нецелевых сайтах; г) рандомизацию от 0 до N нуклеотидов в линкере и от 0 до М нуклеотидов в первой gRNA, и получение второй gRNA, и повторение стадии (е) со второй gRNA; и) идентификацию оптимизированной gRNA на основе последовательности gRNA, которая удовлетворяет критериям конструирования; и и) тестирование оптимизированной gRNA *in vivo* для определения специфичности связывания.

Настоящее изобретение направлено на способ получения оптимизированной направляющей РНК (gRNA). Способ предусматривает: а) идентификацию представляющей интерес целевой области, при этом представляющая интерес целевая область содержит последовательность протоспейсера; б) определение полинуклеотидной последовательности полноразмерной gRNA, которая нацеливается на представляющую интерес целевую область, при этом полноразмерная gRNA содержит нацеливающиеся на протоспейсер последовательность или сегмент; в) определение по меньшей мере одного или нескольких нецелевых сайтов для полноразмерной gRNA; г) получение полинуклеотидной последовательности первой gRNA, при этом первая gRNA содержит полинуклеотидную последовательность полноразмерной gRNA и сегмент РНК, при этом сегмент РНК содержит полинуклеотидную последовательность, имеющую М нуклеотидов в длину, которая комплементарна нуклеотидному сегменту нацеливающихся на протоспейсер последовательности или сегменту, при этом РНК расположена на 3'-конце полинуклеотидной последовательности полноразмерной gRNA, при этом первая gRNA необязательно содержит линкер между 3'-концом полинуклеотидной последовательности полноразмерной gRNA и сегментом РНК, при этом линкер содержит полинуклеотидную последовательность, имеющую N нуклеотидов в длину, при этом первая gRNA способна встраиваться в последовательность протоспейсера, и связываться с последовательностью ДНК, которая комплементарна последовательности протоспейсера, и образовывать дуплекс с протоспейсером, и при этом первая gRNA способна встраиваться в нецелевой сайт, и связываться с последовательностью ДНК, которая комплементарна нецелевому сайту, и образовывать нецелевой дуплекс; е) вычисление оценочного показателя или вычислительное моделирование кинетики встраивания и времени нахождения, в течение которого gRNA остается встроенной в дуплексы протоспейсера и неце-

левого сайта, где динамику встраивания оценивают нуклеотид за нуклеотидом путем определения энергетических различий между дополнительным встраиванием отличающейся gRNA и повторным отжигом первой gRNA на последовательности ДНК, которая комплементарна последовательности протоспейсера; f) сравнение значений предполагаемого времени нахождения в сайтах протоспейсера и/или нецелевых сайтах первой gRNA со значениями предполагаемого времени нахождения полноразмерной gRNA или усеченной gRNA (tru-gRNA) в сайтах протоспейсера и/или нецелевых сайтах; g) рандомизацию от 0 до N нуклеотидов в линкере и от 0 до M нуклеотидов в первой gRNA, и получение второй gRNA, и повторение стадии (e) со второй gRNA; h) идентификацию оптимизированной gRNA на основе последовательности gRNA, которая удовлетворяет критериям конструирования; и i) тестирование оптимизированной gRNA in vivo для определения специфичности связывания.

Настоящее изобретение направлено на оптимизированную gRNA, полученную с помощью описанных выше способов.

Настоящее изобретение направлено на выделенный полинуклеотид, кодирующий описанную выше оптимизированную gRNA.

Настоящее изобретение направлено на вектор, содержащий описанный выше выделенный полинуклеотид.

Настоящее изобретение направлено на клетку, содержащую описанный выше выделенный полинуклеотид или описанный выше вектор.

Настоящее изобретение направлено на набор, содержащий описанный выше выделенный полинуклеотид, описанный выше вектор или описанную выше клетку.

Настоящее изобретение направлено на способ эпигемного редактирования в клетке-мишени или в субъекте. Способ предусматривает приведение клетки или субъекта в контакт с эффективным количеством молекулы оптимизированной gRNA, описанной выше, и слитого белка, при этом слитый белок содержит первый полипептидный домен, содержащий дефицитную по нуклеазной активности Cas9, и второй полипептидный домен с активностью, выбранной из группы, состоящей из активности активации транскрипции, активности репрессии транскрипции, нуклеазной активности, активности фактора освобождения транскриптов, активности модификации гистонов, активности ассоциации нуклеиновых кислот, ДНК-метилазной активности и прямой или непрямой ДНК-деметилазной активности.

Настоящее изобретение направлено на способ сайт-специфического расщепления ДНК в клетке-мишени или в субъекте. Способ предусматривает приведение клетки или субъекта в контакт с эффективным количеством молекулы оптимизированной gRNA, описанной выше, и слитого белка или белка Cas9, при этом слитый белок содержит первый полипептидный домен, содержащий дефицитную по нуклеазной активности Cas9, и второй полипептидный домен с активностью, выбранной из группы, состоящей из активности активации транскрипции, активности репрессии транскрипции, нуклеазной активности, активности фактора освобождения транскриптов, активности модификации гистонов, активности ассоциации нуклеиновых кислот, ДНК-метилазной активности и прямой или непрямой ДНК-деметилазной активности.

Настоящее изобретение направлено на способ редактирования генома в клетке. Способ предусматривает введение в клетку эффективного количества молекулы оптимизированной gRNA, описанной выше, и слитого белка, при этом слитый белок содержит первый полипептидный домен, содержащий дефицитную по нуклеазной активности Cas9, и второй полипептидный домен с активностью, выбранной из группы, состоящей из активности активации транскрипции, активности репрессии транскрипции, нуклеазной активности, активности фактора освобождения транскриптов, активности модификации гистонов, активности ассоциации нуклеиновых кислот, ДНК-метилазной активности и прямой или непрямой ДНК-деметилазной активности.

Настоящее изобретение направлено на способ модулирования экспрессии генов в клетке. Способ предусматривает приведение клетки в контакт с эффективным количеством оптимизированной gRNA, описанной выше, и слитого белка, при этом слитый белок содержит первый полипептидный домен, содержащий дефицитную по нуклеазной активности Cas9, и второй полипептидный домен с активностью, выбранной из группы, состоящей из активности активации транскрипции, активности репрессии транскрипции, нуклеазной активности, активности фактора освобождения транскриптов, активности модификации гистонов, активности ассоциации нуклеиновых кислот, ДНК-метилазной активности и прямой или непрямой ДНК-деметилазной активности.

Краткое описание графических материалов

На фиг. 1A показано схематическое изображение активности Cas9.

На фиг. 1B показано изображение атомно-силовой микроскопии (AFM) dCas9-sgRNA, связанной с последовательностью протоспейсера в пределах одной меченой стрептавидином молекулы ДНК, полученной из локуса AAVS1 человека.

На фиг. 1C-1D показана фракция связанной ДНК, занимаемой Cas9/dCas9-sgRNA по длине полученного из AAVS1 (фиг. 1C) или сконструированного ДНК-субстрата (фиг. 1D), разработанного с рядом полностью комплементарных и частично комплементарных последовательностей протоспейсера. Вертикальные линии представляют сегменты (23 п.о.), где каждый значимый признак расположен на соответ-

ствующих субстратах.

На фиг. 2А-2D показано модулирование сродства и специфичности связывания вариантами направляющих РНК. На фиг. 2А показана схема dCas9, связанной с единой направляющей РНК с усечением двух нуклеотидов с ее 5'-конца (tru-gRNA, фиолетовый). На фиг. 2В показан схематический и предполагаемый механизм dCas9, связанной с единой направляющей РНК с удлинением 5'-конца, которое образует шпильку с РАМ-дистальным сегментом связывания ее нацеливающейся области (hp-gRNA, синий). На фиг. 2С показано сродство моносайтового связывания (K_A) для dCas9 с tru-gRNA (фиолетовый, $n=257$) вдоль сконструированного ДНК-субстрата (см. фиг. 1D). Пунктирная линия показывает моносайтовое сродство dCas9-sgRNA для сравнения. На фиг. 2D показано сродство моносайтового связывания (K_A) для dCas9 с направляющими РНК с 5'-шпильками, которые перекрывают нуклеотиды, комплементарные последним шести (hp6-gRNA, синий) или десяти (hp10-gRNA, зеленый) РАМ-дистальным нуклеотидам протоспейсера.

На фиг. 3А-3D показано, что Cas9 подвергается прогрессивному конформационному переходу, поскольку она связывается с сайтами, которые все в большей степени соответствуют последовательности протоспейсера. На фиг. 3А показана фракция связанной ДНК, занимаемой Cas9/dCas9 по длине ДНК-субстратов, причем цвета, представляющие популяции Cas9/dCas9, сгруппированы по их структурам (по среднеквадратичной разнице после выравнивания, см. текст). Различные признаки на ДНК, которые использовали для сайт-специфического анализа структурных свойств Cas9/Cas9, метили как неспецифические последовательности (α ; "20MM"), сайты, содержащие 10 РАМ-дистальных ошибочных спариваний в протоспейсере (β ; "10MM"), сайты, содержащие 5 РАМ-дистальных ошибочных спариваний в протоспейсере (γ ; "5MM") или сайт полного протоспейсера (δ или ϵ для dCas9 или Cas9, соответственно; "ОММ"). Среднее по ансамблю значение первичных кластеров и цветовое кодирование в соответствии со сгруппированными структурами, которые они представляют, представлено на фиг. 3С. На фиг. 3В показан объем в зависимости от высоты наблюдаемой Cas9/dCas9 с цветовым кодированием по кластеру, к которому относится каждый белок. Пунктирные линии обозначают области, которые, вероятно, состоят из агрегатов (вверху справа) или стрептавидиновых меток, адсорбированных вблизи ДНК (внизу слева). Для сравнения - средняя высота стрептавидиновых концевых меток: $0,92 \text{ нм} \pm 0,006 \text{ нм}$ (SEM); средний объем стрептавидиновых концевых меток: $0,110 \times 10^4 \text{ нм}^3 \pm 0,002 \times 10^4 \text{ нм}^3$ (SEM); $n=1941$. На фиг. 3D показаны средние объемы и высоты Cas9/dCas9 с sgRNA (красные круги, с красными метками для Cas9 и синими метками для dCas9) или tru-gRNA (фиолетовые круги), связанных у каждого признака на субстратах. Обратите внимание, что только dCas9 с tru-gRNA должны взаимодействовать с первыми 3 или 8 РАМ-дистальными ошибочными спариваниями сайтов 5MM и 10MM (здесь, соответственно, обозначены "3MM" и "8MM"). Для стандартных ошибок средних объемов и высот см. табл. 2. Для Cas9/dCas9 с sgRNA их структурные свойства у каждого признака статистически отличаются (δ - ϵ , α - ϵ : $p < 0,05$; α - β : $p < 0,005$; β - γ , γ - δ : $p \ll 0,0005$. T^2 -критерий Хотеллинга).

На фиг. 4А-4D показаны эксперименты согласно кинетическому методу Монте-Карло (КМС), выявляющие различия в стабильности R-петли или структуру, образованную дуплексом протоспейсера с внедряющейся направляющей РНК, в стабильно связанной Cas9 для разных вариантов направляющей РНК. На фиг. 4А показана схема встраивания нити протоспейсера (зеленый цвет) с помощью направляющей РНК (красный цвет) для экспериментов КМС. R-петля выделена. Скорости перехода для встраивания (v_f для скорости $m \rightarrow m+1$, где m представляет собой степень встраивания нити или, что то же самое, длину R-петли) или повторного отжига дуплекса (v_r для скорости $m \rightarrow m-1$) являются функцией ближайших соседних энергий гибридизации ДНК:ДНК и РНК:ДНК. Подробнее см. в тексте и разделе Дополнительные способы. На фиг. 4В показано относительное время, когда R-петля имеет размер m , для sg-RNA (красный цвет) или tru-gRNA (фиолетовый цвет), полученное из экспериментов КМС "в равновесии" (моделирование начато при $m=20$ или 18 соответственно). Моделирование проводили до ≥ 10000 (условных единиц). На фиг. 4С показана динамика "дыхания" R-петли согласно кинетическому методу Монте-Карло для sgRNA (красный цвет) и tru-gRNA (фиолетовый цвет) после полного встраивания (моделирование начато при $m=20$ или 18, соответственно). Звездочки отмечают начальное положение для моделирования (вставка). Гистограмма соответствующей продолжительности сроков, в течение которых R-петля составляет ≥ 16 п.о. в длину. На фиг. 4D показана предполагаемая модель для механизмов, управляющих специфичностью Cas9/dCas9, на основании результатов АФМ-визуализации и экспериментов согласно кинетическому методу Монте-Карло (КМС) (см. основной текст). Cas9/dCas9 связывается с РАМ, и направляющая РНК встраивается в примыкающий к РАМ дуплекс протоспейсера. В ходе этого встраивания нити направляющая РНК должна вытеснить комплементарную нить протоспейсера. Конкуренция между встраиванием и повторным отжигом дуплекса приводит к динамической ("дышащей") структуре R-петли. Стабильность 14^{10} - 17^{10} сайтов взаимодействия протоспейсера с направляющей РНК, которая резко возрастает при связывании в 19^{0M} и 20^{0M} сайтах, способствует конформационному изменению в Cas9/dCas9, которое обеспечивает возможность расщепления ДНК Cas9.

На фиг. 5А-5С показано, что эксперименты согласно кинетическому методу Монте-Карло (КМС) выявляют различия в способности преодолевать ошибочные спаривания (ММ) и внедряться в протоспей-

сер в зависимости от структуры направляющей РНК. На фиг. 5А-5В показано фрагментарное заполнение по времени длины R-петель m для sgRNA (фиг. 5А) или tru-gRNA (фиг. 5В) в ходе встраивания, полученное в экспериментах согласно КМС (начато при $m=10$, отмечено звездочкой). Белые X указывают положения ошибочных спариваний. Моделирование проводили до $t \geq 10000$ (условных единиц) и результаты представляют собой среднее по 100 испытаниям. На фиг. 5С показана типичная динамика согласно КМС для встраивания нити (начало при $m=10$) с ошибочно спаренным сайтом в $m=14$ (стрелка) для sgRNA (красный цвет) и tru-gRNA (фиолетовый цвет). В то время как sgRNA в значительной степени стабильно внедряются после обхода ошибочного спаривания, tru-gRNA повторно задерживаются за ошибочным спариванием в результате присущей им изменчивости R-петель (см. фиг. 4).

На фиг. 6А-6В показано, что экспериментальные (Hsu et al. (2013) Nature biotechnology, 31, 827-832) частоты разрезания в целевых сайтах, содержащих одно ошибочное спаривание rG-dG, rC-dC, rA-dA и rU-dT в PAM-дистальной области ($\geq 10^{10}$ сайта протоспейсера), коррелируют со стабильностью R-петли, определенной в экспериментах согласно кинетическому методу Монте-Карло. На фиг. 6А показаны $\log_{10}(p\text{-значение})$ корреляции между частотой разрезания Cas9 и стабильностью R-петли в сайтах m (фрагмент времени, когда направляющая РНК остается связанной с протоспейсером в сайте m , см. текст) в ходе встраивания нити, начатого в сайте m . (i) Стабильность в сайтах $m=10$ до $m=14$ имеет высокую обратную корреляцию с вероятностью того, что направляющая РНК отойдет от протоспейсера до преодоления ошибочного спаривания (фиг. 15В), тогда как (ii) сайты от $m=14$ до $m=17$ ассоциированы (согласно изображениям AFM) с конформационным изменением, которое индуцирует расщепляющую активность. Цвет соответствует коэффициенту корреляции. На фиг. 6В показано, что экспериментальная частота разрезания не коррелирует в значительной степени с расчетными свободными энергиями равновесного связывания направляющей РНК протоспейсера (ΔG°_{37}) (слева), в то время как она коррелирует со стабильностью сайта $m=14$ в ходе встраивания нити (справа). Планки погрешности представляют собой стандартные ошибки среднего времени пребывания в сайте $m=14$. Для этих экспериментов согласно кинетическому методу Монте-Карло $\max(t)=100$ (условных единиц). Цветная полоса используется для отображения местоположения ошибочно спаренного сайта (MM).

На фиг. 7А-7С показано краткое изложение предполагаемых механизмов, посредством которых структура направляющей РНК влияет на специфичность Cas9/dCas9. На фиг. 7А показано, что для единой направляющей РНК (sgRNA) первые несколько нуклеотидов РНК (которые связываются с 18-20^{ом} сайтами протоспейсера) стабилизируют дыхание и связывание R-петли в 14-17^{ом} сайтах протоспейсера, обеспечивают возможность эффективного конформационного перехода в активное состояние для обеспечения расщепления. Однако данная повышенная стабильность, придаваемая этими основаниями, обеспечивает временную стабилизацию в ошибочно спаренных сайтах и конформационное изменение, позволяющее расщепление. Во многих случаях, преодолев ошибочное спаривание, R-петли остаются стабильно полностью встроенными. На фиг. 7В показано, что для направляющих РНК с первыми несколькими (здесь 2) усеченными нуклеотидами (tru-gRNA) пониженная стабильность R-петли (характеризующейся значительной изменчивостью) снижает вероятность сохранения активной конформации. Когда в протоспейсере есть сайты ошибочного спаривания, изменчивость R-петли гарантирует, что она будет быстро и многократно "повторно остановлена" за ошибочным спариванием и в значительной степени испытает помехи в этих сайтах. На фиг. 7С показано, что хотя оказалось, что "простые" продления 5'-конца направляющей РНК для целенаправленного воздействия на протоспейсер и соседние сайты за протоспейсером *in vivo* отщепляются обратно приблизительно до длины sgRNA (фиг. 7А), направляющие РНК с 5'-шпильками, комплементарными сегментам, целенаправленно воздействующим "PAM-дистально" (hp-gRNA), как ожидается, остаются защищенными в структуре Cas9/dCas9 до встраивания. После связывания сайта PAM и инициирования встраивания нити с помощью hp-gRNA при связывании с полным протоспейсером шпилька открывается и может иметь место полное встраивание нити. Если в целевом сайте имеются PAM-дистальные ошибочные спаривания, то энергетически более выгодно, чтобы шпилька оставалась закрытой и встраивание нити было затруднено. Остается необходимость проверить способность Cas9-hp-gRNA расщеплять РНК.

На фиг. 8А-8В показана чистота экспрессированной Cas9 и dCas9 в геле SDS очищенных продуктов Cas9 (фиг. 8А) и dCas9 (фиг. 8В) (номинальная молекулярная масса: 160 кДа). Элюированные полосы показывают, что продукт имеет чистоту $\sim 95\%$.

На фиг. 9А-9С показаны дополнительные изображения Cas9/dCas9, связанной с ДНК. А) Распределение связывания dCas9 с субстратом, не характеризующимся гомологией с последовательностью протоспейсера AAVs1 (сравните с фиг. 1) ($n=443$). Наложение представляет собой распределение кумулятивных вероятностей (CDF) сайтов PAM (CDF_{PAM}, черный цвет) и CDF оснований, связанных dCas9 (красный цвет, CDF_{Cas9}). Сравнение начинается за 100 оснований с каждого конца, чтобы избежать артефактов, вводимых за счет перекрытия со стрептавидиновой меткой (критерий отбора ДНК) и связывания с незащищенными тупыми концами ДНК (что приводит к ожидаемому увеличению неспецифического связывания). В) Абсолютное различие D_n между CDF связывания белка и сайтов PAM. Пунктирная линия представляет собой критерий Колмогорова-Смирнова для степени соответствия двух распределений.

С) CDF связывания сравнивали с CDF распределений PAM из 100000 случайным образом полученных последовательностей с одинаковыми вероятностями G, A, T и C с использованием MATLAB. Вертикальная красная линия представляет собой экспериментальную $\text{Sup}(D_n)$, указывая, что экспериментальное связывание dCas9 более близко соответствует экспериментальному распределению PAM, чем для 71,20% полученных последовательностей.

На фиг. 10A-10C показано связывание с "нонсенс"-субстратом, не характеризующимся гомологией (>3 п.о.) с последовательностью протоспейсера. (А) Изображения dCas9 в отдельности. (В) Гистограмма (n=423) объема (слева) и высоты (справа) dCas9, изображенной отдельно, с гауссовским сглаживанием основных пиков. Согласно гауссовскому сглаживанию: средняя высота составляет 1,746 нм (95% доверительный интервал: 1,689 нм - 1,802 нм) со стандартным отклонением 0,441 нм, и средний объем составляет 1302 нм^3 (95% доверительный интервал: 1266 нм^3 - 1337 нм^3) со стандартным отклонением $259,1 \text{ нм}^3$ (отметим, что поскольку dCas9 здесь не имеет ДНК в своем канале связывания, их зарегистрированные объемы могут казаться искусственно низкими из-за снижения механической устойчивости к зонду AFM). Высоты измеряли относительно медианного значения 10-пиксельной области, окружающей каждый белок, и объемы записаны в виде смежных признаков, превышающих в два раза стандартное отклонение локальных высот фона. (С) Дополнительные типичные изображения dCas9, связанной с ДНК, которая была помечена на одном конце моновалентным стрептавидином.

На фиг. 11A-11D показана репрезентативная фигура dCas9-sgRNA, связанная с РНК, и пример процессинга структурных свойств белка. На фиг. 11A показано типичное широкопольное изображение dCas9, связанной со сконструированной ДНК. На фиг. 11B показан крупный план выделенной области. Белые стрелки представляют одновалентный стрептавидин, а красные стрелки белки dCas9. На фиг. 11C-11D показан пример извлечения из исходного изображения (фиг. 11C) и выделения (фиг. 11D) структур Cas9/dCas9. Это извлечение повторяли для каждого выделенного белка, связанного с ДНК, затем попарно выравнивали посредством итерационного перевода, вращения и отражения, чтобы свести к минимуму их среднеквадратичное топологическое различие. Из этих сведенных к минимуму среднеквадратичных различий составляли матрицу расстояний, группировали каждый белок по методу Laio и Rodriguez (2014) Science (New York, N.Y.), 344, 1492-1496, затем сопоставляли популяции структур по кластеру обратно с их сайтами на ДНК (фиг. 2A, фиг. 10A-10C).

На фиг. 12A-12B показаны свойства Cas9/dCas9-sgRNA, сопоставленные с их соответствующими сайтами связывания. Верх: расположенные друг над другом гистограммы объема (слева), максимальных высот (посередине) и структур (сгруппированные по среднеквадратичной разнице) после выравнивания (справа, см. текст) для всех условий эксперимента. Популяции окрашены в соответствии со сгруппированным объемом, высотой или структурным кластером, как показано на графике рассеяния ниже. Распределение связывания извлеченных молекул Cas9/dCas9 (фиг. 10A-10C) близко соответствует распределению всего набора данных (фиг. 1C-1D, фиг. 8A-8B), указывая, что процедура отбора является объективной и выбранные белки являются типичными представителями всего набора данных. Ниже: график рассеяния объема в зависимости от максимальной высоты всех Cas9/dCas9 с цветовым кодированием по сгруппированному объему (слева), максимальной высоте (посередине) и структурному кластеру (справа).

На фиг. 13 показаны структурные свойства Cas9/dCas9 с tru-gRNA и hp-gRNA в их соответствующих сайтах связывания. Фракция связанной ДНК, занимаемой Cas9/dCas9, вместе со сконструированным ДНК-субстратом с цветами, представляющими популяции Cas9/dCas9, сгруппированные в соответствии с их структурами (см. фиг. 3C). Структуры белка классифицировали в соответствии с dCas9/Cas9 с sgRNA, которые они наиболее близко напоминали (по среднеквадратическому отклонению после выравнивания, см. текст). Для ссылки, на сконструированных ДНК-субстратах расположение полного сайта протоспейсера: 144-167 п.о.; расположение сайта 10 MM (8MM): 452-465 п.о.; расположение сайта 5MM (3MM): 592-610 п.о. Наблюдались аналогичные тенденции, которые встречались у dCas9/Cas9 и sgRNA: поскольку dCas9 связывается с сайтами, которые все больше спариваются ошибочно, доля популяции, группирующейся с наибольшей группой (желтый) растет, хотя этот эффект подавлен в tru-gRNA со значительной долей популяции, группирующейся с меньшими популяциями (зеленый и синий) даже в полном сайте протоспейсера. Эффект для hp10-gRNA особенно выражен, подчеркивая, что она характеризуется слабым сродством к нецелевым сайтам.

На фиг. 14A-14C показана модель встраивания нити в протоспейсеры ДНК с помощью направляющих РНК и предполагаемые стабильности связывания РНК, встроенной в протоспейсер с PAM-дистальными ошибочными спариваниями. На фиг. 14A показана схематическая модель встраивания нити в протоспейсеры ДНК с помощью направляющих РНК. См. также фиг. 4A. Предполагается, что направляющая РНК диссоциирует при $m=1$. На фиг. 14B показано расчетное распределение вероятностей времени диссоциации для направляющей РНК, первоначально встроенной до $m=5$, для протоспейсеров с различным количеством смежных PAM-дистальных ошибочных спариваний. Длительность этих значений времени диссоциации можно рассматривать как аппроксимацию предрасположенности связывания dCas9 в этих сайтах. Звездочка отмечает времена диссоциации для популяции направляющих РНК, которые изначально не могут полностью внедриться после первоначального встраивания до $m=5$. Встроенные РНК являются высоко нестабильными в сайтах протоспейсера с 15 PAM-дистальными ошибочными

спариваниями (15ММ), и редко наблюдали в эксперименте Cas9/dCas9, связанную в этих сайтах (фиг. 1D). Рассчитано, что встроенная РНК (перед диссоциацией) в сайтах протоспейсера с 10 или 5 РАМ-дистальными ошибочными спариваниями (10ММ и 5ММ) остается значительно дольше, чем в сайтах с 15ММ, но в пределах одного порядка величины друг от друга; обнаружили, что их предрасположенность к связыванию должна быть приблизительно равна и ниже, чем у полных сайтов протоспейсера (0ММ) в экспериментах AFM. Функции плотности распределения вероятности рассчитывали с использованием способа Q-матрицы, как описано (Sakmann et al. (1995) Single-channel recording, Springer; 2nd ed.), с помощью последовательность-специфичных скоростей перехода между t состояниями (v_f и v_r , см. раздел Дополнительные способы). На фиг. 14С показано изучение предполагаемых периодов полужизни связывания РНК-протоспейсер в протоспейсерах с различными количествами РАМ-дистальных ошибочных спариваний, свидетельствуя, что существует приблизительно три режима, при которых стабильности встроенных РНК являются сходными: те, которые несут > 11 РАМ-дистальных ошибочных спариваний (низкая стабильность); те, которые несут 3-11 РАМ-дистальных ошибочных спариваний (средняя стабильность); и те, которые несут < 3 РАМ-дистальных ошибочных спариваний (высокая стабильность). Результаты являются качественно аналогичными распределению dCas9 на сконструированном субстрате, наблюдаемому посредством AFM (фиг. 1D).

На фиг. 15 показано стимулированное среднее время первого прохода для преодоления ошибочно спаренного сайта в ходе встраивания нити с помощью sgRNA и tru-gRNA. Стимулированное (согласно кинетическому методу Монте-Карло) среднее время первого прохода для преодоления ошибочно спаренного сайта в ходе встраивания нити с помощью sgRNA (синий цвет) и tru-gRNA (красный цвет) для разных положений ошибочно спаренного сайта. Планки погрешностей представляют собой стандартные отклонения зарегистрированного времени первого прохода. Последовательность протоспейсера (сайт AAVS1) в рамке.

На фиг. 16А-16В показаны корреляции между частотой расщепления Cas9 (Hsu et al. (2013) Nature Biotechnology, 31, 827-832) и измерениями стабильности R-петли, полученными согласно кинетическому методу Монте-Карло. На фиг. 16А показано, что статистическая мощность и сила корреляционных связей между стабильностью сайтов R-петли (согласно кинетическому методу Монте-Карло, см. основной текст) и экспериментальной частотой расщепления по Hsu et al. (2013) Nature Biotechnology, 31, 827-832, снижается с увеличением длительности моделирования (от $\max(t)=100$ до $\max(t)=1000$ условных единиц). Этот результат свидетельствует о том, что кинетика встраивания нити может быть важным предиктором скорости нецелевого расщепления. На фиг. 16В показана корреляция между фрагментами времени, в течение которого R-петля имеет размер m относительно вероятности того, что испытание согласно кинетическому методу Монте-Карло предсказывает, что внедряющаяся нить будет диссоциировать до преодоления ошибочного спаривания. Связывание в сайтах 10 ~ 14-15 очень сильно отрицательно коррелирует (~0,5-0,85) с вероятностью диссоциации до преодоления ошибочного спаривания, в то время как из экспериментов с AFM-визуализацией обнаружили, что связыванием в сайтах ≈ 16 ассоциировано с конформационным изменением в Cas9/dCas9.

На фиг. 17 показано краткое изложение данных Deep-Seq со сравнением целевых активностей.

На фиг. 18 показано краткое изложение данных Deep-Seq со сравнением увеличения специфичности.

На фиг. 19 показан протоспейсер 1, Dystrophin; на дорожке 1 показан контроль GFP; на дорожке 2 показана полная gRNA; на дорожке 3 показана Tru-gRNA 19 нт; на дорожке 4 показана Tru-gRNA 18 нт; на дорожке 5 показана Tru-gRNA 17 нт; на дорожке 6 показана Tru-gRNA 16 нт; на дорожке 7 показана Hp-gRNA 4 п.о.; на дорожке 8 показана Hp-gRNA 5 п.о.; на дорожке 9 показана Hp-gRNA 6 п.о.; на дорожке 10 показана Hp-gRNA 7 п.о.; на дорожке 11 показана Hp-gRNA 8 п.о., и на дорожке 12 показана Hp-gRNA 9 п.о., шпилька 1 (дорожка 12, 9 нт hp) - GtgagtaggttctgCCTACTCAGACTGTTACTC (SEQ ID NO: 335), где курсивом показана часть шпильки и подчеркнута петля шпильки.

На фиг. 20 показан протоспейсер 1, Dystrophin, внутренние петли.

На фиг. 21 показаны рассчитанные вторичные структуры 5'-концов нацеливающихся на протоспейсер сегментов hp-gRNA, используемых для экспериментов Deep Seq (с помощью программного пакета NuPack). Цвета представляют собой вероятность того, что каждый нуклеотид, существующий в этой вторичной структуре, находится в равновесии.

На фиг. 22 показан Dystrophin, частоты вставок/делеций, все сайты.

На фиг. 23 показан Dystrophin, целевая мишень/сумма (нецелевые мишени).

На фиг. 24 показан протоспейсер 2, EMX1; на дорожке 1 показан контроль GFP; на дорожке 2 показана полная gRNA; на дорожке 3 показана Tru-gRNA; на дорожке 4 показана hp-gRNA 10 п.о., и на дорожке 5 показана hp-gRNA 6 п.о., шпилька 1. Преобразования - Surv_OT1=DS_OT2; Surv_OT53=DS_OT3.

На фиг. 25А и 25В показан протоспейсер 2, EMX1, tru-hps, внутренние петли.

На фиг. 26А-26С показаны структуры шпильки. На фиг. 26А показана шпилька 1, которая пред-

ставляет собой 5'-шпильку в 6 п.о. На фиг. 26В показана шпилька 2, которая представляет собой 5'-шпильку в 5 п.о. на 18 нт (усеченной) gRNA. На фиг. 26С показана шпилька 3, которая представляет собой 5'-шпильку в 3 п.о.

На фиг. 27 показан EMX1, частоты вставок/делеций, все сайты.

На фиг. 28 показан EMX1, частоты вставок/делеций, низкие уровни нецелевых участков.

На фиг. 29 показан EMX1, целевая мишень/сумма (нецелевые мишени).

На фиг. 30 показан протоспейсер 3, VEGFA1. На дорожке 1 показан контроль GFP; на дорожке 2 показана полная gRNA; на дорожке 3 показана Tru-gRNA; на дорожке 4 показана hp-gRNA 10 п.о., и на дорожке 5 показана hp-gRNA 6 п.о.

На фиг. 31 показан протоспейсер 3, VEGFA1: ram-проксимальные шпильки. На дорожке 1 показан контроль GFP; на дорожке 2 показана полная gRNA; на дорожке 3 показана hp-gRNA1; на дорожке 4 показана hp-gRNA2; на дорожке 5 показана hp-gRNA3; на дорожке 6 показана hp-gRNA4; на дорожке 7 показана hp-gRNA5 и на дорожке 8 показана hp-gRNA6.

На фиг. 32 показан протоспейсер 3, VEGFA1: ram-проксимальные шпильки.

На фиг. 33 показан протоспейсер 3, VEGF1, внутренние пегли. На дорожке 1 показан контроль; на дорожке 2 показана полная; на дорожке 3 показана 2 нт hr; на дорожке 4 показана 3 нт hr, шпилька 5; и на дорожке 5 показана 4 нт hr.

На фиг. 34А и 34В показано, что эксперименты Deep-seq для шпилек 1, 2 и 3 не удалась. На фиг. 25А показана шпилька 4 - полученная на основе вычислений шпилька, разработанная для распознавания нецелевого сайта 2 при сохранении целевой активности. На фиг. 25В показана шпилька 5-5'-шпилька в 4 п.о. (обычно gRNA имеет значительную 3'-вторичную структуру).

На фиг. 35 показан VEGF1, частоты вставок/делеций, все сайты.

На фиг. 36 показан VEGF1, частоты вставок/делеций, низкие уровни нецелевых участков.

На фиг. 37 показан VEGF1, целевая мишень/сумма (нецелевые мишени).

На фиг. 38 показан протоспейсер 4, VEGFA3. На дорожке 1 показан контроль GFP; на дорожке 2 показана полная gRNA, на дорожке 3 показана Tru-gRNA; на дорожке 4 показана hp-gRNA в 3 п.о.; на дорожке 5 показана hp-gRNA в 4 п.о.; на дорожке 6 показана hp-gRNA в 5 п.о.; на дорожке 7 показана hp-gRNA в 6 п.о. и на дорожке 8 показана hp-gRNA в 10 п.о.

На фиг. 39 показаны gRNA4, VEGFA3: ram-проксимальные шпильки. На дорожке 1 показан контроль GFP; на дорожке 2 показана полная gRNA; на дорожке 3 показана hp-gRNA1; на дорожке 4 показана hp-gRNA2; на дорожке 5 показана hp-gRNA3; на дорожке 6 показана hp-gRNA4; на дорожке 7 показана hp-gRNA5 и на дорожке 8 показана hp-gRNA6.

На фиг. 40А показана шпилька 1 - шпилька в 4 п.о., нацеливающаяся на 3'-область.

На фиг. 40В показана шпилька 2 - шпилька в 4 п.о., нацеливающаяся на 3'-область с неоднозначностью пар G-U.

На фиг. 40 показана шпилька 3 - шпилька в 4 п.о., нацеливающаяся на 3'-область с неоднозначностью пар G-U (вариант дизайна).

На фиг. 41 показан VEGF3, частоты вставок/делеций, все сайты.

На фиг. 42 показан VEGF3, частоты вставок/делеций, низкие уровни нецелевых участков.

На фиг. 43 показан VEGF3, целевая мишень/сумма (нецелевые мишени).

На фиг. 44А показана шпилька, разработанная для нацеливания на ген EMX1.

На фиг. 44В показана последовательность EMX1-sg1 шпильки с фиг. 44А.

На фиг. 44С показано воздействие уменьшения длины протоспейсера и увеличения длины шпильки на специфичность.

На фиг. 45А-45D показаны последовательности ДНК/РНК.

На фиг. 46 показана фигура, которая описывает анализ Surveyor.

На фиг. 47 показана устойчивость AsCpf1 и LbCpf1 к crRNA с ошибочным спариванием или усечением и эндогенные модификации генов с помощью AsCpf1 и LbCpf1 с использованием crRNA, которые содержат основания с одним ошибочным спариванием. Активность, определяемая с помощью анализа T7E1; планки погрешностей, s.e.m.; n=3 (взято у Kleinstiver et al., Nat. Biotech. 34:869-875).

На фиг. 48 показаны результаты анализа surveyor для hp-gRNA, использованных с системой CRISPR V типа, в котором шпильку добавляют к 3'-концу полноразмерной gRNA для отмены нецелевой активности.

Подробное описание изобретения

В данном документе раскрываются композиция и способы для сайт-специфического целенаправленного воздействия на ДНК и эпигеномного редактирования генов и/или регуляции транскрипции, как например: расщепление ДНК и активация или репрессия гена. Настоящее изобретение направлено на модульный способ конструирования и использования оптимизированных направляющих РНК, имеющих структуры шпильки (hpgRNA), которые могут быть легко встроены в существующую биотехнологическую инфраструктуру и которые приводят к контролируемому уменьшению нецелевой активности при сохранении способности специфично нацеливаться на соответствующую последовательность ДНК. Способы, описанные в данном документе, обеспечивают новый подход к конструированию оптимизирован-

ной gRNA, который работает более эффективно, чем другие доступные способы, и их можно использовать в сочетании с другими белок-специфическими средствами улучшения, особенно повышающими высокоэффективную производительность.

Раскрытые способы и оптимизированные gRNA имеют большое преимущество в том, что они легко адаптируются к существующим методологиям и инфраструктурам, уже имеющим место для осуществления геномной инженерии с направляющими РНК. В некоторых вариантах осуществления Cas9, dCas9 или Crpf1 доставляют в клетку с использованием вирусных векторов вместе с векторами, кодирующими транскрипцию оптимизированных gRNA в клетке. В соответствии с настоящим изобретением потребуется лишь несколько дополнительных нуклеотидов для вектора, кодирующего оптимизированную gRNA, что можно легко адаптировать под современные и стандартные практические методики. Подобно усеченным направляющим РНК (tru-gRNA), оптимизированные gRNA или hpgRNA можно использовать в комбинации с парными никазами, например, или с другими модификациями самих эндонуклеаз для дополнительного улучшения специфичности. Провели серию экспериментов *in vitro*, которая показала, что применение оптимизированных gRNA, полученных с использованием описанных в данном документе способов, увеличивает специфичность связывания ДНК относительно наиболее доступных вариантов gRNA (см. фиг. 2).

Применение оптимизированной gRNA устраняет или значительно ослабляет активность мишеней, содержащих лишь несколько ошибочно спаренных последовательностей ДНК, которые, как правило, являются сайтами, в которых имеет место нецелевая активность РНК-направляемых эндонуклеаз. Оптимизированная gRNA также обеспечивает специфичность расщепляющей активности в клетках млекопитающих в сайтах, которые, как известно, индуцируют нецелевую активность даже при наиболее известных улучшениях в отношении направляющих РНК. Настоящее изобретение представляет собой универсально применимый способ снижения нецелевой активности РНК-направляемых эндонуклеаз, в частности Cas9, путем внесения изменений в структуру направляющей РНК.

1. Определения

Подразумевается, что термины "содержат(содержит)", "включают(включает)", "имеющий", "имеет", "может", "включает(включают)" и их варианты, используемые в данном документе, являются открытыми переходными фразами, терминами или словами, которые не исключают возможность наличия дополнительных действий или структур. Формы единственного числа включают ссылки на множественное число, если контекст явно не указывает на иное. Настоящее изобретение также охватывает другие варианты осуществления, "содержащие" варианты осуществления или элементы, представленные в данном документе, "состоящие из" них и "по сути, состоящие из них, независимо от того, изложены они явным образом или нет.

При упоминании в данном документе числовых диапазонов каждое промежуточное число в них охватывается явным образом с той же степенью точности. Например, в случае диапазона 6-9 в дополнение к 6 и 9 охватываются числа 7 и 8, а в случае диапазона 6,0-7,0 явным образом охватываются числа 6,0, 6,1, 6,2, 6,3, 6,4, 6,5, 6,6, 6,7, 6,8, 6,9 и 7,0.

Если не определено иное, все технические и научные термины, используемые в данном документе, имеют то же значение, которое обычно понимает специалист в данной области. В случае противоречий данный документ, содержащий определения, будет иметь преимущественную силу. Предпочтительные способы и материалы описаны ниже, хотя при практическом осуществлении или тестировании настоящего изобретения можно использовать способы и материалы, аналогичные или эквивалентные описанным в данном документе. Все публикации, заявки на патенты, патенты и другие ссылочные материалы, упомянутые в данном документе, включены посредством ссылки во всей своей полноте. Материалы, способы и примеры, раскрытые в данном документе, являются только иллюстративными и не подразумеваются как ограничивающие.

"Аденоассоциированный вирус" или "AAV", используемый в данном документе взаимозаменяемо, относится к небольшому вирусу, принадлежащему к роду Dependovirus семейства Parvoviridae, который инфицирует людей и некоторые другие виды приматов. В настоящее время AAV не вызывает болезни, и, следовательно, вирус вызывает очень умеренный иммунный ответ.

Используемый в данном документе термин "область связывания" относится к области в целевой области для нуклеазы, которая распознается и связывается нуклеазой, такой как Cas9.

Используемый в данном документе термин "хроматин" относится к организованному комплексу хромосомной ДНК, связанному с гистонами.

"Цис-регуляторные элементы" или "CRE", используемые в данном документе взаимозаменяемо, относятся к областям некодирующей ДНК, которые регулируют транскрипцию близлежащих генов. CRE выявлены рядом с геном или генами, которые они регулируют. Как правило, CRE регулируют транскрипцию гена, функционируя в качестве сайтов связывания для факторов транскрипции. Примеры CRE включают промоторы, энхансеры, супер-энхансеры, сайленсеры, инсуляторы и регуляторные области локуса.

"Короткие палиндромные повторы, регулярно расположенные группами" и "CRISPR", используемые в данном документе взаимозаменяемо, относятся к локусам, содержащим множественные короткие

прямые повторы, которые встречаются в геномах примерно 40% секвенированных бактерий и 90% секвенированных архей.

Используемое в данном документе выражение "кодирующая последовательность" или "кодирующая нуклеиновая кислота" означает нуклеиновые кислоты (молекулу РНК или ДНК), которые содержат нуклеотидную последовательность, кодирующую белок. Кодированная последовательность может дополнительно включать сигналы инициации и терминации, функционально связанные с регуляторными элементами, включая промотор и сигнал полиаденилирования, способный управлять экспрессией в клетках индивидуума или млекопитающего, которому вводят нуклеиновую кислоту. Кодированная последовательность может быть кодон-оптимизированной.

Используемое в данном документе выражение "комплементарная последовательность" или "комплементарное" означает, что нуклеиновая кислота может образовывать пары оснований согласно Уотсону-Крику (например, А-Т/У и С-Г) или Хугстону между нуклеотидами или нуклеотидными аналогами молекул нуклеиновых кислот. "Комплементарность" относится к свойству, разделяемому двумя последовательностями нуклеиновой кислоты, за счет которого, когда они выровнены антипараллельно друг другу, нуклеотидные основания в каждом положении будут комплементарными.

Термины "корректирование", "редактирование генома" и "восстановление", используемые в данном документе, относятся к изменению мутантного гена, который кодирует усеченный белок или вообще не кодирует какой-либо белок, за счет чего достигается экспрессия полноразмерного функционального или полноразмерного частично функционального белка. Корректирование или восстановление мутантного гена может включать в себя замену области гена, которая имеет мутацию, или замену всего мутантного гена копией гена, которая не имеет мутации, с помощью механизма репарации, такого как репарация с участием гомологичной рекомбинации (HDR). Корректирование или восстановление мутантного гена может также включать в себя репарацию мутации со сдвигом рамки, которая является причиной преждевременного стоп-кодона, абберрантного сайта акцептора сплайсинга или абберрантного сайта донора сплайсинга, путем создания двухнитевого разрыва в гене, который затем репарируют с использованием негомологичного соединения концов (NHEJ). За счет NHEJ может добавляться или удаляться по меньшей мере одна пара оснований в ходе репарации, что может восстанавливать соответствующую рамку считывания и устранять преждевременный стоп-кодон. Коррекция или восстановление мутантного гена также может включать в себя разрушение абберрантного сайта акцептора сплайсинга или последовательности донора сплайсинга. Коррекция или восстановление мутантного гена может также включать удаление сегмента гена, не являющегося важным, путем одновременного воздействия двух нуклеаз на одну и ту же нить ДНК, чтобы восстановить соответствующую рамку считывания путем удаления ДНК между двумя сайтами-мишенями для нуклеаз и репарации разрыва ДНК с помощью NHEJ.

Используемый в данном документе термин "деметилазы" относится к ферменту, который удаляет метильные (CH₃-) группы из нуклеиновых кислот, белков (в частности, гистонов) и других молекул. Ферменты-деметилазы важны в механизмах эпигенетической модификации. Белки-деметилазы изменяют регуляцию транскрипции генома за счет осуществления контроля уровней метилирования, которые имеют место на ДНК и гистонах, и, в свою очередь, регулируют состояние хроматина в определенных локусах генов внутри организмов. Выражение "гистоновая деметилаза" относится к деметилазе, которая удаляет метильные группы из гистонов. Существует несколько семейств гистоновых деметилаз, которые действуют на разные субстраты и играют различную роль в клеточной функции. Зависимые от Fe(II) лизиновые деметилазы могут представлять собой деметилазу JMJC. Деметилаза JMJC является гистоновой деметилазой, содержащей домен JumonjiC (JmjC). Деметилаза JMJC может быть представителем семейства гистоновых деметилаз KDM3, KDM4, KDM5 или KDM6.

Выражение "сверхчувствительные к ДНКазе I сайты" или "DHS", используемые в данном документе взаимозаменяемо, относятся к сайтам докинга для факторов транскрипции и модификаторов хроматина, включая p300, которые координируют экспрессию дистального целевого гена.

Термины "донорная ДНК", "донорная матрица" и "матрица репарации", используемые в данном документе взаимозаменяемо, относятся к фрагменту или молекуле двухнитевой ДНК, которая включает в себя по меньшей мере часть представляющего интерес гена. Донорная ДНК может кодировать полнофункциональный белок или частично функциональный белок.

Термин "эндогенный ген", используемый в данном документе, относится к гену, который происходит из организма, ткани или клетки. Эндогенный ген является нативным для клетки, находится в своем нормальном геномном и хроматиновом окружении и не является гетерологичным по отношению к клетке. Такие клеточные гены включают, например, гены животных, гены растений, гены бактерий, гены простейших, гены грибов, митохондриальные гены и гены хлоропластов. Выражение "эндогенный целевой ген", используемое в данном документе, относится к эндогенному гену, на который нацеливается оптимизированная gRNA и система на основе CRISPR/Cas9 или на основе CRISPR/Cpf1.

Термин "энхансер", используемый в данном документе, относится к некодирующим последовательностям ДНК, содержащим множественные сайты связывания активатора и репрессора. Энхансеры изменяются в диапазоне от 50 до 1500 п. о. в длину и могут быть либо проксимальными, в 5'-сторону относительно промотора, в пределах любого интрона регулируемого гена, либо дистальными, в интронах

соседних генов или в межгенных областях вдали от локуса, или в областях на разных хромосомах. Более одного энхансера могут взаимодействовать с промотором. Аналогичным образом энхансеры могут регулировать более одного гена без ограничения сцепления и могут "пропускать" соседние гены для регуляции более отдаленных генов. Регуляция транскрипции может вовлекать элементы, расположенные в хромосоме, отличающейся от той, где находится промотор. Проксимальные энхансеры или промоторы соседних генов могут служить в качестве платформ для привлечения более дистальных элементов.

Термин "мышечная дистрофия Дюшенна" или "DMD", используемый в данном документе взаимозаменяемо, относится к рецессивному, фатальному, X-сцепленному нарушению, которое приводит к дегенерации мышц и впоследствии к смертельному исходу. DMD является распространенным наследственным моногенным заболеванием и встречается у 1 из 3500 мужчин. DMD является результатом врожденных или спонтанных мутаций, которые вызывают нонсенс-мутации или мутации со сдвигом рамки в гене дистрофина. Большинство мутаций дистрофина, которые являются причиной DMD, представляют собой делеции экзонов, которые разрушают рамку считывания и приводят к преждевременной терминации трансляции в гене дистрофина. Пациенты с DMD обычно теряют способность физически поддерживать на протяжении детства, постепенно становятся слабее в подростковом возрасте и умирают в возрасте от двадцати до тридцати лет.

Используемый в данном документе термин "дистрофин" относится к стержнеобразному цитоплазматическому белку, который является частью белкового комплекса, соединяющего цитоскелет мышечного волокна с окружающим внеклеточным матриксом через клеточную мембрану. Дистрофин обеспечивает структурную стабильность дистрогликанового комплекса клеточной мембраны, который отвечает за регуляцию целостности и функции мышечных клеток. Ген дистрофина или "ген DMD", используемые в данном документе взаимозаменяемо, образован 2,2 миллионами пар оснований в локусе Xp21. Размер первичного транскрипта составляет приблизительно 2400 т.о., при этом размер зрелой мРНК составляет приблизительно 14 т.о. 79 экзонов кодируют белок, образованный более чем 3500 аминокислотами.

"Экзон 51", используемый в данном документе, относится к 51^{му} экзону гена дистрофина. Экзон 51 часто является смежным с положениями делеций, разрушающих рамку считывания, у пациентов с DMD, и в клинических испытаниях на него был направлен пропуск экзона, основанный на применении олигонуклеотидов. Недавно в клиническом испытании с пропуском экзона 51 с помощью соединения этерлипсена сообщали о значительном положительном функциональном эффекте в течение 48 недель со средним количеством дистрофин-положительных волокон 47% по сравнению с исходным уровнем. Мутации в экзоне 51 идеально подходят для устойчивой коррекции посредством редактирования генома на основе NHEJ.

Выражения "сдвиг рамки" или "мутация сдвига рамки", используемые в данном документе взаимозаменяемо, относятся к типу мутации гена, при которой добавление или делеция одного или нескольких нуклеотидов вызывает сдвиг в рамке считывания кодонов в мРНК. Сдвиг в рамке считывания может привести к изменению аминокислотной последовательности при трансляции белка, такой как миссенс-мутация или преждевременный стоп-кодон.

Выражение "полноразмерная gRNA" или "стандартная gRNA", используемое в данном документе взаимозаменяемо, относится к gRNA, которая включает в себя "каркас" и последовательность или сегмент, нацеливающиеся на протоспейсер, которые обычно составляют 20 нуклеотидов в длину.

Выражение "функциональный" и "полнофункциональный", используемое в данном документе взаимозаменяемо, описывает белок, который обладает биологической активностью. Выражение "функциональный ген" относится к гену, транскрибируемому в мРНК, которая подвергается трансляции в функциональный белок.

Выражение "слитый белок", используемое в данном документе, относится к химерному белку, созданному путем соединения двух или более генов, которые исходно кодировали отдельные белки. Трансляция гибридного гена приводит к получению одного полипептида с функциональными свойствами, полученными от каждого из исходных белков.

Выражение "генетическая конструкция", используемое в данном документе, относится к молекулам ДНК или РНК, которые содержат нуклеотидную последовательность, кодирующую белок. Кодированная последовательность включает в себя сигналы инициации и терминации, функционально связанные с регуляторными элементами, в том числе промотор и сигнал полиаденилирования, способный управлять экспрессией в клетках индивидуума, которому вводят молекулу нуклеиновой кислоты. Используемое в данном документе выражение "экспрессируемая форма" относится к генным конструкциям, которые содержат необходимые регуляторные элементы, функционально связанные с кодирующей последовательностью, которая кодирует белок, за счет чего, когда он присутствует в клетке индивидуума, кодирующая последовательность будет экспрессироваться.

Используемое в данном документе выражение "генетическое заболевание" относится к заболеванию, частично или полностью, прямо или косвенно, вызванному одним или несколькими нарушениями в геноме, особенно к состоянию, которое присутствует от рождения. Нарушение может быть мутацией, вставкой или делецией. Нарушение может влиять на кодирующую последовательность гена или его регуляторную последовательность. Генетическое заболевание может представлять собой без ограничения

DMD, гемофилию, кистозный фиброз, хорею Гентингтона, семейную гиперхолестеринемию (дефект рецептора LDL), гепатобластому, болезнь Вилсона, врожденную печеночную порфирию, наследственные нарушения обменных процессов в печени, синдром Леша-Нихана, серповидно-клеточную анемию, талассемию, пигментную ксеродермию, анемию Фанкони, пигментный ретинит, атаксию телеангиэктазию, синдром Блума, ретинобластому и болезнь Тай-Сакса.

Термин "геном", используемый в данном документе, относится к полному набору генов или генетического материала, присутствующего в клетке или организме. Геном включает в себя ДНК или РНК в РНК-вирусах. Геном включает в себя как гены (кодирующие области), так и некодирующую ДНК и геномы митохондрий и хлоропластов.

Термины "направляющая РНК", "gRNA", "одиночная gRNA" и "sgRNA", используемые в данном документе взаимозаменяемо, относятся к короткой синтетической РНК, состоящей из последовательности "каркаса", необходимой для связывания с Cas9 или связывания с Cpf1, и определяемого пользователем "спейсера" или "нацеливающейся последовательности" (также называемой в данном документе как последовательность или сегмент, нацеливающейся на протоспейсер), которая определяет геномную мишень, подлежащую модификации. Термины "hpgRNA", "hp-gRNA" и "оптимизированная gRNA", используемые в данном документе взаимозаменяемо, относятся к gRNA, которая имеет дополнительные нуклеотиды с 5'-конца или 3'-конца, которые могут образовывать вторичную структуру со всеми последовательностью или сегментом, нацеливающимися на протоспейсер, или их частью.

Термин "гистоновые ацетилтрансферазы" или "НАТ", используемый в данном документе взаимозаменяемо, относится к ферментам, которые ацетируют консервативные аминокислоты лизина на гистоновых белках путем переноса ацетильной группы с ацетил-CoA с образованием ϵ -N-ацетиллизина. ДНК обернута вокруг гистонов и при переносе ацетильной группы на гистоны гены могут включаться и выключаться. В целом ацетилирование гистонов повышает экспрессию генов, поскольку оно связано с активацией транскрипции и ассоциировано с эухроматином. Гистоновые ацетилтрансферазы также могут ацетилировать негистоновые белки, такие как ядерные рецепторы и другие транскрипционные факторы для облегчения экспрессии генов.

Термин "гистоновые деацетилазы" или "HDAC", используемый в данном документе взаимозаменяемо, относится к классу ферментов, которые удаляют ацетильные группы ($O=C-CH_3$) из аминокислоты ϵ -N-ацетиллизин на гистоне, что позволяет гистонам обертывать ДНК более плотно. HDAC также называют лизиндеацетилазами (KDAC) с целью описания их функции, а не их мишени, которая также включает негистоновые белки.

Термин "гистоновая метилтрансфераза" или "HMT", используемый в данном документе взаимозаменяемо, относится к гистон-модифицирующим ферментам (например, гистон-лизин N-метилтрансферазам и гистон-аргинин N-метилтрансферазам), которые катализируют перенос одной, двух или трех метильных групп на остатки лизина и аргинина гистоновых белков. Присоединение метильных групп происходит преимущественно на конкретных остатках лизина или аргинина на гистонах H3 и H4.

Термин "репарация с участием гомологичной рекомбинации" или "HDR", используемый в данном документе взаимозаменяемо, относится к механизму в клетках для репарации двухнитевых повреждений ДНК, когда в ядре присутствует гомологичная часть ДНК, в основном в фазе клеточного цикла G2 и S. В HDR используется матрица донорной ДНК для управления репарацией, и ее можно использовать для создания специфических изменений последовательности в геноме, включая целенаправленное добавление целых генов. Если донорная матрица предоставляется вместе с сайт-специфической нуклеазой, например с системой на основе CRISPR/Cas9 или системой на основе CRISPR/Cpf1, то клеточный аппарат осуществляет репарацию разрыва путем гомологичной рекомбинации, которая усиливается на несколько порядков в присутствии расщепления ДНК. Когда фрагмент гомологичной ДНК отсутствует, вместо этого может происходить негомологичное соединение концов.

Термин "геном", используемый в данном документе, относится к полному набору генов или генетического материала, присутствующего в клетке или организме. Геном включает в себя ДНК или РНК в РНК-вирусах. Геном включает в себя как гены (кодирующие области), так и некодирующую ДНК и геномы митохондрий и хлоропластов.

Выражение "редактирование генома", используемое в данном документе, относится к изменению гена. Редактирование генома может включать в себя корректирование или восстановление мутантного гена. Редактирование генома может включать в себя нокаут гена, такого как мутантный ген или нормальный ген. Редактирование генома можно использовать для лечения заболеваний или улучшения восстановления мышц путем изменения представляющего интерес гена.

Выражение "идентичная" или "идентичность", используемое в данном документе в контексте двух или более нуклеиновых кислот или полипептидных последовательностей, означает, что последовательности имеют определенный процент остатков, которые являются одинаковыми в указанной области. Процент может быть рассчитан путем оптимального выравнивания двух последовательностей, сравнения двух последовательностей в указанной области, определения количества положений, в которых одинаково-

вый остаток встречается в обеих последовательностях, с получением количества совпадающих положений, разделяя количество совпадающих положений на общее количество положений в указанной области и умножая результат на 100, чтобы получить процент идентичности последовательности. В тех случаях, когда две последовательности имеют разную длину или выравнивание дает в результате одну или несколько ступеней в шахматном порядке, а указанная область сравнения включает только одну последовательность, то при расчете эти остатки одной последовательности включаются в знаменатель, но не в числитель. При сравнении ДНК и РНК тимин (Т) и урацил (U) можно считать эквивалентом. Идентификацию можно проводить вручную или с использованием компьютерного алгоритма для работы с последовательностями, такого как BLAST или BLAST 2.0.

Термин "инсуляторы", используемый в данном документе, относится к генетическому пограничному элементу, который блокирует взаимодействие между энхансерами и промоторами. При расположении между энхансером и промотором инсулятор может ингибировать их последующие взаимодействия. Инсуляторы могут определять набор генов, на которые может оказывать влияние энхансер. Инсуляторы необходимы в тех случаях, когда два соседних гена на хромосоме имеют значительно отличающиеся режимы транскрипции, а индуцирующие или подавляющие механизмы одного гена не препятствуют функции соседнего гена. Было обнаружено, что инсуляторы также группируются на границах доменов топологической ассоциации (TAD) и могут играть роль в разделении генома на "районы хромосом" - геномные области, в пределах которых происходит регуляция. Считается, что активность инсулятора осуществляется главным образом за счет 3D-структуры ДНК, опосредованной белками, в том числе CTCF. Инсуляторы могут функционировать посредством нескольких механизмов. Многие энхансеры образуют петли ДНК, которые при активации транскрипции помещают их в тесной физической близости с промоторными областями. Инсуляторы могут способствовать образованию петель ДНК, которые препятствуют образованию петель промотор-энхансер. Барьерные инсуляторы могут препятствовать распространению гетерохроматина от молчащего гена к активно транскрибируемому гену.

Выражение "встраивание", используемое в данном документе, относится к разрушению дуплекса ДНК в области протоспейсера в целевой области целевого гена, например с помощью gRNA, которая связывается с последовательностью ДНК, комплементарной протоспейсеру.

Термин "кинетика встраивания", используемый в данном документе, относится к скорости, с которой происходит встраивание. Кинетика встраивания может относиться к скорости, с которой направляющая РНК встраивается в дуплекс, либо к "полному встраиванию", так что протоспейсер полностью встроен, или к скорости, с которой сегмент ДНК протоспейсера, связанный с направляющей РНК, удлиняется по мере его вытеснения из его комплементарной нити и связывания с направляющей РНК нуклеотид за нуклеотидом от его РАМ-проксимальной области до полного встраивания.

Выражение "время нахождения", используемое в данном документе, относится к периоду времени, в течение которого gRNA остается встроеной в область в целевой области целевого гена.

Выражение "области контроля локуса", используемое в данном документе, относится к цисрегуляторному элементу дальнего действия, который усиливает экспрессию связанных генов на дистальных участках хроматина. Он функционирует в зависимости от количества копий и является тканеспецифичным, как видно по избирательной экспрессии генов β -глобина в эритроидных клетках. Уровни экспрессии генов могут быть модифицированы с помощью LCR и проксимальных по отношению к гену элементов, таких как промоторы, энхансеры и сайленсеры. LCR функционирует путем привлечения хроматин-модифицирующих, коактиваторных и транскрипционных комплексов. Его последовательность является консервативной у многих позвоночных, при этом консервативность определенных сайтов может указывать на важность функции.

Выражение "ошибочно спаренный" или "ММ", используемое в данном документе взаимозаменяемо, относится к ошибочно спаренным основаниям, которые включают пары G/T или A/C. Ошибочные спаривания обычно связаны с таутомеризацией оснований во время G2 Повреждение репарируется путем распознавания деформации, вызванной ошибочным спариванием, определения матричной и нематричной нити и вырезания ошибочно встроеного основания и замены его соответствующим нуклеотидом.

Используемый в данном документе термин "модуляция" может означать любое изменение активности, такое как регуляция, понижающая регуляция, повышающая регуляция, снижение, ингибирование, увеличение, уменьшение, деактивация или активация.

Термин "мутантный ген" или "мутированный ген", используемый в данном документе взаимозаменяемо, относится к гену, который был подвергнут выявляемой мутации. Мутантный ген подвергся изменению, такому как потеря, получение или обмен генетического материала, что влияет на нормальную передачу и экспрессию гена. Термин "разрушенный ген", используемый в данном документе, относится к мутантному гену, в котором имеется мутация, являющаяся причиной преждевременного стоп-кодона. Продукт разрушенного гена усечен по сравнению с продуктом полноразмерного неразрушенного гена.

Выражение "путь негомологичного соединения концов (NHEJ)", используемое в данном документе, относится к пути репарации двухнитевых разрывов ДНК путем прямого лигирования концов разрыва без необходимости в гомологичной матрице. Независимое от матрицы повторное лигирование концов ДНК с помощью NHEJ представляет собой стохастический, подверженный ошибкам процесс репарации, кото-

рый вводит случайные микровставки и микроделеции (вставки-делеции) в точке разрыва ДНК. Этот способ можно использовать для намеренного разрушения, делеции или изменения рамки считывания последовательностей целевых генов. В NHEJ обычно используют короткие гомологичные последовательности ДНК, называемые микрогомологиями, для управления репарацией. Эти микрогомологии часто присутствуют в одонитевых липких концах на конце двухнитевых разрывов. Когда липкие концы идеально совместимы, то NHEJ обычно репарирует разрыв точно, однако может иметь место неточная репарация, приводящая к потере нуклеотидов, но гораздо более часто липкие концы несовместимы.

Используемое в данном документе выражение "нормальный ген" относится к гену, который не подвергался изменениям, таким как потеря, получение или обмен генетического материала. Нормальный ген подвергается нормальной передаче генов и экспрессии генов.

Выражение "опосредованное нуклеазой NHEJ", используемое в данном документе, относится к NHEJ, которое инициируется после того, как нуклеаза, такая как cas9, разрезает двухнитевую ДНК.

Используемые в данном документе термины "нуклеиновая кислота", или "олигонуклеотид", или "полинуклеотид" означают по меньшей мере два нуклеотида, ковалентно связанных друг с другом. Изображение одиночной нити также определяет последовательность комплементарной нити. Таким образом, нуклеиновая кислота также охватывает комплементарную нить изображенной одиночной нити. Множество вариантов нуклеиновой кислоты можно использовать для той же цели, как и указанную нуклеиновую кислоту. Таким образом, нуклеиновая кислота также охватывает по сути идентичные нуклеиновые кислоты и их комплементарные нити. Одиночная нить обеспечивает зонд, который может гибридизироваться с целевой последовательностью в жестких условиях гибридизации. Таким образом, нуклеиновая кислота также охватывает зонд, который гибридизуется в жестких условиях гибридизации.

Нуклеиновая кислота может быть одонитевой или двухнитевой или может содержать части как двухнитевой, так и одонитевой последовательности. Нуклеиновая кислота может представлять собой ДНК, как геномную, так и кДНК, РНК или гибридную молекулу, где нуклеиновая кислота может содержать комбинации дезоксирибо- и рибонуклеотидов, и при этом комбинации оснований включают урацил, аденин, тимин, цитозин, гуанин, инозин, ксантин, гипоксантин, изоцитозин и изогуанин. Нуклеиновые кислоты можно получать с помощью способов химического синтеза или с помощью рекомбинантных способов.

Термин "целевой сайт", используемый в данном документе, относится к целевой области или последовательности в геноме, на которую предусмотрено нацеливание gRNA. В идеальном случае целевой сайт имеет идеальную гомологию (100% идентичность или гомологию) с последовательностью целевой ДНК без гомологии в другом месте генома.

Термин "нецелевой сайт", используемый в данном документе, относится к области генома, которая имеет частичную гомологию или частичную идентичность с целевым сайтом или целевой областью gRNA, но gRNA не предназначена или не сконструирована для нацеливания на него.

Используемое в данном документе выражение "функционально связанный" означает, что экспрессия гена находится под контролем промотора, с которым он пространственно соединен. Промотор может располагаться 5' (выше) или 3' (ниже) относительно гена под его контролем. Расстояние между промотором и геном может быть приблизительно таким же, как и расстояние между этим промотором и геном, который он контролирует, в гене, из которого получен промотор. Как известно в данной области, изменение этого расстояния может быть осуществлено без потери промоторной функции.

Термины "белок p300", "EP300" или "E1A-связывающий белок p300", используемые в данном документе взаимозаменяемо, относятся к ассоциированному с аденовирусом E1A клеточному белку-коактиватору транскрипции p300, кодируемому геном EP300. p300 представляет собой высококонсервативную ацетилтрансферазу, принимающую участие в широком спектре клеточных процессов. p300 функционирует как гистоновая ацетилтрансфераза, которая регулирует транскрипцию посредством ремоделирования хроматина и принимает участие в процессах клеточной пролиферации и дифференцировки клеток.

Выражение "частично функциональный", используемое в данном документе, описывает белок, который кодируется мутантным геном и характеризуется меньшей биологической активностью, чем функциональный белок, но большей, чем нефункциональный белок.

Выражение "преждевременный стоп-кодон" или "стоп-кодон вне рамки", используемые в данном документе взаимозаменяемо, относятся к нонсенс-мутации в последовательности ДНК, которая является причиной стоп-кодона в месте, которое обычно не обнаруживается в гене дикого типа. Преждевременный стоп-кодон может приводить к усечению или укорочению белка по сравнению с полноразмерной версией белка.

Используемый в данном документе термин "первичная клетка" относится к клеткам, отобраным непосредственно из живой ткани (например, материала-биоптата). Первичные клетки могут быть предназначены для выращивания *in vitro*. Эти клетки претерпели лишь несколько удвоенной популяции и поэтому являются более характерными для основного функционального компонента ткани, из которой они получены, по сравнению с непрерывными (опухолевыми или искусственно immortalizированными) клеточными линиями, что представляет собой в большей степени репрезентативную модель для состоя-

ния *in vivo*. Первичные клетки могут быть отобраны у разных видов, таких как мышь или человек.

Термин "последовательность протоспейсера" или "сегмент протоспейсера", используемое в данном документе взаимозаменяемо, относится к последовательности ДНК, на которую нацеливается нуклеаза Cas9 или нуклеаза Cpf1 в бактериальной адаптивной иммунной системе CRISPR. В системе CRISPR/Cas9 за последовательностью протоспейсера обычно располагается смежный с протоспейсером мотив (PAM); при этом PAM расположен на 5'-конце. В системе CRISPR/Cpf1 за PAM располагается последовательность протоспейсера; PAM расположен на 3'-конце.

Выражение "нацеливающаяся на протоспейсер последовательность" или "нацеливающийся на протоспейсер сегмент", используемое в данном документе взаимозаменяемо, относится к нуклеотидной последовательности gRNA, которая соответствует последовательности протоспейсера и облегчает нацеливание системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1 на последовательность протоспейсера.

Используемый в данном документе термин "промотор" означает молекулу синтетического или природного происхождения, которая способна обеспечивать, активировать или усиливать экспрессию нуклеиновой кислоты в клетке. Промотор может содержать одну или несколько специфических последовательностей регуляции транскрипции для дополнительного усиления экспрессии и/или для изменения пространственной экспрессии и/или временной экспрессии. Промотор может также содержать дистальные энхансерные или репрессорные элементы, которые могут располагаться в нескольких тысячах пар оснований или в любом месте генома от сайта инициации транскрипции. Промотор может быть получен из источников, включая вирусы, бактерии, грибы, растения, насекомых и животных. Промотор может регулировать экспрессию генетического компонента конститутивно или дифференциально с учетом клетки, ткани или органа, в которых происходит экспрессия, или с учетом стадии развития, на которой происходит экспрессия, или в ответ на внешние раздражители, такие как физиологические стрессы, гормоны, токсины, лекарственные средства, патогены, ионы металлов или индуцирующие средства. Типичные примеры промоторов включают промотор бактериофага T7, промотор бактериофага T3, промотор SP6, промотор лактозного оперона, промотор *tac*, поздний промотор SV40, ранний промотор SV40, промотор RSV-LTR, промотор IE CMV, ранний промотор SV40 или поздний промотор SV40 и промотор IE CMV.

Выражение "прилегающий к протоспейсеру мотив" или "PAM", используемое в данном документе, относится к последовательности ДНК, которая располагается непосредственно за последовательностью ДНК, на которую нацеливается Cas9, или располагается непосредственно перед последовательностью ДНК, на которую нацеливается нуклеаза Cpf1 в бактериальной адаптивной иммунной системе CRISPR. PAM является компонентом встраивающихся вируса или плазмиды, но не является компонентом локуса бактериальной CRISPR. Cas9 и Cpf1 не будут успешно связывать или расщеплять последовательность целевой ДНК, если после нее не расположена или ей не предшествует последовательность PAM соответственно. PAM является важным компонентом нацеливания (не выявленным в бактериальном геноме), который отличает собственную бактериальную ДНК от несобственной ДНК, предотвращая тем самым нацеливание на локус CRISPR и его разрушение нуклеазой.

Термин "рекомбинантный" при использовании в отношении, например, клетки или нуклеиновой кислоты, белка или вектора указывает на то, что клетка, нуклеиновая кислота, белок или вектор были модифицированы путем введения гетерологичных нуклеиновой кислоты или белка или изменения нативных нуклеиновой кислоты или белка, или что клетка получена из клетки, модифицированной таким образом. Таким образом, например, рекомбинантные клетки экспрессируют гены, не выявляемые в нативной (природного происхождения) форме клетки, или экспрессируют вторую копию нативного гена, которая при иных обстоятельствах характеризуется нормальной или аномальной экспрессией, недостаточной экспрессией или вовсе не экспрессируется.

Термины "сайленсеры" или "репрессоры", используемые в данном документе взаимозаменяемо, относятся к последовательности ДНК, способной связывать факторы регуляции транскрипции и предотвращать экспрессию генов в виде белков. Сайленсер является специфичным по отношению к последовательности элементом, который индуцирует отрицательное воздействие на транскрипцию своего конкретного гена. Есть множество положений, в которых элемент-сайленсер может располагаться в ДНК. Наиболее распространенное положение находится выше относительно целевого гена, где оно может способствовать подавлению транскрипции гена. Это расстояние может сильно варьироваться от приблизительно -20 п.о. до -2000 п.о. выше относительно гена. Определенные сайленсеры могут находиться ниже по отношению к промотору, расположенному внутри интрона или экзона самого гена. Сайленсеры также были обнаружены в 3'-нетранслируемой области (3' UTR) мРНК. В ДНК существует два основных типа сайленсеров, представляющих собой классический элемент-сайленсер и неклассический отрицательный регуляторный элемент (NRE). В случае классических сайленсеров ген активно подавляется элементом-сайленсером, главным образом, путем вмешательства в сборку общего фактора транскрипции (GTF). NRE пассивно подавляют ген, как правило, путем ингибирования других элементов, которые находятся выше относительно гена.

Термин "скелетная мышца", используемый в данном документе, относится к типу поперечно-

полосатой мышцы, которая находится под контролем соматической нервной системы и прикреплена к костям пучками коллагеновых волокон, известных как сухожилия. Скелетная мышца состоит из отдельных компонентов, известных как миоциты, или "мышечные клетки", иногда в разговорной речи называемые "мышечными волокнами". Миоциты образуются в результате слияния развивающихся миобластов (тип эмбриональной клетки-предшественника, из которой возникает мышечная клетка) в процессе, известном как миогенез. Эти длинные цилиндрические многоядерные клетки также называются мышечными волокнами.

Выражение "состояние скелетных мышц", используемое в данном документе, относится к состоянию, связанному со скелетной мышцей, как например: мышечные дистрофии, старение, дегенерация мышц, заживление ран и мышечная слабость или атрофия.

Термины "субъект" и "пациент", используемые в данном документе взаимозаменяемо, относятся к любому позвоночному животному, включая без ограничения млекопитающее (например, корову, свинью, верблюда, ламу, лошадь, козу, кролика, овцу, хомяков, морскую свинку, кошку, собаку, крысу и мышь, приматов, не относящихся к человеку (например, обезьяну, такую как яванский макак или макак-резус, шимпанзе и т.д.), и человека). В некоторых вариантах осуществления субъект представляет собой человека или не является человеком. Субъекта или пациента можно подвергать другим формам лечения.

Выражение "супер-энхансер", используемое в данном документе, относится к области генома млекопитающего, содержащей множественные энхансеры, которые в совокупности связываются с комплексом белков транскрипционных факторов для управления транскрипцией генов, принимающих участие в формировании клеточной идентичности. Супер-энхансеры часто выявляют рядом с генами, являющимися важными для контроля и определения клеточной идентичности, и их можно использовать для быстрого определения ключевых точек, регулирующих клеточную идентичность. У энхансеров есть несколько количественно измеряемых признаков, которые имеют диапазон значений, и эти признаки обычно повышены у супер-энхансеров. Супер-энхансеры связаны с более высокими уровнями белков, регулирующих транскрипцию, и ассоциированы с более высоко экспрессируемыми генами. Экспрессия генов, ассоциированных с супер-энхансерами, особенно чувствительна к внесению изменений, что может облегчить переходы клеточного состояния или объяснить чувствительность ассоциированных с супер-энхансерами генов к малым молекулам, которые нацелены на транскрипцию.

Термин "энхансер-мишень", используемый в данном документе, относится к энхансеру, на который нацеливается gRNA и система на основе CRISPR/Cas9. Энхансер-мишень может располагаться в пределах целевой области.

Термин "целевой ген", используемый в данном документе, относится к любой нуклеотидной последовательности, кодирующей известный или предполагаемый генный продукт. Целевой ген может представлять собой мутированный ген, вовлеченный в генетическое заболевание.

Термины "целевая область", "целевая последовательность", "протоспейсер" или "последовательность протоспейсера", используемые в данном документе взаимозаменяемо, относятся к области целевого гена, на которую нацеливается система на основе CRISPR/Cas9 или система на основе CRISPR/Cpf1.

Термин "транскрибируемая область", используемый в данном документе, относится к области ДНК, которая транскрибируется в молекулу однонитевой РНК, известную как информационная РНК, что приводит в результате к переносу генетической информации с молекулы ДНК на информационную РНК. В ходе транскрипции РНК-полимераза считывает матричную нить в направлении 3'-5' и синтезирует РНК от 5' до 3'. Последовательность мРНК комплементарна нити ДНК.

Выражение "регуляторный элемент-мишень", используемое в данном документе, относится к регуляторному элементу, на который осуществляется нацеливание gRNA и системы на основе CRISPR/Cas9. Регуляторный элемент-мишень может находиться в целевой области.

Термин "транскрибируемая область", используемый в данном документе, относится к области ДНК, которая транскрибируется в молекулу однонитевой РНК, известную как информационная РНК, что приводит в результате к переносу генетической информации с молекулы ДНК на информационную РНК. В ходе транскрипции РНК-полимераза считывает матричную нить в направлении 3'-5' и синтезирует РНК от 5' до 3'. Последовательность мРНК комплементарна нити ДНК.

Термин "сайт инициации транскрипции" или "TSS", используемый взаимозаменяемо, относится к первому нуклеотиду транскрибируемой последовательности ДНК, где РНК-полимераза начинает синтезировать транскрипт РНК.

Используемый в данном документе термин "трансген" относится к гену или генетическому материалу, содержащему последовательность гена, которая была выделена из одного организма и введена в другой отличающийся организм. Этот ненативный сегмент ДНК может сохранять способность продуцировать РНК или белок в трансгенном организме или может изменять нормальную функцию генетического кода трансгенного организма. Введение трансгена может изменить фенотип организма.

Термин "tru gRNA", используемый в данном документе, относится к полноразмерной направляющей РНК с нуклеотидами, усеченными с ее 5'-конца, как правило, 2 нуклеотидами.

Термин "транс-регуляторные элементы", используемый в данном документе, относится к областям некодирующей ДНК, которые регулируют транскрипцию генов, удаленных от гена, из которого их

транскрибировали. Транс-регуляторные элементы могут находиться на одной и той же или на другой хромосоме относительно целевого гена. Примеры транс-регуляторных элементов включают энхансеры, супер-энхансеры, сайленсеры, инсуляторы и области контроля локуса.

Выражение "вариант", используемое в данном документе, в отношении нуклеиновой кислоты означает (i) часть или фрагмент эталонной нуклеотидной последовательности (в том числе нуклеотидные последовательности, которые имеют вставки или делеции по сравнению с эталонными нуклеотидными последовательностями); (ii) последовательность, комплементарную эталонной нуклеотидной последовательности или ее части; (iii) нуклеиновую кислоту, по сути, идентичную эталонной нуклеиновой кислоте или комплементарной ей последовательности; или (iv) нуклеиновую кислоту, которая гибридизируется в жестких условиях с эталонной нуклеиновой кислотой, последовательностью, комплементарной ей, или последовательностями, по сути, идентичными им.

"Вариант" в отношении пептида или полипептида, который за счет вставки, делеции или консервативной замены аминокислот отличается по аминокислотной последовательности, но сохраняет по меньшей мере один тип биологической активности. Вариант также может означать белок с аминокислотной последовательностью, которая, по сути, идентична аминокислотной последовательности эталонного белка, который сохраняет по меньшей мере один тип биологической активности. В данной области признается, что консервативная замена аминокислоты, т.е. замена аминокислоты другой аминокислотой с аналогичными свойствами (например, гидрофильностью, относительным количеством и распределением заряженных участков), обычно предусматривает незначительное изменение. Эти незначительные изменения можно частично идентифицировать, учитывая индекс гидропатичности аминокислот, как это понимают в уровне техники. Kyte et al., *J. Mol. Biol.* 157:105-132 (1982). Индекс гидропатичности аминокислоты определяется с учетом гидрофобности и заряда. Из уровня техники известно, что аминокислоты с аналогичными индексами гидропатичности можно заменять, и при этом функция белка по-прежнему будет сохраняться. В одном аспекте заменяют аминокислоты с индексами гидропатичности, составляющими ± 2 . Гидрофобность аминокислот также можно использовать для выявления замен, которые в результате будут давать белки, сохраняющие биологическую функцию. Учет гидрофильности аминокислот в контексте пептида позволяет рассчитывать наибольшую локальную усредненную гидрофильность этого пептида. Замены можно проводить с аминокислотами, у которых различие значений гидрофильности находится в пределах ± 2 . Как на индекс гидрофобности, так и на значение гидрофильности аминокислот влияет определенная боковая цепь этой аминокислоты. В соответствии с этим наблюдением понимают, что совместимость аминокислотных замен с биологической функцией зависит от относительного сходства аминокислот и, в частности, боковых цепей этих аминокислот, что выявляют на основе гидрофобности, гидрофильности, заряда, размера и других свойств.

Термин "вектор", используемый в данном документе, означает последовательность нуклеиновой кислоты, содержащую точку начала репликации. Вектор может представлять собой вирусный вектор, бактериофаг, бактериальную искусственную хромосому или искусственную хромосому дрожжей. Вектор может представлять собой ДНК- или РНК-вектор. Вектор может быть самореплицирующимся внехромосомным вектором, а предпочтительно представляет собой плазмидную ДНК. Например, вектор может кодировать Cas9 и по меньшей мере одну оптимизированную нуклеотидную последовательность gRNA любой из SEQ ID NO: 149-315, 321-323 и 326-329.

Если не указано иное, научные и технические термины, используемые применительно к настоящему изобретению, должны иметь значения, которые обычно понимаются специалистами в данной области. Например, любые цифровые или словесные обозначения на иллюстрациях и методики, используемые применительно к клеточной и тканевой культуре, молекулярной биологии, иммунологии, микробиологии, генетике и химии и гибридизации белка и нуклеиновых кислот, описанные в данном документе, являются такими, которые хорошо известны и обычно используются в данной области. Смысл и объем терминов должны быть четкими и понятными; в случае, однако, любой скрытой двусмысленности, определения, представленные в данном документе, превалируют над любым словарным или внешним определением. Кроме того, если иное не требуется по контексту, термины единственного числа включают множественное число, а термины множественного числа включают единственное число.

2. Система CRISPR

Система CRISPR представляет собой нуклеазную систему микроорганизмов, вовлечена в защиту от встраивающихся фагов и плазмид, что обеспечивает форму приобретенного иммунитета. Локусы CRISPR в микробных хозяевах могут содержать комбинацию CRISPR-ассоциированных (Cas) генов, а также элементов некодирующей РНК, способных программировать специфичность CRISPR-опосредованного расщепления нуклеиновой кислоты. Короткие сегменты чужеродной ДНК, называемые спейсерами, включаются в состав генома между повторами CRISPR и служат в качестве "памяти" о прошлых воздействиях. Cas9 образует комплекс с 3'-концом одиночной направляющей РНК ("sgRNA"), а пара белок-РНК распознает свою геномную мишень посредством комплементарного спаривания оснований между 5'-концом последовательности sgRNA и предварительно определенной последовательностью ДНК размером 20 п.о., известной как протоспейсер. Этот комплекс направлен на гомологичные локусы

патогенной ДНК через области, кодируемые в РНК CRISPR ("crRNA"), то есть протоспейсеры и мотивы, смежные с протоспейсером (PAM), в патогенном геноме. Некодирующая матрица CRISPR транскрибируется и расщепляется в пределах прямых повторов на короткие crRNA, содержащие отдельные спейсерные последовательности, которые направляют Cas-нуклеазы в целевой сайт (протоспейсер). Путем простого обмена последовательности распознавания размером 20 п.о. экспрессированной химерной sgRNA нуклеазу Cas9 можно направлять на новые мишени в геноме. Спейсеры CRISPR используют для распознавания и сайленсинга экзогенных генетических элементов с помощью способа, аналогичного RNAi у эукариотических организмов.

Известны три класса систем CRISPR (эффекторные системы типов I, II и III). Эффекторная система типа II осуществляет двухнитевой разрыв в целевой ДНК на четырех последовательных стадиях с использованием одного эффекторного фермента Cas9 для расщепления dsDNA. По сравнению с эффекторными системами типа I и типа III, для которых требуется несколько различных эффекторов, действующих как в виде комплекса, эффекторная система типа II может функционировать в альтернативной среде, такой как эукариотические клетки. Эффекторная система типа II состоит из длинной пре-crRNA, которая транскрибируется из спейсер-содержащего локуса CRISPR, белка Cas9 и tracrRNA, которая участвует в процессинге пре-crRNA. tracrRNA гибридизируются с областями повторов, разделяющими спейсеры пре-crRNA, инициируя тем самым расщепление dsRNA с помощью эндогенной РНКазы III. Это расщепление сопровождается вторым событием расщепления внутри каждого спейсера Cas9, производя зрелые crRNA, которые остаются связанными с tracrRNA и Cas9, образуя комплекс Cas9:crRNA-tracrRNA.

Было показано, что сконструированная форма эффекторной системы типа II из *Streptococcus pyogenes* функционирует в клетках человека для конструирования генома. В этой системе белок Cas9 направляли на целевые сайты в геноме с помощью синтетически восстановленной "направляющей РНК" ("gRNA", также используемой в данном документе взаимозаменяемо как химерная sgRNA, которая для Cas9 представляет собой слитую конструкцию crRNA-tracrRNA, что устраняет необходимость в РНКазе III и процессинге crRNA в целом.

Комплекс Cas9:crRNA-tracrRNA раскручивает ДНК-дуплекс и ищет последовательности, соответствующие crRNA, для расщепления. Распознавание мишени происходит при выявлении комплементарности между последовательностью "протоспейсера" в последовательности целевой ДНК и оставшейся спейсерной последовательности в crRNA. Cas9 опосредует расщепление целевой ДНК, если на 3'-конце протоспейсера также присутствует мотив, смежный с протоспейсером (PAM). Для нацеливания на протоспейсер последовательность должна непосредственно сопровождаться мотивом, смежным с протоспейсером (PAM), короткой последовательностью, распознаваемой нуклеазой Cas9, которая требуется для расщепления ДНК. Различные системы типа II имеют разные требования к PAM. Система CRISPR из *S. pyogenes* может иметь последовательность PAM для этой Cas9 (SpCas9) в виде 5'-NRG-3', где R представляет собой A или G и характеризует специфичность этой системы в клетках человека. Уникальной способностью системы на основе CRISPR/Cas9 является эффективная способность одновременно нацеливаться на несколько различных геномных локусов путем совместной экспрессии одного белка Cas9 с двумя или более sgRNA. Например, система типа II из *Streptococcus pyogenes* в естественных условиях предпочитает применение последовательности "NGG", где "N" может быть любым нуклеотидом, но также принимает другие последовательности PAM, такие как "NAG", в сконструированных системах (Hsu et al. (2013) *Nature Biotechnology*, 31, 827-832). Аналогично Cas9, полученная из *Neisseria meningitidis* (NmCas9), как правило, имеет нативный PAM из NNNNGATT, но имеет активность со множеством PAM, включая PAM NNNNGNNN с высокой степенью вырожденности (Esvelt et al. *Nature Methods* (2013) doi:10.1038/nmeth.2681).

3. Система на основе CRISPR/Cas9

В данном документе представлены системы CRISPR/Cas9, включающие в себя оптимизированную gRNA, такую как шпильчатая gRNA (также называемую в данном документе "hpgRNA" или "hp-gRNA"), которые позволяют улучшить нацеливание на ДНК для применения в эпигемном редактировании и регуляции транскрипции, как, например, в частности, расщепление представляющей интерес целевой области, такой как целевой ген, или активация или репрессия экспрессии целевого гена. Оптимизированные gRNA обеспечивают повышенную специфичность связывания с мишенью, уменьшая в то же время нецелевое связывание и нецелевую активность систем на основе CRISPR/Cas9 и CRISPR/Cpf1 путем модулирования времени нахождения в нецелевых местоположениях с тем, чтобы минимизировать любую активность в этих нецелевых сайтах.

Оптимизированная gRNA может модулировать активности Cas9-слитого белка путем модулирования времени нахождения Cas9 в этих местоположениях и модулирования общей кинетики встраивания без учета активности второго домена. Кроме того, связывание gRNA с протоспейсером на 5'-конце нацеливающего на протоспейсер сегмента также может быть связано с расщеплением за счет Cas9.

Снижение связывания с нецелевыми сайтами ограничивало бы потенциал полного встраивания/расщепления в этих нецелевых сайтах. Было показано, что разработанная форма эффекторной системы типа II из *Streptococcus pyogenes* функционирует в клетках человека для конструирования генома. В этой системе белок Cas9 направляли на геномные целевые сайты с помощью синтетически восстанов-

ленной "направляющей РНК" ("gRNA", также используемой в данном документе взаимозаменяемо как химерная одиночная направляющая РНК ("sgRNA")), которая для Cas9 представляет собой слитую конструкцию crRNA-tracrRNA, которая устраняет необходимость в РНКазе III и процессинге crRNA в целом. В данном документе представлены системы на основе CRISPR/Cas9 для применения в редактировании генома и лечении генетических заболеваний. Системы на основе CRISPR/Cas9 могут быть сконструированы для нацеливания на любой ген, в том числе гены, вовлеченные в генетическое заболевание, старение, регенерацию тканей или заживление ран. Системы на основе CRISPR/Cas9 могут включать в себя белок Cas9 или слитый белок Cas9 и по меньшей мере одну оптимизированную gRNA, как описано ниже. Слитый белок Cas9 может, например, включать в себя домен с активностью, отличающейся от той, которая является эндогенной для Cas9, такой как домен трансактивации.

Целевой ген может иметь мутацию, такую как мутация со сдвигом рамки или нонсенс-мутация. Если целевой ген имеет мутацию, которая является причиной преждевременного стоп-кодона, абберрантного сайта акцептора сплайсинга или абберрантного сайта донора сплайсинга, то система на основе CRISPR/Cas9 может быть сконструирована для распознавания и связывания нуклеотидной последовательности выше или ниже преждевременного стоп-кодона, абберрантного сайта акцептора сплайсинга или абберрантного сайта донора сплайсинга. Система на основе CRISPR-Cas9 может быть также использована для разрушения нормального сплайсинга гена путем целенаправленного воздействия на акцепторы и доноры сплайсинга, чтобы вызвать пропуск преждевременных стоп-кодонов или восстановить поврежденную рамку считывания. Система на основе CRISPR/Cas9 может опосредовать или не опосредовать нецелевые изменения в кодирующих белок областях генома.

i) Cas9

Система на основе CRISPR/Cas9 может включать в себя белок Cas9 или слитый белок Cas9. Белок Cas9 представляет собой эндонуклеазу, которая расщепляет нуклеиновую кислоту и кодируется локусами CRISPR, и принимает участие в системе CRISPR типа II. Белок Cas9 может быть из любых видов бактерий или архей, например *Streptococcus pyogenes*. Белок Cas9 может быть мутирован таким образом, что нуклеазная активность инактивирована. Инактивированный белок Cas9 из *Streptococcus pyogenes* (iCas9, также называемый "dCas9") без эндонуклеазной активности ранее нацеливали на гены в клетках бактерий, дрожжей и человека с помощью gRNA для сайленсинга экспрессии генов посредством стерических затруднений. Оба используемых в данном документе термина "iCas9" и "dCas9" относятся к белку Cas9, который имеет аминокислотные замены D10A и H840A и характеризуется инактивированной нуклеазной активностью. В некоторых вариантах осуществления можно использовать инактивированный белок Cas9 из *Neisseria meningitidis*, такой как NmCas9. Например, система на основе CRISPR/Cas9 может включать в себя iCas9 под SEQ ID NO: 1.

ii) Слитый белок Cas9

Система на основе CRISPR/Cas9 может включать в себя слитый белок из белка Cas9, который не обладает нуклеазной активностью, такого как dCas9, и второго домена. Второй домен может включать в себя домен активации транскрипции, такой как домен VP64 или домен р300, домен репрессии транскрипции, такой как домен KRAB, нуклеазный домен, домен фактора освобождения транскриптов, домен модификации гистонов, домен ассоциации нуклеиновых кислот, ацетилазный домен, деацетилазный домен, метилазный домен, такой как домен ДНК-метилазы, деметилазный домен, домен фосфорилирования, домен убиквитинилирования или домен сумоилирования. Второй домен может быть модификатором метилирования ДНК или петлеобразования хроматина.

В некоторых вариантах осуществления слитый белок может включать домен dCas9 и активатор транскрипции. Например, слитый белок может включать в себя аминокислотную последовательность под SEQ ID NO: 2. В других вариантах осуществления слитый белок может включать в себя домен dCas9 и репрессор транскрипции. Например, слитый белок содержит аминокислотную последовательность под SEQ ID NO: 3. В дополнительных аспектах слитый белок может включать в себя домен dCas9 и сайт-специфическую нуклеазу, которая отличается от нуклеазы Cas9 по активности.

Слитый белок может содержать два гетерологичных полипептидных домена, где первый полипептидный домен содержит белок Cas, а второй полипептидный домен не обладает нуклеазной активностью. Слитый белок может включать в себя белок Cas9 или мутированный белок Cas9, как описано выше, гибридный со вторым полипептидным доменом, который обладает нуклеазной активностью. Второй полипептидный домен может обладать нуклеазной активностью, которая отличается от нуклеазной активности белка Cas9. Нуклеазой или белком, обладающим нуклеазной активностью, является фермент, способный расщеплять фосфодиэфирные связи между нуклеотидными субъединицами нуклеиновых кислот. Нуклеазы обычно дополнительно разделяют на эндонуклеазы и экзонуклеазы, хотя некоторые из ферментов могут попадать в обе категории. Хорошо известными нуклеазами являются дезоксирибонуклеаза и рибонуклеаза.

(1) Система активации генов на основе CRISPR/Cas9

Система на основе CRISPR/Cas9 может представлять собой систему активации генов на основе CRISPR/Cas9, которая может активировать функцию регуляторных элементов с исключительной специфичностью эпигеномного редактирования. Систему активации генов на основе CRISPR/Cas9 можно ис-

пользовать для скрининга энхансеров, инсуляторов, сайленсеров и областей контроля локуса, на которые можно нацеливаться для увеличения или уменьшения экспрессии целевого гена. Эту технологию можно использовать для закрепления функции за предполагаемыми регуляторными элементами, идентифицированными с помощью геномных исследований, таких как проекты ENCODE и Roadmap Epigenomics.

Система активации генов на основе CRISPR/Cas9 может активировать экспрессию гена путем модификации метилирования ДНК, петлеобразования хроматина или катализа ацетилирования лизина 27 гистона H3 на его целевых сайтах, что приводит к устойчивой активации транскрипции целевых генов за счет промоторов и проксимальных и дистальных энхансеров. Система активации генов на основе CRISPR/Cas9 является высокоспецифичной и может нацеливаться на целевой ген с использованием только одной направляющей РНК. Система активации генов на основе CRISPR/Cas9 может активировать экспрессию одного гена или семейства генов путем нацеливания на энхансеры в дистальных местоположениях в геноме.

(a) Белок гистон-ацетилтрансфераза (НАТ)

Система активации генов на основе CRISPR/Cas9 может включать в себя белок гистон-ацетилтрансферазы, такой как белок p300, CREB-связывающий белок (CBP, аналог p300), GCN5 или PCAF или их фрагмент. Ацетилирование гистонов в регуляторных элементах с использованием программируемого слитого белка на основе CRISPR/Cas9 представляет собой эффективную стратегию увеличения экспрессии генов-мишеней. Гистон-ацетилтрансфераза на основе CRISPR/Cas9, которая может быть нацелена на любой сайт в геноме, характеризуется уникальной способностью активировать дистальные регуляторные элементы. Белок гистон-ацетилтрансфераза может включать в себя белок p300 человека или его фрагмент. Белок гистон-ацетилтрансферазы может включать в себя дикий тип белка p300 человека, или мутантный белок p300 человека, или их фрагменты. Белок гистон-ацетилтрансфераза может включать в себя коровый лизин-ацетилтрансферазный домен белка p300 человека, то есть ядро НАТ p300 (также известное как "p300 Core").

(b) Система активации CRISPR/dCas9^{p300 Core}

Белок p300 регулирует активность многих генов в тканях во всем организме. Белок p300 играет роль в регуляции роста и деления клеток, побуждая клетки созревать и принимать специализированные функции (дифференцироваться) и предупреждая рост злокачественных опухолей. Белок p300 может активировать транскрипцию путем связывания факторов транскрипции с комплексом белков, которые осуществляют транскрипцию в ядре клетки. Белок p300 также функционирует как гистон-ацетилтрансфераза, которая регулирует транскрипцию посредством ремоделирования хроматина.

Гибридный белок dCas9^{p300 Core} является высокоактивным и легко программируемым инструментом для искусственного манипулирования ацетилированием в целевых эндогенных локусах, что приводит к регуляции генов, контролируемых проксимальным и дистальным энхансерами. p300 Core ацетилирует лизин 27 на гистоне H3 (H3K27ac) и может обеспечить накопление H3K27ac. Гибридизация каталитического корового домена p300 с dCas9 может приводить к значительно более высокой трансактивации расположенных ниже генов, чем непосредственная гибридизация полноразмерного белка p300, несмотря на устойчивую экспрессию белка. Слитый белок dCas9^{p300 Core} также может проявлять повышенную способность к трансактивации по сравнению с dCas9^{VP64}, в том числе в контексте каркаса Nm-dCas9, особенно в областях дистального энхансера, в которых dCas9^{VP64} демонстрирует незначительную, если таковая имеется, измеримую ниже транскрипционную активность. Кроме того, dCas9^{p300 Core} демонстрирует точную и устойчивую специфичность транскрипции по всему геному. dCas9^{p300 Core} может быть способным к мощной активации транскрипции и одновременному накоплению ацетилирования в промоторах, на которые нацелен эпигенетически модифицированный энхансер.

dCas9^{p300 Core} может активировать экспрессию генов с помощью одиночной gRNA, которая нацеливается на промотор и/или охарактеризованный энхансер и связывает их. Эта технология также дает возможность синтетически трансактивировать дистальные гены из предполагаемых и известных регуляторных областей и упрощает трансактивацию посредством применения одного программируемого эффектора и одного целевого сайта. Эти возможности обеспечивают возможность мультиплексирования для одновременного нацеливания на несколько промоторов и/или энхансеров. p300, полученный от млекопитающих, может обеспечить преимущества над эффекторными доменами, полученными из вирусов, для применений *in vivo* путем минимизации потенциальной иммуногенности.

Активация генов с помощью dCas9^{p300-Core} является высокоспецифичной для целевого гена. В некоторых вариантах осуществления p300 Core включает в себя аминокислоты 1048-1664 из SEQ ID NO: 2 (т.е. SEQ ID NO: 4). В некоторых вариантах осуществления система активации генов на основе CRISPR/Cas9 включает в себя слитый белок dCas9^{p300 Core} под SEQ ID NO: 2 или слитый белок Nm-dCas9^{p300 Core} под SEQ ID NO: 5.

(2) Система репрессии генов на основе CRISPR/Cas9

Система на основе CRISPR/Cas9 может представлять собой систему репрессии генов на основе CRISPR/Cas9, которая может ингибировать функцию регуляторных элементов с исключительной специфичностью эпигеномного редактирования. В некоторых вариантах осуществления система репрессии генов на основе CRISPR/Cas9, такая как включающая в себя dCas9^{KRAB}, может оказывать влияние на ак-

тивность дистального энхансера путем высокоспецифического ремоделирования эпигенетического состояния целевых генетических локусов.

(а) Система репрессии генов CRISPR/dCas9^{KRAB}

Репрессор dCas9^{KRAB} представляет собой высокоспецифичный инструмент для эпигеномного редактирования, который может использоваться в скринингах с потерей функции для изучения функции генов и обнаружения мишеней для разработки лекарственных средств. dCas9^{KRAB} характеризуется исключительной специфичностью нацеливания на конкретный энхансер, осуществляет сайленсинг только целевых генов этого энхансера и создает репрессивное гетерохроматиновое окружение в этом сайте. dCas9^{KRAB} можно использовать для скрининга новых регуляторных элементов в эндогенном геномном контексте путем сайленсинга проксимальных или дистальных регуляторных элементов и соответствующих генных мишеней. Специфичность репрессоров dCas9^{KRAB} позволяет использовать их из-за специфичности по всему транскриптому для сайленсинга эндогенных генов. Эпигенетические механизмы нарушения в целевом локусе, такие как метилирование гистонов.

Домен KRAB, общий гетерохроматин-образующий мотив во встречающихся в природе факторах транскрипции типа "цинковые пальцы", был генетически связан с dCas9 для создания РНК-нацеливаемого синтетического репрессора, dCas9^{KRAB}. Kruppel-ассоциированный бокс ("KRAB") привлекает гетерохроматин-образующие факторы: Kap1, HP1, SETDB1, NuRD. Он индуцирует триметилирование H3K9, деацетилирование гистонов. Синтетические репрессоры на основе KRAB могут вызывать эффективный сайленсинг экспрессии отдельных генов, и их использовали для репрессии онкогенов, ингибирования репликации вирусов и лечения доминантно-негативных заболеваний.

4. Система на основе CRISPR/Cpf1

Раскрытая оптимизированная gRNA может использоваться с системой коротких палиндромных повторов, регулярно расположенных группами, из *Prevotella* и *Francisella* 1 или ("CRISPR/Cpf1"). Система CRISPR/Cpf1, технология редактирования ДНК, аналогичная системе CRISPR/Cas9, обнаружена в бактериях *Prevotella* и *Francisella* и предотвращает генетические повреждения, вызываемые вирусами. Cpf1 представляет собой РНК-направляемую эндонуклеазу системы CRISPR/Cas класса II, содержащей белок из 1300 аминокислот. Гены Cpf1 ассоциированы с локусом CRISPR, кодирующим эндонуклеазу, которая использует направляющую РНК для нахождения и расщепления вирусной ДНК. Cpf1 представляет собой более мелкую и более простую эндонуклеазу, чем Cas9, и имеет молекулу sgRNA меньшего размера (примерно в два раза меньше нуклеотидов по сравнению с Cas9), поскольку функционально Cpf1 не нуждается в tracrRNA и требуется только crRNA. Примеры Cpf1, которую можно использовать с оптимизированной gRNA, включают Cpf1 из бактерий *Acidaminococcus* и *Lachnospiraceae*.

Локусы Cpf1 кодируют белки Cas1, Cas2 и Cas4, которые характеризуются большей схожестью с такими белками системы типа I и III, чем типа II. Локус Cpf1 содержит смешанный альфа/бета-домен RuvC-I, после которого следуют спиральная область RuvC-II и домен, подобный домену типа "цинковые пальцы". Белок Cpf1 имеет RuvC-подобный эндонуклеазный домен, который подобен RuvC-домену Cas9. Cpf1 не имеет эндонуклеазного домена HNH, а на N-конце Cpf1 отсутствует альфа-спиральный блок распознавания Cas9. Структура домена Cpf1 CRISPR-Cas показывает, что Cpf1 является функционально уникальным, и классифицируется как система CRISPR типа V класса 2.

Система CRISPR/Cpf1 состоит из фермента Cpf1 и направляющей РНК, которая находит и располагает комплекс в точном месте на двойной спирали для расщепления целевой ДНК. В механизме активности систем CRISPR/Cpf1 выделяют три стадии: адаптацию, образование crRNA и интерференцию. На стадии адаптации белки Cas1 и Cas2 облегчают адаптацию небольших фрагментов ДНК к комплексу CRISPR. Стадия образования crRNA включает процессинг пре-cr-RNA с получением зрелых crRNA для направления белка Cas. На стадии интерференции Cpf1 связывается с crRNA с образованием бинарного комплекса для идентификации и расщепления последовательности целевой ДНК.

Комплекс Cpf1-crRNA расщепляет целевую ДНК или РНК путем идентификации мотива, смежного с протоспейсером, 5'-YTN-3' (где "Y" представляет собой пиримидин, а "N" представляет собой любое нуклеотидное основание) или 5'-TTN-3', в отличие от богатого G PAM, на который нацеливается Cas9. PAM, на который нацеливается Cpf1, расположен на 5'-стороне направляющей РНК в отличие от PAM, на который нацеливается Cas9, который расположен на 3'-стороне направляющей РНК. После идентификации PAM Cpf1 осуществляет двухнитевой разрыв ДНК, содержащий подобные липким выступающие концы из 4-х или 5 нуклеотидов, в отличие от разрезов, которые осуществляет Cas9, оставляя тупые концы, повышая тем самым эффективность генетических вставок и специфичности во время NHEJ или HDR. TTN-сайты PAM являются более применимым для геномной инженерии человека, чем GGN-сайты PAM, поскольку геном человека в значительно степени более богат по содержанию T, чем по содержанию G. Нацеливающий на протоспейсер сегмент gRNA для Cpf1 находится на ее крайнем 3'-конце, тогда как gRNA Cas9 находятся на ее крайнем 5'-конце.

5. gRNA

Система на основе CRISPR/Cas9 или система на основе CRISPR/Cpf1 может включать в себя по меньшей мере одну gRNA, такую как оптимизированная gRNA, описываемая в данном документе, которая нацеливается на последовательность нуклеиновой кислоты. Эта gRNA обеспечивает специфическое

нацеливание системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1 на целевую область или целевой ген. Для системы на основе CRISPR/Cas9 gRNA является слитой конструкцией из двух некодирующих РНК: crRNA и tracrRNA. gRNA или sgRNA может нацеливаться на любую требуемую последовательность ДНК путем обмена последовательности, кодирующей протоспейсер размером 20 п.о., который обеспечивает специфичность нацеливания, путем комплементарного спаривания оснований с требуемой целевой ДНК. gRNA имитирует встречающийся в природе дуплекс crRNA:tracrRNA, принимающий участие в эффекторной системе типа II. Этот дуплекс, который может включать, например, 42-нуклеотидную crRNA и 75-нуклеотидную tracrRNA, действует как направляющий для Cas9 для расщепления целевой нуклеиновой кислоты. gRNA может нацеливаться на целевую область целевого гена и связываться с ней. Для системы на основе CRISPR/Cpf1 gRNA представляет собой crRNA.

Система на основе CRISPR/Cas9 или система на основе CRISPR/Cpf1 может включать в себя по меньшей мере одну gRNA, такую как оптимизированная gRNA, описываемая в данном документе, где gRNA нацеливаются на различные последовательности ДНК. Целевые последовательности ДНК могут быть переключаемыми. После целевой последовательности или протоспейсера следует последовательность PAM на 3'-конце протоспейсера. Различные системы типа II имеют разные требования к PAM. Например, в системе типа II из *Streptococcus pyogenes* используется последовательность "NGG", где "N" может представлять собой любой нуклеотид.

6. Способы получения оптимизированной направляющей РНК (gRNA)

Настоящее раскрытие направлено на способы получения оптимизированных gRNA, таких как шпильчатые gRNA (также называемые в данном документе "hpgRNA" и "hp-gRNA"). Оптимизированная gRNA включает в себя нуклеотидную последовательность полноразмерной gRNA и нуклеотиды, добавленные на 5'-конце или 3'-конце полноразмерной gRNA. В некоторых вариантах осуществления полноразмерная gRNA может быть сконструирована с использованием такой программы, как SgRNA designer, CRISPR MultiTargeter или SSFinder. Нуклеотиды, добавленные к 5'-концу для системы CRISPR/Cas9 или к 3'-концу для системы CRISPR/Cpf1 полноразмерной gRNA, могут образовывать вторичные структуры путем гибридизации или частичной гибридизации с нуклеотидами в нацеливаемой протоспейсер последовательности полноразмерной gRNA. Вторичная структура модулирует связывание или расщепление ДНК, нарушая встраивание gRNA в ДНК-дуплекс. Вторичная структура оказывает влияние на кинетику встраивания gRNA, а не на энергию связывания gRNA с комплементарной нитью ДНК. Как описано в приведенных ниже примерах, направляющие РНК систем CRISPR-Cas типа II связываются с протоспейсерами посредством процесса, обеспечиваемого Cas9, известного как "встраивание нити", где сам белок Cas9 сначала связывается с мотивом, смежным с протоспейсером (PAM), и расплавляет его путем непосредственных взаимодействий с последующим спариванием оснований 3'-конца gRNA с прилегающими к PAM нуклеотидами ("затравочной" областью), продолжая затем спаривание оснований нуклеотид за нуклеотидом от 3'- до 5'-конца gRNA с протоспейсером. Аналогичные механизмы используются с системой CRISPR/Cpf1.

Нуклеотиды, добавленные к 5'-концу или 3'-концу полноразмерной gRNA, не просто добавляются для гибридизации с нацеливаемым протоспейсером сегментом направляющей РНК (шпильки), а чтобы заблокировать доступ к протоспейсеру при термодинамическом равновесии. Как описано в примерах, равновесные термодинамические свойства вторичной структуры (такие как температура плавления вторичной структуры gRNA) вовсе не коррелируют со специфичностью направляющей РНК. Скорее, в случае расщепления и при последующей программируемой работе для связывания Cas9 (как измерено с помощью ChIP-Seq в клетках (см. doi:10.1038/nbt.2916; doi:10.1038/nbt.2889)) существует значительная и существенная корреляция между ними и оценкой кинетики встраивания нити, а также структурой, конструированием и функцией направляющих РНК, которые модулируют встраивание нити в протоспейсер, которые обязательно отличаются от шпилек, предназначенных для термодинамической конкуренции за связывание в равновесии с целевыми и нецелевыми сайтами. Например, элементы вторичной структуры, которые сконструированы для того, чтобы быть стабильными в равновесии (такие как РНК, которая образует подобную шпильке структуру, содержащую внутренние неоднозначные пары rG-rU в пределах стебля), могут быстро дестабилизироваться в ходе встраивания нити (например, когда неоднозначные пары rG-rU становятся терминальной парой оснований структуры стебля, поскольку соседние нуклеотиды встраиваются в протоспейсер, что приводит к значительному энергетическому ухудшению в отношении вторичной структуры РНК, модулируя встраивание нити и кинетику связывания при помощи полностью отдельного механизма, а не просто блокируя доступ к протоспейсеру при термодинамическом равновесии. Вторичные структуры, которые стабильны в равновесии, но быстро дестабилизируются в ходе встраивания нити, можно сконструировать с использованием способов, описанных в данном документе, таким образом, чтобы они различали целевые и нецелевые сайты с минимальными термодинамическими энергетическими различиями между сайтами (в результате одиночного внутреннего ошибочного спаривания, скажем), которые нельзя различить на практике с помощью цис-блокирующей или термодинамической конкуренции. Когда встраивание в целевой сайт дестабилизирует шпильку, содержащую неоднозначные пары G-U, сайты различают кинетически при встраивании. Например, сайты VEGFA1, описанные в примерах ниже (целевым сайтом является GGGTGGGGGGAGTTTGCTCC, а нецелевым сайтом 2

является GGATGGAGGGAGTTTGCTCC; ошибочные спаривания подчеркнуты), были способны уменьшить нецелевое расщепление на 93% и 98% по сравнению со стандартной или полноразмерной направляющей РНК или усеченной направляющей РНК соответственно, с использованием сконструированных вычислительным путем вторичных структур, которые учитывают встраивание нити.

Кроме того, нуклеотиды могут быть добавлены к 5'-концу или 3'-концу полноразмерной gRNA для разрушения "встречающейся в природе" вторичной структуры нацеливающегося на протоспейсер сегмента gRNA в "затравочной" области для усиления инициации встраивания нити направляющей РНК. Следовательно, добавление этих нуклеотидов, которые образуют вторичные структуры, что изменяет встраивание нити путем гибридизации частично гибридизирующихся нуклеотидов в нацеливающейся на протоспейсер последовательности для модулирования связывания или расщепления ДНК, представляют собой другой класс модификации направляющей РНК.

Оптимизированные gRNA предназначены для минимизации связывания на нецелевом сайте и обеспечения возможности связывания с последовательностью протоспейсера. В некоторых вариантах осуществления нецелевой сайт является известным или прогнозируемым нецелевым сайтом. В некоторых вариантах осуществления способ предусматривает: идентификацию представляющей интерес целевой области, при этом представляющая интерес целевая область содержит последовательность протоспейсера; определение полинуклеотидной последовательности полноразмерной gRNA, которая нацеливается на представляющую интерес целевую область, при этом полноразмерная gRNA содержит нацеливающиеся на протоспейсер последовательность или сегмент; определение по меньшей мере одного или нескольких нецелевых сайтов для полноразмерной gRNA; получение полинуклеотидной последовательности первой gRNA, при этом первая gRNA содержит полинуклеотидную последовательность полноразмерной gRNA и сегмент РНК, при этом сегмент РНК содержит полинуклеотидную последовательность, имеющую M нуклеотидов в длину, которая комплементарна нуклеотидному сегменту нацеливающегося на протоспейсер последовательности или сегменту, при этом РНК расположена на 5'-конце полинуклеотидной последовательности полноразмерной gRNA, при этом первая gRNA необязательно содержит линкер между 5'-концом полинуклеотидной последовательности полноразмерной gRNA и сегментом РНК, при этом линкер содержит полинуклеотидную последовательность, имеющую N нуклеотидов в длину, при этом первая gRNA способна встраиваться в последовательность протоспейсера, и связываться с последовательностью ДНК, которая комплементарна последовательности протоспейсера, и образовывать дуплекс с протоспейсером, и при этом первая gRNA способна встраиваться в нецелевой сайт, и связываться с последовательностью ДНК, которая комплементарна нецелевому сайту, и образовывать нецелевой дуплекс; вычисление оценочного показателя или вычислительное моделирование кинетики встраивания и времени нахождения, в течение которого gRNA остается встроенной в дуплексы протоспейсера и нецелевого сайта, где динамику встраивания оценивают нуклеотид за нуклеотидом путем определения энергетических различий между дополнительным встраиванием отличающейся gRNA и повторным отжигом первой gRNA на последовательности ДНК, которая комплементарна последовательности протоспейсера; сравнение значений предполагаемого времени нахождения в сайтах протоспейсера и/или нецелевых сайтах первой gRNA со значениями предполагаемого времени нахождения полноразмерной gRNA или усеченной gRNA (tru-gRNA) в сайтах протоспейсера и/или нецелевых сайтах; рандомизацию от 0 до N нуклеотидов в линкере и от 0 до M нуклеотидов в первой gRNA, и получение второй gRNA, и повторение стадии (e) со второй gRNA; идентификацию оптимизированной gRNA на основе последовательности gRNA, которая удовлетворяет критериям конструирования; и тестирование оптимизированной gRNA in vivo для определения специфичности связывания.

В некоторых вариантах осуществления способ предусматривает: идентификацию представляющей интерес целевой области, при этом представляющая интерес целевая область содержит последовательность протоспейсера; определение полинуклеотидной последовательности полноразмерной gRNA, которая нацеливается на представляющую интерес целевую область, при этом полноразмерная gRNA содержит нацеливающиеся на протоспейсер последовательность или сегмент; определение по меньшей мере одного или нескольких нецелевых сайтов для полноразмерной gRNA; получение полинуклеотидной последовательности первой gRNA, при этом первая gRNA содержит полинуклеотидную последовательность полноразмерной gRNA и сегмент РНК, при этом сегмент РНК содержит полинуклеотидную последовательность, имеющую M нуклеотидов в длину, которая комплементарна нуклеотидному сегменту нацеливающегося на протоспейсер последовательности или сегменту, при этом РНК расположена на 3'-конце полинуклеотидной последовательности полноразмерной gRNA, при этом первая gRNA необязательно содержит линкер между 3'-концом полинуклеотидной последовательности полноразмерной gRNA и сегментом РНК, при этом линкер содержит полинуклеотидную последовательность, имеющую N нуклеотидов в длину, при этом первая gRNA способна встраиваться в последовательность протоспейсера, и связываться с последовательностью ДНК, которая комплементарна последовательности протоспейсера, и образовывать дуплекс с протоспейсером, и при этом первая gRNA способна встраиваться в нецелевой сайт, и связываться с последовательностью ДНК, которая комплементарна нецелевому сайту, и образовывать нецелевой дуплекс; вычисление оценочного показателя или вычислительное моделирование кинетики встраивания и значений времени нахождения, в течение которого gRNA остается встроенной в

дуплексы протоспейсера и нецелевого сайта, где динамику встраивания оценивают нуклеотид за нуклеотидом путем определения энергетических различий между дополнительным встраиванием отличающейся gRNA и повторным отжигом первой gRNA на последовательности ДНК, которая комплементарна последовательности протоспейсера; сравнение значений предполагаемого времени нахождения в сайтах протоспейсера и/или нецелевых сайтах первой gRNA со значениями предполагаемого времени нахождения полноразмерной gRNA или усеченной gRNA (tru-gRNA) в сайтах протоспейсера и/или нецелевых сайтах; рандомизацию от 0 до N нуклеотидов в линкере и от 0 до M нуклеотидов в первой gRNA, и получение второй gRNA, и повторение стадии (e) со второй gRNA; идентификацию оптимизированной gRNA на основе последовательности gRNA, которая удовлетворяет критериям конструирования; и тестирование оптимизированной gRNA *in vivo* для определения специфичности связывания.

В некоторых вариантах осуществления энергетические балансы дополнительного встраивания отличающейся gRNA определяют путем определения энергетического баланса по меньшей мере одного из (I) нарушения спаривания оснований ДНК-ДНК, (II) образования пар оснований РНК-ДНК, (III) энергетического различия, возникающего в результате разрушения или образования отличающейся вторичной структуры внутри не встроенной направляющей РНК, и (IV) образования или разрушения взаимодействий между вытесненной нитью ДНК, которая комплементарна протоспейсеру, и нуклеотидами любыми неспаренными нуклеотидами направляющей РНК, которые не вовлечены во вторичные структуры. В некоторых вариантах осуществления энергетический баланс повторного отжига первой gRNA с последовательностью ДНК, которая комплементарна последовательности протоспейсера, определяют путем определения энергетического баланса по меньшей мере одного из (I) образования пар оснований ДНК-ДНК, (II) разрушения пар оснований РНК-ДНК, (III) энергетического различия, возникающего в результате разрушения или образования отличающейся вторичной структуры внутри новой не встроенной направляющей РНК, и (IV) образования или разрушения взаимодействий между вытесненной нитью ДНК, которая комплементарна протоспейсеру, и любыми неспаренными нуклеотидами направляющей РНК, которые не вовлечены во вторичные структуры. В некоторых вариантах осуществления способ дополнительно предусматривает определение энергетических факторов по меньшей мере одного из (V) спаривания оснований среди ошибочных спариваний, (VI) взаимодействий с белком Cas9 и/или (VII) дополнительных эвристических показателей, где дополнительные эвристические показатели относятся к продолжительности связывания, степени встраивания, стабильности встраиваемой направляющей РНК или другим расчетным/моделируемым свойствам встраивания gRNA для расщепляющей активности Cas9.

В системе на основе CRISPR/Cas9 или системе на основе CRISPR/Cpf1 можно использовать gRNA, такую как оптимизированная gRNA, описанная в данном документе, с различными последовательностями и длиной. В некоторых вариантах осуществления полноразмерная gRNA может содержать нацеливающийся на протоспейсер сегмент, который соответствует полинуклеотидной последовательности целевой ДНК (т.е. протоспейсеру). В некоторых вариантах осуществления нацеливающийся на протоспейсер сегмент может иметь 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 30 нуклеотидов или по меньшей мере 35 нуклеотидов. gRNA может нацеливаться по меньшей мере на одно из области промотора, области энхансера, области репрессора, области инсулятора, области сайленсера, области, вовлеченной в образование петли ДНК с промоторной областью, области сплайсинга гена или транскрибируемой области целевого гена. В некоторых вариантах осуществления полноразмерная gRNA содержит нацеливающийся на протоспейсер сегмент, имеющий от приблизительно 15 до 20 нуклеотидов.

В некоторых вариантах осуществления сегмент РНК содержит от 2 до 20 нуклеотидов, от 3 до 10 нуклеотидов или от 5 до 8 нуклеотидов. В некоторых вариантах осуществления сегмент РНК содержит от 2 до 20 нуклеотидов, от 3 до 10 нуклеотидов или от 5 до 8 нуклеотидов, которые комплементарны последовательности, нацеливающейся на протоспейсер. В некоторых вариантах осуществления M составляет от 1 до 20, от 1 до 19, от 1 до 18, от 1 до 17, от 1 до 16, от 1 до 15, от 1 до 14, от 1 до 13, от 1 до 12, от 1 до 11, от 1 до 10, от 1 до 9, от 1 до 8, от 1 до 7, от 1 до 6, от 1 до 5, от 2 до 20, от 2 до 19, от 2 до 18, от 2 до 17, от 2 до 16, от 2 до 15, от 2 до 14, от 2 до 13, от 2 до 12, от 2 до 11, от 2 до 10, от 2 до 9, от 2 до 8, от 2 до 7, от 2 до 6, от 2 до 5, от 3 до 20, от 3 до 19, от 3 до 18, от 3 до 17, от 3 до 16, от 3 до 15, от 3 до 14, от 3 до 13, от 3 до 12, от 3 до 11, от 3 до 10, от 3 до 9, от 3 до 8, от 3 до 7, от 3 до 6, от 3 до 5, от 4 до 20, от 4 до 19, от 4 до 18, от 4 до 17, от 4 до 16, от 4 до 15, от 4 до 14, от 4 до 13, от 4 до 12, от 4 до 11, от 4 до 10, от 4 до 9, от 4 до 8, от 4 до 7, от 4 до 6, от 4 до 5, от 5 до 20, от 5 до 19, от 5 до 18, от 5 до 17, от 5 до 16, от 5 до 15, от 5 до 14, от 5 до 13, от 5 до 12, от 5 до 11, от 5 до 10, от 5 до 9, от 5 до 8, от 5 до 7, от 5 до 6, от 6 до 20, от 6 до 19, от 6 до 18, от 6 до 17, от 6 до 16, от 6 до 15, от 6 до 14, от 6 до 13, от 6 до 12, от 6 до 11, от 6 до 10, от 6 до 9, от 6 до 8, от 6 до 7, от 7 до 20, от 7 до 19, от 7 до 18, от 7 до 17, от 7 до 16, от 7 до 15, от 7 до 14, от 7 до 13, от 7 до 12, от 7 до 11, от 7 до 10, от 7 до 9, от 7 до 8, от 8 до 20, от 8 до 19, от 8 до 18, от 8 до 17, от 8 до 16, от 8 до 15, от 8 до 14, от 8 до 13, от 8 до 12, от 8 до 11, от 8 до 10, от 8 до 9,

от 9 до 20, от 9 до 19, от 9 до 18, от 9 до 17, от 9 до 16, от 9 до 15, от 9 до 14, от 9 до 13, от 9 до 12, от 9 до 11 или от 9 до 10. Например, M может составлять 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 или 20. В некоторых вариантах осуществления сегмент РНК может иметь от 1 до 20, от 1 до 19, от 1 до 18, от 1 до 17, от 1 до 16, от 1 до 15, от 1 до 14, от 1 до 13, от 1 до 12, от 1 до 11, от 1 до 10, от 1 до 9, от 1 до 8, от 1 до 7, от 1 до 6, от 1 до 5, от 2 до 20, от 2 до 19, от 2 до 18, от 2 до 17, от 2 до 16, от 2 до 15, от 2 до 14, от 2 до 13, от 2 до 12, от 2 до 11, от 2 до 10, от 2 до 9, от 2 до 8, от 2 до 7, от 2 до 6, от 2 до 5, от 3 до 20, от 3 до 19, от 3 до 18, от 3 до 17, от 3 до 16, от 3 до 15, от 3 до 14, от 3 до 13, от 3 до 12, от 3 до 11, от 3 до 10, от 3 до 9, от 3 до 8, от 3 до 7, от 3 до 6, от 3 до 5, от 4 до 20, от 4 до 19, от 4 до 18, от 4 до 17, от 4 до 16, от 4 до 15, от 4 до 14, от 4 до 13, от 4 до 12, от 4 до 11, от 4 до 10, от 4 до 9, от 4 до 8, от 4 до 7, от 4 до 6, от 4 до 5, от 5 до 20, от 5 до 19, от 5 до 18, от 5 до 17, от 5 до 16, от 5 до 15, от 5 до 14, от 5 до 13, от 5 до 12, от 5 до 11, от 5 до 10, от 5 до 9, от 5 до 8, от 5 до 7, от 5 до 6, от 6 до 20, от 6 до 19, от 6 до 18, от 6 до 17, от 6 до 16, от 6 до 15, от 6 до 14, от 6 до 13, от 6 до 12, от 6 до 11, от 6 до 10, от 6 до 9, от 6 до 8, от 6 до 7, от 7 до 20, от 7 до 19, от 7 до 18, от 7 до 17, от 7 до 16, от 7 до 15, от 7 до 14, от 7 до 13, от 7 до 12, от 7 до 11, от 7 до 10, от 7 до 9, от 7 до 8, от 8 до 20, от 8 до 19, от 8 до 18, от 8 до 17, от 8 до 16, от 8 до 15, от 8 до 14, от 8 до 13, от 8 до 12, от 8 до 11, от 8 до 10, от 8 до 9, от 9 до 20, от 9 до 19, от 9 до 18, от 9 до 17, от 9 до 16, от 9 до 15, от 9 до 14, от 9 до 13, от 9 до 12, от 9 до 11 или от 9 до 10 нуклеотидов, некоторые или все из которых комплементарны нацеливаемой на протоспейсер последовательности. В некоторых вариантах осуществления сегмент РНК может иметь 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 или 20 нуклеотидов.

В некоторых вариантах осуществления N составляет от 1 до 20, от 1 до 19, от 1 до 18, от 1 до 17, от 1 до 16, от 1 до 15, от 1 до 14, от 1 до 13, от 1 до 12, от 1 до 11, от 1 до 10, от 1 до 9, от 1 до 8, от 1 до 7, от 1 до 6, от 1 до 5, от 2 до 20, от 2 до 19, от 2 до 18, от 2 до 17, от 2 до 16, от 2 до 15, от 2 до 14, от 2 до 13, от 2 до 12, от 2 до 11, от 2 до 10, от 2 до 9, от 2 до 8, от 2 до 7, от 2 до 6, от 2 до 5, от 3 до 20, от 3 до 19, от 3 до 18, от 3 до 17, от 3 до 16, от 3 до 15, от 3 до 14, от 3 до 13, от 3 до 12, от 3 до 11, от 3 до 10, от 3 до 9, от 3 до 8, от 3 до 7, от 3 до 6, от 3 до 5, от 4 до 20, от 4 до 19, от 4 до 18, от 4 до 17, от 4 до 16, от 4 до 15, от 4 до 14, от 4 до 13, от 4 до 12, от 4 до 11, от 4 до 10, от 4 до 9, от 4 до 8, от 4 до 7, от 4 до 6, от 4 до 5, от 5 до 20, от 5 до 19, от 5 до 18, от 5 до 17, от 5 до 16, от 5 до 15, от 5 до 14, от 5 до 13, от 5 до 12, от 5 до 11, от 5 до 10, от 5 до 9, от 5 до 8, от 5 до 7, от 5 до 6, от 6 до 20, от 6 до 19, от 6 до 18, от 6 до 17, от 6 до 16, от 6 до 15, от 6 до 14, от 6 до 13, от 6 до 12, от 6 до 11, от 6 до 10, от 6 до 9, от 6 до 8, от 6 до 7, от 7 до 20, от 7 до 19, от 7 до 18, от 7 до 17, от 7 до 16, от 7 до 15, от 7 до 14, от 7 до 13, от 7 до 12, от 7 до 11, от 7 до 10, от 7 до 9, от 7 до 8, от 8 до 20, от 8 до 19, от 8 до 18, от 8 до 17, от 8 до 16, от 8 до 15, от 8 до 14, от 8 до 13, от 8 до 12, от 8 до 11, от 8 до 10, от 8 до 9, от 9 до 20, от 9 до 19, от 9 до 18, от 9 до 17, от 9 до 16, от 9 до 15, от 9 до 14, от 9 до 13, от 9 до 12, от 9 до 11 или от 9 до 10. Например, N может составлять 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 или 20. В некоторых вариантах осуществления линкер содержит от 1 до 20 нуклеотидов, от 3 до 10 нуклеотидов или от 5 до 8 нуклеотидов. Например, линкер может иметь от 1 до 20, от 1 до 19, от 1 до 18, от 1 до 17, от 1 до 16, от 1 до 15, от 1 до 14, от 1 до 13, от 1 до 12, от 1 до 11, от 1 до 10, от 1 до 9, от 1 до 8, от 1 до 7, от 1 до 6, от 1 до 5, от 2 до 20, от 2 до 19, от 2 до 18, от 2 до 17, от 2 до 16, от 2 до 15, от 2 до 14, от 2 до 13, от 2 до 12, от 2 до 11, от 2 до 10, от 2 до 9, от 2 до 8, от 2 до 7, от 2 до 6, от 2 до 5, от 3 до 20, от 3 до 19, от 3 до 18, от 3 до 17, от 3 до 16, от 3 до 15, от 3 до 14, от 3 до 13, от 3 до 12, от 3 до 11, от 3 до 10, от 3 до 9, от 3 до 8, от 3 до 7, от 3 до 6, от 3 до 5, от 4 до 20, от 4 до 19, от 4 до 18, от 4 до 17, от 4 до 16, от 4 до 15, от 4 до 14, от 4 до 13, от 4 до 12, от 4 до 11, от 4 до 10, от 4 до 9, от 4 до 8, от 4 до 7, от 4 до 6, от 4 до 5, от 5 до 20, от 5 до 19, от 5 до 18, от 5 до 17, от 5 до 16, от 5 до 15, от 5 до 14, от 5 до 13, от 5 до 12, от 5 до 11, от 5 до 10, от 5 до 9, от 5 до 8, от 5 до 7, от 5 до 6, от 6 до 20, от 6 до 19, от 6 до 18, от 6 до 17, от 6 до 16, от 6 до 15, от 6 до 14, от 6 до 13, от 6 до 12, от 6 до 11, от 6 до 10, от 6 до 9, от 6 до 8, от 6 до 7, от 7 до 20, от 7 до 19, от 7 до 18, от 7 до 17, от 7 до 16, от 7 до 15, от 7 до 14, от 7 до 13, от 7 до 12, от 7 до 11, от 7 до 10, от 7 до 9, от 7 до 8, от 8 до 20, от 8 до 19, от 8 до 18, от 8 до 17, от 8 до 16, от 8 до 15, от 8 до 14, от 8 до 13, от 8 до 12, от 8 до 11, от 8 до 10, от 8 до 9, от 9 до 20, от 9 до 19, от 9 до 18, от 9 до 17, от 9 до 16, от 9 до 15, от 9 до 14, от 9 до 13, от 9 до 12, от 9 до 11 или от 9 до 10 нуклеотидов, некоторые или все из которых комплементарны нацеливаемой на протоспейсер последовательности. В некоторых вариантах осуществления линкер может иметь 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 или 20 нуклеотидов. В некоторых вариантах осуществления линкер может включать в себя стабилизирующий линкер, такой как тетрапетля. Примеры тетрапетли включают без ограничения ANYA, CUYG, GNRA, UMAC и UNCG.

В некоторых вариантах осуществления сегмент РНК и/или целенаправленно воздействующая на протоспейсер последовательность обеспечивает вторичную структуру. В некоторых вариантах осуществления вторичная структура образуется за счет частичной гибридизации целенаправленно воздействующей на протоспейсер последовательности с сегментом РНК. В некоторых вариантах осуществления вторичная структура модулирует связывание или расщепление ДНК посредством Cas9 путем нарушения встраивания оптимизированной gRNA в дуплекс протоспейсера или нецелевой дуплекс. В некоторых вариантах осуществления вторичная структура сохраняет 5'-конец gRNA стабильным в белке и защищает оптимизированную gRNA в Cas9 с целью предотвращения разрушения.

В некоторых вариантах осуществления вторичная структура образуется путем гибридизации всего

или части сегмента РНК с нуклеотидами на 5'-конце нацеливающихся на протоспейсер последовательности или сегмента, нуклеотидами в середине нацеливающихся на протоспейсер последовательности или сегмента, и/или нуклеотидами на 3'-конце нацеливающихся на протоспейсер последовательности или сегмента. В некоторых вариантах осуществления сегменты, смежные по отношению к сегменту РНК, гибридизируются с нацеливающимися на протоспейсер последовательностью или сегментом. В некоторых вариантах осуществления сегменты, которые не являются смежными по отношению к сегменту РНК, гибридизируются с нацеливающимися на протоспейсер последовательностью или сегментом. В некоторых вариантах осуществления вторичная структура представляет собой шпильку.

В некоторых вариантах осуществления вторичная структура является стабильной при комнатной температуре или при 37°C. В некоторых вариантах осуществления общая равновесная свободная энергия вторичной структуры составляет менее приблизительно 2 ккал/моль при температуре от приблизительно 4°C до приблизительно 50°C, как, например, при комнатной температуре или при 37°C. Например, общая равновесная свободная энергия вторичной структуры может составлять менее приблизительно 10 ккал/моль, менее приблизительно 5 ккал/моль, менее приблизительно 4 ккал/моль, менее приблизительно 3 ккал/моль, менее приблизительно 2 ккал/моль, менее приблизительно 1 ккал/моль или менее приблизительно 0,5 ккал/моль при температуре от приблизительно 4°C до приблизительно 50°C, от приблизительно 4°C до приблизительно 40°C, от приблизительно 4°C до приблизительно 37°C, от приблизительно 4°C до приблизительно 30°C, от приблизительно 4°C до приблизительно 25°C, от приблизительно 4°C до приблизительно 20°C, от приблизительно 4°C до приблизительно 10°C, от приблизительно 5°C до приблизительно 50°C, от приблизительно 5°C до приблизительно 40°C, от приблизительно 5°C до приблизительно 37°C, от приблизительно 5°C до приблизительно 30°C, от приблизительно 5°C до приблизительно 25°C, от приблизительно 5°C до приблизительно 20°C, от приблизительно 5°C до приблизительно 10°C, от приблизительно 10°C до приблизительно 50°C, от приблизительно 10°C до приблизительно 40°C, от приблизительно 10°C до приблизительно 37°C, от приблизительно 10°C до приблизительно 30°C, от приблизительно 10°C до приблизительно 25°C, от приблизительно 10°C до приблизительно 20°C, от приблизительно 20°C до приблизительно 50°C, от приблизительно 20°C до приблизительно 40°C, от приблизительно 20°C до приблизительно 37°C, от приблизительно 20°C до приблизительно 30°C, от приблизительно 25°C до приблизительно 50°C, от приблизительно 25°C до приблизительно 40°C, от приблизительно 25°C до приблизительно 37°C или от приблизительно 25°C до приблизительно 30°C. В некоторых вариантах осуществления сегмент РНК гибридизируется или образует неканонические пары оснований по меньшей мере с двумя нуклеотидами нацеливающихся на протоспейсер последовательности или сегмента. В некоторых вариантах осуществления неканоническая пара оснований представляет собой *gU-gG*.

В некоторых вариантах осуществления от 1 до 20 нуклеотидов в линкере отбирают случайным образом. Например, от 1 до 20, от 1 до 15, от 1 до 10, от 1 до 9, от 1 до 8, от 1 до 7, от 1 до 6, от 1 до 5, от 1 до 4, от 1 до 3, от 1 до 2, от 2 до 20, от 2 до 15, от 2 до 10, от 2 до 9, от 2 до 8, от 2 до 7, от 2 до 6, от 2 до 5, от 2 до 4, от 3 до 20, от 3 до 15, от 3 до 10, от 3 до 9, от 3 до 8, от 3 до 7, от 3 до 6, от 3 до 5, от 3 до 4, от 4 до 20, от 4 до 15, от 4 до 10, от 4 до 9, от 4 до 8, от 4 до 7, от 4 до 6, от 4 до 5, от 5 до 20, от 5 до 15, от 5 до 10, от 5 до 9, от 5 до 8, от 5 до 7, от 5 до 6, от 6 до 20, от 6 до 15, от 6 до 10, от 6 до 9, от 6 до 8, от 6 до 7, от 7 до 20, от 7 до 15, от 7 до 10, от 7 до 9, от 7 до 8, от 8 до 20, от 8 до 15, от 8 до 10, от 8 до 9, от 9 до 20, от 9 до 15 или от 9 до 10, от 10 до 20, от 10 до 15 или от 15 до 20 нуклеотидов в линкере могут быть отобраны случайным образом.

В некоторых вариантах осуществления от 1 до 20 нуклеотидов в сегменте РНК отбирают случайным образом. Например, от 1 до 20, от 1 до 15, от 1 до 10, от 1 до 9, от 1 до 8, от 1 до 7, от 1 до 6, от 1 до 5, от 1 до 4, от 1 до 3, от 1 до 2, от 2 до 20, от 2 до 15, от 2 до 10, от 2 до 9, от 2 до 8, от 2 до 7, от 2 до 6, от 2 до 5, от 2 до 4, от 3 до 20, от 3 до 15, от 3 до 10, от 3 до 9, от 3 до 8, от 3 до 7, от 3 до 6, от 3 до 5, от 3 до 4, от 4 до 20, от 4 до 15, от 4 до 10, от 4 до 9, от 4 до 8, от 4 до 7, от 4 до 6, от 4 до 5, от 5 до 20, от 5 до 15, от 5 до 10, от 5 до 9, от 5 до 8, от 5 до 7, от 5 до 6, от 6 до 20, от 6 до 15, от 6 до 10, от 6 до 9, от 6 до 8, от 6 до 7, от 7 до 20, от 7 до 15, от 7 до 10, от 7 до 9, от 7 до 8, от 8 до 20, от 8 до 15, от 8 до 10, от 8 до 9, от 9 до 20, от 9 до 15 или от 9 до 10, от 10 до 20, от 10 до 15 или от 15 до 20 нуклеотидов в сегменте РНК могут быть отобраны случайным образом.

В некоторых вариантах осуществления стадию (g) повторяют X раз, получая тем самым gRNA в количестве X и повторяя стадию (e) с каждым количеством X gRNA, где X составляет от 0 до 20. В некоторых вариантах осуществления X может составлять от 1 до 20, от 1 до 19, от 1 до 18, от 1 до 17, от 1 до 16, от 1 до 15, от 1 до 14, от 1 до 13, от 1 до 12, от 1 до 11, от 1 до 10, от 1 до 9, от 1 до 8, от 1 до 7, от 1 до 6, от 1 до 5, от 2 до 20, от 2 до 19, от 2 до 18, от 2 до 17, от 2 до 16, от 2 до 15, от 2 до 14, от 2 до 13, от 2 до 12, от 2 до 11, от 2 до 10, от 2 до 9, от 2 до 8, от 2 до 7, от 2 до 6, от 2 до 5, от 3 до 20, от 3 до 19, от 3 до 18, от 3 до 17, от 3 до 16, от 3 до 15, от 3 до 14, от 3 до 13, от 3 до 12, от 3 до 11, от 3 до 10, от 3 до 9, от 3 до 8, от 3 до 7, от 3 до 6, от 3 до 5, от 4 до 20, от 4 до 19, от 4 до 18, от 4 до 17, от 4 до 16, от 4 до 15, от 4 до 14, от 4 до 13, от 4 до 12, от 4 до 11, от 4 до 10, от 4 до 9, от 4 до 8, от 4 до 7, от 4 до 6, от 4 до 5, от 5 до 20, от 5 до 19, от 5 до 18, от 5 до 17, от 5 до 16, от 5 до 15, от 5 до 14, от 5 до 13, от 5 до 12, от 5 до 11, от 5 до 10, от 5 до 9, от 5 до 8, от 5 до 7, от 5 до 6, от 6 до 20, от 6 до 19, от 6 до 18, от 6 до 17, от 6 до 16,

от 6 до 15, от 6 до 14, от 6 до 13, от 6 до 12, от 6 до 11, от 6 до 10, от 6 до 9, от 6 до 8, от 6 до 7, от 7 до 20, от 7 до 19, от 7 до 18, от 7 до 17, от 7 до 16, от 7 до 15, от 7 до 14, от 7 до 13, от 7 до 12, от 7 до 11, от 7 до 10, от 7 до 9, от 7 до 8, от 8 до 20, от 8 до 19, от 8 до 18, от 8 до 17, от 8 до 16, от 8 до 15, от 8 до 14, от 8 до 13, от 8 до 12, от 8 до 11, от 8 до 10, от 8 до 9, от 9 до 20, от 9 до 19, от 9 до 18, от 9 до 17, от 9 до 16, от 9 до 15, от 9 до 14, от 9 до 13, от 9 до 12, от 9 до 11 или от 9 до 10. Например, X может составлять 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 или 20.

В некоторых вариантах осуществления кинетику встраивания и время нахождения вычисляют с использованием кинетической методики Монте-Карло или алгоритма Гиллеспи. В некоторых вариантах осуществления кинетику встраивания и время нахождения можно определить с использованием "детерминистских" способов, таких как дифференциальные уравнения, моделирующие встраивание нити, которые известны специалисту в данной области. Кинетическая методика Монте-Карло (КМС) представляет собой компьютерное моделирование по методике Монте-Карло, предназначенное для моделирования временной эволюции некоторых процессов, происходящих в природе. Данные процессы, как правило, представляют собой процессы, которые происходят при известных скоростях перехода между состояниями. Эти известные скорости перехода представляют собой исходные вводимые данные для алгоритма КМС. Алгоритм Гиллеспи (также известный как алгоритм Дуба-Гиллеспи) генерирует статистически правильную траекторию (возможное решение) стохастического уравнения. Алгоритм Гиллеспи можно использовать для моделирования еще более сложных систем. Этот алгоритм особенно полезен для моделирования реакций внутри клеток, где количество реагентов обычно исчисляется десятками молекул (или меньше). Математически он представляет собой вариант динамической методики Монте-Карло и подобен кинетическим методикам Монте-Карло. Алгоритм Гиллеспи позволяет проводить дискретное и стохастическое моделирование системы с несколькими реагентами, поскольку каждая реакция моделируется в явном виде. Траектория, соответствующая одиночному моделированию Гиллеспи, представляет собой точную выборку из функции распределения масс, являющейся решением основного уравнения.

В некоторых вариантах осуществления критериями конструирования могут быть специфичность, модулирование продолжительности связывания и/или расчетная специфичность расщепления. Например, оптимизированная gRNA может быть сконструирована так, чтобы ее продолжительность связывания была больше или равная продолжительности связывания полной gRNA в целевом сайте и/или продолжительность связывания была меньше или равная продолжительности связывания полноразмерной gRNA в нецелевом сайте. В некоторых вариантах осуществления оптимизированную gRNA выбирают так, чтобы ее продолжительность связывания была меньше или равная продолжительности связывания полноразмерной gRNA по меньшей мере с тремя нецелевыми сайтами, где нецелевые сайты прогнозируют как ближайшие нецелевые сайты, или прогнозируется, что они имеют наивысшую идентичность с целевыми сайтами. В некоторых вариантах осуществления критерии конструирования предусматривают время нахождения или скорость расщепления в нецелевом сайте, которые меньше или равны продолжительности нахождения или скорости расщепления полноразмерной gRNA или усеченной gRNA в нецелевом сайте, и/или прогнозируемую степень целевой активности, которая превышает 10% от прогнозируемой степени целевой активности полноразмерной gRNA или усеченной gRNA.

В некоторых вариантах осуществления оптимизированную gRNA тестируют на стадии i) с использованием чувствительной к ошибочному спариванию нуклеазы для определения активности CRISPR, например с использованием анализа *surveyor*, или анализа с эндонуклеазой I из T7 (T7E1), или методик секвенирования нового поколения, таких как Illumina MiSeq или GUIDE-Seq. В некоторых вариантах осуществления оптимизированную gRNA тестируют на стадии i) с использованием анализа с репортером, где активность Cas9-слитого белка изменяет экспрессию репортерного белка, такого как GFP. GUIDE-Seq представляет собой анализ, который был разработан для анализа нецелевых расщеплений.

В некоторых вариантах осуществления целевую область можно определить на основе близости последовательности к последовательности PAM с использованием программы, такой как CRISPR design (Ran, et al. *Nature Protocols* (2013) 8:2281-2308) и инструмента CCTop (Stemmer, *PLoS One* (2015) 10:e0124633). В некоторых вариантах осуществления целевые сайты могут включать в себя промоторы, сайты гиперчувствительности к ДНКазе I, сайты хроматина, доступные для транспозазы, сайты метилирования ДНК, сайты связывания факторов транскрипции, эпигенетические метки, локусы количественных признаков экспрессии и/или области, связанные с признаками человека или фенотипами в исследованиях генетических ассоциаций. Целевые сайты могут быть определены путем секвенирования сайтов расщепления ДНКазой (DNase-seq), анализа хроматина, доступного для транспозазы, с высокопроизводительным секвенированием (ATAC-seq), ChIP-секвенирования, секвенирования самотранскрибирующихся активных регуляторных областей (STARR-Seq), секвенирования одиночных молекул в реальном времени (SMRT), секвенирования выделенных с помощью формальдегида регуляторных элементов (FAIRE-seq), секвенирования сайтов расщепления микрококковой нуклеазы (MNase-seq), секвенирования после обработки бисульфитом при сниженном представителе (RRBS-seq), полногеномного секвенирования с предварительной обработкой бисульфитом, секвенирования после иммунопреципитации метилированной ДНК (MEDIP-seq) или исследований генетических ассоциаций. В некоторых вариантах осуществления нецелевые сайты могут быть определены с использованием CasOT (PKU Zebrafish Functional

Genomics group, Пекинский университет), CHOPCHOP (Гарвардский университет), CRISPR Design, (Массачусетский технологический институт), CRISPR Design tool (Институт Броудов при МТИ и Гарварде), CRISPR/Cas9 gRNA finder (Университет Колорадо), CRISPRfinder (Университет Париж-юг), E-CRISP (DKFZ, Немецкий центр исследования рака), CRISPR gRNA Design tool (DNA 2.0), PROGNOS (Университет Эмори/Технологический институт Джорджии), ZiFiT (Массачусетская больница общего профиля). Примеры инструментов, которые можно использовать для определения целевых областей и нецелевых сайтов, описаны в международной патентной заявке № WO 2016109255, которая включена в данный документ посредством ссылки во всей ее полноте.

7. Целевой ген

Раскрываемая в данном документе система на основе CRISPR/Cas9 или система на основе CRISPR/Cpf1 может быть сконструирована для нацеливания на любой целевой ген и его расщепления. Например, gRNA, такая как оптимизированная gRNA, описанная в данном документе, может нацеливаться на целевую область в целевом гене и связываться с ней. Целевой ген может представлять собой эндогенный ген, трансген или вирусный ген в клеточной линии. В некоторых вариантах осуществления целевой ген может представлять собой известный ген. В некоторых вариантах осуществления целевой ген является неизвестным геном. gRNA может нацеливаться на любую последовательность нуклеиновой кислоты. Целевая последовательность нуклеиновой кислоты может представлять собой ДНК. ДНК может представлять собой любой ген. Например, gRNA может нацеливаться на ген, такой как DMD, EMX1, или VEGFA.

В некоторых аспектах целевой ген представляет собой ген, связанный с заболеванием. В некоторых вариантах осуществления клетка-мишень представляет собой клетку млекопитающего. В некоторых вариантах осуществления геном включает в себя геном человека. В некоторых вариантах осуществления целевой ген может представлять собой прокариотический ген или эукариотический ген, такой как ген млекопитающего. Например, система на основе CRISPR/Cas9 или система на основе CRISPR/Cpf1 может нацеливаться на ген млекопитающего, такой как DMD (ген дистрофина), EMX1, VEGFA, IL1RN, MYOD1, OCT4, EBE, HBG, HBD, HBB, MYOCD (миокардин), PAX7 (белок парного бокса Pax-7), гены FGF1 (фактор роста фибробластов-1), такие как FGF1A, FGF1B и FGF1C. Другие целевые гены включают без ограничения Atf3, Axud1, Btg2, c-Fos, c-Jun, Cxcl1, Cxcl2, Edn1, Ereg, Fos, Gadd45b, Ier2, Ier3, Ifrd1, Il1b, Il6, Irf1, Junb, Lif, Nfkb1a, Nfkb2, Ptg2, Sle25a25, Sqstm1, Tieg, Tnf, Tnfaip3, Zfp36, Birc2, Ccl2, Ccl20, Ccl7, Cebpd, Ch25h, CSF1, Cx3cl1, Cxcl10, Cxcl5, Gch, Icam1, Ifi47, Ifngr2, Mmp10, Nfkbie, Npal1, p21, Relb, Ripk2, Rnd1, Slpr3, Stx11, Tgtp, Tlr2, Tmem140, Tnfaip2, Tnfrsf6, Vcam1, 1110004C05Rik (номер доступа в GenBank BC010291), Abca1, AI561871 (номер доступа в GenBank B1143915), AI882074 (номер доступа в GenBank BB730912), Arts1, AW049765 (номер доступа в GenBank BC026642.1), C3, Casp4, Ccl5, Ccl9, Cdsn, Enpp2, Gbp2, H2-D1, H2-K, H2-L, Ifit1, Ii, Il13ra1, Il1rl1, Lcn2, Lhfp12, LOC677168 (номер доступа в GenBank AK019325), Mmp13, Mmp3, Mt2, Naf1, Ppicap, Prnd, Psmb10, Saa3, Serpina3g, Serpinf1, Sod3, Stat1, Tapbp, U90926 (номер доступа в GenBank NM_020562), Ubd, A2AR (аденозиновый рецептор A2A), B7-H3 (также называемый CD276), B7-H4 (также называемый VTCN1), BTLA (аттенюатор В- и Т-лимфоцитов; также называемый CD272), CTLA-4 (ассоциированный с цитотоксическими Т-лимфоцитами белок 4; также называемый CD152), IDO (индоламин-2,3-диоксигеназа) KIR (иммуноглобулиноподобный рецептор клеток-киллеров), LAG3 (ген активации лимфоцитов-3), PD-1 (рецептор программируемой смерти 1 (PD-1)), TIM-3 (Т-клеточный иммуноглобулиновый домен и муциновый домен 3) и VISTA (иммуноглобулиновый V-домен, супрессор активации Т-клеток). В некоторых вариантах осуществления целевой ген представляет собой ген DMD (дистрофии), EMX1 или VEGFA.

8. Композиции для редактирования генома

Настоящее изобретение направлено на композиции для редактирования генома, геномной перестройки или изменения экспрессии целевого гена. Композиции включают оптимизированную gRNA, получаемую с помощью раскрытого способа, с системой на основе CRISPR/Cas9 или системой на основе CRISPR/Cpf1. В некоторых вариантах осуществления gRNA может различать целевые и нецелевые сайты с минимальными термодинамическими энергетическими различиями между сайтами и обеспечивать повышенную специфичность. В некоторых вариантах осуществления оптимизированная gRNA модулирует встраивание нити в протоспейсер.

Увеличение специфичности достигается за счет добавления удлинения к 5'-концу или 3'-концу полноразмерной или стандартной gRNA, так что образуется структура "шпильки", которая является самокомплементарной сегменту полноразмерной или стандартной gRNA, которая нацеливается на протоспейсер, например, нацеливающейся на протоспейсер последовательности. См. фиг. 1В и 2В. Шпильки служат кинетическим барьером для встраивания нити в протоспейсер, но шпильки вытесняются в ходе встраивания нити в полные целевые сайты, в результате чего может происходить полное встраивание.

Как показано на фиг. 2D, предпочтительно происходит связывание dCas9 с полными протоспейсерами, что убедительно указывает на то, что шпильки фактически вытесняются во в ходе встраивания. Раскрытые оптимизированные gRNA, которые представляют собой шпильки, сконструировали для увеличения специфичности связывания с целевыми сайтами путем ингибирования встраивания, если были ошибочные спаривания между мишенью и РАМ-дистальной нацеливающей областью направляющей

РНК. В этих случаях энергетически более выгодно, чтобы шпильки оставались закрытыми, а присутствие шпильки, очевидно, способствует плавлению и отсоединению Cas9/dCas9 от этих сайтов.

Оптимизированные gRNA с 5'-шпильками или 3'-шпильками (hpgRNA) значительно повышали специфичность связывания по сравнению со стандартными направляющими РНК и с наилучшими доступными вариантами направляющих РНК (см. примеры), а также устраняли или значительно ослабляли связывание в сайтах протоспейсера, содержащих ошибочные спаривания. Увеличение длины шпильки увеличивало специфичность связывания dCas9.

Оптимизированные gRNA и hpgRNA можно использовать для редактирования сродства и специфичности связывания Cas9/dCas9 или Cpf1. Исходя из размера и структуры шпильки, шпилька hpgRNA может быть размещена в канале связывания ДНК молекулы Cas9/dCas9 и защищена от разрушения. В некоторых вариантах осуществления длину шпильки, длину петли и состав петли можно изменять для обеспечения более точного контроля этих свойств. В некоторых вариантах осуществления шпилька может составлять от приблизительно 1 до приблизительно 20 нуклеотидов или от приблизительно 3 до приблизительно 10 нуклеотидов в длину. Например, шпилька может составлять в длину от 1 до 20, от 1 до 19, от 1 до 18, от 1 до 17, от 1 до 16, от 1 до 15, от 1 до 14, от 1 до 13, от 1 до 12, от 1 до 11, от 1 до 10, от 1 до 9, от 1 до 8, от 1 до 7, от 1 до 6, от 1 до 5, от 2 до 20, от 2 до 19, от 2 до 18, от 2 до 17, от 2 до 16, от 2 до 15, от 2 до 14, от 2 до 13, от 2 до 12, от 2 до 11, от 2 до 10, от 2 до 9, от 2 до 8, от 2 до 7, от 2 до 6, от 2 до 5, от 3 до 20, от 3 до 19, от 3 до 18, от 3 до 17, от 3 до 16, от 3 до 15, от 3 до 14, от 3 до 13, от 3 до 12, от 3 до 11, от 3 до 10, от 3 до 9, от 3 до 8, от 3 до 7, от 3 до 6, от 3 до 5, от 4 до 20, от 4 до 19, от 4 до 18, от 4 до 17, от 4 до 16, от 4 до 15, от 4 до 14, от 4 до 13, от 4 до 12, от 4 до 11, от 4 до 10, от 4 до 9, от 4 до 8, от 4 до 7, от 4 до 6, от 4 до 5, от 5 до 20, от 5 до 19, от 5 до 18, от 5 до 17, от 5 до 16, от 5 до 15, от 5 до 14, от 5 до 13, от 5 до 12, от 5 до 11, от 5 до 10, от 5 до 9, от 5 до 8, от 5 до 7, от 5 до 6, от 6 до 20, от 6 до 19, от 6 до 18, от 6 до 17, от 6 до 16, от 6 до 15, от 6 до 14, от 6 до 13, от 6 до 12, от 6 до 11, от 6 до 10, от 6 до 9, от 6 до 8, от 6 до 7, от 7 до 20, от 7 до 19, от 7 до 18, от 7 до 17, от 7 до 16, от 7 до 15, от 7 до 14, от 7 до 13, от 7 до 12, от 7 до 11, от 7 до 10, от 7 до 9, от 7 до 8, от 8 до 20, от 8 до 19, от 8 до 18, от 8 до 17, от 8 до 16, от 8 до 15, от 8 до 14, от 8 до 13, от 8 до 12, от 8 до 11, от 8 до 10, от 8 до 9, от 9 до 20, от 9 до 19, от 9 до 18, от 9 до 17, от 9 до 16, от 9 до 15, от 9 до 14, от 9 до 13, от 9 до 12, от 9 до 11 или от 9 до 10. Например, шпилька может составлять 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 или 20 или от приблизительно 5 до приблизительно 8 нуклеотидов в длину.

В некоторых вариантах осуществления петля может составлять от приблизительно 1 до приблизительно 20 нуклеотидов, от приблизительно 3 до приблизительно 10 нуклеотидов или от приблизительно 5 до приблизительно 8 нуклеотидов в длину. Например, петля может составлять в длину от 1 до 20, от 1 до 19, от 1 до 18, от 1 до 17, от 1 до 16, от 1 до 15, от 1 до 14, от 1 до 13, от 1 до 12, от 1 до 11, от 1 до 10, от 1 до 9, от 1 до 8, от 1 до 7, от 1 до 6, от 1 до 5, от 2 до 20, от 2 до 19, от 2 до 18, от 2 до 17, от 2 до 16, от 2 до 15, от 2 до 14, от 2 до 13, от 2 до 12, от 2 до 11, от 2 до 10, от 2 до 9, от 2 до 8, от 2 до 7, от 2 до 6, от 2 до 5, от 3 до 20, от 3 до 19, от 3 до 18, от 3 до 17, от 3 до 16, от 3 до 15, от 3 до 14, от 3 до 13, от 3 до 12, от 3 до 11, от 3 до 10, от 3 до 9, от 3 до 8, от 3 до 7, от 3 до 6, от 3 до 5, от 4 до 20, от 4 до 19, от 4 до 18, от 4 до 17, от 4 до 16, от 4 до 15, от 4 до 14, от 4 до 13, от 4 до 12, от 4 до 11, от 4 до 10, от 4 до 9, от 4 до 8, от 4 до 7, от 4 до 6, от 4 до 5, от 5 до 20, от 5 до 19, от 5 до 18, от 5 до 17, от 5 до 16, от 5 до 15, от 5 до 14, от 5 до 13, от 5 до 12, от 5 до 11, от 5 до 10, от 5 до 9, от 5 до 8, от 5 до 7, от 5 до 6, от 6 до 20, от 6 до 19, от 6 до 18, от 6 до 17, от 6 до 16, от 6 до 15, от 6 до 14, от 6 до 13, от 6 до 12, от 6 до 11, от 6 до 10, от 6 до 9, от 6 до 8, от 6 до 7, от 7 до 20, от 7 до 19, от 7 до 18, от 7 до 17, от 7 до 16, от 7 до 15, от 7 до 14, от 7 до 13, от 7 до 12, от 7 до 11, от 7 до 10, от 7 до 9, от 7 до 8, от 8 до 20, от 8 до 19, от 8 до 18, от 8 до 17, от 8 до 16, от 8 до 15, от 8 до 14, от 8 до 13, от 8 до 12, от 8 до 11, от 8 до 10, от 8 до 9, от 9 до 20, от 9 до 19, от 9 до 18, от 9 до 17, от 9 до 16, от 9 до 15, от 9 до 14, от 9 до 13, от 9 до 12, от 9 до 11 или от 9 до 10. В некоторых вариантах осуществления петля может составлять 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 или 20 или от приблизительно 5 до приблизительно 8 нуклеотидов в длину.

В некоторых вариантах осуществления в состав петли могут входить от приблизительно 1 до приблизительно 20 нуклеотидов, от приблизительно 3 до приблизительно 10 нуклеотидов или от приблизительно 5 до приблизительно 8 нуклеотидов. Например, в состав петли могут входить от 1 до 20, от 1 до 19, от 1 до 18, от 1 до 17, от 1 до 16, от 1 до 15, от 1 до 14, от 1 до 13, от 1 до 12, от 1 до 11, от 1 до 10, от 1 до 9, от 1 до 8, от 1 до 7, от 1 до 6, от 1 до 5, от 2 до 20, от 2 до 19, от 2 до 18, от 2 до 17, от 2 до 16, от 2 до 15, от 2 до 14, от 2 до 13, от 2 до 12, от 2 до 11, от 2 до 10, от 2 до 9, от 2 до 8, от 2 до 7, от 2 до 6, от 2 до 5, от 3 до 20, от 3 до 19, от 3 до 18, от 3 до 17, от 3 до 16, от 3 до 15, от 3 до 14, от 3 до 13, от 3 до 12, от 3 до 11, от 3 до 10, от 3 до 9, от 3 до 8, от 3 до 7, от 3 до 6, от 3 до 5, от 4 до 20, от 4 до 19, от 4 до 18, от 4 до 17, от 4 до 16, от 4 до 15, от 4 до 14, от 4 до 13, от 4 до 12, от 4 до 11, от 4 до 10, от 4 до 9, от 4 до 8, от 4 до 7, от 4 до 6, от 4 до 5, от 5 до 20, от 5 до 19, от 5 до 18, от 5 до 17, от 5 до 16, от 5 до 15, от 5 до 14, от 5 до 13, от 5 до 12, от 5 до 11, от 5 до 10, от 5 до 9, от 5 до 8, от 5 до 7, от 5 до 6, от 6 до 20, от 6 до 19, от 6 до 18, от 6 до 17, от 6 до 16, от 6 до 15, от 6 до 14, от 6 до 13, от 6 до 12, от 6 до 11, от 6 до 10, от 6 до 9, от 6 до 8, от 6 до 7, от 7 до 20, от 7 до 19, от 7 до 18, от 7 до 17, от 7 до 16, от 7 до 15, от 7 до 14, от 7 до 13, от 7 до 12, от 7 до 11, от 7 до 10, от 7 до 9, от 7 до 8, от 8 до 20, от 8 до 19, от 8 до 18, от 8 до 17, от 8 до 16, от 8 до 15, от 8 до 14, от 8 до 13, от 8 до 12, от 8 до 11, от 8 до 10, от 8 до 9, от 9 до 20, от 9 до

19, от 9 до 18, от 9 до 17, от 9 до 16, от 9 до 15, от 9 до 14, от 9 до 13, от 9 до 12, от 9 до 11 или от 9 до 10. В некоторых вариантах осуществления в состав петли могут входить 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 или 20 или от приблизительно 5 до приблизительно 8 нуклеотидов.

Композиции могут включать в себя вирусный вектор и систему на основе CRISPR/Cas9 или систему на основе CRISPR/Cpf1 по меньшей мере с одной gRNA, такой как оптимизированная gRNA, описанная в данном документе. В некоторых вариантах осуществления композиция включает в себя модифицированный вектор на основе AAV и нуклеотидную последовательность, кодирующую систему на основе CRISPR/Cas9 по меньшей мере с одной gRNA, такой как оптимизированная gRNA, описанная в данном документе. Композиция может дополнительно содержать донорную ДНК или трансген. Эти композиции могут быть использованы в редактировании генома, в конструировании генома и в коррекции или уменьшении эффектов мутаций в генах, вовлеченных в генетические заболевания.

Целевой ген может быть вовлечен в дифференцировку клетки или любой другой процесс, в котором может быть желательна активация, репрессия или разрушение гена, или может иметь мутацию, такую как делеция, мутация со сдвигом рамки или нонсенс-мутация. Если целевой ген имеет мутацию, которая является причиной преждевременного стоп-кодона, абберрантного сайта акцептора сплайсинга или абберрантного сайта донора сплайсинга, то система на основе CRISPR/Cas9 или система на основе CRISPR/Cpf1 по меньшей мере с одной gRNA, такой как оптимизированная gRNA, описанная в данном документе, может быть сконструирована для распознавания и связывания нуклеотидной последовательности выше или ниже преждевременного стоп-кодона, абберрантного сайта акцептора сплайсинга или абберрантного сайта донора сплайсинга. Система на основе CRISPR/Cas9 или система на основе CRISPR/Cpf1 по меньшей мере с одной gRNA, такой как оптимизированная gRNA, описанная в данном документе, может также использоваться для разрушения нормального сплайсинга генов путем целенаправленного воздействия на акцепторы и доноры сплайсинга, чтобы вызвать пропуск преждевременных стоп-кодонов или восстановить поврежденную рамку считывания. Система на основе CRISPR/Cas9 или система на основе CRISPR/Cpf1 по меньшей мере с одной gRNA, такой как оптимизированная gRNA, описанная в данном документе, может опосредовать или не опосредовать нецелевые изменения в областях генома, кодирующих белок.

В некоторых вариантах осуществления система на основе CRISPR/Cas9 индуцирует или осуществляет репрессию целевого гена в по меньшей мере приблизительно 1 раз, по меньшей мере приблизительно 2 раза, по меньшей мере приблизительно 3 раза, по меньшей мере приблизительно 4 раза, по меньшей мере приблизительно 5 раз, по меньшей мере приблизительно 6 раз, по меньшей мере приблизительно 7 раз, по меньшей мере приблизительно 8 раз, по меньшей мере приблизительно 9 раз, по меньшей мере приблизительно 10 раз, по меньшей мере 15 раз, по меньшей мере 20 раз, по меньшей мере 30 раз, по меньшей мере 40 раз, по меньшей мере 50 раз, по меньшей мере 60 раз, по меньшей мере 70 раз, по меньшей мере 80 раз, по меньшей мере 90 раз, по меньшей мере 100 раз, по меньшей мере приблизительно 110 раз, по меньшей мере 120 раз, по меньшей мере 130 раз, по меньшей мере 140 раз, по меньшей мере 150 раз, по меньшей мере 160 раз, по меньшей мере 170 раз, по меньшей мере 180 раз, по меньшей мере 190 раз, по меньшей мере 200 раз, по меньшей мере приблизительно 300 раз, по меньшей мере 400 раз, по меньшей мере 500 раз, по меньшей мере 600 раз, по меньшей мере 700 раз, по меньшей мере 800 раз, по меньшей мере 900 раз, по меньшей мере 1000 раз, по меньшей мере 1500 раз, по меньшей мере 2000 раз, по меньшей мере 2500 раз, по меньшей мере 3000 раз, по меньшей мере 3500 раз, по меньшей мере 4000 раз, по меньшей мере 4500 раз, по меньшей мере 5000 раз, по меньшей мере 6000 раз, по меньшей мере 7000 раз, по меньшей мере 8000 раз, по меньшей мере 9000 раз, по меньшей мере 10000 раз, по меньшей мере 100000 раз по сравнению с контрольным уровнем экспрессии гена. Контрольным уровнем генной экспрессии целевого гена может быть уровень генной экспрессии целевого гена в клетке, на которую не оказывают воздействия какой-либо из систем на основе CRISPR/Cas9.

а) Модифицированный лентивирусный вектор

Композиции для редактирования генома, геномной перестройки или изменения генной экспрессии целевого гена могут включать в себя модифицированный лентивирусный вектор. Модифицированный лентивирусный вектор включает в себя первую полинуклеотидную последовательность, кодирующую систему, нацеливающуюся на ДНК, и вторую полинуклеотидную последовательность, кодирующую по меньшей мере одну sgRNA. Первая полинуклеотидная последовательность может быть функционально связана с промотором. Промотор может представлять собой конститутивный промотор, индуцируемый промотор, репрессируемый промотор или регулируемый промотор.

Вторая полинуклеотидная последовательность кодирует по меньшей мере 1 gRNA, такую как оптимизированная gRNA, описанная в данном документе. Например, вторая полинуклеотидная последовательность может кодировать по меньшей мере 1 gRNA, по меньшей мере 2 gRNA, по меньшей мере 3 gRNA, по меньшей мере 4 gRNA, по меньшей мере 5 gRNA, по меньшей мере 6 gRNA, по меньшей мере 7 gRNA, по меньшей мере 8 gRNA, по меньшей мере 9 gRNA, по меньшей мере 10 gRNA, по меньшей мере 11 gRNA, по меньшей мере 12 gRNA, по меньшей мере 13 gRNA, по меньшей мере 14 gRNA, по меньшей мере 15 gRNA, по меньшей мере 16 gRNA, по меньшей мере 17 gRNA, по меньшей мере 18 gRNA, по меньшей мере 19 gRNA, по меньшей мере 20 gRNA, по меньшей мере 25 gRNA, по меньшей

мере 30 gRNA, по меньшей мере 35 gRNA, по меньшей мере 40 gRNA, по меньшей мере 45 gRNA или по меньшей мере 50 gRNA. Вторая полинуклеотидная последовательность может кодировать от 1 gRNA до 50 gRNA, от 1 gRNA до 4 5 gRNA, от 1 gRNA до 40 gRNA, от 1 gRNA до 35 gRNA, от 1 gRNA до 30 gRNA, от 1 gRNA до 25 разных gRNA, от 1 gRNA до 20 gRNA, от 1 gRNA до 16 gRNA, от 1 gRNA до 8 разных gRNA, от 4 разных gRNA до 50 разных gRNA, от 4 разных gRNA до 4 5 разных gRNA, от 4 разных gRNA до 40 разных gRNA, от 4 разных gRNA до 35 разных gRNA, от 4 разных gRNA до 30 разных gRNA, от 4 разных gRNA до 25 разных gRNA, от 4 разных gRNA до 20 разных gRNA, от 4 разных gRNA до 16 разных gRNA, от 4 разных gRNA до 8 разных gRNA, от 8 разных gRNA до 50 разных gRNA, от 8 разных gRNA до 4 5 разных gRNA, от 8 разных gRNA до 40 разных gRNA, от 8 разных gRNA до 35 разных gRNA, от 8 разных gRNA до 30 разных gRNA, от 8 разных gRNA до 25 разных gRNA, от 8 разных gRNA до 20 разных gRNA, от 8 разных gRNA до 16 разных gRNA, от 16 разных gRNA до 50 разных gRNA, от 16 разных gRNA до 45 разных gRNA, от 16 разных gRNA до 40 разных gRNA, от 16 разных gRNA до 35 разных gRNA, от 16 разных gRNA до 30 разных gRNA, от 16 разных gRNA до 25 разных gRNA или от 16 разных gRNA до 20 разных gRNA. Каждая из полинуклеотидных последовательностей, кодирующих разные gRNA, может быть функционально связана с промотором. Промоторы, которые функционально связаны с разными gRNA, могут быть одним и тем же промотором. Промоторы, которые функционально связаны с разными gRNA, могут быть разными промоторами. Промотор может представлять собой конститутивный промотор, индуцируемый промотор, репрессируемый промотор или регулируемый промотор. По меньшей мере одна gRNA может связываться с геном-мишенью или целевым локусом. Если включено более одной gRNA, каждая из gRNA связывается с другой целевой областью в пределах одного целевого локуса, или каждая из gRNA связывается с другой целевой областью в пределах разных генных локусов.

b) Векторы на основе аденоассоциированного вируса

AAV может использоваться для доставки композиций в клетку с использованием различных конфигураций конструкции. Например, AAV может доставлять кассеты экспрессии системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1 и gRNA в разных векторах. В качестве альтернативы, если используются белки Cas9 малого размера, полученные из таких видов, как *Staphylococcus aureus* или *Neisseria meningitidis*, то и кассета экспрессии Cas9, и до двух кассет экспрессии gRNA могут быть объединены в одном векторе на основе AAV с пределом упаковки, составляющим 4,7 т.о.

Композиция, описанная выше, включает в себя модифицированный вектор на основе аденоассоциированного вируса (AAV). Модифицированный вектор на основе AAV может быть способен к доставке и экспрессировать систему на основе CRISPR/Cas9 или систему на основе CRISPR/Cpf1 в клетке млекопитающего. Например, модифицированный вектор на основе AAV может представлять собой вектор AAV-SASTG (Piacentino et al. (2012) *Human Gene Therapy* 23:635-646). Модифицированный вектор на основе AAV может быть на основе одного или нескольких из ряда типов капсида, включая AAV1, AAV2, AAV5, AAV6, AAV8 и AAV9. Модифицированный вектор на основе AAV может быть на основе псевдо-типа AAV2 с альтернативными тропными к мышцам капсидами AAV, таким как векторы AAV2/1, AAV2/6, AAV2/7, AAV2/8, AAV2/9, AAV2.5 и AAV/SASTG, которые эффективно транслируют скелетную мышцу или сердечную мышцу путем системной и местной доставки (Seto et al. *Current Gene Therapy* (2012) 12:139-151).

9. Клетки-мишени

Раскрываемая в данном документе gRNA, такая как описанная в данном документе оптимизированная gRNA, может быть использована с системой CRISPR/Cas9 с любым типом клетки. В некоторых вариантах осуществления клетка представляет собой клетку бактерий, клетку грибов, клетку архей, клетку растений или клетку животных, как, например, клетку млекопитающего. В некоторых вариантах осуществления это может быть орган или организм животного. В некоторых вариантах осуществления клетка может представлять собой любой тип клетки или линию клеток, в том числе без ограничения клетки 293-T, клетки 3T3, клетки 721, клетки 9L, клетки A2780, клетки A2780ADR, клетки A2780cis, клетки A172, клетки A20, клетки A253, клетки A431, клетки A-549, клетки ALC, клетки B16, клетки B35, клетки BCP-1, клетки BEAS-2B, клетки bEnd.3, клетки BHK-21, клетки BR 293, клетки VxPC3, клетки C2C12, клетки C3H-10T1/2, клетки C6/36, клетки Cal-27, клетки CHO, клетки COR-L23, клетки COR-L23/CPR, клетки COR-L23/5010, клетки COR-L23/R23, клетки COS-7, клетки COV-434, клетки CML T1, клетки CMT, клетки CT26, клетки D17, клетки DH82, клетки DU145, клетки DuCaP, клетки EL4, клетки EM2, клетки EM3, клетки EMT6/AR1, клетки EMT6/AR10.0, клетки FM3, клетки H1299, клетки H69, клетки HB54, клетки HB55, клетки HCA2, клетки HEK-293, клетки HeLa, клетки Heralc1c7, клетки HL-60, клетки HMEC, клетки HT-29, клетки Jurkat, клетки J558L, клетки JY, клетки K562, клетки Ku812, клетки KCL22, клетки KG1, клетки KYO1, клетки LNCaP, клетки Ma-Mel 1, 2, 3... 48, клетки MC-38, клетки MCF-7, клетки MCF-10A, клетки MDA-MB-231, клетки MDA-MB-468, клетки MDA-MB-435, клетки MDCK II, клетки MDCK II, клетки MG63, клетки MOR/0.2R, клетки MONO-MAC 6, клетки MRC5, клетки MTD-1A, клетки MuEnd, клетки NCI-H69/CPR, клетки NCI-H69/LX10, клетки NCI-H69/LX20, клетки NCI-H69/LX4, клетки NIH-3T3, клетки NALM-1, клетки NW-145, клетки OPCN/OPCT, клетки Peer, клетки PNT-1A/PNT 2, клетки Raji, клетки RBL, клетки RenCa, клетки RIN-5F, клетки RMA/RMAS, клетки Saos-

2, клетки Sf-9, клетки SiHa, клетки SkBr3, клетки T2, клетки T-47D, клетки T84, клетки THP1, клетки U373, клетки U87, клетки U937, клетки VCaP, клетки Vero, клетки WM39, клетки WT-49, клетки X63, клетки YAC-1, клетки YAR, GM12878, K562, эмбриональные стволовые клетки человека H1, HeLa-S3, HepG2, HUVeC, SK-N-SH, IMR90, A549, MCF7, HMEC или LHCM, CD14+, CD20+, первичные клетки сердца или печени, дифференцированные клетки H1, 8988T, Adult CD4 naive, Adult CD4 Th0, Adult CD4 Th1, AG04449, AG04450, AG09309, AG09319, AG10803, AoAF, AoSMC, BC_Adipose_UHN00001, BC_Adrenal_Gland_H12803N, BC_Bladder_01-11002, BC_Brain_H11058N, BC_Breast_02-03015, BC_Colon_01-11002, BC_Colon_H12817N, BC_Esophagus_01-11002, BC_Esophagus_H12817N, BC_Jejunum_H12817N, BC_Kidney_01-11002, BC_Kidney_H12817N, BC_Left_Ventricle_N41, BC_Leukocyte_UHN00204, BC_Liver_01-11002, BC_Lung_01-11002, BC_Lung_H12817N, BC_Pancreas_H12817N, BC_Penis_H12817N, BC_Pericardium_H12529N, BC_Placenta_UHN00189, BC_Prostate_Gland_H12817N, BC_Rectum_N29, BC_Skeletal_Muscle_01-11002, BC_Skeletal_Muscle_H12817N, BC_Skin_01-11002, BC_Small_Intestine_01-11002, BC_Spleen_H12817N, BC_Stomach_01-11002, BC_Stomach_H12817N, BC_Testis_N30, BC_Uterus_BN0765, BE2_C, BG02ES, BG02ES-EBD, BJ, bone_marrow_HS27a, bone_marrow_HS5, bone_marrow_MSC, Breast_OC, Caco-2, CD20+_RO01778, CD20+_RO01794, CD34+_Mobilized, CD4+_Naive_Wb11970640, CD4+_Naive_Wb78495824, Cerebellum_OC, Cerebrum_frontal_OC, Chorion, CLL, CMK, Colo829, Colon_BC, Colon_OC, Cord_CD4 naive, Cord_CD4_Th0, Cord_CD4_Th1, децидуальной оболочки, Dnd41, ECC-1, Endometrium_OC, Esophagus_BC, фибробласты, Fibrobl_GM03348, FibroP, FibroP_AG08395, FibroP_AG08396, FibroP_AG20443, Frontal_cortex_OC, GC_B_cell, глиобластомы, GM04503, GM04504, GM06990, GM08714, GM10248, GM10266, GM10847, GM12801, GM12812, GM12813, GM12864, GM12865, GM12866, GM12867, GM12868, GM12869, GM12870, GM12871, GM12872, GM12873, GM12874, GM12875, GM12878-XiMat, GM12891, GM12892, GM13976, GM13977, GM15510, GM18505, GM18507, GM18526, GM18951, GM19099, GM19193, GM19238, GM19239, GM19240, GM20000, H0287, H1-нейроны, H7-hESC, H9ES, H9ES-AFP-, H9ES-AFP+, H9ES-CM, H9ES-E, H9ES-EB, H9ES-EBD, HAc, HAEpiC, HA-h, HAL, HAoAF, HAoAF_6090101.11, HAoAF_6111301.9, HAoEC, HAoEC_7071706.1, HAoEC_8061102.1, HA-sp, HBMEC, HBVP, HBVSMC, HCF, HCFaa, HCH, HCH_0011308.2P, HCH_8100808.2, HCM, HConF, HCPEpiC, HCT-116, Heart_OC, Heart_STL003, HEEpiC, HEK293, HEK293T, HEK293-T-REX, гепатоциты, HFDPC, HFDPC_0100503.2, HFDPC_0102703.3, HFF, HFF-Myc, HFL11W, HFL24W, HGF, HHSEC, HPEpiC, HL-60, HMEpC, HMEpC_6022801.3, HMF, hMNC-CB, hMNC-CB_8072802.6, hMNC-CB_9111701.6, hMNC-PB, hMNC-PB_0022330.9, hMNC-PB_0082430.9, hMSC-AT, hMSC-AT_0102604.12, hMSC-AT_9061601.12, hMSC-BM, hMSC-BM_0050602.11, hMSC-BM_0051105.11, hMSC-UC, hMSC-UC_0052501.7, hMSC-UC_0081101.7, HMVEC-dAd, HMVEC-dBI-Ad, HMVEC-dBI-Neo, HMVEC-dLy-Ad, HMVEC-dLy-Neo, HMVEC-dNeo, HMVEC-LB1, HMVEC-LLy, HNPCEpiC, HOB, HOB_0090202.1, HOB_0091301, HPAEC, HPAEpiC, HPAF, HPC-PL, HPC-PL_0032601.13, HPC-PL_0101504.13, HPDE6-E6E7, HPdLF, HPIEpC, HPIEpC_9012801.2, HPIEpC_9041503.2, HRCEpiC, HRE, HRGEC, HRPEpiC, HSaVEC, HSaVEC_0022202.16, HSaVEC_9100101.15, HSMM, HSMM_emb, HSMM_FSHD, HSMMtube, HSMMtube_emb, HSMMtube_FSHD, HT-1080, HTR8svn, Huh-7, Huh-7.5, HVMF, HVMF_6091203.3, HVMF_6100401.3, HWP, HWP_0092205, HWP_8120201.5, iPS, iPS_CWRU1, iPS_hFib2_iPS4, iPS_hFib2_iPS5, iPS_NiHi11, iPS_NiHi7, Ishikawa, Jurkat, Kidney_BC, Kidney_OC, LHCN-M2, LHSR, Liver_OC, Liver_STL004, Liver_STL011, LNCaP, Loucy, Lung_BC, Lung_OC, Lymphoblastoid_cell_line, M059J, MCF10A-Er-Src, MCF-7, MDA-MB-231, Medullo, Medullo_D341, Mel_2183, меланоциты, Monocytes-CD14+, Monocytes-CD14+_RO01746, Monocytes-CD14+_RO01826, MRT_A204, MRT_G401, MRT_TTC549, миометрия, Native_B_cell, NB4, NH-A, NHBE, NHBE_RA, NHDF, NHDF_0060801.3, NHDF_7071701.2, NHDF-Ad, NHDF-neo, NHEK, NHEM.f_M2, NHEM.f_M2_5071302.2, NHEM.f_M2_6022001, NHEM_M2, NHEM_M2_7_011001.2, NHEM_M2_7012303, NHLF, NT2-D1, Olf_neurosphere, остеобласты, ovcar-3, PANC-1, Pancreas_OC, PanIsletD, PanIslets, PBDE, PBDEFetal, PBMC, PFSK-1, pHTE, Pons_OC, PrEC, ProgFib, предстательной железы, Prostate_OC, Psoas_muscle_OC, Raji, RCC_7860, RPMI-7951, RPTEC, RWPE1, SAEC, SH-SY5Y, Skeletal_Muscle_BC, SkMC, SKMC, SkMC_8121902.17, SkMC_9011302, SK-N-MC, SK-N-SH_RA, Small_intestine_OC, Spleen_OC, звездчатые, Stomach_BC, T_cells_CD4+, T-47D, T98G, TBEC, Th1, Th1_Wb33676984, Th1_Wb54553204, Th17, Th2, Th2_Wb33676984, Th2_Wb54553204, Treg_Wb78495824, Treg_Wb83319432, U2OS, U87, UCH-1, уротелиальные, WERI-Rb-1 и WI-38. В некоторых вариантах осуществления клетка-мишень может представлять собой любую клетку, такую как первичная клетка, клетка HEK293, клетка 293Ts, клетка SKBR3, клетка A431, клетка K562, клетка HCT116, клетка HepG2 или группы K-Ras-зависимых и K-Ras-независимых клеток.

10. Способы эпигеномного редактирования

Настоящее изобретение относится к способу эпигеномного редактирования в клетке-мишени или субъекте с использованием системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1. Способ может быть использован для активации или подавления целевого гена. Способ включает приведение в контакт клетки или субъекта с эффективным количеством оптимизированной молекулы gRNA, описываемой в данном документе, и системой на основе CRISPR/Cas9 или системой на основе CRISPR/Cpf1. В

некоторых вариантах осуществления оптимизированная gRNA кодируется полинуклеотидной последовательностью и упакована в лентивирусный вектор. В некоторых вариантах осуществления лентивирусный вектор содержит кассету экспрессии, содержащую промотор, функционально связанный с полинуклеотидной последовательностью, кодирующей sgRNA. В некоторых вариантах осуществления промотор, функционально связанный с полинуклеотидом, кодирующим оптимизированную gRNA, является индуцируемым.

11. Способы сайт-специфического расщепления ДНК

Настоящее раскрытие относится к способу сайт-специфического расщепления ДНК в клетке-мишени или субъекте с использованием системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1. Способ включает приведение в контакт клетки или субъекта с эффективным количеством оптимизированной молекулы gRNA, описываемой в данном документе, и системой на основе CRISPR/Cas9 или системой на основе CRISPR/Cpf1. В некоторых вариантах осуществления оптимизированная gRNA кодируется полинуклеотидной последовательностью и упакована в лентивирусный вектор. В некоторых вариантах осуществления лентивирусный вектор содержит кассету экспрессии, содержащую промотор, функционально связанный с полинуклеотидной последовательностью, кодирующей sgRNA. В некоторых вариантах осуществления промотор, функционально связанный с полинуклеотидом, кодирующим оптимизированную gRNA, является индуцируемым.

Количество gRNA, вводимой в клетку или образец, может составлять по меньшей мере 1 gRNA, по меньшей мере 2 разные gRNA, по меньшей мере 3 разные gRNA, по меньшей мере 4 разные gRNA, по меньшей мере 5 разных gRNA, по меньшей мере 6 разных gRNA, по меньшей мере 7 разных gRNA, по меньшей мере 8 разных gRNA, по меньшей мере 9 разных gRNA, по меньшей мере 10 разных gRNA, по меньшей мере 11 разных gRNA, по меньшей мере 12 разных gRNA, по меньшей мере 13 разных gRNA, по меньшей мере 14 разных gRNA, по меньшей мере 15 разных gRNA, по меньшей мере 16 разных gRNA, по меньшей мере 17 разных gRNA, по меньшей мере 18 разных gRNA, по меньшей мере 18 разных gRNA, по меньшей мере 20 разных gRNA, по меньшей мере 25 разных gRNA, по меньшей мере 30 разных gRNA, по меньшей мере 35 разных gRNA, по меньшей мере 40 разных gRNA, по меньшей мере 45 разных gRNA или по меньшей мере 50 разных gRNA. Количество gRNA, вводимой в клетку, может составлять от по меньшей мере 1 gRNA до по меньшей мере 50 разных gRNA, от по меньшей мере 1 gRNA до по меньшей мере 45 разных gRNA, от по меньшей мере 1 gRNA до по меньшей мере 40 разных gRNA, от по меньшей мере 1 gRNA до по меньшей мере 35 разных gRNA, от по меньшей мере 1 gRNA до по меньшей мере 25 разных gRNA, от по меньшей мере 1 gRNA до по меньшей мере 20 разных gRNA, от по меньшей мере 1 gRNA до по меньшей мере 16 разных gRNA, от по меньшей мере 1 gRNA до по меньшей мере 12 разных gRNA, от по меньшей мере 1 gRNA до по меньшей мере 8 разных gRNA, от по меньшей мере 1 gRNA до по меньшей мере 4 разных gRNA, от по меньшей мере 4 gRNA до по меньшей мере 50 разных gRNA, от по меньшей мере 4 разных gRNA до по меньшей мере 40 разных gRNA, от по меньшей мере 4 разных gRNA до по меньшей мере 35 разных gRNA, от по меньшей мере 4 разных gRNA до по меньшей мере 30 разных gRNA, от по меньшей мере 4 разных gRNA до по меньшей мере 25 разных gRNA, от по меньшей мере 4 разных gRNA до по меньшей мере 20 разных gRNA, от по меньшей мере 4 разных gRNA до по меньшей мере 16 разных gRNA, от по меньшей мере 4 разных gRNA до по меньшей мере 12 разных gRNA, от по меньшей мере 4 разных gRNA до по меньшей мере 8 разных gRNA, от по меньшей мере 8 разных gRNA до по меньшей мере 50 разных gRNA, от по меньшей мере 8 разных gRNA до по меньшей мере 45 разных gRNA, от по меньшей мере 8 разных gRNA до по меньшей мере 40 разных gRNA, от по меньшей мере 8 разных gRNA до по меньшей мере 35 разных gRNA, от по меньшей мере 8 разных gRNA до по меньшей мере 30 разных gRNA, от по меньшей мере 8 разных gRNA до по меньшей мере 25 разных gRNA, от 8 разных gRNA до по меньшей мере 20 разных gRNA, от по меньшей мере 8 разных gRNA до по меньшей мере 16 разных gRNA или от 8 разных gRNA до по меньшей мере 12 разных gRNA.

gRNA может содержать полинуклеотидную последовательность, комплементарную последовательности целевой ДНК, после которой следует последовательность PAM. gRNA может содержать "G" на 5'-конце комплементарной полинуклеотидной последовательности. gRNA может содержать полинуклеотидную последовательность из по меньшей мере 10 пар оснований, по меньшей мере 11 пар оснований, по меньшей мере 12 пар оснований, по меньшей мере 13 пар оснований, по меньшей мере 14 пар оснований, по меньшей мере 15 пар оснований, по меньшей мере 16 пар оснований, по меньшей мере 17 пар оснований, по меньшей мере 18 пар оснований, по меньшей мере 19 пар оснований, по меньшей мере 20 пар оснований, по меньшей мере 21 пары оснований, по меньшей мере 22 пар оснований, по меньшей мере 23 пар оснований, по меньшей мере 24 пар оснований, по меньшей мере 25 пар оснований, по меньшей мере 30 пар оснований или по меньшей мере 35 пар оснований, комплементарную последовательности целевой ДНК, после которой следует последовательность PAM.

Последовательность PAM может представлять собой "NGG", где "N" может представлять собой любой нуклеотид. gRNA может нацеливаться по меньшей мере на одно из промоторной области, энхансерной области или транскрибируемой области целевого гена. В некоторых вариантах осуществления

gRNA нацеливается на последовательность нуклеиновой кислоты, имеющую полинуклеотидную последовательность по меньшей мере одного из SEQ ID NO: 13-148, 316, 317 или 320. gRNA может включать в себя последовательность нуклеиновой кислоты по меньшей мере одного из SEQ ID NO: 149-315, 321-323 или 326-329.

12. Способы коррекции мутантного гена и лечения субъекта

Настоящее раскрытие также направлено на способ коррекции мутантного гена у субъекта. Способ предусматривает введение в клетку субъекта композиции, описанной выше. Применение композиции для доставки системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1 по меньшей мере с одной gRNA, такой как оптимизированная gRNA, описанная в данном документе, в клетку может восстанавливать экспрессию полнофункционального или частично функционального белка с помощью матрицы репарации или донорной ДНК, которая может заменять весь ген или область, содержащую мутацию. Система на основе CRISPR/Cas9 или система на основе CRISPR/Cpf1, содержащая по меньшей мере одну gRNA, такую как оптимизированная gRNA, описанная в данном документе, может использоваться для введения сайт-специфических двухнитевых разрывов в целевых геномных локусах. Сайт-специфические двухнитевые разрывы создаются тогда, когда система на основе CRISPR/Cas9 или система на основе CRISPR/Cpf1 по меньшей мере с одной gRNA, такой как оптимизированная gRNA, описанная в данном документе, связывается с последовательностями целевой ДНК, позволяя тем самым расщепление целевой ДНК. Это расщепление ДНК может стимулировать естественный механизм репарации ДНК, сводящийся к одному из двух возможных путей репарации: репарации с участием гомологичной рекомбинации (HDR) или негомологичного соединения концов (NHEJ).

Настоящее раскрытие направлено на редактирование генома с использованием системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1 по меньшей мере с одной gRNA, такой как оптимизированная gRNA, описанная в данном документе, без матрицы репарации, которые могут эффективно корректировать рамку считывания и восстанавливать экспрессию функционального белка, вовлеченного в генетическое заболевание. Раскрытая система на основе CRISPR/Cas9 или система на основе CRISPR/Cpf1 по меньшей мере с одной gRNA, такой как оптимизированная gRNA, описанная в данном документе, может предусматривать использование подходов коррекции на основе репарации с участием гомологичной рекомбинации или опосредованного нуклеазой негомологичного соединения концов (NHEJ), обеспечивающих эффективную коррекцию в линиях первичных клеток с ограниченной пролиферацией, которые могут не поддаваться гомологичной рекомбинации или генной коррекции на основе отбора. Эта стратегия объединяет быструю и надежную сборку активной системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1 по меньшей мере с одной gRNA, такой как оптимизированная gRNA, описанная в данном документе, с эффективным способом редактирования генов для лечения генетических заболеваний, вызванных мутациями в несущественных кодирующих областях, которые являются причиной сдвига рамки, преждевременных стоп-кодонов, абберрантных сайтов доноров сплайсинга или абберрантных сайтов акцепторов сплайсинга.

а) Опосредованное нуклеазой негомологичное соединение концов

Восстановление экспрессии белка из эндогенного мутированного гена может проводиться путем NHEJ-опосредованной репарации ДНК без участия матрицы. В отличие от временного способа нацеливания на РНК целевого гена коррекция рамки считывания целевого гена в геноме с помощью временно экспрессирующейся системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1 по меньшей мере с одной gRNA, такой как оптимизированная gRNA, описанная в данном документе, может приводить к полному восстановлению экспрессии целевого гена каждой модифицированной клеткой и во всем ее потомстве.

Коррекция гена с помощью опосредованного нуклеазой NHEJ может исправлять мутированный целевой ген и предлагает несколько потенциальных преимуществ над путем HDR. Например, NHEJ не требует донорной матрицы, которая может вызывать неспецифический инсерционный мутагенез. В отличие от HDR, NHEJ эффективно работает на всех стадиях клеточного цикла и поэтому может эффективно использоваться как в делящихся, так и в постмитотических клетках, как, например, в мышечных волокнах. Это обеспечивает надежное, полное восстановление гена, являющееся альтернативой пропуску экзонов на основе олигонуклеотидов или фармакологически иницируемому сквозному прохождению стоп-кодонов, и теоретически может потребоваться всего лишь обработка одним лекарственным средством. Коррекцию гена на основе NHEJ с использованием системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1, а также других сконструированных нуклеаз, включая мегануклеазы и нуклеазы типа "цинковые пальцы", можно комбинировать с другими существующими платформами *ex vivo* и *in vivo* для клеточной и генной терапии в дополнение к описанному в данном документе способу электропорации плазмид. Например, доставка системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1 путем переноса генов на основе мРНК или в виде очищенных белков, способных проникать через клетку, может обеспечить возможность использования подхода к редактированию генома без ДНК, что позволит избежать любой возможности инсерционного мутагенеза.

б) Репарация с участием гомологичной рекомбинации

Восстановление экспрессии белка из эндогенного мутированного гена может предусматривать ре-

парацию с участием гомологичной рекомбинации. Способ, описанный выше, дополнительно включает введение донорной матрицы в клетку. Донорная матрица может включать в себя нуклеотидную последовательность, кодирующую полнофункциональный белок или частично функциональный белок. Например, донорная матрица может включать в себя конструкцию дистрофина с уменьшенным размером, называемую минидистрофином ("минидис"), полнофункциональную конструкцию дистрофина для восстановления мутантного гена дистрофина или фрагмент гена дистрофина, который после гомологичной репарации приводит к восстановлению мутантного гена дистрофина.

13. Способы редактирования генома

Настоящее раскрытие также направлено на редактирование генома с помощью системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1, описанных выше, для восстановления экспрессии полнофункционального или частично функционального белка с использованием матрицы репарации или донорной ДНК, которая может заменить весь ген или область, содержащую мутацию. Система на основе CRISPR/Cas9 или система на основе CRISPR/Cpf1 может использоваться для введения сайт-специфических двухнитевых разрывов в целевых геномных локусах. Сайт-специфические двухнитевые разрывы создаются тогда, когда система на основе CRISPR/Cas9 или система на основе CRISPR/Cpf1 связывается с последовательностями целевой ДНК с помощью gRNA, позволяя тем самым расщепление целевой ДНК. Система на основе CRISPR/Cas9 и система на основе CRISPR/Cpf1 имеют преимущество усовершенствованного редактирования генома из-за их высокого уровня успешной и эффективной генетической модификации. Это расщепление ДНК может стимулировать естественный механизм репарации ДНК, сводящийся к одному из двух возможных путей репарации: репарации с участием гомологичной рекомбинации (HDR) или негомологичного соединения концов (NHEJ).

Настоящее раскрытие направлено на редактирование генома с использованием системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1 без матрицы репарации, которые могут эффективно корректировать рамку считывания и восстанавливать экспрессию функционального белка, вовлеченного в генетическое заболевание. Раскрытая система на основе CRISPR/Cas9 или система на основе CRISPR/Cpf1 и способы могут предусматривать использование подходов коррекции на основе репарации с участием гомологичной рекомбинации или опосредованного нуклеазой негомологичного соединения концов (NHEJ), что обеспечивает возможность эффективной коррекции в линиях первичных клеток с ограниченной пролиферацией, которые могут не поддаваться гомологичной рекомбинации или генной коррекции на основе отбора. Эта стратегия объединяет быструю и надежную сборку активной системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1 с эффективным способом редактирования генов для лечения генетических заболеваний вызванных мутациями в несущественных кодирующих областях, которые являются причиной сдвига рамки, преждевременных стоп-кодонах, абберрантных сайтов доноров сплайсинга или абберрантных сайтов акцепторов сплайсинга.

Настоящее раскрытие предусматривает способы коррекции мутантного гена в клетке и лечения субъекта, страдающего от генетического заболевания, такого как DMD. Способ может включать в себя введение в клетку или субъекту системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1, полинуклеотида или вектора, кодирующего указанную систему на основе CRISPR/Cas9 или систему на основе CRISPR/Cpf1, или композицию, содержащую указанную систему на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1, как описано выше. Способ может включать в себя введение системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1, такое как введение белка Cas9, белка Cpf1, слитого белка Cas9, содержащего второй домен, нуклеотидной последовательности, кодирующей указанный белок Cas9, белок Cpf1 или слитый белок Cas9, и/или по меньшей мере одной gRNA, где gRNA нацеливаются на разные последовательности ДНК. Целевые последовательности ДНК могут быть перекрывающимися. Количество gRNA, вводимой в клетку, может составлять по меньшей мере 1 gRNA, по меньшей мере 2 разные gRNA, по меньшей мере 3 разные gRNA, по меньшей мере 4 разные gRNA, по меньшей мере 5 разных gRNA, по меньшей мере 6 разных gRNA, по меньшей мере 7 разных gRNA, по меньшей мере 8 разных gRNA, по меньшей мере 9 разных gRNA, по меньшей мере 10 разных gRNA, по меньшей мере 15 разных gRNA, по меньшей мере 20 разных gRNA, по меньшей мере 30 разных gRNA или по меньшей мере 50 разных gRNA, описанных выше. gRNA может включать в себя последовательность нуклеиновой кислоты по меньшей мере одного из SEQ ID NO: 149-315, 321-323 или 326-329. Способ может предусматривать репарацию с участием гомологичной рекомбинации или негомологичное соединение концов.

14. Конструкции и плазмиды

Композиции, описанные выше, могут содержать генетические конструкции, которые кодируют систему на основе CRISPR/Cas9 или систему на основе CRISPR/Cpf1, раскрываемую в данном документе. Генетическая конструкция, такая как плазида, может содержать нуклеиновую кислоту, которая кодирует систему на основе CRISPR/Cas9 или систему на основе CRISPR/Cpf1, как например: белок Cas9, белок Cpf1 и слитые белки Cas9 и/или по меньшей мере одну из оптимизированных gRNA, описанных в данном документе. Композиции, как описано выше, могут содержать генетические конструкции, которые кодируют модифицированный вектор на основе AAV и последовательность нуклеиновой кислоты, которая кодирует систему на основе CRISPR/Cas9 или систему на основе CRISPR/Cpf1, по меньшей мере с

одной gRNA, такой как оптимизированная gRNA, описанная в данном документе.

Генетическая конструкция, такая как плазида, может содержать нуклеиновую кислоту, которая кодирует систему на основе CRISPR/Cas9 или систему на основе CRISPR/Cpf1, по меньшей мере с одной gRNA, такой как оптимизированная gRNA, описанная в данном документе. Композиции, как описано выше, могут содержать генетические конструкции, которые кодируют модифицированный лентивирусный вектор, раскрываемый в данном документе. Генетическая конструкция, такая как плазида, может содержать нуклеиновую кислоту, которая кодирует Cas9-слитый белок, и по меньшей мере одну sgRNA. Генетическая конструкция может присутствовать в клетке в виде функционирующей внехромосомной молекулы. Генетическая конструкция может быть линейной минихромосомой, включающей центромеру, теломеры или плазмиды или космиды.

Генетическая конструкция также может быть частью генома рекомбинантного вирусного вектора, в том числе рекомбинантного лентивируса, рекомбинантного аденовируса и рекомбинантного аденоассоциированного вируса. Генетическая конструкция может быть частью генетического материала в аттенуированных живых микроорганизмах или рекомбинантных микробных векторах, которые заселяют клетки. Генетические конструкции могут содержать регуляторные элементы для экспрессии генов кодирующих последовательностей нуклеиновой кислоты. Регуляторные элементы могут представлять собой промотор, энхансер, иницирующий кодон, стоп-кодон или сигнал полиаденилирования.

Последовательности нуклеиновой кислоты могут составлять генетическую конструкцию, которая может представлять собой вектор. Вектор может быть способным к экспрессированию слитого белка, такого как Cas9-слитый белок, в клетке млекопитающего. Вектор может быть рекомбинантным. Вектор может содержать гетерологичную нуклеиновую кислоту, кодирующую Cas9-слитый белок. Вектор может представлять собой плазмиду. Вектор может быть применим для трансфекции клеток нуклеиновой кислотой, кодирующей Cas9-слитый белок, при этом трансформированную клетку-хозяина культивируют и поддерживают в условиях, при которых происходит экспрессия системы Cas9-слитого белка.

Кодирующие последовательности могут быть оптимизированы для стабильности и высоких уровней экспрессии. В некоторых случаях кодоны выбирают для уменьшения образования вторичной структуры РНК, такой как структура, образовавшаяся вследствие внутримолекулярного связывания.

Вектор может содержать гетерологичную нуклеиновую кислоту, кодирующую систему на основе CRISPR/Cas9 или систему на основе CRISPR/Cpf1, и может дополнительно содержать иницирующий кодон, который может располагаться выше кодирующей последовательности системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1, и стоп-кодон, который может располагаться ниже кодирующей последовательности системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1. Иницирующий кодон и кодон терминации могут находиться в рамке с кодирующей последовательностью системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1. Вектор также может содержать промотор, который функционально связан с кодирующей последовательностью системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1. Промотор, функционально связанный с кодирующей последовательностью системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1, может представлять собой промотор из обезьяньего вируса 40 (SV40), промотор вируса опухоли молочной железы мыши (MMTV), промотор вируса иммунодефицита человека (ВИЧ), такой как промотор длинного концевой повтора (LTR) вируса иммунодефицита крупного рогатого скота (BIV), промотор вируса Молони, промотор вируса птичьего лейкоза (ALV), промотор цитомегаловируса (CMV), такой как промотор гена немедленного раннего ответа CMV, промотор вируса Эпштейна-Барр (EBV) или промотор вируса саркомы Рауса (RSV). Промотор также может представлять собой промотор из гена человека, такого как ген убиквитина С человека (hUbC), актина человека, миозина человека, гемоглобина человека, креатина мышц человека или металлотронеина человека. Промотор также может быть тканеспецифическим промотором, таким как специфический для мышц или кожи промотор, природный или синтетический. Примеры таких промоторов описаны в публикации патентной заявки США № US 20040175727, содержание которой включено в данный документ во всей своей полноте.

Вектор также может содержать сигнал полиаденилирования, который может располагаться ниже системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1. Сигнал полиаденилирования может представлять собой сигнал полиаденилирования SV40, сигнал полиаденилирования LTR, сигнал полиаденилирования бычьего гормона роста (bGH), сигнал полиаденилирования гормона роста человека (hGH) или сигнал полиаденилирования β -глобина человека. Сигнал полиаденилирования SV40 может представлять собой сигнал полиаденилирования из вектора pCEP4 (Invitrogen, Сан-Диего, Калифорния).

Вектор также может содержать энхансер выше системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1, т.е. кодирующей последовательности белка Cas9, белка Cpf1 или Cas9-слитого белка или sgRNA, такой как оптимизированная gRNA, описанная в данном документе. Энхансер может быть необходим для экспрессии ДНК. Энхансер может представлять собой энхансер гена актина человека, миозина человека, гемоглобина человека, креатина мышц человека, или энхансер из вируса, такого как один из CMV, HA, RSV или EBV. Энхансеры для осуществления функции полинуклеотидами описаны в патентах США №№ 5593972, 5962428 и в заявке WO 94/016737, содержание каждого из которых полностью включено посредством ссылки. Вектор также может содержать точку начала репликации млекопи-

тающих, чтобы поддерживать вектор вне хромосомы и создавать множественные копии вектора в клетке. Вектор также может содержать регуляторную последовательность, которая может хорошо подходить для экспрессии генов в клетке млекопитающего или человека, в которую вводится вектор. Вектор также может содержать репортерный ген, такой как ген зеленого флуоресцентного белка ("GFP"), и/или селективный маркер, такой как ген резистентности к гигромицину ("Hygro").

Вектор может представлять собой векторы экспрессии или системы для получения белка с помощью стандартных методик и с помощью общедоступных исходных материалов, включенных в справочник Sambrook et al., *Molecular Cloning and Laboratory Manual*, Second Ed., Cold Spring Harbor (1989), который полностью включен посредством ссылки. В некоторых вариантах осуществления вектор может содержать последовательность нуклеиновой кислоты, кодирующую систему на основе CRISPR/Cas9 или систему на основе CRISPR/Cpf1, в том числе последовательность нуклеиновой кислоты, кодирующую белок Cas9, белок Cpf1 или Cas9-слитый белок, и последовательность нуклеиновой кислоты, кодирующую по меньшей мере одну gRNA, содержащую последовательность нуклеиновой кислоты, по меньшей мере одного из SEQ ID NO: 149-315, 321-323 или 326-329.

15. Фармацевтические композиции

Композиция может быть в составе фармацевтической композиции. Фармацевтическая композиция может содержать от приблизительно 1 нг до приблизительно 10 мг ДНК, кодирующей систему на основе CRISPR/Cas9, систему на основе CRISPR/Cpf1 или белковый компонент системы на основе CRISPR/Cas9, т.е. белок Cas9, белок Cpf1 или Cas9-слитый белок. Фармацевтическая композиция может содержать от приблизительно 1 нг до приблизительно 10 мг ДНК модифицированного вектора на основе AAV и нуклеотидной последовательности, кодирующей систему на основе CRISPR/Cas9, по меньшей мере с одной gRNA, такой как оптимизированная gRNA, описанная в данном документе.

Фармацевтическая композиция может содержать от приблизительно 1 нг до приблизительно 10 мг ДНК модифицированного лентивирусного вектора. Фармацевтические композиции в соответствии с настоящим изобретением составляют в соответствии с используемым способом введения. В тех случаях, когда фармацевтические композиции представляют собой инъекционные фармацевтические композиции, они являются стерильными, свободны от пирогенов и не содержат частиц. Предпочтительно использование изотонического состава. Как правило, добавки для изотоничности могут включать хлорид натрия, декстрозу, маннит, сорбит и лактозу. В некоторых случаях предпочтительными являются изотонические растворы, такие как фосфатно-буферный солевой раствор. Стабилизаторы включают желатин и альбумин. В некоторых вариантах осуществления в состав добавляют сосудосуживающее средство.

Композиция может дополнительно содержать фармацевтически приемлемое вспомогательное вещество.

Фармацевтически приемлемое вспомогательное вещество может представлять собой функциональные молекулы, такие как наполнители, адъюванты, носители или разбавители.

Фармацевтически приемлемое вспомогательное вещество может представлять собой средство, облегчающее трансфекцию, которое может включать поверхностно-активные средства, такие как иммуностимулирующие комплексы (ISCOMS), неполный адъювант Фрейнда, аналог LPS, в том числе монофосфорилилипид А, мурамиловые пептиды, аналоги хинона, везикулы, такие как сквален и сквален, гиалуроновая кислота, липиды, липосомы, ионы кальция, вирусные белки, полианионы, поликатионы или наночастицы или другие известные средства, облегчающие трансфекцию.

Средство, облегчающее трансфекцию, представляет собой полианион, поликатион, в том числе поли-L-глутамат (LGS), или липид. Средство, облегчающее трансфекцию, представляет собой поли-L-глутамат, а более предпочтительно поли-L-глутамат присутствует в композиции для редактирования генома в концентрации, составляющей менее 6 мг/мл. Средство, облегчающее трансфекцию, может также включать поверхностно-активные средства, такие как иммуностимулирующие комплексы (ISCOMS), неполный адъювант Фрейнда, аналог LPS, в том числе монофосфорилилипид А, мурамиловые пептиды, аналоги хинона и везикулы, такие как сквален и сквален, причем также можно использовать гиалуроновую кислоту при введении в сочетании с генетической конструкцией. В некоторых вариантах осуществления ДНК-вектор, кодирующий композицию, также может включать в себя средство, облегчающее трансфекцию, такое как липиды, липосомы, в том числе лецитиновые липосомы, или другие липосомы, известные в данной области, в виде смеси ДНК-липосом (см., например, WO 9324640), ионы кальция, вирусные белки, полианионы, поликатионы или наночастицы или другие известные средства, облегчающие трансфекцию. Предпочтительно средство, облегчающее трансфекцию, представляет собой полианион, поликатион, в том числе поли-L-глутамат (LGS), или липид.

16. Конструкции и плазмиды

Композиции, описанные выше, могут содержать генетические конструкции, которые кодируют систему на основе CRISPR/Cas9 или систему на основе CRISPR/Cpf1, раскрываемую в данном документе. Генетическая конструкция, такая как плаزمид или вектор экспрессии, может содержать нуклеиновую кислоту, которая кодирует систему на основе CRISPR/Cas9 или систему на основе CRISPR/Cpf1, и/или по меньшей мере с одной gRNA, такой как оптимизированная gRNA, описанная в данном документе. Композиции, описанные выше, могут содержать генетические конструкции, которые кодируют модифи-

цированный лентивирусный вектор и последовательность нуклеиновой кислоты, которая кодирует систему на основе CRISPR/Cas9 или систему на основе CRISPR/Cpf1, раскрываемую в данном документе. Генетическая конструкция, такая как плазида, может содержать нуклеиновую кислоту, которая кодирует систему на основе CRISPR/Cas9 или систему на основе CRISPR/Cpf1. Композиции, как описано выше, могут содержать генетические конструкции, которые кодируют модифицированный лентивирусный вектор. Генетическая конструкция, такая как плазида, может содержать нуклеиновую кислоту, которая кодирует систему на основе CRISPR/Cas9 или систему на основе CRISPR/Cpf1, и по меньшей мере одну sgRNA, такую как оптимизированная gRNA, описанная в данном документе.

Генетическая конструкция может присутствовать в клетке в виде функционирующей внехромосомной молекулы. Генетическая конструкция может быть линейной минихромосомой, включающей центромеру, теломеры или плазмиды или космиды.

Генетическая конструкция также может быть частью генома рекомбинантного вирусного вектора, в том числе рекомбинантного лентивируса, рекомбинантного аденовируса и рекомбинантного аденоассоциированного вируса. Генетическая конструкция может быть частью генетического материала в аттенуированных живых микроорганизмах или рекомбинантных микробных векторах, которые заселяют клетки. Генетические конструкции могут содержать регуляторные элементы для экспрессии генов кодирующих последовательностей нуклеиновой кислоты. Регуляторные элементы могут представлять собой промотор, энхансер, иницирующий кодон, стоп-кодон или сигнал полиаденилирования.

Последовательности нуклеиновой кислоты могут составлять генетическую конструкцию, которая может представлять собой вектор. Вектор может быть способным к экспрессии слитого белка, такого как система на основе CRISPR/Cas9 или система на основе CRISPR/Cpf1, в клетке млекопитающего. Вектор может быть рекомбинантным. Вектор может содержать гетерологичную нуклеиновую кислоту, кодирующую слитый белок, такой как система на основе CRISPR/Cas9. Вектор может представлять собой плазмиду. Вектор может быть применим для трансфекции клеток нуклеиновой кислотой, кодирующей систему на основе CRISPR/Cas9 или систему на основе CRISPR/Cpf1, при этом трансформированную клетку-хозяина культивируют и поддерживают в условиях, при которых происходит экспрессия системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1.

Кодирующие последовательности могут быть оптимизированы для стабильности и высоких уровней экспрессии. В некоторых случаях кодоны выбирают для уменьшения образования вторичной структуры РНК, такой как структура, образовавшаяся вследствие внутримолекулярного связывания.

Вектор может содержать гетерологичную нуклеиновую кислоту, кодирующую систему на основе CRISPR/Cas9 или систему на основе CRISPR/Cpf1, и может дополнительно содержать иницирующий кодон, который может располагаться выше кодирующей последовательности системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1, и стоп-кодон, который может располагаться ниже кодирующей последовательности системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1. Иницирующий кодон и кодон терминации могут находиться в рамках с кодирующей последовательностью системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1. Вектор также может содержать промотор, который функционально связан с кодирующей последовательностью системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1. Система на основе CRISPR/Cas9 или система на основе CRISPR/Cpf1 могут находиться под индуцируемым светом или химически индуцируемым контролем для обеспечения возможности динамического контроля в пространстве и времени. Промотор, функционально связанный с кодирующей последовательностью системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1, может представлять собой промотор из обезьяньего вируса 40 (SV40), промотор вируса опухоли молочной железы мыши (MMTV), промотор вируса иммунодефицита человека (ВИЧ), такой как промотор длинного концевой повтора (LTR) вируса иммунодефицита крупного рогатого скота (BIV), промотор вируса Молони, промотор вируса птичьего лейкоза (ALV), промотор цитомегаловируса (CMV), такой как промотор гена немедленного раннего ответа CMV, промотор вируса Эпштейна-Барр (EBV) или промотор вируса саркомы Рауса (RSV). Промотор также может представлять собой промотор из гена человека, такого как ген убиквитина С человека (hUbC), актина человека, миозина человека, гемоглобина человека, креатина мышц человека или металлотионеина человека. Промотор также может быть тканеспецифическим промотором, таким как специфический для мышц или кожи промотор, природный или синтетический. Примеры таких промоторов описаны в публикации патентной заявки США № US 20040175727, содержание которой включено в данный документ во всей своей полноте.

Вектор также может содержать сигнал полиаденилирования, который может располагаться ниже системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1. Сигнал полиаденилирования может представлять собой сигнал полиаденилирования SV40, сигнал полиаденилирования LTR, сигнал полиаденилирования бычьего гормона роста (bGH), сигнал полиаденилирования гормона роста человека (hGH) или сигнал полиаденилирования β -глобина человека. Сигнал полиаденилирования SV40 может представлять собой сигнал полиаденилирования из вектора pCEP4 (Invitrogen, Сан-Диего, Калифорния).

Вектор также может содержать энхансер выше системы на основе CRISPR/Cas9, или системы на основе CRISPR/Cpf1, и/или sgRNA, такой как оптимизированная gRNA, описанная в данном документе. Энхансер может быть необходим для экспрессии ДНК. Энхансер может представлять собой энхансер

гена актина человека, миозина человека, гемоглобина человека, креатина мышц человека или энхансер из вируса, такого как один из CMV, HA, RSV или EBV. Энхансеры для осуществления функции полинуклеотидами описаны в патентах США №№ 5593972, 5962428 и в заявке WO 94/016737, содержание каждого из которых полностью включено посредством ссылки. Вектор также может содержать точку начала репликации млекопитающих, чтобы поддерживать вектор вне хромосомы и создавать множественные копии вектора в клетке. Вектор также может содержать регуляторную последовательность, которая может хорошо подходить для экспрессии генов в клетке млекопитающего или человека, в которую вводится вектор. Вектор также может содержать репортерный ген, такой как ген зеленого флуоресцентного белка ("GFP"), и/или селектируемый маркер, такой как ген резистентности к гигромицину ("Hygro").

Вектор может представлять собой векторы экспрессии или системы для получения белка с помощью стандартных методик и с помощью общедоступных исходных материалов, включенных в справочник Sambrook et al., *Molecular Cloning and Laboratory Manual*, Second Ed., Cold Spring Harbor (1989), который полностью включен посредством ссылки. В некоторых вариантах осуществления вектор может содержать последовательность нуклеиновой кислоты, кодирующую систему на основе CRISPR/Cas9 или систему на основе CRISPR/Cpf1, и последовательность нуклеиновой кислоты, кодирующую по меньшей мере одну gRNA, такую как оптимизированная gRNA, описанная в данном документе.

В некоторых вариантах осуществления gRNA, такая как оптимизированная gRNA, описанная в данном документе, кодируется полинуклеотидной последовательностью и упакована в лентивирусный вектор. В некоторых вариантах осуществления лентивирусный вектор включает в себя кассету экспрессии. Кассета экспрессии может включать в себя промотор, функционально связанный с полинуклеотидной последовательностью, кодирующей gRNA, такую как оптимизированная gRNA, описанная в данном документе. В некоторых вариантах осуществления промотор, функционально связанный с полинуклеотидом, кодирующим gRNA, является индуцируемым.

i) Векторы на основе аденоассоциированного вируса

Композиция, описанная выше, включает в себя модифицированный вектор на основе аденоассоциированного вируса (AAV). Модифицированный вектор на основе AAV может обладать повышенным тропизмом к сердечной и скелетной мышечной ткани. Модифицированный вектор на основе AAV может быть способен доставлять и экспрессировать систему на основе CRISPR/Cas9 или систему на основе CRISPR/Cpf1 по меньшей мере с одной gRNA, такой как оптимизированная gRNA, описанная в данном документе, в клетке млекопитающего. Например, модифицированный вектор на основе AAV может представлять собой вектор AAV-SASTG (Piacentino et al. (2012) *Human Gene Therapy* 23:635-646). Модифицированный вектор на основе AAV может доставлять нуклеазы в скелетную и сердечную мышцу *in vivo*. Модифицированный вектор на основе AAV может быть на основе одного или нескольких из ряда типов капсида, включая AAV1, AAV2, AAV5, AAV6, AAV8 и AAV9. Модифицированный вектор на основе AAV может быть на основе псевдотипа AAV2 с альтернативными тропными к мышцам капсидами AAV, таким как векторы AAV2/1, AAV2/6, AAV2/7, AAV2/8, AAV2/9, AAV2.5 и AAV/SASTG, которые эффективно трансдуцируют скелетную мышцу или сердечную мышцу путем системной и местной доставки (Seto et al. *Current Gene Therapy* (2012) 12:139-151).

17. Способы доставки

В данном документе предусмотрен способ доставки системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1 и оптимизированной gRNA, описанной в данном документе, для обеспечения генетических конструкций и/или белков системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1. Доставка системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1 и оптимизированной gRNA, описанной в данном документе, может осуществляться с помощью трансфекции или электропорации системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1 и оптимизированной gRNA, описанной в данном документе, в виде одной или нескольких молекул нуклеиновой кислоты, которые экспрессируются в клетке и доставляются на поверхность клетки. Белок системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1 можно доставлять в клетку. Молекулы нуклеиновых кислот можно подвергать электропорации с использованием устройств BioRad Gene Pulser Xcell или Маха Nucleofector Pb или другого устройства для электропорации. Можно использовать несколько различных буферов, в том числе раствор для электропорации BioRad, фосфатно-буферный солевой раствор от Sigma с № D8537 (PBS), Invitrogen OptiMEM I (ОМ) или Амаха Nucleofector solution V (N.V.). Трансфекции могут предусматривать реагент для трансфекции, такой как Lipofectamine 2000.

Вектор, кодирующий белок системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1, можно доставлять в модифицированную клетку-мишень в ткани или субъекте путем инъекции ДНК (также называемой ДНК-вакцинацией) с электропорацией *in vivo* и без нее, опосредованно липосомами, с помощью наночастиц и/или рекомбинантных векторов. Рекомбинантный вектор можно доставить с помощью любой вирусной формы. Вирусная форма может представлять собой рекомбинантный лентивирус, рекомбинантный аденовирус и/или рекомбинантный аденоассоциированный вирус.

Нуклеотид, кодирующий белок системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1, можно вводить в клетку для индуцирования генной экспрессии целевого гена. Например, одну или несколько нуклеотидных последовательностей, кодирующих систему на основе CRISPR/Cas9

или систему на основе CRISPR/Cpf1, нацеливающуюся на целевой ген, можно вводить в клетку млекопитающего. После доставки системы на основе CRISPR/Cas9 или системы на основе CRISPR/Cpf1 в клетку и, как результат, вектора в клетки млекопитающего, трансфицированные клетки будут экспрессировать систему на основе CRISPR/Cas9 или систему на основе CRISPR/Cpf1. Систему на основе CRISPR/Cas9 или систему на основе CRISPR/Cpf1 можно вводить млекопитающему для индуцирования или модулирования генной экспрессии целевого гена у млекопитающего. Млекопитающее может представлять собой человека, примата, отличного от человека, корову, свинью, овцу, козу, антилопу, бизона, водного буйвола, быка, оленя, ежа, слона, ламу, альпаку, мышей, крыс или курицу, предпочтительно человека, корову, свинью или курицу.

Способы введения нуклеиновой кислоты в клетку-хозяина известны в данной области, и любой известный способ можно применять для введения нуклеиновой кислоты (например, конструкции для экспрессии) в клетку. Подходящие способы включают в себя, например, вирусную трансфекцию или инфекцию бактериофагом, трансфекцию, конъюгацию, слияние протопластов, липофекцию, электропорацию, осаждение фосфатом кальция, опосредованную полиэтиленгликолем (PEG) трансфекцию, опосредованную DEAE-декстраном трансфекцию, опосредованную липосомами трансфекцию, технологию биобаллистической пушки, осаждение фосфатом кальция, прямую микроинъекцию, опосредованную наночастицами доставку нуклеиновой кислоты и им подобные. В некоторых вариантах осуществления композицию можно доставлять с помощью мРНК-доставки и доставки рибонуклеопротеинового (RNP) комплекса.

18. Пути введения

Композиции могут вводиться субъекту различными путями, включая пероральный, парентеральный, сублингвальный, трансдермальный, ректальный, трансмукозальный, местный, посредством ингаляции, посредством буккального введения, внутривенный, внутриаартериальный, внутривенный, подкожный, внутримышечный, интраназальный, интратекальный и внутрисуставной или их комбинации. Для ветеринарного применения композицию можно вводить в виде отвечающего требованиям приемлемого состава в соответствии с обычной ветеринарной практикой. Ветеринар может легко определить схему применения и способ введения, который наиболее подходит для конкретного животного. Композиции можно вводить традиционными шприцами, безыгольными устройствами для инъекций, "пушками для проведения баллистической трансфекции" или другими физическими способами, такими как электропорация ("EP"), "гидродинамический метод" или ультразвук.

Композицию можно доставить млекопитающему с помощью нескольких технологий, включая инъекцию ДНК (также называемую вакцинацией ДНК) с электропорацией *in vivo* и без нее, опосредованную липосомами, с помощью наночастиц, рекомбинантных векторов, таких как рекомбинантный лентивирус, рекомбинантный аденовирус и рекомбинантный адено-ассоциированный вирус. Композицию можно вводить в скелетную мышцу или сердечную мышцу. Например, композицию можно вводить в переднюю большеберцовую мышцу.

19. Наборы

В данном документе предусмотрен набор, который можно использовать для сайт-специфического связывания ДНК. Набор содержит композицию, как описано выше, и инструкции для использования указанной композиции. Инструкции, включенные в наборы, могут быть прикреплены к упаковочному материалу или могут быть включены в виде листка-вкладыша в упаковке. Хотя инструкции обычно представляют собой письменные или печатные материалы, они не ограничены таковыми. Любой носитель, способный хранить такие инструкции и сообщать их конечному пользователю, рассматривается в данном изобретении. Такие носители включают в себя без ограничения электронные носители (например, магнитные диски, кассеты, картриджи, чипы), оптические носители (например, CD ROM) и тому подобное. Используемый в данном документе термин "инструкции" может включать адрес интернет-сайта, который предоставляет инструкции.

Композиция может включать модифицированный вектор на основе лентивируса и нуклеотидную последовательность, кодирующую систему на основе CRISPR/Cas9 и оптимизированную gRNA, как описано выше. Система на основе CRISPR/Cas9, как описано выше, может быть включена в набор для специфического связывания и нацеливания на конкретную регуляторную область гена-мишени.

20. Примеры

Смысл вышеизложенного может быть лучше раскрыт при рассмотрении следующих примеров, которые представлены в иллюстративных целях и не предназначены для ограничения объема настоящего изобретения.

Пример 1.

Материалы и способы

Материалы. Буфер трис-HCl (pH 7,6) получали от Corning Life Sciences. Моногидрат монокалевой соли L-глутаминовой кислоты, дитиотреитол (DTT) и хлорид магния получали из Sigma Aldrich Co., LLC.

Клонирование плазмид экспрессии Cas9, dCas9 и sgRNA.

Плазмиды, кодирующие Cas9, dCas9 и sgRNA, которые нацеливаются на локус AAVS1 хромосомы 19 человека, клонировали, экспрессировали и очищали с использованием стандартных методов. Субстра-

ты ДНК, используемые для визуализации - (i) субстрат 1198 п.о., полученный из сегмента локуса AAVS1 хромосомы 19 человека; (ii) "сконструированный" ДНК-субстрат 989 п.о., содержащий серию из шести полных, частичных или ошибочно спаренных целевых сайтов; и (iii) "нонсенс"-субстрат 1078 п.о., не содержащий гомологии с протоспейсером (> 3 п.о.) - также получали с использованием стандартных методов. Плазмиды, кодирующие Cas9 дикого типа и dCas9, получали из Addgene (плазида 39312 и плазида 47106). Плазмиды для экспрессии Cas9 и dCas9 в бактериях клонировали с использованием Gateway Cloning (Life Technologies). Коротко, ПЦР использовали для амплификации генов Cas9 и dCas9 и добавления фланкирующих сайтов attL1 и attL2. ВР-рекомбинацию проводили для переноса этих генов на челночный вектор, после чего проводили LP-рекомбинацию для переноса этих генов на pDest17, которая добавляет N-концевую гексагистидиновую метку (Life Technologies). Плазмиды, кодирующие химерную sgRNA и варианты sgRNA (описанные ниже), клонировали, как описано выше (Perez-Pinera et al., (2013) Nature methods, 10, 973-976).

Экспрессия и очистка Cas9, dCas9.

Плазмидами, кодирующими Cas9 или dCas9, трансформировали компетентные клетки SoluBL21 (Genlantis) в соответствии со стандартными методами (Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular cloning. Cold Spring Harbor Laboratory Press, New York). Отдельные колонии использовали для инокуляции 25 мл стартовых культур. 25 мл стартовых культур выращивали в течение ночи и использовали для инокуляции 1 л культур. Инокулированные культуры объемом 1 л выращивали в течение 5 ч при 25°C, после чего температуру снижали до 16°C и индуцировали экспрессию белка добавлением 0,1 mM IPTG. Индуцированные культуры выращивали еще в течение 12 ч при 16°C. Клетки собирали центрифугированием при 4 000 × g и хранили при -80°C с целью длительного хранения.

Клеточный осадок ресуспендировали в 30 мл буфера для лизиса (50 mM трис-HCl, 500 mM NaCl, 10 mM MgCl₂, 10% об. глицерина, 0,2% Тритона-1000 и 1 mM PMSF). Клеточную суспензию лизировали ультразвуком при 30% рабочем цикле в течение 5 мин. Затем суспензию центрифугировали в течение 30 мин при 12000 × g. Затем супернатант отбирали и инкубировали со смолой Ni-NTA (Qiagen) в течение 30 мин при осторожном перемешивании. Затем смолу загружали в колонку, промывали промывочным буфером (35 mM имидизола, 50 mM трис-HCl, 500 mM NaCl, 10 mM MgCl₂, 10% об. глицерина) и элюировали элюирующим буфером (120 mM имидизола, 50 mM Трис-HCl, 500 mM NaCl, 10 mM MgCl₂, 10% об./об. глицерина). Затем центрифужные фильтры Ultracel-30k использовали для обмена растворителей на буфер для хранения (50 mM трис-HCl, 500 mM NaCl, 10 mM MgCl₂, 10% об./об. глицерина). Затем образцы делили на аликваты и замораживали при -80°C. Типичные полиакриламидные гели с SDS очищенных Cas9 и dCas9 представлены на фиг. S1, указывая чистоту приблизительно > 95%.

Экспрессия и очистка вариантов sgRNA и направляющей RNA. Направляющие РНК транскрибировали *in vitro* с помощью набора для транскрипции MEGAShortscript T7 (Life Technologies). ДНК-матрицы с промотором T7 получали с помощью ПЦР из плазмид для направляющих РНК и реакции устанавливали в соответствии с инструкциями изготовителя. Матрицы T7 для направляющих РНК с 2 усеченными нуклеотидами с 5'-конца (tru-gRNA) и направляющих РНК с 5'-удлинением, которые формируют шпильки (hp-gRNA), получали с помощью ПЦР из плазмид стандартных gRNA. Затем РНК очищали с использованием экстракции фенол-хлороформом с помощью стандартных методов (Sambrook et al. (1989) Molecular cloning. Cold Spring Harbor Laboratory Press, New York).

Получение ДНК-субстратов.

Геномную ДНК экстрагировали и очищали из клеточной линии HEK293T с использованием набора DNeasy (Qiagen), следуя протоколу изготовителя. Затем локус AAVS1 амплифицировали с использованием ПЦР. Субстрат размером 1198 п.о., полученный из AAVS1, конструировали посредством прямой ПЦР из геномной ДНК с использованием праймеров из Integrated DNA Technologies (IDT): 5'-/Bt/-CCAGGATCAGTGAACGCAC-3' и 5'-GAGCTCTACTGGCTTCTGCG-3', где /Bt/ представляет биотинирование праймера на 5'-конце. 'Сконструированный' ДНК-субстрат, который содержит ряд PAM и полных или частичных сайтов протоспейсера, заказывали в виде двух gBlock-фрагментов, каждый из которых содержит сайт рестрикции EcoRI на одном конце. Субстраты расщепляли, лигировали вместе и затем обогащали посредством ПЦР с праймерами (Integrated DNA Technologies, IDT): 5'-\Bt/-CATGACGTGCAGCAAGC-3' и 5'-CGACGATGCGCTGAATC-3'. Для конструирования 'нонсенс'-субстрата, не содержащего гомологичных сайтов (более чем 3 п.о.) для протоспейсера, синтезировали конструкцию ДНК размером 690 п.о. (GeneScript, Inc.), содержащую ряд сайтов рестрикции, и дополнительную длину ДНК из лямбда-ДНК (New England Biolabs) субклонировали в конструкцию; затем субстрат размером 1078 п.о. амплифицировали в ПЦР с использованием праймеров (IDT): 5'-\Bt/-GACCTGCAGGCATGCAAGCTTGG-3' и 5'-CAGCGTCCCGGTTGTGAATCT-3'. Все ДНК очищали в геле, разбавляли до 25 нМ в рабочем буфере (20 mM трис-HCl (pH 7,6), 100 mM глутамата калия, 5 mM MgCl₂ и 0,4 mM DTT) и инкубировали с 40-кратным избытком мономерного стрептавидина (Howarth et al., (2006) Nature methods, 3, 267-273) в течение 10 мин до инкубации с Cas9/dCas9.

Натрийдодecilсульфат-полиакриламидные гели очищенного Cas9 и dCas9 представлены на фиг. 8A-8B, указывая чистоту приблизительно 95%.

Атомно-силовая микроскопия.

Атомно-силовую микроскопию (AFM) проводили на воздухе с использованием Bruker (née Veeco) Nanoscope V Multimode с зондами RTSEP (Bruker) (номинальный коэффициент жесткости 40 Н/м, резонансная частота 300 кГц). До экспериментов белок и направляющие РНК смешивали в соотношении 1:1,5 в течение 10 мин. Белок и ДНК смешивали в растворе рабочего буфера в течение по меньшей мере 10 мин (до 35 мин) при комнатной температуре, осаждали в течение 8 с на свежерасщепленной слюде (Ted Pella, Inc.) которую обрабатывали 3-аминопропилсилоксаном (как описано ранее (24)), промывали ультрачистой ($> 17 \text{ M}\Omega$) водой и сушили на воздухе. Белки предварительно недолго центрифугировали перед инкубацией с ДНК. При использовании стандартной sgRNA получили изображение по меньшей мере четырех препаратов для каждого экспериментального условия и по меньшей мере двух для экспериментов с другими вариантами направляющей РНК. В общем случае изображения получили с разрешающей способностью 1024×1024 на квадратных площадках размером 2,75 мкм или 2048×2048 на квадратных площадках размером 5,5 мкм при 1-1,5 линии/с для каждого образца. Изображения нескольких тысяч (~ 2500 -6000) молекул ДНК были проанализированы для каждого экспериментального условия.

Визуализация и детализация ДНК с экстраразрешающей способностью.

Полученные изображения AFM сплющивали и выравнивали (по плоскости, построчно и с помощью выравнивания полинома 3^{го} порядка) с использованием программного обеспечения для анализа изображений с открытым исходным кодом для сканирующей зондовой микроскопии Gwyddion (<http://gwyddion.net/>), а затем экспортировали в MATLAB (Mathworks, Inc.). Области 151×151 пикселей ($405 \text{ нм} \times 405 \text{ нм}$), содержащие каждую молекулу ДНК, сортировали путем проверки четко идентифицируемой метки стрептавидина, наличия по меньшей мере одной связанной молекулы Cas9/dCas9 и четкого сквозного пути для обеспечения отсутствия агрегации или наложения на другие молекулы ДНК. Контур ДНК прослеживали вручную и отмечали оцененные границы стрептавидина и Cas9/dCas9. Затем кривую записи алгоритмически уточняли с помощью способа, основанного на Wiggins et al. (2006) Nature nanotechnology, 1, 137-141. Начиная с взвешенного центра тяжести стрептавидина (x_1), оценивали положение следующего элемента каркаса (x_2) с шагом 2,5 нм по направлению к ближайшим прослеженным вручную точкам за пределами расчетной границы стрептавидина. 11-пиксельная линия нарисована на двукратной линейной интерполяции изображения ДНК перпендикулярно сегменту линии (x_1 - x_2) в x_2 . x_2 перемещается в положение на нормальной линии с максимальной топографической высотой, а затем корректируется до 2,5 нм из x_1 на новой линии (x_1 - x_2). Затем положения $x_3 \dots x_n$ итерационно оценивают с использованием ближайших прослеженных вручную точек для создания исходного предположения для следующего положения в каркасе, затем корректируют, как и раньше, и процесс коррекции продолжают до тех пор, пока точка x_n не будет находиться менее чем в 2,5 нм от конца отслеживания молекулы ДНК. Когда очищенная кривая записи входит в расчетные границы молекулы Cas9/dCas9 в точке x_i , положение ДНК вместо этого рассчитывают как точку на кубическом сплайне Эрмита (с использованием точек x_{i-1} , x_i , x_j и x_{j+1} , где x_j представляет собой первую точку нарисованной кривой записи за расчетной границей Cas9/dCas9), расположенную на расстоянии 2,5 нм от x_i .

По завершении кривой записи извлекают высоту ДНК вдоль контура (относительно средней высоты пикселя изображения локальной области). Расчетные границы стрептавидина и Cas9/dCas9 итерационно расширяли и стягивали вокруг исходной оценки, пока они не расширились до смежной области более чем на $(\mu_d + \sigma_d)$, где μ_d и σ_d представляют собой среднее и стандартное отклонение высоты визуализируемой ДНК за расчетными положениями связанных белков, и оценку сводили.

Чтобы учесть любой инструментальный гистерезис, который может исказить кажущуюся длину ДНК, длину ДНК нормализовали, и только молекулы ДНК, первоначально измеренные как 20% от их расчетной длины (учитывая известное количество пар оснований, 0,33 нм на пару оснований), использовали для дальнейшего анализа (для субстрата AAVS1 - отслеживаемое количество: 804; номинальная длина: 1198 п.о., средняя длина записи: 1283 п.о., станд. откл: 154 п.о.; для сконструированного субстрата - отслеживаемое количество: 1520, номинальная длина: 986 п.о., средняя длина записи: 1071 п.о., станд. откл: 124 п.о.; для 'нонсенс'-субстрата - отслеживаемое количество: 616, номинальная длина: 1078 п.о., средняя длина записи: 1217 п.о., станд. откл: 135 п.о.). Эта стадия не позволяет нам ненадлежащим образом анализировать, например, две молекулы ДНК, которые оказались на одной прямой, ДНК, которая может быть фрагментирована, или ДНК, которая могла быть расщеплена Cas9 и разделена (что было редко, см. основной текст).

Гистограммы связывания на фиг. 1C-1D, фиг. 2C-2D и фиг. 9 получили путем сопоставления относительного местоположения каждого связанного белка с основаниями, перекрываемыми (интерполяция по соседним элементам) белком, и суммированием общего количества белков, связанных с каждым сайтом (если один Cas9/dCas9 можно интерпретировать как находящийся в контакте с несколькими (k) сайтами, каждую область контакта утяжеляли на $1/k$ в гистограмме связывания). Пики в гистограмме связывания согласовывали с эмпирической функцией Гаусса $\exp(-((x-\mu)/w)^2)$, где μ среднее положение пика и w - параметр ширины пика ($w = \sqrt{2}\sigma$, где σ - стандартное отклонение), с использованием MATLAB.

Определение кажущихся констант диссоциации dCas9. Кажущиеся константы диссоциации dCas9 с различными вариантами направляющих РНК определяли, как описано ранее (Yang et al. (2005) *Nucleic Acids Res.*, 33, 4322-4334). Коротко, при известных концентрациях раствора dCas9-направляющая RNA ($[dCas9]_0$) и молекул ДНК ($[DNA]_0$) подсчитывали соответствующее количество 'сконструированных' молекул ДНК со связанными белками и без них (фракция ДНК, связанной белками Θ_{dCas9}). После визуализации ДНК со связанными белками (см. выше) определяли среднее количество белков, связанных молекулой ДНК (n_{dCas9}). Общие константы диссоциации рассчитывают как $K_{d,ДНК} = [ДНК][dCas9] / [ДНК \cdot dCas9] = (1 - \Theta_{dCas9}) [dCas9]_0 - n_{dCas9} [ДНК]_0 / (\Theta_{dCas9})$.

Протоспейсер-специфические константы диссоциации $K_{d,протоспейсер}$ рассчитывают аналогично, используя вместо этого $\Theta_{dCas9,протоспейсер}$, фракции ДНК с dCas9, связанным в пределах ширины одного пика гауссовского распределения в их соответствующих гистограммах связывания (т.е. см. табл. 1), тогда как сайт-специфические константы ассоциации $K_{a,ss} = K_{d,ss}^{-1}$ с использованием фракций каждого сайта на ДНК со связанным dCas9 $\Theta_{dCas9,ss}$.

Выравнивание и группирование белков.

Извлекали изображения белков Cas9 и dCas9, которые были выделены и наблюдались только при контакте с ДНК в одном местоположении. Эти признаки выбирали как те, у которых характеристики более чем $(\mu_d + 2\sigma_d)$, которые полностью помещаются в контур $134 \text{ нм} \times 134 \text{ нм}$, где μ_d и σ_d представляют собой среднее и стандартное отклонение высоты ДНК, на которой связаны белки; этот этап по существу вызывал удаление большей части агрегированного/плотно упакованного Cas9/dCas9 из набора, а также этих белков из изображений с более выраженными внешними шумами. После четырехкратной интерполяции по соседним элементам каждую характеристику белка с топографической высотой, превышающей $(\mu_d + \sigma_d)$, выравнивали путем повторного смещения, вращения и отражения относительно друг друга, чтобы свести к минимуму среднеквадратические отклонения между их топографическими высотами. Матрицу расстояний составили по этому сведенному к минимуму среднеквадратическому отклонению, затем белки со стандартной sgRNA группировали в соответствии с этим критерием с помощью способа Rodriguez и Laio (27); белки с вариантами направляющих РНК группировали в соответствии с ближайшей структурой Cas9/dCas9 со стандартной sgRNA. Средние по ансамблю структуры извлекали путем выполнения безэталонного выравнивания по каждому представителю отдельных кластеров способом Penczek, Radermacher и Frank (28). Свойства популяций Cas9/dCas9 в каждом признаке (таком как сайты протоспейсера) на ДНК определяли с использованием белков, связанных в пределах ширины одного пика гауссовских распределений, соответствующих гистограммам связывания (т.е., см. табл. 1).

Кинетический метод Монте-Карло (КМС) для встраивания нити направляющей РНК и 'дыхания' R-петли. Эксперименты согласно кинетическому методу Монте-Карло (КМС) для моделирования встраивания нити направляющими РНК в сайтах протоспейсера выполняли с использованием алгоритма Гиллеспи (непрерывное время, дискретное состояние) ((Gillespie (1976) *Journal of computational physics*, 22, 403-434), реализованного в MATLAB. Встраивание нити моделируют как одномерное случайное блуждание в зависящем от положения потенциале, определяемом по относительным, зависимым от соседних элементов свободным энергиям связывания ДНК:ДНК и РНК:ДНК. См., например, фиг. 4А. То есть, основания направляющей РНК спаривают с протоспейсером до сайта протоспейсера ($m \geq m \geq 20$ для sgRNA и $1 \geq m \geq 18$ для усеченной sgRNA (tru-gRNA)) и, в первом приближении, прямую скорость (скорость дополнительного встраивания направляющей РНК) v_f оценивают с помощью симметричной аппроксимации согласно $\exp(-(\Delta G^\circ(m+1)_{РНК:ДНК} - \Delta G^\circ(m+1)_{РНК:ДНК}) / 2RT)$, где R представляет собой постоянную Больцмана, T представляет собой температуру (здесь 37°C для соответствия используемому набору параметров), $\Delta G^\circ(m+1)_{РНК:ДНК}$ представляет собой свободную энергию спаривания оснований между РНК и протоспейсером в сайте $m+1$ и $\Delta G^\circ(m+1)_{ДНК:ДНК}$ представляет собой свободную энергию спаривания оснований между протоспейсером и комплементарной ему нитью ДНК (поправочный член $1/2$ включен для соответствия детальному равновесию). v_f в состоянии $m=20$ или 18 для sgRNA или tru-gRNA устанавливали равным 0. Обратную скорость (скорость повторной гибридизации между протоспейсером и его комплементарной нитью ДНК) v_r рассчитывали аналогичным образом пропорционально $\exp(-(\Delta G^\circ(m)_{ДНК:ДНК} - \Delta G^\circ(m)_{РНК:ДНК}) / 2RT)$; если состояние $m=1$, моделирование останавливали (что означает диссоциацию направляющей РНК - протоспейсера). Начиная с момента времени $t=0$ (в условных единицах времени), для каждой итерации алгоритма определяют m -зависимые скорости и два случайных числа r_1 и r_2 получают из равномерного распределения между 0 и 1. t продвигается на $\Delta t = \log(r_1) / (v_f + v_r)$. Состояние m увеличивается до $m+1$, если $r_2 \geq v_f / (v_f + v_r)$, или снижается до $m - 1$ в противном случае. Для 'равновесных' измерений 'дыхания' R-петли m инициировали при $m=20$ (или 18 в случае tru-gRNA) и алгоритм повторяли до $t \geq 10000$. Для измерений динамики "кинетики встраивания" (такой как в присутствии ошибочно спаренных оснований), m инициировали при $m=10$ (до $t=1000$).

Параметры свободной энергии получают из литературных источников для экспериментов при 1M NaCl и 37°C . Свободные энергии зависимой от последовательности гибридизации ДНК:ДНК $\Delta G^\circ(x)_{ДНК:ДНК}$ получали из SantaLucia et al. (1996) *Biochemistry*, 35, 3555-3562; свободные энергии зави-

симой от последовательности гибридизации РНК:РНК $\Delta G^\circ(x)_{\text{РНК:РНК}}$ получали из Sugimoto et al. (1995) *Biochemistry*, 34, 11211-11216; и значения $\Delta G^\circ(x)_{\text{РНК:ДНК}}$ в случаях введенных точечных ошибочных спариваний rG-dG, rC-dC, rA-dA и rU-dT получали из Watkins et al. (2011) *Nucleic acids research*, 39, 1894-1902 (при условиях с несколько большей концентрацией соли). Последовательность использованного протоспейсера представляет собой 'ATCCTGTCCCTAGTGGCCCC' (SEQ ID NO: 336), целевой сайт AAVS1, как в экспериментах AFM; последовательность комплементарной протоспейсеру ДНК представляет собой 'GGGGCCACTAGGGACAGGAT' (SEQ ID NO: 337) и последовательность направляющей РНК представляет собой 'GGGGCCACUAGGGACAGGAU' (SEQ ID NO: 338) для sgRNA или 'GGCCA-CUAGGGACAGGAU' (SEQ ID NO: 339) для усеченной РНК.

Корреляции между стабильностью R-петли, полученной в КМС, и экспериментальными скоростями расщепления Cas9. Для анализа корреляций между взаимодействиями направляющей РНК протоспейсера и скоростями расщепления Cas9 *in vivo* извлекали последовательности направляющих РНК и целевой ДНК по Hsu et al. (2013) *Nature biotechnology*, 31, 827-832 и их экспериментально определенную оценку максимального правдоподобия (MLE) частот разрезания с помощью Cas9. Последовательности направляющих РНК и целевой ДНК по Hsu et al. (2013) *Nature Biotechnology*, 31, 827-832 с однонуклеотидными РАМ-дистальными (≥ 10 п.о. от сайта РАМ) ошибочными спариваниями по типу rG-dG, rC-dC, rA-dA и rU-dT и экспериментально определенную оценку максимального правдоподобия (MLE) частот разрезания с помощью Cas9 в этих сайтах импортировали ($n=136$) в сценарий выполнения КМС. Моделирования встраивания нити, начатые при $m=10$, повторяли 1000 раз для каждой последовательности (до $t=100$) с получением среднего отрезка времени $m \geq 16$ и сопоставляли с эмпирическими скоростями расщепления. Значимость определяли путем взятия бустрап-выборки средней доли занятости с MLE частот разрезания посредством перестановки 100000 раз, затем пересчета коэффициентов корреляции и значений. Свободные энергии связывания направляющей РНК - протоспейсера оценивали путем суммирования по энергиям ближайших соседних элементов с использованием наборов параметров, перечисленных выше, и корректировали с помощью фактора инициации $-3,1$ ккал моль⁻¹.

Данные dCas9-tru-gRNA и dCas9-hp-gRNA для сравнения со структурными свойствами dCas9-sgRNA.

При сравнении измерений высоты и объема белков в экспериментах условия визуализации AFM должны оставаться в основном одинаковыми, чтобы не вводить артефакты. Это, как правило, не представляет проблемы, например, при сравнении высот и объемов dCas9, связанной с различными сайтами на сконструированных молекулах ДНК, но представляет собой проблему при сравнении структурных свойств dCas9/Cas9 при использовании разных направляющих РНК или ДНК-субстратов. В качестве контроля использовали высоты и объемы белков стрептавидина, используемых для маркировки концов отслуживаемых молекул ДНК, которые должны оставаться неизменными при всех экспериментальных условиях для разных экспериментов. Для экспериментов с sgRNA средние значения высоты стрептавидинов отличались менее чем на 0,1 нм (среднее различие: 0,087 нм; стандартное отклонение различий: 0,052 нм) и их средние объемы (1098 нм³) отличались менее чем на 15 нм³ (среднее различие: 14,461 нм³; стандартное отклонение различий: 10,419 нм³). Однако средние значения высоты и объема в экспериментах с tru-gRNA и hp-gRNA отличались от средних значений высоты и объема с sgRNA на вплоть до 0,14 нм и 225 нм³ соответственно. Чтобы непосредственно сравнить результаты этих экспериментов, значения высоты dCas9 с tru-gRNA и hp-gRNA на сконструированной ДНК смещали на их разницу в средних высотах по сравнению со значениями высоты с sgRNA и объемы пересчитывали с помощью процентной разницы средних объемов.

Пример 2

Атомно-силовая микроскопия фиксирует специфическое и неспецифическое связывание Cas9/dCas9 вдоль сконструированных ДНК-субстратов с высоким разрешением.

Анализ кристаллографических и биохимических экспериментов показывает, что специфичность связывания и расщепления протоспейсера придается сначала путем распознавания сайтов РАМ самим Cas9 с последующим встраиванием нити связанным комплексом РНК и прямым спариванием оснований по Уотсону-Крику с протоспейсером (фиг. 1А), хотя полная механистическая картина еще не выяснена. Чтобы непосредственно исследовать относительные предрасположенности к связыванию с сайтами протоспейсера и нецелевыми сайтами с одномолекулярным разрешением, 50 нМ комплексов Cas9-sgRNA или dCas9-sgRNA, нацеливающихся на локус AAVS1 хромосомы 19 человека, визуализировали с помощью AFM на воздухе после инкубации с одним из трех ДНК-субстратов (2,5 нМ):

(i) сегмент размером 1198 п.о. из локуса AAVS1, содержащий полный целевой сайт после РАМ (hrc 'TGG') (фиг. 1С);

(ii) сконструированный ДНК-субстрат размером 989 п.о., содержащий серию из шести полных, частичных или ошибочно спаренных целевых сайтов, каждый из которых отделен с помощью приблизительно 150 п.о. (фиг. 1D). Ошибочные спаривания в этих сайтах могут охватывать как 'затравочные' (РАМ-проксимальные, приблизительно 12 п.о.), так и 'незатравочные' (РАМ-дистальные) области протоспейсера. Только сайты РАМ в этом сконструированном субстрате были в этих специально разрабо-

танных местоположениях и

(iii) "нонсенс"-ДНК размером 1078 п.о. без гомологии (последовательности свыше 3 п.о.) с целевой последовательностью (фиг. 9А-9С).

На фиг. 1С показано, что dCas9 и Cas9 демонстрируют почти одинаковые распределения связывания на субстрате AAVS1 (n=404 и n=250, соответственно). На фиг. 1D показано, что на скоструированном субстрате (n=536) dCas9 связывается с наивысшей предрасположенностью с полным протоспейсером без ошибочно спаренных (ММ) сайтов (пик 1, позже называемый полным или "0ММ" сайтом), а также с сайтами с 5 или 10 ошибочно спаренными основаниями, дистальными относительно сайта РАМ (третий и четвертый признаки от метки стрептавидина, упоминаемые позже как сайты "5ММ" или "10ММ", соответственно), хотя и со сниженным сродством. Сайты, содержащие большее количество ошибочных спариваний (второй и пятый признаки) или имеющие два РАМ-проксимальных ошибочно спаренных нуклеотида (шестой признак), связываются со значительно более низкими скоростями. (ниже) Распределение сайтов РАМ ('TGG') в каждом из субстратов.

Структурно Cas9 из *S. pyogenes* представляет собой мономерный белок размером 160 кДа, приблизительно 10 нм × 10 нм × 5 нм (согласно кристаллическим структурам), примерно разделенный на две блок-подобные половины, каждая из которых содержит нуклеазный домен. В соответствии с рентгеновскими структурами dCas9 - sgRNA, визуализированные посредством АФМ, выглядят как крупные овальные структуры (фиг. 10А-10С), после инкубирования Cas9 или dCas9 с ДНК эти структуры, связанные вдоль ДНК, наблюдали и устанавливали как Cas9 или dCas9 соответственно (фиг. 1В, 10А-10С и 11А-11D). Чтобы однозначно определить последовательность сайтов, связанных Cas9 и dCas9, молекулы биотинилированной ДНК метили с одного конца с помощью моновалентной стрептавидиновой метки до АФМ-визуализации. Молекулы ДНК, которые наблюдали со связанными белками Cas9 или dCas9, отбирали для дальнейшего анализа и визуализировали с экстрапаразизирующей способностью в соответствии с модифицированным протоколом, адаптированным из протокола Wiggins et al. (25), и сайты, связанные с Cas9/dCas9, извлекали (подробности см. в разделе Дополнительные способы).

Этот способ оказался чрезвычайно устойчивым (табл. 1): на ДНК, связанной Cas9 или dCas9, наблюдается отчетливое накопление белков, сконцентрированных именно в месте расположения сайтов протоспейсера с прилегающим РАМ (в пределах ожидаемых 23 п.о., фиг. 1С-Д), и проявляется в виде острых пиков. Никаких таких очевидных пиков не наблюдается в ДНК-субстрате, не содержащем целевых сайтов (фиг. 9А-9С). Стандартное отклонение ширины пиков составляло от 36 до 60 п.о., что является значительным улучшением по сравнению с экспериментами по связыванию с использованием одномолекулярной флуоресценции, которые приводят к стандартным отклонениям ширины пика σ приблизительно 1000 п.о.). Средняя кажущаяся занимаемая Cas9/dCas9 площадь на ДНК покрывает 78,1 п.о. $\pm 37,9$ п.о.; это расширение кажущейся занимаемой площади на ~ 20 п.о. занимаемой Cas9 площади на ДНК, определяемое биохимическими и кристаллографическими способами, является хорошо известным результатом свертывания изображений с шириной острия АФМ. Ранее *in vitro* наблюдалось, что Cas9 остается связанным с целевой ДНК в течение длительных периодов времени (> 10 мин) после предполагаемого расщепления ДНК в качестве однооборотной эндонуклеазы и не может быть вытеснена из расщепленных нитей без жесткой химической обработки. Большинство молекул ДНК, наблюдаемых со связанным Cas9, выглядят как полученные из полноразмерного AAVS1 субстраты с очень небольшим ($\sim 5\%$) процентом субстратов, которые были расщеплены и разделены. После отслеживания этих молекул ДНК, наблюдали, что Cas9 связывается с этими "полноразмерными" субстратами с почти одинаковым распределением, как и dCas9 (двухсторонний тест Колмогорова-Смирнова, уровень значимости 5%) (фиг. 1С).

Таблица 1. Пики, зарегистрированные в гистограммах связывания на фиг. 1C-D для Cas9/dCas9-sgRNA и фиг. 2C для dCas9 с sgRNA, имеющими усечение 2 нуклеотидов на 5'-конце (tru-gRNA), на основании эмпирической подгонки к функции Гаусса $\alpha \exp(-(x-\mu)/w)^2$

Направляющая РНК:	sgRNA ^a			tru-gRNA ^b			sgRNA	
	Сконструированная ДНК:			Сконструированная ДНК:			ДНК, полученная из AAVs1:	
Общее количество визуализированных молекул ДНК: ^c	n=536			n=257			n=404	n=250
Имя местоположения:	Полный сайт	Сайт 10ММ ^d	Сайт 5ММ ^e	Полный сайт	Сайт 10ММ ^f	Сайт 5ММ ^g	Полный сайт	Полный сайт
Cas9/dCas9	dCas9	dCas9	dCas9	dCas9	dCas9	dCas9	dCas9	Cas9
Местоположение: ^h	144-167	452-465	592-610	144-167	452-465	592-610	316-339	316-339
μ^i пика (95% дов. инт.):	151,3 (151,1 , 151,6)	467,6 (466,6 , 468,5)	600,6 (599,5, 601,7)	159,0 (158, 2, 159,7)	462,9 (462,1, 463,6)	592,0 (590,9, 593,0)	327,7 (327, 3, 328,2)	315,0 (314, 4, 315, 7)
Ширина пика ^j $w=\sqrt{2\sigma}$ (95% дов. инт.):	51,46 (51,53 , 52,38)	57,5 (55,84 , 59,16)	70,8 (68,72, 72,89)	53,98 (52,2 , 55,76)	54,44 (52,07, 56,81)	67,88 (64,27, 71,49)	84,10 (83,1 2, 85,27)	58,7 (56, 8, 60,6 3)
# dCas9 ^k :	287	180,5	211,9	84,5	58,75	74,33		
#/(2w) (пересчитывается до плотности на полном сайте, 95% конф. инт.):	1	0,5688	0,5399	1	0,6894	0,6994		

^a Стандартная единая направляющая РНК (sgRNA)

^b Единая направляющая РНК с 2 нуклеотидами, усеченными с 5'-конца (tru-gRNA)

^c Количество молекул ДНК, наблюдаемых как с моновалентной стрептавидиновой меткой, так и со связанным белком, которые затем визуализировали (подробнее см. раздел Вспомогательные способы)

^d Целевой сайт с 10 PAM-дистальными ошибочно спаренными нуклеотидами

^e Целевой сайт с 5 PAM-дистальными ошибочно спаренными нуклеотидами

^f Предполагается, что на сконструированном ДНК-субстрате tru-gRNA будет взаимодействовать только с первыми 8 из 10 PAM-дистальных ошибочно спаренных нуклеотидов на сайте 10ММ

^g Предполагается, что на сконструированном ДНК-субстрате tru-gRNA будет взаимодействовать только с первыми 3 из 5 PAM-дистальных ошибочно спаренных нуклеотидов на сайте 5ММ

^h п.о. с меченого стрептавидином конца (от PAM до конца сайта)

ⁱ Максимум пика в гистограмме связывания (из гауссовского распределения)

^j Ширина пика - $\sqrt{2}\sigma$, σ - стандартное отклонение

^k Количество молекул dCas9, наблюдаемых в пределах ширины 1 пика ($\sqrt{2}\sigma$) сайта связывания. Если Cas9/dCas9 вступает в контакт с ДНК в n сайтах, такую молекулу утяжеляют на 1/n. Если молекула перекрывает одновременно сайты 10ММ и 5ММ, # утяжеляли дополнительно на 1/2.

Изучая время пребывания dCas9 связанным с различными местоположениями вдоль сконструиро-

ванного субстрата, можно определить относительную предрасположенность связывания dCas9 с различными ошибочно спаренными и частичными целевыми сайтами (фиг. 1D, табл. 1). Общую константу диссоциации между dCas9 и полным ДНК-субстратом оценили как 2,70 нМ ($\pm 1,58$ нМ, 95% доверительный интервал, табл. 2). Константа диссоциации dCas9, особенно в сайте полного (идеально совпадающего) протоспейсера (в пределах ширины одного пика в гистограмме связывания), расположенного в субстрате, составляет 44,67 нМ ($\pm 1,04$ нМ, 95% доверительный интервал). В ранее проведенных анализах электрофоретической подвижности (EMSA) оценили связывание dCas9-sgRNA с сайтами протоспейсера на коротких молекулах ДНК (~ 50 п.о.), составляющее от 0,5 нМ до 2 нМ. В то время как увеличение константы диссоциации в наблюдаемых сайтах протоспейсера может быть связано с наличием множественных нецелевых сайтов на сконструированном ДНК-субстрате, типично, что константы диссоциации, определенные с помощью AFM, почти на порядок выше, чем те, которые определены в традиционных анализах (26). Эта разница часто объясняется неспецифическими взаимодействиями белков с тупыми концами более короткой ДНК, которые не учитываются в EMSA.

Таблица 2. Кажущиеся константы диссоциации для dCas9 с различными вариантами направляющей РНК из 'сконструированных' ДНК-субстратов размером 989 п.о. (например, фиг. 1D, 2C и 2D), которые содержат ряд полностью или частично комплементарных сайтов протоспейсера

Вариант направляющей РНК	Общая константа диссоциации между dCas9 и сконструированным ДНК-субстратом (\pm 95% доверительный интервал)	Протоспейсер-специфическая константа диссоциации для dCas9 и полной мишени на сконструированном субстрате (\pm 95% доверительный интервал)
sgRNA ^a	2,70 нМ (\pm 1,58 нМ)	44,67 нМ (\pm 1,04 нМ)
tru-gRNA ^b	17,89 нМ (\pm 0,45 нМ)	136,4 нМ (\pm 2,30 нМ)
hp6-gRNA ^c	16,61 нМ (\pm 0,40 нМ)	164,4 нМ (\pm 13,63 нМ)
hp10-gRNA ^d	35,84 нМ (\pm 0,63 нМ)	164,8 нМ (\pm 15,60 нМ)

^a Полноразмерная единая направляющая РНК (sgRNA)

^b Усеченная sgRNA (первые два нуклеотида на 5'-конце усечены)

^c sgRNA с дополнительной 5'-шпилькой, которая перекрывает шесть PAM-дистальных нацеливающихся нуклеотидов (см. текст)

^d sgRNA с дополнительной 5'-шпилькой, которая перекрывает десять PAM-дистальных нацеливающихся нуклеотидов (см. текст)

На сконструированном субстрате dCas9 относительно устойчива к дистальным ошибочным спариваниям (проявляя 50-60% предрасположенность к связыванию по сравнению с полным целевым сайтом, фиг. 1D и табл. 1) и имеет такое же кажущееся сродство (в пределах доверительного интервала) к целевым сайтам, содержащим 5 и 10 дистальных ошибочных спариваний (MM). Однако связывание с сайтами протоспейсера, содержащими только два прилегающих к PAM ошибочных спаривания, происходило с аналогичной предрасположенностью, как с сайтами с 15 или даже 20 (только сайт PAM) дистальными ошибочными спариваниями (приблизительно 5-10% предрасположенность к связыванию по сравнению с идеальной мишенью, приблизительно такая, как фоновый сигнал связывания), открытие, согласующееся с предыдущими биохимическими исследованиями. Хотя на сконструированном субстрате нет сайтов PAM, за исключением прилегающих к протоспейсеру сайтов, на полученном из AAVS1 субстрате имеется отчетливый "плечевой пик" повышенного связывания Cas9 и dCas9 вблизи мишени AAVS1, которая особенно обогащена сайтами PAM. На "нонсенс"-субстрате и сегментах полученного из AAVS1 субстрата вдали от целевых сайтов едва заметные накопления dCas9 тесно отражали распределение сайтов PAM (двухсторонний тест Колмогорова-Смирнова, уровень значимости 5%) и распределение dCas9 на "нонсенс"-субстрате более точно отражало экспериментальное распределение PAM, чем для 71,20% из 100 000 случайно генерируемых последовательностей с теми же распределениями dA, dT, dC и dG (фиг. 9A-9C). Поскольку связывание dCas9 вдоль 'нонсенс'-субстрата (с 879 сайтами PAM в 1079 п.о.) так хорошо соответствует распределению сайтов PAM, это интерпретировали как измерение реальных взаимодействий dCas9-PAM. По оценкам, средняя константа моносайтовой диссоциации для связывания dCas9 вдоль "неспецифического" субстрата составляла приблизительно 867 нМ (стандартное отклонение ± 209 нМ). Это можно понять как оценку константы диссоциации связи dCas9 на ДНК без гомологии протоспейсера.

Пример 3. sgRNA с усечением двух нуклеотидов на 5'-концах (tru-gRNA) не увеличивают специфичность связывания dCas9 in vitro

Было обнаружено, что Cas9 по-прежнему проявляет расщепляющую активность, даже если до че-

тырех нуклеотидов направляющего (нацеливающегося на протоспейсер) сегмента sgRNA или crRNA усекали с 5'-концов и Fu et al. (21) недавно показали, что применение sgRNA с этими 5'-усечениями (оптимально на 2-3 нуклеотида) может фактически приводить к увеличению на порядок точности расщепления Cas9 *in vivo*. Это может свидетельствовать о том, что повышенная чувствительность к сайтам ошибочного спаривания (ММ) с использованием этих усеченных sgRNA (называемых "tru-gRNA", фиг. 2А) была результатом их уменьшенной энергии связи между направляющей РНК и сайтами протоспейсера. Это означает, что энергия связи, придаваемая дополнительными 5'-нуклеотидами на sgRNA, может компенсировать любые ошибочно спаренные нуклеотиды и стабилизировать Cas9 в некорректных сайтах, тогда как tru-gRNA будут относительно менее стабильными на ДНК, если есть ошибочные спаривания.

В качестве теста этого предложенного механизма dCas9 визуализировали с tru-gRNA с двухнуклеотидным 5'-усечением относительно ранее использованной sgRNA. Комплексы dCas9-tru-gRNA инкубировали с сконструированными субстратами, которые содержали ряд полных и частичных сайтов протоспейсера. Снова обнаружили отчетливый пик точно в полном сайте протоспейсера (фиг. 2С и табл. 1), хотя кажущаяся константа ассоциации по отношению к dCas9 с полной sgRNA в этом сайте значительно уменьшается (т.е., константа диссоциации увеличивается, см. табл. 2). Однако по сравнению со связыванием в полных сайтах протоспейсера нецелевое связывание dCas9 с tru-gRNA на сайтах протоспейсера с РАМ-дистальными ошибочными спариваниями фактически увеличивается по сравнению с dCas9 с sgRNA (фиг. 2С и табл. 1). Подобно dCas9 с sgRNA, dCas9 с tru-gRNA связывается с протоспейсером с 10 или 5 РАМ-дистальными сайтами ошибочного спаривания с приблизительно равными предрасположенностями (следует отметить, что только предполагается, что tru-gRNA будет взаимодействовать с первыми 8 и 3 ошибочными спариваниями в этих сайтах, соответственно). Эти результаты свидетельствуют о том, что повышенная точность расщепления с использованием tru-gRNA не обязательно обеспечивается при относительном снижении предрасположенности связывания в нецелевых сайтах или снижении относительной стабильности при наличии ошибочных спариваний. Скорее всего, хотя могут быть некоторые "пороговые" эффекты, когда уменьшение константы ассоциации ниже $\sim 4\text{-}5 \times 10^6$ М эффективно устраняет расщепляющую активность *in vivo*, эти и дополнительные результаты, представленные ниже, показывают, что повышенная специфичность, проявляемая tru-gRNA, может быть вызвана различием в самом механизме расщепления. Более того, эти данные свидетельствуют о том, что, хотя tru-gRNA могут улучшить специфичность расщепления у активного Cas9, они могут не улучшать специфичность их связывающей активности для применений с привлечением dCas9 (или химерных производных) *in vivo*.

Кроме того, в предыдущих сообщениях было показано, что tru-gRNA, которые имеют 5'-усечения (оптимально на 2-3 нуклеотида) в нацеливаемом на протоспейсер сегменте, могут приводить к увеличению точности расщепления Cas9 *in vivo* на порядок (фиг. 2А), при этом результаты, показанные в разделе Примеры, указывают на то, что усеченные gRNA не улучшают специфичности связывания dCas9 (фиг. 2С). На фиг. 2С показано сродство связывания dCas9 со стандартной gRNA (пунктирная линия) по сравнению со сродством связывания dCas9 с tru-gRNA (tru-gRNA, фиолетовая линия) на молекуле ДНК, которая содержит полный протоспейсер (сайт i), а также сайты протоспейсеров с 5 и 10 РАМ-дистальными ошибочными спариваниями (сайты ii и iii соответственно). На фиг. 2С показано, что стандартные направляющие РНК сохраняют значительную способность связываться с этими нецелевыми сайтами (содержащими ошибочные спаривания) и что tru-gRNA не проявляют относительного усиления специфичности связывания в сайтах, которые содержат ошибочные спаривания в 5 и 10 нуклеотидах на РАМ-дистальном конце протоспейсера. Распределение связывания dCas9 с tru-gRNA демонстрирует отчетливые пики ее сродства точно в сайтах протоспейсеров с 10 РАМ-дистальными ошибочными спариваниями и 5 РАМ-дистальными ошибочными спариваниями, демонстрируя, что она не обладает повышенной специфичностью связывания по сравнению с полными sgRNA (см. табл. 1). "Пики" в гистограмме связывания указывают на специфическое стабильное связывание в этих нецелевых сайтах. Фактически, связывание dCas9-tru-gRNA в нецелевых сайтах в действительности увеличивается относительно связывания с протоспейсером со стандартной направляющей РНК. Это неизбежное связывание может ограничить их полезность для dCas9 и химерных производных dCas9. Это может также отражать отмеченное для данной системы нецелевое расщепление, которое, хотя и улучшено по сравнению со стандартными направляющими РНК, по-прежнему значимо в некоторых нецелевых сайтах. Для сравнения, не было обнаружено никакого специфического связывания hpgRNA в этих сайтах с ошибочными спариваниями (Фиг. 2D). hpgRNA связаны в этих сайтах с примерно одинаковым сродством, так как они неспецифически связаны с ДНК без гомологии с протоспейсером с уменьшением на $\sim 22\%$ максимального наблюдаемого сродства нецелевого связывания относительно усеченных gRNA. Кроме того, исходя из узкой геометрии канала связывания ДНК в Cas9, мы ожидаем, что наличие закрытой шпильки в ошибочно спаренных протоспейсерах может ингибировать конформационное изменение в Cas9, необходимое для выполнения расщепления (фиг. 1В).

Значительные усилия были предприняты для того, чтобы охарактеризовать эту нецелевую активность и улучшить специфичность Cas9/dCas9 посредством разумного выбора целевых последовательностей протоспейсера; оптимизации структуры sgRNA, например, путем усечения первых двух 5'-

нуклеотидов в sgRNA; и применения "вносящих два однонитевых разрыва" ферментов Cas9, но четкое понимание точного механизма РНК-направляемого расщепления, поскольку оно относится к структурной биологии Cas9, будет иметь важное значение для разработки производных Cas9 и направляющих РНК с повышенной точностью для их новых приложений в медицине и биологии.

В соответствии с этой целью в данном случае применяют атомно-силовую микроскопию (AFM) для анализа отдельных белков Cas9 и dCas9 *S. pyogenes*, поскольку они связываются с мишенями вдоль сконструированных ДНК-субстратов после инкубации с различными вариантами sgRNA. Этот метод позволяет непосредственно анализировать одновременно сайт связывания и структуру отдельных белков Cas9/dCas9, предоставляя большой объем механистической информации относительно специфичности Cas9/dCas9 с одномолекулярным разрешением. В соответствии с традиционными биохимическими исследованиями мы обнаружили, что значительное связывание Cas9/dCas9 с sgRNA происходит в сайтах, содержащих до 10 ошибочно спаренных пар оснований в целевой последовательности. Однако, хотя применение направляющих РНК с двумя нуклеотидами, усеченными с 5'-конца (tru-gRNA), как ранее показано *in vivo*, приводило к снижению нецелевого мутагенеза с помощью Cas9 вплоть до 5000 раз, *in vitro* обнаружили сходные особенности связывания dCas9 и tru-gRNA с ошибочно спаренными мишенями, как для стандартной sgRNA. Обнаружено, что добавление шпильки к 5'-концу sgRNA, которая частично перекрывает область связывания с мишенью в направляющей РНК, увеличивает специфичность dCas9 за счет общего снижения предрасположенности связывания с ДНК. Результаты авторов настоящего изобретения показывают, что общая стабильность связывания направляющей РНК-ДНК не обязательно определяет специфичность расщепления Cas9, когда ошибочные спаривания расположены на расстоянии более 10 п.о. от PAM.

Пример 4.

Направляющие РНК с 5'-шпильками, комплементарными "PAM-дистально" нацеливающимся сегментам (hp-gRNA), модулируют абсолютную связывающую предрасположенность и профиль dCas9, связанных с ДНК с ошибочно спаренными протоспейсерами *In vitro*

Специфичность dCas9 можно увеличить путем удлинения 5'-конца sgRNA таким образом, чтобы она образовала структуру шпильки, которая перекрывает "PAM-дистально" нацеливающийся (или "незатравочный") сегмент sgRNA (фиг. 2B). После того, как сайт PAM связан и начато встраивание направляющей РНК в ДНК, шпилька открывается при связывании с полным протоспейсером и может произойти полное встраивание нити. Если в целевом сайте имеются PAM-дистальные ошибочные спаривания, то энергетически более выгодно, чтобы шпилька оставалась закрытой и встраивание нити было затруднено. В последнее время аналогичные топологии использовались для "динамических циклов ДНК", которые вызываются встраиванием нити. В этих системах шпильки служат кинетическими барьерами для встраивания, причем скорости встраивания олигонуклеотидов замедлялись на несколько порядков в случаях попытки встраивания в мишени с ошибочными спариваниями. Шпильки здесь могут вытесняться во время встраивания в полные целевые сайты, но ингибируют встраивание, если имеются ошибочные спаривания между мишенью и незатравочной нацеливающейся областью направляющей РНК (Фиг. 2B). В этих случаях энергетически более выгодно, чтобы шпильки оставались закрытыми. Несмотря на предыдущие усилия, которые добавили 5'-удлинения к sgRNA, чтобы обеспечить комплементарность дополнительным нуклеотидам за протоспейсером, эти направляющие РНК не показали повышенной специфичности расщепления Cas9 *in vivo*. Скорее всего они расщеплялись приблизительно до своей стандартной длины в живых клетках. Основываясь на размере и структуре шпильки, шпилька может быть размещена в канале связывания ДНК молекулы Cas9/dCas9 и защищена от разрушения.

sgRNA получали с 5'-шпильками (hp-gRNA), которые перекрывали нуклеотиды, комплементарные последним шести (hp6-gRNA) или десяти (hp10-gRNA) PAM-дистальным сайтам протоспейсера. При сопоставлении наблюдаемых мест связывания dCas9-hp-gRNA на сконструированном ДНК-субстрате (фиг. 2D) острые пики наблюдали точно в сайте протоспейсера (PAM и протоспейсер расположены в сайтах 144-167, с пиком связывания в сайте 154,0 (95% доверительный интервал: 153,3-154,8) для dCas9-hp6-gRNA и в 158,3 (95% доверительный интервал: 157,6-158,9) для dCas9-hp10-gRNA). Специфические пики в сайтах с 5 и 10 дистальными ошибочными спариваниями значительно сплюснены, причем dCas9 и hp10-gRNA проявляют значительно уменьшенное сродство к нецелевым сайтам (снижение на 22% относительно dCas9 с tru-gRNA). Пики сродства на полных сайтах протоспейсера подразумевают, что шпильки действительно открыты при полном встраивании. n=243 для hp6-gRNA и n=212 для hp10-gRNA. dCas9 с hp-gRNA показывают сходное снижение сродства к целевому сайту, как и с tru-gRNA, однако в отличие от dCas9 с tru-gRNA, dCas9 с hp-gRNA не демонстрирует каких-либо острых пиков связывания в нецелевых сайтах, что в противном случае указывало бы на сильное специфическое связывание. С hp6-gRNA наблюдали накопление связывания вокруг сайтов протоспейсеров с 5 или 10 ошибочно спаренными PAM-дистальными сайтами. Поскольку у них отсутствуют острые пики связывания, наблюдаемые с sgRNA и tru-gRNA, эти накопления вряд ли указывают на специфическое связывание, а скорее могут указывать на то, что dCas9 диссоциировался из этих сайтов при адсорбции на поверхности. Это означало бы очень слабое связывание в этих нецелевых сайтах в случае hp6-gRNA.

В случае hp10-gRNA связывание с этими ошибочно спаренными сайтами находится приблизительно-

но на уровне неспецифического связывания в любом другом месте на субстрате, что представляет собой уменьшение на 22% максимального наблюдаемого сродства нецелевого связывания по сравнению с tr-gRNA (уменьшение максимальной наблюдаемой константы ассоциации с $3,18 \times 10^6$ М до $2,48 \times 10^6$ М, фиг. 2D). Это увеличение специфичности hr10-gRNA также отражается в сходной константе диссоциации связи, как у hr6-gRNA, с сайтами протоспейсера, но в значительном увеличении общей константы диссоциации ко всему (специфическому+неспецифическому) сконструированному субстрату (табл. 2).

Отчетливое накопление именно в полных сайтах протоспейсера предполагает, что при встраивании в полные сайты протоспейсера шпильки в hr-gRNA фактически открываются, так как в противном случае нуклеотиды, которые связывают PAM-дистальные сайты протоспейсера, были бы захвачены в шпильке. Вероятным механизмом для улучшения специфичности связывания является то, что при отсутствии открытия в сайтах протоспейсеров с PAM-дистальными ошибочными спариваниями присутствие шпильки способствует выплавлению направляющей РНК из этих нецелевых сайтов. Результаты показывают, что hr-gRNA можно использовать для настройки сродства и специфичности связывания Cas9/dCas9, а дополнительная манипуляция с длиной шпильки, длиной петли и композицией петли может обеспечить возможность более точного контроля этих свойств.

Пример 5.

Cas9 и dCas9 подвергаются прогрессивному структурному переходу, поскольку они связываются с сайтами ДНК, которые все в большей степени соответствуют целевой последовательности протоспейсера.

С использованием просвечивающей электронной микроскопии с негативным окрашиванием (ТЕМ) было обнаружено, что при связывании sgRNA структура dCas9 уплотняется и поворачивается, чтобы открыть предполагаемый канал связывания ДНК между двумя его блоками. После связывания с ДНК, содержащей последовательность PAM и протоспейсера, dCas9 подвергается второй структурной переориентации с расширением конформации. Предполагалось, что роль этого второго перехода связана с встраиванием нити sgRNA или выравниванием двух основных нуклеазных сайтов Cas9 с двумя разделенными нитями ДНК. Однако эти исследования проводили только в присутствии или отсутствии ДНК, содержащей полностью совпадающие последовательности протоспейсеров, и изучение перехода между этими конформациями в частично совпадающих сайтах протоспейсеров может дать представление о механизме нецелевого связывания и расщепления. Поэтому помимо определения относительной предрасположенности связывания использовали АФМ-визуализацию для фиксации этих предполагаемых конформационных переходов Cas9 и dCas9, поскольку они связаны с ДНК в сайтах с различной комплементарностью с протоспейсером. Извлекали объемы и максимальные топографические высоты белков Cas9 и dCas9 с sgRNA, которые находились отдельно на ДНК ($n=839$), и сопоставляли эти значения с их соответствующими сайтами связывания на ДНК (фиг. 3, фиг. 11A-11D и фиг. 12A-12B). Распределение сайтов связывания почти идентично распределению полного набора данных, указывая на то, что этот отбор был объективным и репрезентативным. Записанное изображение каждого из этих белков извлекали (фиг. 11C-11D) и попарно выравнивали с помощью итерационного вращения, отражения и перевода. Структуры белка группировали в соответствии с их попарным топографическим среднеквадратическим отклонением (фиг. 12A-12B и табл. 3). Явное преимущество этого метода заключается в том, что он естественным образом группирует любой моновалентный стрептавидин или любые агрегированные белки Cas9/dCas9, которые совместно локализируются на поверхности с ДНК, отдельно от тех, которые установлены как индивидуальные молекулы Cas9/dCas9, что позволяет провести объективный анализ структурных свойств этих белков на ДНК. Анализ распределения сайтов связывания с предполагаемыми молекулами стрептавидина либо агрегированными белками показывает, что они являются редкими и равномерно распределенными вдоль ДНК и, следовательно, не мешают анализу распределения сайтов связывания (фиг. 12A-12B).

В сайтах, не содержащих гомологии к мишеням, например, на "нонсенс"-ДНК-субстрате, молекулы dCas9 с sgRNA были преимущественно меньшего размера и яйцевидными (фиг. 3C (iii) и табл. 3). Но поскольку белки dCas9 связываются с целевыми последовательностями со возрастающей комплементарностью (фиг. 3 (α - δ)), их высота и объем значительно увеличиваются (фиг. 3D и 12A-12B, табл. 2) относительно неспецифического связывания, достигая максимального размера в последовательности протоспейсера. Это увеличение также сопровождается сдвигом в популяции dCas9 (фиг. 3A и 12A-12B, табл. 2) из сгруппированных структур с более плоскими и яйцевидными конформациями (фиг. 3C (ii) и C3 (iii), синий и зеленый) к тем, которые все чаще группируются с более округлыми структурами, имеющими крупный центральный выступ (фиг. 3C (i), желтый цвет). Эта последняя наблюдаемая конформация, вероятно, представляет собой расширенную конформацию, ранее наблюдаемую посредством ТЕМ, а в последнее время и с помощью эксклюзионной хроматографии, и, по-видимому, является активным состоянием, где нуклеазные домены Cas9 расположены надлежащим образом вокруг ДНК, так что расщепление может происходить наиболее эффективно.

Каталитически активный Cas9 подвергается значительному увеличению размера, поскольку он также связывается с протоспейсерной последовательностью (фиг. 3(ϵ)); однако имеется небольшое, но

статистически значимое уменьшение размера по сравнению с dCas9, и конформация Cas9 в полных сайтах протоспейсера имеет тенденцию группироваться с более плоскими (зелеными) структурами. Поскольку одновременно не контролировали, была ли расщеплена ДНК во время визуализации, неясно, является ли это еще одним конформационным изменением после расщепления ДНК, или является результатом мутационных различий между Cas9 и dCas9; однако, поскольку связывание и встраивание нити ранее были определены как ограничивающие скорость этапы, вероятно, что ДНК внутри Cas9 расщепляется во время этих измерений.

Таблица 3. Свойства dCas9/Cas9 с различными вариантами направляющей РНК в полностью и частично комплементарных и не комплементарных сайтах протоспейсера

Сайт	ДНК	Направляющая РНК	n ^a	Средний объем (нм ³ × 10 ⁴) ± SEM ^b	Средняя высота (нм) ± SEM
Протоспейсер (dCas9)	Сконструированная+AAVsI ^c	sgRNA	201 Y: 41% (± 6,8%) G: 22% (± 5,8%) B: 21% (± 5,6%)	0,6226 ± 0,016	1,932 ± 0,041
Протоспейсер	AAVsI	sgRNA	65 Y: 17% (±	0,5784 ± 0,035	1,753 ± 0,076

(Cas9)			9,1%) G: 32% (± 11,4%) B: 26% (± 10,7%)		
10 ММ (dCas9)	Сконстру ированна я	sgRNA	76 Y: 25% (± 8,9%) G: 31% (± 9,7%) B: 34% (± 9,9%)	0,5510 ± 0,011	1,601 ± 0,026
5 ММ (dCas9)	Сконстру ированна я	sgRNA	85 Y: 34% (± 8,8%) G: 34% (± 8,8%) B: 25% (± 8,1%)	0,6055 ± 0,024	1,790 ± 0,049
Неспециф ический (Cas9+dC as9)	AAVsI+но нсенс ^c	sgRNA	274 Y: 21% (± 4,8%) G: 17% (± 4,5%) B: 39% (± 5,8%)	0,4780 ± 0,015	1,553 ± 0,034
Протоспе йсер (dCas9) д	Сконстру ированна я	tru- gRNA ^g	47 Y: 26% (± 12,5%) G: 17% (± 10,7%) B: 34% (± 13,6%)	0,5421 ± 0,041	1,761 ± 0,079

(10 ММ) (dCas9) ^d ,e	Сконструированная	tru-gRNA	32 Y: 13% (±11,5%) G: 38% (±16,7%) B: 19% (±13,5%)	0,5123 ± 0,049	1,665 ± 0,099
(5 ММ) (dCas9) ^d ,f	Сконструированная	tru-gRNA	34 Y: 18% (±12,8%) G: 29% (±15,3%) B: 24% (±14,2%)	0,5346 ± 0,048	1,705 ± 0,084
Неспецифический (dCas9)	Сконструированная	tru-gRNA	72 Y: 14% (±8,0%) G: 17% (±8,6%) B: 29% (±10,5%)	0,4554 ± 0,035	1,532 ± 0,059
Протоспейсер (dCas9)	Сконструированная	hp6-gRNA ^g	47 Y: 26% (±12,5%) G: 17% (±10,7%) B: 34% (±13,6%)	0,5940 ± 0,043	1,860 ± 0,109
Неспецифический (dCas9)	Сконструированная	hp6-gRNA	32 Y: 13% (±11,5%) G: 38% (±16,7%) B: 19% (±13,5%)	0,4656 ± 0,024	1,572 ± 0,047
Протоспейсер (dCas9)	Сконструированная	hp10-gRNA ^g	47 Y: 26% (±12,5%) G: 17% (±10,7%) B: 34% (±13,6%)	0,6304 ± 0,038	1,837 ± 0,076
Неспецифический (dCas9)	Сконструированная	hp10-gRNA	32 Y: 13% (±11,5%) G: 38% (±16,7%) B: 19% (±13,5%)	0,5181 ± 0,027	1,644 ± 0,050

^a Общее количество молекул, наблюдаемых в пределах двух стандартных отклонений от этих сайтов. Ниже: доля популяции в основных трех структурных кластерах ($\pm 95\%$ доверительный интервал при биномиальном распределении), окрашенная как на фиг. 2 в основном тексте (Y=желтый кластер, G=зеленый кластер, B=светло-синий кластер). Полное распределение свойств по кластерам на фиг. 12A-12B.

^b Стандартная ошибка среднего

^c Стандартная ошибка среднего

^d Отклоненная нулевая гипотеза о распределении высота-объем отличается ($p > 0,05$; T^2 -критерий Хотеллинга)

^e Предполагается, что на сконструированном ДНК-субстрате tru-gRNA будет взаимодействовать только с первыми 8 из 10 РАМ-дистальных ошибочно спаренных нуклеотидов на сайте 10ММ (отмечены "8ММ" на фиг. 3D).

^f Предполагается, что на сконструированном ДНК-субстрате tru-gRNA будет взаимодействовать только с первыми 3 из 5 РАМ-дистальных ошибочно спаренных нуклеотидов на сайте 5ММ (отмечены "3ММ" на фиг. 3D).

^g См. дополнительный комментарий 1 в поддержку информации относительно коррекции высот и объемов белков с tru-gRNA и hp-gRNA, чтобы их можно было сравнить с таковыми с sgRNA.

Пример 6.

Взаимодействия между направляющей РНК и ДНК-мишенью внутри или вблизи 16^{го} сайта протоспейсера стабилизирует конформационное изменение Cas9/dCas9

AFM-визуализация непосредственно показывает, что хотя dCas9/Cas9 сохраняет значительную предрасположенность связывать сайты протоспейсера с наличием до десяти дистальных ошибочных спариваний, связывание с сайтами ДНК, которые все более комплементарны протоспейсеру, приводит к увеличению сдвига в популяции белков dCas9/Cas9 в сторону того, что похоже на активную конформацию. Примечательно, что наблюдается аналогичный сдвиг в структуре между нецелевыми сайтами и идеально соответствующими сайтами для dCas9 с hp-gRNA (табл. 2 и фиг. 13). Известно, что наличие комплементарных РАМ-дистальных последовательностей связано с повышенной стабильностью Cas9 на ДНК. Недавно было обнаружено, что связывание Cas9 с однонитевой ДНК с увеличением РАМ-дистальной комплементарности к протоспейсеру (от 10 до 20 сайтов) приводит к увеличению изменения размера белка. Это также было связано с переходом активности Cas9 от внесения однонитевого разрыва к полному расщеплению. Здесь мы можем непосредственно определить объемы Cas9/dCas9, связанной на сайтах двухнитевой ДНК. Анализ структурных свойств отдельных белков Cas9/dCas9 на двухнитевой ДНК обнаруживает устойчивый конформационный переход со все более соответствующими целевыми последовательностями, который согласуется с механизмом "конформационного отбора", где спаривание оснований sgRNA с этими дистальными сайтами также стабилизирует активную конформацию, так что может происходить эффективное расщепление, тогда как связывание с сайтами с многочисленными дистальными ошибочными спариваниями сдвигает равновесие от активной структуры (т.е. см. фиг. 4D).

По этому принципу мы видим, что данный эффект значительно подавлен для dCas9 с tru-gRNA (фиг. 3D и табл. 3) с меньшими сдвигами между структурными популяциями, в пределах которых группируются белки (фиг. 13). Кроме того, хотя мы видим статистическую разницу между свойствами высоты-объема dCas9-tru-gRNA, которые неспецифически связаны, и теми, которые связаны в полном или частичном сайтах протоспейсера ($p < 0,05$; T^2 -критерий Хотеллинга), в сайтах, которые все более соответствуют протоспейсеру (10ММ, 5ММ и полные сайты протоспейсера), их структурные свойства не являются статистически различимыми (фиг. 3D и табл. 3). Недавно возникло предположение, что, хотя встраивание в первые 10 п.о. протоспейсера инициирует конформационное изменение в Cas9, полное встраивание направляющей РНК в протоспейсер помогает привести к дальнейшему смещению в полностью активное состояние. Поэтому предполагали, что наблюдаемое подавление конформационного изменения во все более соответствующих сайтах протоспейсера для dCas9 с tru-gRNA (по сравнению с таковыми с sgRNA) было результатом снижения стабильности этих направляющих РНК в РАМ-дистальных сайтах.

Чтобы изучить относительную стабильность sgRNA и tru-gRNA в этих сайтах, проводили исследование согласно кинетическому методу Монте-Карло (КМС) динамической структуры R-петли, то есть структуры, образованной встраивающейся направляющей РНК, связанной с сегментом непрерывной ДНК, раскрывая однонитевую петлю комплементарной ДНК этого сегмента (фиг. 4A) - во время и после встраивания нити. Для более подробной информации см. раздел Дополнительные способы. Коротко, используя алгоритм Гиллеспи, моделировали встраивание нити направляющей РНК, связанной с сайтом протоспейсера m , в виде последовательной нуклеотид за нуклеотидом конкуренции между встраиванием (нарушение спаривания оснований между протоспейсером и комплементарной ему нитью ДНК, затем замена парами оснований протоспейсер-направляющая РНК) и повторным отжигом (обратимым), с зависимыми от последовательности скоростями встраивания и повторного отжига v_f и v_r , соответственно (фиг. 4A). В первом приближении аппроксимируют скорость перехода из состояния m в $m+1$, v_f , пропорционально $\exp(-(\Delta G^\circ(m+1)_{\text{РНК-ДНК}} - \Delta G^\circ(m+1)_{\text{ДНК-ДНК}})/2RT)$, где $\Delta G^\circ(m+1)_{\text{РНК-ДНК}}$ представляет собой свободную энергию спаривания оснований между РНК и протоспейсером в сайте $m+1$ и $\Delta G^\circ(m+1)_{\text{ДНК-ДНК}}$ представляет собой свободную энергию спаривания оснований между протоспейсером и комплементарной ему нитью ДНК в $m+1$ (R идеальная газовая постоянная, T - температура, а член $1/2$ включен для соответствия детальному равновесию). v_r аналогично оценивают пропорционально $\exp(-(\Delta G^\circ(m)_{\text{ДНК-ДНК}} - \Delta G^\circ(m)_{\text{РНК-ДНК}})/2RT)$.

$\Delta G^\circ(m)_{\text{РНК:ДНК}}/2RT$). Скорости перехода этого типа ранее использовали для вычислительных исследований нуклеотидного спаривания и стабильности, и здесь они позволили нам фиксировать общую динамику R-петли зависимым от последовательности образом.

В общем случае пары оснований РНК:ДНК энергетически сильнее, чем пары оснований ДНК:ДНК, и при равновесии по интегральным кривым КМС видно, что направляющие РНК стабильно связаны с протоспейсером, как и ожидалось (фиг. 4С). Тем не менее, хотя sgRNA довольно стабильна и остается почти полностью встроенной - в ходе 95% моделируемой динамики нить остается встроенной до 19^{го} сайта протоспейсера (фиг. 4В) - tru-gRNA проявляет значительные колебания повторного отжига протоспейсера в PAM-дистальных сайтах (фиг. 4В и 4С). Поскольку единственным различием между dCas9-sgRNA и dCas9-tru-gRNA является простое усечение двух 5'-нуклеотидов в направляющей РНК, и поскольку мы наблюдаем ингибирование конформационного изменения dCas9-sgRNA в сайтах, содержащих 5 PAM-дистальных ошибочных спариваний, эти результаты показывают, что конформационное изменение в полностью активном состоянии стабилизируется взаимодействием между направляющей РНК и протоспейсером вблизи 16^{го} сайта протоспейсера, которое нарушается при нестабильности tru-gRNA в этой области. Фактически эксперименты КМС показывают, что средняя продолжительность нахождения между полным встраиванием и повторным отжигом ДНК обратно к 16^{му} сайту уменьшается на два порядка при замене sgRNA на tru-gRNA (фиг. 4С, вставка). Этот результат согласуется с более ранним свидетельством того, что хотя активность Cas9 с вариантами tru-gRNA с усечениями 2 или 3 нуклеотидов (нт) модулировалась в зависимости от особенности последовательности, расщепление во всех тестируемых случаях резко снижалось на ~ 90% - 100% при усечении 4 нт и устранялось после усечения 5 нт. Конформационное изменение для состояния активации белка стабилизируется этими взаимодействиями на 16м сайте протоспейсера или вблизи него. Это свидетельство подтверждается стабильностью gRNA в 14^м - 17^м положениях протоспейсера, которую оценивали в дополнительных экспериментах КМС, описанных ниже, и которая коррелировала с экспериментальным нецелевым расщеплением *in vivo* (см. ниже), в то время как стабильности направляющей РНК в сайтах протоспейсера 18-20 не было.

Пример 7.

Флуктуации R-петли направляющей РНК-протоспейсера предлагают механизм устойчивости к ошибочному спариванию под влиянием Cas9/dCas9 и повышенной специфичности расщепления под влиянием tru-gRNA

Чтобы исследовать механизмы, с помощью которых Cas9 или dCas9 могут переносить ошибочные спаривания в протоспейсерах или приобретать чувствительность к ним, провели серию экспериментов КМС с использованием сайта протоспейсера AAVS1, куда ввели один или два PAM-дистальных (≥ 10 п.о. от PAM) ошибочных спаривания (фиг. 5). Cas9, как правило, более устойчив к PAM-дистальным ошибочным спариваниям, чем к PAM-проксимальным ошибочным спариваниям. Однако Hsu et al. (2013) *Nature Biotechnology*, 31, 827-832 выявили значительные и вариабельные различия в расчетных скоростях расщепления Cas9 в протоспейсерах, содержащих PAM-дистальные ошибочные спаривания, в зависимости от особенности последовательности, типа ошибочного спаривания и сайта ошибочного спаривания. Основываясь на наших экспериментах АФМ и более ранних экспериментах КМС, мы предположили, что различия в скорости расщепления также могут быть результатом разной стабильности направляющей РНК вблизи 16^{го} сайта протоспейсера. Для этого моделирования изучали только последовательности с парами протоспейсер-направляющая РНК, которые приводили бы к изолированным ошибочным спариваниям rG-dG, rC-dC, rA-dA и rU-dT, для которых зависящие от последовательности термодинамические данные являются наиболее полными и подходят для модели КМС. Не ожидается, что эффекты этих ошибочно спаренных оснований существенно снижают общую энергию связи между sgRNA и протоспейсером (табл. 4); например, одиночные rG-dG, rC-dC, rA-dA и rU-dT ошибочные спаривания снижают температуру плавления РНК:ДНК в среднем на 1,7°C. Скорее всего их эффект, как ожидается, будет кинетическим, а не термодинамическим по своей природе, препятствуя смещению нити при ошибочном спаривании. Поэтому инициировали эксперименты согласно кинетическому методу Монте-Карло, исходя из 10^{го} сайта протоспейсера (начальная длина R-петли $m=10$), такого как будет иметь место при встраивании нити.

Таблица 4. Последовательности и оценка максимального правдоподобия (MLE) частот разрезания по Hsu et al. (2013) Nature Biotechnology, 31, 827-832, применяемые для корреляционного анализа (сайт ошибочного спаривания в целевой последовательности выделен жирным шрифтом)

Целевая последовательность	SEQ ID NO	Нацеливающаяся на протоспейсер область направляющей РНК	SEQ ID NO	MLE частоты разрезания (Hsu et al. (2013))	Расчетная ΔG_{37}° (ккал/моль)
TTCTTCTTCTGCTCGG A CTC	13	GUGUCCGAGCAGAAGAAGAA	149	0,10384	-32,16
TTCTTCTTCTGCTCGG A CTC	14	GACUCCGAGCAGAAGAAGAA	150	0,12609	-31,4
TTCTTCTTCTGCTCGG A CTC	15	GAGACCGAGCAGAAGAAGAA	151	0,13145	-32,69
TTCTTCTTCTGCTCGG A CTC	16	GAGUGCGAGCAGAAGAAGAA	152	0,097464	-32,33
TTCTTCTTCTGCTCGG A CTC	17	GAGUCGGAGCAGAAGAAGAA	153	0,12704	-33,43
TTCTTCTTCTGCTCGG A CTC	18	GAGUCCCAGCAGAAGAAGAA	154	0,079556	-31,37
TTCTTCTTCTGCTCGG A CTC	19	GAGUCCGUGCAGAAGAAGAA	155	0,11197	-32,36
TTCTTCTTCTGCTCGG A CTC	20	GAGUCCGACCAGAAGAAGAA	156	0,04788	-31,9
TTCTTCTTCTGCTCGG A CTC	21	GAGUCCGAGGAGAAGAAGAA	157	0,085461	-32,83
TTCTTCTTCTGCTCGG A CTC	22	GAGUCCGAGCUGAAGAAGAA	158	0,074938	-32,22
TTCTTCTTCTGCTCGG A CTC	23	GUGUCCGAGCAGAAGAAGAA	159	0,15588	-32,16
TTCTTCTTCTGCTCGG A CTC	24	GACUCCGAGCAGAAGAAGAA	160	0,11015	-31,4
TTCTTCTTCTGCTCGG A CTC	25	GAGACCGAGCAGAAGAAGAA	161	0,11435	-32,69

TTCTTCTTCTGCTCGGACTC	26	GAGUGCGAGCAGAAGAAGAA	162	0,15072	-32,33
TTCTTCTTCTGCTCGGACTC	27	GAGUCGGAGCAGAAGAAGAA	163	0,11567	-33,43
TTCTTCTTCTGCTCGGACTC	28	GAGUCCCAGCAGAAGAAGAA	164	0,070181	-31,37
TTCTTCTTCTGCTCGGACTC	29	GAGUCCGUGCAGAAGAAGAA	165	0,10538	-32,36
TTCTTCTTCTGCTCGGACTC	30	GAGUCCGACCAGAAGAAGAA	166	0,064145	-31,9
TTCTTCTTCTGCTCGGACTC	31	GAGUCCGAGGAGAAGAAGAA	167	0,085148	-32,83
TTCTTCTTCTGCTCGGACTC	32	GAGUCCGAGCUGAAGAAGAA	168	0,064903	-32,22
CCCTAGTCATTGGAGGTGAC	33	GACACCUCCAUGACUAGGG	169	0,062949	-32,19
CCCTAGTCATTGGAGGTGAC	34	GUGACCUCCAUGACUAGGG	170	0,063313	-31,73
CCCTAGTCATTGGAGGTGAC	35	GUCUCCUCCAUGACUAGGG	171	0,068655	-31,72
CCCTAGTCATTGGAGGTGAC	36	GUCAGCUCCAUGACUAGGG	172	0,073003	-32
CCCTAGTCATTGGAGGTGAC	37	GUCACGUCCAUGACUAGGG	173	0,037401	-32,63
CCCTAGTCATTGGAGGTGAC	38	GUCACCACCAUGACUAGGG	174	0,038197	-32,11
CCCTAGTCATTGGAGGTGAC	39	GUCACCUGCAAUGACUAGGG	175	0,041758	-31,63
CCCTAGTCATTGGAGGTGAC	40	GUCACCUCGAAUGACUAGGG	176	0,067751	-32,23
CCCTAGTCATTGGAGGTGAC	41	GUCACCUCCAUGACUAGGG	177	0,031653	-31,62
CCCTAGTCATTGGAGGTGAC	42	GUCACCUCCAUGACUAGGG	178	0,027161	-31,77
ATGGGGAGGACATCGATGTC	43	GUCAUCGAUGUCCUCCCAU	179	0,027124	-31,26
ATGGGGAGGACATCGATGTC	44	GAGAUCGAUGUCCUCCCAU	180	0,022366	-31,7
ATGGGGAGGACATCGATGTC	45	GACUUCGAUGUCCUCCCAU	181	0,01127	-30,92
ATGGGGAGGACATCGATGTC	46	GACAACGAUGUCCUCCCAU	182	0,011836	-31,44
ATGGGGAGGACATCGATGTC	47	GACAUGGAUGUCCUCCCAU	183	0,009146	-31,83
ATGGGGAGGACATCGATGTC	48	GACAUCCAUGUCCUCCCAU	184	0,006333	-30,27
ATGGGGAGGACATCGATGTC	49	GACAUCGUUGUCCUCCCAU	185	0,006232	-31,06
ATGGGGAGGACATCGATGTC	50	GACAUCGAAGUCCUCCCAU	186	0,007085	-31,64
ATGGGGAGGACATCGATGTC	51	GACAUCGAUCUCCUCCCAU	187	0,001545	-30,32
ATGGGGAGGACATCGATGTC	52	GACAUCGAUGACUCCUCCCAU	188	0,00025	-31,59
ATCACATCAACCCGTGCCGC	53	GCGCCACCGGUUGAUGUGAU	189	0,15963	-35,23
ATCACATCAACCCGTGCCGC	54	GCCCCACCGGUUGAUGUGAU	190	0,14121	-32,17
ATCACATCAACCCGTGCCGC	55	GCGGCACCGGUUGAUGUGAU	191	0,18743	-33,43
ATCACATCAACCCGTGCCGC	56	GCGGCACCGGUUGAUGUGAU	192	0,1634	-33,63
ATCACATCAACCCGTGCCGC	57	GCGCCUCCGGUUGAUGUGAU	193	0,15877	-33,12
ATCACATCAACCCGTGCCGC	58	GCGCCAGCGGUUGAUGUGAU	194	0,029249	-33,4
ATCACATCAACCCGTGCCGC	59	GCGCCACGGUUGAUGUGAU	195	0,12208	-34,13
ATCACATCAACCCGTGCCGC	60	GCGCCACCGUUGAUGUGAU	196	0,051622	-31,57
ATCACATCAACCCGTGCCGC	61	GCGCCACCGUUGAUGUGAU	197	0,004914	-31,74
ATCACATCAACCCGTGCCGC	62	GCGCCACCGAUGAUGUGAU	198	0,032227	-33,79
GAGTTTCTCATCTGTGCCCC	63	GGGCCACAGAAGAAACUC	199	0,015879	-33,54
CCAGTTTCTGCCGTTGTAC	64	GUUCAACGGCAGAAGCUGG	200	0,037469	-33,17
CCAGTTTCTGCCGTTGTAC	65	GUACUAACGGCAGAAGCUGG	201	0,059921	-32,92
CCAGTTTCTGCCGTTGTAC	66	GUACAAACGGGAGAAGCUGG	202	0,032605	-33,43

046214

TTCTCTCCAGCTTCTGCC	67	GCCAGAAGCUGGAGGAGGAA	203	0,000481	-35,94
TTCTCTCCAGCTTCTGCC	68	GGCACAAGCUGGAGGAGGAA	204	0,041538	-37,4
TTCTCTCCAGCTTCTGCC	69	GGCAGAACCUGGAGGAGGAA	205	0,047874	-37,5
TTCTCTCCAGCTTCTGCC	70	GGCAGAAGCAGGAGGAGGAA	206	0,050381	-38,61
TTCTCTCCAGCTTCTGCC	71	GGCAGAAGCUCGAGGAGGAA	207	0,006459	-36,92
CCGTTGATGTGATGGCAGC	72	GCACCCAUCACAUAACCGG	208	0,03967	-33,31
CCGTTGATGTGATGGGAGC	73	GCUCCCUACACAUAACCGG	209	0,033426	-32,52
CCGTTGATGTGATGGGAGC	74	GCUCCCAACACAUAACCGG	210	0,035651	-33,04
CCGTTGATGTGATGGGAGC	75	GCUCCCAUCAGAUAACCGG	211	0,03209	-33,3
GCAGCAAGCAGCACTCTGCC	76	GGCAGUGUGCUGCUUGCUGC	212	0,004014	-32,46
GCAGCAAGCAGCACTCTGCC	77	GGCAGAGUGCAGCUUGCUGC	213	0,000219	-33,11
GCTTGGGCCACGCAGGGGC	78	GCCCCAGCGUGGCCCAAGC	214	0,001487	-38,81
GCTTGGGCCACGCAGGGGC	79	GCCCCUGCCUGGCCCAAGC	215	0,003322	-36,77
GCTTCGTGGCAATGCCAC	80	GUGGCCAUUGCCACGAAGC	216	0,000463	-32,67
GCTTGGGCCACGCAGGGGC	81	GCCCCUGCGUCGCCCAAGC	217	0	-37,12
AAGCTGGACTCTGGCCACTC	82	GAGUGGCCUGAGUCCAGCUU	218	0,010169	-33,02
TTCTTCTCTGCTCGGACTC	83	GAGACCGAGCAGAAGAAGAA	219	0,084395	-32,69
TTCTTCTCTGCTCGGACTC	84	GAGUCCGAGGAGAAGAAGAA	220	0,051852	-32,83
TTCTTCTCTGCTCGGACTC	85	GAGUCCGAGCUGAAGAAGAA	221	0,050685	-32,22
GAGTTTCTCATCTGTGCC	86	GGGGCACAGUUGAGAAACUC	222	0,004503	-34,16
TTCTCTCCAGCTTCTGCC	87	GGCAGAAGGUGGAGGAGGAA	223	0,006035	-38,83
TTCTCTCCAGCTTCTGCC	88	GGCAGAAGCAGGAGGAGGAA	224	0,011364	-38,61
AGCAGAAGAAGAGGGCTCC	89	GGAGCCCUUGUUCUUCUGCU	225	0,007206	-29,83
AAGCTGGACTCTGGCCACTC	90	GAGUGGCCUGAGUCCAGCUU	226	0	-33,02
CCCTAGTCATTGGAGGTGAC	91	GACACCUCCAUAUGACUAGGG	227	0,053611	-32,19
CCCTAGTCATTGGAGGTGAC	92	GUGACCUCCAUAUGACUAGGG	228	0,05399	-31,73
CCCTAGTCATTGGAGGTGAC	93	GUCUCCUCCAUAUGACUAGGG	229	0,070404	-31,72
CCCTAGTCATTGGAGGTGAC	94	GUCAGCUCCAUAUGACUAGGG	230	0,067678	-32
CCCTAGTCATTGGAGGTGAC	95	GUCACGUCCAUAUGACUAGGG	231	0,03597	-32,63
CCCTAGTCATTGGAGGTGAC	96	GUCACCACCAUAUGACUAGGG	232	0,025207	-32,11
CCCTAGTCATTGGAGGTGAC	97	GUCACCUCCAUAUGACUAGGG	233	0,056019	-31,63
CCCTAGTCATTGGAGGTGAC	98	GUCACCUCCAUAUGACUAGGG	234	0,065347	-32,23
CCCTAGTCATTGGAGGTGAC	99	GUCACCUCCAUAUGACUAGGG	235	0,063769	-31,62
CCCTAGTCATTGGAGGTGAC	100	GUCACCUCCAUAUGACUAGGG	236	0,052644	-31,77
ATGGGGAGGACATCGATGTC	101	GUCAUCGAUGUCCUCCCAU	237	0,020295	-31,26
ATGGGGAGGACATCGATGTC	102	GAGAUCGAUGUCCUCCCAU	238	0,012126	-31,7
ATGGGGAGGACATCGATGTC	103	GACUUCGAUGUCCUCCCAU	239	0,007202	-30,92
ATGGGGAGGACATCGATGTC	104	GACAACGAUGUCCUCCCAU	240	0,010912	-31,44
ATGGGGAGGACATCGATGTC	105	GACAUGGAUGUCCUCCCAU	241	0,009292	-31,83
ATGGGGAGGACATCGATGTC	106	GACAUCGAUGUCCUCCCAU	242	0,006125	-30,27
ATGGGGAGGACATCGATGTC	107	GACAUCGUUGUCCUCCCAU	243	0,007805	-31,06

ATGGGGAGGAC A TCGATGTC	108	GACAUCGAAGUCCUCCCAU	244	0,010174	-31,64
ATGGGGAGGAC A TCGATGTC	109	GACAUCGAUCCUCCCAU	245	0,003595	-30,32
ATGGGGAGG A TCGATGTC	110	GACAUCGAUGACCUCCCAU	246	0,000206	-31,59
ATCACATCAACCGGTGG C GC	111	GGCCACCGGUUGAUGUGAU	247	0,18977	-35,23
ATCACATCAACCGGTGG C GC	112	GCCCCACCGGUUGAUGUGAU	248	0,13525	-32,17
ATCACATCAACCGGTGG C GC	113	GCGGCACCGGUUGAUGUGAU	249	0,14749	-33,43
ATCACATCAACCGGTGG C GC	114	GCGCGACCGGUUGAUGUGAU	250	0,13952	-33,63
ATCACATCAACCGGTGG C GC	115	GCGCCUCCGGUUGAUGUGAU	251	0,13949	-33,12
ATCACATCAACCG G TGGCGC	116	GCGCCAGCGGUUGAUGUGAU	252	0,031221	-33,4
ATCACATCAACCG G TGGCGC	117	GCGCCACCGGUUGAUGUGAU	253	0,14776	-34,13
ATCACATCAACCGGTGG C GC	118	GCGCCACCGGUUGAUGUGAU	254	0,050539	-31,57
ATCACATCAACCGGTGG C GC	119	GCGCCACCGCUUGAUGUGAU	255	0,003982	-31,74
ATCACATCAACCGGTGG C GC	120	GCGCCACCGGAUGAUGUGAU	256	0,015494	-33,79
GAGTTTCTCATCTGTG C CCC	121	GGCCACAGAUGAGAAACUC	257	0,025334	-33,54
CCAGCTTCTGCCGTTT G TAC	122	GUUCAACGGCAGAAGCUGG	258	0,062094	-33,17
CCAGCTTCTGCCGTTT G TAC	123	GUACUAACGGCAGAAGCUGG	259	0,080429	-32,92
CCAGCTTCT G CCGTTTGTAC	124	GUACAAACGGGAGAAGCUGG	260	0,032505	-33,43
TTCTCTCCAGCTTCT G CC	125	GCCAGAAGCUGGAGGAGGAA	261	0,00117	-35,94
TTCTCTCCAGCTTCT G CC	126	GGCACAAGCUGGAGGAGGAA	262	0,034381	-37,4
TTCTCTCCAG C TTCTGCC	127	GGCAGAACCUGGAGGAGGAA	263	0,059128	-37,5
TTCTCTCC A GCTTCTGCC	128	GGCAGAAGCAGGAGGAGGAA	264	0,05162	-38,61
TTCTCTCC A GCTTCTGCC	129	GGCAGAAGCUCGAGGAGGAA	265	0,007682	-36,92
CCGGTTGATGTGATGG A GC	130	GCACCAUCAUCAACCGG	266	0,093725	-33,31
CCGGTTGATGTGAT G GGAGC	131	GCUCCCUUCAUCAACCGG	267	0,075435	-32,52
CCGGTTGATGTG A TGGGAGC	132	GCUCCCAACAUCAACCGG	268	0,091723	-33,04
CCGGTTGAT G TGATGGGAGC	133	GCUCCCAUCAGAUCAACCGG	269	0,070319	-33,3
GCAGCAAGCAGCA T CTGCC	134	GGCAGUGUGCUGCUUGCUC	270	0,006754	-32,46
CCAGCAAG C ACTCTCTGCC	135	GGCAGAGUCGAGCUUGCUC	271	0,000545	-33,11
GCTTGGGCCACGC A GGGGC	136	GCCCCAGCUGGGCCCAAGC	272	0,004676	-38,81
GCTTGGGCC C AGGGGGC	137	GCCCCUGCCUGGGCCCAAGC	273	0,001918	-36,77
GCTTCGTGGCAAT G CCAC	138	GUGGCCCAUUGCCACGAAGC	274	0,001045	-32,67
GCTTGGGCC C ACGCAGGGGC	139	GCCCCUGCGUCGGCCCAAGC	275	0	-37,12
AAGCTGGACT T GGCCACTC	140	GAGUGGCCUGAGUCCAGCUU	276	0,008891	-33,02
TTCTTCTTCTGCTCG G ACTC	141	GAGACCGAGCAGAAGAAGAA	277	0,091861	-32,69
TTCTTCTTCT G CTCGGACTC	142	GAGUCCGAGGAGAAGAAGAA	278	0,062783	-32,83
TTCTTCTTCT T GCTCGGACTC	143	GAGUCCGAGCUGAAGAAGAA	279	0,044444	-32,22
GAGTTTCTCAT T GTGCCCC	144	GGGCACAGUUGAGAAACUC	280	0,0053	-34,16
TTCTCTCC A GCTTCTGCC	145	GGCAGAAGGUGGAGGAGGAA	281	0,00714	-38,83
TTCTCTCC A GCTTCTGCC	146	GGCAGAAGCAGGAGGAGGAA	282	0,019945	-38,61
AGCAGAAGA A GAGGGCTCC	147	GGAGCCUUGUUCUUCUCU	283	0,007996	-29,83
AAGCTGGACT T GGCCACTC	148	GAGUGGCCUGAGUCCAGCUU	284	0,006102	-33,02

Затем проводили эксперименты КМС для исследования кинетики встраивания нити в присутствии РАМ-дистальных ошибочных спариваний. Во всех случаях (по 1000 испытаний каждый) направляющие РНК остаются достаточно стабильно связанными даже тогда, когда имеются ошибочные спаривания (т.е. не наблюдаются при полном расплавлении) и часто могут быстро обходить эти сайты для завершения полного встраивания (фиг. 5С и фиг. 14А-14С), хотя среднее время первого прохождения полного встраивания нити значительно варьировало в зависимости от положения сайта ошибочного спаривания (фиг. 14А-14С). R-петли довольно стабильны во время встраивания (фиг. 5А), так как sgRNA часто способны оставаться полностью встроенными даже при наличии множественных ошибочных спариваний. Результаты качественно напоминают результаты ранее проведенных *in vitro* исследований связывания и расщепления dCas9/Cas9 на ошибочно спаренных мишенях. Однако в случае tru-gRNA (фиг. 5В) R-петли часто останавливаются за сайтами ошибочного спаривания. Среднее время первого прохода через ошибочное спаривание аналогично как для sgRNA, так и для tru-gRNA (фиг. 14А-14С), но проверка динамики для

КМС показывает, что из-за присущей изменчивости R-петли tru-gRNA, tru-gRNA часто быстро "повторно останавливаются" за ошибочным спариванием (фиг. 5C). Для sgRNA эта повторная остановка является гораздо более редкой. Следовательно, в сочетании с AFM-визуализацией результаты экспериментов КМС предполагают, что происхождение повышенной специфичности tru-gRNA заключается не в различии во время связывания, а в изменчивости ее R-петли (фиг. 4D), так что она повторно останавливается за ошибочными спариваниями даже после первоначального их обхода, что делает Cas9 менее склонным принимать активную конформацию. Относительно sgRNA, как только ошибочное спаривание обойдено, она может оставаться полностью встроенной с относительно небольшими отклонениями, что свидетельствует о механизме толерантности к ошибочному спариванию.

Пример 8.

Стабильность взаимодействия направляющей РНК с 14-17-м положениями протоспейсера коррелирует с экспериментальными значениями скорости нецелевого расщепления Cas9, в то время как общие энергии связи направляющей РНК и протоспейсера не коррелируют

Чтобы проверить, связана ли стабильность R-петли на или вблизи 16^{го} положения протоспейсера, которую связали согласно исследованиям AFM с конформационным изменением в Cas9, с активностью Cas9 *in vivo*, выполняли анализ согласно кинетическому методу Монте-Карло (КМС) стабильности R-петли на последовательностях, использованных Hsu et al. (2013) Nature Biotechnology, 31, 827-832. Набор данных Hsu et al. (2013) Nature Biotechnology, 31, 827-832 состоял из измерений частоты расщепления в пятнадцати различных мишенях-протоспейсерах, содержащих различные точечные мутации, относительно направляющей РНК, которые выполняли для исследования специфичности расщепления Cas9. Этот набор данных содержал 136 пар протоспейсер-направляющая РНК, имеющих одно изолированное ошибочное спаривание по типу rG·dG, rC·dC, rA·dA и rU·dT в РАМ-дистальной области (табл. 4), которые мы исследовали с использованием способов КМС, инициированных при размере R-петли $m=10$, для моделирования встраивания. Включение одного сайта ошибочного спаривания из этого набора уменьшало величину их общей свободной энергии связи направляющей РНК-протоспейсера в среднем только приблизительно на 6% по сравнению с идеально соответствующими мишенями, хотя, как уже упоминалось, имело место широкое распределение частот разрезания Cas9, наблюдаемое для этих пар направляющих РНК и протоспейсеров, происхождение которого не было очевидным.

Средний отрезок времени, в течение которого РНК стабильно связывалась с каждым сайтом протоспейсера, определяли для каждой направляющей РНК более чем в 1000 испытаний, который затем сопоставляли с оценкой расщепляющей активности Cas9 с максимальным правдоподобием (табл. 4, фиг. 6 и фиг. 15). Умеренную (0,433), но статистически значимую ($p < 1 \times 10^{-6}$) корреляцию обнаружили между стабильностью направляющей РНК в 16^м положении протоспейсера и описанной нецелевой расщепляющей активностью. Примечательно, что не была обнаружена статистически значимая корреляция между скоростью расщепления и предсказанными энергиями связи ДНК:РНК в отдельности (0,0786; $p=0,3631$) (фиг. 6А и 6В). В дополнение к стабильности R-петли в 16^м положении также обнаружили значимую корреляцию для стабильности 17^{го} сайта протоспейсера и описанного расщепления (табл. 5), но это не относилось к сайтам ≥ 18 сайта (фиг. 6). Хотя представленная в данном документе кинетическая модель Монте-Карло основана на относительно простой модели встраивания нити, эти результаты также свидетельствуют о том, что стабильность 16^{го} - 17^{го} сайтов протоспейсера и, следовательно, сопутствующие конформационные изменения, которые мы наблюдали, связаны с расщепляющей активностью Cas9 *in vivo* (фиг. 4D).

Таблица 5. Корреляции между экспериментальными (Hsu et al. (2013) Nature Biotechnology, 31, 827-832) частотами разрезания в целевых сайтах, содержащих одно ошибочное спаривание rG-dG, rC-dC, rA-dA и rU-dT в РАМ-дистальной области ($\geq 10^{10}$ сайта протоспейсера)^a и измерения стабильности направляющая РНК протоспейсер

		<u>$\log_{10}(p\text{-значение})$</u>	<u>Коэффициент корреляции</u>
Hsu et al. (2013) расчетная частота разрезания в зависимости от энергии связи направляющей РНК - протоспейсера ^b		-0,4400	(0,0786)
Расчетная частота разрезания по Hsu et al. в зависимости от положения сайта ошибочного спаривания		-5,8258	0,3990
Расчетная частота разрезания по Hsu et al. в зависимости от фрагментарного времени, в течение которого направляющая РНК связана в сайтах $\geq m^{10}$ сайта протоспейсера в смоделированной R-петле (КМС) ^c	$m=14$	-9,5550	0,5078
	$m=15$	-7,4854	0,4522
	$m=16$	-6,9510	0,4333
	$m=17$	-3,9270	0,3191
	$m=18$	-0,7639	(0,1159)
	$m=19$	-0,5546	(0,1058)
	$m=20$	-0,2346	(-0,0176)

^a $n=136$.

^b Для подробной информации см. табл. 4.

^c Для подробной информации см. текст. $\text{Max}(t)=100$.

Большая часть анализа была ограничена взаимодействиями с 16^M - 18^M нуклеотидами протоспейсера из-за наблюдаемых структурных различий между dCas9 с tru-gRNA и sgRNA. Однако также наблюдалось увеличение силы и статистической значимости корреляций между расщеплением и стабильностью 14^{10} и 15^{10} сайтов протоспейсера (фиг. 6) с наибольшей значимостью для корреляции в 14^M сайте. Поскольку R-петля является динамической структурой (фиг. 4D), возможно, что взаимодействия с этими сайтами являются теми критическими взаимодействиями, которые, как считается, ответственны за расщепление ДНК. Усечение направляющей РНК на 4 или 5 нуклеотидов может устранить расщепляющую активность при достаточной дестабилизации R-петли в 14^M или 15^M положении во многом таким же образом, как tru-gRNA дестабилизирует R-петлю в 16^M - 17^M сайтах. Однако, поскольку в данной модели в 14^M и 15^M сайты необходимо встраивание всякий раз, когда 16^M сайт связан с sgRNA, вполне вероятно, что эти положения дополнительно информативны, поскольку они также сильно отрицательно коррелировали с вероятностью диссоциации sgRNA из дуплекса до обхода сайта ошибочного спаривания (фиг. 6A и 6B), еще один механизм, с помощью которого расщепление не сможет произойти. В настоящее время нет кристаллографических свидетельств, которые непосредственно связывают встраивание нити с наблюдаемым конформационным изменением, которое, как полагают, обеспечивает возможность расщепления. Однако на основании доказательств, обеспеченных экспериментами AFM, представленными в данном документе, и результатов кинетических моделирований Монте-Карло мы заключаем, что стабильность направляющей gRNA в 14^M - 17^M сайтах протоспейсера во время встраивания имеет решающее значение для этого конформационного изменения и, в конечном счете, специфичности расщепления Cas9.

Кроме того, R-петля как динамическая структура в конкуренции между встраиванием нити и повторным отжигом ДНК может быть полезна в понимании механизмов нецелевого расщепления и толерантности к ошибочному спариванию. Никакой статистически значимой корреляции не было обнаружено между скоростью расщепления и предсказанными энергиями связи ДНК-РНК в отдельности (фиг. 6B), указывая на то, что кинетику встраивания нити можно рассмотреть при попытке определения активности Cas9 в нецелевых сайтах. Хотя расщепление устраняется, когда 4 или 5 нуклеотидов усечены в

направляющей РНК, Cas9 все еще способен расщеплять ДНК с вплоть до 6 дистальными сайтами ошибочного спаривания. Временные неспецифические взаимодействия в этих РАМ-дистальных сайтах могут достаточно стабилизировать конформационные сдвиги, необходимые для расщепления. Поскольку мы наблюдаем популяции меньшинства dCas9-sgRNA в частичных сайтах протоспейсера со структурами, аналогичными тем, что находятся в полном протоспейсере (желтый, фиг. 3С (i)), эта популяция может представлять долю Cas9 во временно стабилизированной активной конформации. Таким образом, эта популяция может нести ответственность за нецелевое расщепление.

В то время как специфичность связывания Cas9/dCas9 во многом определяется взаимодействием с РАМ-проксимальной областью, специфичность расщепления ДНК, скорее всего, определяется конформационным изменением в активированную структуру, которая стабилизируется путем взаимодействия направляющей РНК в области 14^и-17^и п.о. протоспейсера (фиг. 4D). Эксперименты согласно кинетическому методу Монте-Карло показывают, что R-петля, образованная во время встраивания нити направляющей РНК, может быть довольно динамичной структурой, даже когда направляющая РНК остается стабильно связанной, что предполагает механизм улучшения специфичности tru-gRNA и происхождения нецелевого расщепления посредством временной стабильности направляющей РНК-протоспейсера в критической области вокруг сайтов ошибочного спаривания. Предлагаемые механизмы для воздействия каждого из вариантов sgRNA на специфичность Cas9/dCas9 суммированы на фиг. 7.

С использованием АФМ было обнаружено, что hp-gRNA значительно ослабили или устранили специфическое связывание в гомеологичных мишенях. hp-gRNA могут быть полезными для модулирования сродства связывания и специфичности dCas9 в их потенциальных применениях в биологии и медицине. В частности, исходя из узкой геометрии канала связывания Cas9, наличие закрытой шпильки в ошибочно спаренных протоспейсерах может ингибировать конформационное изменение Cas9 в активное состояние. Открытие шпильки в hp-gRNA при связывании также может быть использовано в качестве зависимого от связывания сигнала *in vivo*, например, для образования динамических структур ДНК/РНК только при связывании с определенными сайтами.

Ранее проведенные исследования усечения направляющей РНК подняли вопрос о том, почему в естественных системах Cas9 используется crRNA, которая нацеливается на сайты протоспейсера размером 20 п.о., когда для расщепления требуется только направляющая последовательность из 16 нуклеотидов, а дополнительные нуклеотиды (> 18) не улучшают специфичность расщепления *in vivo*. Эти результаты свидетельствуют о том, что присутствие "лишних" 5'-нуклеотидов, которые связываются с 19^м и 20^м сайтами протоспейсера, буферизует этот неустойчивый повторный отжиг на критических 14^м-17^м сайтах протоспейсера, обеспечивая возможность осуществления эффективного конформационного изменения в активное состояние и последующего расщепления. Результаты экспериментов АФМ и КМС показывают, что стабильность направляющей РНК в этих сайтах сдвигает равновесную структуру Cas9 в сторону активной конформации при полном встраивании (фиг. 4А), тогда как изменчивость R-петель для "усеченных" направляющих РНК уменьшает давление для смещения равновесия в активное состояние. Неизбирательная активность Cas9 с sgRNA относительно tru-gRNA также может иметь эволюционные преимущества в своей роли средства адаптивного иммунитета у прокариот к инвазивной ДНК, поскольку ДНК внедряющихся фагов подвергается быстрым точечным мутациям в сайтах, на которые нацелен Cas9, чтобы избежать расщепления.

Конструирование последовательностей направляющей РНК для применений Cas9/dCas9 *in vivo* было сфокусировано главным образом на предотвращении мишеней с множеством сайтов с подобными последовательностями в геноме. Тем не менее, недавнее исследование, изучающее нецелевое расщепление, показало, что существующие способы прогнозирования нецелевой активности были в значительной степени неэффективными. Стабильность R-петли во время встраивания коррелирует со скоростями нецелевого расщепления значительно лучше, чем энергии связи направляющей РНК-протоспейсера в отдельности или положение ошибочного спаривания (еще один важный критерий, используемый при проектировании направляющей РНК, табл. 3). Стабильность R-петли в более короткие моменты времени после начала встраивания коррелировала с экспериментальной скоростью расщепления намного лучше, чем долгосрочная стабильность в экспериментах КМС (фиг. 16А), указывая на то, что кинетика встраивания нити является фактором в предсказании нецелевой активности.

Пример 9. Тестирование *In vivo*

Активность оптимизированной gRNA тестировали в живых клетках для исследования специфичности связывания dCas9.

Несколько gRNA-шпилек (hp-gRNA) разрабатывали для каждого из четырех целевых местоположений (протоспейсеров) в геноме человека (фиг. 17 и 18). Одно из них было в гене дистрофина (фиг. 19-23), другое было в гене EMX1 (фиг. 24-29 и 44), а две мишени были в гене VEGFA, помечены VEGFA1 (фиг. 30-37) и VEGFA3 (фиг. 38-43). Все эксперименты проводили в клетках HEK293T.

Дополнительные нуклеотиды (нт) добавляли к 5'-концу полной направляющей РНК (gRNA, полная длина 20 нт) и разрабатывали для образования шпилек и вторичных структур путем гибридизации с нуклеотидами, целенаправленно воздействующими на 5'-протоспейсер, или с нуклеотидами в середине или на 3'-конце области, нацеливающейся на протоспейсер, для модулирования активности связывания и

расщепления Cas9 относительно протоспейсеров.

Одну вторичную структуру нацеливающейся на VEGFA1 hp-gRNA компьютерно разрабатывали с использованием способов, описанных в данном документе, для предотвращения связывания в известном нецелевом сайте, позволяя при этом связывание с полным протоспейсером (фиг. 44A-44C). Hp-gRNA выбирали так, чтобы иметь продолжительность связывания, большую чем продолжительность связывания полной gRNA в целевом сайте или равную ей, и продолжительность связывания, меньшую чем продолжительность связывания полноразмерной gRNA в 3 главных нецелевых сайтах или равную ей. Другие 5'-структуры разрабатывали с включением неоднозначных пар dG-rU для модулирования энергии вторичных структур hp-gRNA или добавляли к концу усеченных gRNA (tru-gRNA, <20 нт), которые сами, как было показано, стимулировали высокую специфичность активности Cas9.

Работа с клетками. Для проведения анализа глубокого секвенирования клетки 293T трансфицировали плазмидами, которые экспрессировали Cas9 и представляющую интерес gRNA. Клетки инкубировали в течение 4 дней, позволяя Cas9 и gRNA проявлять свою максимальную активность. Затем клетки собирали и их геномную ДНК очищали. Использовали gRNA, которые были хорошо охарактеризованы в литературе (т.е. были известны их целевые и нецелевые сайты).

Анализ Surveyor. По сравнению с глубоким секвенированием анализ surveyor ниже по производительности и менее чувствителен. Однако анализ surveyor является более быстрым и менее техническим при анализе данных, обеспечивая изображения на геле. Таким образом, surveyor выполняли в качестве первого прохода и наилучшие условия анализировали в трех повторах с глубоким секвенированием. Как глубокое секвенирование, так и Surveyor представляют собой способы количественной оценки мутационных событий, вызванных Cas9+gRNA.

Работа с клетками для Surveyor была такой же, как описано выше. После очистки геномной ДНК разрабатывали праймеры для амплификации целевого сайта. В этом эксперименте использовали пул из 200 тыс. клеток, и каждая из них имела другую мутацию, поскольку репарация ДНК была стохастической. Сайт по всем 200 тыс. клеткам амплифицировали для создания гетерогенного продукта ПЦР: некоторые ампликоны имели делеции, некоторые имели вставки, а некоторые были дикого типа и немодифицированы из-за того, что каждая клетка стохастически (то есть случайно, со сниженной точностью) репарировала сайты разреза Cas9.

Гетерогенный ПЦР-пул нагревали и репарировали, а в некоторых случаях разные нити отжигали друг с другом: нить ДНК дикого типа могла связаться с ДНК со вставкой, или вставка могла связаться с делецией. Когда это происходит, образуется небольшой "пузырь", и эта структура называется гетеродуплексом ДНК (см. фиг. 46).

Нуклеазу surveyor использовали для обнаружения этих гетеродуплексов путем их расщепления. Затем расщепление ДНК было посредником для мутационной активности Cas9. ПЦР-пул разделяли на геле и интенсивность этих расщепленных полос использовали для количественной оценки уровня активности Cas9.

Глубокое секвенирование. Праймеры разрабатывали для амплификации этих известных мишеней/нецелевых участков. В этой PCR использовали высоконадежную полимеразу. На этих праймерах также присутствовали адаптеры Illumina, так что их можно было пометить штрихкодом и загрузить на платформу Illumina Mi-Seq. На фигурах и в кратком описании графических материалов описано количество шпилек, количество мишеней, количество нецелевых участков, охват секвенирования и т.д. Хороший охват получили во всех образцах, используемых в анализе. Среднее число прочтений/образцов составило 20000. Наименьшее количество прочтений образца составило 1700. Очень небольшое количество мишеней не генерировало достаточное количество выровненных прочтений и не было включено в анализ.

Полученные данные секвенирования анализировали с использованием программного обеспечения CRISPResso (Pinello et al. Nat Biotechnol. (2016) 34(7):695-697), которое выравнивает прочтения глубокого секвенирования с определенными сайтами известных нецелевых или целевых местоположений. Результаты этого программного обеспечения сопоставляли с внутрилабораторными сценариями, в которых выполнялось глобальное выравнивание прочтений глубокого секвенирования с геномом человека и очень хорошо коррелировало. Мутационные частоты определяли количественно с использованием CRISPResso, и полученные данные отображали на демонстрируемых гистограммах для каждого гена-мишени.

Сначала конструкции тестировали с использованием анализа Surveyor для проверки вставок/делеций после экспрессии Cas9 и hp-gRNA в клетках НЕК в целевом сайте и нецелевых сайтах, на которые, как известно, целенаправленно воздействовали с использованием стандартных gRNA (см. табл. 6). Активность в этих сайтах сравнивали со стандартной gRNA и усеченными gRNA (tru-gRNA). Они показаны ниже в виде гелей, демонстрирующих расщепление нуклеазой Surveyor ПЦР-амплифицированной геномной ДНК, где расщепление указывает на мутагенез при помощи Cas9.

Таблица 6
Протоспейсеры **Геномные мишени**

Дистрофин	1 целевая, 1 нецелевая
EMX1	1 целевая, 7 нецелевых
VEGFA1	1 целевая, 10 нецелевых
VEGFA3	1 целевая, 22 нецелевых

Наиболее перспективные конструкции hp-gRNA выбрали для дополнительного количественного анализа с использованием секвенирования нового поколения для оценки активности Cas9 в целевых и нецелевых сайтах в клетках НЕК. Специфичность определяли как целевые хиты/сумма (нецелевые хиты).

В то время как активность Cas9 в целом была равна или несколько уменьшалась при использовании hp-gRNA, каждая hp-gRNA, выбранная для экспериментов глубокого секвенирования, показала повышенную специфичность по сравнению с полными gRNA и в большинстве случаев была равна или больше, чем tru-gRNA, с точки зрения специфичности.

В одном случае шпилька hp-gRNA, нацеливающаяся на EMX1, демонстрировала > 6000-кратное улучшение специфичности по сравнению с полной gRNA (по сравнению с tru-gRNA с "100-кратным улучшением по сравнению с gRNA). Нацеливающаяся на VEGFA1 hp-gRNA с вычислительно разработанной вторичной структурой с использованием внутрिलाбораторного алгоритма значительно превосходила активность tru-gRNA с точки зрения специфичности (в 18 раз по сравнению с 3-кратным улучшением при использовании gRNA). Эти hp-gRNA тестировали в сочетании с Cas9 из *S. pyogenes*. На фиг. 44A-44C показаны анализы Surveyoq нацеливающихся на EMX1 hp-gRNA с Cas9 из *S. aureus*, проявляющих целевую активность и не обнаруживающих нецелевой активности в отличие от tru-gRNA, которые проявили значительную нецелевую активность.

Пример 10. hp-gRNA для системы CRISPR/Cpf1

Эксперименты разработали для воспроизведения результатов Kleinstiver et al., Nat. Biotech. (2016) 34:869-874. Kleinstiver и соавт. использовали полноразмерные gRNA, чтобы показать, что Cpf1 из *Lachnospiraceae* восприимчив к разрезанию в нецелевых сайтах с ошибочными спариваниями 8-9 нуклеотидов в дополнение к PAM-дистальным сайтам при использовании gRNA, которые имели ошибочные спаривания с целевым сайтом в разных местоположениях (фиг. 47). В этом примере направляющие РНК-шпильки, используемые с системой CRISPR-Cas типа V CRISPR-Cpf1, разрабатывали и испытывали, как описано выше, с использованием способов по настоящему изобретению.

Для тестирования нецелевой активности Cpf1 с дополнительными элементами вторичной структуры и без них на ген DNMT1 (TTTC CTGATGGGTCCATGTCTGTACTC (SEQ ID NO: 330)) нацеливались для расщепления с помощью Cpf1. "Нецелевую активность" тестировали с использованием направляющих РНК, которые имели ошибочно спаренный нуклеотид в положении 9, например, CTGATGGTgCATGTCTGTTA (SEQ ID NO: 331), с использованием полноразмерных направляющих РНК длиной 20 нуклеотидов или усеченных gRNA длиной 17 нуклеотидов CTGATGGTgCATGTCTTG (SEQ ID NO: 332).

Элементы вторичной структуры длиной 9 нуклеотидов добавляли к 3'-концу направляющих РНК с Cpf1 для гибридизации с сегментом направляющей РНК, окружающим ошибочно спаренный нуклеотид, где в этом случае "линкерный" элемент состоял из 4 3'-нт сегмента, нацеливающегося на протоспейсер, т.е.

CTGATGGTgCATGTCT GTTA AGACATGcACCA (SEQ ID NO: 333) и CTGATGGTgCATG TCTG CATGcACCA (SEQ ID NO: 334).

Анализ

Surveyoq показывает, что включение этих дополнительных 3'-элементов уменьшило или устранило нецелевую активность в сайте DNMT1, проявляемую полными или усеченными gRNA.

hp-gRNA разрабатывали с конструкцией "внутренней" шпильки, в которой 4 PAM-дистальных нуклеотида служили в качестве петли. Шпильку добавляли к 3'-концу gRNA. В табл. 7 показаны последовательности hp-gRNA с промежутком в последовательностях, который разделяет эту область. Ошибочное спаривание показано в нижнем регистре.

Результаты исследования Surveyoq этих hp-gRNA показаны на фиг. 48 и демонстрируют, что добавление шпильки к 3'-концу устраняло нецелевую активность. Дорожка 1 показывает контроль; дорожка 2 показывает полноразмерную gRNA, содержащую ошибочно спаренный нуклеотид в положении 9; дорожка 3 показывает полноразмерную gRNA, содержащую ошибочно спаренный нуклеотид в положении 9 и дополнительную структуру 3'-шпильки; дорожка 4 показывает усеченную gRNA, содержащую ошибочно спаренный нуклеотид в положении 9; и дорожка 5 показывает усеченную gRNA, содержащую ошибочно спаренный нуклеотид в положении 9 и дополнительную структуру 3'-шпильки. Используемые

праймеры Surveyor также показаны в табл. 7.

Cpf1 переносит ошибочные спаривания в нуклеотидах 8-10 при использовании нормальных направляющих РНК и расщепляет ДНК в этих нецелевых сайтах (фиг. 47). Как показано на фиг. 48, hp-gRNA Cpf1 смогли устранить нецелевую активность, продемонстрированную у Kleinstiver, тогда как усеченные gRNA не могли.

Таблица 7

Праймеры Surveyor		
Метка	Последовательность	Ожидаемый размер продукта
CN391 DNMT1 (прямой)	CTGGGACTCAGGCGGGTCAC (SEQ ID NO: 324)	606 п.о.
CN406 DNMT1 обратный зафиксированный	CCTCACACAACAGCTTTCATGTACAGC (SEQ ID NO: 325)	
Последовательности протоспейсера		
Метка	Последовательность	
LbCpf1_9mm_20nt_S	CTGATGGTgCATGTCTGTTA (SEQ ID NO: 326)	
LbCpf1_9mm_17nt_S	CTGATGGTgCATG TCTG (SEQ ID NO: 327)	
LbCpf1_9mm_20nt_hp_S	CTGATGGTgCATGTCT GTTA AGACATGcACCA (SEQ ID NO: 328)	
LbCpf1_9mm_17nt_hp_S	CTGATGGTgCATG TCTG CATGcACCA (SEQ ID NO: 329)	

Понятно, что приведенное выше подробное описание и сопровождающие примеры являются просто иллюстративными и не должны рассматриваться в качестве ограничений объема настоящего изобретения, которое определяется исключительно прилагаемой формулой изобретения и ее эквивалентами.

Различные изменения и модификации раскрытых вариантов осуществления будут очевидны для специалистов в данной области техники. Такие изменения и модификации, включая без ограничения те, которые связаны с химическими структурами, заместителями, производными, промежуточными соединениями, синтезами, композициями, составами или способами применения настоящего изобретения, могут быть сделаны без отхода от его сути и объема.

В целях обеспечения полноты различные аспекты настоящего изобретения изложены в следующих пронумерованных пунктах.

Пункт 1.

Способ получения оптимизированной направляющей РНК (gRNA), при этом способ предусматривает: а) идентификацию представляющей интерес целевой области, при этом представляющая интерес целевая область содержит последовательность протоспейсера; б) определение полинуклеотидной последовательности полноразмерной gRNA, которая нацеливается на представляющую интерес целевую область, при этом полноразмерная gRNA содержит нацеливающиеся на протоспейсер последовательность или сегмент; в) определение по меньшей мере одного или нескольких нецелевых сайтов для полноразмерной gRNA; г) получение полинуклеотидной последовательности первой gRNA, при этом первая gRNA содержит полинуклеотидную последовательность полноразмерной gRNA и сегмент РНК, при этом сегмент РНК содержит полинуклеотидную последовательность, имеющую М нуклеотидов в длину, которая комплементарна нуклеотидному сегменту нацеливающимся на протоспейсер последовательности или сегменту, при этом РНК расположена на 5'-конце полинуклеотидной последовательности полноразмерной gRNA, при этом первая gRNA необязательно содержит линкер между 5'-концом полинуклеотидной последовательности полноразмерной gRNA и сегментом РНК, при этом линкер содержит полинуклеотидную последовательность, имеющую N нуклеотидов в длину, при этом первая gRNA способна встраиваться в последовательность протоспейсера, и связываться с последовательностью ДНК, которая комплементарна последовательности протоспейсера, и образовывать дуплекс с протоспейсером, и при этом первая gRNA способна встраиваться в нецелевой сайт и связываться с последовательностью ДНК, которая комплементарна нецелевому сайту, и образовывать нецелевой дуплекс; е) вычисление оценочного показателя или вычислительное моделирование кинетики встраивания и времени нахождения, в течение которого gRNA остается встроенной в дуплексы протоспейсера и нецелевого сайта, где динамику встраивания оценивают нуклеотид за нуклеотидом путем определения энергетических различий между дополнительным встраиванием отличающейся gRNA и повторным отжигом первой gRNA на последова-

тельности ДНК, которая комплементарна последовательности протоспейсера; f) сравнение значений предполагаемого времени нахождения в сайтах протоспейсера и/или нецелевых сайтах первой gRNA со значениями предполагаемого времени нахождения полноразмерной gRNA или усеченной gRNA (tru-gRNA) в сайтах протоспейсера и/или нецелевых сайтах; g) рандомизацию от 0 до N нуклеотидов в линкере и от 0 до M нуклеотидов в первой gRNA и получение второй gRNA, и повторение стадии (e) со второй gRNA; h) идентификацию оптимизированной gRNA на основе последовательности gRNA, которая удовлетворяет критериям конструирования; и i) тестирование оптимизированной gRNA *in vivo* для определения специфичности связывания.

Пункт 2.

Способ получения оптимизированной направляющей РНК (gRNA), при этом способ предусматривает: а) идентификацию представляющей интерес целевой области, при этом представляющая интерес целевая область содержит последовательность протоспейсера; б) определение полинуклеотидной последовательности полноразмерной gRNA, которая нацеливается на представляющую интерес целевую область, при этом полноразмерная gRNA содержит нацеливающиеся на протоспейсер последовательность или сегмент; в) определение по меньшей мере одного или нескольких нецелевых сайтов для полноразмерной gRNA; д) получение полинуклеотидной последовательности первой gRNA, при этом первая gRNA содержит полинуклеотидную последовательность полноразмерной gRNA и сегмент РНК, при этом сегмент РНК содержит полинуклеотидную последовательность, имеющую M нуклеотидов в длину, которая комплементарна нуклеотидному сегменту нацеливающихся на протоспейсер последовательности или сегменту, при этом РНК расположена на 3'-конце полинуклеотидной последовательности полноразмерной gRNA, при этом первая gRNA необязательно содержит линкер между 3'-концом полинуклеотидной последовательности полноразмерной gRNA и сегментом РНК, при этом линкер содержит полинуклеотидную последовательность, имеющую N нуклеотидов в длину, при этом первая gRNA способна встраиваться в последовательность протоспейсера, и связываться с последовательностью ДНК, которая комплементарна последовательности протоспейсера, и образовывать дуплекс с протоспейсером, и при этом первая gRNA способна встраиваться в нецелевой сайт, и связываться с последовательностью ДНК, которая комплементарна нецелевому сайту, и образовывать нецелевой дуплекс; е) вычисление оценочного показателя или вычислительное моделирование кинетики встраивания и времени нахождения, в течение которого gRNA остается встроенной в дуплексы протоспейсера и нецелевого сайта, где динамику встраивания оценивают нуклеотид за нуклеотидом путем определения энергетических различий между дополнительным встраиванием отличающейся gRNA и повторным отжигом первой gRNA на последовательности ДНК, которая комплементарна последовательности протоспейсера; f) сравнение значений предполагаемого времени нахождения в сайтах протоспейсера и/или нецелевых сайтах первой gRNA со значениями предполагаемого времени нахождения полноразмерной gRNA или усеченной gRNA (tru-gRNA) в сайтах протоспейсера и/или нецелевых сайтах; g) рандомизацию от 0 до N нуклеотидов в линкере и от 0 до M нуклеотидов в первой gRNA, и получение второй gRNA, и повторение стадии (e) со второй gRNA; h) идентификацию оптимизированной gRNA на основе последовательности gRNA, которая удовлетворяет критериям конструирования; и i) тестирование оптимизированной gRNA *in vivo* для определения специфичности связывания.

Пункт 3.

Способ по п.1 или 2, где энергетический баланс дополнительного встраивания отличающейся gRNA определяют путем определения энергетического баланса по меньшей мере одного из (I) нарушения спаривания оснований ДНК-ДНК, (II) образования пар оснований РНК-ДНК, (III) энергетического различия, возникающего в результате разрушения или образования отличающейся вторичной структуры внутри не встроенной направляющей РНК, и (IV) образования или разрушения взаимодействий между вытесненной нитью ДНК, которая комплементарна протоспейсеру, и любыми неспаренными нуклеотидами направляющей РНК, которые не вовлечены во вторичные структуры.

Пункт 4.

Способ по любому из пп.1-3, где энергетический баланс повторного отжига первой gRNA с последовательностью ДНК, которая комплементарна последовательности протоспейсера, определяют путем определения энергетического баланса по меньшей мере одного из (I) образования пар оснований ДНК-ДНК, (II) разрушения пар оснований РНК-ДНК, (III) энергетического различия, возникающего в результате разрушения или образования отличающейся вторичной структуры внутри новой не встроенной направляющей РНК, и (IV) образования или разрушения взаимодействий между вытесненной нитью ДНК, которая комплементарна протоспейсеру, и любыми неспаренными нуклеотидами направляющей РНК, которые не вовлечены во вторичные структуры.

Пункт 5.

Способ по п.3 или 4, дополнительно предусматривающий определение энергетических факторов по меньшей мере одного из (V) спаривания оснований среди ошибочных спариваний, (VI) взаимодействий с белком Cas9 и/или (VII) дополнительных эвристических показателей, где дополнительные эвристические показатели относятся к продолжительности связывания, степени встраивания, стабильности встраивающейся направляющей РНК или другим расчетным/моделируемым свойствам встраивания gRNA для

расщепляющей активности Cas9.

Пункт 6. Способ по любому из пп.1-5, где полноразмерная gRNA содержит от приблизительно 15 до 20 нуклеотидов.

Пункт 7. Способ по любому из пп.1-5, где M составляет от 1 до 20.

Пункт 8. Способ по п.7, где M составляет от 4 до 10.

Пункт 9. Способ по любому из пп.1-8, где сегмент РНК содержит от 2 до 15 нуклеотидов, которые комплементарны нацеливающейся на протоспейсер последовательности.

Пункт 10. Способ по любому из пп.1-9, где N составляет от 1 до 20.

Пункт 11. Способ по п.10, где N составляет от 3 до 10.

Пункт 12. Способ по любому из пп.1-11, где сегмент РНК и/или нацеливающаяся на протоспейсер последовательность обеспечивают вторичную структуру.

Пункт 13. Способ по п.12, где вторичная структура образуется путем частичной гибридизации нацеливающейся на протоспейсер последовательности с сегментом РНК.

Пункт 14. Способ по п.13, где вторичная структура модулирует связывание или расщепление ДНК посредством Cas9 путем нарушения встраивания оптимизированной gRNA в дуплекс протоспейсера или нецелевой дуплекса.

Пункт 15. Способ по любому из пп.12-14, где вторичная структура образуется путем гибридизации всего или части сегмента РНК с нуклеотидами на 5'-конце нацеливающихся на протоспейсер последовательности или сегмента, нуклеотидами в середине нацеливающихся на протоспейсер последовательности или сегмента, и/или нуклеотидами на 3'-конце нацеливающихся на протоспейсер последовательности или сегмента.

Пункт 16. Способ по любому из пп.12-15, где вторичная структура представляет собой шпильку.

Пункт 17. Способ по любому из пп.12-16, где вторичная структура является стабильной при комнатной температуре или при 37°C.

Пункт 18. Способ по любому из пп.12-17, где общая равновесная свободная энергия вторичной структуры составляет менее приблизительно 2 ккал/моль при комнатной температуре или при 37°C.

Пункт 19. Способ по любому из пп.1-18, где сегмент РНК гибридизируется или образует неканонические пары оснований по меньшей мере с двумя нуклеотидами нацеливающихся на протоспейсер последовательности или сегмента.

Пункт 20. Способ по п.19, где неканоническая пара оснований представляет собой rU-rG.

Пункт 21. Способ по любому из пп.1-20, где оптимизированную gRNA применяют с системой на основе CRISPR/Cas9 или с системой на основе CRISPR/Cpf1 в клетке.

Пункт 22. Способ по любому из пп.1-21, где вторичная структура защищает оптимизированную gRNA в системе на основе CRISPR/Cas9 или системе на основе CRISPR/Cpf1 с целью предотвращения разрушения в клетке.

Пункт 23. Способ по любому из пп.1-22, где 1-20 нуклеотидов рандомизированы в линкере.

Пункт 24. Способ по любому из пп.1-23, где 1-20 нуклеотидов рандомизированы в сегменте РНК.

Пункт 25. Способ по любому из пп.1-24, где стадию (g) повторяют X раз, с получением тем самым gRNA в количестве X, и повторяют стадию (e) с каждым количеством X gRNA, где X составляет от 0 до 20.

Пункт 26. Способ по любому из пп.1-25, где кинетику встраивания и время нахождения вычисляют с применением кинетической методики Монте-Карло или алгоритма Гиллеспи.

Пункт 27. Способ по любому из пп.1-26, где кинетика встраивания представляет собой скорость, с которой направляющая РНК встраивается в дуплекс протоспейсера с обеспечением полного встраивания таким образом, что встраивание в протоспейсер является полностью осуществленным, и/или скорость, с которой сегмент ДНК протоспейсера, связанной с gRNA, удлиняется по мере его вытеснения из его комплементарной нити и связывания с gRNA нуклеотид за нуклеотидом от его PAM-проксимальной области до полного встраивания.

Пункт 28. Способ по любому из пп.1-27, где критерии конструирования предусматривают специфичность, модуляцию продолжительности связывания и/или расчетную специфичность расщепления.

Пункт 29. Способ по п.28, где критерии конструирования предусматривают оптимизированную gRNA с продолжительностью связывания, которая больше или равна продолжительности связывания полноразмерной gRNA с целевым сайтом, и/или с продолжительностью связывания, которая меньше или равна продолжительности связывания полноразмерной gRNA с нецелевым сайтом.

Пункт 30. Способ по п.29, где критерии конструирования предусматривают оптимизированную gRNA с продолжительностью связывания, которая меньше или равна продолжительности связывания полноразмерной gRNA по меньшей мере с тремя нецелевыми сайтами, где нецелевые сайты согласно прогнозированию представляют собой ближайшие нецелевые сайты, или прогнозируется, что они имеют наивысшую степень идентичности с целевыми сайтами.

Пункт 31. Способ по п.28, где критерии конструирования предусматривают время нахождения или скорость расщепления в нецелевом сайте, которые меньше или равны времени нахождения или скорости расщепления полноразмерной gRNA или усеченной gRNA в нецелевом сайте, и/или прогнозируемый

показатель целевой активности, который составляет более 10% от прогнозируемого показателя целевой активности полноразмерной gRNA или усеченной gRNA.

Пункт 32. Способ по любому из пп.1-31, где оптимизированную gRNA тестируют на стадии i) с применением анализа surveyor, методик секвенирования нового поколения или GUIDE-Seq.

Пункт 33. Способ по любому из пп.1-32, где оптимизированная gRNA сконструирована для минимизации связывания в нецелевом сайте и обеспечения возможности связывания с последовательностью протоспейсера.

Пункт 34. Способ по любому из пп.1-33, где нецелевой сайт представляет собой известный или прогнозируемый нецелевой сайт.

Пункт 35. Способ по любому из пп.1-34, где полноразмерная gRNA нацелена на ген млекопитающего.

Пункт 36. Способ по любому из пп.1-35, где целевой ген предусматривает эндогенный целевой ген или трансген.

Пункт 37. Способ по любому из пп.1-36, где целевой ген представляет собой ген, связанный с заболеванием.

Пункт 38. Способ по любому из пп.1-37, где геном-мишенью является ген DMD, EMX1 или VEGFA.

Пункт 39. Способ по п.38, где ген VEGFA представляет собой VEGFA1 или VEGFA3.

Пункт 40. Оптимизированная gRNA, полученная с помощью способа по любому из пп.1-39.

Пункт 41. Оптимизированная gRNA по п.40, где gRNA может различать целевые и нецелевые сайты с минимальными термодинамическими энергетическими различиями между сайтами.

Пункт 42. Оптимизированная gRNA по п.40 или 41, где оптимизированная gRNA модулирует встраивание нити в протоспейсер.

Пункт 43. Оптимизированная gRNA по любому из пп.40-42, где оптимизированная gRNA содержит нуклеотидную последовательность по меньшей мере из одного из SEQ ID NO: 149-315, 321-323 и 326-329.

Пункт 44. Выделенный полинуклеотид, кодирующий оптимизированную gRNA по любому из пп.40-43.

Пункт 45. Вектор, содержащий выделенный полинуклеотид по п.44.

Пункт 46. Клетка, содержащая выделенный полинуклеотид по п.44 или вектор по п.45.

Пункт 47. Набор, содержащий выделенный полинуклеотид по п.44, вектор по п.45 или клетку по п.46.

Пункт 48. Способ эпигеномного редактирования в клетке-мишени или субъекте, при этом способ предусматривает приведение клетки или субъекта в контакт с эффективным количеством молекулы оптимизированной gRNA по любому из пп.40-43 или слитого нуклеотида по п.44 и слитого белка, при этом слитый белок содержит первый полипептидный домен, содержащий дефицитную по нуклеазной активности Cas9, и второй полипептидный домен с активностью, выбранной из группы, состоящей из активности активации транскрипции, активности репрессии транскрипции, нуклеазной активности, активности фактора освобождения транскриптов, активности модификации гистонов, активности ассоциации нуклеиновых кислот, ДНК-метиلاзной активности и прямой или непрямой ДНК-деметилазной активности.

Пункт 49. Способ сайт-специфического расщепления ДНК в клетке-мишени или субъекте, при этом способ предусматривает приведение клетки или субъекта в контакт с эффективным количеством молекулы оптимизированной gRNA по любому из пп.40-43 или выделенного полинуклеотида по п.44 и слитого белка или белка Cas9, при этом слитый белок содержит первый полипептидный домен, содержащий дефицитную по нуклеазной активности Cas9, и второй полипептидный домен с активностью, выбранной из группы, состоящей из активности активации транскрипции, активности репрессии транскрипции, нуклеазной активности, активности фактора освобождения транскриптов, активности модификации гистонов, активности ассоциации нуклеиновых кислот, ДНК-метилазной активности и прямой или непрямой ДНК-деметилазной активности.

Пункт 50. Способ редактирования генома в клетке, при этом способ предусматривает введение в клетку эффективного количества молекулы оптимизированной gRNA по любому из пп.40-43 или выделенного полинуклеотида по п.44 и слитого белка, при этом слитый белок содержит первый полипептидный домен, содержащий дефицитную по нуклеазной активности Cas9, и второй полипептидный домен с активностью, выбранной из группы, состоящей из активности активации транскрипции, активности репрессии транскрипции, нуклеазной активности, активности фактора освобождения транскриптов, активности модификации гистонов, активности ассоциации нуклеиновых кислот, ДНК-метилазной активности и прямой или непрямой ДНК-деметилазной активности.

Пункт 51. Способ по п.50, где редактирование генома предусматривает коррекцию мутантного гена или вставку трансгена.

Пункт 52. Способ по п.51, где исправление мутантного гена включает делецию, перестройку или замену мутантного гена.

Пункт 53. Способ по любому из п.51 или 52, где коррекция мутантного гена предусматривает опосредованное нуклеазой негомологичное соединение концов или репарацию с участием гомологичной рекомбинации.

Пункт 54. Способ модулирования экспрессии гена в клетке, при этом способ предусматривает приведение клетки в контакт с эффективным количеством молекулы оптимизированной gRNA по любому из пп.40-43 или выделенного полинуклеотида по п.44 и слитого белка, при этом слитый белок содержит первый полипептидный домен, содержащий дефицитную по нуклеазной активности Cas9, и второй полипептидный домен с активностью, выбранной из группы, состоящей из активности активации транскрипции, активности репрессии транскрипции, нуклеазной активности, активности фактора освобождения транскриптов, активности модификации гистонов, активности ассоциации нуклеиновых кислот, ДНК-метилазной активности и прямой или непрямой ДНК-деметилазной активности.

Пункт 55. Способ по п.54, где экспрессию гена, по меньшей мере одного целевого гена, модулируют, если уровни экспрессии гена, по меньшей мере одного целевого гена, увеличены или уменьшены по сравнению с уровнями экспрессии нормального гена, по меньшей мере для одного целевого гена.

Пункт 56. Способ по п.54 или 55, где слитый белок содержит домен dCas9 и активатор транскрипции.

Пункт 57. Способ по п.56, где слитый белок содержит аминокислотную последовательность под SEQ ID NO: 2.

Пункт 58. Способ по п.54 или 55, где слитый белок содержит домен dCas9 и репрессор транскрипции.

Пункт 59. Способ по п.58, где слитый белок содержит аминокислотную последовательность под SEQ ID NO: 3.

Пункт 60. Способ по п.54 или 55, где слитый белок содержит домен dCas9 и сайт-специфическую нуклеазу.

Пункт 61. Способ по любому из пп.48-60, где оптимизированная gRNA кодируется полинуклеотидной последовательностью и упакована в лентивирусный вектор.

Пункт 62. Способ по п.61, где лентивирусный вектор содержит кассету экспрессии, содержащую промотор, функционально связанный с полинуклеотидной последовательностью, кодирующей gRNA.

Пункт 63. Способ по п.62, где промотор, функционально связанный с полинуклеотидом, кодирующим оптимизированную gRNA, является индуцируемым.

Пункт 64. Способ по любому из пп.61-63, где лентивирусный вектор дополнительно содержит полинуклеотидную последовательность, кодирующую белок Cas9 или слитый белок.

Пункт 65. Способ по любому из пп.48-64, где по меньшей мере один целевой ген представляет собой ген, связанный с заболеванием.

Пункт 66. Способ по любому из пп.48-65, где клетка-мишень представляет собой эукариотическую клетку.

Пункт 67. Способ по любому из пп.48-66, где клетка-мишень представляет собой клетку млекопитающего.

Способ по любому из пп.48-67, где клетка-мишень представляет собой клетку HEK293T.

Последовательности

Cas 9 из Streptococcus pyogenes (с D10A, H840A) (SEQ ID NO:

1)

MDKKYSIGLAI^AIGTNSVGWAVITDEYKVPSSKFKVLGN^TDRHSIKKNLIGALLFDSGETA
 EATRLKRTARRRYTRRKNRICYLQEIFS^NEMAKVDD^SFFHRLEESFLVEEDKKHERHPIFGNIV
 DEVAYHEKYPTIYHLRKKLV^DSTDKADLR^LIYLALAHMIKFRGHFLIEGDLNPDNSDV^DKLFIQ
 LVQTYNQLFEENPINASGVDAKAILSARLSKSRREN^LIAQLPGEKKNLFGNLI^ALSLGLTPN
 FKS^NFDLAEDA^KLQLSKD^TYDDDLNLLAQIGDQYADLFLAAKNLSDAI^LLLSDILRVNTEITKA
 PLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPI
 LEKMDGTEELLV^KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFY^PFLKDNREKIEKI
 LTRIPYVYVGPLARGNSRFAWMTRKSEETITPWNFE^EVVDKGASAQSFIERMTNFDKNLPNEKV
 LPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLS^GEQKKAIVDLLFKTNRKVTVKQ^LKEDYFKK
 IECFDSVEISGVEDRFNASLGT^YHDL^LKIKDKDFLDNEENEDILEDIVLTLTLFEDREMI^EER
 LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFL^KSDGFANRNF^MQLIHD
 DSLTFKEDIQKAQVSGQ^GDSLHEHIANLAGSPA^IKKGILQTVKVVDELVKVMGRHKPENIV^IEM
 ARENQTQKQKNSRERMKRIE^EGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQ^E
 LDINRLSDYDVDAIVPQSFLKDDSIDNKV^LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKL
 ITQRKFDNLTKAERGGLSEL^DKAGFIKRQLVETRQITKHVAQILDSR^MNTKYDENDKLI^REVKV
 ITL^KSKLVSDFRKDFQFYK^VREINNYHHAHDAYLNAVVG^TALIKKYPKLESEFVYGDYK^VYDVR
 KMI^AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRK^RPLIETNGETGEIVWDKGRDFATV
 R^KVL^SM^PQVNI^VKKTEVQTGGFSKESILPKRNSDKLIAR^KKDWD^PKKYGGFDSPTVAYS^VLVVA
 KVEK^GSK^KLK^SVKELLGITIMERSSEK^NPIDFLEAKGYKEV^KKD^LIKLPK^YSLFELENGR^K
 RMLASAGELQKGNELALPSKYVNF^LYLASHYEK^LKGSPEDNEQKQLFVEQHKHYLDEIIEQISE
 FSKRVILADANLDK^VLSAYN^KHRDKPIREQAENIHLFTLTNLGAPAAFKYFD^TTIDR^KRYTST
 KEVL^DATLIHQ^SITGLYETRIDL^SQLGGD

dCas9^{p300} Core: (Addgene Plasmid 61357) аминокислотная последовательность; 3X эпитоп "Flag", Последовательность ядерной локализации, Cas9 *Streptococcus pyogenes* (D10A, H840A), эффектор p300 Core, эпитоп "HA" (SEQ ID NO: 2)

MDYKDHGDYKD^HDI^DYKDDDDKMAP^KKKR^KVGRGM^DKKYSIGLAI^AIGTNSVGWAVITDE
 YKVPSSKFKVLGN^TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS^N
 E^MAKVDD^SFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV^DSTDKADLR
 LIYLALAHMIKFRGHFLIEGDLNPDNSDV^DKLFIQLVQTYNQLFEENPINASGVDAKAILSARL
 SKSRREN^LIAQLPGEKKNLFGNLI^ALSLGLTPNFKS^NFDLAEDA^KLQLSKD^TYDDDLNLLA
 QIGDQYADLFLAAKNLSDAI^LLLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLP
 EKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLV^KLNREDLLRKQRTFDNG

SIPHQIHLGELHAILRRQEDFYPLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEET
ITPWNFEVVDKASQAQSFIERMTNFDKNLPNEKVLPHKSLLEYEFTVYNELTKVKYVTEGMRK
PAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKLI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSR
KLINGIRKQSGKTIILDFLKSDFANRNFQLIHDDSLTFKEDIQKAQVSGQDLSLHEHIANLA
GSPAIKKGIILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKQKNSRERMKRIEIEGKELG
SQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVAIVPQSFLKDDSIDNKV
LTRSDKNRGSNDNPSEEVVKKMKNYWRQLLNAKLI TORKFDNLTKAERGGLSELDKAGFIKRO
LVE TRQITKHVAQILDSRMNTKYDENDKLIREVKVI TLKSKLVSDFRKDFQFYKVRE INNYHHA
HDAYLNAVGTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKT
EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIKKTEVQTGGFSKESILP
KRNSDKLIARKKDWDPKYGGFDSPTVAYSVLVVAKVEKGSKLLKSVKELLGITIMERSSEFK
NPIDFLEAKGYKEVKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLAS
HYEKLKGSPEQNEQQLFVEQHKHYLDEIEQISEFSKRVI LADANLDKVL SAYNKHRDKPIRE
QAENI IHLFTLNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGD
PIAGSKASPKKRRKVGRAIFKPEELRQALMPTLEALYRQDPE SLPFRQVPDQLLGI PDYFDIV
KSPMDLSTIKRKLDTGQYQEPWQYVDDIWL MFNNAWLYNRKTSRVYKYCSKLSEVFEQEIDPVM
QSLGYCCGRKLEFSPQTLCCYGKQLCTIPR DATYYSYQNRHYHFCEKCFNEIQGESVSLGDDPSQ
PQTTINKEQFSKRKNDTLDPELFVECTEGRKMHQICVLHHEI IWPAGFVCDGCLKKSARTRKE
NKFSAKRLPSTRLGTFLENRVNDFLRRQNHPESGEVTVRVVHASDKTVEVKPGMKARFVDSGEM
AESFPYRTKALFAFEEIDGVDLCFFGMHVQEYGSDCPPPNQRRVYISYLDVHFFRPKCLR TAV
YHEILIGYLEYVKKLGYTTGHIWACPPSEGDDYIFHCHPPDQKIPKPKRLQEWYKMLDKAVSE
RIVHDYKDI FKQATEDRLTSAKELPYFEGDFWPNVLEESI KELEQEEERKRENTSNESTDVT
KGD SKNAKKNNKTSKNKSSLSRGNKKKPGMPNVSNDLSQKLYATMEKHKEVFFVIRLIAGPA
ANSLPPIVDPDPLIPCDLMDGRDAFLTLARDKHLEFSSLRRAQWSTMCMLVELHTQSODYPYDV
PDYAS

dCas9^{KRAB} (SEQ ID NO: 3)

MDYKDHDGDYKDHDIDYKDDDDKMAPKKRKRKVGGMDDKYSIGLAIGTNSVWAVITDE
YKVP SKKFKVLGNTDRHS IKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSN
EMAKVDDSFHRLEESFLVEEDKKHERHP IFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLR
LIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARL
SKSRRENLI AQLPGEKKNLFGNLI ALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLP
EKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG
SIPHQIHLGELHAILRRQEDFYPLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEET
ITPWNFEVVDKASQAQSFIERMTNFDKNLPNEKVLPHKSLLEYEFTVYNELTKVKYVTEGMRK

PAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGYHDLLKI
 IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSR
 KLINGIRDKQSGKTIIDFLKSDGFANRNFQMLIHDDSLTFKEDIQKAQVSGGDSLHEHIANLA
 GSPAIAKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG
 SQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVAIVPQSFLKDDSIDNKV
 LTRSDKNRGSNDVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGSELKAGFIKQ
 LVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKREINNYHHA
 HDAYLNAVGTALIKKYPKLESEFVYGDYKVDVRKMIKSEQEI GKATAKYFFYSNIMNFFKT
 EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNI VKKTEVQTGGFSKESILP
 KRNSDKLIARKKDWDPKPYGGFDSPTVAYSVLVVAKEGKSKLKS VKELLGITIMERSSEFEK
 NPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLAS
 HYEKLGKSPEDNEQQLFVEQHKHYLDEIEQISEFSKRVILADANLDKLSAYNKHRDKPIRE
 QAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGD
 SRADPKKRRKVASDAKSLTAWSR TLVTFKDV FVDF TREEWKLLDTAQOILYRNVMLENYKNLVS
 LGYQLTKPDVILRLEKGEEPWLVEREIHQETHPDSETAFEIKSSVPKRRKVAS

Nm-dCas9³⁰⁰ Core: (Addgene Plasmid 61365) аминокислотная
 последовательность; Cas9 *Neisseria meningitidis* (D16A, D587A,
H588A, N611A), последовательность ядерной локализации, эффектор
 p300 Core, эпитоп "HA" (SEQ ID NO: 5)

MAAFKPNPINYILGLAIGIASVGWAMVEIDEDENPICLIDLGVRVFERAEVPKTGDSLA
 MARRLARSVRRLTRRAHRLLRARRLLKREGVLQAADFDENGLIKSLPNTPWQLRAAALDRKLT
 PLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVADNAHALQTGDFRTPAELALNKFEK
 ESGHIRNQRGDYSHTFSRKDLQAE LILLFEKQKEFGNPHVSGGLKEGIE TLLMTQRPALSGDAV
 QKMLGHCTFEPAEPKAAKNTYTAERFIWLTCLNLRILEQGSERPLDTERATLMDEPYRKS KL
 TYAQARKLLGLEDTAFFKGLRYGKDNAEASTLMEKAYHAI SRALEKEGLKDKKSPNLNLSPELQ
 DEIGTAFSLFKTDEDITGRLKDRIQPEILEALLKHISFDKQVQISL KALRRIVPLMEQ GKRYDE
 ACAE IYGDHYGKKNTEEKIYLPPIPADEIRNPVVLRLALSQARKVINGVRRY GSPARIHIE TAR
 EVGKSFKDRKEIEKRQEENRKDREKAAAKFREYFPNFVGEPKSKDILKRLRYEQQH GKCLYSGK
 EINLGR LNEKGYVEIAAALPFSRTWDDSFNNKVLVLGSEAQNKGNQTPYEYFNGKDNSREWQEF
 KARVETS RFPFRSKQRILLQKFDEDFGFKERNLNDTRYVNRFLCQFVADRMRLTGK GKRVFASN
 GQITNLLRGFWGLRKVRAENDRHHALDAVVACSTVAMQOKITRFVRYKEMNAFDGKTIDKE TG
 EVLHQKTHFPQWFFAQEV MIRVFGKPDGKPEFEEADTPEKLR TLLAEKLSRPEAVHEVYVTP
 LFVSRAPNRKMSGQGHME TVKSAKRLDEGVSVLRVPLTQLKLDLEKMNREREP KLYEALKAR
 LEAHKDDPAKAFAPFYKYDKAGNRTQVQKAVRVEQVQKTGVVVRNHNGIADNATMVRVDVFEK
 GDKYYLVPIYSWQVAKGILPDRAVVQKDEEDWQLIDDSFNFKFSLHPNDLVEVITKKARMGY
 FASCHRGTGNINIRIHDLDHKIGKNGILEGIGVKTALSFQKYQIDELGKEIRPCRLKRRPPVRS
 RADPKKRRKVEASGRAIFKPEELRQALMPTLEALYRQDPESLPFRQFVDPQLLGI PDYFDIVKS
 PMDLSTIKRKLDTGQYQEPWQYVDDIWL MFNNAWLYNRKTSRVYKYCSKLSSEVFEQEIDPVMQS
 LGYCCGRKLEFSPQTLCCYQKQLCTIPRDATYYSYQNRHYHFCEKCFNEIQGESVSLGDDPSQPQ
 TITNKEQFSKRKNDTLDPELFVECTECGRKMHIQICVLHHEI IWPAGFVCDGCLKKSARTKENK
 FSAKRLPSTRLGTFLNRVNDFLRRQNHPESEVTVRVVHASDKTVEVKPMKARFVDSGEMAE
 SFPYRTKALFAFEEIDGVDLCFFGMHVQEYGSDCPPPNQRRVYISYLDVHFFRPKCLRTAVYH
 EILIGYLEYVKKLGYTTGHIWACPPSEGDDYIFHCHPPDQKIPKPKRLQEWYKMLDKAVSERI
 VHDYKIDFKQATEDRLTSAKELPYFEGDFWPNVLEESIKELEQEEERKREENTSNESTDVTKG
 DSKNAKKNNKTSKNKSSLSRGNKKKPGMPNVSNDLSQKLYATMEKHKEVFFVIRLIAGPAAN
 SLPPIVDPDPLIPCDLMDGRDAFLTLARDKHLEFSSLRRAQWSTMCMLVELHTQSQDY PYDVPD
YAS

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Способ получения оптимизированной направляющей РНК (gRNA), при этом способ предусматривает:
 - а) идентификацию представляющей интерес целевой области, при этом представляющая интерес

целевая область содержит последовательность протоспейсера;

b) определение полинуклеотидной последовательности полноразмерной gRNA, которая нацеливается на представляющую интерес целевую область, при этом полноразмерная gRNA содержит нацеливающиеся на протоспейсер последовательность или сегмент;

c) определение по меньшей мере одного или нескольких нецелевых сайтов для полноразмерной gRNA;

d) получение полинуклеотидной последовательности первой gRNA, при этом первая gRNA содержит полинуклеотидную последовательность полноразмерной gRNA и сегмент РНК, при этом сегмент РНК содержит полинуклеотидную последовательность, имеющую М нуклеотидов в длину, которая комплементарна нуклеотидному сегменту нацеливающихся на протоспейсер последовательности или сегмента, при этом сегмент РНК расположен на 5'-конце или 3'-конце полинуклеотидной последовательности полноразмерной gRNA, при этом первая gRNA способна встраиваться в последовательность протоспейсера и связываться с последовательностью ДНК, которая комплементарна последовательности протоспейсера, и образовывать дуплекс с протоспейсером, и при этом первая gRNA способна встраиваться в нецелевой сайт и связываться с последовательностью ДНК, которая комплементарна нецелевому сайту, и образовывать нецелевой дуплекс;

e) вычисление оценочного показателя или вычислительное моделирование кинетики встраивания и времени нахождения, в течение которого первая gRNA остается встроенной в дуплексы протоспейсера и нецелевого сайта, где кинетику встраивания и время нахождения оценивают нуклеотид за нуклеотидом путем определения различий свободной энергии между дополнительным встраиванием отличающейся gRNA и повторным отжигом первой gRNA с последовательностью ДНК, которая комплементарна последовательности протоспейсера;

f) сравнение значений предполагаемого времени нахождения в сайтах протоспейсера и/или нецелевых сайтах первой gRNA со значениями предполагаемого времени нахождения полноразмерной gRNA или усеченной gRNA (tru-gRNA) в сайтах протоспейсера и/или нецелевых сайтах;

g) рандомизацию от 1 до М нуклеотидов в сегменте РНК первой gRNA и получение второй gRNA, и повторение стадии (e) со второй gRNA;

h) идентификацию оптимизированной gRNA, где оптимизированная gRNA имеет продолжительность связывания, которая больше или равна продолжительности связывания полноразмерной gRNA в протоспейсере, и/или продолжительность связывания была меньше или равная продолжительности связывания полноразмерной gRNA в нецелевом сайте; и

i) тестирование оптимизированной gRNA *in vivo* для определения специфичности связывания.

2. Способ по п.1, где на стадии (d) первая gRNA дополнительно содержит линкер между 5'-концом или 3'-концом полинуклеотидной последовательности полноразмерной gRNA и сегментом РНК, при этом линкер содержит полинуклеотидную последовательность, имеющую N нуклеотидов в длину, и где на стадии (g) способ дополнительно включает рандомизацию от 1 до N нуклеотидов в линкере.

3. Способ по п.1, где свободную энергию дополнительного встраивания отличающейся gRNA определяют путем определения свободной энергии по меньшей мере одного из: (I) нарушения спаривания оснований ДНК-ДНК, (II) образования пар оснований РНК-ДНК, (III) различия свободной энергии, возникающего в результате разрушения или образования отличающейся вторичной структуры внутри не встроенной направляющей РНК, и (IV) образования или разрушения взаимодействий между вытесненной нитью ДНК, которая комплементарна протоспейсеру, и любыми неспаренными нуклеотидами направляющей РНК, которые не вовлечены во вторичные структуры.

4. Способ по п.3, где свободная энергия повторного отжига первой gRNA с последовательностью ДНК, которая комплементарна последовательности протоспейсера, определяют путем определения свободной энергии по меньшей мере одного из: (I) образования пар оснований ДНК-ДНК, (II) разрушения пар оснований РНК-ДНК, (III) различия свободной энергии, возникающего в результате разрушения или образования отличающейся вторичной структуры внутри новой не встроенной направляющей РНК, и (IV) образования или разрушения взаимодействий между вытесненной нитью ДНК, которая комплементарна протоспейсеру, и любыми неспаренными нуклеотидами направляющей РНК, которые не вовлечены во вторичные структуры.

5. Способ по п.4, дополнительно предусматривающий определение свободной энергии по меньшей мере одного из (V) спаривания оснований среди ошибочных спариваний, (VI) взаимодействий с белком Cas9 и/или (VII) дополнительных эвристических показателей, где дополнительные эвристические показатели относятся к продолжительности связывания, степени встраивания, стабильности встраивающейся направляющей РНК или другим расчетным/моделируемым свойствам встраивания gRNA для расщепляющей активности Cas9.

6. Способ по п.2, где полноразмерная gRNA содержит от 15 до 20 нуклеотидов, М составляет от 1 до 20, сегмент РНК содержит от 2 до 15 нуклеотидов, которые комплементарны нацеливающейся на протоспейсер последовательности, N составляет от 1 до 20 или сегмент РНК и/или нацеливающаяся на протоспейсер последовательность обеспечивают вторичную структуру.

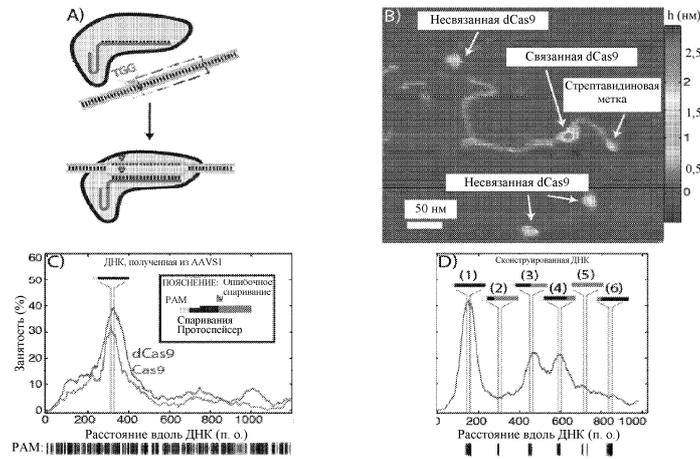
7. Способ по п.1, где оптимизированную gRNA применяют с системой на основе CRISPR/Cas9 или

с системой на основе CRISPR/Cpf1 в клетке.

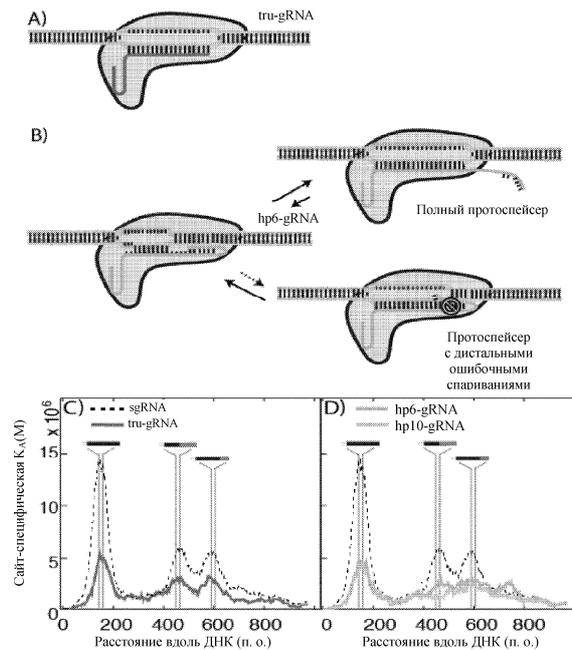
8. Способ по п.2, где 1-20 нуклеотидов рандомизированы в линкере, 1-20 нуклеотидов рандомизированы в сегменте РНК, стадию (g) повторяют X раз, с получением тем самым gRNA в количестве X, и повторяют стадию (e) с каждым количеством X gRNA, где X составляет от 0 до 20, кинетику встраивания и время нахождения вычисляют с применением кинетической методики Монте-Карло или алгоритма Гиллеспы, кинетика встраивания представляет собой скорость, с которой направляющая РНК встраивается в дуплекс протоспейсера с обеспечением полного встраивания таким образом, что встраивание в протоспейсер является полностью осуществленным, и/или скорость, с которой сегмент ДНК протоспейсера, связанной с gRNA, удлиняется по мере его вытеснения из его комплементарной нити и связывания с gRNA нуклеотид за нуклеотидом от его РАМ-проксимальной области до полного встраивания, или критерии конструирования предусматривают специфичность, модуляцию продолжительности связывания и/или расчетную специфичность расщепления.

9. Способ по п.1, где сегмент РНК и нацеливающаяся на протоспейсер последовательность или сегмент обеспечивают вторичную структуру,

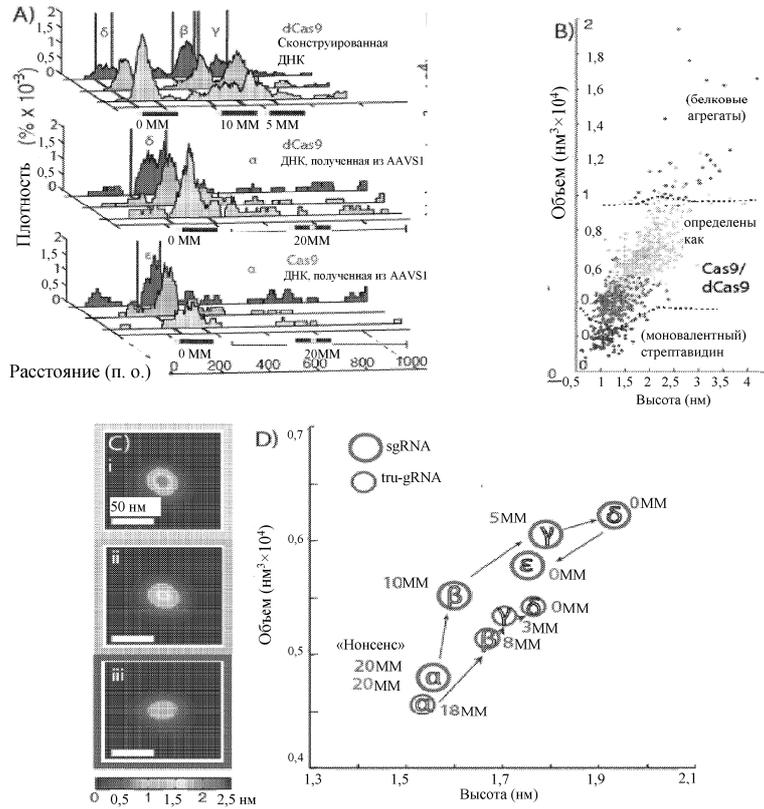
где вторичная структура образуется путем гибридизации всего или части сегмента РНК с нуклеотидами на 5'-конце нацеливающейся на протоспейсер последовательности или сегмента, нуклеотидами в середине нацеливающейся на протоспейсер последовательности или сегмента и/или нуклеотидами на 3'-конце нацеливающейся на протоспейсер последовательности или сегмента, и где вторичная структура представляет собой шпильку.



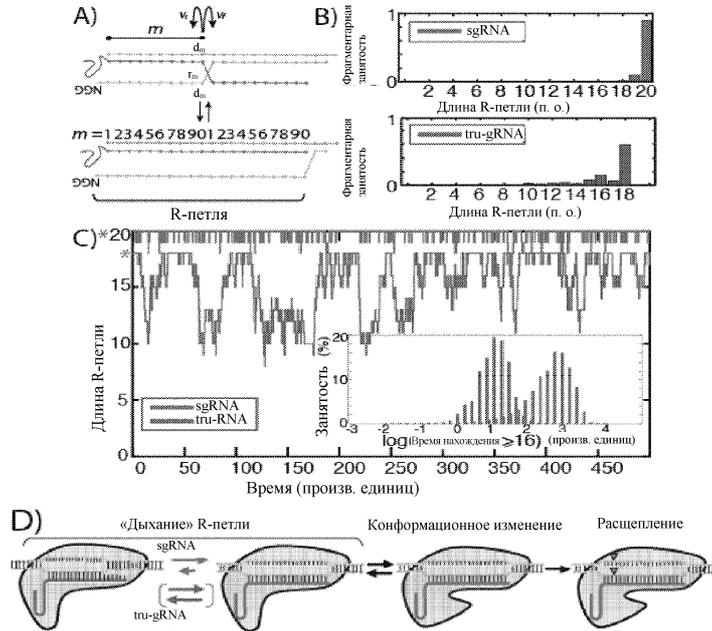
Фиг. 1A-1D



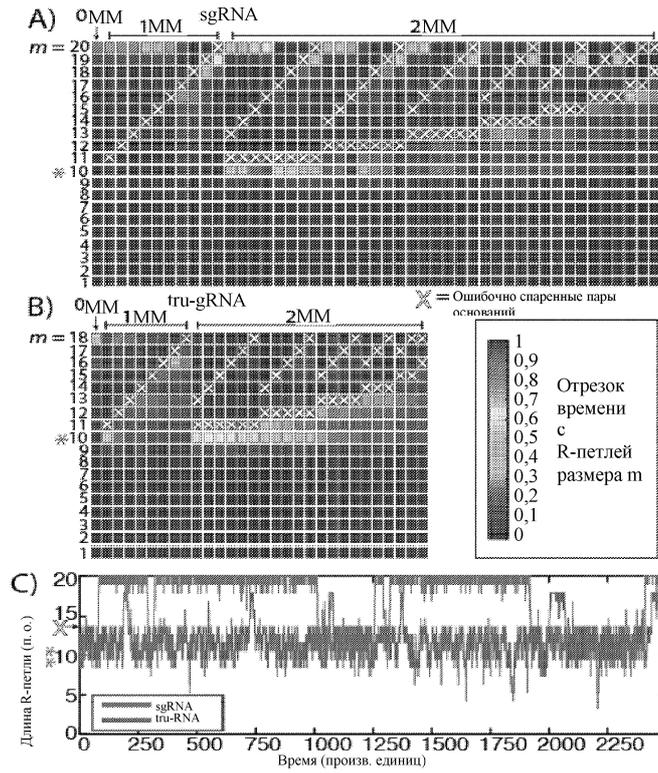
Фиг. 2A-D



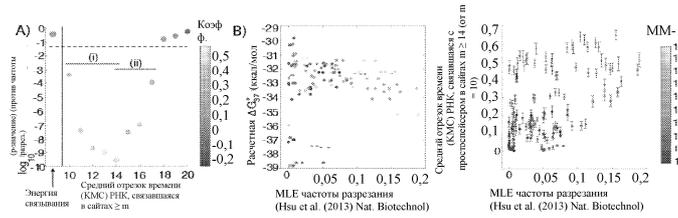
Фиг. 3C-D



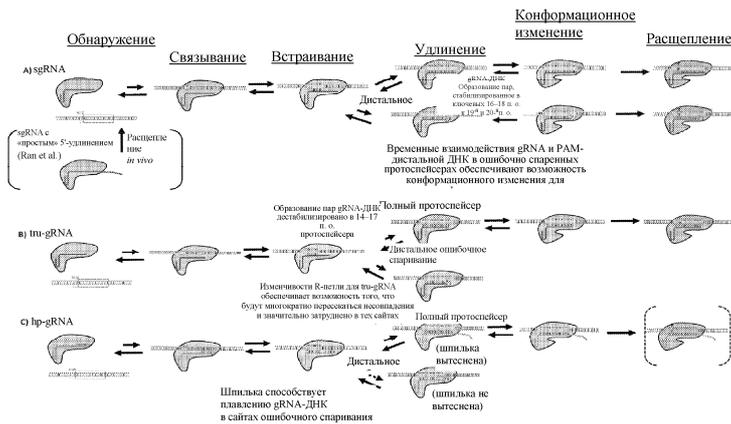
Фиг. 4A-4D



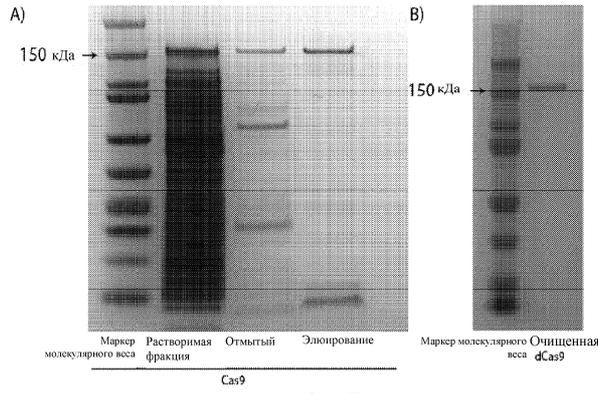
Фиг. 5A-C



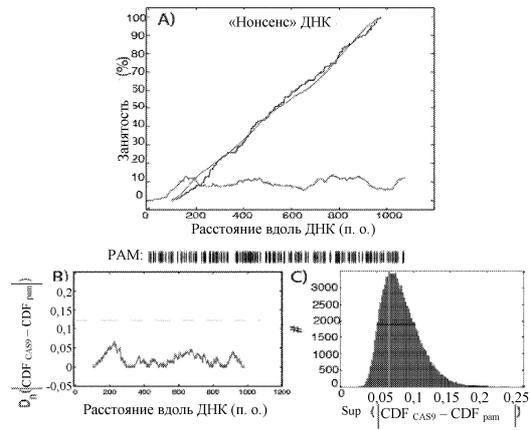
Фиг. 6A-B



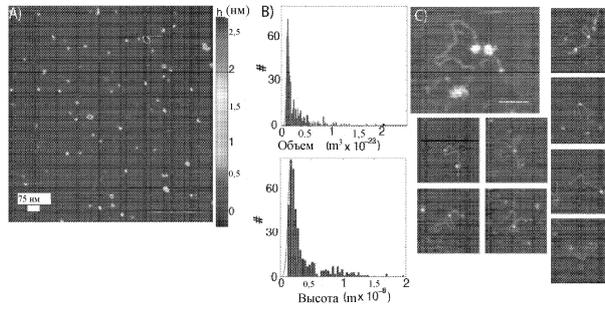
Фиг. 7A-C



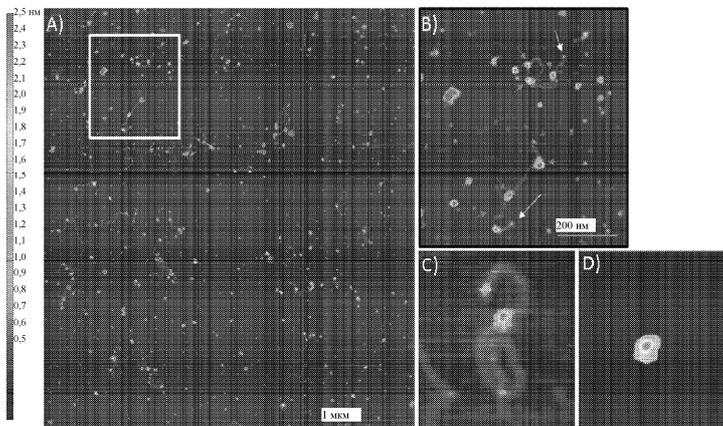
Фиг. 8А-В



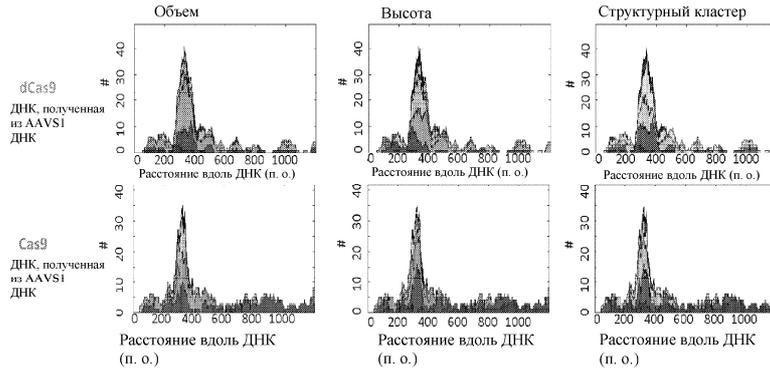
Фиг. 9А-С



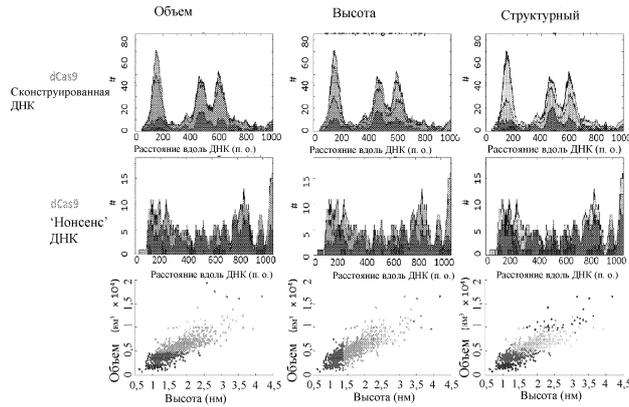
Фиг. 10А-С



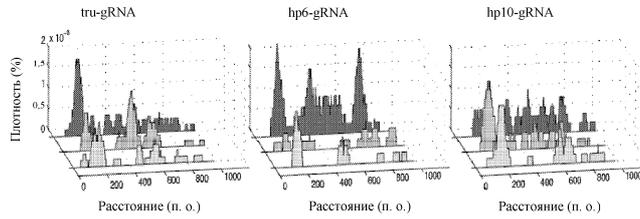
Фиг. 11А-Д



Фиг. 12А

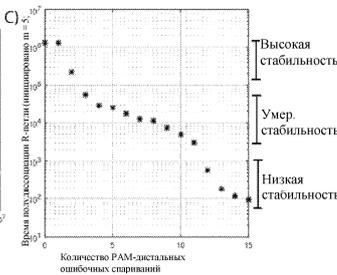
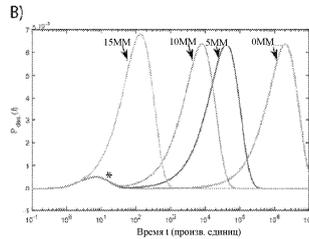
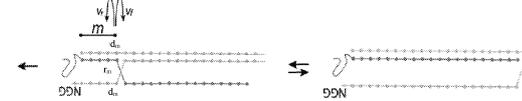


Фиг. 12В

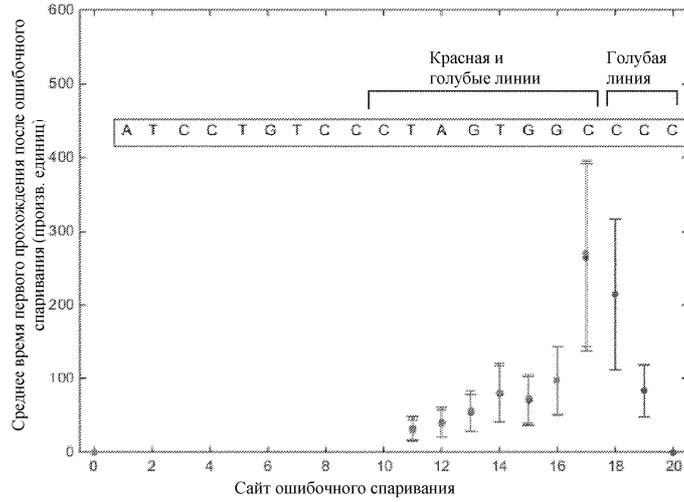


Фиг. 13

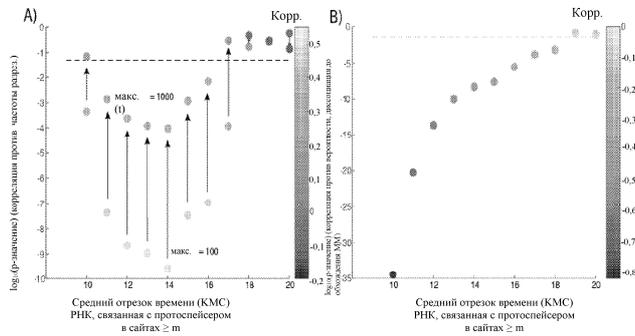
А) Модель встраивания нити протоспейсера ДНК с помощью направляющей РНК



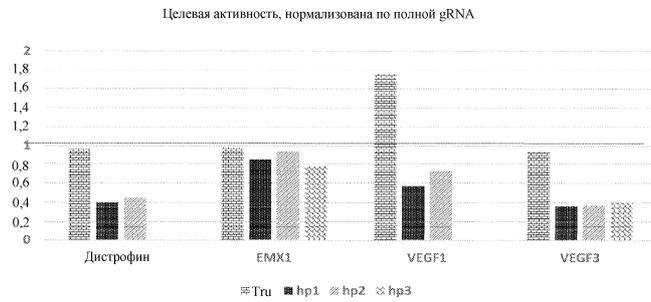
Фиг. 14А-С



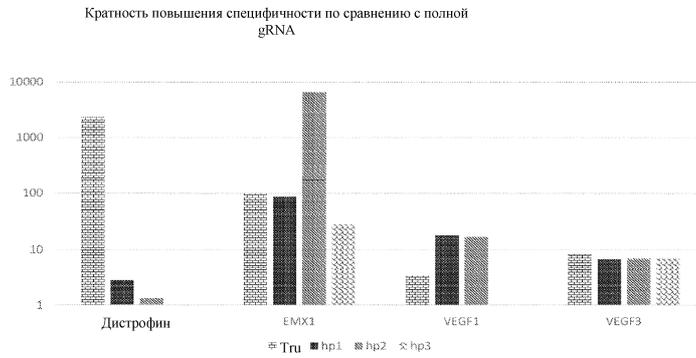
Фиг. 15



Фиг. 16А - 16В

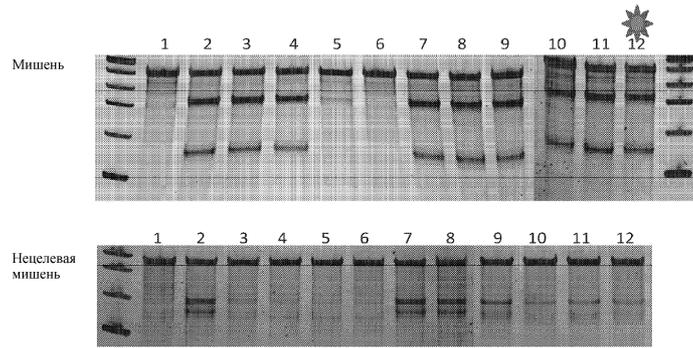


Фиг. 17

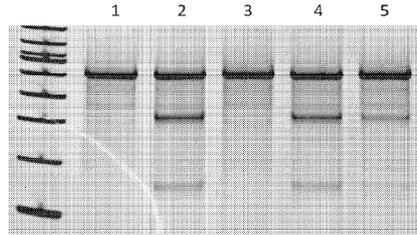


Фиг. 18

046214



Фиг. 19



Дорожка	Контроль.	SEQ ID NO:
Дорожка 1	Контроль.	
Дорожка 2	Полная	
Дорожка 3	Dysint_4L_4H_S	g agt TCCT ACTCAGACTGTTACTC 285
Дорожка 4	Dysint_4L_5H_S, шпилька 2	g tgagt TCCT ACTCAGACTGTTACTC 286
Дорожка 5	Dysint_4L_8H_S	g tctgagt TCCT ACTCAGACTGTTACTC 287

Фиг. 20

Шпилька 1-
9 п. о., 5'-шпилька

Структура MFE при



Вероятность возникновения равновесия

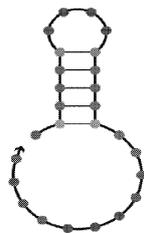


Свободная энергия вторичной структуры: - 12,50 ккал/моль

Фиг. 21А

Шпилька 2-
5 п. о., 5'-шпилька

Структура MFE при 37,0°C

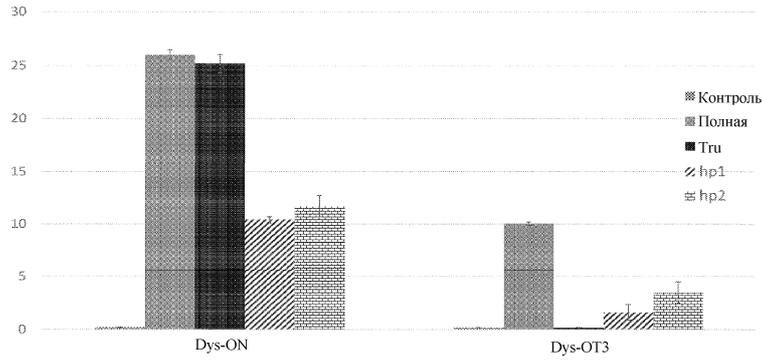


Вероятность возникновения равновесия

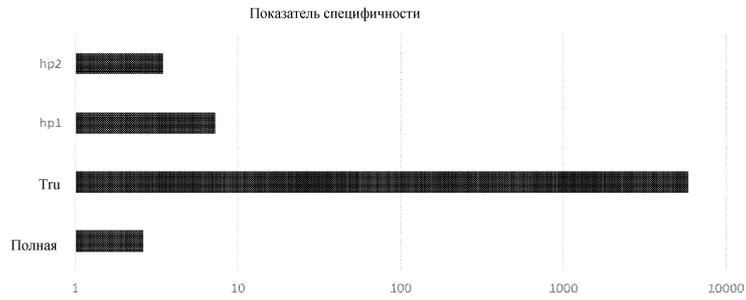


Свободная энергия вторичной структуры: - 4,50 ккал/моль

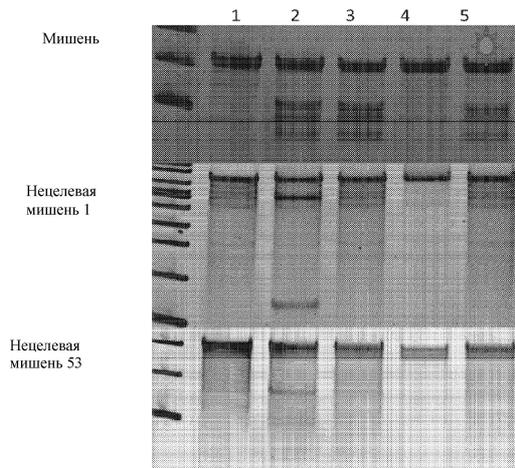
Фиг. 21В



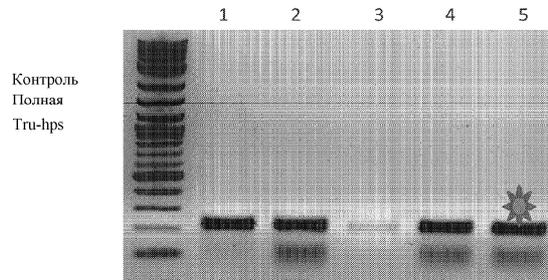
Фиг. 22



Фиг. 23



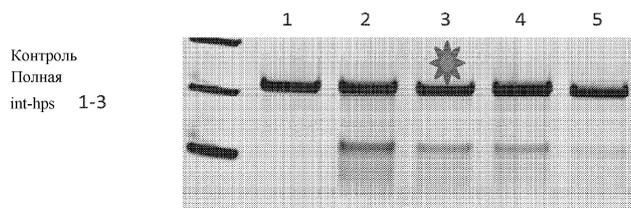
Фиг. 24



Дорожка	SEQ ID NO:	Sequence	SEQ ID NO:
Дорожка 1	DYS1_3hp-T18_S	G ta ttcg TACTCAGACTGTTACTC	288
Дорожка 2	DYS1_5hp-T18_S	G agta ttcg TACTCAGACTGTTACTC	289
Дорожка 3	EMX1_3hp-T18_S	g ac ttcg gtccgagcagaagaaga	290
Дорожка 4	EMX1_4hp-T18_S	gg ac ttcg gtccgagcagaagaaga	291
Дорожка 5	EMX1_5hp-T18_S, Шпилька 2	g cgg ac ttcg gtccgagcagaagaaga	292

Фиг. 25А

046214

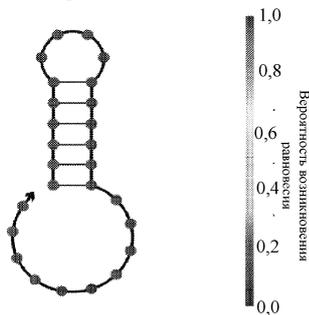


SEQ ID NO:

Дорожка 1	GFP Контроль			
Дорожка 2	Полная			
Дорожка 3	EMX1int_4L_3H_5.Шпилька 3	g	cgg GAGT CCGAGCAGAAGAAGAA	293
Дорожка 4	EMX1int_4L_4H_5	g	tcgg GAGT CCGAGCAGAAGAAGAA	294
Дорожка 5	EMX1int_4L_6H_5	g	ctcgg GAGT CCGAGCAGAAGAAGAA	295

Фиг. 25В

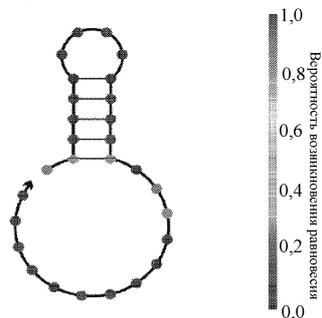
Структура MFE при 37,0°C



Свободная энергия вторичной структуры: - 10,60 ккал/моль

Фиг. 26А

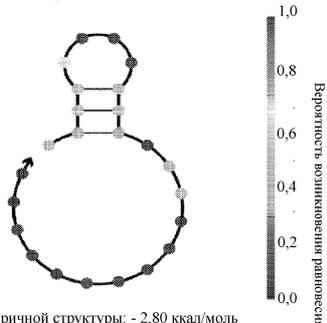
Структура MFE при



Свободная энергия вторичной структуры: - 9,10 ккал/моль

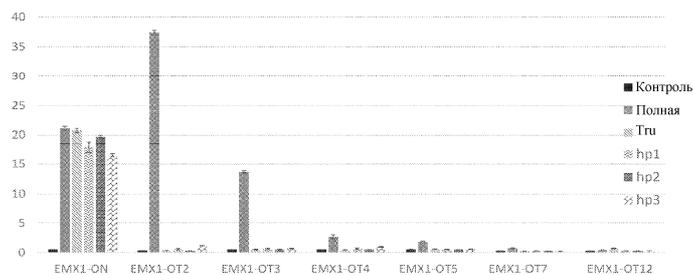
Фиг. 26В

Структура MFE при 37,0°C

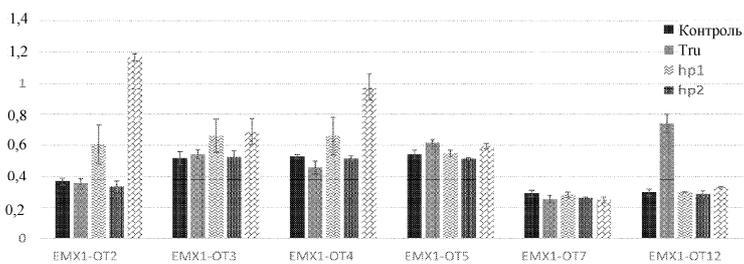


Свободная энергия вторичной структуры: - 2,80 ккал/моль

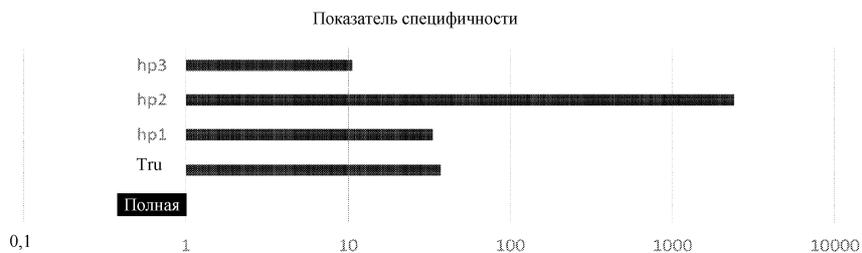
Фиг. 26С



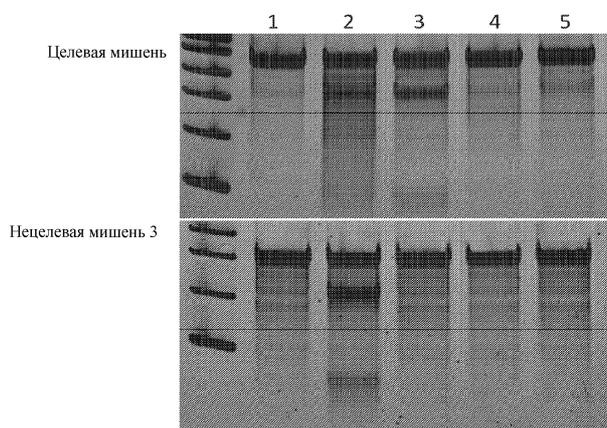
Фиг. 27



Фиг. 28

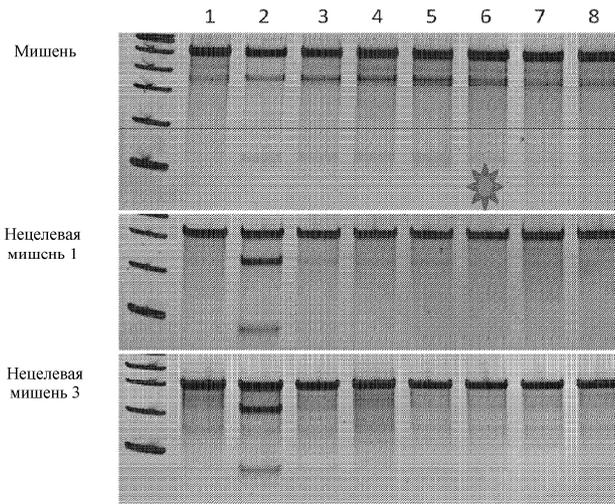


Фиг. 29



Фиг. 30

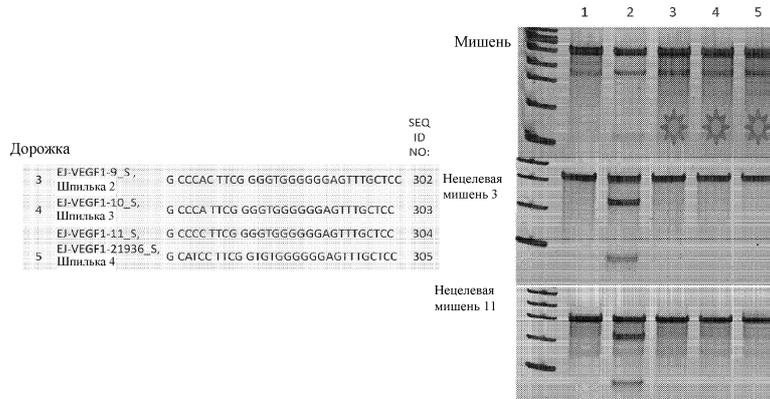
046214



Фиг. 31А

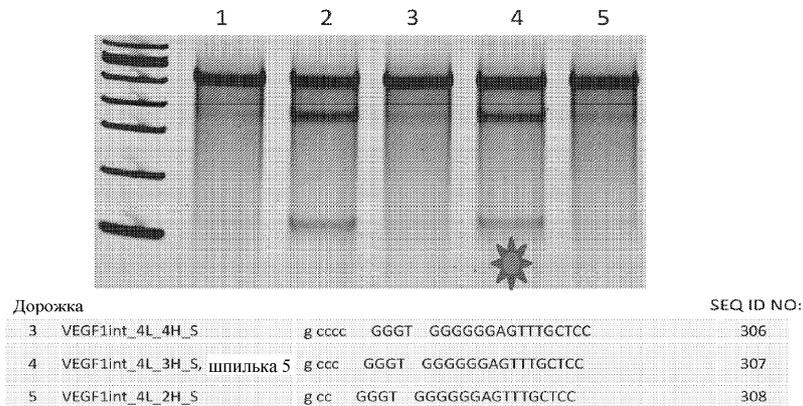
Дорожка		SEQ ID NO:
3	EJ-VEGF1-1_5	G TTCAC TTCG GGGTGGGGGAGTTTGCTCC 296
4	EJ-VEGF1-2_5	G TTCA TTCG GGGTGGGGGAGTTTGCTCC 297
5	EJ-VEGF1-3_5	G TCC TTCG GGGTGGGGGAGTTTGCTCC 298
6	EJ-VEGF1-4_5 шпилька1	G CTCC TTCG GGGTGGGGGAGTTTGCTCC 299
7	EJ-VEGF1-5_5	G TTCC TTCG GGGTGGGGGAGTTTGCTCC 300
8	EJ-VEGF1-6_5	G CTCC TTCG GGGTGGGGGAGTTTGCTCC 301

Фиг. 31В



Дорожка		SEQ ID NO:
3	EJ-VEGF1-9_5, Шпилька 2	G CCCAC TTCG GGGTGGGGGAGTTTGCTCC 302
4	EJ-VEGF1-10_5, Шпилька 3	G CCCA TTCG GGGTGGGGGAGTTTGCTCC 303
5	EJ-VEGF1-11_5, Шпилька 4	G CCCC TTCG GGGTGGGGGAGTTTGCTCC 304
5	EJ-VEGF1-21936_5, Шпилька 4	G CAICC TTCG GTGTGGGGGAGTTTGCTCC 305

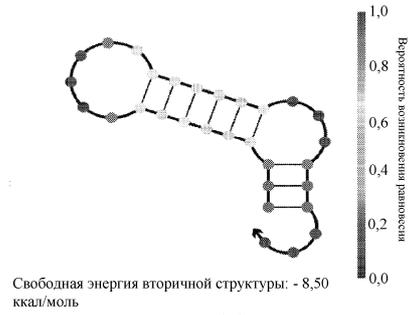
Фиг. 32



Дорожка		SEQ ID NO:
3	VEGF1int_4L_4H_5	g cccc GGGT GGGGGGAGTTTGCTCC 306
4	VEGF1int_4L_3H_5, шпилька 5	g ccc GGGT GGGGGGAGTTTGCTCC 307
5	VEGF1int_4L_2H_5	g cc GGGT GGGGGGAGTTTGCTCC 308

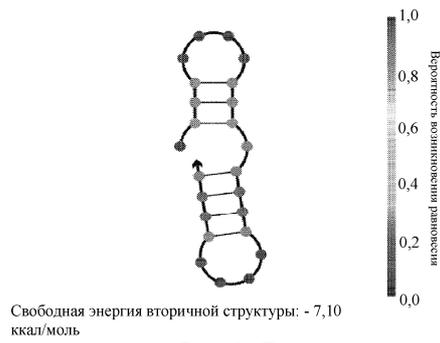
Фиг. 33

Структура MFE при 37,0°C

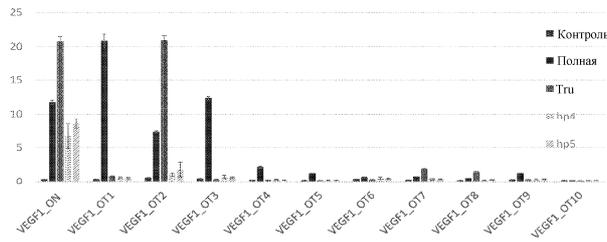


Фиг. 34А

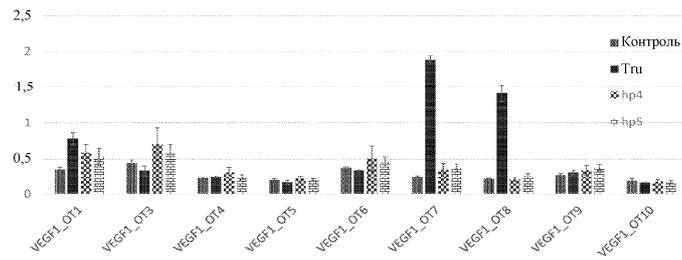
Структура MFE при 37,0°C



Фиг. 34В

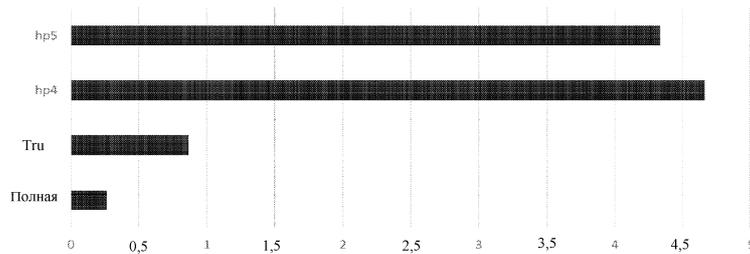


Фиг. 35



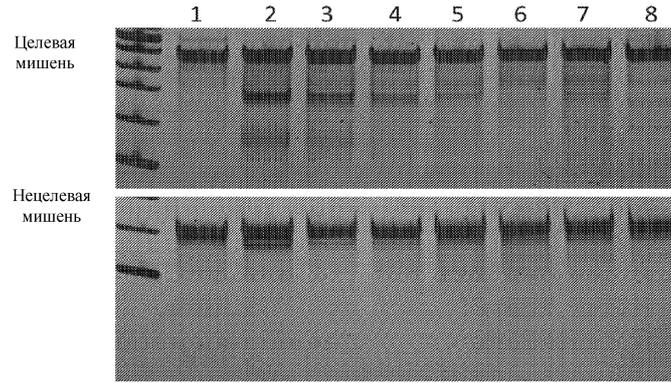
Фиг. 36

Показатель специфичности

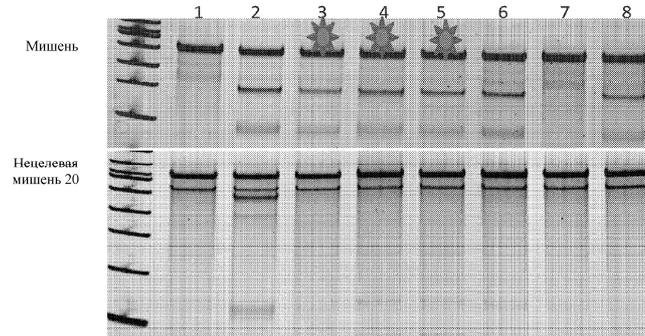


Фиг. 37

046214



Фиг. 38



Преобразование

Surv_OT20=DS_OT2

Фиг. 39А

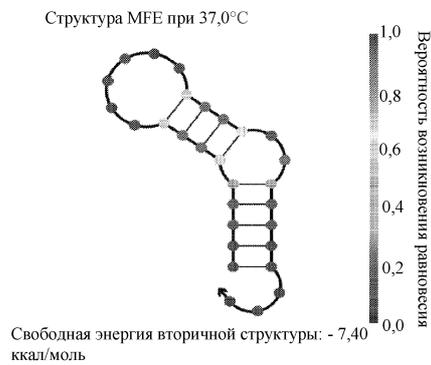
Дорожка

SEQ ID NO:

3	EJ-VEGF3-1_S_hp3	G CACG TTCG GGTGAGTGAGTGTGCGTG	309
4	EJ-VEGF3-2_S_hp2	G TACG TTCG GGTGAGTGAGTGTGCGTG	310
5	EJ-VEGF3-3_S_hp3	G CATG TTCG GGTGAGTGAGTGTGCGTG	311
6	EJ-VEGF3-4_S	G TATG TTCG GGTGAGTGAGTGTGCGTG	312
7	EJ-VEGF3-5_S	G CGCG TTCG GGTGAGTGAGTGTGCGTG	313
8	EJ-VEGF3-6_S	G CGTG TTCG GGTGAGTGAGTGTGCGTG	314

Полная	22,17609557	Полная	23,44021291
E1	14,62010307	E17	12,8315672
E2	17,96688778	E18	16,5260316
E3	15,45320881	E19	14,90816768
E14	18,09621804	E110	15,88223119
E15	0	E111	16,7829789
E16	11,03398795	E112	10,77677544

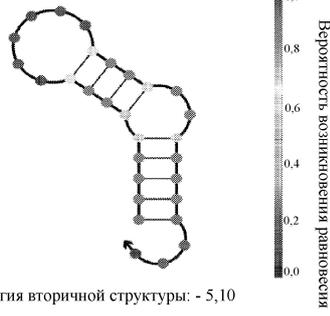
Фиг. 39В



Фиг. 40А

046214

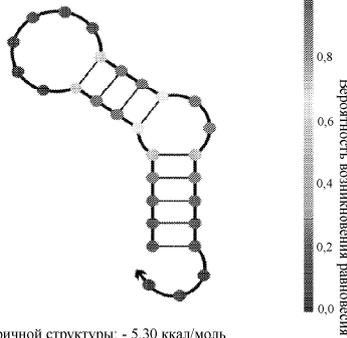
Структура MFE при 37,0°C



Свободная энергия вторичной структуры: - 5,10 ккал/моль

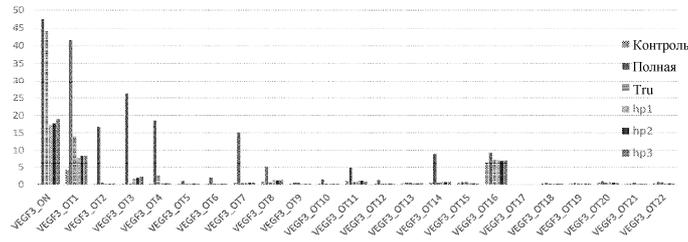
Фиг. 40В

Структура MFE при 37,0°C

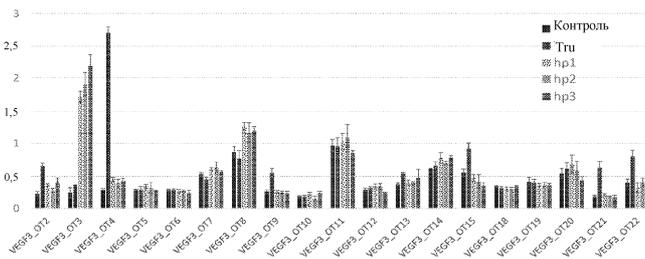


Свободная энергия вторичной структуры: - 5,30 ккал/моль

Фиг. 40С

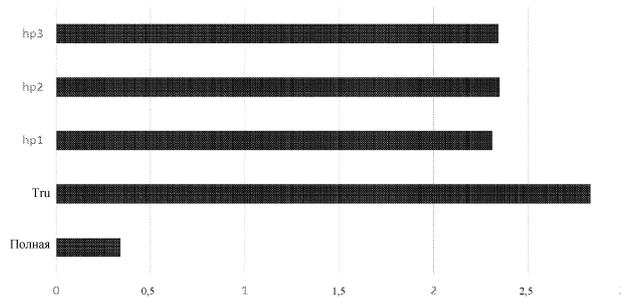


Фиг. 41

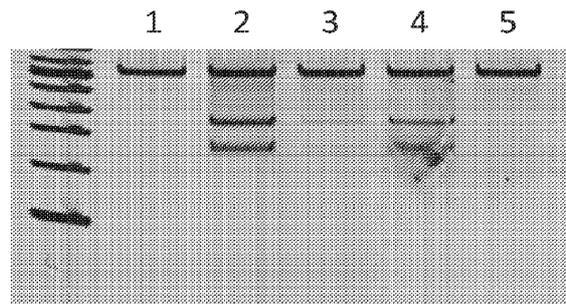


Фиг. 42

Показатель специфичности



Фиг. 43



Фиг. 48

