

(19)



Евразийское
патентное
ведомство

(21) 202191875 (13) A1

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ

(43) Дата публикации заявки
2022.04.04

(51) Int. Cl. C12N 15/90 (2006.01)
C12N 9/22 (2006.01)

(22) Дата подачи заявки
2020.02.03

(54) ИНГИБИРОВАНИЕ НЕПРЕДУСМОТРЕННЫХ МУТАЦИЙ ПРИ РЕДАКТИРОВАНИИ ГЕНОВ

(31) PCT/CN2019/074577

(32) 2019.02.02

(33) CN

(86) PCT/CN2020/074218

(87) WO 2020/156575 2020.08.06

(71) Заявитель:
ШАНХАЙТЕК ЮНИВЕРСИТИ (CN)

(72) Изобретатель:

Чэнь Цзя, Ян Бэй, Ян Ли, Хуан
Синсюй, Ван Лицзе (CN)

(74) Представитель:

Нилова М.И. (RU)

(57) Предложены слитые белки и родственные молекулы, полезные для осуществления редактирования оснований со сниженной частотой или отсутствием нецелевых мутаций. Слитые белки могут содержать первый фрагмент, содержащий дезаминазу азотистых оснований или ее каталитический домен, второй фрагмент, содержащий ингибиторный домен дезаминазы азотистых оснований, и сайт расщепления протеазой между первым фрагментом и вторым фрагментом. Также предложены улучшенные системы праймированного редактирования, содержащие направляющую РНК для праймированного редактирования с улучшенной стабильностью.

A1

202191875

202191875

A1

ИНГИБИРОВАНИЕ НЕПРЕДУСМОТРЕННЫХ МУТАЦИЙ ПРИ РЕДАКТИРОВАНИИ ГЕНОВ

УРОВЕНЬ ТЕХНИКИ

[0001] Редактирование генома представляет собой тип генной инженерии, при котором ДНК вставляют, удаляют или заменяют в геноме живого организма, применяя сконструированные нуклеазы (молекулярные ножницы). Инструменты редактирования генома для генетических манипуляций с геномом клеток и живым организмом находят широкое применение в исследованиях в области медико-биологических наук, разработках в области биотехнологии/агротехники и, важнее всего, в фармацевтических/клинических инновационных разработках. Например, редактирование генома можно применять для исправления «драйверных» мутаций, лежащих в основе генетических заболеваний, приводя к полному излечению от данных заболеваний у живых организмов. Редактирование генома также можно применять для конструирования генома сельскохозяйственных культур, повышения урожайности сельскохозяйственных культур и придания сельскохозяйственным культурам устойчивости к загрязнениям окружающей среды или инфицированию патогенами. Кроме того, трансформация генома микроорганизмов посредством точного редактирования генома имеет огромное значение для разработки возобновляемых источников биологической энергии.

[0002] Система CRISPR/Cas (короткие палиндромные повторы, регулярно расположенные группами/CRISPR-ассоциированный белок) стала наиболее эффективным инструментом редактирования генома с момента ее изобретения благодаря эффективности ее уникального редактирования, удобству и потенциальным применениям в живых организмах. Направляемая направляющей РНК (нРНК), нуклеаза Cas может создавать двухцепочечные разрывы (ДЦР) ДНК в целевых участках генома в различных клетках (как в линиях клеток, так и в клетках из живых организмов). Данные ДЦР затем репарируются эндогенной системой репарации ДНК, которую можно использовать для осуществления желательного редактирования генома.

[0003] Как правило, два основных пути репарации ДНК могут активироваться ДЦР: негомологичное соединение концов (НГСК) и направляемая гомологией репарация (НГР). НГСК может ввести произвольные вставки/делеции (indels) в участок геномной

ДНК вокруг ДЦР, приводя к сдвигу открытой рамки считывания (ОРС) и, в конечном счете, к инаktivации гена. Напротив, когда запускается НГР, последовательность геномной ДНК в целевом сайте можно заменить на последовательность матрицы экзогенной донорной ДНК посредством механизма гомологичной рекомбинации, что может привести к исправлению генетической мутации. Тем не менее, практическая эффективность опосредованного НГР исправления генов низка (обычно < 5%), поскольку возникновение гомологичной рекомбинации как специфично в отношении типа клеток, так и зависимо от клеточного цикла, и НГСК запускается чаще, чем НГР. Относительно низкая эффективность НГР, следовательно, ограничила переход инструментов редактирования генома CRISPR/Cas в область точной генотерапии (коррекции вызывающих заболевания генов).

[0004] Недавно были разработаны редакторы оснований (BE), в которых объединена система CRISPR/Cas с семейством APOBEC (подобный каталитическому полипептиду белок, редактирующий мРНК аполипопротеина В)/AID (индуцированная активацией цитидиндезаминаза), которые сильно повышают эффективность опосредованной CRISPR/Cas коррекции генов. Посредством соединения с нисказой Cas9 (nCas9) или каталитически неактивным Cpf1 (dCpf1 также известным как dCas12a), активность дезаминирования цитозина (C) представителями семейства APOBEC/AID можно специально направить на целевые основания в геноме и катализировать замены C на тимин (T) в данных основаниях.

[0005] Тем не менее, так как представители семейства APOBEC/AID могут вызывать мутации с заменой основания C на T в участках одноцепочечной ДНК (оцДНК), специфичность современной системы редактирования оснований ненадежна, что ограничивает применения, например, применение BE для обращения мутации T на C, которая приводит к заболеваниям человека, в терапевтических целях. Следовательно, требуется создание новых BE, которые могут специфично редактировать цитозины на целевом участке, но не вызывают мутации C на T на других участках оцДНК. Такие новые BE позволят осуществить более специфичное редактирование оснований в различных живых организмах. Важно, что высокая специфичность таких BE будет стимулировать потенциальный переход в клиническую практику, в частности, в области генотерапии, которая включает обращение связанных с заболеванием мутаций T на C.

КРАТКОЕ ОПИСАНИЕ ИЗОБРЕТЕНИЯ

[0006] В некоторых вариантах реализации настоящего изобретения предложены редакторы оснований, пригодные для редактирования генома, которые вызывают меньшее количество или даже не вызывают нецелевых мутаций, характерных для современных редакторов оснований. В некоторых вариантах реализации ингибитор дезаминазы азотистых оснований соединен расщепляемой связью с дезаминазой азотистых оснований, участвующей в редактировании генома. В присутствии ингибитора дезаминазы азотистых оснований, дезаминаза азотистых оснований неспособна (менее способна) реагировать с молекулой нуклеотида. В целевом положении редактирования ингибитор дезаминазы азотистых оснований может быть отщеплен, высвобождая полностью активную дезаминазу азотистых оснований, которая затем может необходимым образом осуществить редактирование.

[0007] Соответственно, в одном варианте реализации предложен слитый белок, содержащий: первый фрагмент, содержащий дезаминазу азотистых оснований или ее каталитический домен, второй фрагмент, содержащий ингибитор дезаминазы азотистых оснований, и сайт расщепления протеазой между первым фрагментом и вторым фрагментом.

[0008] В некоторых вариантах реализации дезаминаза азотистых оснований представляет собой аденозиндезаминазу. В некоторых вариантах реализации аденозиндезаминаза выбрана из группы, состоящей из тРНК-специфичной аденозиндезаминазы (TadA), тРНК-специфичной аденозиндезаминазы 1 (ADAT1), тРНК-специфичной аденозиндезаминазы 2 (ADAT2), тРНК-специфичной аденозиндезаминазы 3 (ADAT3), РНК-специфичной аденозиндезаминазы B1 (ADARB1), РНК-специфичной аденозиндезаминазы B2 (ADARB2), аденозинмонофосфат-дезаминазы 1 (AMPD1), аденозинмонофосфат-дезаминазы 2 (AMPD2), аденозинмонофосфат-дезаминазы 3 (AMPD3), аденозиндезаминазы (ADA), аденозиндезаминазы 2 (ADA2), подобного аденозиндезаминазе белка (ADAL), содержащего домен аденозиндезаминазы белка 1 (ADAD1), содержащего домен аденозиндезаминазы белка 2 (ADAD2), РНК-специфичной аденозиндезаминазы (ADAR) и РНК-специфичной аденозиндезаминазы B1 (ADARB1).

[0009] В некоторых вариантах реализации дезаминаза азотистых оснований представляет собой цитидиндезаминазу. В некоторых вариантах реализации цитидиндезаминаза выбрана из группы, состоящей из АРОВЕС3В (А3В), АРОВЕС3С (А3С), АРОВЕС3Д (А3Д), АРОВЕС3F (А3F), АРОВЕС3G (А3G), АРОВЕС3Н (А3Н), АРОВЕС1 (А1), АРОВЕС3 (А3), АРОВЕС2 (А2), АРОВЕС4 (А4) и АICDA (АID). В некоторых вариантах реализации цитидиндезаминаза представляет собой цитидиндезаминазу человека или мыши. В некоторых вариантах реализации каталитический домен представляет собой домен 1 цитидиндезаминазы А3 мыши (CDA1) или домен 2 цитидиндезаминазы А3В человека (CDA2).

[0010] В некоторых вариантах реализации ингибитор дезаминазы азотистых оснований представляет собой ингибиторный домен дезаминазы азотистых оснований. В некоторых вариантах реализации ингибитор дезаминазы азотистых оснований представляет собой ингибиторный домен цитидиндезаминазы. В некоторых вариантах реализации ингибитор дезаминазы азотистых оснований представляет собой ингибиторный домен аденозиндезаминазы. В некоторых вариантах реализации ингибитор дезаминазы азотистых оснований содержит последовательность аминокислот, выбранную из SEQ ID NO: 1 - 2 и последовательностей в таблицах 1 и 2 (SEQ ID NO: 48 - 135), или последовательность аминокислот, идентичную по меньшей мере на 85% любой из последовательностей аминокислот, выбранных из SEQ ID NO: 1 - 2 и последовательностей в таблицах 1 и 2. В некоторых вариантах реализации ингибитор дезаминазы азотистых оснований содержит последовательность аминокислот SEQ ID NO: 1, остатки аминокислот АК76 - АК149 из последовательности SEQ ID NO: 1 или последовательность аминокислот SEQ ID NO: 2.

[0011] В некоторых вариантах реализации первый фрагмент дополнительно содержит белок, ассоциированный с короткими палиндромными повторами, регулярно расположенными группами (CRISPR) (Cas). В некоторых вариантах реализации белок Cas выбран из группы, состоящей из SpCas9, FnCas9, St1Cas9, St3Cas9, NmCas9, SaCas9, AsCpf1, LbCpf1, FnCpf1, VQR SpCas9, EQR SpCas9, VRER SpCas9, xSpCas9, SpCas9-NG, RHA FnCas9, KKH SaCas9, NmeCas9, StCas9, CjCas9, AsCpf1, FnCpf1, SsCpf1, PcCpf1, BpCpf1, CmtCpf1, LiCpf1, PmCpf1, Pb3310Cpf1, Pb4417Cpf1, BsCpf1, EeCpf1, BhCas12b, AkCas12b, EbCas12b, LsCas12b, RfCas13d, LwaCas13a, PspCas13b, PguCas13b и RanCas13b.

[0012] В некоторых вариантах реализации сайт расщепления протеазой представляет собой сайт расщепления протеазой, выбранной из группы, состоящей из протеазы TuMV, протеазы PPV, протеазы PVY, протеазы ZIKV и протеазы WNV.

[0013] В некоторых вариантах реализации сайт расщепления протеазой представляет собой сайт саморасщепления. В некоторых вариантах реализации сайт расщепления протеазой представляет собой сайт расщепления протеазой TEV. В некоторых вариантах реализации слитый белок дополнительно содержит третий фрагмент, содержащий протеазу TEV или ее фрагмент. В некоторых вариантах реализации третий фрагмент содержит фрагмент протеазы TEV, который отдельно не способен расщеплять сайт расщепления протеазой TEV.

[0014] Также предложен, в другом варианте реализации, слитый белок, содержащий: первый фрагмент, содержащий цитидиндезаминазу или ее каталитический домен, белок, ассоциированный с короткими палиндромными повторами, регулярно расположенными группами (CRISPR) (Cas), и первый фрагмент протеазы TEV, второй фрагмент, содержащий ингибитор цитидиндезаминазы, и сайт расщепления протеазой TEV между первым фрагментом и вторым фрагментом, причем первый фрагмент протеазы TEV отдельно не способен расщеплять сайт расщепления протеазой TEV.

[0015] В некоторых вариантах реализации слитый белок дополнительно содержит ингибитор урацилгликозилазы (UGI). В некоторых вариантах реализации ингибитор цитидиндезаминазы, сайт расщепления протеазой TEV, цитидиндезаминаза или ее каталитический домен, белок Cas и первый фрагмент протеазы TEV расположены в указанном порядке от N-конца к С-концу. В некоторых вариантах реализации первый фрагмент протеазы TEV представляет собой N-концевой домен (SEQ ID NO: 3) или С-концевой домен (SEQ ID NO: 4) протеазы TEV. В некоторых вариантах реализации сайт расщепления протеазой TEV имеет последовательность аминокислот SEQ ID NO: 5.

[0016] Дополнительно предложен, в одном варианте реализации, способ осуществления редактирования генома в клетке в целевом сайте, включающий введение в клетку: (a) слитого белка согласно настоящему изобретению, (b) направляющей РНК, которая нацелена на целевой сайт, или крРНК, которая нацелена на целевой сайт, и тракрРНК, и дополнительно содержит последовательность метки, и (c) второго фрагмента протеазы

TEV, соединенного с распознающим РНК пептидом, который способен связываться с последовательностью метки.

[0017] В некоторых вариантах реализации одну или более указанных молекул вводят в клетку с помощью полинуклеотида, кодирующего указанную молекулу. В некоторых вариантах реализации первый фрагмент протеазы TEV и второй фрагмент протеазы TEV, когда они взаимодействуют, способны расщеплять сайт расщепления протеазой TEV. В некоторых вариантах реализации второй фрагмент протеазы TEV соединен с распознающим РНК пептидом.

[0018] В некоторых вариантах реализации последовательность метки содержит последовательность MS2 (SEQ ID NO: 16). В некоторых вариантах реализации распознающий РНК пептид содержит белок оболочки MS2 (MCP, SEQ ID NO: 22). В некоторых вариантах реализации последовательность метки содержит последовательность PP7 (SEQ ID NO: 18), и распознающий РНК пептид содержит белок оболочки PP7 (PCP, SEQ ID NO: 23), или последовательность метки содержит последовательность boxB (SEQ ID NO: 20), и распознающий РНК пептид содержит белок оболочки boxB (N22p, SEQ ID NO: 24).

[0019] Также предложен, в одном варианте реализации, набор или упаковка для осуществления редактирования генов, содержащая: (a) слитый белок согласно настоящему изобретению, и (b) второй фрагмент протеазы TEV, соединенный с распознающим РНК пептидом, который способен связываться с последовательностью РНК.

[0020] В еще одном варианте реализации предложен слитый белок, содержащий: первый фрагмент, содержащий первую цитидиндезаминазу или ее каталитический домен, и второй фрагмент, содержащий ингибиторный домен второй цитидиндезаминазы, причем первая цитидиндезаминаза такая же или отличная от второй цитидиндезаминазы.

[0021] В другом варианте реализации предложен слитый белок, содержащий первый фрагмент, содержащий: дезаминазу азотистых оснований или ее каталитический домен, ингибитор дезаминазы азотистых оснований, первый распознающий РНК пептид и сайт расщепления протеазой TEV между дезаминазой азотистых оснований или ее каталитическим доменом и ингибитором дезаминазы азотистых оснований.

[0022] В некоторых вариантах реализации слитый белок дополнительно содержит второй фрагмент, содержащий: фрагмент протеазы TEV, который отдельно не способен расщеплять сайт расщепления протеазой TEV, и второй распознающий РНК пептид. В некоторых вариантах реализации слитый белок дополнительно содержит сайт саморасщепления между первым фрагментом и вторым фрагментом.

[0023] В некоторых вариантах реализации слитый белок дополнительно содержит третий фрагмент, содержащий второй фрагмент протеазы TEV, причем первый фрагмент протеазы TEV способен расщеплять сайт расщепления протеазой TEV в присутствии второго фрагмента протеазы TEV. В некоторых вариантах реализации слитый белок дополнительно содержит второй сайт саморасщепления между вторым фрагментом и третьим фрагментом, и после расщепления второго сайта саморасщепления, слитый белок высвобождает второй фрагмент протеазы TEV, который не соединен с каким-либо распознающим РНК пептидом.

[0024] Также предложена, в одном варианте реализации, система двух направляющих РНК, содержащая: целевую единую направляющую РНК, содержащую первый спейсер, последовательность которого комплементарна целевой последовательности нуклеиновой кислоты, ближайшей к первому сайту мотива, смежного с протоспейсером (PAM), вспомогательную единую направляющую РНК, содержащую второй спейсер, последовательность которого комплементарна второй последовательности нуклеиновой кислоты, ближайшей ко второму сайту PAM, белок, ассоциированный с короткими палиндромными повторами, регулярно расположенными группами (CRISPR) (Cas), и дезаминазу азотистых оснований, причем второй сайт PAM расположен на расстоянии от 34 до 91 основания от первого сайта PAM. В некоторых вариантах реализации длина второго спейсера составляет 8 - 15 оснований. В некоторых вариантах реализации длина второго спейсера составляет 9 - 12 оснований.

[0025] В одном варианте реализации предложена направляющая РНК, содержащая остов, который содержит, в направлении от 5' к 3', первую структуру шпильки (структуру "петля-на-стебле"), вторую структуру шпильки, третью структуру шпильки и четвертую структуру шпильки, причем третья шпилька содержит в себе пять спаренных оснований. В другом варианте реализации в настоящем изобретении предложена направляющая РНК, содержащая остов, полученный из SEQ ID NO: 31 путем введения спаривания между основаниями в положениях 45 и 55. В некоторых

вариантах реализации остов содержит последовательность, выбранную из группы, состоящей из SEQ ID NO: 32 - 43. В некоторых вариантах реализации длина направляющей РНК составляет по меньшей мере 100 или 120 нуклеотидов.

[0026] В другом варианте реализации предложен способ осуществления генетического редактирования в клетке в целевом сайте, включающий введение в клетку: первой вирусной частицы, заключающей в себе первую конструкцию, кодирующую белок, ассоциированный с короткими палиндромными повторами, регулярно расположенными группами (CRISPR) (Cas), и второй вирусной частицы, заключающей в себе вторую конструкцию, кодирующую обратную транскриптазу, соединенную с распознающим РНК пептидом.

[0027] В некоторых вариантах реализации вторая конструкция дополнительно кодирует направляющую РНК, содержащую сайт распознавания РНК, с которым связывается распознающий РНК пептид. В некоторых вариантах реализации белок Cas представляет собой SpCas9-NG (SEQ ID NO: 46) или xSpCas9 (SEQ ID NO: 47).

[0028] Также предложены полинуклеотиды, кодирующие слитые белки согласно настоящему изобретению, конструкции, содержащие указанные полинуклеотиды, клетки, содержащие указанные полинуклеотиды или конструкции, и композиции, содержащие любые из упомянутых выше компонентов, без ограничения.

КРАТКОЕ ОПИСАНИЕ ЧЕРТЕЖЕЙ

[0029] ФИГ. 1А-С. Непредусмотренные замены оснований, вызванные современными ВЕ на участке оцДНК Sa-SITE31. 1А. Схематическая диаграмма, иллюстрирующая коэкспрессию никазы SaD10A и Sa-sgSITE31 для запуска образования участка оцДНК в целевом сайте Sa-sgSITE31. **1В.** Схематическая диаграмма, иллюстрирующая котрансфекцию плазмиды, экспрессирующей Sa-sgSITE31, и плазмиды, экспрессирующей никазу SaD10A, с плазмидой, экспрессирующей ВЕ3, плазмидой, экспрессирующей hA3A-ВЕ3, или пустым вектором. **1С.** Непредусмотренные замены оснований, вызванные ВЕ3 и hA3A-ВЕ3. Пунктирные прямоугольники показывают положения непредусмотренных замен оснований в целевом сайте Sa-sgSITE31.

[0030] ФИГ. 2А-С. Непредусмотренные замены оснований, вызванные современными ВЕ на участке оцДНК Sa-SITE42. **2А.** Схематическая диаграмма, иллюстрирующая коэкспрессию никазы SaD10A и Sa-sgSITE42 для запуска образования участка оцДНК в целевом сайте Sa-sgSITE42. **2В.** Схематическая диаграмма, иллюстрирующая котрансфекцию плазмиды, экспрессирующей Sa-sgSITE42, и плазмиды, экспрессирующей никазу SaD10A, с плазмидой, экспрессирующей ВЕ3, плазмидой, экспрессирующей hA3A-ВЕ3, или пустым вектором. **2С.** Непредусмотренные замены оснований, вызванные ВЕ3 и hA3A-ВЕ3. Пунктирные прямоугольники показывают положения непредусмотренных замен оснований в целевом сайте Sa-sgSITE42.

[0031] ФИГ. 3А-С. Непредусмотренные замены оснований, вызванные современными ВЕ на участке оцДНК Sa-F1. **3А.** Схематическая диаграмма, иллюстрирующая коэкспрессию никазы SaD10A и Sa-sgF1 для запуска образования участка оцДНК в целевом сайте Sa-sgF1. **3В.** Схематическая диаграмма, иллюстрирующая котрансфекцию плазмиды, экспрессирующей Sa-sgF1, и плазмиды, экспрессирующей никазу SaD10A, с плазмидой, экспрессирующей ВЕ3, плазмидой, экспрессирующей hA3A-ВЕ3, или пустым вектором. **3С.** Непредусмотренные замены оснований, вызванные ВЕ3 и hA3A-ВЕ3. Пунктирные прямоугольники показывают положения непредусмотренных замен оснований в целевом сайте Sa-sgF1.

[0032] ФИГ. 4А-С. mA3CDA2 ингибирует активность редактирования оснований С на Т на участке TET1. **4А.** Схематическая диаграмма, иллюстрирующая участки доменов CDA в mA3, rA1 и hA3A. **4В.** Схематическая диаграмма, иллюстрирующая котрансфекцию плазмиды, экспрессирующей sgTET1, с плазмидой, экспрессирующей mA3-ВЕ3, плазмидой, экспрессирующей mA3CDA1-ВЕ3, плазмидой, экспрессирующей mA3rev-ВЕ3, плазмидой, экспрессирующей mA3rev-2A-ВЕ3, плазмидой, экспрессирующей ВЕ3, плазмидой, экспрессирующей mA3CDA2-ВЕ3, плазмидой, экспрессирующей mA3CDA2-2A-ВЕ3, плазмидой, экспрессирующей hA3A-ВЕ3, плазмидой, экспрессирующей mA3CDA2-hA3A-ВЕ3, или плазмидой, экспрессирующей mA3CDA2-2A-hA3A-ВЕ3. **4С.** mA3CDA2 ингибирует активность редактирования С на Т mA3CDA1-ВЕ3, ВЕ3 и hA3A-ВЕ3. Пунктирные прямоугольники показывают положения редактирования оснований С на Т в целевом сайте sgTET1.

[0033] ФИГ. 5А-С. mA3CDA2 ингибирует активность редактирования оснований С на Т на участке RNF2. 5А. Схематическая диаграмма, иллюстрирующая участки доменов CDA в mA3, rA1 и hA3A. **5В.** Схематическая диаграмма, иллюстрирующая котрансфекцию плазмиды, экспрессирующей sgRNF2, с плазмидой, экспрессирующей mA3-BE3, плазмидой, экспрессирующей mA3CDA1-BE3, плазмидой, экспрессирующей mA3rev-BE3, плазмидой, экспрессирующей mA3rev-2A-BE3, плазмидой, экспрессирующей BE3, плазмидой, экспрессирующей mA3CDA2-BE3, плазмидой, экспрессирующей mA3CDA2-2A-BE3, плазмидой, экспрессирующей hA3A-BE3, плазмидой, экспрессирующей mA3CDA2-hA3A-BE3, или плазмидой, экспрессирующей mA3CDA2-2A-hA3A-BE3. **5С.** mA3CDA2 ингибирует активность редактирования С на Т mA3CDA1-BE3, BE3 и hA3A-BE3. Пунктирные прямоугольники показывают положения редактирования оснований С на Т в целевом сайте sgRNF2.

[0034] ФИГ. 6А-С. mA3CDA2 ингибирует активность редактирования оснований С на Т на участке SITE3. 6А. Схематическая диаграмма, иллюстрирующая участки доменов CDA в mA3, rA1 и hA3A. **6В.** Схематическая диаграмма, иллюстрирующая котрансфекцию плазмиды, экспрессирующей sgSITE3, с плазмидой, экспрессирующей mA3-BE3, плазмидой, экспрессирующей mA3CDA1-BE3, плазмидой, экспрессирующей mA3rev-BE3, плазмидой, экспрессирующей mA3rev-2A-BE3, плазмидой, экспрессирующей BE3, плазмидой, экспрессирующей mA3CDA2-BE3, плазмидой, экспрессирующей mA3CDA2-2A-BE3, плазмидой, экспрессирующей hA3A-BE3, плазмидой, экспрессирующей mA3CDA2-hA3A-BE3, или плазмидой, экспрессирующей mA3CDA2-2A-hA3A-BE3. **6С.** mA3CDA2 ингибирует активность редактирования С на Т mA3CDA1-BE3, BE3 и hA3A-BE3. Пунктирные прямоугольники показывают положения редактирования оснований С на Т в целевом сайте sgSITE3.

[0035] ФИГ. 7А-С. hA3BCDA1 ингибирует активность редактирования оснований С на Т на участке TET1. 7А. Схематическая диаграмма, иллюстрирующая участки доменов CDA в hA3B. **7В.** Схематическая диаграмма, иллюстрирующая котрансфекцию плазмиды, экспрессирующей sgTET1, с плазмидой, экспрессирующей hA3B-BE3, плазмидой, экспрессирующей hA3BCDA2-BE3, или плазмидой, экспрессирующей hA3B-2A-BE3. **7С.** hA3BCDA1 ингибирует активность редактирования С на Т hA3BCDA2-BE3. Пунктирные прямоугольники показывают положения редактирования оснований С на Т в целевом сайте sgTET1.

[0036] ФИГ. 8А-С. hA3BCDA1 ингибирует активность редактирования оснований С на Т на участке RNF2. 8А. Схематическая диаграмма, иллюстрирующая участки доменов CDA в hA3B. **8В.** Схематическая диаграмма, иллюстрирующая котрансфекцию плазмиды, экспрессирующей sgRNF2, с плазмидой, экспрессирующей hA3B-BE3, плазмидой, экспрессирующей hA3BCDA2-BE3, или плазмидой, экспрессирующей hA3B-2A-BE3. **8С.** hA3BCDA1 ингибирует активность редактирования С на Т hA3BCDA2-BE3. Пунктирные прямоугольники показывают положения редактирования оснований С на Т в целевом сайте sgRNF2.

[0037] ФИГ. 9А-С. hA3BCDA1 ингибирует активность редактирования оснований С на Т на участке SITE3. 9А. Схематическая диаграмма, иллюстрирующая участки доменов CDA в hA3B. **9В.** Схематическая диаграмма, иллюстрирующая котрансфекцию плазмиды, экспрессирующей sgSITE3, с плазмидой, экспрессирующей hA3B-BE3, плазмидой, экспрессирующей hA3BCDA2-BE3, или плазмидой, экспрессирующей hA3B-2A-BE3. **9С.** hA3BCDA1 ингибирует активность редактирования С на Т hA3BCDA2-BE3. Пунктирные прямоугольники показывают положения редактирования оснований С на Т в целевом сайте sgSITE3.

[0038] ФИГ. 10А-С. Картирование сайта разделения mA3 путем анализа эффективности редактирования оснований на участке FANCF. 10А. Схематическая диаграмма, иллюстрирующая участки двух доменов CDA в mA3 и сайты (AK196/AK197, AK207/AK208, AK215/AK216, AK229/AK230, AK237/AK238), используемые для разделения mA3. **10В.** Схематическая диаграмма, иллюстрирующая котрансфекцию плазмиды, экспрессирующей sgFANCF, с плазмидой, экспрессирующей mA3rev-BE3-196, плазмидой, экспрессирующей mA3rev-2A-BE3-196, плазмидой, экспрессирующей mA3rev-BE3, плазмидой, экспрессирующей mA3rev-2A-BE3, плазмидой, экспрессирующей mA3rev-BE3-215, плазмидой, экспрессирующей mA3rev-2A-BE3-215, плазмидой, экспрессирующей mA3rev-BE3-229, плазмидой, экспрессирующей mA3rev-2A-BE3-229, плазмидой, экспрессирующей mA3rev-BE3-237, или плазмидой, экспрессирующей mA3rev-2A-BE3-237. **10С.** Сайты разделения, располагающиеся от AK196/AK197 до AK237/AK238, как правило, сохраняют эффективности редактирования С на Т. Пунктирные прямоугольники показывают положения редактирования оснований С на Т в целевом сайте sgFANCF.

[0039] ФИГ. 11А-С. Картирование сайта разделения mA3 путем анализа эффективности редактирования оснований на участке SITE2. **11А.** Схематическая диаграмма, иллюстрирующая участки двух доменов CDA в mA3 и сайты (AK196/AK197, AK207/AK208, AK215/AK216, AK229/AK230, AK237/AK238), используемые для разделения mA3. **11В.** Схематическая диаграмма, иллюстрирующая котрансфекцию плазмиды, экспрессирующей sgSITE2, с плазмидой, экспрессирующей mA3rev-BE3-196, плазмидой, экспрессирующей mA3rev-2A-BE3-196, плазмидой, экспрессирующей mA3rev-BE3, плазмидой, экспрессирующей mA3rev-2A-BE3, плазмидой, экспрессирующей mA3rev-BE3-215, плазмидой, экспрессирующей mA3rev-2A-BE3-215, плазмидой, экспрессирующей mA3rev-BE3-229, плазмидой, экспрессирующей mA3rev-2A-BE3-229, плазмидой, экспрессирующей mA3rev-BE3-237, или плазмидой, экспрессирующей mA3rev-2A-BE3-237. **11С.** Сайты разделения, располагающиеся от AK196/AK197 до AK237/AK238, как правило, сохраняют эффективности редактирования С на Т. Пунктирные прямоугольники показывают положения редактирования оснований С на Т в целевом сайте sgSITE2.

[0040] ФИГ. 12А-С. Картирование сайта разделения mA3 путем анализа эффективности редактирования оснований в участке SITE4. **12А.** Схематическая диаграмма, иллюстрирующая участки двух доменов CDA в mA3 и сайты (AK196/AK197, AK207/AK208, AK215/AK216, AK229/AK230, AK237/AK238), используемые для разделения mA3. **12В.** Схематическая диаграмма, иллюстрирующая котрансфекцию плазмиды, экспрессирующей sgSITE4, с плазмидой, экспрессирующей mA3rev-BE3-196, плазмидой, экспрессирующей mA3rev-2A-BE3-196, плазмидой, экспрессирующей mA3rev-BE3, плазмидой, экспрессирующей mA3rev-2A-BE3, плазмидой, экспрессирующей mA3rev-BE3-215, плазмидой, экспрессирующей mA3rev-2A-BE3-215, плазмидой, экспрессирующей mA3rev-BE3-229, плазмидой, экспрессирующей mA3rev-2A-BE3-229, плазмидой, экспрессирующей mA3rev-BE3-237, или плазмидой, экспрессирующей mA3rev-2A-BE3-237. **12С.** Сайты разделения, располагающиеся от AK196/AK197 до AK237/AK238, как правило, сохраняют эффективности редактирования С на Т. Пунктирные прямоугольники показывают положения редактирования оснований С на Т в целевом сайте sgSITE4.

[0041] ФИГ. 13А-В. Картирование минимального участка mA3, который оказывает ингибиторное влияние на редактирование оснований на участке

FANCF. 13A. Схематическая диаграмма, иллюстрирующая котрансфекцию плазмиды, экспрессирующей sgFANCF, с плазмидой, экспрессирующей mA3rev-BE3-237, плазмидой, экспрессирующей mA3rev-BE3-237-Del-255, плазмидой, экспрессирующей mA3rev-BE3-237-Del-285, или плазмидой, экспрессирующей mA3rev-BE3-237-Del-333.

13B. Участок, располагающийся от АК334 до АК429, в mA3 оказывает ингибиторное влияние на редактирование оснований. Пунктирные прямоугольники показывают положения редактирования оснований С на Т в целевом сайте sgFANCF.

[0042] ФИГ. 14A-B. Картирование минимального участка mA3, который оказывает ингибиторное влияние на редактирование оснований на участке SITE2.

14A. Схематическая диаграмма, иллюстрирующая котрансфекцию плазмиды, экспрессирующей sgSITE2, с плазмидой, экспрессирующей mA3rev-BE3-237, плазмидой, экспрессирующей mA3rev-BE3-237-Del-255, плазмидой, экспрессирующей mA3rev-BE3-237-Del-285, или плазмидой, экспрессирующей mA3rev-BE3-237-Del-333.

14B: Участок, располагающийся от АК334 до АК429, в mA3 оказывает ингибиторное влияние на редактирование оснований. Пунктирные прямоугольники показывают положения редактирования оснований С на Т в целевом сайте sgSITE2.

[0043] ФИГ. 15A-B. Картирование минимального участка mA3, который оказывает ингибиторное влияние на редактирование оснований в участке SITE4.

15A. Схематическая диаграмма, иллюстрирующая котрансфекцию плазмиды, экспрессирующей sgSITE4, с плазмидой, экспрессирующей mA3rev-BE3-237, плазмидой, экспрессирующей mA3rev-BE3-237-Del-255, плазмидой, экспрессирующей mA3rev-BE3-237-Del-285, или плазмидой, экспрессирующей mA3rev-BE3-237-Del-333.

15B. Участок, располагающийся от АК334 до АК429, в mA3 оказывает ингибиторное влияние на редактирование оснований. Пунктирные прямоугольники показывают положения редактирования оснований С на Т в целевом сайте sgSITE4.

[0044] ФИГ. 16A-B. Схематическая диаграмма, иллюстрирующая рабочий процесс BEsafe и BE3 или hA3A-BE3.

16A. BEsafe вызывает редактирование оснований С на Т в целевом сайте и позволяет избежать образования мутаций в нецелевых участках оцДНК. **16B.** BE3 или hA3A-BE3 вызывает редактирование оснований С на Т в целевом сайте, но вызывает мутации С на Т в нецелевых участках оцДНК.

[0045] ФИГ. 17A-D. Сравнение hA3A-BE3 и BEsafe в нецелевом участке оцДНК Sa-SITE31 и в целевом сайте TET1. **17A.** Схематическая диаграмма, иллюстрирующая коэкспрессию никазы SaD10A и Sa-sgSITE31 для запуска образования участка оцДНК в целевом сайте Sa-sgSITE31. **17B.** Схематическая диаграмма, иллюстрирующая котрансфекцию плазмиды, экспрессирующей Sa-sgSITE31, и плазмиды, экспрессирующей никазу SaD10A, с плазмидой, экспрессирующей hA3A-BE3, и плазмидой, экспрессирующей sgTET1, с плазмидой, экспрессирующей BEsafe, и плазмидой, экспрессирующей MS2-sgTET1 и MCP-TEVc, или с плазмидой, экспрессирующей MCP-TEVc, и плазмидой, экспрессирующей MS2-sgTET1 и BEsafe. **17C.** Сравнение частот непредусмотренной мутации С на Т, инициируемой hA3A-BE3 и BEsafe в нецелевом участке оцДНК Sa-SITE31. Пунктирные прямоугольники показывают положения непредусмотренных замен оснований в целевом сайте Sa-sgSITE31. **17D.** Сравнение эффективностей редактирования оснований hA3A-BE3 и BEsafe в сайте TET1. Пунктирные прямоугольники показывают положения редактирования оснований С на Т в целевом сайте sgTET1.

[0046] ФИГ. 18A-D. Сравнение hA3A-BE3 и BEsafe в нецелевом участке оцДНК Sa-SITE32 и в целевом сайте RNF2. **18A.** Схематическая диаграмма, иллюстрирующая коэкспрессию никазы SaD10A и Sa-sgSITE32 для запуска образования участка оцДНК в целевом сайте Sa-sgSITE32. **18B.** Схематическая диаграмма, иллюстрирующая котрансфекцию плазмиды, экспрессирующей Sa-sgSITE32, и плазмиды, экспрессирующей никазу SaD10A, с плазмидой, экспрессирующей hA3A-BE3, и плазмидой, экспрессирующей sgRNF2, с плазмидой, экспрессирующей BEsafe, и плазмидой, экспрессирующей MS2-sgRNF2 и MCP-TEVc, или с плазмидой, экспрессирующей MCP-TEVc, и плазмидой, экспрессирующей MS2-sgRNF2 и BEsafe. **18C.** Сравнение частоты непредусмотренной мутации С на Т, инициируемой hA3A-BE3 и BEsafe, в нецелевом участке оцДНК Sa-SITE32. Пунктирные прямоугольники показывают положения непредусмотренных замен оснований в целевом сайте Sa-sgSITE32. **18D.** Сравнение эффективностей редактирования оснований hA3A-BE3 и BEsafe в сайте RNF2. Пунктирные прямоугольники показывают положения редактирования оснований С на Т в целевом сайте sgRNF2.

[0047] ФИГ. 19A-D. Сравнение hA3A-BE3 и BEsafe в нецелевом участке оцДНК Sa-F1 и в целевом сайте SITE3. **19A.** Схематическая диаграмма, иллюстрирующая

коэкспрессию никазы SaD10A и Sa-sgF1 для запуска образования участка оцДНК в целевом сайте Sa-sgF1. **19B.** Схематическая диаграмма, иллюстрирующая котрансфекцию плазмиды, экспрессирующей Sa-sgF1, и плазмиды, экспрессирующей никазу SaD10A, с плазмидой, экспрессирующей hA3A-BE3, и плазмидой, экспрессирующей sgSITE3, с плазмидой, экспрессирующей BEsafe, и плазмидой, экспрессирующей MS2-sgSITE3 и MCP-TEVc, или с плазмидой, экспрессирующей MCP-TEVc, и плазмидой, экспрессирующей MS2-sgSITE3 и BEsafe. **19C.** Сравнение частоты непредусмотренной мутации С на Т, инициируемой hA3A-BE3 и BEsafe, в нецелевом участке оцДНК Sa-F1. Пунктирные прямоугольники показывают положения непредусмотренных замен оснований в целевом сайте Sa-sgF1. **19D.** Сравнение эффективностей редактирования оснований hA3A-BE3 и BEsafe в сайте SITE3. Пунктирные прямоугольники показывают положения редактирования оснований С на Т в целевом сайте sgSITE3.

[0048] ФИГ. 20a-f. Идентификация ингибиторов цитидиндезаминазы. 20a. На схематических диаграммах проиллюстрированы представители семейства APOBEC, которые содержат одинарные или двойные домены CDA (слева), и спаренные редакторы оснований, которые сконструировали с одним или двумя CDA из APOBEC с двойными доменами (справа). **20b.** Частоты редактирования, вызванного указанными BE в одном типичном геномном локусе. **20c.** Статистический анализ нормированных частот редактирования с установкой таковых, вызванных BE, содержащих один CDA, как 100%. $n = 78$ из трех независимых экспериментов в 26 редактируемых сайтах цитозина, показанных в (b). **20d.** На схематических диаграммах проиллюстрировано конъюгирование различных ингибиторов цитидиндезаминазы (CDI) с N-концом mA3CDA1-nSpCas9-BE. **20e.** Частоты редактирования, вызванного указанными BE в одном типичном геномном локусе. **20f.** Статистический анализ нормированных частот редактирования с установкой таковых, вызванных BE без CDI, как 100%. $n = 57$ из трех независимых экспериментов в 19 редактируемых сайтах цитозина, показанных в (e). (b), (e) Средние $\pm$ СО по трем независимым экспериментам. НТ - нетрансфицированный контроль. (c), (f) *P*-значение, односторонний критерий Стьюдента. Показаны медиана и межквартильный размах (IQR).

[0049] ФИГ. 21a-f. Конъюгирование mA3CDI уменьшало непредусмотренное редактирование оснований в енРНК-независимых сайтах НЦоц. 21a. На

схематических диаграммах проиллюстрировано, что BE3 вызывает мутации С на Т, но конъюгированный с CDI iBE1 остается неактивным в енРНК-независимых сайтах НЦоц. **21b.** Сравнение частот редактирования С на Т, вызванного BE3 и iBE1 на участках оцДНК, запускаемого вызванными nSaCas9 одноцепочечными разрывами (ОЦР). **21c.** Статистический анализ нормированных кумулятивных частот редактирования в четырех сайтах оцДНК, показанных в (b), с установкой таковых, вызванных BE3, как 100%. $n = 12$ из трех независимых экспериментов. **21d.** На схематических диаграммах проиллюстрировано, что опосредованное енРНК расщепление CDI восстанавливает активность редактирования iBE в целевых сайтах. **21e.** Сравнение частот редактирования С на Т, вызванных BE3 и iBE1, в целевых сайтах. **21f.** Статистический анализ нормированных кумулятивных частот редактирования в четырех целевых сайтах, показанных в (e), с установкой таковых, вызванных BE3, как 100%. $n = 12$ из трех независимых экспериментов. (c), (f) Средние $\pm$ СО по трем независимым экспериментам. (d), (g) P -значение, односторонний критерий Стьюдента. Показаны медиана и межквартильный размах (IQR).

[0050] ФИГ. 22a-e. neSpCas9 уменьшала непредусмотренное редактирование iBE1 в сайтах НЦен. **22a.** На схематических диаграммах проиллюстрировано, что iBE1, но не iBE2 вызывает редактирование С на Т в сайтах НЦен, которые частично комплементарны енРНК. **22b.** Сравнение частот редактирования С на Т, вызванных iBE1 и iBE с улучшенной специфичностью нацеливания в указанных сайтах НЦен. **22c.** Статистический анализ нормированных кумулятивных частот редактирования в сайтах НЦен для двух енРНК, используемых в (b), с установкой таковых, вызванных iBE1, как 100%. $n = 6$ из трех независимых экспериментов. **22d.** Сравнение частот редактирования С на Т, вызванных iBE1 и iBE с улучшенной специфичностью нацеливания в целевых сайтах. **22e.** Статистический анализ нормированных кумулятивных частот редактирования в шести целевых сайтах, показанных в (d), с установкой таковых, вызванных iBE1, как 100%. $n = 18$ из трех независимых экспериментов. (b), (d) Средние $\pm$ СО по трем независимым экспериментам. (c), (e) P -значение, односторонний критерий Стьюдента. Показаны медиана и межквартильный размах (IQR).

[0051] ФИГ. 23a-e. Сравнение редактирования оснований, вызванного hA3A-BE3 и iBE2. **23a.** Сравнение частот редактирования С на Т, вызванных hA3A-BE3 и iBE2 в типичных НЦоц, НЦен и целевых сайтах. **23b-c.** Статистический анализ нормированных

кумулятивных частот редактирования в НЦоц, НЦен (b) и целевых (c) сайтах для трех енРНК, используемых в (a), с установкой таковых, вызванных hA3A-BE3, как 100%. n = 9 из трех независимых экспериментов. **23d.** Статистический анализ нормированных отношений частот целевого редактирования к частотам всех редактирований в сайтах НЦоц и НЦен для трех енРНК, используемых в (a), с установкой таковых, вызванных hA3A-BE3, на 1. n = 9 из трех независимых экспериментов. **23e:** На схематических диаграммах проиллюстрировано, что iBE2 вызывает специфическое редактирование оснований в целевых сайтах, но не в сайтах НЦоц и НЦен, тогда как hA3A-BE3 вызывает редактирование оснований в целевых сайтах и в обоих сайтах НЦоц и НЦен. (a) Средние $\pm$ СО по трем независимым экспериментам. (b-d) *P*-значение, односторонний критерий Стьюдента. Показаны медиана и IQR.

[0052] ФИГ. 24А-В. Схематическая диаграмма, иллюстрирующая рабочий процесс isplitBE и обычных редакторов оснований. 24А. isplitBE вызывает редактирование оснований С на Т только в целевом сайте и избегает индуцирования мутаций в случайных нецелевых участках оцДНК (НЦоц) или в нецелевых сайтах с подобием последовательности спейсерному участку енРНК (НЦен). **24В.** BE3 или hA3A-BE3 вызывает редактирование оснований С на Т в целевом сайте, но вызывает мутации С на Т в участках НЦоц и НЦен.

[0053] ФИГ. 25. Схематическая диаграмма, иллюстрирующая различные стратегии для удаления ингибитора цитидиндезаминазы (mA3CDA2) с целевого сайта.

[0054] ФИГ. 26А-В. Редактирование С на Т в сайтах EMX1-Ц, Sa-SITE31-НЦоц и EMX1-НЦен, вызванное различными комбинациями nCas9 (D10A), цитидиндезаминазы APOBEC, ингибитора цитидиндезаминазы (CDI), ингибитора урацил-ДНК-гликозилазы (UGI) и протеазы TEV. 26А. Схематическая диаграмма, иллюстрирующая котрансфекцию плазмиды, экспрессирующей Sa-sgSITE31, и плазмиды, экспрессирующей никазу SaD10A, с указанными десятью парами плазмид, экспрессирующих различные редакторы оснований. **26В.** Сравнение эффективностей редактирования в сайтах EMX1-Ц, Sa-SITE31-НЦоц и EMX1-НЦен. isplitBE-rA1 (пара 9) вызвала значительное редактирование в сайте Ц, но не вызвала редактирования в сайтах НЦоц и НЦен.

[0055] ФИГ. 27А-В. Редактирование С на Т в сайтах FANCF-Ц, Sa-VEGFA-7-НЦоц и FANCF-НЦен, вызванное различными комбинациями nCas9 (D10A), цитидиндезаминазы APOBEC, ингибитора цитидиндезаминазы (CDI), ингибитора урацил-ДНК-гликозилазы (UGI) и протеазы TEV. 27А. Схематическая диаграмма, иллюстрирующая котрансфекцию плазмиды, экспрессирующей Sa-sgVEGFA-7, и плазмиды, экспрессирующей никазу SaD10A, с указанными десятью парами плазмид, экспрессирующих различные редакторы оснований. **27В.** Сравнение эффективностей редактирования в сайтах FANCF-Ц, Sa-VEGFA-7-НЦоц и FANCF-НЦен. isplitBE-rA1 (пара 9) вызывала значительное редактирование в сайте Ц, но не вызывала редактирования в сайтах НЦоц и НЦен.

[0056] ФИГ. 28А-В. Редактирование С на Т в сайтах V1B-Ц, Sa-SITE42-НЦоц и V1B-НЦен, вызванное различными комбинациями nCas9 (D10A), цитидиндезаминазы APOBEC, ингибитора цитидиндезаминазы (CDI), ингибитора урацил-ДНК-гликозилазы (UGI) и протеазы TEV. 28А. Схематическая диаграмма, иллюстрирующая котрансфекцию плазмиды, экспрессирующей Sa-sgSITE42, и плазмиды, экспрессирующей никазу SaD10A, с указанными десятью парами плазмид, экспрессирующих различные редакторы оснований. **28В.** Сравнение эффективностей редактирования в сайтах V1B-Ц, Sa-SITE42-НЦоц и V1B-НЦен. isplitBE-rA1 (пара 9) вызывала значительное редактирование в сайте Ц, но не вызывала редактирования в сайтах НЦоц и НЦен.

[0057] ФИГ. 29А-С. Влияние расстояния между вспомогательной енРНК (венРНК) и енРНК на эффективность редактирования оснований. 29А. Схематическая диаграмма, иллюстрирующая расстояние между венРНК и енРНК в сайтах DNTET1, EMX1 и FANCF. **29В.** Частоты редактирования оснований, вызванного указанными енРНК и венРНК. **29С.** Сводное представление влияния расстояния между венРНК и енРНК. Диапазон расстояний для лучшей эффективности редактирования оснований составляет от -91 до -34 п.о. от РАМ венРНК до РАМ енРНК.

[0058] ФИГ. 30А-С. Влияние длины спейсера венРНК на эффективность редактирования оснований. 30А. Схематическая диаграмма, иллюстрирующая котрансфекцию енРНК и венРНК с различными длинами спейсеров в сайтах DNEMX1, FANCF и V1A. **30В.** Частоты редактирования оснований, вызванного указанными енРНК и венРНК в целевых для венРНК и енРНК сайтах. **30С.** Статистический анализ

влияния длины спейсера венРНК. Применение венРНК со спейсером 10 п.о. сильно снижает эффективность редактирования в целевых сайтах венРНК, но сохраняет эффективность редактирования в целевых сайтах енРНК.

[0059] ФИГ. 31. Сравнение эффективности редактирования isplitBE-rA1 и BE3.

Частоты редактирования, вызванные указанными редакторами оснований в различных целевых сайтах.

[0060] ФИГ. 32А-С. Сравнение мутаций С на Т по всему геному, вызванных isplitBE-rA1 и BE3.

32А. Уровни экспрессии мРНК в клетках 293FT дикого типа и клетках 293FT с нокаутом APOBEC3 (293FT-A3KO). **32В.** Схематическая диаграмма, иллюстрирующая процедуры определения мутаций С на Т по всему геному, вызванных редакторами оснований. **32С.** Эффективности целевого редактирования (слева) и количество мутаций С на Т по всему геному, вызванных Cas9, BE3, hA3A-BE3-Y130F (Y130F) и isplitBE-rA1.

[0061] ФИГ. 33А-С. Сравнение мутаций С на U по всему транскриптому, вызванных isplitBE-mA3, BE3 и hA3A-BE3-Y130F (Y130F).

33А. Количество мутаций С на U по всему транскриптому, вызванных Cas9, BE3, hA3A-BE3-Y130F (Y130F) и isplitBE-mA3. **33В.** Частоты редактирования РНК С на U, вызванного Cas9, BE3, hA3A-BE3-Y130F (Y130F) и isplitBE-mA3. **33С.** Распределение редактирования РНК С на U, вызванного BE3 в повторе 1 и isplitBE-mA3 в повторе 1.

[0062] ФИГ. 34А-D. Стоп-кодон, вызванный isplitBE-mA3 в гене PCSK9 человека.

34А. Схематическая диаграмма, иллюстрирующая котрансфекцию енРНК и венРНК с isplitBE-mA3 и nCas9. **34В - 34D.** Эффективность редактирования, вызванного isplitBE-mA3 в указанных сайтах.

[0063] ФИГ. 35А-В. Влияние ингибирования mA3CDA2 на эффективность редактирования редактором основания аденина (ABE).

35А. Схематическая диаграмма, иллюстрирующая котрансфекцию енРНК и ABE, слитым с mA3CDA2 или нет. **35В.** Эффективность редактирования, вызванного указанными ABE в сайтах RNF2 и FANCF.

[0064] ФИГ. 36А-G. Улучшенное праймированное редактирование путем манипулирования над направляющей РНК для праймированного редактирования

(прнРНК). 36А. Схематическая диаграмма, иллюстрирующая замену пар оснований РНК, чтобы повысить стабильность стебля улучшенной прнРНК (упрнРНК). **36В.** Схематическая диаграмма, иллюстрирующая котрансфекцию PE2, вводящей одноцепочечный разрыв в енРНК, с прнРНК или упрнРНК-GC. **36С-36D.** Сравнение эффективностей праймированного редактирования, вызванного прнРНК и упрнРНК-GC. **36Е.** Схематическая диаграмма, иллюстрирующая замену пар оснований РНК для повышения стабильности стебля улучшенной прнРНК (упрнРНК). **36F.** Схематическая диаграмма, иллюстрирующая котрансфекцию PE2, вводящей одноцепочечный разрыв в енРНК, с прнРНК или упрнРНК-CG. **36G.** Сравнение эффективностей праймированного редактирования, вызванного прнРНК и упрнРНК-CG.

[0065] ФИГ. 37А-В. Система праймированного редактирования путем применения PE, содержащих различные белки Cas9. 37А. Схематическая диаграмма, иллюстрирующая котрансфекцию прнРНК, вводящей одноцепочечный разрыв в енРНК, с PE2-NG или xPE2. **37В.** Эффективности праймированного редактирования, вызванного PE2-NG и xPE2.

[0066] ФИГ. 38А-С. Разделенная система праймированного редактирования (разделенное PE). 38А. Схематическая диаграмма, иллюстрирующая рабочий процесс системы PE и разделенной системы PE. **38В.** Схематическая диаграмма, иллюстрирующая котрансфекцию PE и разделенных систем PE. **38С.** Эффективность редактирования, вызванного системой PE и разделенной системой PE в сайте EMX1.

[0067] ФИГ. 39А-С. Выравнивание коровой области mA3CDA2 с другими доменами цитидиндезаминазы.

[0068] ФИГ. 40А-D. Выравнивание hA3BCDA1 с другими доменами цитидиндезаминазы.

ПОДРОБНОЕ ОПИСАНИЕ ИЗОБРЕТЕНИЯ

Определения

[0069] Следует отметить, что термин объекта в единственном числе относится к одному или более таким объектам; например, понимают, что “антитело” представляет одно или более антител. По этой причине термины в единственном числе, термины “один или

более” и “по меньшей мере один” можно использовать взаимозаменяемо в данной заявке.

[0070] В данной заявке предполагается, что в объем термина “полипептид” входит термин “полипептид” в единственном числе, а также термин “полипептиды” во множественном числе, и он относится к молекуле, состоящей из мономеров (аминокислот), линейно связанных амидными связями (также известными как пептидные связи). Термин “полипептид” относится к любой цепи или цепям из двух или более аминокислот и не относится к конкретной длине продукта. Таким образом, пептиды, дипептиды, трипептиды, олигопептиды, “белок”, “цепь аминокислот” или любой другой термин, применяемый в отношении цепи или цепей из двух или более аминокислот, входит в объем определения “полипептид”, и термин “полипептид” можно применять вместо или взаимозаменяемо с любым из данных терминов. Также предполагается, что термин “полипептид” относится к продуктам постэкспрессионных модификаций полипептида, включая, без ограничения, гликозилирование, ацетилирование, фосфорилирование, амидирование, получение производных с помощью известных защитных/блокирующих групп, протеолитическое расщепление или модификацию с помощью не встречающихся в природе аминокислот. Полипептид можно получить из природного биологического источника или получить с помощью рекомбинантной технологии, но не обязательно транслировать с указанной последовательности нуклеиновой кислоты. Его можно получить любым способом, включая химический синтез.

[0071] “Гомология”, или “идентичность”, или “подобие” относится к сходству между последовательностями двух пептидов или между последовательностями двух молекул нуклеиновых кислот. Гомологию можно определить путем сравнения положений в каждой последовательности, которые можно выровнять с целью сравнения. Если положение в сравниваемой последовательности занято таким же основанием или аминокислотой, то молекулы гомологичны в данном положении. Степень гомологии между последовательностями является функцией количества совпадающих или гомологичных положений, общих для последовательностей. “Неродственная” или “негомологичная” последовательность менее чем на 40% идентична, даже предпочтительно менее чем на 25% идентична одной из последовательностей в соответствии с настоящим изобретением.

[0072] Если полинуклеотид или участок полинуклеотида (или полипептид или участок полипептида) на некоторый процент (например, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 % или 99 %) “идентичен по последовательности” другой последовательности, то это означает, что при выравнивании указанный процент оснований (или аминокислот) одинаковы при сравнении двух последовательностей. Такое выравнивание можно осуществить и определить процент гомологии или идентичности последовательностей, применяя программное обеспечение, известное в данной области, например, описанное у Ausubel и др. ред. (2007) *Current Protocols in Molecular Biology*. Предпочтительно для выравнивания используют параметры по умолчанию. Одной программой для выравнивания является BLAST, с применением параметров по умолчанию.

[0073] Термин “эквивалентная нуклеиновая кислота или полинуклеотид” относится к нуклеиновой кислоте, имеющей последовательность нуклеотидов с некоторой степенью гомологии или идентичности с последовательностью нуклеотидов нуклеиновой кислоты или ее комплемента. Предполагается, что гомолог двухцепочечной нуклеиновой кислоты содержит нуклеиновые кислоты с последовательностью нуклеотидов, которая в некоторой степени гомологична ее комплементу. В одном аспекте гомологи нуклеиновых кислот способны гибридизоваться с нуклеиновой кислотой или ее комплементом. Аналогичным образом, “эквивалентный полипептид” относится к полипептиду с некоторой степенью гомологии или идентичности с последовательностью аминокислот эталонного полипептида. В некоторых аспектах идентичность последовательностей составляет по меньшей мере приблизительно 70%, 75%, 80%, 85%, 90%, 95%, 98% или 99%. В некоторых аспектах эквивалентный полипептид или полинуклеотид содержит одну, две, три, четыре или пять вставок, делеций, замен и их комбинаций по сравнению с эталонным полипептидом или полинуклеотидом. В некоторых аспектах эквивалентная последовательность сохраняет активность (например, связывание с эпитопом) или структуру (например, солевой мостик) эталонной последовательности.

[0074] Термин “кодируют”, если он используется по отношению к полинуклеотидам, относится к полинуклеотиду, про который говорят, что он “кодирует” полипептид, если в его нативном состоянии или после обработки с помощью способов, хорошо известных специалистам в данной области, его можно транскрибировать и/или транслировать с

получением мРНК полипептида и/или его фрагмента. Антисмысловая цепь комплементарна такой нуклеиновой кислоте, и кодирующую последовательность можно установить по ней.

Применение ингибитора дезаминазы азотистых оснований для уменьшения количества произвольных вставок и делеций

[0075] В экспериментальных примерах и на **ФИГ. 1 - 3** показано, что широко применяемые на сегодняшний день редакторы оснований BE3 и hA3A-BE3 вызывали мутации С на Т в нецелевых участках одноцепочечной ДНК.

[0076] Тем не менее, неожиданно обнаружили, что применение APOBEC3 мыши (mA3) в mA3-BE3 (**ФИГ. 4В, 5В, 6В**), как правило, не вызывало редактирования С на Т в исследованных целевых сайтах (**ФИГ. 4С, 5С, 6С**). mA3 содержит два домена цитидиндезаминазы (CDA): CDA1 и CDA2 (**ФИГ. 4А, 5А, 6А**). Когда домен CDA2 удаляли из полноразмерного mA3, полученный в результате этого редактор оснований mA3CDA1-BE3 (**ФИГ. 4В, 5В, 6В**) вызывал значительное редактирование С на Т (**ФИГ. 4С, 5С, 6С**). Данные результаты указывают на то, что домен mA3-CDA2 является ингибитором редактирования оснований.

[0077] Также неожиданно домен mA3-CDA2 не только мог ингибировать активность редактирования оснований mA3-CDA1, он также мог ингибировать другие дезаминазы азотистых оснований. Например, когда mA3-CDA2 был слит с N-концом каждого из трех активных BE - mA3CDA1-BE3, BE3 и hA3A-BE3, - у слитых белков mA3rev-BE3, mA3-CDA2-BE3 и mA3-CDA2-hA3A-BE3 (**ФИГ. 4В, 5В, 6В**) выявляли явно пониженные эффективности редактирования оснований (**ФИГ. 4С, 5С, 6С**).

[0078] Более того, отщепление mA3-CDA2 от слитых белков восстанавливало эффективность редактирования оснований (**ФИГ. 4С, 5С, 6С**), позволяя предложить, что ингибирование mA3-CDA2 связано с его ковалентным соединением с BE.

[0079] Подобно mA3, APOBEC3В человека (hA3В) также содержит два домена цитидиндезаминазы (CDA): CDA1 и CDA2 (**ФИГ. 7А, 8А, 9А**). Включение полноразмерного hA3В в hA3В-BE3 (**ФИГ. 7В, 8В, 9В**) вызывало лишь относительно низкие уровни редактирования С на Т в трех исследованных целевых сайтах (**ФИГ. 7С, 8С, 9С**). Тем не менее, hA3В-CDA2-BE3, который получали путем удаления домена

hA3B-CDA1 (ФИГ. 7В, 8В, 9В), вызывал большие уровни редактирования С на Т (ФИГ. 7С, 8С, 9С). Данные результаты указывают на то, что hA3B-CDA1 является другим ингибитором редактирования оснований, и ингибирование hA3B-CDA1 связано с его ковалентным соединением с BE.

[0080] Используя последовательности mA3-CDA2 и hA3B-CDA1, авторы настоящего изобретения смогли определить дополнительные ингибиторы дезаминазы азотистых оснований/их домены в базе данных белков. В таблице 1 показано 44 белка/домена со значительной гомологией последовательности с коровой последовательностью mA3-CDA2 (ФИГ. 39), и в таблице 2 показано 43 белка/домена со значительной гомологией последовательности с hA3B-CDA1 (ФИГ. 40). Предполагают, что все данные белки и домены, а также их варианты и эквиваленты обладают активностями ингибирования дезаминазы азотистых оснований.

Слитые белки

[0081] На основании данных неожиданных и ожидаемых открытий разработан слитый белок, который можно применять для получения редактора оснований с улучшенной специфичностью и эффективностью редактирования оснований. В одном варианте реализации настоящего изобретения предложен слитый белок, который содержит первый фрагмент, содержащий дезаминазу азотистых оснований или ее каталитический домен, второй фрагмент, содержащий ингибитор дезаминазы азотистых оснований, и сайт расщепления протеазой между первым фрагментом и вторым фрагментом.

[0082] Редактор оснований, который содержит такой слитый белок, проявляет пониженную способность или даже не проявляет способности к редактированию и, соответственно, будет создавать меньшее количество или не будет создавать нецелевые мутации. После расщепления сайта расщепления протеазой и высвобождения ингибитора дезаминазы азотистых оснований из слитого белка в целевом сайте, редактор оснований, который находится в целевом сайте, затем будет способен эффективно редактировать целевой сайт.

[0083] Термин “дезаминаза азотистых оснований” в данной заявке относится к группе ферментов, которые катализируют гидролитическое дезаминирование азотистых оснований, таких как цитидин, дезоксицитидин, аденозин и дезоксиаденозин. Лишь

некоторые из примеров дезаминаз азотистых оснований включают цитидиндезаминазы и аденозиндезаминазы.

[0084] “Цитидиндезаминаза” относится к ферментам, которые катализируют необратимое гидролитическое дезаминирование цитидина и дезоксицитидина с образованием уридина и дезоксиуридина, соответственно. Цитидиндезаминазы поддерживают запас в клетке пиримидина. Семейством цитидиндезаминаз является АРОВЕС (“подобный каталитическому полипептиду белок, редактирующий мРНК аполипопротеина В”). Представителями данного семейства являются ферменты, редактирующие С на U. Некоторые представители семейства АРОВЕС содержат два домена: один домен АРОВЕС-подобных белков представляет собой каталитический домен, тогда как другой домен представляет собой псевдокаталитический домен. В частности, каталитический домен представляет собой цинк-зависимый домен цитидиндезаминазы и важен для дезаминирования цитидина. Для редактирования РНК посредством АРОВЕС-1 требуется гомодимеризация, и данный комплекс взаимодействует с РНК-связывающими белками с образованием эдитосомы.

[0085] Неограничивающие примеры белков АРОВЕС включают АРОВЕС1, АРОВЕС2, АРОВЕС3А, АРОВЕС3В, АРОВЕС3С, АРОВЕС3D, АРОВЕС3F, АРОВЕС3G, АРОВЕС3Н, АРОВЕС4 и индуцируемую активацией (цитидин)дезаминазу (AID).

[0086] Также известны различные мутации белков АРОВЕС, которые придают редакторам оснований различные свойства редактирования. Например, для АРОВЕС3А человека, некоторые мутанты (например, W98Y, Y130F, Y132D, W104A, D131Y и P134Y) даже превосходят человеческий АРОВЕС3А дикого типа с точки зрения эффективности редактирования или окна редактирования. Соответственно, в объем термина АРОВЕС и каждого из представителей данного семейства также входят варианты и мутанты с некоторым уровнем (например, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99%) идентичности последовательностей с соответствующим белком АРОВЕС дикого типа или каталитическим доменом, у которых сохранилась дезаминирующая цитидин активность. Указанные варианты и мутанты можно получить путем вставок, делеций и/или замен аминокислот. Такие замены в некоторых вариантах реализации представляют собой консервативные замены.

[0087] “Аденозиндезаминаза”, также известная как аденозинаминогидролаза, или ADA, представляет собой фермент (EC 3.5.4.4), участвующий в метаболизме пурина. Он необходим для катаболизма аденозина из пищи и для метаболизма нуклеиновых кислот в тканях.

[0088] Неограничивающие примеры аденозиндезаминазы включают тРНК-специфичную аденозиндезаминазу (TadA), тРНК-специфичную аденозиндезаминазу 1 (ADAT1), тРНК-специфичную аденозиндезаминазу 2 (ADAT2), тРНК-специфичную аденозиндезаминазу 3 (ADAT3), РНК-специфичную аденозиндезаминазу B1 (ADARB1), РНК-специфичную аденозиндезаминазу B2 (ADARB2), аденозинмонофосфат-дезаминазу 1 (AMPD1), аденозинмонофосфат-дезаминазу 2 (AMPD2), аденозинмонофосфат-дезаминазу 3 (AMPD3), аденозиндезаминазу (ADA), аденозиндезаминазу 2 (ADA2), подобный аденозиндезаминазе белок (ADAL), содержащий домен аденозиндезаминазы белок 1 (ADAD1), содержащий домен аденозиндезаминазы белок 2 (ADAD2), РНК-специфичную аденозиндезаминазу (ADAR) и РНК-специфичную аденозиндезаминазу B1 (ADARB1).

[0089] Некоторые из дезаминаз азотистых оснований содержат один домен - каталитический, тогда как другие также содержат другие домены, такие как ингибиторный домен, что теперь открыли авторы настоящего изобретения. В некоторых вариантах реализации, следовательно, первый фрагмент содержит только каталитический домен, такой как mA3-CDA1 и hA3B-CDA2. В некоторых вариантах реализации первый фрагмент содержит по меньшей мере каталитический кор каталитического домена. Например, что продемонстрировано в примерах экспериментов, когда mA3-CDA1 был укорочен в остатках 196/197, домен CDA1 все еще сохранял значительные эффективности редактирования (**ФИГ. 10C, 11C, 12C**).

[0090] В настоящем описании исследовали два ингибитора дезаминазы азотистых оснований - mA3-CDA2 и hA3B-CDA1, - которые представляют собой ингибиторные домены соответствующей дезаминазы азотистых оснований. Дополнительные ингибиторы дезаминазы азотистых оснований и ингибиторные домены также были идентифицировали в базах данных белков (см. таблицы 1 и 2). Их биологические эквиваленты (например, с идентичностью последовательностей по меньшей мере приблизительно на 80%, 85%, 90%, 95%, 97%, 98%, 99%, 99,5%, или с одной, двумя или тремя вставками/делециями/заменами аминокислот, и обладающие активностью

ингибирования дезаминазы азотистых оснований) также можно получить с помощью известных в данной области способов, таких как консервативные замены аминокислот. “Ингибитор дезаминазы азотистых оснований”, соответственно, относится к белку или к домену белка, который ингибирует дезаминазную активность дезаминазы азотистых оснований. В некоторых вариантах реализации второй фрагмент содержит по меньшей мере ингибиторный кор ингибиторного белка/домена. Например, что продемонстрировано в примерах экспериментов, когда в mA3-CDA2 сохраняли остатки 334 - 429, CDA2 все еще проявлял эффект ингибирования редактирования оснований (ФИГ. 13В, 14В, 15В).

[0091] В некоторых вариантах реализации слитый белок дополнительно содержит белок, ассоциированный с короткими палиндромными повторами, регулярно расположенными группами (CRISPR) (Cas), необязательно в первом фрагменте, рядом с дезаминазой азотистых оснований или ее каталитическим доменом.

[0092] Термин “белок Cas” или “белок, ассоциированный с короткими палиндромными повторами, регулярно расположенными группами (CRISPR) (Cas),” относится к направляемым РНК ферментам ДНК-эндонуклеазам, связанным с CRISPR (короткие палиндромные повторы, регулярно расположенные группами), системы приобретенного иммунитета у *Streptococcus pyogenes*, а также других бактерий. Белки Cas включают белки Cas9, белки Cas12a (Cpf1), белки Cas12b (ранее известные как C2c1), белки Cas13 и различные сконструированные аналоги. Примеры белков Cas включают SpCas9, FnCas9, St1Cas9, St3Cas9, NmCas9, SaCas9, AsCpf1, LbCpf1, FnCpf1, VQR SpCas9, EQR SpCas9, VRER SpCas9, SpCas9-NG, xSpCas9, RHA FnCas9, KKH SaCas9, NmeCas9, StCas9, CjCas9, AsCpf1, FnCpf1, SsCpf1, PcCpf1, BpCpf1, CmtCpf1, LiCpf1, PmCpf1, Pb3310Cpf1, Pb4417Cpf1, BsCpf1, EeCpf1, BhCas12b, AkCas12b, EbCas12b, LsCas12b, RfCas13d, LwaCas13a, PspCas13b, PguCas13b, RanCas13b и белки, представленные в **таблице А** ниже.

Таблица А. Пример белков Cas

Типы белков Cas	Белки Cas
Белки Cas9	Cas9 из <i>Staphylococcus aureus</i> ... (SaCas9) Cas9 из <i>Neisseria meningitidis</i> ... (NmeCas9) Cas9 из <i>Streptococcus thermophilus</i> ... (StCas9) Cas9 из <i>Campylobacter jejuni</i> ... (CjCas9)

Белки Cas12a (Cpf1)	Cas12a (Cpf1) из <i>Acidaminococcus</i> sp BV3L6 (AsCpf1) Cas12a (Cpf1) из <i>Francisella novicida</i> sp BV3L6 (FnCpf1) Cas12a (Cpf1) из <i>Smithella</i> sp SC K08D17 (SsCpf1) Cas12a (Cpf1) из <i>Porphyromonas crevioricanis</i> (PcCpf1) Cas12a (Cpf1) из <i>Butyrivibrio proteoclasticus</i> (BpCpf1) Cas12a (Cpf1) из <i>Candidatus Methanoplasma termitum</i> (CmtCpf1) Cas12a (Cpf1) из <i>Leptospira inadai</i> (LiCpf1) Cas12a (Cpf1) из <i>Porphyromonas macacae</i> (PmCpf1) Cas12a (Cpf1) из <i>Peregrinibacteria bacterium GW2011 WA2 33 10</i> (Pb3310Cpf1) Cas12a (Cpf1) из <i>Parcubacteria bacterium GW2011 GWC2 44 17</i> (Pb4417Cpf1) Cas12a (Cpf1) из <i>Butyrivibrio</i> sp. NC3005 (BsCpf1) Cas12a (Cpf1) из <i>Eubacteria eligens</i> (EeCpf1)
Белки Cas12b (C2c1)	Cas12b (C2c1) <i>Bacillus hisashii</i> (BhCas12b) Cas12b (C2c1) <i>Bacillus hisashii</i> с мутацией с приобретением функции (см., например, Strecker и др., Nature Communications 10 (статья 212) (2019)) Cas12b (C2c1) <i>Alicyclobacillus kakegawensis</i> (AkCas12b) Cas12b (C2c1) <i>Elusimicrobia bacterium</i> (EbCas12b) Cas12b (C2c1) <i>Laceyella sediminis</i> (Ls) (LsCas12b)
Белки Cas13	Cas13d из <i>Ruminococcus flavefaciens</i> XPD3002 (RfCas13d) Cas13a из <i>Leptotrichia wadei</i> (LwaCas13a) Cas13b из <i>Prevotella</i> sp. P5-125 (PspCas13b) Cas13b из <i>Porphyromonas gulae</i> (PguCas13b) Cas13b из <i>Riemerella anatipestifer</i> (RanCas13b)
Сконструированные белки Cas	Никазы (мутация в одном нуклеазном домене) Каталитически неактивный мутант (dCas9; мутации в обоих нуклеазных доменах) Улучшенные варианты с улучшенной специфичностью (см., например, Chen и др., Nature, 550, 407-410 (2017))

[0093] Сайт расщепления протеазой между первым фрагментом и вторым фрагментом может представлять собой любой известный сайт расщепления протеазой (пептид) для любой протеазы. Неограничивающие примеры протеаз включают протеазу TEV, протеазу TuMV, протеазу PPV, протеазу PVY, протеазу ZIKV и протеазу WNV. Белковые последовательности типичных протеаз и соответствующих их сайтов расщепления представлены в **таблице В**.

Таблица В. Типичные последовательности

Название	Последовательность	SEQ ID NO:
Домен 2 цитидиндезаминазы АРОВЕС3 мыши	MSSSTLSNICLTGKGLPETRFWVEGRMDPLSEEEFYSSQFYNQRVKHLCY YHRMKPYLCYQLEQFNGQAPLKGCLLSEKKGQHAIEILFLDKIRSMELSQ VTITCYLTWSPCPCNAWQLAAFKRDRPDLLIHIYTSRLYFHWKRPFQKG	1

	LCSLWQSGILVDVMDLPQFTDCWTNFMNPKRPFWPWKGLEIISRRTQRR LRRIKESWGLQDLVNDFGNLQLGPPMS	
Домен 1 цитидиндезамина зы АРОВЕС3В человека	MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLL WDTGVFRGQVYFKPQYHAEMCFLSWFCGNQLPAYKCFQITWVSWTPCP DCVAKLAEFLSEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVKI MDYEEFAYCWENFVYNEGQ	2
N-концевой домен протеазы TEV	MGESLFGKPRDYNPISSITICHLTNESDGHTTSLYGIGFGPFIITNKHLE RRNNGTLLVQSLHGVFKVKNTTTLQQHLIDGRDMIIRMPKDFPPFPQK LKFREPQREERICLVTTNFQT	3
C-концевой домен протеазы TEV	MKSMSSMVSDTSCTFPSSDGI FWKHWIQT KDGCQSPLVSTRDGFIVGI HSASNFTNTNNYFTSVPKNFMELLTNQEAQQWVSGWRLNADSVLWGGHK VFMVKPEEPFQPVKEATQ	4
Сайт расщепления протеазой TEV	ENLYFQS	5
Протеаза TuMV	MASSNSMFRGLRDYNPISSNICHLTNVSDGASNSLYGVGFGPLILTNRH LFERNNGELVIKSRHGEFVIKNTTQLHLLPIPDRLLLIRLPKDVPPFP QKLGFRQPEKGERICMVGSNFQTKSITSIVSETSTIMPVENSQFWKHWI STKDGQCGSPMVSTKDGKILGLHSLANFQNSINYFAAFPDDFAEKYLHT IEAHEWVKHWKYNTSAISWGSINIQASQPSGLFKVSKLISDLSTAVYA Q	6
Сайт расщепления протеазой TuMV	GGCSHQ	7
Протеаза PPV	MASKSLFRGLRDYNPIASSICQLNNSGARQSEMFGFGGLIVTNQH LFKRNDGELTIRSHHGEFVVKDTKTLKLLPCKGRDIVIIRLPKDFPPFP RRLQFRTPTTEDRVCLIGSNFQTKSISSTMSETSATYPVDNSHFWKHWI STKDGHCGLPIVSTRDGSILGLHSLANSTNTQNFYAAFPDNFETTYLSN QDNDNWIKQWRYPNDEVCWGSLLKRDIPQSPFTICKLLTDLDGEFVYT Q	8
Сайт расщепления протеазой PPV	QVVVHQSK	9
Протеаза PVY	MASAKSLMRGLRDNFPIAQTVCRKVSVEYGASEMYGFGGAYIVANHH LFRSYNGSMEVQSMHGTFRVKNLHSLSVLPKGRDILIKMPKDFPVFP QKLHFRAPTQNERICLVGTNFQEKYASSIITETSTTYNIPGSTFWKHWI ETDNGHCGLPVVSTADGCIVGIHSLANNAHTTNYSAFDEDFESKYLR NEHNEWVKSQWVYNPDVTLWGFLKLDSTPKGLFKTTKLVDLIDHDVVV EQ	10
Сайт расщепления протеазой PVY	YDVRHQSR	11
Протеаза ZIKV	MASDMYIERAGDITWEKDAEVTGNSPRLDVALDESGDFSIVEEDGPPMR EGGGGSGGGGSGALWDVPAPKEVKKGETTDGVYRVMTRRLLGSTQVGVG VMQEGVFHTMWHVTKGAALRSGEGRDPYWGDKQDLVSYCGPWKLDAA WDGLSEVQLLAVPPGERARNIQTLPGIFKTKDGDIGAVALDYPAGTSGS PILDKCGRVIGLYGNGVVIKNGSYVSAITQKREEETPVECFE	12
Сайт расщепления протеазой ZIKV	KERKRRGA	13
Протеаза WNV	MASSTDMWIERTADISWESDAEITGSSSERVDVRLDDDGNGFQLMNDPGAP WKGSGGGGSGGGGVLWDTPSPKEYKKGDTTGVYRIMTRGLLGSYQAGAG VMVEGVFHTLWHTTKGAALMSGEGRLDPYWGSKEDRLCYGGPWKLQHK WNGQDEVQMIIVVEPGKNVKNVQTKPGVFKTPEGEIGAVTLDFTGTSGS PIVDKNGDVIGLYGNGVIMPNGSYISAIVQGERMDEPIPAGEFEPEML	14

Сайт расщепления протеазой WNV	KQKKRGGK	15
MS2	ACAUGAGGAUCACCCAUGU	16
Остов еНРНК с 2×MS2	GUUUGAGAGCUAGGCCA <u>ACAUGAGGAUCACCCAUGU</u> CUGCAGGGCCUA GCAAGUCAAUAAGGCUAGUCCGUUAUCAACUUG.GCCA <u>ACAUGAGGA</u> <u>UCACCCAUGU</u> CUGCAGGGCC.AAGUGGCACCGAGUCGGUGC	17
PP7	GGAGCAGACGAUAUGGCGUCGCUCC	18
Остов еНРНК с 2×PP7	GUUUGAGAGCUACCGGAGCAGACGAUAUGGCGUCGCUCCGGUAGCAAGU UCAAAUAAGGCUAGUCCGUUAUCAACUUGGAGCAGACGAUAUGGCGUCG CUCCAAGUGGCACCGAGUCGGUGC	19
boxB	GCCCUGAAGAAGGGC	20
Остов еНРНК с 2×boxB	GUUUGAGAGCUAGGGCCCCUGAAGAAGGGCCCCUAGCAAGUCAAUAAGG CUAGUCCGUUAUCAACUUGGGCCCCUGAAGAAGGGCCCCAAGUGGCACCGA GUCGGUGC	21
Белок оболочки MS2 (MCP)	MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSV RQSSAQNRKYTIKVEVPKGAWRSYLNMEITPIFATNSDCELI VKAMQG LLKDGNIPIPSAIAANSIY	22
Белок оболочки PP7 (PCP)	MGSKTIVLSVGEATRTLTEIQSTADRQIFEEKVGPLVGRRLRLTASLRQN GAKTAYRVNLKLDQADVDSGLPKVRYTQVWSHDVTIVANSTEASRKS YDLTKSLVATSQVEDLVVNLVPLGR	23
Белок оболочки boxB (N22p)	MGNARTRRRERRAEKQQAQWKAAN	24
UGI	TNLSDIIEKETGKQLVIQESILMLPEEVEEVI GNKPESDILVHTAYDES TDENVMLLTSDAPEYKFWALVIQDSNGENKIKML	25
P2A	GSGATNFSLLKQAGDVEENPGP	26
T2A	GSGEGRGSLTTCGDVEENPGP	27
E2A	GSGQCTNYALLKLAGDVESNPGP	28

[0094] В некоторых вариантах реализации сайт расщепления протеазой представляет собой саморасщепляющийся пептид, такой как пептиды 2А. “Пептиды 2А” представляют собой вирусные олигопептиды длиной 18 - 22 аминокислоты, которые опосредуют “расщепление” полипептидов в процессе трансляции в эукариотических клетках. Обозначение “2А” относится к особому участку вирусного генома, и различные вирусные 2А, как правило, названы в честь вируса, из которого они происходят. Первым открытым 2А был F2А (вирус ящура), после которого также обнаружили E2А (вирус ринита А лошадей), P2А (2А тешовируса-1 свиньи) и T2А (2А вируса *Thosea asigna*). Несколько неограничивающих примеров пептидов 2А представлены в SEQ ID NO: 26 - 28.

[0095] В некоторых вариантах реализации сайт расщепления протеазой представляет собой сайт расщепления (например, SEQ ID NO: 5) протеазой TEV. В некоторых

вариантах реализации слитый белок дополнительно содержит третий фрагмент, который содержит протеазу TEV или ее фрагмент. В некоторых вариантах реализации фрагмент протеазы TEV в слитом белке неактивен, то есть неспособен сам расщеплять сайт расщепления TEV. Тем не менее, в присутствии остальной части протеазы TEV, данный фрагмент будет способен осуществлять расщепление. Ниже дополнительно описано, что такое расположение обеспечивает дополнительный контроль и гибкость способностей редактирования оснований. Фрагмент TEV может представлять собой N-концевой домен TEV (например, SEQ ID NO: 3) или C-концевой домен TEV (например, SEQ ID NO: 4).

[0096] Можно расположить фрагменты в различном порядке. Неограничивающие примеры включают, с N-концевого участка к C-концевому участку:

- (1) первый фрагмент (например, каталитический домен) – сайт расщепления протеазой – второй фрагмент (например, ингибиторный домен);
- (2) первый фрагмент (например, каталитический домен и белок Cas) – сайт расщепления протеазой – второй фрагмент (например, ингибиторный домен);
- (3) первый фрагмент (например, каталитический домен, белок Cas и N-концевой домен TEV) – сайт расщепления протеазой (например, сайт расщепления TEV) – второй фрагмент (например, ингибиторный домен);
- (4) второй фрагмент (например, ингибиторный домен) – сайт расщепления протеазой (например, сайт расщепления TEV) – первый фрагмент (например, каталитический домен, белок Cas и N-концевой домен TEV); и
- (5) второй фрагмент (например, ингибиторный домен) – сайт расщепления протеазой (например, сайт расщепления TEV) – первый фрагмент (например, белок Cas, каталитический домен и C-концевой домен TEV).

[0097] В некоторых вариантах реализации предложены слитые белки, содержащие первый фрагмент, содержащий первую дезаминазу азотистых оснований (например, цитидиндезаминазу) или ее каталитический домен, и второй фрагмент, содержащий ингибиторный домен второй дезаминазы азотистых оснований, причем первая дезаминаза азотистых оснований отлична от второй дезаминазы азотистых оснований. В некоторых вариантах реализации каждая из первой и второй дезаминазы азотистых оснований независимо выбрана из группы, состоящей из человеческой и мышьиной APOBEC3B (A3B), APOBEC3C (A3C), APOBEC3D (A3D), APOBEC3F (A3F),

АРОВЕС3G (A3G), АРОВЕС3Н (A3H), АРОВЕС1 (A1), АРОВЕС3 (A3), АРОВЕС2 (A2), АРОВЕС4 (A4) и AICDA (AID).

[0098] Указанные слитые белки могут содержать другие фрагменты, такие как ингибитор урацил-ДНК-гликозилазы (UGI) и последовательности ядерной локализации (NLS).

[0099] “Ингибитор урацилгликозилазы” (UGI), который можно получить из бактериофага PBS1 *Bacillus subtilis*, представляет собой малый белок (9,5 кДа) который ингибирует урацил-ДНК-гликозилазу (UDG) *E. coli*, а также UDG из других видов. Ингибирование UDG происходит путем обратимого связывания белков со стехиометрией 1:1 UDG:UGI. UGI способен разделять комплексы UDG-ДНК. Неограничивающий пример UGI обнаружен в фaге AR9 *Bacillus* (YP_009283008.1). В некоторых вариантах реализации UGI содержит последовательность аминокислот SEQ ID NO: 25 или идентичен по меньшей мере на 70%, 75%, 80%, 85%, 90% или 95% последовательности SEQ ID NO: 25 и сохраняет активность ингибирования урацилгликозилазы.

[0100] Указанный слитый белок в некоторых вариантах реализации может содержать одну или более последовательностей ядерной локализации (NLS).

[0101] “Сигнал или последовательность ядерной локализации” (NLS) представляет собой последовательность аминокислот, которая помечает белок для импорта в ядро клетки путем ядерного транспорта. Обычно такой сигнал состоит из одной или более коротких последовательностей положительно заряженных лизинов или аргининов, выставленных на поверхность белка. У различных белков, локализующихся в ядре, могут быть одинаковые NLS. NLS осуществляет функцию, противоположную функции сигнала ядерного экспорта (NES), который направляет белки наружу из ядра. Неограничивающим примером NLS является последовательность внутренней ядерной локализации (iNLS) SV40.

[0102] В некоторых вариантах реализации между каждым из фрагментов в слитом белке необязательно присутствует пептидный линкер. В некоторых вариантах реализации пептидный линкер содержит от 1 до 100 аминокислотных остатков (или 3 - 20, 4 - 15, без ограничения). В некоторых вариантах реализации по меньшей мере 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80% или 90% аминокислотных остатков пептидного линкера

представляют собой аминокислотные остатки, выбранные из группы, состоящей из аланина, глицина, цистеина и серина.

[0103] Для любого слитого белка согласно настоящему изобретению также предложены его биологические эквиваленты. В некоторых вариантах реализации биологические эквиваленты по меньшей мере приблизительно на 70%, 75%, 80%, 85%, 90%, 95%, 98% или 99% идентичны по последовательности эталонному слитому белку. Предпочтительно, биологические эквиваленты сохраняли желательную активность эталонного слитого белка. В некоторых вариантах реализации биологические эквиваленты получают путем включения одной, двух, трех, четырех, пяти или более вставок, делеций, замен аминокислот или их комбинаций. В некоторых вариантах реализации замена представляет собой консервативную замену аминокислоты.

[0104] “Консервативная замена аминокислоты” представляет собой замену, при которой аминокислотный остаток замещают на аминокислотный остаток, содержащий сходную боковую цепь. В данной области определили семейства аминокислотных остатков, содержащих сходные боковые цепи, включая основные боковые цепи (например, лизин, аргинин, гистидин), кислотные боковые цепи (например, аспарагиновая кислота, глутаминовая кислота), незаряженные полярные боковые цепи (например, глицин, аспарагин, глутамин, серин, треонин, тирозин, цистеин), неполярные боковые цепи (например, аланин, валин, лейцин, изолейцин, пролин, фенилаланин, метионин, триптофан), бета-разветвленные боковые цепи (например, треонин, валин, изолейцин) и ароматические боковые цепи (например, тирозин, фенилаланин, триптофан, гистидин). Таким образом, остаток заменимой аминокислоты в полипептиде иммуноглобулина предпочтительно заменяют на другой аминокислотный остаток из того же семейства боковых цепей. В другом варианте реализации цепочку аминокислот можно заменить на сходную по структуре цепочку, которая отличается по порядку и/или составу представителей семейств боковых цепей.

[0105] Неограничивающие примеры консервативных замен аминокислот представлены в таблице ниже, где балл подобия, равный 0 или выше, указывает на консервативную замену между двумя аминокислотами.

Таблица С. Матрица подобия аминокислот

	C	G	P	S	A	T	D	E	N	Q	H	K	R	V	M	I	L	F	Y	W
W	-8	-7	-6	-2	-6	-5	-7	-7	-4	-5	-3	-3	2	-6	-4	-5	-2	0	0	17
Y	0	-5	-5	-3	-3	-3	-4	-4	-2	-4	0	-4	-5	-2	-2	-1	-1	7	10	
F	-4	-5	-5	-3	-4	-3	-6	-5	-4	-5	-2	-5	-4	-1	0	1	2	9		
L	-6	-4	-3	-3	-2	-2	-4	-3	-3	-2	-2	-3	-3	2	4	2	6			
I	-2	-3	-2	-1	-1	0	-2	-2	-2	-2	-2	-2	-2	4	2	5				
M	-5	-3	-2	-2	-1	-1	-3	-2	0	-1	-2	0	0	2	6					
V	-2	-1	-1	-1	0	0	-2	-2	-2	-2	-2	-2	-2	4						
R	-4	-3	0	0	-2	-1	-1	-1	0	1	2	3	6							
K	-5	-2	-1	0	-1	0	0	0	1	1	0	5								
H	-3	-2	0	-1	-1	-1	1	1	2	3	6									
Q	-5	-1	0	-1	0	-1	2	2	1	4										
N	-4	0	-1	1	0	0	2	1	2											
E	-5	0	-1	0	0	0	3	4												
D	-5	1	-1	0	0	0	4													
T	-2	0	0	1	1	3														
A	-2	1	1	1	2															
S	0	1	1	1																
P	-3	-1	6																	
G	-3	5																		
C	12																			

Таблица D. Консервативные замены аминокислот

Для аминокислоты	Замена на
Аланин	D-Ala, Gly, Aib, β -Ala, L-Cys, D-Cys
Аргинин	D-Arg, Lys, D-Lys, Orn D-Orn
Аспарагин	D-Asn, Asp, D-Asp, Glu, D-Glu Gln, D-Gln
Аспарагиновая кислота	D-Asp, D-Asn, Asn, Glu, D-Glu, Gln, D-Gln
Цистеин	D-Cys, S-Me-Cys, Met, D-Met, Thr, D-Thr, L-Ser, D-Ser
Глутамин	D-Gln, Asn, D-Asn, Glu, D-Glu, Asp, D-Asp
Глутаминовая Кислота	D-Glu, D-Asp, Asp, Asn, D-Asn, Gln, D-Gln
Глицин	Ala, D-Ala, Pro, D-Pro, Aib, β -Ala
Изолейцин	D-Ile, Val, D-Val, Leu, D-Leu, Met, D-Met
Лейцин	Val, D-Val, Met, D-Met, D-Ile, D-Leu, Ile
Лизин	D-Lys, Arg, D-Arg, Orn, D-Orn
Метионин	D-Met, S-Me-Cys, Ile, D-Ile, Leu, D-Leu, Val, D-Val
Фенилаланин	D-Phe, Tyr, D-Tyr, His, D-His, Trp, D-Trp
Пролин	D-Pro
Серин	D-Ser, Thr, D-Thr, allo-Thr, L-Cys, D-Cys
Треонин	D-Thr, Ser, D-Ser, allo-Thr, Met, D-Met, Val, D-Val
Тирозин	D-Tyr, Phe, D-Phe, His, D-His, Trp, D-Trp
Валин	D-Val, Leu, D-Leu, Ile, D-Ile, Met, D-Met

Целевая активация слитых белков

[0106] В настоящем изобретении также предложены композиции и способы, в которых слитый белок согласно настоящему изобретению, который содержит как дезаминазу азотистых оснований или ее каталитический домен, так и ингибитор, активируется, когда его активность желательна. Технология показана на **ФИГ. 16**.

[0107] В иллюстративной конфигурации слитый белок (A) содержит (a) первый фрагмент, содержащий дезаминазу азотистых оснований (например, цитидиндезаминазу) или ее каталитический домен, необязательно с белком, ассоциированным с короткими палиндромными повторами, регулярно расположенными группами (CRISPR) (Cas), и первый фрагмент протеазы TEV, (b) второй фрагмент, содержащий ингибитор дезаминазы азотистых оснований, и (c) сайт расщепления протеазой TEV между первым фрагментом и вторым фрагментом. В некоторых вариантах реализации первый фрагмент протеазы TEV отдельно не способен расщеплять сайт расщепления протеазой TEV.

[0108] Когда используют слитый белок, *in vitro* или *in vivo*, для осуществления редактирования генов в клетке, можно ввести две дополнительные молекулы. В одном примере, одна молекула (B) представляет собой единую направляющую РНК (енРНК), которая дополнительно содержит последовательность метки, которая может распознаваться распознающим РНК пептидом. енРНК, в качестве альтернативы, можно заменить на крРНК, которая нацелена на целевой сайт, и РНК CRISPR (крРНК) отдельно или в комбинации с трансктивирующей РНК CRISPR (тракрРНК). Примеры последовательностей метки и соответствующих распознающих РНК пептидов включают MS2/белок оболочки MS2 (MCP), PP7/белок оболочки PP7 (PCP) и boxB/белок оболочки boxB (N22p), последовательности которых представлены в **таблице В**. Молекула (B) может быть предложена в виде последовательности ДНК, кодирующей молекулу РНК.

[0109] Другая дополнительная молекула (C), в некоторых вариантах реализации, содержит второй фрагмент протеазы TEV, соединенный с распознающим РНК пептидом (например, MCP, PCP, N22p). Первый фрагмент TEV и второй фрагмент TEV, в некоторых вариантах реализации, когда они присутствуют вместе, способны расщеплять сайт расщепления протеазой TEV.

[0110] Такое совместное присутствие можно запустить путем связывания молекулы (С) с молекулой (В) благодаря взаимодействию последовательности метки и распознающего РНК белка. Тем временем, слитый белок (А) и молекула (В) будут оба присутствовать в целевом локусе генома для редактирования генов. Следовательно, молекула (В) сближает оба фрагмента протеазы TEV из слитого белка (А) и молекулы (С), что будет активировать протеазу TEV, приводя к удалению ингибитора дезаминазы азотистых оснований из слитого белка и активации редактора оснований. Можно легко понять, что такая активация происходит лишь в целевом сайте генома, не в нецелевых участках одноцепочечной ДНК. По этой причине, редактирование оснований не происходит в участках одноцепочечной ДНК, с которыми не связывается енРНК (что продемонстрировано на **ФИГ. 17 - 19**).

[0111] “Направляющие РНК” представляют собой некодирующие короткие последовательности РНК, которые связываются с комплементарными целевыми последовательностями ДНК. Направляющая РНК сначала связывается с ферментом Cas, и последовательность нРНК направляет указанный комплекс путем спаривания с определенным положением на ДНК, где Cas осуществляет свою эндонуклеазную активность путем разрезания цепи целевой ДНК. “Единая направляющая РНК”, часто называемая просто “направляющей РНК”, относится к синтетической или экспрессируемой единой направляющей РНК (енРНК), которая состоит как из крРНК, так и из тракрРНК в виде единой конструкции. Часть тракрРНК отвечает за активность эндонуклеазы Cas, а часть крРНК связывается со специфическим для мишени участком ДНК. Следовательно, трансактивирующая РНК (тракрРНК, или участок остова) и крРНК являются двумя ключевыми компонентами и соединены друг с другом тетрапетлей, что приводит к образованию енРНК.

[0112] Остов направляющей РНК сам по себе имеет шпильчатую структуру и присоединяется к ферменту эндонуклеазе. Обычный остов имеет структуру, показанную на **ФИГ. 36А** (вверху), которая содержит, от 5'- к 3'-концу, (а) участок повтора, (b) тетрапетлю, (с) антиповтор, который по меньшей мере частично комплементарен участку повтора, (d) шпильку 1, (е) линкер, (f) шпильку 2 и (g) шпильку 3. Последовательность остова, как правило, консервативна, но петли в шпильке 1 и шпильке 3 могут иметь различные последовательности. Что ещё более важно, петли тетрапетли и шпильки 2 можно полностью заменить на еще более длинные

последовательности. Последовательности, такие как РНК-метки (например, MS2, PP7, boxB), можно вставить сюда, позволяя распознавание соответствующим распознающим пептидом. Примеры последовательностей остова показаны ниже.

Таблица Е. Примеры последовательностей остова енРНК

SEQ ID NO:	Последовательность
29	GUUUUAGAGCUA.GAAA.UAGCAAGUUA AAAUAAGGC.UA.GUCCGUUAUCAACUU.GAAA.AAGUGGCACCGAGUCGGUGC
30	GUUUUAGAGCUA.GAAA.UAGCAAGUUA AAAUAAGGC.AU.GUCCGUUAUCAACUU.GAAA.AAGUGGCACCGAUUCGGUGC
196	GUUUUAGAGCUA.GAAA.UAGCAAGUUA AAAUAAGGC.UA.GUCCGUUAUCAACUU.GAAA.AAGUGGCACCGAGUCGGUGC

[0113] Ссылаясь на данные примеры последовательностей остова, фрагмент в положениях 1 - 12 (например, GUUUUAGAGCUA, SEQ ID NO: 197; GUUUGAGAGCUA, SEQ ID NO: 198) представляет собой участок повтора, который образует приблизительно 8 - 12 спариваний оснований с антиповтором, который включает положения 17 - 30 (например, UAGCAAGUUA AAAAU, SEQ ID NO: 199). Петля GAAA (SEQ ID NO: 200) между ними представляет собой тетрапетлю. Как показано в последовательности SEQ ID NO: 17, всю данную петлю можно заменить на последовательность MS2. Шпилька 1 приблизительно содержит положения 31 - 39 и содержит малую петлю (например, UA, AU, AA или UU, без ограничения). Шпилька 1, как правило, содержит 3 - 4 спаривания оснований в стебле. Шпилька 2, включая положения 48 - 61 (например, AACUUGAAAAAGUG, SEQ ID NO: 201), как правило, содержит 4 спаривания оснований в стебле и петлю GAAA (SEQ ID NO: 200), которую можно полностью заменить. Остальные положения 62 - 76 (например, GCACCGAGUCGGUGC, SEQ ID NO: 202; GCACCGAUUCGGUGC; SEQ ID NO: 203) составляют шпильку 3, которая, как правило, содержит 4 спаривания оснований в стебле. Малая петля (U и G здесь в примере) может представлять собой любой нуклеотид.

[0114] Соответственно, последовательность остова можно выразить в виде: GUUUUAGAGCUA $\underline{X_1}$ UAGCAAGUUA AAAUAAGGC $\underline{NN}$ GUCCGUUAUCAACUU $\underline{X_2}$ AAGUGGCACCGANUCGGUGC (SEQ ID NO: 31), где N представляет собой любое основание, и X1 и X2 обозначают любую последовательность нуклеотидов длиной 2 - 50 оснований. В объем терминов “направляющая РНК” и “единая направляющая РНК” входят таковые, содержащие дополнительные последовательности, такие как MS2, PP7 и boxB, вставленные в одну или более петель в РНК.

[0115] Различные варианты реализации и примеры дезаминаз азотистых оснований, каталитических доменов, ингибиторов дезаминазы азотистых оснований и белков Cas представлены в описании. Например, дезаминаза азотистых оснований может представлять собой цитидиндезаминазы и аденозиндезаминазы. Неограничивающие примеры цитидиндезаминаз включают APOBEC1, APOBEC2, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D, APOBEC3F, APOBEC3G, APOBEC3H, APOBEC4 и индуцируемую активацией (цитидин)дезаминазу.

[0116] Неограничивающие примеры аденозиндезаминаз включают тРНК-специфичную аденозиндезаминазу (TadA), тРНК-специфичную аденозиндезаминазу 1 (ADAT1), тРНК-специфичную аденозиндезаминазу 2 (ADAT2), тРНК-специфичную аденозиндезаминазу 3 (ADAT3), РНК-специфичную аденозиндезаминазу B1 (ADARB1), РНК-специфичную аденозиндезаминазу B2 (ADARB2), аденозинмонофосфат-дезаминазу 1 (AMPD1), аденозинмонофосфат-дезаминазу 2 (AMPD2), аденозинмонофосфат-дезаминазу 3 (AMPD3), аденозиндезаминазу (ADA), аденозиндезаминазу 2 (ADA2), подобный аденозиндезаминазе белок (ADAL), содержащий домен аденозиндезаминазы белок 1 (ADAD1), содержащий домен аденозиндезаминазы белок 2 (ADAD2), РНК-специфичную аденозиндезаминазу (ADAR) и РНК-специфичную аденозиндезаминазу B1 (ADARB1).

[0117] Пример белков Cas включают SpCas9, FnCas9, St1Cas9, St3Cas9, NmCas9, SaCas9, AsCpf1, LbCpf1, FnCpf1, VQR SpCas9, EQR SpCas9, VRER SpCas9, RHA FnCas9 и KKH SaCas9 и белки, представленные в **таблице А**.

[0118] Указанные слитые белки могут содержать другие фрагменты, такие как ингибитор урацил-ДНК-гликозилазы (UGI) и последовательности ядерной локализации (NLS), каждый из которых обсуждается в данной заявке.

[0119] Редакторы оснований и способы редактирования оснований, описанные в настоящем описании, можно применять для осуществления высокоспецифичного и высокоэффективного редактирования оснований в геноме различных эукариот.

[0120] В настоящем изобретении предложены композиции и способы. Такие композиции содержат эффективное количество слитого белка, и приемлемый носитель. В некоторых вариантах реализации композиция дополнительно содержит

направляющую РНК с необходимой комплементарностью к целевой ДНК. Такую композицию можно применять для редактирования оснований в образце.

[0121] Слитые белки и композиции можно применять для редактирования оснований. В одном варианте реализации предложен способ редактирования целевого полинуклеотида, включающий приведение в контакт с целевым полинуклеотидом слитого белка согласно настоящему изобретению и направляющей РНК с по меньшей мере частичной комплементарностью последовательности к таковой у целевого полинуклеотида, причем редактирование включает дезаминирование цитозина (С) в целевом полинуклеотиде.

[0122] В одном варианте реализации предложен способ редактирования цитозина на последовательности нуклеиновой кислоты в образце. В некоторых вариантах реализации указанный способ включает приведение в контакт образца со слитым белком согласно настоящему изобретению, или полинуклеотидом, кодирующим слитый белок. В некоторых вариантах реализации дополнительно добавляют подходящую направляющую РНК. Разработка направляющей РНК легко доступна квалифицированному специалисту.

[0123] Приведение в контакт слитого белка (и направляющей РНК) и целевого полинуклеотида может происходить *in vitro*, в частности, в культуре клеток. Если приведение в контакт происходит *ex vivo*, или *in vivo*, слитые белки могут иметь клиническую/терапевтическую значимость. Приведением в контакт *in vivo* может быть введение живому субъекту, такому как человек, животное, дрожжи, растение, бактерия, вирус, без ограничения.

Конфигурации индуцированных и разделенных редакторов оснований

[0124] Исследовали различные конфигурации конструкций, чтобы реализовать разработку индуцированного и разделенного редактора оснований (isplitBE) (ФИГ. 24). Среди исследованных конфигураций (ФИГ. 25) у пары 9 из примера 3 выявили лучшую эффективность редактирования и минимизированное нецелевое редактирование (сильно улучшенную специфичность). В паре 9 используется система двойной енРНК, в которой используют вспомогательную енРНК (венРНК) для нацеливания на сайт, находящийся

в непосредственной близости от основного целевого сайта. Такое двойное нацеливание улучшает специфичность (ФИГ. 32 - 33).

[0125] В конфигурации пары 9 (ФИГ. 25 - 28) ингибитор дезаминазы азотистых оснований высвобождается, только когда обе енРНК связаны с целевыми последовательностями, гарантируя, что дезаминаза азотистых оснований не редактирует нецелевые сайты. Конфигурация пары 9 содержит 6 различных молекул, которые можно получить из двух отдельных конструкций, например (ФИГ. 26А и 34А).

[0126] Первая молекула может содержать лишь белок Cas, который имеет подходящий размер для упаковки в обычный носитель - AAV. Вторая молекула содержит, среди прочего, дезаминазу азотистых оснований (например, APOBEC), ингибитор дезаминазы азотистых оснований (например, mA3-CDA2) и распознающий РНК пептид (например, MCP). Между дезаминазой азотистых оснований и ингибитором дезаминазы азотистых оснований вставлен сайт расщепления протеазой (например, сайт TEV), который позволяет удаление ингибитора дезаминазы азотистых оснований в нужный момент/в нужном положении. Необязательно, вторая молекула дополнительно содержит UGI.

[0127] Третья молекула представляет собой неактивную часть протеазы (например, TEVc), слитую с отличным распознающим РНК пептидом (например, N22p). Четвертая молекула представляет собой обособленную TEVn, которая, в комбинации с первой частью, может осуществлять протеазную активность для удаления ингибитора дезаминазы азотистых оснований из второй молекулы.

[0128] Пятая молекула представляет собой вспомогательную енРНК, содержащую сайт распознавания РНК (например, MS2), распознаваемый распознающим РНК пептидом во 2^{ой} молекуле. Шестая молекула представляет собой обычную енРНК, которая содержит сайт распознавания РНК (например, boxB), распознаваемый распознающим РНК пептидом в 3^{ей} молекуле.

[0129] В правильном целевом сайте в геноме (или РНК), обе венРНК и енРНК будут связываться, и каждая будет привлекать белок Cas к сайту связывания. ВенРНК также будет привлекать 2^{ую} молекула благодаря связыванию MS2-MCP, и енРНК будет привлекать 3^{ью} молекулу благодаря связыванию boxB-N22p. Следовательно, TEVc 3^{ей}РР молекулы находится в контакте с сайтом TEV. Поскольку обособленный TEVn присутствует во всей клетке, он также может присутствовать здесь, что позволяет

гарантировать, что TEVc активна и отщепляет ингибитор дезаминазы азотистых оснований от дезаминазы азотистых оснований в молекуле 2, тем самым активируя дезаминазу азотистых оснований.

[0130] Дополнительно обнаружили, что оптимальное расстояние между сайтом связывания венРНК и сайтом связывания обычной енРНК составляет 34 - 91 п.о. (от РАМ до РАМ), причем венРНК располагается против хода транскрипции от енРНК.

[0131] Более того, даже несмотря на необходимость правильного связывания как венРНК, так и обычной енРНК для предполагаемого редактирования в целевом сайте для обычной енРНК, редактирование в целевом сайте для венРНК нежелательно. В данной заявке обнаружили, что если длина спейсера венРНК (спейсер представляет собой комплементарный мишени участок) составляет 8 - 15 оснований, такой венРНК все еще достаточно, чтобы позволить двойное распознавание для гарантии специфичности связывания, но сильно снизить редактирование в целевом сайте для венРНК.

[0132] В соответствии с одним вариантом реализации настоящего изобретения, следовательно, предложен слитый белок, содержащий первый фрагмент, содержащий: дезаминазу азотистых оснований или ее каталитический домен, ингибитор дезаминазы азотистых оснований, первый распознающий РНК пептид и сайт расщепления протеазой TEV между дезаминазой азотистых оснований или ее каталитическим доменом и ингибитором дезаминазы азотистых оснований.

[0133] В некоторых вариантах реализации слитый белок дополнительно содержит второй фрагмент, содержащий: фрагмент протеазы TEV, который отдельно не способен расщеплять сайт расщепления протеазой TEV, и второй распознающий РНК пептид. В некоторых вариантах реализации слитый белок дополнительно содержит сайт саморасщепления между первым фрагментом и вторым фрагментом.

[0134] В некоторых вариантах реализации слитый белок дополнительно содержит третий фрагмент, содержащий второй фрагмент протеазы TEV, причем первый фрагмент протеазы TEV способен расщеплять сайт расщепления протеазой TEV в присутствии второго фрагмента протеазы TEV. В некоторых вариантах реализации слитый белок дополнительно содержит второй сайт саморасщепления между вторым фрагментом и третьим фрагментом, и после расщепления второго сайта

саморасщепления, слитый белок высвобождает второй фрагмент протеазы TEV, который не соединен с каким-либо распознающим РНК пептидом.

[0135] Также предложена, в одном варианте реализации, система двух направляющих РНК, содержащая: целевую единую направляющую РНК, содержащую первый спейсер, последовательность которого комплементарна целевой последовательности нуклеиновой кислоты, ближайшей к первому сайту РАМ, вспомогательную единую направляющую РНК, содержащую второй спейсер, последовательность которого комплементарна второй последовательности нуклеиновой кислоты, ближайшей ко второму сайту РАМ, белок, ассоциированный с короткими палиндромными повторами, регулярно расположенными группами (CRISPR) (Cas), и дезаминазу азотистых оснований.

[0136] В некоторых вариантах реализации второй сайт РАМ расположен в пределах 150 оснований или, в качестве альтернативы, в пределах 140, 130, 120, 110, 100, 95, 94, 93, 92, 91, 90, 89, 88, 87, 86, 85, 84, 83, 82, 81, 80, 75 или 70 оснований от второго сайта РАМ. В некоторых вариантах реализации второй сайт РАМ расположен по меньшей мере в 10 основаниях или, в качестве альтернативы, по меньшей мере в 15, 20, 25, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 45, 50, 55 или 60 основаниях от первого РАМ. В некоторых вариантах реализации второй сайт РАМ расположен против хода транскрипции от первого сайта РАМ. В некоторых вариантах реализации второй сайт РАМ расположен по ходу транскрипции от первого сайта РАМ. В некоторых вариантах реализации указанное расстояние составляет 20 - 100, 25 - 95, 30 - 95, 34 - 95, 34 - 91, 34 - 90, 35 - 90, 40 - 90, 40 - 84, 45 - 85 или 50 - 80 оснований, без ограничения.

[0137] В некоторых вариантах реализации длина второго (вспомогательного) спейсера составляет 8 - 15 оснований. В некоторых вариантах реализации длина второго спейсера составляет 8 - 14, 8 - 13, 8 - 12, 8 - 11, 8 - 10, 9 - 15, 9 - 14, 9 - 13, 9 - 12, 9 - 11, 9 - 10, 10 - 15, 10 - 14, 10 - 13, 10 - 12, 10 - 11, 11 - 15, 11 - 14, 11 - 13, 11 - 12, 12 - 15, 12 - 14, 12 - 13, или 13 - 15 оснований. Длина первого спейсера, напротив, составляет по меньшей мере 16, 17, 18 или 19 оснований.

[0138] Также здесь описаны различные “разделенные” системы редактирования оснований, которые позволяют упаковать белок Cas и дезаминазу азотистых оснований в отдельные средства доставки (например, AAV).

[0139] В некоторых вариантах реализации предложены пары обычных енРНК и венРНК, которые могут опосредовать эффективное редактирование для создания ранних стоп-кодонов в генах PCSK9, что может принести пользу в клинике. На основании описанных здесь открытий, выбрали подходящие целевые сайты для енРНК и венРНК для превращения кодона, не являющегося стоп-кодоном, в стоп-кодон. Если взять в качестве примера редактирование С на Т/У, то кодоном, не являющимся стоп-кодоном, может быть CAG, CAA или CGA.

[0140] Примеры таких целевых сайтов показаны в таблице 4. Легко понять, что последовательности в таблице 4 используют, чтобы показать расположение мишени. Фактическим енРНК и венРНК, тем не менее, нет необходимости связываться со всей последовательностью. В действительности, для венРНК, например, связывания 8 - 15 нуклеотидов может быть вполне достаточно, как объясняется выше. Соответственно, последовательность спейсера на венРНК может быть комплементарна подпоследовательности любой последовательности, показанной в таблице 4, или даже перекрываться с любой из них. Это также справедливо для енРНК, с предпочтительной длиной спейсера 18 - 24 нуклеотида, без ограничения.

[0141] В одном варианте реализации предложена пара вспомогательной направляющей РНК/направляющей РНК для редактирования последовательности нуклеиновой кислоты PCSK9 человека, причем направляющая РНК специфично нацеливается на первый сайт на нуклеиновой кислоте PCSK9, позволяя осуществлять редактирование оснований, чтобы превратить кодон, не являющийся стоп-кодоном, в стоп-кодон, и причем вспомогательная направляющая РНК специфично нацеливается на второй сайт на нуклеиновой кислоте PCSK9, который находится на расстоянии 20 - 100 оснований от первого сайта. В некоторых вариантах реализации второй сайт находится на расстоянии приблизительно 20 - 100, 25 - 95, 30 - 95, 34 - 95, 34 - 91, 34 - 90, 35 - 90, 40 - 90, 40 - 84, 45 - 85 или 50 - 80 оснований от первого сайта.

[0142] В некоторых вариантах реализации венРНК содержит спейсер, длина которого составляет 8 - 15 оснований. В некоторых вариантах реализации длина спейсера составляет 8 - 14, 8 - 13, 8 - 12, 8 - 11, 8 - 10, 9 - 15, 9 - 14, 9 - 13, 9 - 12, 9 - 11, 9 - 10, 10 - 15, 10 - 14, 10 - 13, 10 - 12, 10 - 11, 11 - 15, 11 - 14, 11 - 13, 11 - 12, 12 - 15, 12 - 14, 12 - 13 или 13 - 15 оснований. В некоторых вариантах реализации енРНК содержит спейсер, длина которого составляет по меньшей мере 16, 17, 18 или 19 оснований.

[0143] Последовательности спейсера для енРНК/венРНК можно легко разработать. Например, для каждого целевого сайта, показанного в таблице 4, спейсер может представлять собой комплементарную последовательность желательной длины (т.е., комплементарен подпоследовательности любой из последовательностей SEQ ID NO: 166 - 180 или 181 - 195). Конкретные примеры пар сайтов связывания включают, без ограничения, SEQ ID NO: 166 и 181; SEQ ID NO: 167 и 182; SEQ ID NO: 168 и 183; SEQ ID NO: 169 и 184; SEQ ID NO: 170 и 185; SEQ ID NO: 171 и 186; SEQ ID NO: 172 и 187; SEQ ID NO: 173 и 188; SEQ ID NO: 174 и 189; SEQ ID NO: 175 и 190; SEQ ID NO: 176 и 191; SEQ ID NO: 177 и 192; SEQ ID NO: 178 и 193; SEQ ID NO: 179 и 194; и SEQ ID NO: 180 и 195.

[0144] Типичные последовательности енРНК/венРНК также были разработаны и исследованы. См. таблицу 3. Кроме того, также предложены полинуклеотидные последовательности, кодирующие вспомогательную направляющую РНК и направляющую РНК.

[0145] С помощью таких пар последовательностей енРНК/венРНК можно осуществить способы инактивации гена PSCK9 в клетке. В некоторых вариантах реализации указанный способ включает приведение в контакт клетки с парой вспомогательной направляющей РНК и направляющей РНК согласно настоящему изобретению, белком, ассоциированным с короткими палиндромными повторами, регулярно расположенными группами (CRISPR) (Cas), и дезаминазой азотистых оснований. Каждый из данных элементов был дополнительно описан в настоящем описании.

Улучшенное праймированное редактирование

[0146] В некоторых вариантах реализации также предложены улучшенные системы праймированного редактирования. В частности, некоторые молекулы направляющей РНК для праймированного редактирования (прнРНК), предложенные в данной заявке, обладают улучшенной стабильностью. Данные прнРНК содержат остов, который, по сравнению с обычной направляющей РНК, содержит одно дополнительное спаривание оснований (см. ФИГ. 36А и 36Е). Если использовать стандартный остов (SEQ ID NO: 31) в качестве матрицы, то улучшенный остов может иметь любую последовательность, представленную в SEQ ID NO: 32 - 43.

[0147] Что обсуждалось выше, обычный остов направляющей РНК имеет структуру, которая содержит, от 5' - к 3'-концу, (а) участок повтора, (b) тетрапетлю, (с) антиповтор, который по меньшей мере частично комплементарен участку повтора, (d) шпильку 1, (е) линкер, (f) шпильку 2 и (g) шпильку 3. Другими словами, остов содержит 4 шпильки. Третья шпилька (при счете от 5' к 3'), также называемая “шпилькой 2”, содержит 4 спаривания оснований в обычном дизайне. В новом дизайне эта шпилька содержит 5 спариваний оснований.

[0148] В одном варианте реализации предложена направляющая РНК, содержащая остов, который содержит, в направлении от 5' к 3', первую структуру шпильки, вторую структуру шпильки, третью структуру шпильки и четвертую структуру шпильки, причем третья шпилька содержит в себе пять спариваний оснований.

[0149] Последовательность остова можно выразить в виде: GUUUNAGAGCUAX₁UAGCAAGUUNAAAUAAGGCNNGUCCGUUAUCAACUUX₂AAGUGGCACCGANUCGGUGC (SEQ ID NO: 31), где N представляет собой любое основание, и X1 и X2 обозначают любую последовательность нуклеотидов длиной 2 - 50 оснований (или 2 - 40, 3 - 40, 4 - 40, 4 - 30, 2 - 30, 4 - 20 оснований). Соответственно, в некоторых вариантах реализации спаривания оснований включают одно между положениями 45 и 55, в соответствии с положениями в SEQ ID NO: 31. В некоторых вариантах реализации остов идентичен по меньшей мере на 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% или 99% последовательности SEQ ID NO: 31 и содержит пять спариваний оснований в третьей шпильке.

[0150] В одном варианте реализации, следовательно, предложена направляющая РНК, содержащая остов, полученный из SEQ ID NO: 31 путем введения спаривания оснований между основаниями в положении 45 и положении 55, и необязательно допускающая одну, две, три, четыре или пять вставок, делеций, замен оснований или их комбинацию, при условии, что сохраняется структура шпилек или функциональность остова/направляющей РНК. В некоторых вариантах реализации остов содержит последовательность, выбранную из группы, состоящей из SEQ ID NO: 32 - 43. В некоторых вариантах реализации длина направляющей РНК составляет по меньшей мере 100 нуклеотидов, или 105, 110, 115, 120, 125, 130, 140 или 150 нуклеотидов. В некоторых вариантах реализации направляющая РНК дополнительно содержит спейсер

(например, 8 - 25 нуклеотидов), матрицу для обратной транскриптазы и/или сайт связывания праймера.

[0151] В некоторых вариантах реализации также предложены улучшенные белки праймирующих редакторов. В одном варианте реализации праймирующий редактор содержит белок Cas и обратную транскриптазу, связанные посредством линкера, который проверили на оптимальность для эффективности функционирования праймирующего редактора. В одном варианте реализации праймирующий редактор содержит последовательность аминокислот SEQ ID NO: 44. В одном варианте реализации праймирующий редактор содержит последовательность аминокислот SEQ ID NO: 45. Исследовали оба данных праймирующих редактора и показали, что они проявляют лучшую эффективность и специфичность редактирования.

[0152] В данной заявке также описаны различные “разделенные” системы праймированного редактирования, которые позволяют упаковать белок Cas и обратную транскриптазу в отдельные средства доставки (например, AAV).

[0153] Также предложены способы осуществления генетического редактирования в клетке в целевом сайте с помощью разделенных систем праймированного редактирования. В некоторых вариантах реализации указанный способ включает введение в клетку первой вирусной частицы, заключающей в себе первую конструкцию, кодирующую белок, ассоциированный с короткими палиндромными повторами, регулярно расположенными группами (CRISPR) (Cas), и второй вирусной частицы, заключающей в себе вторую конструкцию, кодирующую обратную транскриптазу, соединенную с распознающим РНК пептидом. В некоторых вариантах реализации вторая конструкция дополнительно кодирует направляющую РНК, содержащую сайт распознавания РНК, с которым связывается распознающий РНК пептид.

[0154] В некоторых вариантах реализации вторая конструкция дополнительно кодирует направляющую РНК, содержащую сайт распознавания РНК, с которым связывается распознающий РНК пептид. В некоторых вариантах реализации белок Cas выбран из группы, состоящей из SpCas9, FnCas9, St1Cas9, St3Cas9, NmCas9, SaCas9, AsCpf1, LbCpf1, FnCpf1, VQR SpCas9, EQR SpCas9, VRER SpCas9, SpCas9-NG, xSpCas9, RHA FnCas9, KKH SaCas9, NmeCas9, StCas9, CjCas9, AsCpf1, FnCpf1, SsCpf1, PcCpf1, BpCpf1, CmtCpf1, LiCpf1, PmCpf1, Pb3310Cpf1, Pb4417Cpf1, BsCpf1, EeCpf1, BhCas12b,

AkCas12b, EbCas12b, LsCas12b, RfCas13d, LwaCas13a, PspCas13b, PguCas13b и RanCas13b. В некоторых вариантах реализации белок Cas представляет собой SpCas9-NG или xSpCas9.

[0155] Неограничивающие примеры обратных транскриптаз включают обратную транскриптазу вируса иммунодефицита человека (ВИЧ), обратную транскриптазу вируса мышиного лейкоза Молони (MMLV) и обратную транскриптазу вируса миелобластома птиц (AMV).

ПРИМЕРЫ

Пример 1. Слитые редакторы оснований со сниженной активностью нецелевого редактирования.

[0156] Единые направляющие РНК (енРНК) и редакторы оснований (BE), упомянутые в примерах, предназначены для SpCas9, если конкретно не указано, например, енРНК для SaCas9 (Sa-енРНК). Для того чтобы проверить, может ли современная система редактирования оснований вызывать мутации С на Т в участках оцДНК, использовали нишау SaD10A и Sa-енРНК для получения одноцепочечного разрыва ДНК (ОЦР), который может запускать рецессию конца с образованием участка оцДНК. (фиг. 1А, 2А и 3А). Котрансфицировали SaD10A, Sa-енРНК (Sa-sgSITE31, Sa-sgSITE42 и Sa-sgF1) с двумя опубликованными BE, т.е., BE3 и hA3A-BE3, или с пустым вектором (фиг. 1В, 2В и 3В) и определили мутагенез вокруг участков оцДНК, запускаемый SaD10A. В трех исследованных сайтах (Sa-SITE31, Sa-SITE42 и Sa-F1) экспрессия BE3 или hA3A-BE3 вызывала мутации С на Т, тогда как экспрессия пустого вектора не вызывала их (фиг. 1С, 2С и 3С). Данные результаты свидетельствовали о том, что современные редакторы оснований, которые содержат каталитически активные цитидиндезаминазы, действительно вызывают непредусмотренные мутации в нецелевых участках оцДНК (фиг. 1, 2 и 3).

[0157] Для ингибирования активности цитидиндезаминазы в нецелевых сайтах, например, участках оцДНК, было предложено соединить редакторы оснований с ингибитором редактирования оснований. APOBEC3 мыши (mA3) содержит два домена цитидиндезаминазы (CDA) (CDA1 и CDA2, фиг. 4А, 5А, 6А), и использование полноразмерного mA3 в mA3-BE3 (фиг. 4В, 5В, 6В) не вызывало редактирования С на

Т в трех исследованных целевых сайтах (фиг. 4С, 5С, 6С). Тем не менее, mA3CDA1-BE3, который получали путем удаления mA3CDA2 из mA3-BE3 (фиг. 4В, 5В, 6В), вызывал значительное редактирование С на Т (фиг. 4С, 5С, 6С). Данные результаты позволяют предположить, что mA3CDA2 является природным ингибитором редактирования оснований. Следовательно, добавляли mA3CDA2 к N-концу трех активных BE, т.е., mA3CDA1-BE3, BE3 и hA3A-BE3, с получением mA3rev-BE3, mA3CDA2-BE3 и mA3CDA2-hA3A-BE3 (фиг. 4В, 5В, 6В). Как и ожидалось, добавление mA3CDA2 к N-концу явно снижало эффективности редактирования оснований (фиг. 4С, 5С, 6С).

[0158] Затем было проанализировано, может ли расщепление mA3CDA2 восстановить эффективность редактирования оснований. Саморасщепляющиеся пептиды 2А вставляли между mA3CDA2 и остальной частью BE в mA3rev-BE3, mA3CDA2-BE3 и mA3CDA2-hA3A-BE3 с получением mA3rev-2A-BE3, mA3CDA2-2A-BE3 и mA3CDA2-2A-hA3A-BE3 (фиг. 4В, 5В, 6В). Соответственно, эффективности редактирования оснований восстанавливались в mA3rev-2A-BE3, mA3CDA2-2A-BE3 и mA3CDA2-2A-hA3A-BE3 (фиг. 4С, 5С, 6С), свидетельствуя о том, что ингибирование mA3CDA2 зависит от его ковалентного соединения с BE. Также был проведен поиск в базах данных белков в отношении доменов, сходных с коровой последовательностью mA3CDA2, и было обнаружено, что по меньшей мере 44 белка содержат сходные домены (таблица 1).

[0159] АРОВЕС3В человека (hA3В) также содержит два домена цитидиндезаминазы (CDA) (CDA1 и CDA2, фиг. 7А, 8А, 9А), и применение полноразмерного hA3В в hA3В-BE3 (фиг. 7В, 8В, 9В) вызывало лишь относительно низкие уровни редактирования С на Т в трех исследованных целевых сайтах (фиг. 7С, 8С, 9С). Тем не менее, hA3BCDA2-BE3, который получали путем удаления hA3BCDA1 из hA3В-BE3 (фиг. 7В, 8В, 9В), вызывал более высокий уровень редактирования С на Т (фиг. 7С, 8С, 9С). Кроме того, между hA3BCDA1 и hA3BCDA2 вставляли саморасщепляющиеся пептиды 2А с получением hA3В-2A-BE3 (фиг. 7В, 8В, 9В), который вызывал более высокие эффективности редактирования С на Т, чем hA3В-BE3 (фиг. 7С, 8С, 9С). Данные результаты указывают на то, что hA3BCDA1 является еще одним ингибитором редактирования оснований, и ингибирование hA3BCDA1 зависит от его ковалентного соединения с BE. Также был проведен поиск в базе данных белков доменов, сходных с hA3BCDA1, и было обнаружено, что по меньшей мере 43 белка содержат сходные домены (таблица 2).

[0160] Далее планировалось использовать mA3 для разработки новых BE. Два BE - mA3rev-BE3 и mA3rev-2A-BE3 – получали путем разделения mA3 между аминокислотами (AK) 207 и AK 208, а затем определяли, где расщеплять, чтобы mA3CDA2 мог сохранить наибольшую эффективность редактирования (фиг. 10A, 11A, 12A). Так как mA3CDA1 заканчивается на аминокислоте (AK) 154 и mA3CDA2 начинается с AK238, расщепляли mA3CDA2 в AK196/AK197, AK215/AK216, AK229/AK230 и AK237/AK238 с получением mA3rev-BE3-196, mA3rev-2A-BE3-196, mA3rev-BE3-215, mA3rev-2A-BE3-215, mA3rev-BE3-229, mA3rev-2A-BE3-229, mA3rev-BE3-237 и mA3rev-2A-BE3-237 (фиг. 10B, 11B, 12B). Хотя разделение mA3 в AK207/AK208 и AK215/AK216 сохраняло наибольшие эффективности редактирования, результаты также показали, что сайты разделения, располагающиеся от AK196/AK197 до AK237/AK238, как правило, сохраняют значительные эффективности редактирования (фиг. 10C, 11C, 12C).

[0161] Кроме того, пытались определить минимальный участок mA3, который оказывает ингибиторное действие на редактирование оснований. Удаляли различные N-концевые части mA3CDA2 в mA3rev-BE-237, чтобы разработать mA3rev-BE-237-Del-255, mA3rev-BE-237-Del-285 и mA3rev-BE-237-Del-333, которые содержат AK256-AK429, AK286-AK429 и AK334-AK429 части mA3 в качестве ингибитора редактирования оснований, соответственно (фиг. 13A, 14A, 15A). По сравнению с mA3rev-BE-237, который содержит AK238-AK429 часть mA3, для mA3rev-BE-237-Del-255, mA3rev-BE-237-Del-285 и mA3rev-BE-237-Del-333 выявили сходные эффективности редактирования (фиг. 13B, 14B, 15B). Данные результаты свидетельствовали о том, что AK334-AK429 часть mA3 все еще оказывала ингибиторное действие на редактирование оснований.

[0162] Для того чтобы разработать редакторы оснований, которые не вызывают мутацию С на Т в нецелевых участках оцДНК, заменяли сайт саморасщепления 2A на сайт расщепления протеазы TEV в mA3rev-2A-BE3, а затем соединяли N-концевую часть протеазы TEV (TEVn) [Gray и др., 2010, Cell, doi: 10.1016/j.cell.2010.07.014] с С-концом mA3rev-2A-BE3 с другим сайтом расщепления TEV. Вновь разработанный BE назвали BEsafe. Кроме того, поместили одну петлю MS2 в енРНК с получением MS2-енРНК [Ma и др., 2016, Nature Biotechnology, doi: 10.1038/nbt.3526], а затем соединили С-концевую часть протеазы TEV (TEVc) с белком оболочки MS2 (MCP), который может

связывать петлю MS2 (фиг. 16A). Когда BEsafe, MS2-енРНК и MCP-TEVc коэкспрессировали, TEVn в составе BEsafe и TEVc из MCP-TEVc, которую может привлечь MS2-енРНК, будут соединяться и восстанавливать протеазную активность в целевом сайте. Последующие расщепления в сайтах TEV удалят mA3CDA2 и TEVn с N- и C-концов BEsafe, и полученный в результате этого mA3CDA1-BE3 может вызывать эффективное редактирование оснований в целевом сайте (фиг. 16A). Напротив, BEsafe не будет вызывать мутации С на Т в нецелевых участках оцДНК, так как цитидиндезаминазная активность mA3CDA1 ингибируется mA3CDA2 (фиг. 16B).

[0163] Затем сравнивали эффективность функционирования BEsafe и hA3A-BE3 в целевых сайтах и нецелевых участках оцДНК (фиг. 17, 18, 19). Котрансфицировали плазмиду, экспрессирующую Sa-енРНК и SaD10A, которая может приводить к образованию оцДНК в целевых сайтах Sa-енРНК (фиг. 17A, 18A, 19A), с плазмидой экспрессии hA3A-BE3 и плазмидой экспрессии енРНК, с плазмидой экспрессии BEsafe и плазмидой, экспрессирующей MS2-енРНК и MCP-TEVc, или с плазмидой экспрессии MCP-TEVc и плазмидой, экспрессирующей MS2-енРНК и BEsafe (фиг. 17B, 18B, 19B). Исследовали частоты мутации С на Т в нецелевых участках оцДНК (целевых сайтах Sa-енРНК, ортогональных таковым для SpCas9) (фиг. 17C, 18C, 19C) и эффективности редактирования оснований в целевых сайтах енРНК hA3A-BE3 и BEsafe, оба из которых происходили из SpCas9 (фиг. 17D, 18D, 19D). Было обнаружено, что BEsafe не вызывал какую-либо мутацию С на Т в нецелевых участках оцДНК (целевых сайтах Sa-енРНК), но hA3A-BE3 вызывал очевидные мутации (фиг. 17C, 18C, 19C). В целевых сайтах енРНК BEsafe вызывал редактирование оснований, сравнимое с таковым для hA3A-BE3, тогда как экспрессия обоих MS2-енРНК и BEsafe с одной единой плазмиды привела к более высоким эффективностям редактирования оснований, чем экспрессия только BEsafe с одной плазмиды (фиг. 17D, 18D, 19D).

[0164] Редакторы оснований и способ редактирования оснований, описанные в настоящем описании, можно применять для осуществления высокоспецифичного и высокоэффективного редактирования оснований в геноме различных эукариот.

[0165] Впервые была создана система редактирования оснований, позволяющая избежать возникновения мутаций С на Т в нецелевых участках оцДНК и вызвать эффективное редактирование оснований в целевых сайтах. Систему редактирования оснований BEsafe и сопроводительные способы, описанные в настоящем изобретении,

можно применять для осуществления высокоспецифичного редактирования оснований, которое невозможно осуществить с помощью существующих на данный момент ВЕ, так как цитидиндезаминазы в современных ВЕ могут вызывать непредусмотренные мутации в нецелевых участках оцДНК. Важно отметить, что высокая специфичность и эффективность данной системы редактирования оснований ВЕsafe будет стимулировать потенциальный переход в клиническую практику, особенно в области генотерапии, которая включает обращение связанных с заболеванием мутаций.

Таблица 1. Домены, сходные с коровой последовательностью mA3CDA2.

Название	Последовательность	SEQ ID NO:
Кор домена 2 цитидиндезаминазы АРОВЕС3 мыши (AK282-AK355)	SEKGGQHAIEILFLDKIRSMELSQVTITCYLTWSPCPNCAWQLAAFKRDRP DLILHIYTSRLYFHWKRPFQKGLC	48
A3 Mus spicilegus (AK248-AK321)	SEKGGQHAIEILFLDKIRSMELSQVTITCYLTWSPCPNCAWQLAAFKRDRP DLIPHIYTSRLYFHWKRPFQKGLC	49
A3 Cricetulus longicaudatus (AK249-AK322)	SEKGGQHAIEILFLDKIRSMELSQVTITCYLTWSPCPNCAWRLAAFKRDRP DLILHIYTSRLYFHWKRPFQKGLC	50
A3 Mus terricolor (AK248-AK321)	SEKGGQHAIEILFLNKRSMELSQVTITCYLTWSPCPNCAWQLAAFKKDRP DLILHIYTSRLYFHWKRPFQKGLC	51
A3 Mus caroli (AK260-AK333)	SKKGGQHAIEILFLDKIRSMELSQVTITCYLTWSPCPNCAWQLAAFKRDHP DLILHIYTSRLYFHWKRPFQKGLC	52
A3 Mus pahari (AK263-AK336)	SKKGGQHAIEILFLEKIRSMELSQMRITCYLTWSPCPNCAWQLAAFKKDRP DLILHIYTSRLYFHWRRIFQKGLC	53
A3 Mus shortridgei (AK233-AK306)	SKKGGQHAIEILFLEKIRSMELSQMRITCYLTWSPCPNCAWQLAAFKKDRP DLILHIYTSRLYFHWRRIFQKGLC	54
A3 Mus setulosus (AK29-AK302)	SKKGGQHAIEILFLDKIRSMELSQVRITCYLTWSPCPNCAWQLETFKKDRP DLILHIYTSRLYFHWKRAFQEGLC	55
A3 Grammomys surdaster (AK270- AK344)	SKKGGKHAIEILFLDKMWSMEELSQVRITCYLTWSPCPNCARQLAAFKKDH PGLILRIYTSRLYFYWRRKFQKGLC	56
A3 Rattus norvegicus (AK256-AK328)	KKGEQHVEILFLEKMRSMEELSQVRITCYLTWSPCPNCARQLAAFKKDHPD LILRIYTSRLYFYWRKKFQKGLC	57
A3 Mastomys coucha (AK258-AK331)	SKKGRQHAIEILFLEKVRSMQLSQVRITCYLTWSPCPNCAWQLAAFKMDHP DLILRIYASRLYFHWRRAFQKGLC	58
A3B Cricetulus griseus (AK235- AK307)	NKKGKHAIEILFIDEMRSLELGQVQITCYLTWSPCPNCAQELAAFKSDHPD LVLRIYTSRLYFHWRRKYQEGLC	59
A3 Peromyscus leucopus (AK266- AK338)	NKKGKHAIEILFIDEMRSLELGQARITCYLTWSPCPNCAQKLAFAFKKDHPD LVLRVYTSRLYFHWRRKYQEGLC	60
A3 Mesocricetus auratus (AK268- AK340)	NKDKKHAIEILFIDKMRSLELCQVRITCYLTWSPCPNCAQELAAFKKDHPD LVLRIYTSRLYFHWRRKYQEGLC	61
A3B Microtus ochrogaster (AK266- AK338)	NKKGKHAIEILFIDEMRSKLKSQERITCYLTWSPCPNCAQELAAFKRDHPG LVLRIYASRLYFHWRRKYQEGLC	62
A3 Nannospalax galili (AK231- AK302)	NKRAKHAIEILLIDMMRSMELGQVQITCYITWSPCPTCAQELAAFKQDHPD LVLRIYASRLYFHWKRKFQKGL	63

A3 Meriones unguiculatus (AK233-AK305)	NKKGRHAEICLIDEMRSLGLGKAQITCYLTWSPCRKCAQELATFKKDHPD LVLRVYASRLYFHWRSRYQQGLC	64
A3 Dipodomys ordii (AK256-AK330)	NKKGHAEIRFIERIRSMGLDPSQDYQITCYLTWSPCLDCAFKLAKLKDD FPRLTLRIFTSRLYFHWIRKFQKGL	65
A3 Jaculus jaculus (AK303-AK374)	NKKKGHAEARFVDKMRSMQLDHALITCYLTWSPCLDCSQKLAALKRDHPG LTLRIFTSRLYFHWVKKFQEGEGL	66
A3H Chinchilla lanigera (AK86-AK161)	SPQKGHHAESRFIKRISSMDLDRSRSYQITCFLTWSPPSCAQELASFKR AHPHLRFQIFVSRLYFHWKRSYQAGL	67
A3 Heterocephalus glaber (AK277-AK350)	KKGYHAESRFIKRICSMGLGQDQSYQVTCFLTWSPPHCAQELVSEFKRAH PHLRLQIFFTARLFFHWKRSYQEGEGL	68
A3 Octodon degus (AK256-AK329)	KKGQHAIEIRFIERIHSMDLDQARSYQITCFLTWSPPFCAQELASFKSTH PRVHLQIFVSRLYFHWKRSYQEGEGL	69
A3 Urocitellus parryii (AK256-AK330)	NKKGHAEIRFIKKIRSLDLQSQNYEVTCTWSPCPDCAQELVALTRS HPHVRLRLFTSRLYFHWFWFSQEGEGL	70
A3H Aotus nancymae (AK75-AK146)	NRHAEICFIDEIESMGLDKTQCYEVTCTWSPCPSCAQKLAFTKAQVH LNLRI FASRLYYHWRSSYQKGL	71
A3H Cebus capucinus imitator (AK55-AK126)	NRHAEICFIDEIESMGLDKTQCYEVTCTWSPCPSCAQKLVAFKAQDH LNLRI FASRLYYHWRRRYKEGL	72
A3H Saimiri boliviensis boliviensis (AK56-AK125)	HVEICFIDKIASMELDKTQCYDVTCYLTWSPCPSCAQKLAFAKAQDHLN LRIFASRLYYHWRSSYQKGL	73
A3H Homo sapiens (AK49-AK123)	NKKKCHAEICFINEIKSMGLDETQCYQVTCYLTWSPCSSCAWELVDFIKA HDHLNLGIFASRLYYHWCKPQQKGL	74
ARP10 Homo sapiens (AK48-AK123)	ENKKKCHAEICFINEIKSMGLDETQCYQVTCYLTWSPCSSCAWELVDFIK AHDHLNLGIFASRLYYHWCKPQQKGL	75
A3H Pan paniscus (AK49-AK123)	NKKKCHAEICFINEIKSMGLDETQCYQVTCYLTWSPCSSCAWKLVDIFIQA HDHLNLRI FASRLYYHWCKPQQEGEGL	76
A3H Symphalangus syndactylus (AK49-AK123)	NKKKRHAEIRFINKIKSMGLDETQCYQVTCYLTWSPCPSCAWELVDFIKA HDHLNLGIFASRLYYHWCRRHQEGEGL	77
A3H Macaca mulatta (AK49-AK123)	NKKKDHAIEIRFINKIKSMGLDETQCYQVTCYLTWSPCPSCAGELVDFIKA HRHLNLRI FASRLYYHWRPNYQEGEGL	78
A3H Theropithecus gelada (AK54-AK128)	NKKKEHAEIRFINKIKSMGLDETQCYQVTCYLTWSPCPSCAGKLVDIFIKA HHHLNLRI FASRLYYHWRPNYQEGEGL	79
A3H Mandrillus leucophaeus (AK49-AK123)	NKKKHAEIHFIKIKSMGLDETQCYQVTCYLTWSPCPSCARELVDFIKA HRHLNLRI FASRLYYHWRPHYQEGEGL	80
A3 Bos grunniens (AK74-AK148)	NKKQRHAEIRFIDKINSLDLNPSQSYKIICYITWSPCPNCANELVNFITR NNHLKLEIFASRLYFHWIKPFKMGL	81
A3 Bubalus bubalis (AK74-AK148)	NKKQRHAEIRFIDKINSLDLNPSQSYKIICYITWSPCPNCASELVDFITR NDHLDLQIFASRLYFHWIKPFKRGL	82
A3H Odocoileus virginianus texanus (AK209-AK283)	NKKQRHAEIRFIDKINSLNDRRQSYKIICYITWSPCPRCASELVDFITG NDHLNLQIFASRLYFHWKKPFQKGL	83
A3 Sus scrofa (AK51-AK125)	NKKKRHAEIRFIDKINSLNLDQNCYRIICYVTWSPCHNCAKELVDFISN RHHLNLQIFASRLYFHWVRCYQKGL	84
A3B Ceratotherium simum simum (AK232-AK306)	NKKKRHAEIRFIDKIKSLGLDRVQSYEITCYITWSPCPTCALELVAFTRD YPRLSLQIFASRLYFHWRRRSIQGL	85

A3H Equus caballus (AK79-AK153)	NKKKRHAEIRFIDKINSLGLDQDQSYEITCYVTWSPCATCACKLIKFTRK FPNLSLRIFVSRLYYHWFRQNQQGL	86
A3B Enhydra lutris kenyonii (AK243-AK316)	KKKRHAEIRFIDSIRALQLDQSRFEITCYLTWSPCPTCAKELAMFVQDH PHISLRLFASRLYFHWLRKYQEGEGL	87
A3H Leptonychotes weddellii (AK50-AK123)	KKKRHAEIRFIDNIKALRLDTSQRFEITCYVTWSPCPTCAKELVAFVRDH RHISLRLFASRLYFHWLRNKKGL	88
A3F Ursus arctos horribilis (AK552-AK626)	NKKKRHAEIRFIDKIRSLQDSSQTFEITCYVTWSPCPTCAEELVAFVRD HPHVRLRLFASRLYFHWLRKYQEGEGL	89
A3H Panthera leo bleyenberghi (AK50-AK124)	NKKKRHAEICFIDKIKSLTRDTSQRFEIICYITWSPCFCAEELVAFVKD NPHLSLRIFASRLYVHWLRKYQQGL	90
A3H Panthera tigris sumatrae (AK50-AK124)	NKKKRHAEICFIDKIKSLTRDTSQRFEIICYITWSPCFCAEELVAFVKD NPHLSLRIFASRLYVHWLRKYQQGL	91
A3 Tupaia belangeri (AK46-AK120)	NKKHRHAEVRFIAKIRSMSLDLDQKHQLTCYLTWSPCPSCAQELVTFMAE SRHLNLQVFVSRLYFHWQRDFQQGL	92

Таблица 2. Домены, сходные с hA3BCDA1.

Название	Последовательность	SEQ ID NO:
A3B Gorilla, (AK29-AK138)	GRSYNWLCYEVKIKRGRSNLLWNTGVFRGQMYSQPEHHAEMCFLSWFCGN QLPAYKCFQITWVSWTTPCPDCVAKLAEFLEAHPNVTLTI STARLYYYWE RDYRRALCRL	93
A3B Pan paniscus (AK29-AK138)	GRSYTWLCYEVKIRRGHSNLLWDTGVFRGQMYSQPEHHAEMYFLSWFCGN QLPAYKCFQITWVSWTTPCPDCVAKLAEFLEAHPNVTLTI SAARLYYYWE RDYRRALCRL	94
A3B Pan troglodytes (AK29-AK138)	GRSYTWLCYEVKIRRGHSNLLWDTGVFRGQMYSQPEHHAEMCFLSWFCGN QLSAYKCFQITWVSWTTPCPDCVAKLAKFLAHPNVTLTI SAARLYYYWE RDYRRALCRL	95
A3F Gorilla (AK30-AK137)	RNTVWLCYEVKTKGPSRPLDAKI FRGQVYFEPQYHAEMCFLSWFCGNQL PAYKCFQITWVSWTTPCPDCVAKLAEFLEAHPNVTLTI SAARLYYYWE	96
A3F Pan troglodytes (AK30-AK137)	RNTVWLCYEVKTKGPSRPLDTKI FRGQVYFEPQYHAEMCFLSWFCGNQL PAYKCFQITWVSWTTPCPDCVAKLAEFLEAHPNVTLTI SAARLYYYWERD YRRALCRL	97
A3F Homo sapiens (AK30-AK137)	RNTVWLCYEVKTKGPSRPLDAKI FRGQVYSQPEHHAEMCFLSWFCGNQL PAYKCFQITWVSWTTPCPDCVAKLAEFLEAHPNVTLTI SAARLYYYWERD YRRALCRL	98
A3F Macaca leonine (AK30-AK137)	RNTVWLCYEVKTRGPSMTWGTKI FRGQVCFEPQYHAEMCFLSRFCGNQL PAYKRFQITWVSWTTPCPDCVAKVAEFLEAHPNVTLTI SAARLYYYWETD YRRALCRL	99
A3F Macaca nemestrina (AK30-AK137)	RNTVWLCYEVKTRGPSMTWGTKI FRGQVCFEPQYHAEMCFLSRFCGNQL PAYKRFQITWVSWTTPCPDCVAKVAEFLEAHPNVTLTI SAARLYYYWETD YRRALCRL	100
A3F Rhinopithecus roxellana (AK30-AK137)	RNTVWLCYEVKTRGPSMTWGTKI FRGQVYFEPQYHAEMCFLSWFCGNQL PAYKRFQITWVSWTTPCPDCVAKVAEFLEAHPNVTLTI SAARLYYYWETD YRRALCRL	101
A3F Mandrillus leucophaeus (AK30-AK130)	RNTVWLCYKVKTRGPSMTWGTKI FRGQVYFQPYHAEMCFLSWFCGNQL PAYKRFQITWVSWTTPCPDCVVKVAEFLEAHPNVTLTI SAARLYYYWETD Y	102
A3F Macaca mulatta (AK30-AK137)	RNTVWLCYEVKTRGPSMTWDTKI FRGQVYSKPEHHAEMCFLSRFCGNQL PAYKRFQITWVSWTTPCPDCVAKVAEFLEAHPNVTLTI SAARLYYYWETD YRRALCRL	103
A3F Theropithecus gelada (AK30-AK137)	RNTVWLCYEVKTRGPSMTWGTKI FRGQVYFQPYHAEMCFLSRFCGNQL PAYKRFQITWVSWNCPDCVAKVIEFLAHPNVTLTI SAARLYYYWGRD WRRALRRL	104
A3B Cercopithecus atys (AK29-AK138)	GRSYTWLCYEVKIRKDPKLPWYTGTVFRGQVYSKPEHHAEMCFLSRFCGN QLPAYKRFQITWVSWNCPDCVAKVIEFLAHPNVTLTI SAARLYYYWS RDWQRALCRL	105

A3B <i>Macaca fascicularis</i> (AK29-AK138)	GRSYTWLCYEVKIRKDP SKLPWDTGVFRGQVYSKPEHHAEMCFLSRFCGNQLPAYKRFQITWVFSWNP CPDCVAKVIEFLAEHPNVTLTISTARLYYYWG RDWQRALCRL	106
A3B <i>Macaca mulatta</i> (AK29-AK138)	GRSYTWLCYEVKIRKDP SKLPWDTGVFRGQVYSKPEHHAEMCFLSRFCGNQLPAYKRFQITWVFSWNP CPDCVAKVIEFLAEHPNVTLTISTARLYYYWG RDWQRALCRL	107
A3B <i>Macaca leonina</i> (AK29-AK138)	GRSYTWLCYEVKIRKDP SKLPWDTGVFRGQVYSKPEHHAEMCFLSRFCGNQLPAYKRFQITWVFSWNP CPDCVAKVIEFLAEHPNVTLTISTARLYYYWG RDWQRALCRL	108
A3B <i>Mandrillus leucophaeus</i> (AK29-AK138)	GRSYTWLCYEVKIRKDP SKLPWDTGVFRGQVYSKPEHHAEMCFLSRFCGNQLPAYKRFQITWVFSWNP CPDCVAKVIEFLAEHPNVTLTIFTARLYYYWG RDWQRALCRL	109
A3B <i>Macaca nemestrina</i> (AK29-AK138)	GRSYTWLCYEVKIRKDP SKLPWDTGVFRGQVYSKPEHHAEMCFLSRFCGNQLPAYKRFQITWVFSWNP CPDCVAKVTEFLAEHPNVTLTISTARLYYYWG RDWQRALCRL	110
A3F <i>Rhinopithecus bieti</i> (AK29-AK138)	GRSYTWLCYEVKIRKDP SKLPWDTGVFRGQVYSEPEHHAEMYFLSWFCGNQLPAYKRFQITWVFSWTP CPDCVAKVAEFLTEHPNVTLTISAA RLYYYRG RDWRRALCRL	111
A3B <i>Rhinopithecus roxellana</i> (AK29-AK138)	GRSYTWLCYEVKIRKDP SKLPWDTGVFRGQVYSEPEHHAEMYFLSWFCGNQLPAYKRFQITWVFSWTP CPDCVAKVAEFLTEHPNVTLTISAA RLYYYRG RDWRRALCRL	112
A3B <i>Chlorocebus sabaeus</i> (AK29-AK138)	GRSYTWLCYEVKIRKDP SKLPWDTGVFRGQMYSKPEHHAEMCFLSWFCGNQLPAHKRFQITWVFSWTP CPDCVAKVAEFLAEYPNVTLTISAA RLYYYWE TDYRRALCRL	113
A3B <i>Nomascus leucogenys</i> (AK30-AK138)	RSYTWLCYEVKIRKDP SKLPWDTGVFRGQMYFQPEYHAEMCFLSWFCGNQLPAYKRFQITWVFSWTP CPDCVAKVAVFLAEHPNVTLTISAA RLYYYWEK DWQRALCRL	114
A3F <i>Cercopithecus atys</i> (AK29-AK138)	GRSYTWLCYEVKIKKPSKLLWDTGVFQGGQVYFQPYHAEMCFLSRFCGNQLPAYKRFQITWVFSWNP CPDCVAKVTEFLAEHPNVTLTISAA RLYYYWE KDXRRALRRL	115
A3F <i>Papio anubis</i> (AK29-AK138)	GRSYTWLCYEVKIKEDPSKLLWDTGVFQGGQVYFQPYHAEMCFLSRFCGNQLPAYKRFQITWVFSWNP CPDCVAKVTEFLAEHPNVTLTISAA RLYYYWG RDWRRALRRL	116
A3D <i>Chlorocebus aethiops</i> (AK29-AK150)	GRRYTWLCYEVKIKKDP SKLPWDTGVFPQGQVRPKFQSNRRYEVYFQPYHAEMYFLSWFCGNQLPAYKH FQITWVFSWNP CPDCVAKVTEFLAEHRNVTLTISAA RLYYYWGKDWRRALCRL	117
A3D <i>Chlorocebus sabaeus</i> (AK29-AK134)	GRRYTWLCYEVKIKKDP SKLPWDTGVFPQGQPYHAEMYFLSWFCGNQLPAYKH FQITWVFSWNP CPDCVAKVTEFLAEHRNVTLTISAA RLYYYWGKDWRRALCRL	118
A3F <i>Chlorocebus sabaeus</i> (AK29-AK150)	GRRYTWLCYEVKIKKDP SKLPWDTGVFPQGQVRPKFQSNRRQKVYFQPYHAEMYFLSWFCGNQLPAYKH FQITWVFSWNP CPDCVAKVTEFLAEHRNVTLTISAA RLYYYWGKDWRRALCRL	119
A3D <i>Erythrocebus patas</i> (AK29-AK150)	GRRYTWLCYEVKIKKDP SKLPWDTGVFQGGQVRPKFQSNRRYEVYFQPYHAEMYFLSWFCGNQLPAYKH FQITWVFSWNP CPDCVAKVTEFLAEHPNVTLTISAA RLYYYWGKDWRRALCRL	120
A3D <i>Macaca fascicularis</i> (AK29-AK159)	GRSYTWLCYEVKIRKDP SKLPWDTGVFRGQVRPKLQSNRRYELSNWECRKRVYFQPYHAEMYFLSWFCGNQLPANKRFQITWFASWNP CPDCVAKVTEFLAEHPNVTLTISVARLYYYRGKDWRRALRRL	121
A3F <i>Macaca fascicularis</i> (AK29-AK138)	GRSYTWLCYEVKIRKDP SKLPWDTGVFRGQVYFQPYHAEMYFLSWFCGNQLPANKRFQITWFASWNP CPDCVAKVTEFLAEHPNVTLTISVARLYYYRGKDWRRALRRL	122
A3D <i>Macaca nemestrina</i> (AK29-AK138)	GRSYTWLCYEVKIRKDP SKLPWDTGVFRDQVYFQPYHAEMCFLSWFCGNQLPANKRFQITWVFSWNP CPDCVTKVTEFLAEHPNVTLTISVARLYYYRGKDWRRALRRL	123
A3D <i>Macaca leonina</i> (AK29-AK138)	GRSYTWLCYEVKIRKDP SKLPWDTGVFRGQVYFQPYHAEMCFLSWFCGNQLPANKRFQITWVFSWNP CPDCVAKVTEFLAEHPNVTLTISVARLYYYRGKDWRRALRRL	124
A3D <i>Macaca mulatta</i> (AK29-AK138)	GRSYTWLCYEVKIRKDP SKLPWDTGVFRGQVYFQPYHAEMCFLSWFCGNQLPAYKRFQITWVFSWNP CPDCVAKVTEFLAEHPNVTLTISVARLYYYRGKDWRRALCRL	125

A3D Gorilla (AK29-AK150)	GRSYTWLCYEVKIRRGSSNLLWNTGVFRGPVPPKLQSNHRQEVYFQFENH AEMCFLSWFCGNRLPANRRFQITWVFSWNPCLPCVVKVTKFLAEHPNVT TISAARLYYYRDREWRRVLRL	126
A3D Pan paniscus (AK29-AK150)	GRSYTWLCYEVKIRGCSNLIWDTGVFRGPVLPKLQSNHRQEVYFQFENH AEMCFFSWFCGNRLPANRRFQITWVFSWNPCLPCVVKVTKFLAEHPNVT TISAARLYYYQDREWRRVLRL	127
A3D Pan troglodytes (AK29-AK150)	GRSYTWLCYEVKIRGCSNLIWDTGVFRGPVLPKLQSNHRQEVYFQFENH AEMCFFSWFCGNRLPANRRFQITWVFSWNPCLPCVVKVTKFLAEHPNVT TISAARLYYYQDREWRRVLRL	128
A3D Homo sapiens (AK29-AK150)	GRSYTWLCYEVKIRGRSNLLWDTGVFRGPVLPKRQSNHRQEVYFRFENH AEMCFLSWFCGNRLPANRRFQITWVFSWNPCLPCVVKVTKFLAEHPNVT TISAARLYYYRDRDWRVLLRL	129
A3D Nomascus leucogenys (AK29-AK150)	GRSYTWLCYEVKIRKDPKSLPDKGVFRGQVLPKFQSNHRQEVYFQLENH AEMCFLSWFCGNQLPANRRFQITWVFSWNPCLPCVAKVTEFLAEHPNVT TISAARLYYYRGRDWRRALRL	130
A3C Saimiri boliviensis (AK29-AK138)	GKKYTWLCYEVKIKKDTSKLPWNTGVFRGQVNFNPEHHAEMYFLSWFRGK LLPACKRSQITWVFSWNPCLYCVAKVAEFLAEHPNVTTLTVSTARLYCYWK KDWRRALRKL	131
A3F Saimiri boliviensis (AK29-AK138)	GKKYTWLCYEVKIKKDTSKLPWNTGVFRGQVNFNPEHHAEMYFLSWFRGK LLPACKRSQITWVFSWNPCLYCVAKVAEFLAEHPNVTTLTVSTARLYCYWK KDWRRALRKL	132
A3F Piliocolobus tephrosceles (AK36-AK145)	GRRYTWLCYEVKIMKDHKSLPWTGVFRGQVYFEPQNHAEMLSWFCGN QLPAYECCQITWVFSWTPCPDCVAKVTEFLAEHPNVTTLTISAARLYYYRG RDWRRALRL	133
A3F Colobus angolensis palliatus (AK29-AK138)	GRRYTWLCYEVKISKDPSKLPWDTGIFRGQVYFEPQYHAEMCFLSWYCGN QLPAYKCFQITWVFSWTPCPDCVGKVAEFLAEHPNVTTLTISAARLYYYWE TDYRRALCRL	134
A3F Pongo abelii (AK30-AK150)	RNYTWLCYEVKIRKDPKSLAWDTGVFRGQVLPKLQSNHRREVYFEPQYHA EMCFLSWFCGNQLSAYERFQITWVFSWTPCPDCVAMLAFLAEHPNVTLT VSAARLYYYWERDYRGALRL	135

Пример 2. Дальнейшая оценка конъюгированных с ингибитором редакторов оснований

[0166] В данном примере описан разработанный эффективный способ демонстрации того, что молекула АРОВЕС редакторов оснований (BE) непосредственно вызывала мутации в нецелевых сайтах одноцепочечной ДНК (НЦоц) независимым от енРНК образом. Тестируя ряд белков АРОВЕС с двумя доменами цитидиндезаминазы (CDA), было обнаружено, что каталитически неактивные домены CDA некоторых АРОВЕС с двумя доменами действуют как ингибиторы цитидиндезаминазы (CDI). Пользуясь преимуществом данного открытия и идеей разделенной протеазы TEV, был разработан индуцируемый редактор оснований (iBE) посредством направляемого енРНК расщепления CDI, в котором nSpCas9-BE и CDI связаны с помощью сайта расщепления TEV. В енРНК-независимых сайтах НЦоц, iBE1 оставался неактивным по причине ковалентно связанного CDI. Тогда как, в целевых сайтах, iBE1 активировался направляемым енРНК расщеплением CDI посредством TEV, приводя к эффективному редактированию оснований. Применяя никазу SpCas9 с «повышенной специфичностью», iBE2 дополнительно усовершенствовали, чтобы снизить частоту

непредусмотренных мутаций НЦен. Благодаря минимальным побочным действиям и ненарушенной эффективности целевого редактирования, специфичность редактирования iBE была значительно выше, чем таковая у ранее описанных BE. Таким образом, система iBE, описанная в данном примере, обеспечивает новый уровень регуляции специфичности современной системы редактирования оснований и предохраняет от нецелевых мутаций при ее применении.

МЕТОДЫ

Культура клеток и трансфекция

[0167] Клетки HEK293FT из ATCC поддерживали в DMEM (10566, Gibco/Thermo Fisher Scientific) + 10% ФБР (16000-044, Gibco/Thermo Fisher Scientific) и периодически проверяли, чтобы исключить заражение микоплазмой.

[0168] Для редактирования оснований в геномной ДНК, клетки HEK293FT высевали в 24-луночный планшет при плотности $1,1 \times 10^5$ на лунку и трансфицировали 250 мкл бессывороточной Opti-MEM, которая содержала 5,35 мкл липофектамина LTX (Life, Invitrogen), 2,14 мкл липофектамина плюс (Life, Invitrogen), 1 мкг вектора экспрессии pCMV-BE3 (или hA3B-BE3, hA3BCDA2-nSpCas9-BE, hA3D-BE3, hA3DCDA2-nSpCas9-BE, hA3F-BE3, hA3FCDA2-nSpCas9-BE, hA3G-BE3, hA3GCDA2-nSpCas9-BE, mA3-BE3, mA3CDA1-nSpCas9-BE, mA3CDA2-mA3CDA1-nSpCas9-BE, hA3FCDA1-mA3CDA1-nSpCas9-BE, hA3BCDA1-mA3CDA1-nSpCas9-BE, mA3CDA2-rA1-nSpCas9-BE, hA3FCDA1-rA1-nSpCas9-BE, hA3BCDA1-rA1-nSpCas9-BE, hA3A-BE3, mA3CDA2-hA3A-nSpCas9-BE, hA3FCDA1-hA3A-nSpCas9-BE, hA3BCDA1-hA3A-nSpCas9-BE, mA3CDA2F1-mA3CDA1-nSpCas9-BE, mA3CDA2F2-mA3CDA1-nSpCas9-BE, mA3CDA2F3-mA3CDA1-nSpCas9-BE, mA3CDAI-T2A-mA3CDA1-nSpCas9-BE, EGFP-mA3CDA1-nSpCas9-BE, EGFP-T2A-mA3CDA1-nSpCas9-BE, mA3CDAI-T2A-rA1-nSpCas9-BE, EGFP-rA1-nSpCas9-BE, EGFP-T2A-rA1-nSpCas9-BE, mA3CDAI-T2A-hA3A-nSpCas9-BE, EGFP-hA3A-nSpCas9-BE, EGFP-T2A-hA3A-nSpCas9-BE, pCMV-dSpCas9, iBE1, iBE2, mA3CDAI-TS-mA3CDA1-nSpCas9HF1-BE-NTEV или mA3CDAI-TS-mA3CDA1-nHypaSpCas9-BE-NTEV), 0,64 мкг вектора экспрессии енPHK с 0,5 мкг вектора экспрессии Sa-sg-SaD10A или без него. Через 24 ч в среду добавляли пурамицин (ant-pr-1, InvivoGen) при конечной концентрации 4 мкг/мл. Через еще 48 ч геномную

ДНК выделяли из клеток с помощью раствора для выделения ДНК QuickExtract™ (QE09050, Epicentre) для последующего анализа секвенированием.

Получение библиотек ДНК и секвенирование

[0169] Целевые геномные последовательности амплифицировали путем ПЦР с помощью высокоточной ДНК-полимеразы PrimeSTAR HS (Clontech) с наборами праймеров, фланкирующих исследуемые целевые сайты енРНК. Индексированные библиотеки ДНК получали, применяя набор для подготовки проб TruSeq ChIP (Illumina), с незначительными модификациями. Вкратце, продукты ПЦР, амплифицированные с участков геномной ДНК, фрагментировали с помощью Covaris S220. Фрагментированные ДНК затем амплифицировали с помощью ПЦР, применяя набор для подготовки проб TruSeq ChIP (Illumina). После количественного анализа с помощью набора Qubit High-Sensitivity DNA (Invitrogen), продукты ПЦР с различными метками объединяли для глубокого секвенирования с применением Illumina HiSeq X10 (2×150) или NextSeq 500 (2×150) в CAS-MPG Partner Institute for Computational Biology Omics Core, Шанхай, Китай. Качество исходных прочтений оценивали с помощью FastQC. Для секвенирования спаренных концов использовали только прочтения R1. Последовательности адаптеров и прочитанные последовательности на обоих концах с баллом качества Phred менее 30 отбрасывали. Укороченные прочтения затем картировали с помощью алгоритма BWA-MEM (BWA версии 0.7.17) на целевые последовательности. После сведения с помощью SAMtools (версии 1.9) далее рассчитывали замены оснований.

Расчет замен оснований

[0170] Замены оснований выбирали в каждом положении исследованных целевых сайтов енРНК, которые картировали с помощью по меньшей мере 1000 независимых прочтений, и очевидные замены оснований наблюдали лишь в целевых сайтах редактирования оснований. Частоты замены оснований рассчитывали путем деления прочтений замен оснований на общее число прочтений. Для каждой енРНК, отношение замен основания С на Т к indel рассчитывали путем деления суммы частот замены основания С на Т во всех сайтах редактирования на частоту indel на участке 50 п.о. вокруг целевого сайта енРНК (от восьми нуклеотидов против хода транскрипции от целевого сайта до 19 нуклеотидов по ходу транскрипции от сайтов РАМ).

Результаты

[0171] Были разработаны редакторы оснований цитозина или аденина (CBE/BE или ABE), в которых нативные цитидиндезаминазы или созданные *in vitro* аденозиндезаминазы слиты с CRISPR-Cas9, чтобы вызвать направленные превращения С в Т или аденина в гуанин (А в G) с высокими эффективностями. Так как в BE применяется каталитически неактивный белок Cas9 (dCas9) или никаза Cas9 (nCas9), чтобы направить их связывание с геномной ДНК, ожидается, что будут индуцироваться непредусмотренные замены оснований в сайтах НЦен, которые частично комплементарны енРНК. В данных условиях, применение высокоточной Cas9 в BE может уменьшить количество таких мутаций НЦен. Тем временем, поскольку свободные АРОВЕС могут вызывать неожиданные мутации С на Т в участках одноцепочечной ДНК (оцДНК), молекула АРОВЕС в BE может непосредственно запускать неожиданные мутации в сайтах НЦоц. Другими словами, нецелевые мутации, вызванные BE, могут также возникать в сайтах НЦоц независимо от направления посредством енРНК; тем не менее, мутации НЦоц не были выявлены вследствие отсутствия количественного и воспроизводимого способа детектирования.

[0172] В данном примере описан эффективный способ количественной оценки вызванных BE мутаций НЦоц путем коэкспрессии ортологов Cas9 *S. aureus* и *S. pyogenes* (co-expressing *S. aureus* and *S. pyogenes* Cas9 orthologs, CESSCO). При CESSCO экспрессия пар nSaCas9/Sa-енРНК приводила к образованию одноцепочечных разрывов (ОЦР) ДНК в определенных геномных локусах и приводила к образованию участка геномной оцДНК программируемым способом. В то же время коэкспрессированный BE3 в отсутствие енРНК (енРНК далее означает Sp-енРНК) использовали для исследования того, могут ли енРНК-независимые замены оснований С на Т быть вызваны BE3 отдельно на участках оцДНК, созданных вокруг введенного nSaCas9/Sa-енРНК ОЦР. После глубокого секвенирования участков генома, на которые нацелена nSaCas9/Sa-енРНК, было явно показано, что мутации С на Т в сайтах НЦоц были индуцированы содержащим АРОВЕС1 крысы (rA1) BE3, но не dSpCas9 в отсутствие енРНК, подтверждая, что мутации НЦоц вызываются молекулой АРОВЕС из BE независимым от енРНК образом.

[0173] В данном примере затем стремились уменьшить количество мутаций НЦоц путем применения представителей семейства АРОВЕС, подходящих для высокоспецифичного

конструирования ВЕ. Большинство широко применяемых ВЕ были сконструированы с однодоменными АРОВЕС, такими как гА1 в ВЕ3, но не с двухдоменными АРОВЕС. Обычно, в АРОВЕС с двумя доменами CDA один из них каталитически активен, тогда как другой каталитически неактивен и играет регуляторную роль в активности дезаминирования цитидина и, следовательно, может быть пригоден для конструирования высокоспецифичных ВЕ с пониженным действием НЦоц. Для того чтобы опробовать данную возможность, мы сконструировали и сравнили эффективности редактирования С на Т десяти спаренных ВЕ, которые содержат либо один каталитически активный домен CDA, либо два домена CDA из пяти АРОВЕС с двойным доменом (фиг. 20а), т.е., АРОВЕС3В человека (hA3B), АРОВЕС3D человека (hA3D), АРОВЕС3F человека (hA3F), АРОВЕС3G человека (hA3G) и АРОВЕС3 мыши (mA3).

[0174] На фиг. 20b,c показано, что ВЕ, сконструированные с некоторыми АРОВЕС (hA3B, hA3F и mA3), содержащими два домена CDA, вызывали значительно более низкие эффективности редактирования, чем их спаренные ВЕ, содержащие только активный домен CDA. Данный результат показывает, что каталитически неактивные домены CDA из данных АРОВЕС с двойным доменом, т.е., hA3B, hA3F и mA3, оказывают ингибиторное действие на соответствующие активные домены CDA.

[0175] Для того чтобы проверить, является ли ингибиторная функция универсальной, ковалентно соединяли каталитически неактивные домены CDA из mA3, hA3F или hA3B отдельно с N-концом mA3CDA1-nSpCas9-BE (фиг. 20d) и двумя другими широко применяемыми ВЕ, т.е., ВЕ3 и hA3A-BE3. У всех данных каталитически неактивных доменов CDA выявили широкий спектр ингибиторного действия на все исследованные ВЕ, и среди них у CDA2 из mA3 (mA3CDA2) выявили наиболее сильное ингибиторное действие (фиг. 20e,f). Анализ путем подробного картирования дополнительно выявил, что остатки 282 - 355 mA3CDA2 оказывали ингибиторное действие, сходное с таковым у полноразмерного mA3CDA2. В совокупности данные результаты показали, что каталитически неактивные домены некоторых АРОВЕС с двойным доменом в действительности проявляли универсальное ингибиторное действие на активность цитидиндезаминазы, и, следовательно, их назвали ингибиторами цитидиндезаминазы (CDI).

[0176] Далее пытались проверить, может ли отщепление mA3CDI (mA3CDA2) от ковалентно связанных с ним BE восстановить их способность редактирования оснований. Использовали саморасщепляющиеся пептиды (T2A) для соединения mA3CDI и mA3CDA1-nSpCas9-BE для исследования. После саморазделения mA3CDI эффективность редактирования mA3CDI-T2A-mA3CDA1-nSpCas9-BE восстанавливалась до уровней, сходных с таковыми для EGFP-mA3CDA1-nSpCas9-BE или EGFP-T2A-mA3CDA1-nSpCas9-BE, в ~10 раз более высоких, чем для BE, слитого с неотщепляемым mA3CDI. Самоотщепление mA3CDI от BE3 и hA3A-BE3 также повышает их эффективности редактирования, хоть и до отличных степеней.

[0177] Полученные результаты служили важным доказательством правильности концепции для разработки системы iBE для точного редактирования оснований с низкой частотой мутаций НЦоц. iBE1 сконструировали, используя сайт расщепления протеазой TEV (TS) для соединения трех важных молекул, т.е., mA3CDI, mA3CDA1-nSpCas9-BE и N-концевой половины протеазы TEV (NTEV) (фиг. 21a). В теории, благодаря ковалентной связи CDI, iBE1 остается неактивным, когда он связывается с сайтами НЦоц посредством своей молекулы APOBEC (фиг. 21a). Стоит отметить, что NTEV сама по себе неактивна, но образует функциональную протеазу TEV только когда привлечена С-концевая половина (CTEV). Таким образом, iBE1 может направляться своей составляющей CRISPR-Cas для осуществления эффективного редактирования оснований в целевых сайтах, где CDI расщепляется после вызванной енРНК сборки функциональной протеазы TEV (фиг. 21d).

[0178] После экспрессии в клетках iBE1 оставался неактивным в енРНК-независимых участках НЦоц, как и ожидалось (фиг. 21b), и вызывал гораздо более низкий (~20%) уровень мутаций С на Т по сравнению с BE3 (фиг. 21c). В целевых сайтах слитый с РНК-связывающим белком (MCP) CTEV может быть привлечен MS2-слитой енРНК (фиг. 21d), что приводит к удалению mA3CDI из iBE1 и, следовательно, позволяет эффективное редактирование оснований. Сравнение целевой эффективности редактирования, вызванного BE3 и iBE1 во множестве локусов в геноме (фиг. 21e), продемонстрировало, что iBE1 вызывал целевое редактирование оснований на уровнях, сходных с таковыми для BE3 (фиг. 21f, ~80% от BE3). В общей сложности, в данном примере показано, что была разработана система iBE, которая катализирует

эффективное редактирование оснований в целевых сайтах, при этом подавляя образование мутаций НЦоц, посредством манипулирования над CDI.

[0179] Так как известно, что Cas9 вызывает непредусмотренное редактирование в сайтах НЦен с частичной комплементарностью последовательностей с енРНК, нашей целью также было дополнительное уменьшение частоты мутаций НЦен путем замены немодифицированной nSpCas9 в iBE1 на ее сконструированные варианты, которые обладают улучшенной специфичностью нацеливания (фиг. 22a). Исследовали три сконструированных варианта nSpCas9, т.е., neSpCas9, nSpCas9HF1 и nHypaSpCas9, и обнаружили, что применение любого из данных белков Cas9 с улучшенной специфичностью нацеливания сильно снижало частоту мутаций НЦен (фиг. 22b,c). Тем временем, применение neSpCas9 не снижало эффективности целевого редактирования, тогда как применение других двух снижало эффективности целевого редактирования (фиг. 22d,e). В данном варианте, приступили к замене nSpCas9 на neSpCas9 для конструирования iBE2.

[0180] Как и для ранее разработанных BE, эффективность редактирования BE3 ограничена при некоторых условиях, и позже были разработаны дополнительные BE с улучшенными эффективностями редактирования, например, AncBE4max или hA3A-BE3. hA3A-BE3 представляет собой высоко активный BE в различных контекстах и, следовательно, сравнивали эффективность iBE2 с таковой у hA3A-BE3 с точки зрения эффективности и специфичности редактирования (фиг. 23a). Хотя средняя частота целевого редактирования iBE2 составляла ~50% от таковой для hA3A-BE3 (фиг. 23a,c), мутации С на Т, вызванные iBE2 в сайтах НЦоц и НЦен, были близки к фоновому уровню, тогда как hA3A-BE3 вызывал значительную частоту мутаций в данных нецелевых сайтах (фиг. 23a,b). В совокупности, средняя специфичность редактирования iBE2 была в ~40 раз выше, чем таковая для hA3A-BE3 (фиг. 23d).

[0181] В данном примере был впервые разработан эффективный способ (CESSCO) количественной оценки енРНК-независимых мутаций НЦоц и было подтверждено, что BE с обычным остовом APOBEC-nCas9 в действительности вызывал мутации НЦоц независимым от енРНК образом (фиг. 21a, 21b, 23a, 23b). В соответствии с нашими открытиями, недавние исследования полногеномного секвенирования также показали, что BE3 вызывал значительную частоту нецелевых мутаций у мышей и растения риса, предположительно также независимым от енРНК образом. Важно отметить, что было

использовано преимущество открытых авторами CDI для разработки iBE, которые остаются неактивными в сайтах HЦоц благодаря ковалентной связи CDI, но могут быть активированы опосредованным енРНК расщеплением CDI в целевых сайтах (фиг. 21a,d). iBE вызывали значимо более низкий уровень непредусмотренных мутаций в енРНК-независимых участках оцДНК, при этом эффективно осуществляя целевое редактирование (фиг. 21b,c,e,f).

[0182] Путем замены nSpCas9 на enSpCas9 с улучшенной специфичностью, был разработан высокоспецифичный iBE2, чтобы дополнительно снизить непредусмотренное редактирование в сайтах HЦен (фиг. 22 и 23e). Система iBE совместима с BE, содержащими различные молекулы Cas, и со сконструированными BE с улучшенной эффективностью, и не изменяет свойств сконструированных BE, таких как окно редактирования. Кроме того, поскольку существует множество представителей семейства APOBEC, другие CDI могут быть обнаружены в будущем, которые дополнительно обогатят репертуар систем, конъюгированных с CDI iBE. Так как точность и эффективность редактирования важны для редакторов оснований, особенно для их применения в терапии, система iBE, разработанная здесь, обеспечит новый уровень регуляции специфичности современной системы редактирования оснований и закрепит их применение против нецелевых мутаций.

Пример 3. Тестирование различных конфигурации индуцированных и разделенных редакторов оснований

[0183] В данном примере исследовали множество различных конфигураций молекул для реализации системы индуцированного и разделенного редактора оснований (isplitBE).

[0184] Рабочий процесс isplitBE показан на ФИГ. 24А, по сравнению с обычными BE, показанными на ФИГ. 24В. В проиллюстрированной системе isplitBE, конструкция nCas9-D10A упакована в носитель AAV. У обычного носителя AAV емкость составляет 4,7 т.п.н., и длина конструкции nCas9 составляет приблизительно 4,7 т.п.н. Другой носитель AAV может упаковывать нуклеиновые кислоты (общая длина приблизительно 4,4 т.п.н.) для кодирования: (а) слитого белка, который содержит MCP, UGI, APOBEC, сайт распознавания TEV (сайт TEV) и mA3CDA2; (b) слитого белка с TEVc и N22p; (c)

обособленного TEV_n, (d) вспомогательной енРНК (венРНК) с MS2 tag и (e) другой енРНК с меткой boxB.

[0185] В целевом сайте (Ц, левая нижняя ветвь), каждая из венРНК и енРНК связывается с двумя смежными сайтами на целевой ДНК, и содержащие MCP и N22p слитые белки связываются с меткой MS2 и меткой boxB венРНК и енРНК, соответственно. Благодаря пространственной близости TEV_c (в присутствии свободной TEV_n) и сайта TEV, TEV_c/TEV_n расщепляет сайт TEV, удаляя mA3CDA2 из APOBEC. Без присоединенного mA3CDA2 APOBEC может высокоэффективно осуществлять желательное редактирование.

[0186] В нецелевых сайтах, которые могут представлять собой сайт неспецифического связывания (НЦоц, нижняя средняя ветвь) или сайт, который связывает только одну из направляющих РНК, комплекс TEV_c/TEV_n не привлекается к содержащему сайт TEV слитому белку, и, следовательно, APOBEC не может быть активирована. В противоположность этому, в обычной системе BE (ФИГ. 24В) APOBEC уже активна и может вызывать редактирование С на Т во всех случаях, когда она привлечена к одноцепочечной последовательности нуклеотидов.

[0187] Получали и исследовали десять различных конфигураций (пары 1 - 10), представленных на ФИГ. 25. Например, на ФИГ. 26А показано, что пара 1 содержит две конструкции, первая из которых содержит gA1, слитый с nCas9-D10A (spD10A), наряду с UGI и NLS, и вторая из которых содержит енРНК, нацеленную на EMX1. Пара 2 похожа на пару 1, но gA1 заменен на hA3A. пара 3 также похожа, но в ней использовали взамен мутантный hA3A (Y130F).

[0188] В паре 4 белки gA1 и nCas9 были помещены на отличные конструкции. gA1 дополнительно соединяли с белком MCP, который распознает метку MS2 на вспомогательной енРНК. В паре 5 mA3CDA2 дополнительно соединяли с gA1 посредством сайта распознавания TEV (черный закрашенный прямоугольник). В паре 6 белок TEV дополнительно соединяли, посредством сайта саморасщепления 2A, со слитым белком gA1-mA3CDA2. Саморасщепление 2A высвобождает TEV из слитого белка.

[0189] Пара 7 отличается от пары 6 слиянием TEV с белком N22p, который распознает метку boxB на енРНК. В паре 8 белок TEV был разделен на TEV_n и TEV_c, отделенные

друг от друга сайтом саморасщепления 2A. В паре 9 только TEVc был слит с N22p, тогда как TEVn была свободна от каких-либо связывающих РНК-метку белков. В паре 10 вспомогательная енРНК была нацелена на GFP, но не на расположенный поблизости сайт.

[0190] Конструкции на ФИГ. 26А были разработаны для редактирования С на Т в целевом сайте EMX1-Ц, но нецелевое редактирование в сайтах Sa-SITE31-НЦоц и EMX1-НЦен также исследовали. Результаты исследования представлены на ФИГ. 26В. isplitBE-гA1 (пара 9) вызывала значительное редактирование в сайте Ц, но не вызывала редактирования в сайтах НЦоц и НЦен.

[0191] Аналогичным образом, все из данных конфигураций исследовали в сайтах FANCF-Ц, Sa-VEGFA-7-НЦоц и FANCF-НЦен (см. схематическую диаграмму на ФИГ. 27А). На ФИГ. 27В показано сравнение эффективностей редактирования для различных редакторов оснований в сайтах FANCF-Ц, Sa-VEGFA-7-НЦоц и FANCF-НЦен. Снова, isplitBE-гA1 (пара 9) вызывала значительное редактирование в сайте Ц, но не вызывала редактирования в сайтах НЦоц и НЦен.

[0192] Дополнительное тестирование проводили в сайтах V1B-Ц, Sa-SITE42-НЦоц и V1B-НЦен (см. схематическую диаграмму на ФИГ. 28А). Снова, как представлено на ФИГ. 28В, isplitBE-гA1 (пара 9) вызывала значительное редактирование в сайте Ц, но не вызывала редактирования в сайтах НЦоц и НЦен.

Пример 4. Регулировка параметров в системах isplitBE

[0193] Из 10 исследованных конфигураций, для пары 9 выявили наилучшую эффективность с точки зрения специфичности редактирования. В паре 9 используется две енРНК: вспомогательная енРНК (венРНК) и обычная енРНК. Применение двух енРНК дополнительно повышает специфичность, так как при этом требуется, чтобы оба целевых сайта находились в пространственной близости друг к другу.

[0194] В первом анализе данного примера оценивали оптимальное расстояние между двумя целевыми сайтами. На ФИГ. 29А представлена схематическая диаграмма, иллюстрирующая расстояние между венРНК и енРНК в сайтах DNTET1, EMX1 и FANCF. На ФИГ. 29В показаны частоты редактирования оснований, вызванного указанными енРНК и венРНК. В кратком представлении на ФИГ. 29С показано влияние

расстояния между венРНК и енРНК. На основании краткого представления, оптимальный диапазон расстояний для лучшей эффективности редактирования оснований составляет от -91 до -34 п.о. от РАМ венРНК до РАМ енРНК.

[0195] Во втором анализе исследовали влияние длины спейсера венРНК на эффективность и точность редактирования оснований. На ФИГ. 30А представлена схематическая диаграмма, иллюстрирующая котрансфекцию енРНК и венРНК с различными длинами спейсеров в сайтах DNEMX1, FANCF и V1A. На ФИГ. 30В показаны частоты редактирования оснований, вызванного указанными енРНК и венРНК в целевых для венРНК и енРНК сайтах. Статистический анализ на ФИГ. 30С показывает влияние длины спейсера венРНК. Как показано, применение венРНК со спейсером 10 н.т. сильно снижало эффективность редактирования в целевых сайтах венРНК, но сохраняло эффективность редактирования в целевых сайтах енРНК. Соответственно, длина спейсера 9 - 15 нт в последовательности вспомогательной енРНК может представлять собой хороший диапазон, чтобы гарантировать эффективное редактирование в целевом сайте енРНК, при этом минимизируя редактирование в целевом сайте венРНК.

Пример 5. Оценка по всему геному и транскриптому

[0196] Общую эффективность системы isplitBE сравнивали с таковой для обычного BE3. Результаты представлены на ФИГ. 31 (частоты редактирования, вызванного указанными редакторами оснований в различных целевых сайтах). Не наблюдали явного ухудшения эффективности, даже когда isplitBE проявлял сильно улучшенную специфичность.

[0197] В нормальных клетках наблюдали фоновый уровень мутаций С на Т, обусловленный их эндогенными активностями АРОВЕС3. Для получения более точного измерения нецелевых мутаций С на Т применяли линию клеток 293FT с нокаутом АРОВЕС3 (293FT-АЗКО). На ФИГ. 32А показаны уровни экспрессии мРНК в клетках 293FT дикого типа и клетках 293FT с нокаутом АРОВЕС3. На ФИГ. 32В представлена схематическая диаграмма, иллюстрирующая процедуры определения мутаций С на Т по всему геному, индуцированных редакторами оснований, и результаты тестирования представлены на ФИГ. 32С (эффективности целевого редактирования (слева) и количество мутаций С на Т по всему геному, вызванных Cas9, BE3, hA3A-BE3-Y130F

(Y130F) и isplitBE-rA1). Хотя для обоих BE3 и Y130F выявили достаточно высокий уровень побочного редактирования, уровень побочного редактирования isplitBE-rA1 был близок к фоновому (только Cas9).

[0198] В данном примере затем сравнивали мутации С на U по всему транскриптому, вызванные isplitBE-mA3, BE3 и hA3A-BE3-Y130F (Y130F). Количества мутаций С на Т(У) по всему транскриптому, вызванных Cas9, BE3, hA3A-BE3-Y130F (Y130F) и isplitBE-mA3, представлены на ФИГ. 33А. На ФИГ. 33В показаны частоты редактирования РНК С на U, вызванного Cas9, BE3, hA3A-BE3-Y130F (Y130F) и isplitBE-mA3. На ФИГ. 33С показано распределение редактирования С на U в РНК, вызванного BE3, повтор 1, и isplitBE-mA3, повтор 1. Снова isplitBE вызывал гораздо более низкий уровень редактирования С на U, чем BE3.

Пример 6. Нокауты PCSK9.

[0199] Пропропротеиновая конвертаза субтилизин-кексинового типа 9 (PCSK9) представляет собой фермент, кодируемый у людей геном PCSK9 на хромосоме 1. Это 9ый представитель семейства белков пропротеиновых конвертаз, которые активируют другие белки. PCSK9 неактивна, когда впервые синтезируется, так как фрагмент пептидных цепей блокирует ее активность; пропротеиновые конвертазы удаляют данный фрагмент, чтобы активировать фермент. Ген PCSK9 содержит один из 27 локусов, ассоциированных с повышенным риском ишемической болезни сердца.

[0200] PCSK9 повсеместно экспрессируется во множестве тканей и типов клеток. PCSK9 связывается с рецептором частиц липопротеинов низкой плотности (ЛПНП), который обычно транспортирует от 3000 до 6000 молекул жира (включая холестерин) на частицу, внутри внеклеточной жидкости. Рецептор ЛПНП (ЛПНП-Р), на мембранах клеток печени и других клеток, связывает и инициирует поглощение частиц ЛПНП из внеклеточной жидкости в клетки, таким образом уменьшая концентрации частиц ЛПНП. Если PCSK9 заблокирован, то большее количество ЛПНП-Р циркулирует и присутствует на поверхности клеток, чтобы удалить частицы ЛПНП из внеклеточной жидкости. Следовательно, блокирование PCSK9 может снизить концентрации частиц ЛПНП в крови.

[0201] В данном примере исследовали подход к инактивации PCSK9 путем введения стоп-кодонов посредством редактирования оснований с применением технологии в соответствии с настоящим изобретением. Используемые последовательности енРНК/венРНК показаны в таблице 3, и целевые сайты на PCSK9 показаны в таблице 4.

[0202] Количества стоп-кодонов, созданных путем редактирования оснований, измеряли для гена PCSK9 человека. На ФИГ. 34А представлена схематическая диаграмма, иллюстрирующая котрансфекцию енРНК и венРНК с isplitBE-mA3 и nCas9. Эффективности редактирования, вызванного isplitBE-mA3 в указанных сайтах, представлены на ФИГ. 34В - D. Данные результаты демонстрируют высокую эффективность и специфичность указанного способа.

Таблица 3. Обычные остовы енРНК и венРНК и целевые сайты в генах PCSK9

Название	Последовательность	SEQ ID NO:
hsgCAG2-MS2	GAGGUUGCCUGGCACCUACGGUUUGAGAGCUAGGCCAACAUAGGGAUCACCCAU GUCUGCAGGGCCUAGCAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAA GUGGCACCGAGUCGGUGC	136
hsgCAG3-MS2	GAGACCCACCUCUCGCAGUCGUUUGAGAGCUAGGCCAACAUAGGGAUCACCCAU GUCUGCAGGGCCUAGCAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAA GUGGCACCGAGUCGGUGC	137
hsgCAG4-MS2	GCCCCAUGUCGACUACAUCGGUUUGAGAGCUAGGCCAACAUAGGGAUCACCCAU GUCUGCAGGGCCUAGCAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAA GUGGCACCGAGUCGGUGC	138
hsgCAG7-MS2	AUGGUCACCGACUUCGAGAAGUUUGAGAGCUAGGCCAACAUAGGGAUCACCCAU GUCUGCAGGGCCUAGCAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAA GUGGCACCGAGUCGGUGC	139
hsgCAG8-MS2	ACCUUGGCUUUGUCCUCCGUUUGAGAGCUAGGCCAACAUAGGGAUCACCCAU GUCUGCAGGGCCUAGCAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAA GUGGCACCGAGUCGGUGC	140
hsgCAG9-MS2	GGCUUUGUUCUCCAGGCCGUUUGAGAGCUAGGCCAACAUAGGGAUCACCCAU GUCUGCAGGGCCUAGCAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAA GUGGCACCGAGUCGGUGC	141
hsgCAG10-MS2	GUGGUGCUGCGCCUUGGCGUUUGAGAGCUAGGCCAACAUAGGGAUCACCCAU GUCUGCAGGGCCUAGCAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAA GUGGCACCGAGUCGGUGC	142
hsgCAG11-MS2	UGCUGCGCCCCUGGCGGGUUGUUGAGAGCUAGGCCAACAUAGGGAUCACCCAU GUCUGCAGGGCCUAGCAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAA GUGGCACCGAGUCGGUGC	143
hsgCAG12-MS2	ACCCACCUCUCCACCUUCCGUUUGAGAGCUAGGCCAACAUAGGGAUCACCCAU GUCUGCAGGGCCUAGCAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAA GUGGCACCGAGUCGGUGC	144
hsgCAG14-MS2	AGCGACUGCAGCACCUGCUUGUUGAGAGCUAGGCCAACAUAGGGAUCACCCAU GUCUGCAGGGCCUAGCAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAA GUGGCACCGAGUCGGUGC	145
hsgCAG15-MS2	AACGCUUUUGGGGGUGAGGGGUUGAGAGCUAGGCCAACAUAGGGAUCACCCAU GUCUGCAGGGCCUAGCAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAA GUGGCACCGAGUCGGUGC	146
hsgCAG16-MS2	CCACACAGCUCACCAGCUGGUUUGAGAGCUAGGCCAACAUAGGGAUCACCCAU GUCUGCAGGGCCUAGCAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAA GUGGCACCGAGUCGGUGC	147
hsgCAG17-MS2	CACUGGGAGGUGGAGGACCUUGUUGAGAGCUAGGCCAACAUAGGGAUCACCCAU GUCUGCAGGGCCUAGCAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAA GUGGCACCGAGUCGGUGC	148

hsgCAG18-MS2	CCCACAAGCCGCCUGUGCUGGUUUGAGAGCUAGGCCAACAUAGAGGAUCACCCAU GUCUGCAGGGCCUAGCAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAA GUGGCACCGAGUCGGUGC	149
hsgCAG19-MS2	AGGUCUGGAAUGCAAAGUCAGUUUGAGAGCUAGGCCAACAUAGAGGAUCACCCAU GUCUGCAGGGCCUAGCAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAA GUGGCACCGAGUCGGUGC	150
sgCAG2-boxB	CUCUCGCAGUCAGAGCGCACGUUUGAGAGCUAGGGCCCCUGAAGAAGGGCCCCUAG CAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGGGCCCCUGAAGAAGGGCCCCAA GUGGCACCGAGUCGGUGC	151
sgCAG3-boxB	CAGGCCCAGGCUGCCCCGCCGUUUGAGAGCUAGGGCCCCUGAAGAAGGGCCCCUAG CAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGGGCCCCUGAAGAAGGGCCCCAA GUGGCACCGAGUCGGUGC	152
sgCAG4-boxB	UCUUUGCCCCAGAGCAUCCCGGUUUGAGAGCUAGGGCCCCUGAAGAAGGGCCCCUAG CAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGGGCCCCUGAAGAAGGGCCCCAA GUGGCACCGAGUCGGUGC	153
sgCAG7-boxB	CACAGACAGGUAAGCACGGCGUUUGAGAGCUAGGGCCCCUGAAGAAGGGCCCCUAG CAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGGGCCCCUGAAGAAGGGCCCCAA GUGGCACCGAGUCGGUGC	154
sgCAG8-boxB	AAGCCAGCUGGUCCAGCCUGGUUUGAGAGCUAGGGCCCCUGAAGAAGGGCCCCUAG CAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGGGCCCCUGAAGAAGGGCCCCAA GUGGCACCGAGUCGGUGC	155
sgCAG9-boxB	GGUCCAGCCUGUGGGGCCACGUUUGAGAGCUAGGGCCCCUGAAGAAGGGCCCCUAG CAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGGGCCCCUGAAGAAGGGCCCCAA GUGGCACCGAGUCGGUGC	156
sgCAG10-boxB	CGCCUGCCAGCGCCUGGCGAGUUUGAGAGCUAGGGCCCCUGAAGAAGGGCCCCUAG CAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGGGCCCCUGAAGAAGGGCCCCAA GUGGCACCGAGUCGGUGC	157
sgCAG11-boxB	UGCCAGCGCCUGGCGAGGGCGUUUGAGAGCUAGGGCCCCUGAAGAAGGGCCCCUAG CAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGGGCCCCUGAAGAAGGGCCCCAA GUGGCACCGAGUCGGUGC	158
sgCAG12-boxB	AAGACCAGCCGUGACCCUGGUUUGAGAGCUAGGGCCCCUGAAGAAGGGCCCCUAG CAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGGGCCCCUGAAGAAGGGCCCCAA GUGGCACCGAGUCGGUGC	159
sgCAG14-boxB	AUCACAGGCGUGCCACGGUUUGAGAGCUAGGGCCCCUGAAGAAGGGCCCCUAG CAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGGGCCCCUGAAGAAGGGCCCCAA GUGGCACCGAGUCGGUGC	160
sgCAG15-boxB	CUACCCCAAGGCCAACUGCAGGUUUGAGAGCUAGGGCCCCUGAAGAAGGGCCCCUAG CAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGGGCCCCUGAAGAAGGGCCCCAA GUGGCACCGAGUCGGUGC	161
sgCAG16-boxB	CAACAGGGCCACGUCCUCACGUUUGAGAGCUAGGGCCCCUGAAGAAGGGCCCCUAG CAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGGGCCCCUGAAGAAGGGCCCCAA GUGGCACCGAGUCGGUGC	162
sgCAG17-boxB	AGGUCAGCCCAACAGUGCGGUUUGAGAGCUAGGGCCCCUGAAGAAGGGCCCCUAG CAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGGGCCCCUGAAGAAGGGCCCCAA GUGGCACCGAGUCGGUGC	163
sgCAG18-boxB	CCAACCAGUGCGUGGGCCACGUUUGAGAGCUAGGGCCCCUGAAGAAGGGCCCCUAG CAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGGGCCCCUGAAGAAGGGCCCCAA GUGGCACCGAGUCGGUGC	164
sgCAG19-boxB	CCCCUCAGGAGCAGGUAGAAGGUUUGAGAGCUAGGGCCCCUGAAGAAGGGCCCCUAG CAAGUUCAAAUAAGGCUAGUCCGUUAUCAACUUGGGCCCCUGAAGAAGGGCCCCAA GUGGCACCGAGUCGGUGC	165

Таблица 4. Целевые сайты в генах PCSK9

Название	Последовательность	SEQ ID NO:
hsgCAG2	GAGGTTGCCTGGCACCTACGTGG	166
hsgCAG3	GAGACCCACCTCTCGCAGTCAGA	167
hsgCAG4	GCCCCATGTCTGACTACATCGAGG	168
hsgCAG7	ATGGTCACCGACTTCGAGAATGT	169
hsgCAG8	ACCTTGGCTTTGTTCTCCAGG	170
hsgCAG9	GGCTTTGTTCTCCAGGCTGG	171
hsgCAG10	GTGGTGCTGCTGCCCTGGCGGG	172

hsgCAG11	TGCTGCTGCCCCCTGGCGGGTGGG	173
hsgCAG12	ACCCACCTCCTCACCTTTCCAGG	174
hsgCAG14	AGCGACTGCAGCACCTGCTTTGT	175
hsgCAG15	AACGCTTTTGGGGGTGAGGGTGT	176
hsgCAG16	CCACACAGCTCCACCAGCTGAGG	177
hsgCAG17	CACTGGGAGGTGGAGGACCTTGG	178
hsgCAG18	CCCACAAGCCGCTGTGCTGAGG	179
hsgCAG19	AGGTCTGGAATGCAAAGTCAAGG	180
sgCAG2	CTCTCGCAGTCAGAGCGCACTGC	181
sgCAG3	CAGGCCCAGGCTGCCCCGCCGGGG	182
sgCAG4	TCTTTGCCCAGAGCATCCCCTGG	183
sgCAG7	CACAGACAGGTAAGCACGGCCGT	184
sgCAG8	AAGCCAGCTGGTCCAGCCTGTGG	185
sgCAG9	GGTCCAGCCTGTGGGGCCACTGG	186
sgCAG10	CGCCTGCCAGCGCCTGGCGAGGG	187
sgCAG11	TGCCAGCGCCTGGCGAGGGCTGG	188
sgCAG12	AAGACCAGCCGGTGACCCTGGGG	189
sgCAG14	ATCACAGGCTGCTGCCCACGTGG	190
sgCAG15	CTACCCCAGGCCAACTGCAGCGT	191
sgCAG16	CAACAGGGCCACGTCTTCACAGG	192
sgCAG17	AGGTCAGCCCAACCAAGTGCCTGG	193
sgCAG18	CCAACCAAGTGCCTGGGGCCACAGG	194
sgCAG19	CCCCTCAGGAGCAGGTGAAGAGG	195

Пример 7. Применимость дизайна isplitBE для редакторов основания аденина

[0203] Данный пример подтверждает применимость дизайна индуцированного и разделенного редактора оснований (isplitBE) в другом типе редакторов оснований. Используемый ингибитор представлял собой mA3CDA2, и редактор представлял собой редактор основания аденина (ABE).

[0204] Схематическая диаграмма, иллюстрирующая котрансфекцию енРНК и ABE, слитого с mA3CDA2 (или не слитого в качестве контроля), представлена на ФИГ. 35А. Эффективности редактирования, вызванные указанными ABE в сайтах RNF2 и FANCF, представлены на ФИГ. 35В. При присоединении mA3CDA2 к ABE эффективность редактирования снижалась по сравнению с ABE отдельно. Когда mA3CDA2 отщеплялся с помощью 2А, эффективность редактирования ABE восстанавливалась, подтверждая правильность подхода isplitBE для ABE.

Пример 8. Улучшенное праймированное редактирование

[0205] Применение обычных редакторов оснований ограничено транзициями оснований, но не трансверсиями, вставками или делециями оснований. Недавно была предложена система праймированного редактирования, в которой используют

праймирующий редактор (РЕ) путем конъюгирования никазы Cas9 с обратной транскриптазой (ОТаза). Система РЕ может написать геномы с практически любыми предполагаемыми изменениями, включая все типы замен оснований, небольшие indel и их комбинации. Общая эффективность и специфичность систем РЕ, тем не менее, все еще ограничены.

[0206] В первом анализе, в данном примере исследовали новый дизайн направляющей РНК для праймированного редактирования (прнРНК). Обычно каждая направляющая РНК содержит остов. Широко используемая последовательность остова представляет собой

GUUUUAGAGCUAGAAAUAGCAAGUUA AAAUAAGGCUAGUCCGUUAUCAACUU
GAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO: 29). Другой пример представляет собой

GUUUUAGAGCUAGAAAUAGCAAGUUA AAAUAAGGCAUGUCCGUUAUCAACUU
GAAAAAGUGGCACCGAUUCGGUGC (SEQ ID NO: 30). Более общая консенсусная последовательность

представляет собой
GUUUNAGAGCUAX₁UAGCAAGUNAAAUAAGGCNNGUCCGUUAUCAACUUX₂
AAGUGGCACCGANUCGGUGC (SEQ ID NO: 31), где N представляет собой любое основание, и X1 и X2 обозначают любую последовательность нуклеотидов длиной 2 - 50 оснований.

[0207] Ожидается, что остов будет образовывать вторичную структуру (показанную на ФИГ. 36А, SEQ ID NO: 30) благодаря внутренним комплементарным последовательностям. Обычная енРНК, применяемая в редакторах оснований, имеет длину приблизительно 96 нт, включая спейсер, длина которого составляет приблизительно 20 нт, и связывается с целевым сайтом. В прнРНК в 3'-конец остова дополнительно добавлены матрица для обратной транскрипции и сайт связывания праймера. Неожиданно, в рамках настоящего изобретения обнаружили, что исходный остов недостаточно стабилен в контексте прнРНК.

[0208] Следовательно, получили новый остов, в котором образуется новое спаривание между положениями 48 (например, А в SEQ ID NO: 30) и 61 (например, G в SEQ ID NO: 30). В примерах, представленных на ФИГ. 36А и 36Е, новый остов вместо этого содержит G и C или C и G (SEQ ID NO: 36, 37). Данный и дополнительные типичные мутантные остовы представлены в таблице 5 ниже.

Таблица 5. Последовательности остовов направляющей РНК

Описание	Остов направляющей РНК	SEQ ID NO:
Исходный	GUUUNAGAGCUA-X1-UAGCAAGUUNAAAUAAGGCNNGUCCGUUAUC <u>A</u> AC UU-X2-AAGU <u>G</u> GCACCGANUCGGUGC	31
Новый 1	GUUUNAGAGCUA-X1-UAGCAAGUUNAAAUAAGGCNNGUCCGUUAUC <u>G</u> AC UU-X2-AAGU <u>C</u> GCACCGANUCGGUGC	32
Новый 2	GUUUNAGAGCUA-X1-UAGCAAGUUNAAAUAAGGCNNGUCCGUUAUC <u>C</u> AC UU-X2-AAGU <u>G</u> GCACCGANUCGGUGC	33
Новый 3	GUUUNAGAGCUA-X1-UAGCAAGUUNAAAUAAGGCNNGUCCGUUAUC <u>A</u> AC UU-X2-AAGU <u>U</u> GCACCGANUCGGUGC	34
Новый 4	GUUUNAGAGCUA-X1-UAGCAAGUUNAAAUAAGGCNNGUCCGUUAUC <u>U</u> AC UU-X2-AAGU <u>A</u> GCACCGANUCGGUGC	35
Исходный 2	GUUUUAGAGCUAGAAAUAGCAAGUUAUUAAUAAGGCAUGUCCGUUAUC <u>A</u> AC UUGAAAAAGU <u>G</u> GCACCGAUUCGGUGC	30
Новый 5	GUUUUAGAGCUAGAAAUAGCAAGUUAUUAAUAAGGCAUGUCCGUUAUC <u>G</u> AC UUGAAAAAGU <u>C</u> GCACCGAUUCGGUGC	36
Новый 6	GUUUUAGAGCUAGAAAUAGCAAGUUAUUAAUAAGGCAUGUCCGUUAUC <u>C</u> AC UUGAAAAAGU <u>G</u> GCACCGAUUCGGUGC	37
Новый 7	GUUUUAGAGCUAGAAAUAGCAAGUUAUUAAUAAGGCAUGUCCGUUAUC <u>A</u> AC UUGAAAAAGU <u>U</u> GCACCGAUUCGGUGC	38
Новый 8	GUUUUAGAGCUAGAAAUAGCAAGUUAUUAAUAAGGCAUGUCCGUUAUC <u>U</u> AC UUGAAAAAGU <u>A</u> GCACCGAUUCGGUGC	39
Исходный 3	GUUUUAGAGCUAGAAAUAGCAAGUUAUUAAUAAGGCUAGUCCGUUAUC <u>A</u> AC UUGAAAAAGU <u>G</u> GCACCGAGUCGGUGC	29
Новый 9	GUUUUAGAGCUAGAAAUAGCAAGUUAUUAAUAAGGCUAGUCCGUUAUC <u>G</u> AC UUGAAAAAGU <u>C</u> GCACCGAGUCGGUGC	40
Новый 10	GUUUUAGAGCUAGAAAUAGCAAGUUAUUAAUAAGGCUAGUCCGUUAUC <u>C</u> AC UUGAAAAAGU <u>G</u> GCACCGAGUCGGUGC	41
Новый 11	GUUUUAGAGCUAGAAAUAGCAAGUUAUUAAUAAGGCUAGUCCGUUAUC <u>A</u> AC UUGAAAAAGU <u>U</u> GCACCGAGUCGGUGC	42
Новый 12	GUUUUAGAGCUAGAAAUAGCAAGUUAUUAAUAAGGCUAGUCCGUUAUC <u>U</u> AC UUGAAAAAGU <u>A</u> GCACCGAGUCGGUGC	43

[0209] Получали конструкции для тестирования обычных прнРНК и вновь разработанных улучшенных прнРНК (упрнРНК), как показано на ФИГ. 36В и 36F для PE2, и результаты тестирования представлены на ФИГ. 36С-36D и 36G. Сравнение эффективностей праймированного редактирования, вызванного прнРНК и упрнРНК. Для упрнРНК, с сильно улучшенной стабильностью стебля, выявили гораздо более высокую эффективность редактирования, чем у прнРНК, без исключения.

[0210] Аналогичным образом, в соответствии со схематической диаграммой на ФИГ. 37А, котрансфекцию прнРНК, введение одноцепочечного разрыва в енРНК с помощью PE2-NG (SEQ ID NO: 132) или xPE2 (SEQ ID NO: 133) осуществляли, чтобы проверить эффективность редактирования делеции TGATG. Результаты представлены на ФИГ. 37В. PE2-NG содержит сконструированную Cas9, которая может распознавать

релаксированный NG PAM (см., например, Nishimasu и др., Science 361, 1259-62 (2018)). xPE2 содержит сконструированную Cas9, которая может распознавать релаксированные NG, GAA и GAT PAM (см., например, Hu и др., Nature 556, 57-63 (2018)). Последовательности PE2-NG (SEQ ID NO: 44), xPE2 (SEQ ID NO: 45), SpCas9-NG (SEQ ID NO: 46) и xSpCas9 (SEQ ID NO: 47) представлены ниже в таблице 6.

Таблица 6. Последовательности Cas и PE

Название	Последовательность	SEQ ID NO:
PE2-NG	MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVSDTKADLRILIYLAHAMIKFRGHFLIEGDLNPDNSDVKLFI QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIALSLGL TPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLK DNREKIEKILTFRIPIYVVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASASQSFIERMT NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKLIKDKDFLDNEENEDILEDIV LTLTFLFEDREMIERLKTIAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF LKSDGFANRNFQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVV DELVKVMGRHKPENIVIEARENQTTQKGQKNSRERMKRIEEGKELGSQILKEHPVENTQL QNEKLYLYLLQNGRDMYVDQELDINRLSDYDVAIVPQSFLKDDSIDNKVLTRSDKNRGKSD NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH VAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVINNYHHAHDAYLNAV VGTALIKKYPKLESEFVYGDYKVDVRKMIASEQEI GKATAYFFYSNIMNFFKTEITLAN GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESIRPKRNS DKLIARKKDWDPKKYGGFVSPVAYSVLVAKVEKGSKKLKSVKELLGITIMERSSEFKNP IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARFLQKGNELALPSKYVNFYLAS HYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFSKRVI LADANLDKVL SAYNKHDRDKPI REQAENI IHLFTLTNLGAPRAFKYFDTTIDRKVYRSTKEVLDTLIHQSI TGLYETRIDL SQ LGGDGTSGGSSGGSSGSETPGTSESATPESGGSSGGSSSTLNIEDEYRLHETSKEPDVSLGS TWLSDFPQAWAETGGMGLAVRQAPLI I PLKATSTPVS I KQYPMSEARLGI KPHIQRLLDQG ILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGI SGQLTWTRLPQGFKNSPTLFNEALHRDLAD FRIQHFDLILQYVDDLLAATSELDCCQGTALLQTLGNLGYRASAKKAQICQKQVKYLG LLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFI PGFAEMAAPLYPLTKPGTL FNWGPDQQAQYQEI KQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGWRRPVAYL SKKLDPVAAGWP PCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARM THYQALLDTRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPADH TWYTDGSSLLQEGQRKAGAAVTETETEWAKALPAGTSAQRAELIALTQALKMAEGKKLVY TDSRYAFATAHIGEIYRRRGWLTSEGKEIKNKDEI LALLKALFLPKRLSIIHCPGHQKGHS AEARGNRMAQAARKAAITETPDTSTLLIENSSP	44
xPE2	MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVSDTKADLRILIYLAHAMIKFRGHFLIEGDLNPDNSDVKLFI QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIALSLGL TPNFKSNFDLAEDTKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT EITKAPLSASMIKLYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQEDFYFPLK DNREKIEKILTFRIPIYVVGPLARGNSRFAWMTRKSEETITPWNFEKVVDKGASASQSFIERMT NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGDQKKAIVDLLFKTNRK VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKLIKDKDFLDNEENEDILEDIV LTLTFLFEDREMIERLKTIAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF LKSDGFANRNFQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVV DELVKVMGRHKPENIVIEARENQTTQKGQKNSRERMKRIEEGKELGSQILKEHPVENTQL QNEKLYLYLLQNGRDMYVDQELDINRLSDYDVAIVPQSFLKDDSIDNKVLTRSDKNRGKSD NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH VAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVINNYHHAHDAYLNAV VGTALIKKYPKLESEFVYGDYKVDVRKMIASEQEI GKATAYFFYSNIMNFFKTEITLAN GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNS	45

	DKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGSKKLKSVKELLGITIMERSSEFEKNP IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGVLQKGNELALPSKYVNFLYLAS HYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKVL SAYNKH RDKPI REQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQ LGGDGTSGGSSGGSSGSETPGTSESATPESGGSSGGSSSTLNIEDEYRLHETSKEPDVSLGS TWLSDFPQAWAETGGMGLAVRQAPLI I PLKATSTPVSIKQYPMSEQEARLGIKPHIQRLLDQG ILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLAD FRIQHFDLILLQYVDDLLLAATSELDCCQGGTRALLQTLGNLGYRASAKKAQICQKQVKYLG LLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTL FNWGPDQKAYQEI KQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPPWRPVAYL SKKLDPVAAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARM THYQALLDTRVQFGPVVALNPATLLPLPEEGLQHNCLDILA EAHGTRPDLTQPLPADH TWYTDGSSLLQEGQRKAGAAVTETETEWAKALPAGTSAQRAELIALTQALKMAEGKKLVY TDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSI IHCPGHQKGHS AEARGNRMADQAARKAAITETPDTSTLLIENSSP	
SpCas9-NG	MDKKYSIGLDIGTNSVGWAVITDEYKVP SKKFVVLGNTDRHSIKKNLIGALLFDSGETAEAT RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRL EESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVSDTKADLR LIYLA LAHMIKFRGHFLIEGDLNPDNSDVKLFI QLVQTYNQLFEE NPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGL TPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQI GDQYADFLAAKNLSDAI LLSDI LRVT EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI FFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI PHQIHLGELHAILRRQEDFY PFLK DNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWNFEVV DKGASAQSFIERMT NFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL LFKTNRK VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF LKSDGFANRNFQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA I KKGILQTVKV DELVKVMGRHKPENIVIE MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSIDNKVLRSDKNRGKSD NVPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGSEL DKA GFIKRQLVETRQITKH VAQILDSRMNTKYDENDKLI REVKVI TLKSKLVSDFRKDFQFYK VREINNYHHAHDAYLNAV VGTALIKKYPKLESEFVYGDYKVDVRKMIAKSEQEI GKATAKYFFYSNIMNFFKTEITLAN GEIRKRPLIETNGETGEI VWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESIRPKRNS DKLIARKKDWDPKKYGGFVSPTVAYSVLVAKVEKGSKKLKSVKELLGITIMERSSEFEKNP IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARFLQKGNELALPSKYVNFLYLAS HYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKVL SAYNKH RDKPI REQAENI IHLFTLTNLGAPAAFKYFDTTIDRKVYRSTKEVL DATLIHQ SITGLYETRIDLSQ LGGD	46
xSpCas9	MDKKYSIGLDIGTNSVGWAVITDEYKVP SKKFVVLGNTDRHSIKKNLIGALLFDSGETAEAT RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRL EESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVSDTKADLR LIYLA LAHMIKFRGHFLIEGDLNPDNSDVKLFI QLVQTYNQLFEE NPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGL TPNFKSNFDLAEDTKLQLSKDTYDDDLNLLAQI GDQYADFLAAKNLSDAI LLSDI LRVT EITKAPLSASMIKLYDEHHQDLTLLKALVRQQLPEKYKEI FFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGI I PHQIHLGELHAILRRQEDFY PFLK DNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWNFEKVVDKGASAQSFIERMT NFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGDQKKAIVDL LFKTNRK VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF LKSDGFANRNFQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA I KKGILQTVKV DELVKVMGRHKPENIVIE MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSIDNKVLRSDKNRGKSD NVPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGSEL DKA GFIKRQLVETRQITKH VAQILDSRMNTKYDENDKLI REVKVI TLKSKLVSDFRKDFQFYK VREINNYHHAHDAYLNAV VGTALIKKYPKLESEFVYGDYKVDVRKMIAKSEQEI GKATAKYFFYSNIMNFFKTEITLAN GEIRKRPLIETNGETGEI VWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNS DKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGSKKLKSVKELLGITIMERSSEFEKNP IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGVLQKGNELALPSKYVNFLYLAS HYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKVL SAYNKH RDKPI REQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQ LGGD	47

[0211] Для полноразмерного праймирующего редактора требуется конструкция (приблизительно 11 т.п.о.), которая гораздо больше, чем может вместить носитель AAV. Соответственно, была разработана и исследована разделенная система PE. Исходная система PE проиллюстрирована на левой панели на ФИГ. 38А, и вновь разработанная разделенная система PE проиллюстрирована на правой панели, в которой никаза и ОТаза упакованы в различные частицы AAV. ОТаза слита с РНК-связывающим белком МСР, и прнРНК содержит сайт связывания MS2. После поглощения клеткой, ОТаза может привлекаться прнРНК, посредством связывания MS2-МСР, и контактировать с никазой.

[0212] Пример системы котрансфекции показан на ФИГ. 38В, и результаты тестирования представлены на ФИГ. 38С, в сайте EMX1.

* * *

[0213] Настоящее изобретение не должно быть ограничено в объеме конкретными описанными вариантами реализации, которые предназначены для единичного иллюстрирования отдельных аспектов настоящего изобретения, и любые композиции или способы, которые функционально эквивалентны, входят в объем настоящего изобретения. Для специалистов в данной области должно быть очевидно, что можно сделать различные модификации и варианты способов и композиций согласно настоящему изобретению, не отклоняясь от сущности или объема настоящего изобретения. Таким образом, предполагается, что в объем настоящего изобретения входят модификации и варианты настоящего изобретения, при условии, что они попадают под объем прилагаемых пунктов формулы изобретения и их эквивалентов.

[0214] Все публикации и заявки на патент, упомянутые в настоящем описании, включены в данную заявку посредством ссылки в той же степени, как если бы каждая отдельная публикация или заявка на патент была конкретно и отдельно указана как включенная посредством ссылки.

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Слитый белок, содержащий:
первый фрагмент, содержащий дезаминазу азотистых оснований или ее каталитический домен,
второй фрагмент, содержащий ингибитор дезаминазы азотистых оснований, и
сайт расщепления протеазой между первым фрагментом и вторым фрагментом.
2. Слитый белок по п. 1, отличающийся тем, что дезаминаза азотистых оснований представляет собой аденозиндезаминазу.
3. Слитый белок по п. 2, отличающийся тем, что аденозиндезаминаза выбрана из группы, состоящей из тРНК-специфичной аденозиндезаминазы (TadA), тРНК-специфичной аденозиндезаминазы 1 (ADAT1), тРНК-специфичной аденозиндезаминазы 2 (ADAT2), тРНК-специфичной аденозиндезаминазы 3 (ADAT3), РНК-специфичной аденозиндезаминазы В1 (ADARB1), РНК-специфичной аденозиндезаминазы В2 (ADARB2), аденозинмонофосфат-дезаминазы 1 (AMPD1), аденозинмонофосфат-дезаминазы 2 (AMPD2), аденозинмонофосфат-дезаминазы 3 (AMPD3), аденозиндезаминазы (ADA), аденозиндезаминазы 2 (ADA2), подобного аденозиндезаминазе белка (ADAL), содержащего домен аденозиндезаминазы белка 1 (ADAD1), содержащего домен аденозиндезаминазы белка 2 (ADAD2), РНК-специфичной аденозиндезаминазы (ADAR) и РНК-специфичной аденозиндезаминазы В1 (ADARB1).
4. Слитый белок по п. 1, отличающийся тем, что дезаминаза азотистых оснований представляет собой цитидиндезаминазу.
5. Слитый белок по п. 4, отличающийся тем, что цитидиндезаминаза выбрана из группы, состоящей из АРОВЕС3В (А3В), АРОВЕС3С (А3С), АРОВЕС3Д (А3Д),

АРОВЕС3F (A3F), АРОВЕС3G (A3G), АРОВЕС3H (A3H), АРОВЕС1 (A1), АРОВЕС3 (A3), АРОВЕС2 (A2), АРОВЕС4 (A4) и AICDA (AID).

6. Слитый белок по п. 4, отличающийся тем, что цитидиндезаминаза представляет собой цитидиндезаминазу человека или мыши.

7. Слитый белок по п. 6, отличающийся тем, что каталитический домен представляет собой домен 1 цитидиндезаминазы A3 мыши (CDA1) или домен 2 цитидиндезаминазы A3B человека (CDA2).

8. Слитый белок по любому из пп. 1 - 7, отличающийся тем, что ингибитор дезаминазы азотистых оснований представляет собой ингибиторный домен дезаминазы азотистых оснований.

9. Слитый белок по п. 8, отличающийся тем, что ингибитор дезаминазы азотистых оснований представляет собой ингибиторный домен цитидиндезаминазы.

10. Слитый белок по п. 9, отличающийся тем, что ингибитор дезаминазы азотистых оснований содержит последовательность аминокислот, выбранную из SEQ ID NO: 1 - 2 и 48 - 135, или последовательность аминокислот, идентичную по меньшей мере на 85% любой из последовательностей аминокислот, выбранных из SEQ ID NO: 1 - 2 и 48 - 135.

11. Слитый белок по п. 9, отличающийся тем, что ингибитор дезаминазы азотистых оснований содержит последовательность аминокислот SEQ ID NO: 1, остатки аминокислот АК128 - АК223 из последовательности SEQ ID NO: 1 или последовательность аминокислот SEQ ID NO: 2.

12. Слитый белок по любому из пп. 1 - 11, отличающийся тем, что первый фрагмент дополнительно содержит белок, ассоциированный с короткими палиндромными повторами, регулярно расположенными группами (CRISPR) (Cas).

13. Слитый белок по п. 12, отличающийся тем, что белок Cas выбран из группы, состоящей из SpCas9, FnCas9, St1Cas9, St3Cas9, NmCas9, SaCas9, AsCpf1, LbCpf1, FnCpf1, VQR SpCas9, EQR SpCas9, VRER SpCas9, SpCas9-NG, xSpCas9, RHA FnCas9, KKH SaCas9, NmeCas9, StCas9, CjCas9, AsCpf1, FnCpf1, SsCpf1, PcCpf1, BpCpf1, CmtCpf1, LiCpf1, PmCpf1, Pb3310Cpf1, Pb4417Cpf1, BsCpf1, EeCpf1, BhCas12b, AkCas12b, EbCas12b, LsCas12b, RfCas13d, LwaCas13a, PspCas13b, PguCas13b и RanCas13b.

14. Слитый белок по любому из пп. 1 - 13, отличающийся тем, что сайт расщепления протеазой представляет собой сайт расщепления протеазой, выбранной из группы, состоящей из протеазы TuMV, протеазы PPV, протеазы PVY, протеазы ZIKV и протеазы WNV.

15. Слитый белок по любому из пп. 1 - 13, отличающийся тем, что сайт расщепления протеазой представляет собой сайт саморасщепления.

16. Слитый белок по любому из пп. 1 - 13, отличающийся тем, что сайт расщепления протеазой представляет собой сайт расщепления протеазой TEV.

17. Слитый белок по п. 16, дополнительно содержащий третий фрагмент, содержащий протеазу TEV или ее фрагмент.

18. Слитый белок по п. 17, отличающийся тем, что третий фрагмент содержит фрагмент протеазы TEV, который отдельно не способен расщеплять сайт расщепления протеазой TEV.

19. Слитый белок, содержащий:

первый фрагмент, содержащий дезаминазу азотистых оснований или ее каталитический домен, белок, ассоциированный с короткими палиндромными повторами, регулярно расположенными группами (CRISPR) (Cas), и первый фрагмент протеазы TEV,

второй фрагмент, содержащий ингибитор дезаминазы азотистых оснований, и сайт расщепления протеазой TEV между первым фрагментом и вторым фрагментом,

причем первый фрагмент протеазы TEV отдельно не способен расщеплять сайт расщепления протеазой TEV.

20. Слитый белок по п. 19, дополнительно содержащий ингибитор урацилгликозилазы (UGI).

21. Слитый белок по п. 19, отличающийся тем, что ингибитор дезаминазы азотистых оснований, сайт расщепления протеазой TEV, дезаминаза азотистых оснований или ее каталитический домен, белок Cas и первый фрагмент протеазы TEV расположены в указанном порядке от N-конца к С-концу.

22. Слитый белок по п. 19, отличающийся тем, что первый фрагмент протеазы TEV представляет собой N-концевой домен (SEQ ID NO: 3) или С-концевой домен (SEQ ID NO: 4) протеазы TEV.

23. Слитый белок по п. 22, отличающийся тем, что сайт расщепления протеазой TEV имеет последовательность аминокислот SEQ ID NO: 5.

24. Слитый белок по п. 19, отличающийся тем, что дезаминаза азотистых оснований выбрана из группы, состоящей из АРОВЕС3В (А3В), АРОВЕС3С (А3С), АРОВЕС3Д (А3Д), АРОВЕС3F (А3F), АРОВЕС3G (А3G), АРОВЕС3Н (А3Н), АРОВЕС1 (А1), АРОВЕС3 (А3), АРОВЕС2 (А2), АРОВЕС4 (А4) и АICDA (АID).

25. Слитый белок по п. 19, отличающийся тем, что ингибитор дезаминазы азотистых оснований содержит последовательность аминокислот, выбранную из SEQ ID NO: 1 - 2 и 48 - 135, или последовательность аминокислот, идентичную по меньшей мере на 85% любой из последовательностей аминокислот, выбранных из SEQ ID NO: 1 - 2 и 48 - 135.

26. Слитый белок по п. 19, отличающийся тем, что белок Cas выбран из группы, состоящей из SpCas9, FnCas9, St1Cas9, St3Cas9, NmCas9, SaCas9, AsCpf1, LbCpf1, FnCpf1, VQR SpCas9, EQR SpCas9, VRER SpCas9, SpCas9-NG, xSpCas9, RHA FnCas9, KKH SaCas9, NmeCas9, StCas9, CjCas9, AsCpf1, FnCpf1, SsCpf1, PcCpf1, BpCpf1, CmtCpf1, LiCpf1, PmCpf1, Pb3310Cpf1, Pb4417Cpf1, BsCpf1, EeCpf1, BhCas12b, AkCas12b, EbCas12b, LsCas12b, RfCas13d, LwaCas13a, PspCas13b, PguCas13b и RanCas13b.

27. Способ осуществления генетического редактирования в клетке в целевом сайте, включающий введение в клетку:

(а) слитого белка по любому из пп. 19 - 26,

(б) направляющей РНК, которая нацелена на целевой сайт, или крРНК, которая нацелена на целевой сайт, и тракрРНК, и дополнительно содержит последовательность метки, и

- (с) второго фрагмента протеазы TEV, соединенного с распознающим РНК пептидом, который способен связываться с последовательностью метки.
28. Способ по п. 27, отличающийся тем, что одну или более указанных молекул вводят в клетку с помощью полинуклеотида, кодирующего указанную молекулу.
29. Способ по п. 27, отличающийся тем, что первый фрагмент протеазы TEV и второй фрагмент протеазы TEV, когда они взаимодействуют, способны расщеплять сайт расщепления протеазой TEV.
30. Способ по п. 27, отличающийся тем, что второй фрагмент протеазы TEV соединен с распознающим РНК пептидом.
31. Способ по любому из пп. 27 - 30, отличающийся тем, что последовательность метки содержит последовательность MS2 (SEQ ID NO: 16).
32. Способ по п. 31, отличающийся тем, что распознающий РНК пептид содержит белок оболочки MS2 (MCP, SEQ ID NO: 22).
33. Способ по любому из пп. 27 - 30, отличающийся тем, что последовательность метки содержит последовательность PP7 (SEQ ID NO: 18), и распознающий РНК пептид содержит белок оболочки PP7 (PCP, SEQ ID NO: 23), или последовательность метки содержит последовательность boxB (SEQ ID NO: 20), и распознающий РНК пептид содержит белок оболочки boxB (N22p, SEQ ID NO: 24).
34. Способ по п. 27, отличающийся тем, что клетка находится внутри живого организма.

35. Способ по п. 34, отличающийся тем, что клетка представляет собой клетку человека внутри пациента - человека.
36. Набор или упаковка для осуществления редактирования генов, содержащая:
- (a) слитый белок по любому из пп. 19 - 26, и
 - (b) второй фрагмент протеазы TEV, соединенный с распознающим РНК пептидом, который способен связываться с последовательностью РНК.
37. Слитый белок, содержащий:
- первый фрагмент, содержащий первую дезаминазу азотистых оснований или ее каталитический домен, и
- второй фрагмент, содержащий ингибиторный домен второй дезаминазы азотистых оснований,
- причем первая дезаминаза азотистых оснований такая же или отличная от второй дезаминазы азотистых оснований.
38. Слитый белок по п. 37, отличающийся тем, что каждая из первой и второй дезаминазы азотистых оснований независимо выбрана из группы, состоящей из человеческой и мышинной АРОВЕС3В (A3B), АРОВЕС3С (A3C), АРОВЕС3D (A3D), АРОВЕС3F (A3F), АРОВЕС3G (A3G), АРОВЕС3Н (A3H), АРОВЕС1 (A1), АРОВЕС3 (A3), АРОВЕС2 (A2), АРОВЕС4 (A4) и AICDA (AID).
39. Слитый белок, содержащий первый фрагмент, содержащий:
- дезаминазу азотистых оснований или ее каталитический домен,
- ингибитор дезаминазы азотистых оснований,
- первый распознающий РНК пептид, и

сайт расщепления протеазой TEV между дезаминазой азотистых оснований или ее каталитическим доменом и ингибитором дезаминазы азотистых оснований.

40. Слитый белок по п. 39, дополнительно содержащий второй фрагмент, содержащий:

фрагмент протеазы TEV, который отдельно не способен расщеплять сайт расщепления протеазой TEV, и второй распознающий РНК пептид.

41. Слитый белок по п. 40, дополнительно содержащий сайт саморасщепления между первым фрагментом и вторым фрагментом.

42. Слитый белок по п. 41, дополнительно содержащий третий фрагмент, содержащий второй фрагмент протеазы TEV, причем первый фрагмент протеазы TEV способен расщеплять сайт расщепления протеазой TEV в присутствии второго фрагмента протеазы TEV.

43. Слитый белок по п. 42, дополнительно содержащий второй сайт саморасщепления между вторым фрагментом и третьим фрагментом, и после расщепления второго сайта саморасщепления, слитый белок высвобождает второй фрагмент протеазы TEV, который не соединен с каким-либо распознающим РНК пептидом.

44. Система двух направляющих РНК, содержащая:
нацеливающую единую направляющую РНК, содержащую первый спейсер, последовательность которого комплементарна целевой

последовательности нуклеиновой кислоты, ближайшей к первому сайту РАМ,
вспомогательную единую направляющую РНК, содержащую второй спейсер,
последовательность которого комплементарна второй
последовательности нуклеиновой кислоты, ближайшей ко второму сайту РАМ,
белок, ассоциированный с короткими палиндромными повторами, регулярно
расположенными группами (CRISPR) (Cas), и
деаминазу азотистых оснований,
причем второй сайт РАМ расположен на расстоянии от 34 до 91 основания от
первого сайта РАМ.

45. Система двух направляющих РНК по п. 44, отличающаяся тем, что длина второго спейсера составляет 7 - 23 основания.

46. Система двух направляющих РНК по п. 44, отличающаяся тем, что длина второго спейсера составляет 8 - 15 оснований.

47. Система двух направляющих РНК по п. 44, отличающаяся тем, что длина второго спейсера составляет 9 - 12 оснований.

48. Направляющая РНК, содержащая остов, который содержит, в направлении от 5' к 3',
первую структуру шпильки,
вторую структуру шпильки,
третью структуру шпильки, и
четвертую структуру шпильки,
причем третья шпилька содержит в себе пять спаренных оснований.

49. Направляющая РНК по п. 48, отличающаяся тем, что спаривания оснований включают таковое между положениями 45 и 55, в соответствии с положениями в SEQ ID NO: 31.

50. Направляющая РНК по п. 48 или 49, отличающаяся тем, что остов содержит последовательность, выбранную из группы, состоящей из SEQ ID NO: 32 - 43.

51. Направляющая РНК по п. 48, дополнительно содержащая один или более из спейсера для распознавания целевой последовательности, матрицы для обратной транскриптазы или сайта связывания праймера.

52. Способ осуществления генетического редактирования в клетке в целевом сайте, включающий введение в клетку:

первой вирусной частицы, заключающей в себе первую конструкцию, кодирующую белок, ассоциированный с короткими палиндромными повторами, регулярно расположенными группами (CRISPR) (Cas), и второй вирусной частицы, заключающей в себе вторую конструкцию, кодирующую обратную транскриптазу, соединенную с распознающим РНК пептидом.

53. Способ по п. 52, отличающийся тем, что вторая конструкция дополнительно кодирует направляющую РНК, содержащую сайт распознавания РНК, с которым связывается распознающий РНК пептид.

54. Способ по п. 52, отличающийся тем, что белок Cas выбран из группы, состоящей из SpCas9, FnCas9, St1Cas9, St3Cas9, NmCas9, SaCas9, AsCpf1, LbCpf1, FnCpf1, VQR SpCas9, EQR SpCas9, VRER SpCas9, SpCas9-NG, xSpCas9, RHA FnCas9, KKH SaCas9,

NmeCas9, StCas9, CjCas9, AsCpf1, FnCpf1, SsCpf1, PcCpf1, BpCpf1, CmtCpf1, LiCpf1, PmCpf1, Pb3310Cpf1, Pb4417Cpf1, BsCpf1, EeCpf1, BhCas12b, AkCas12b, EbCas12b, LsCas12b, RfCas13d, LwaCas13a, PspCas13b, PguCas13b и RanCas13b.

55. Способ по п. 52, отличающийся тем, что обратная транскриптаза выбрана из группы, состоящей из обратной транскриптазы вируса иммунодефицита человека (ВИЧ), обратной транскриптазы вируса мышинного лейкоза Молони (MMLV) и обратной транскриптазы вируса миелобластома птиц (AMV).

56. Пара вспомогательной направляющей РНК/направляющей РНК для редактирования последовательности нуклеиновой кислоты PCSK9 человека, причем направляющая РНК специфично нацеливается на первый сайт на нуклеиновой кислоте PCSK9, позволяя осуществлять редактирование оснований для превращения кодона, не являющегося стоп-кодоном, в стоп-кодон, и причем вспомогательная направляющая РНК специфично нацеливается на второй сайт на нуклеиновой кислоте PCSK9, который находится на расстоянии 20 - 100 оснований от первого сайта.

57. Пара вспомогательной направляющей РНК/направляющей РНК по п. 56, отличающаяся тем, что кодон, не являющийся стоп-кодоном, представляет собой CAG, CAA или CGA.

58. Пара вспомогательной направляющей РНК/направляющей РНК по п. 56 или 57, отличающаяся тем, что вспомогательная направляющая РНК специфично связывается с последовательностью, длина которой составляет 7 - 23 нуклеотида.

59. Пара вспомогательной направляющей РНК/направляющей РНК по п. 56 или 57, отличающаяся тем, что вспомогательная направляющая РНК специфично связывается с последовательностью, длина которой составляет 8 - 15 нуклеотидов.

60. Пара вспомогательной направляющей РНК/направляющей РНК по п. 58 или 59, отличающаяся тем, что вспомогательная направляющая РНК специфично связывается с последовательностью внутри первого сайта, и направляющая РНК специфично связывается с последовательностью внутри второго сайта.

61. Пара вспомогательной направляющей РНК/направляющей РНК по п. 60, отличающаяся тем, что второй сайт содержит любую последовательность из SEQ ID NO: 166 - 180.

62. Пара вспомогательной направляющей РНК/направляющей РНК по п. 60, отличающаяся тем, что второй сайт и первый сайт, соответственно, содержит последовательности:

SEQ ID NO: 166 и 181;

SEQ ID NO: 167 и 182;

SEQ ID NO: 168 и 183;

SEQ ID NO: 169 и 184;

SEQ ID NO: 170 и 185;

SEQ ID NO: 171 и 186;

SEQ ID NO: 172 и 187;

SEQ ID NO: 173 и 188;

SEQ ID NO: 174 и 189;

SEQ ID NO: 175 и 190;

SEQ ID NO: 176 и 191;

SEQ ID NO: 177 и 192;

SEQ ID NO: 178 и 193;

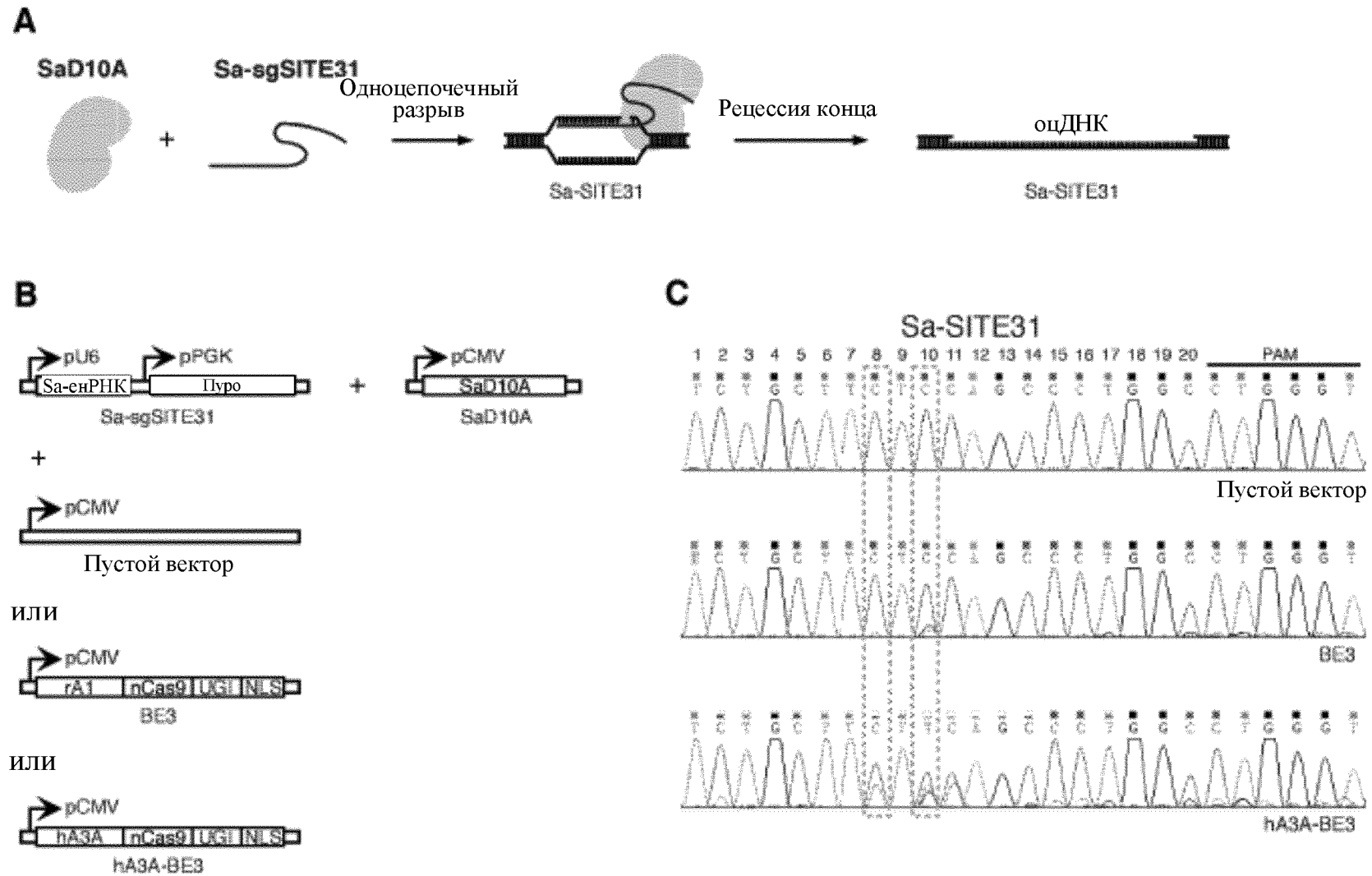
SEQ ID NO: 179 и 194; или

SEQ ID NO: 180 и 195.

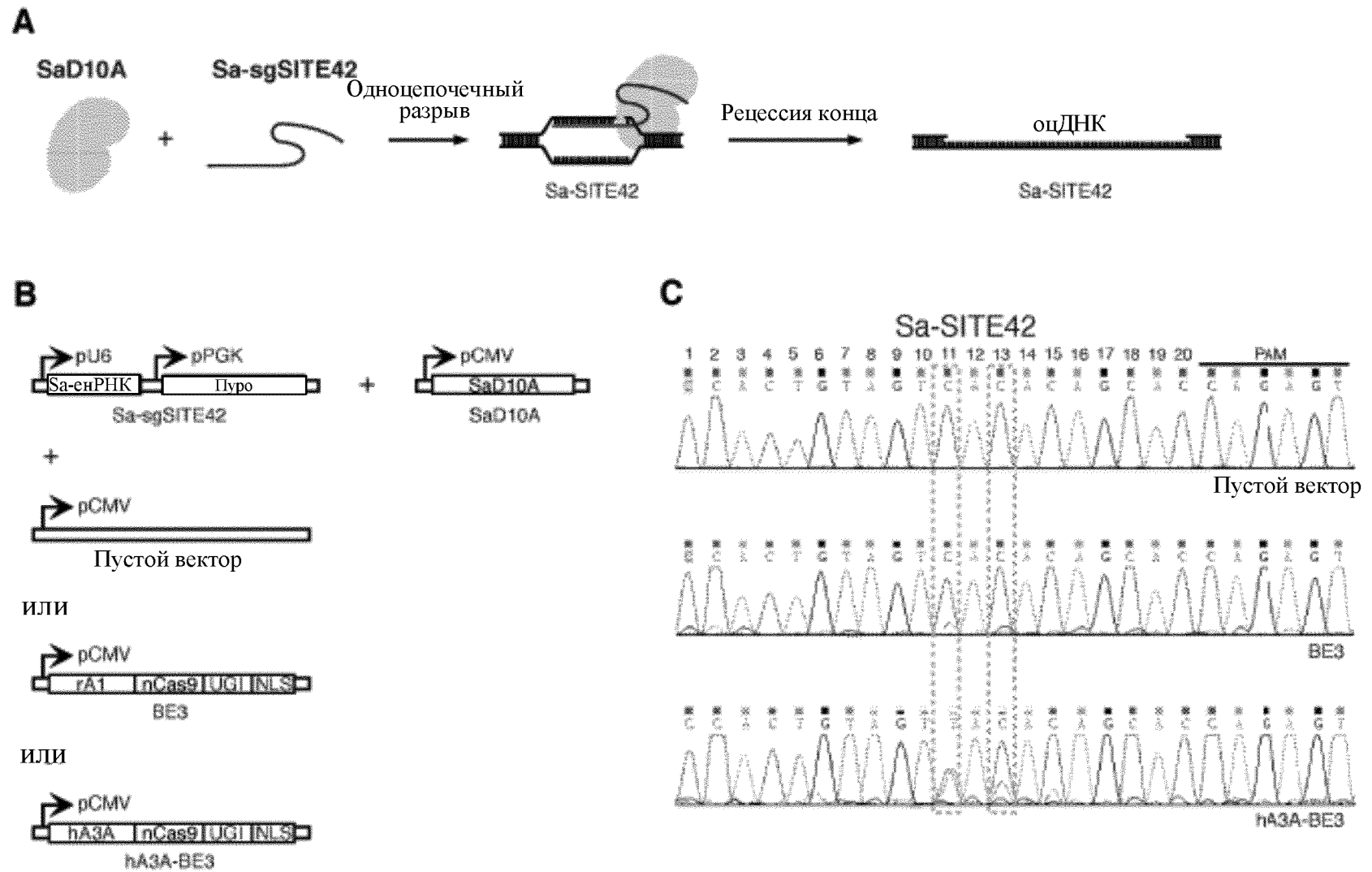
63. Одна или более полинуклеотидных последовательностей, кодирующих вспомогательную направляющую РНК и направляющую РНК по любому из пп. 56 - 62.

64. Способ инактивации гена PSCK9 в клетке, включающий приведение в контакт клетки с парой вспомогательной направляющей РНК и направляющей РНК по любому из пп. 56 - 62, белком, ассоциированным с короткими палиндромными повторами, регулярно расположенными группами (CRISPR) (Cas), и дезаминазой азотистых оснований.

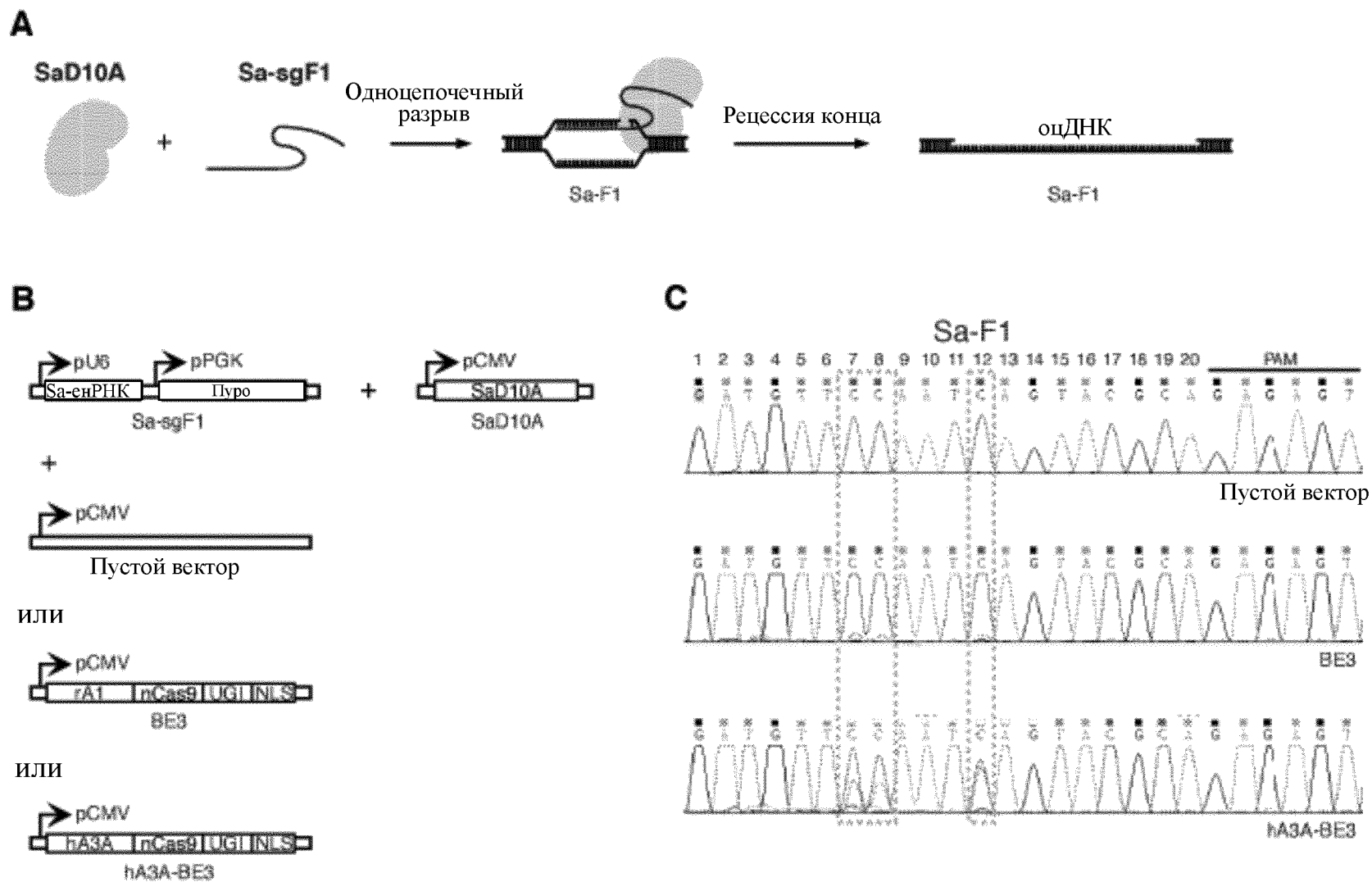
Фигура 1



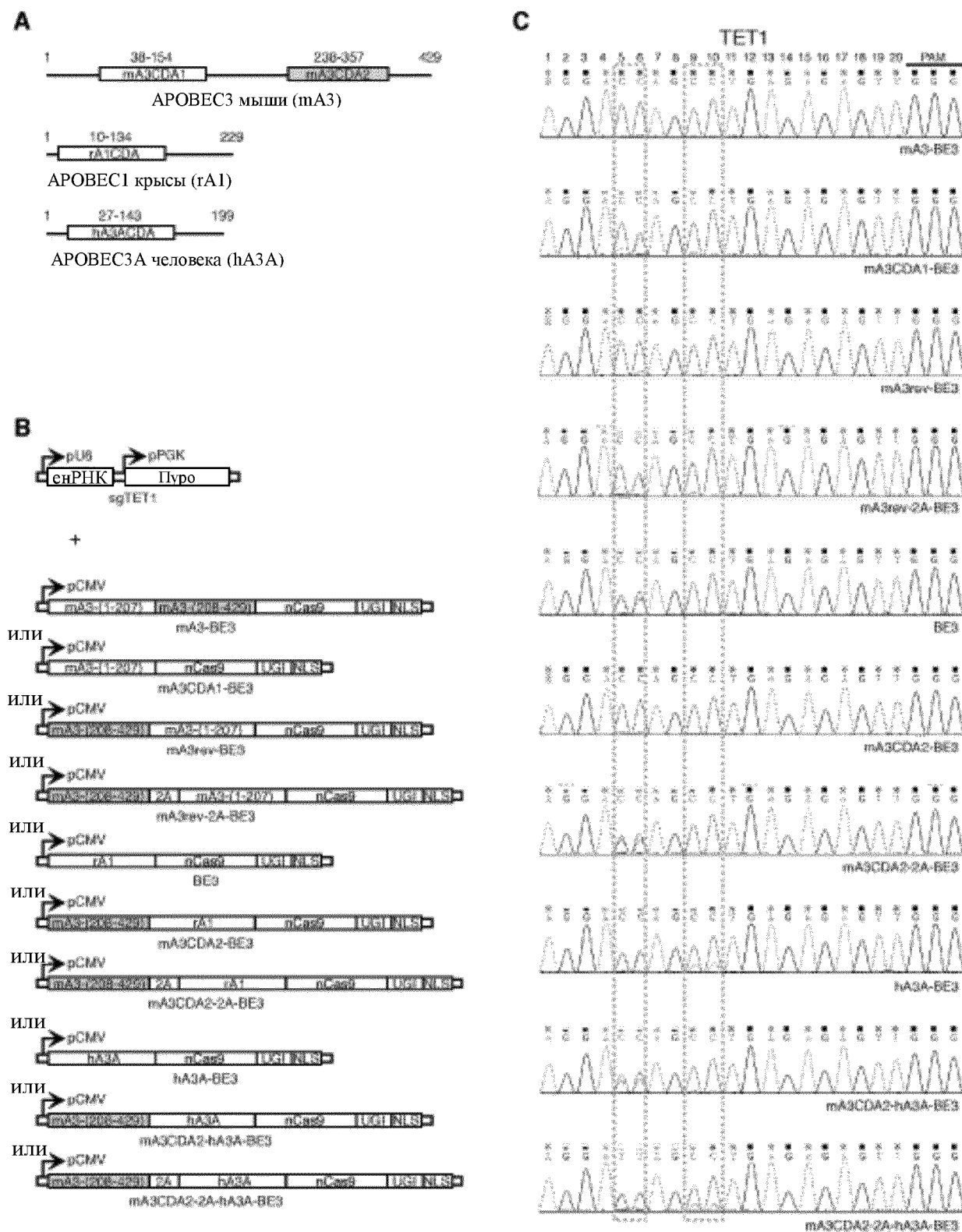
Фигура 2



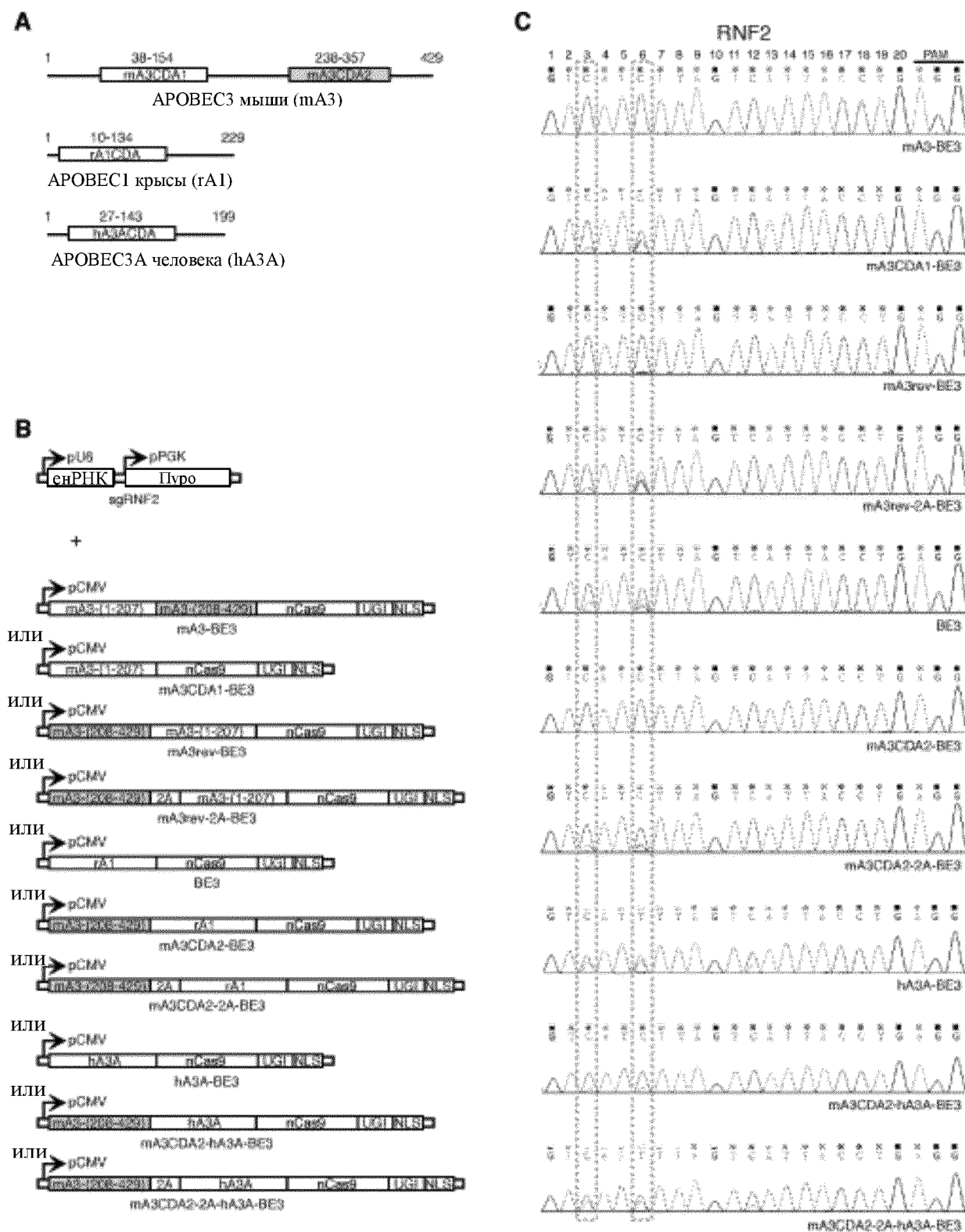
Фигура 3



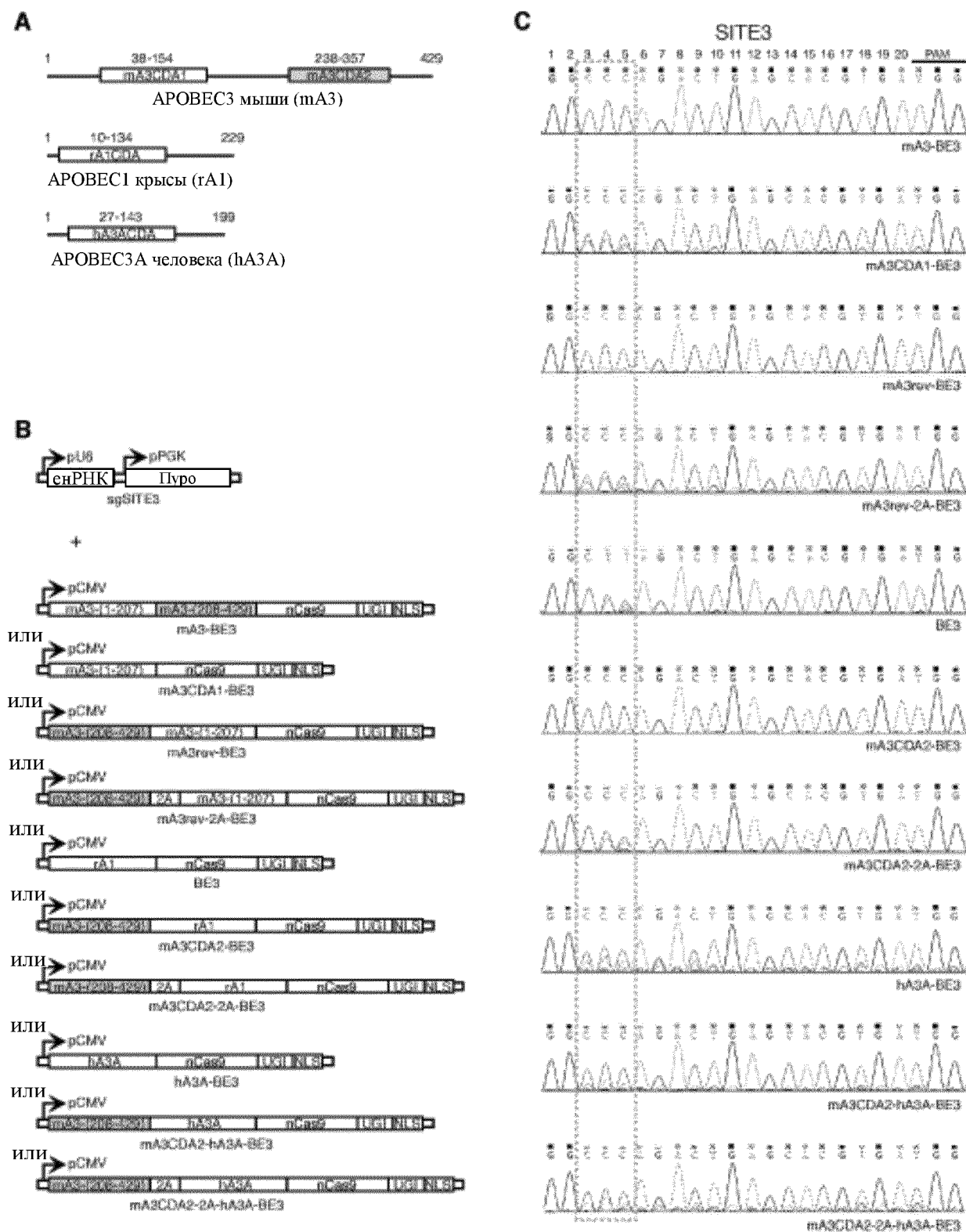
Фигура 4



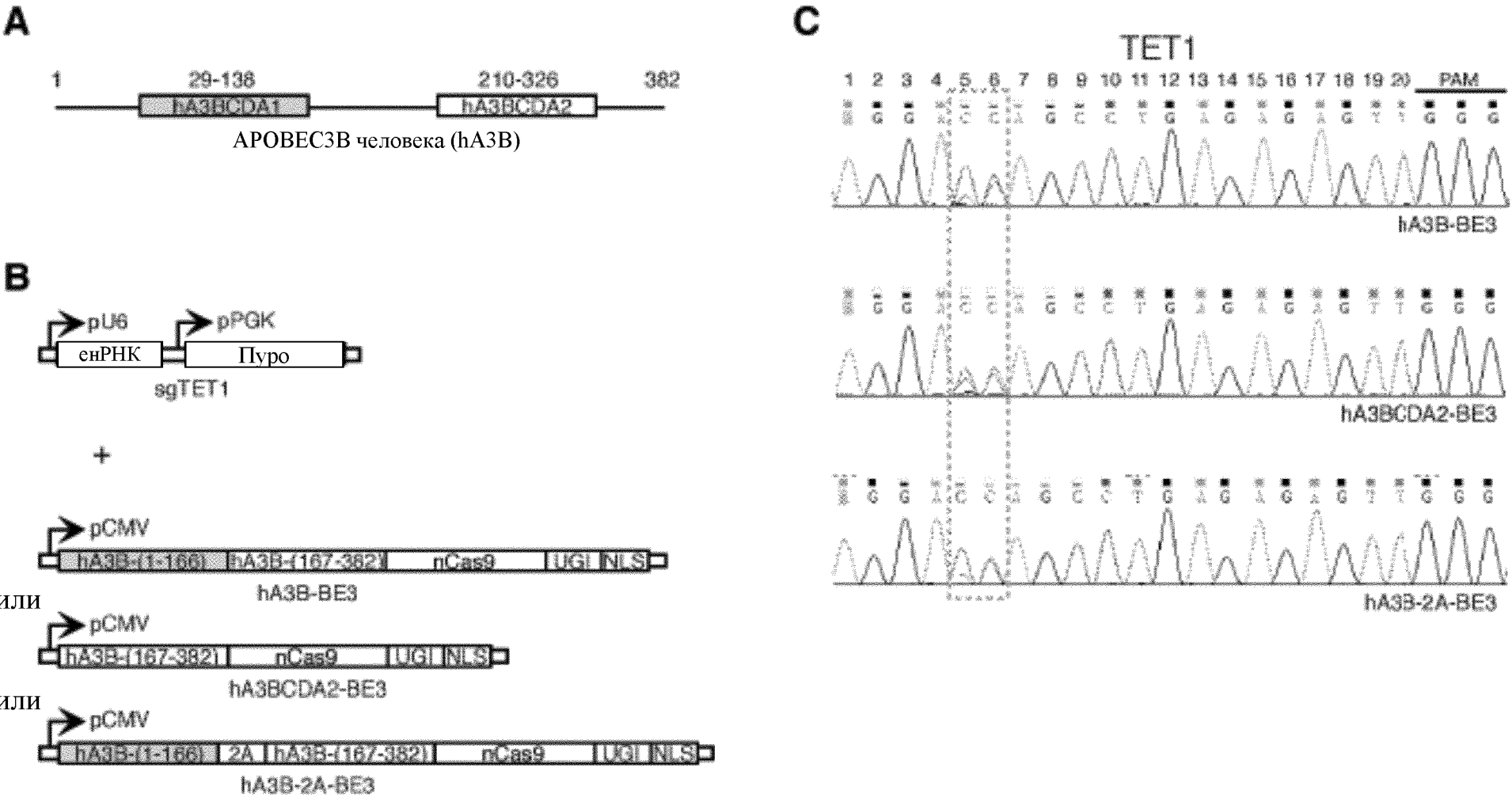
Фигура 5



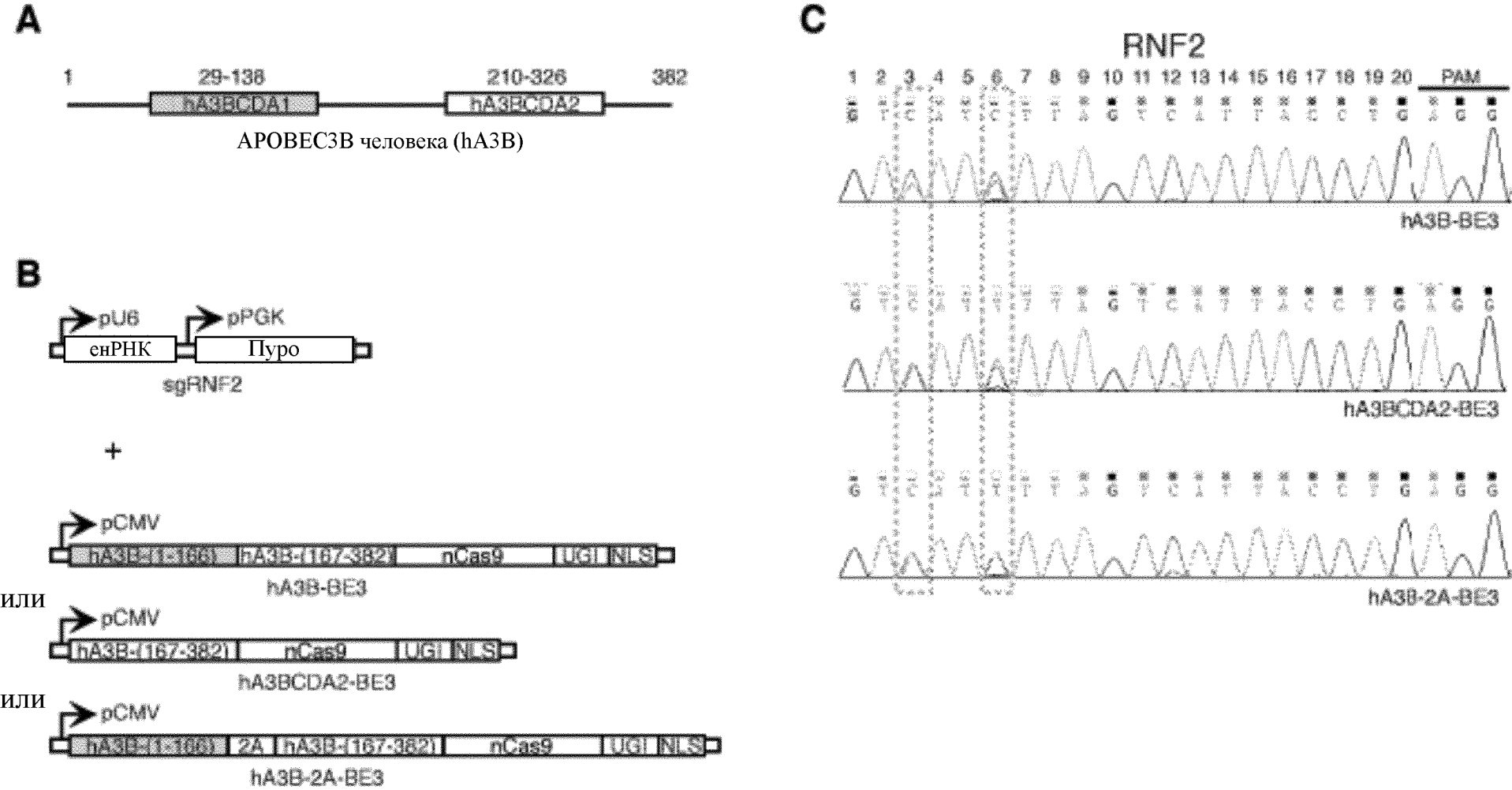
Фигура 6



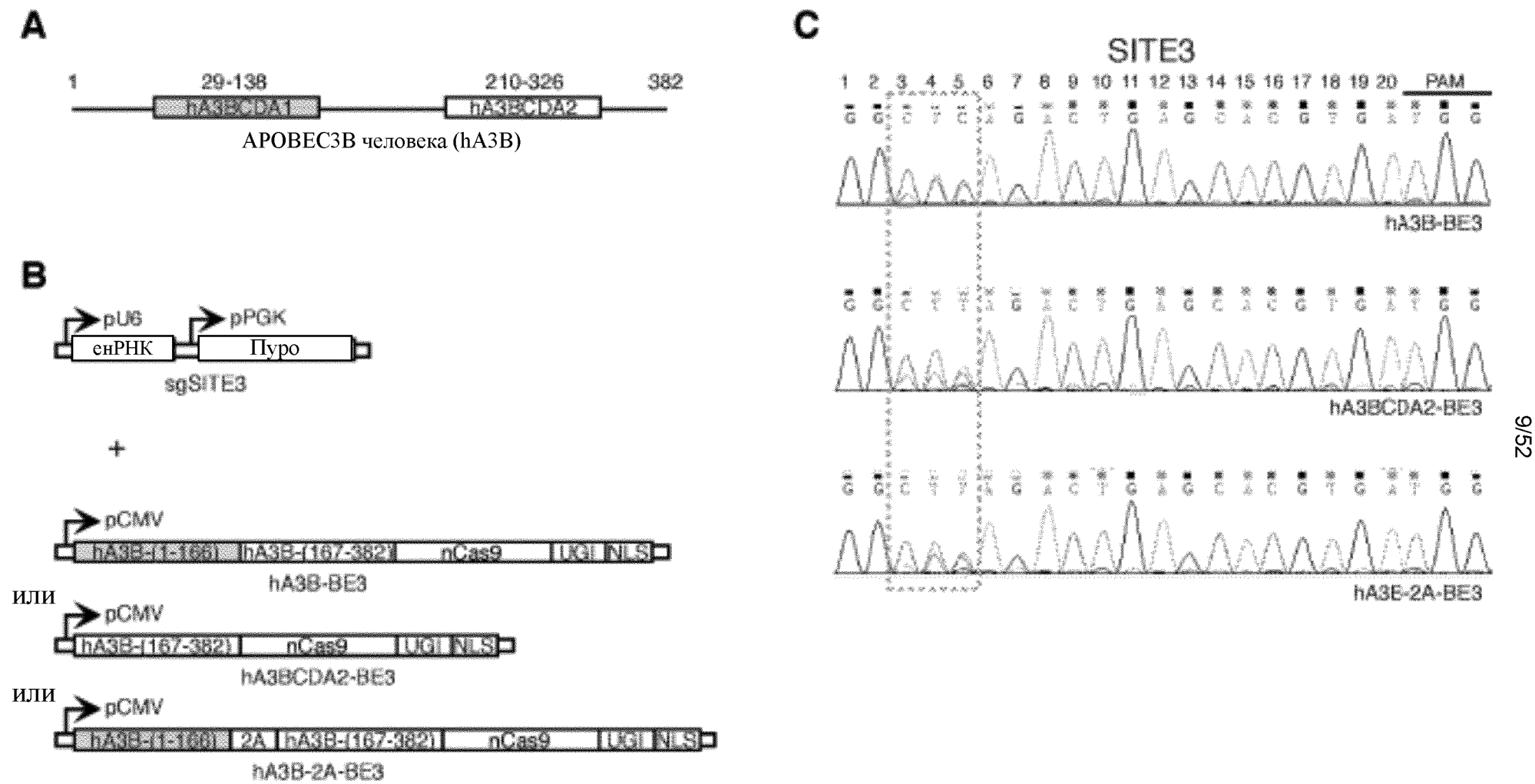
Фигура 7



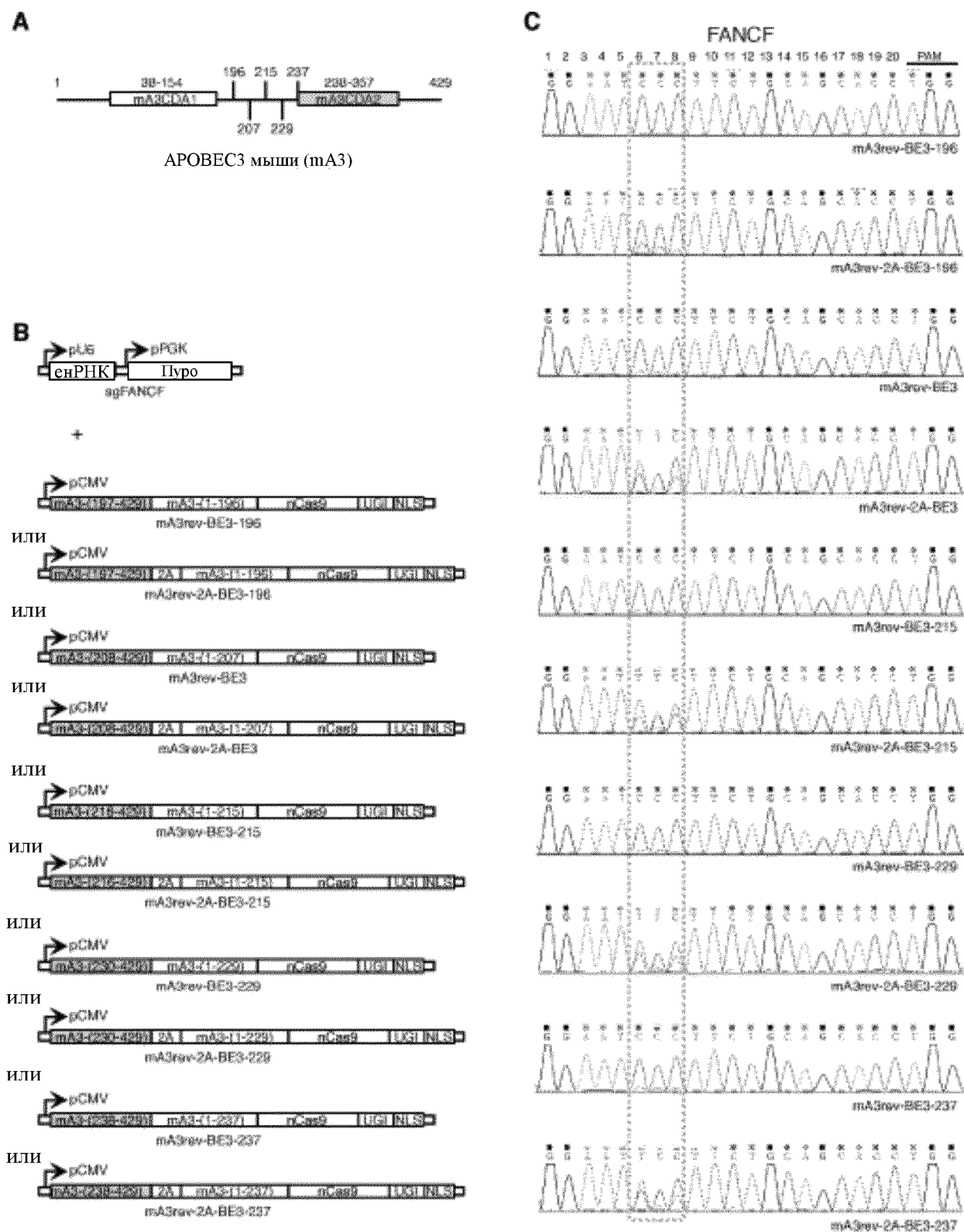
Фигура 8



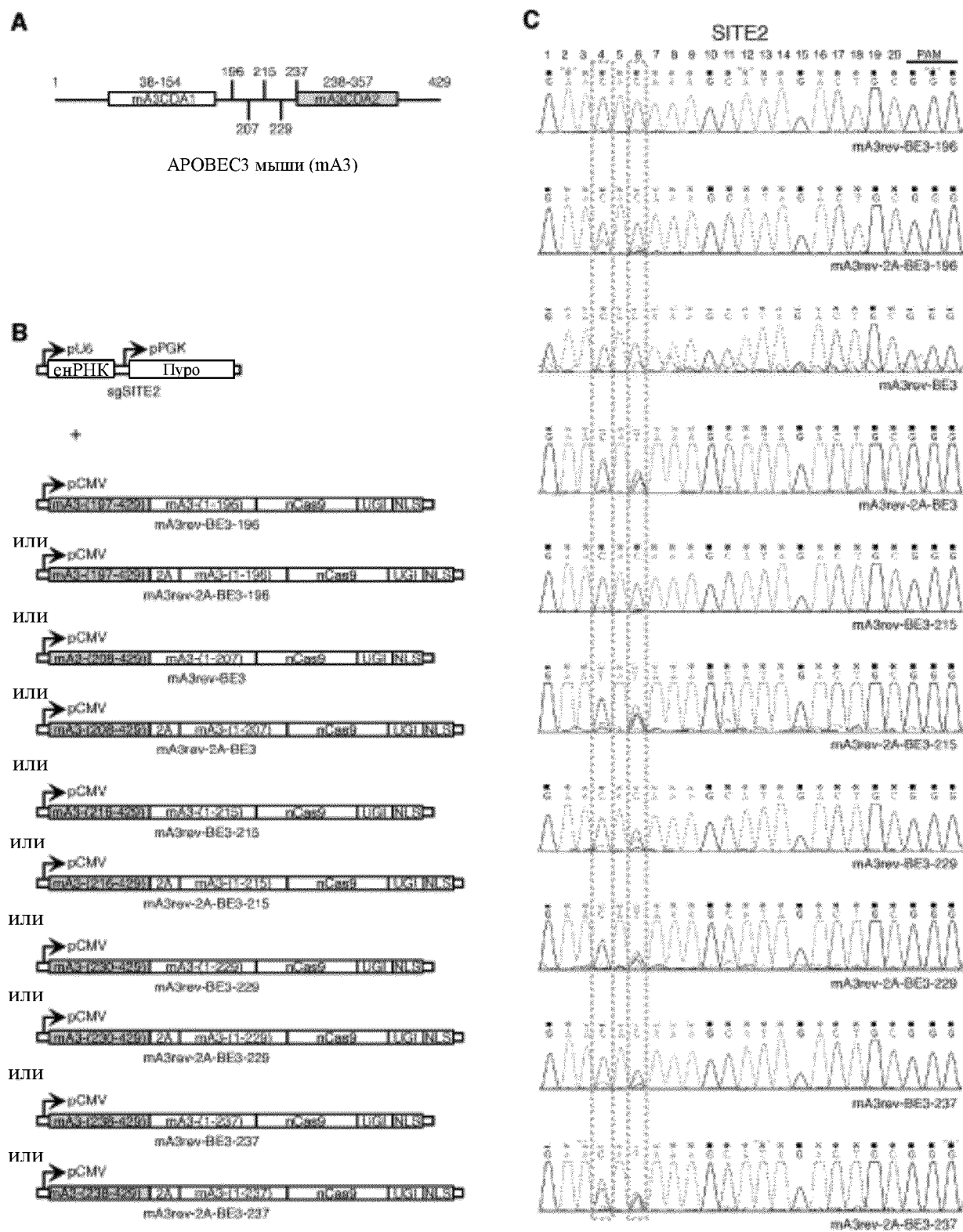
Фигура 9



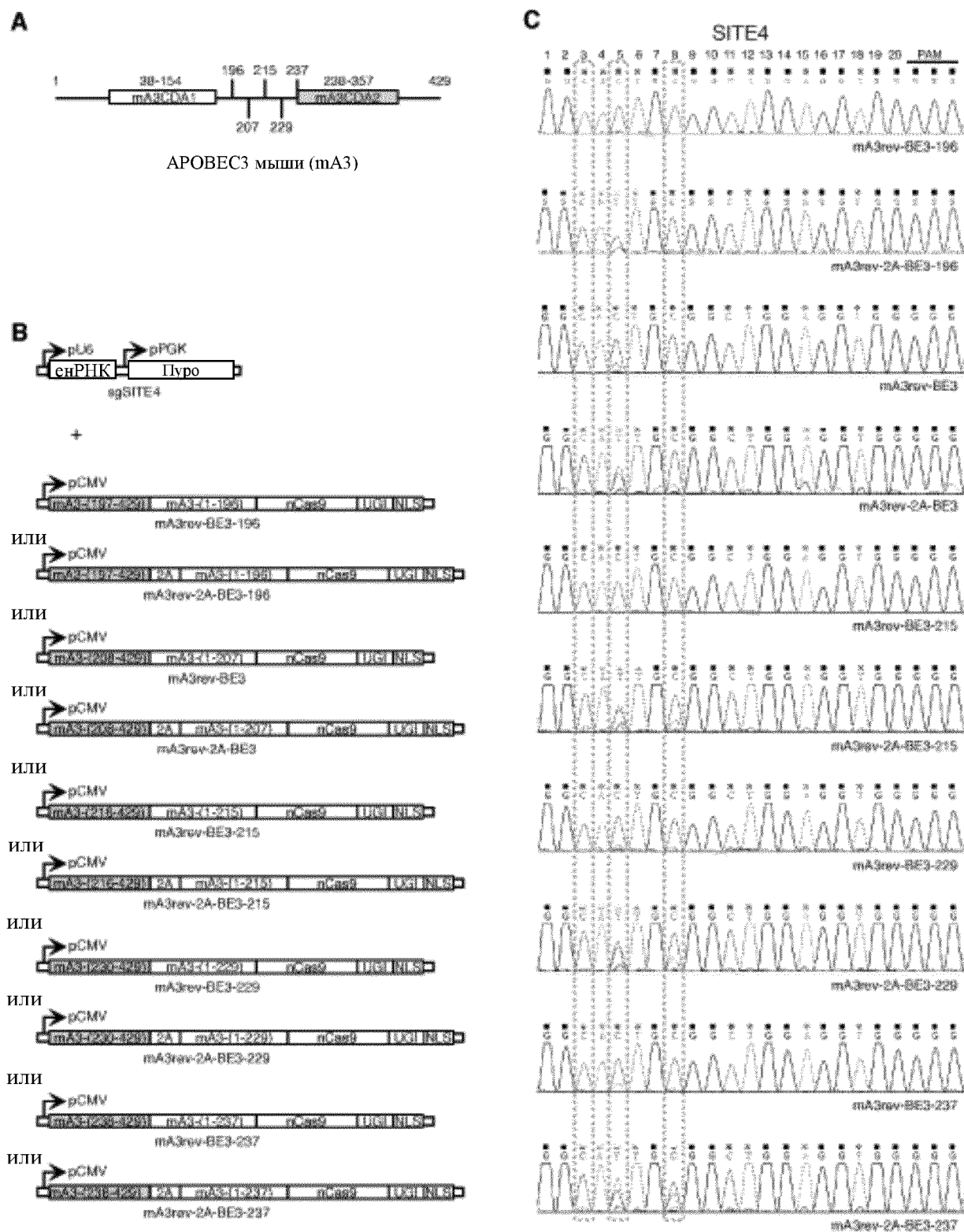
Фигура 10



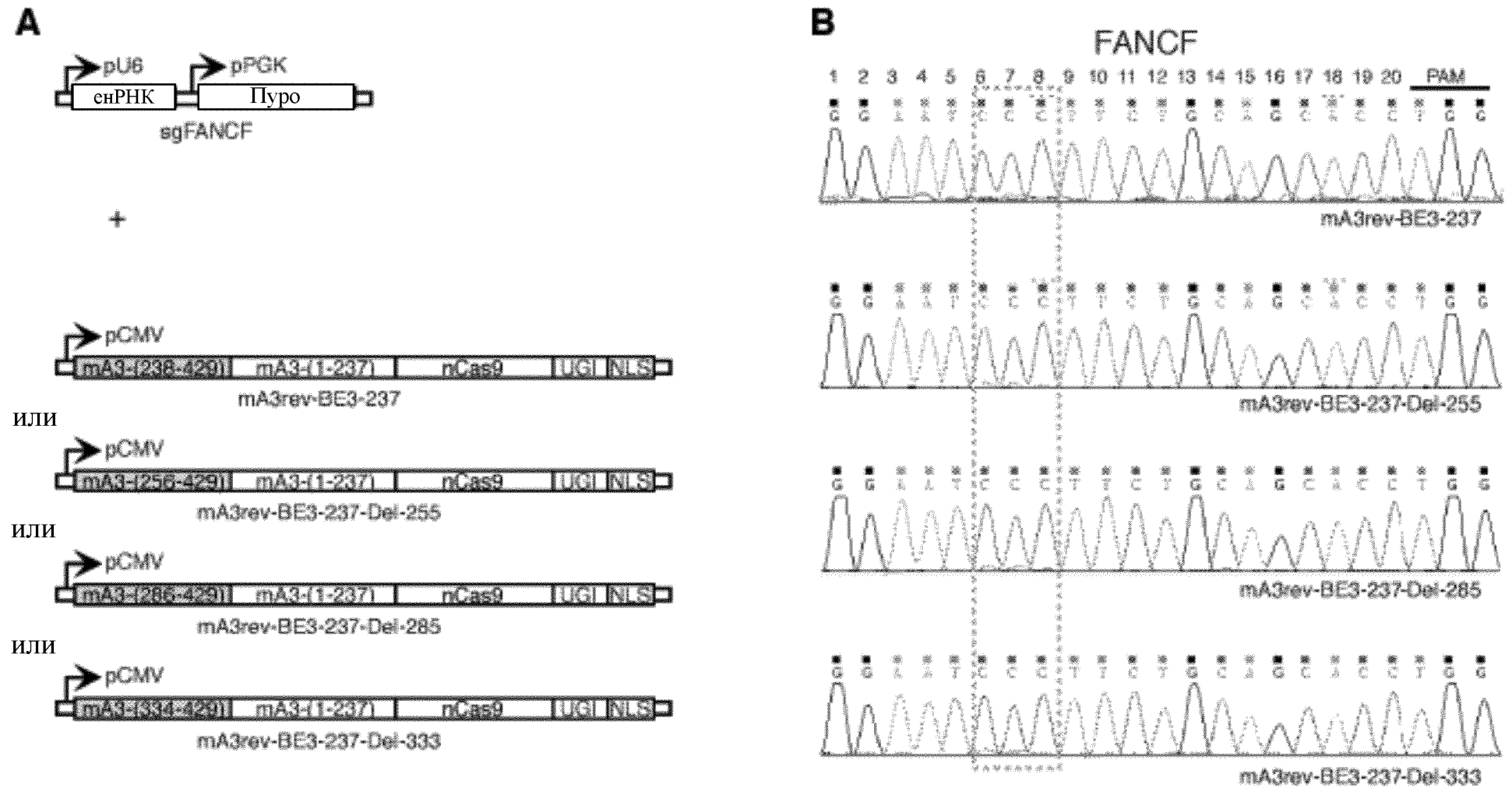
Фигура 11



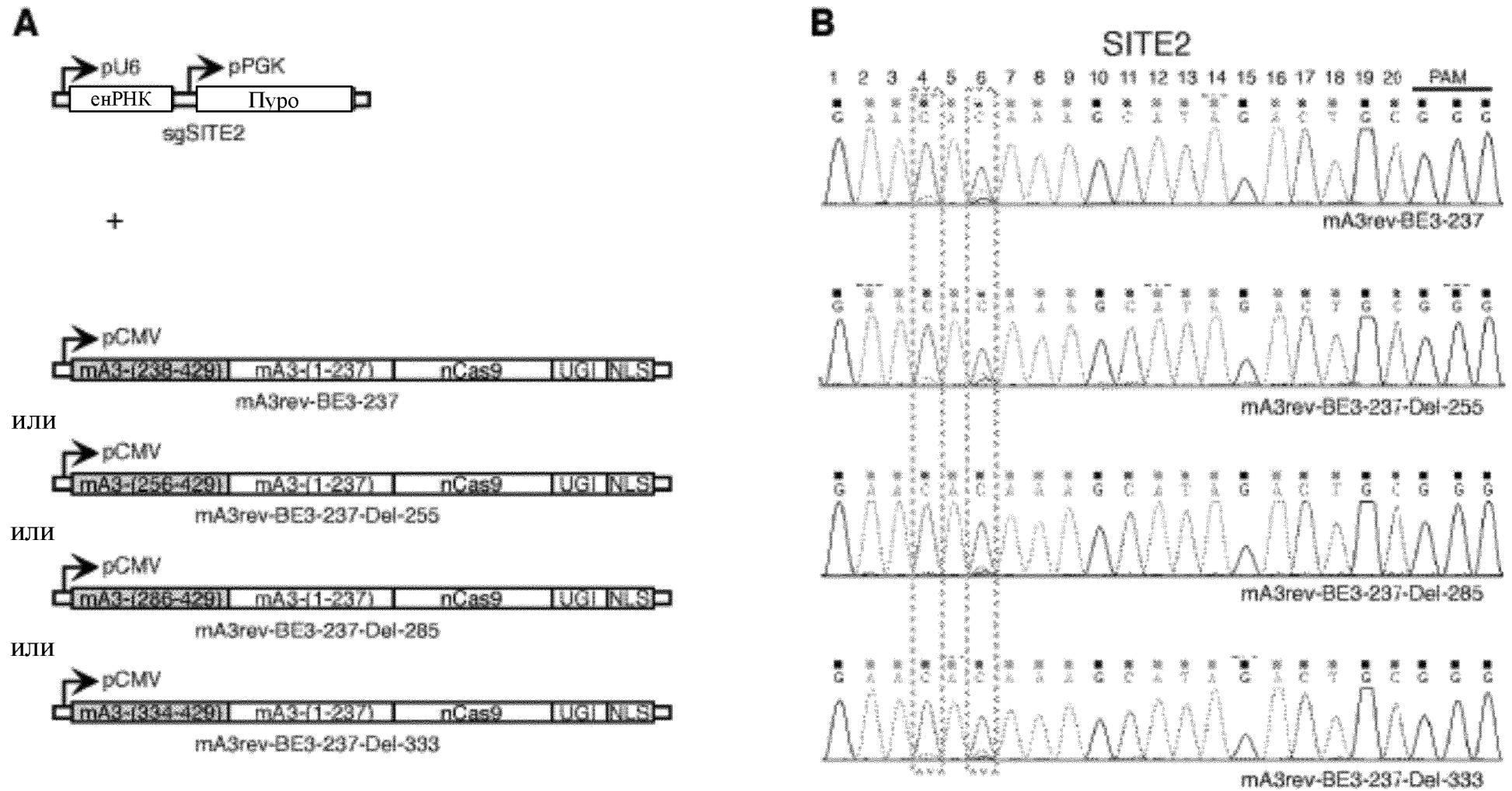
Фигура 12



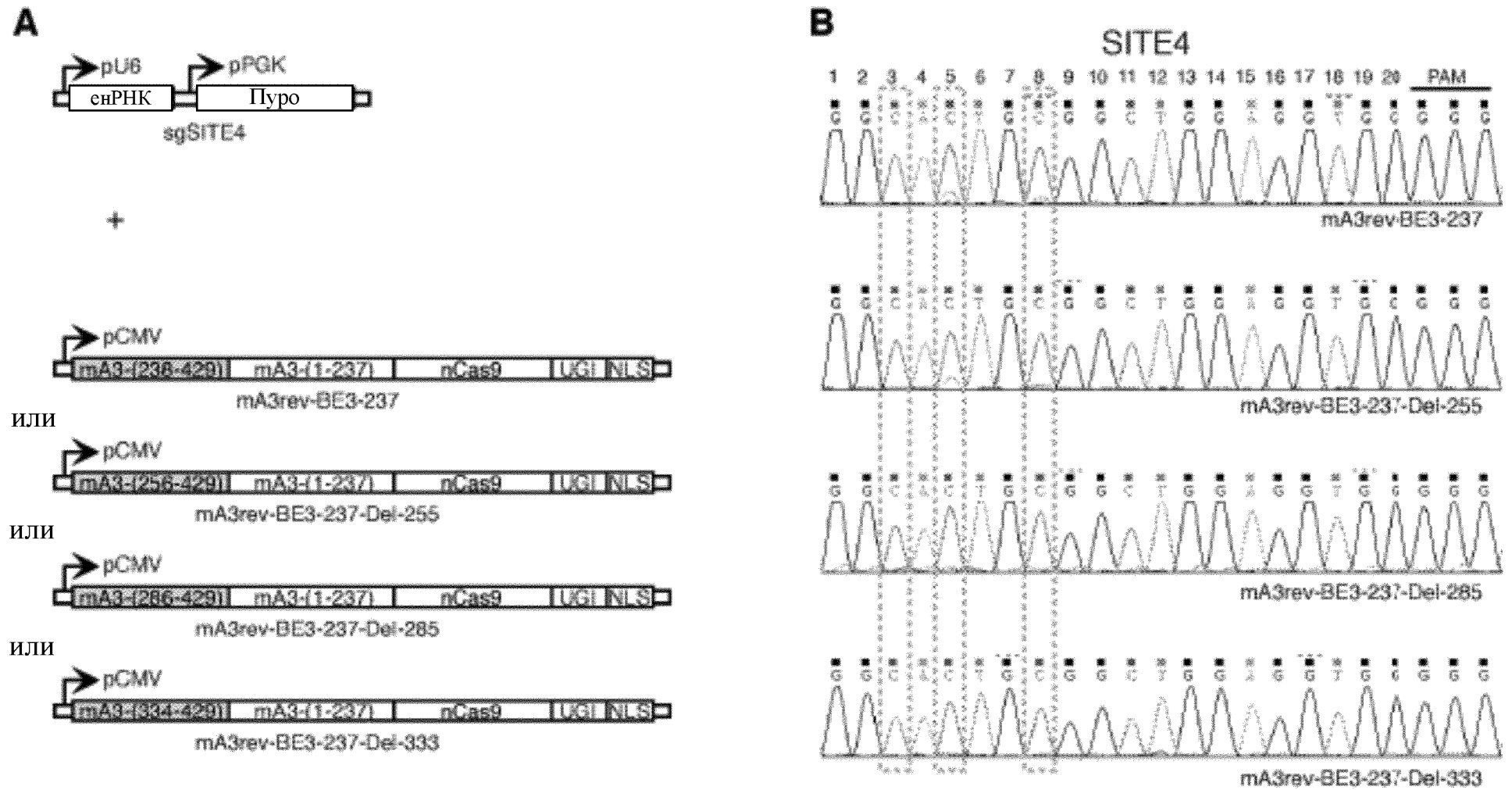
Фигура 13



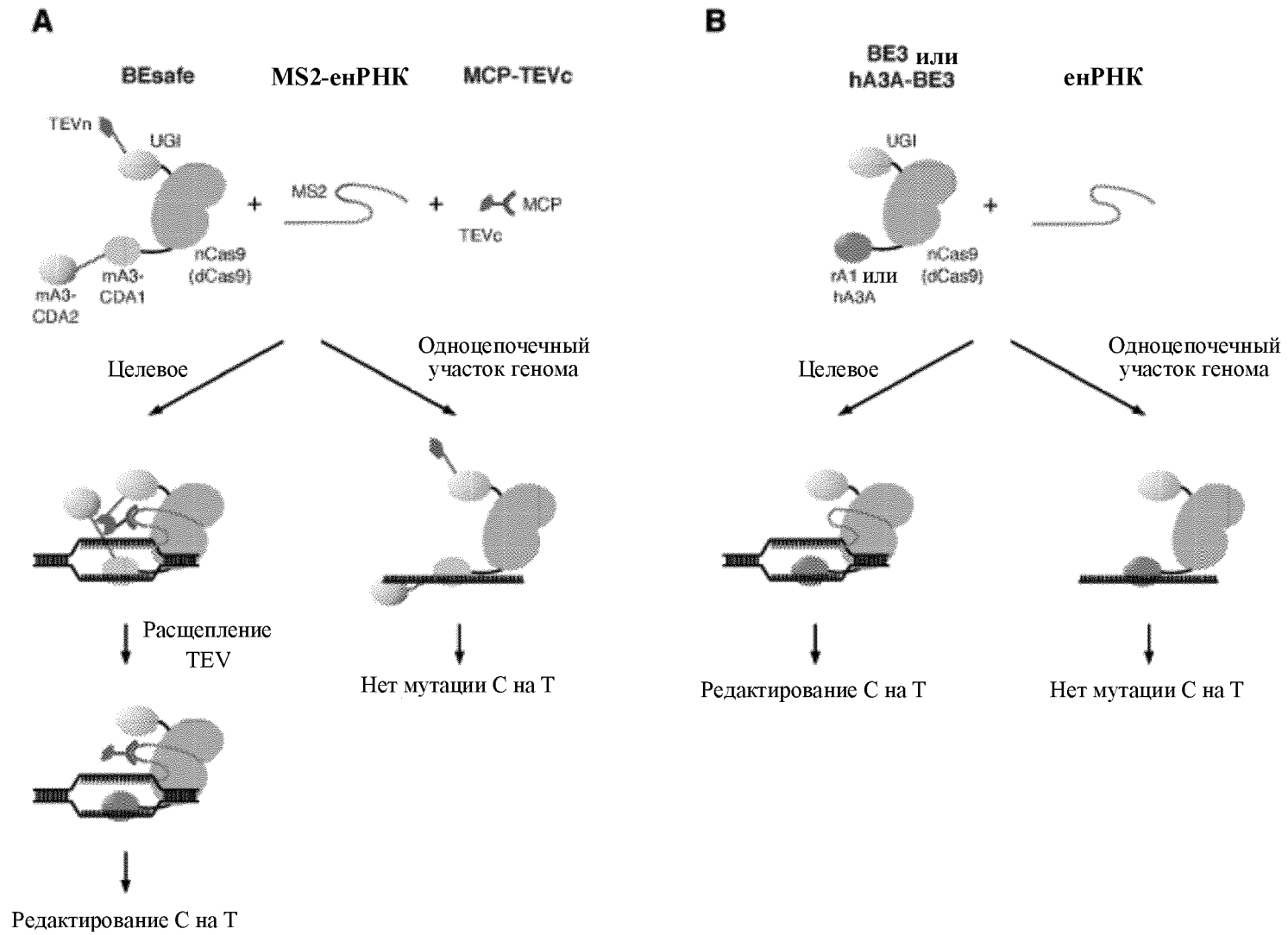
Фигура 14



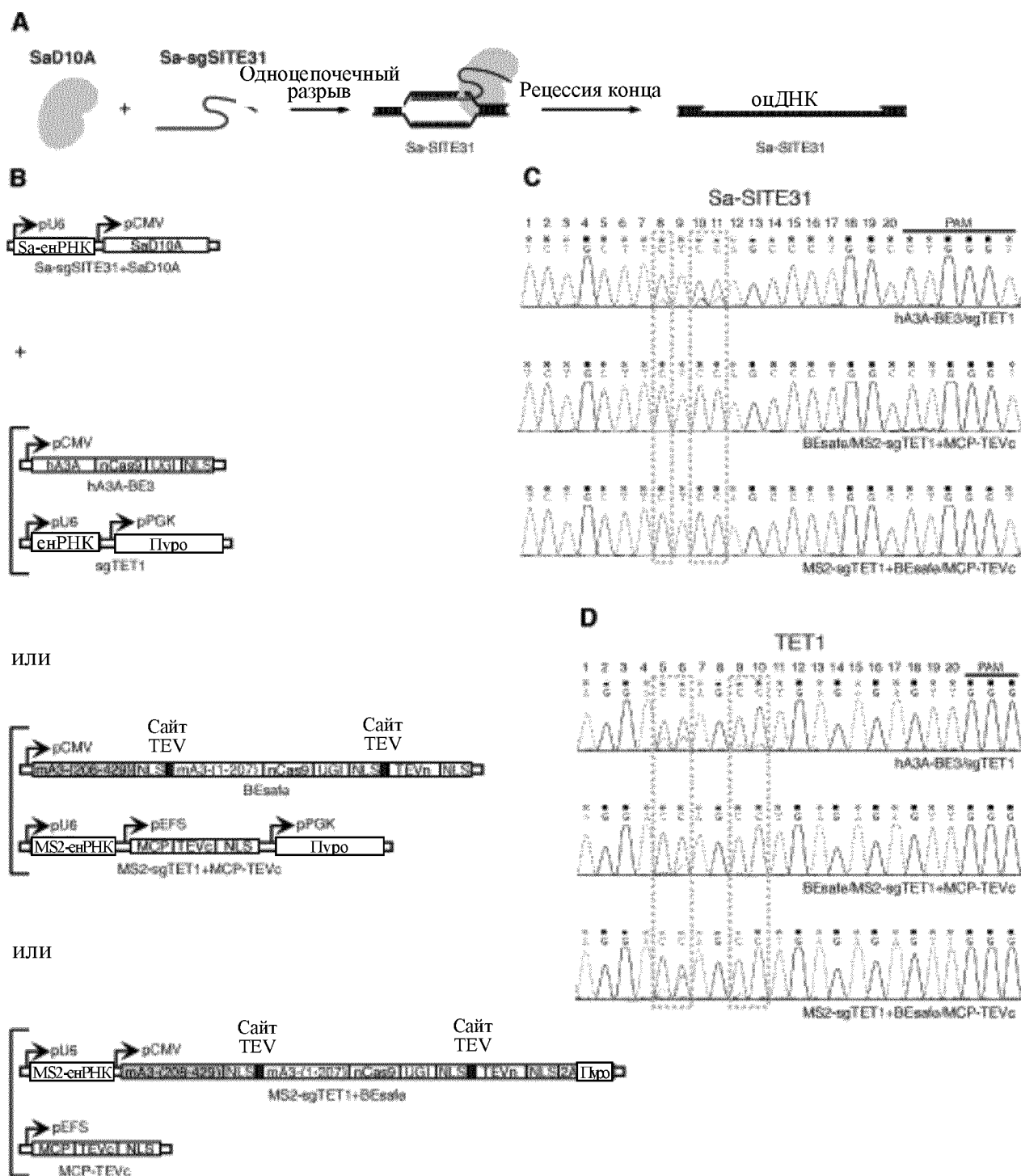
Фигура 15



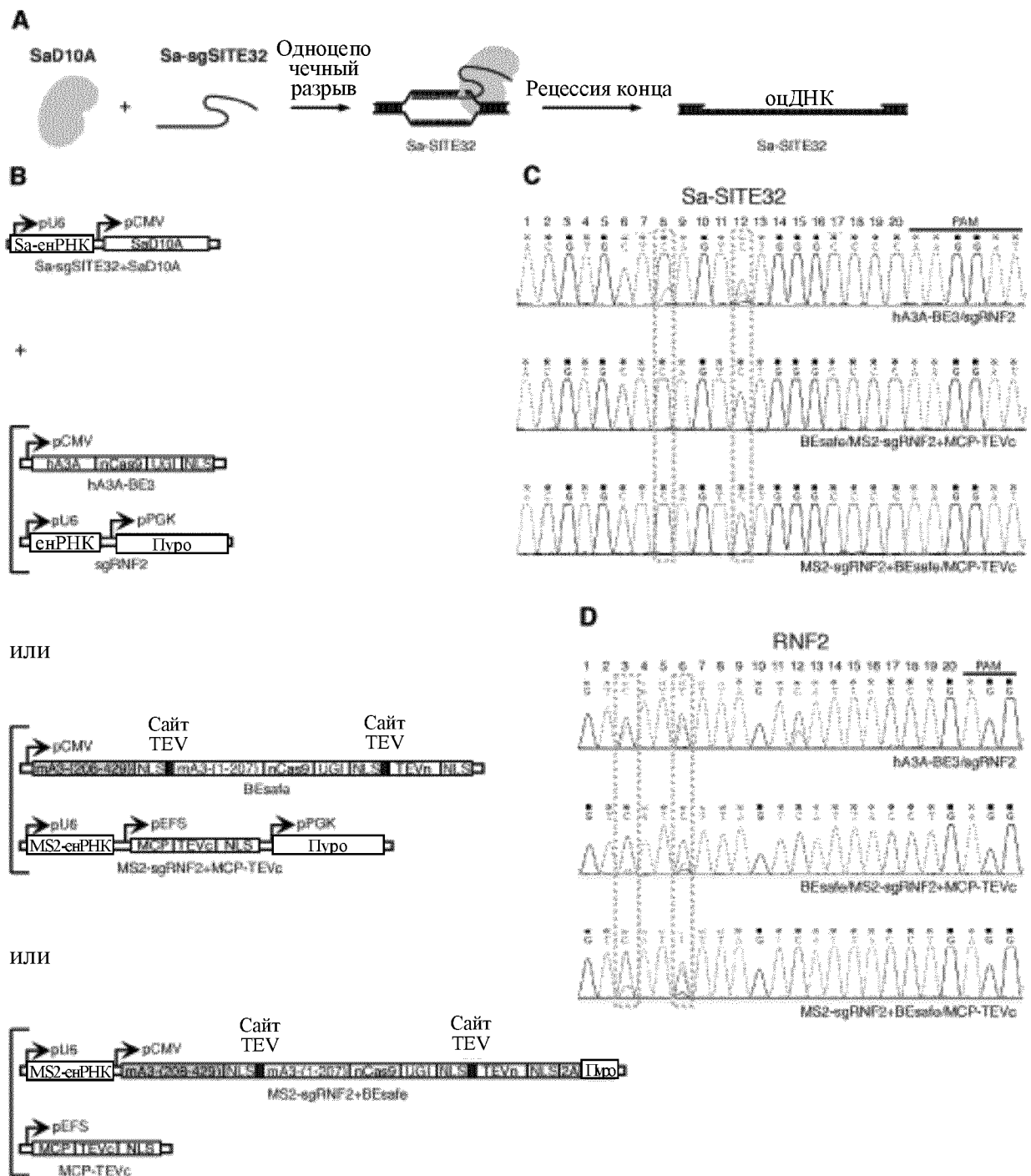
Фигура 16



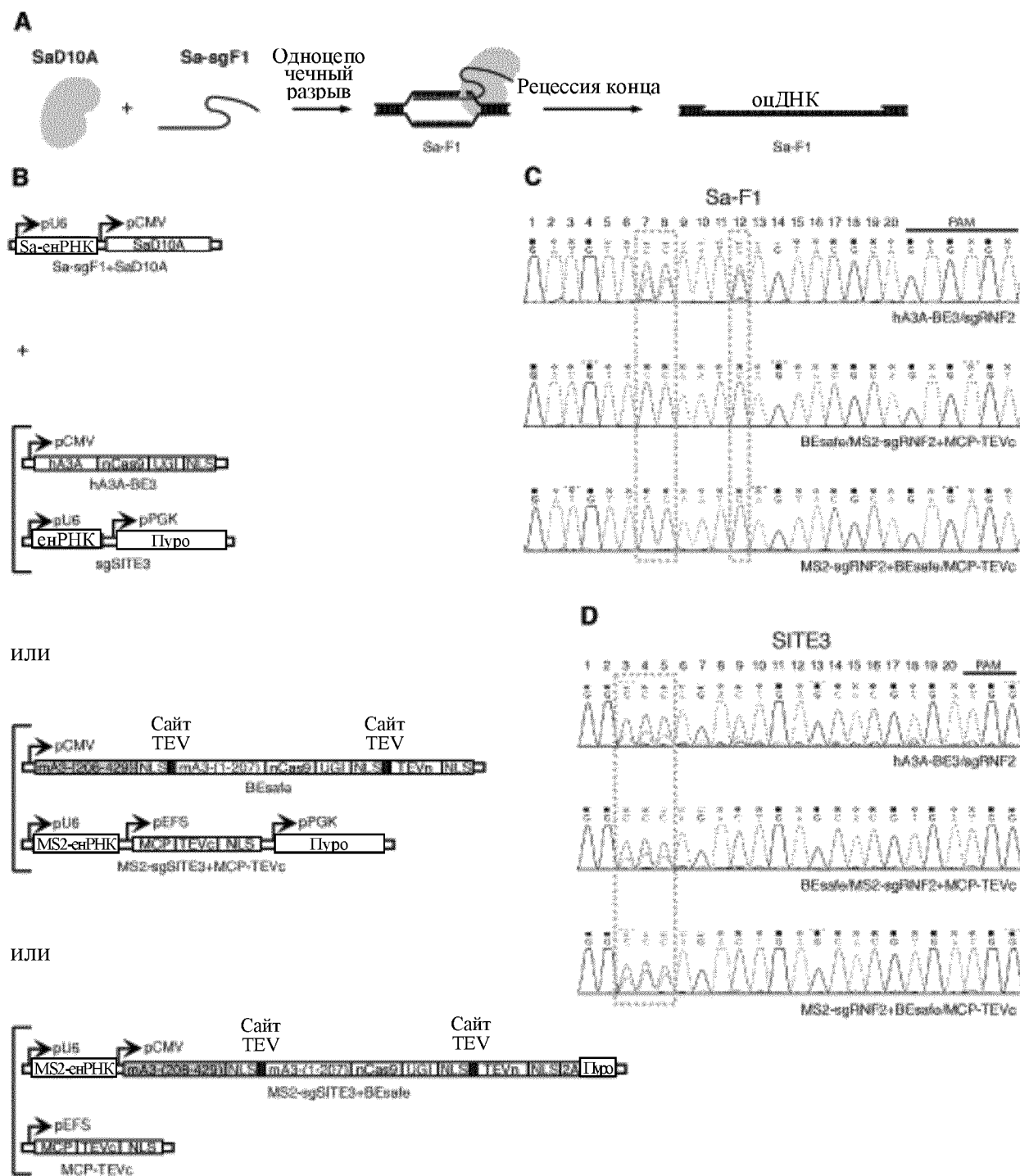
Фигура 17



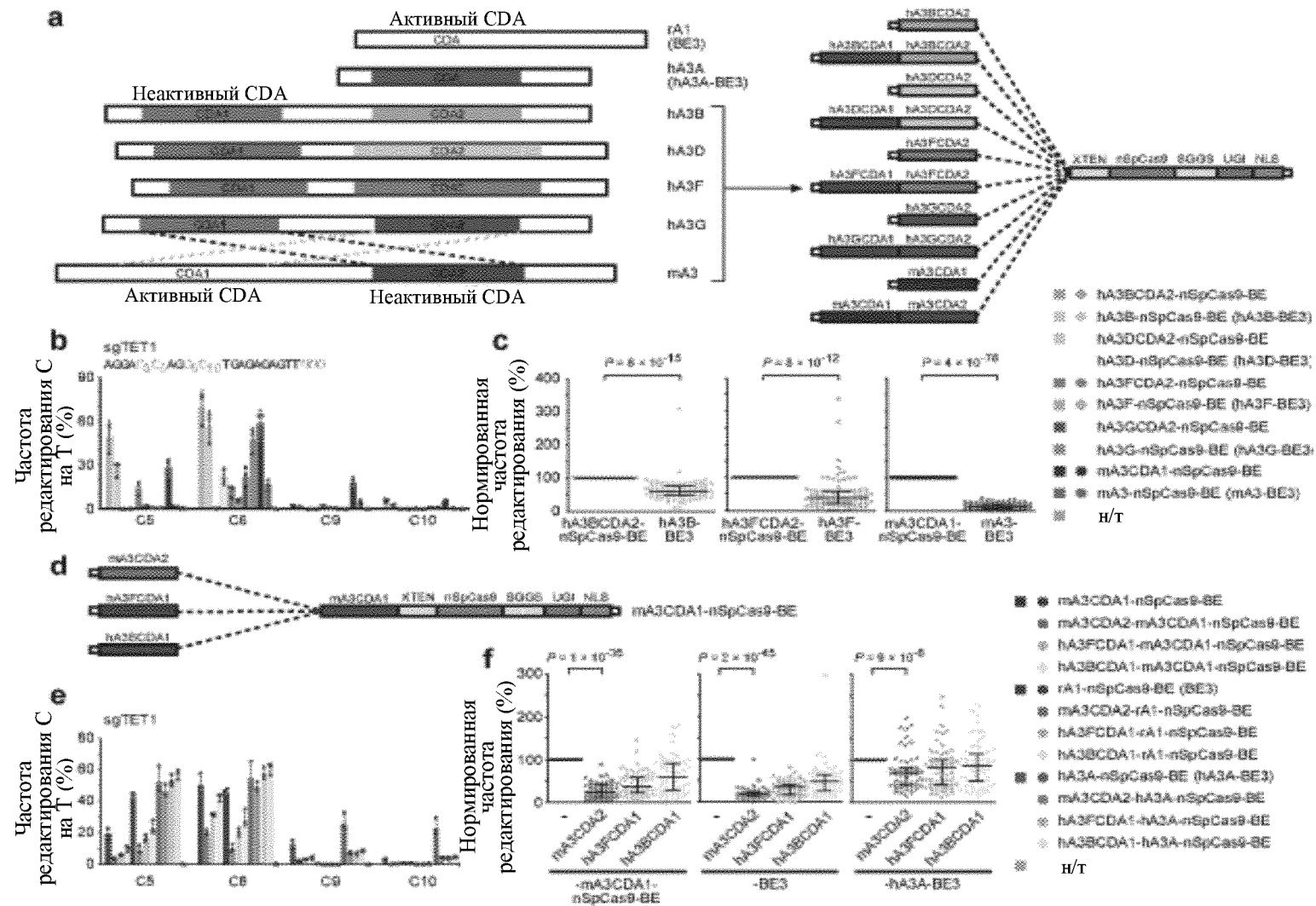
Фигура 18



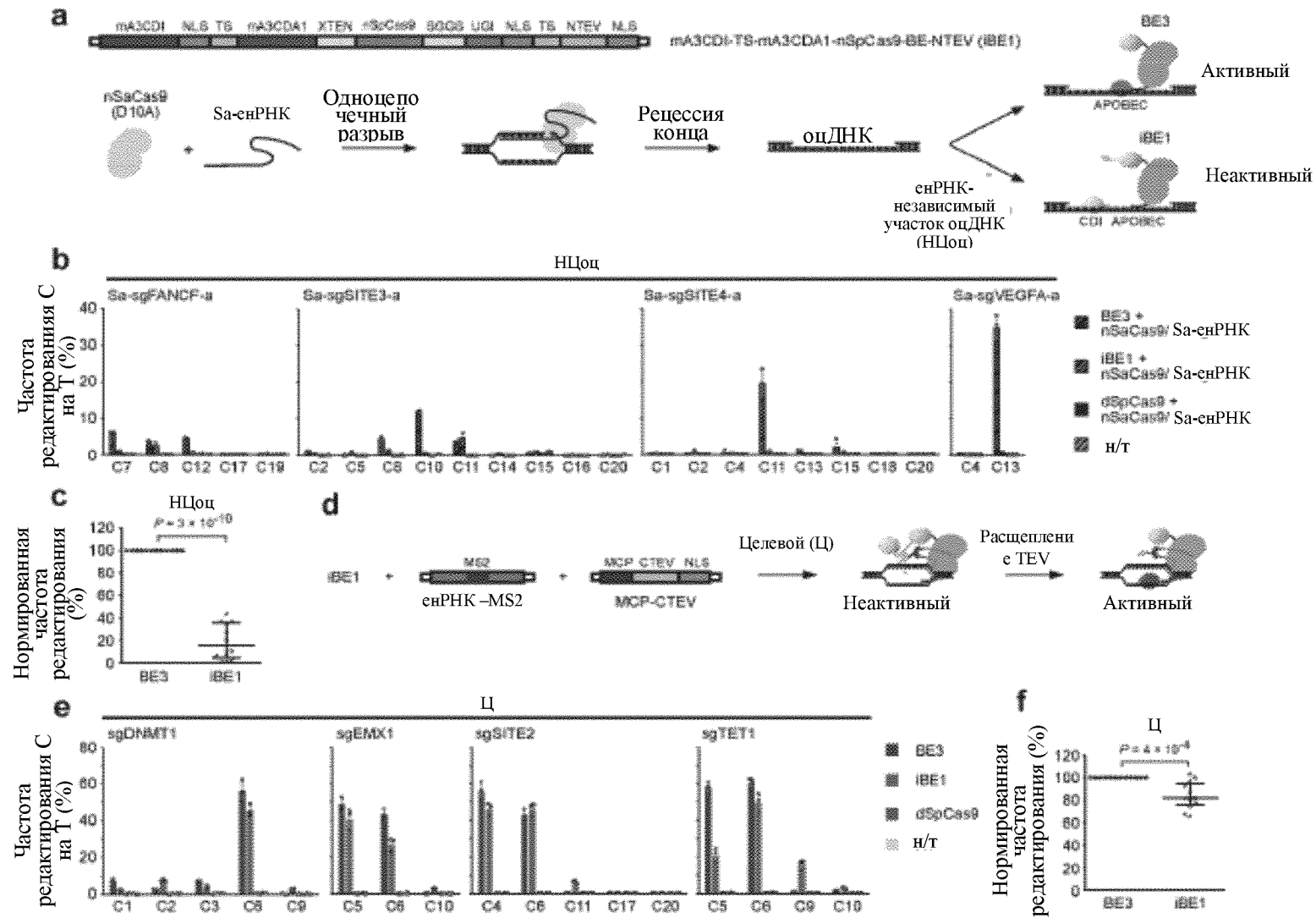
Фигура 19



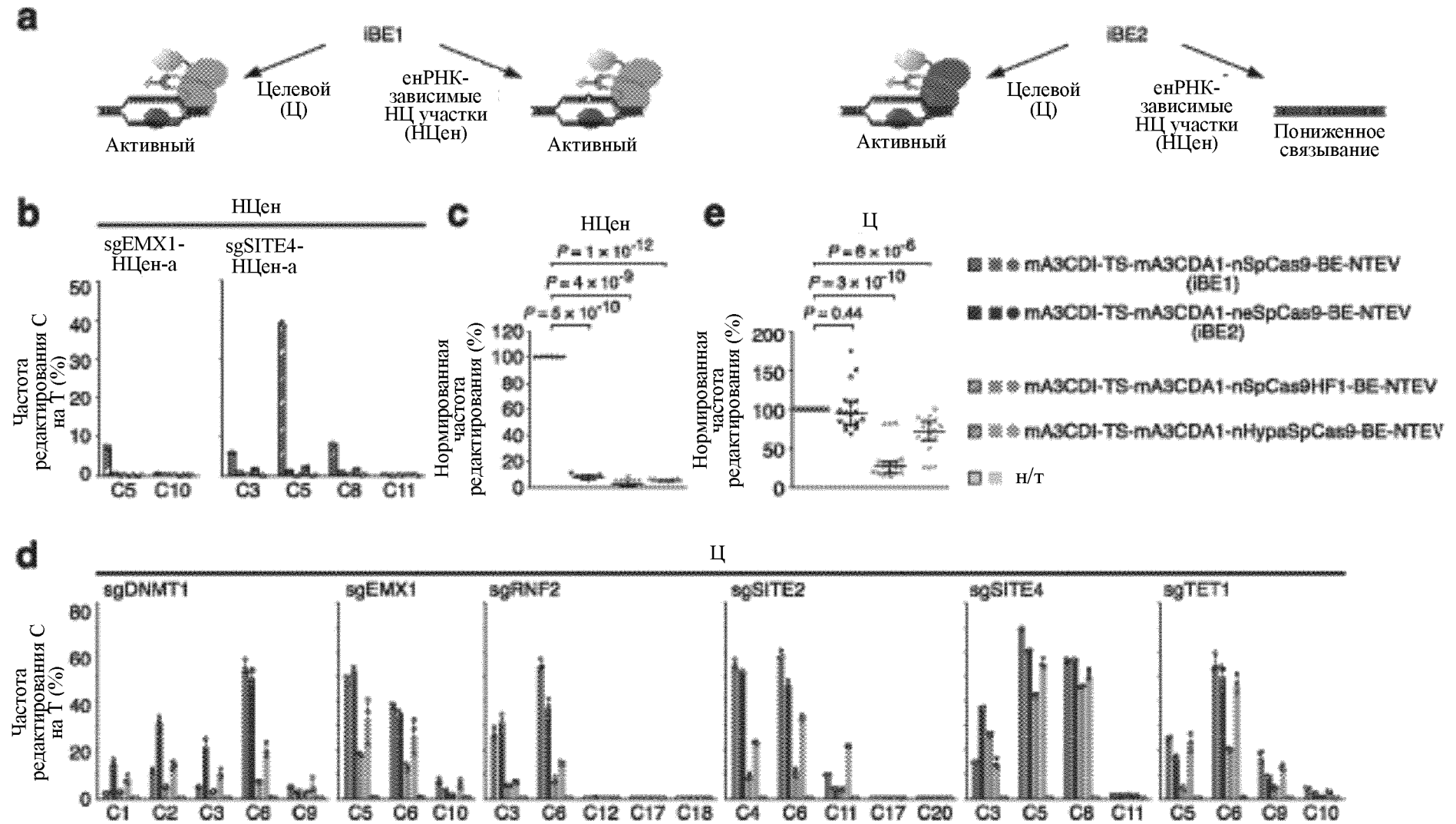
Фигура 20



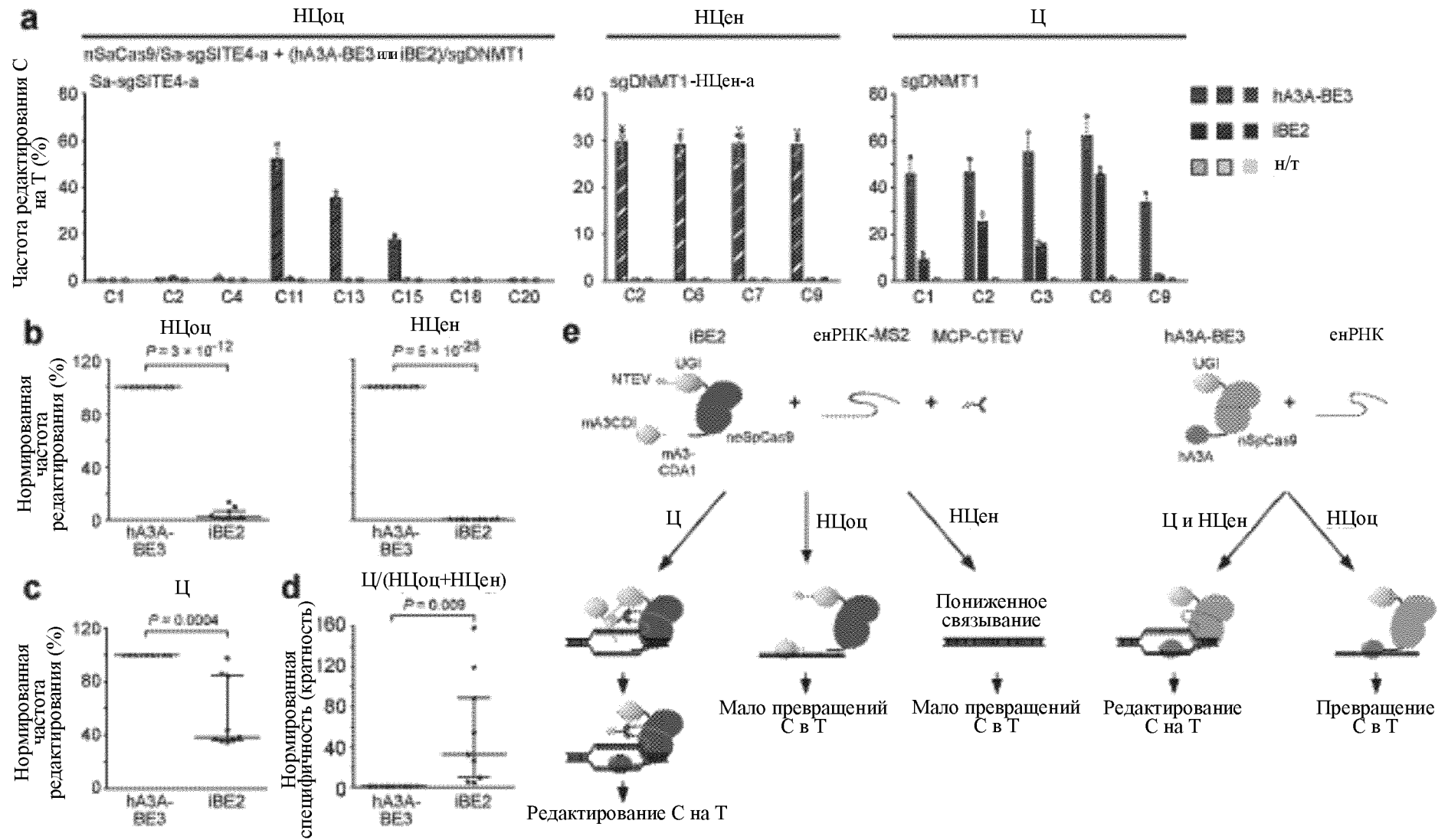
Фигура 21



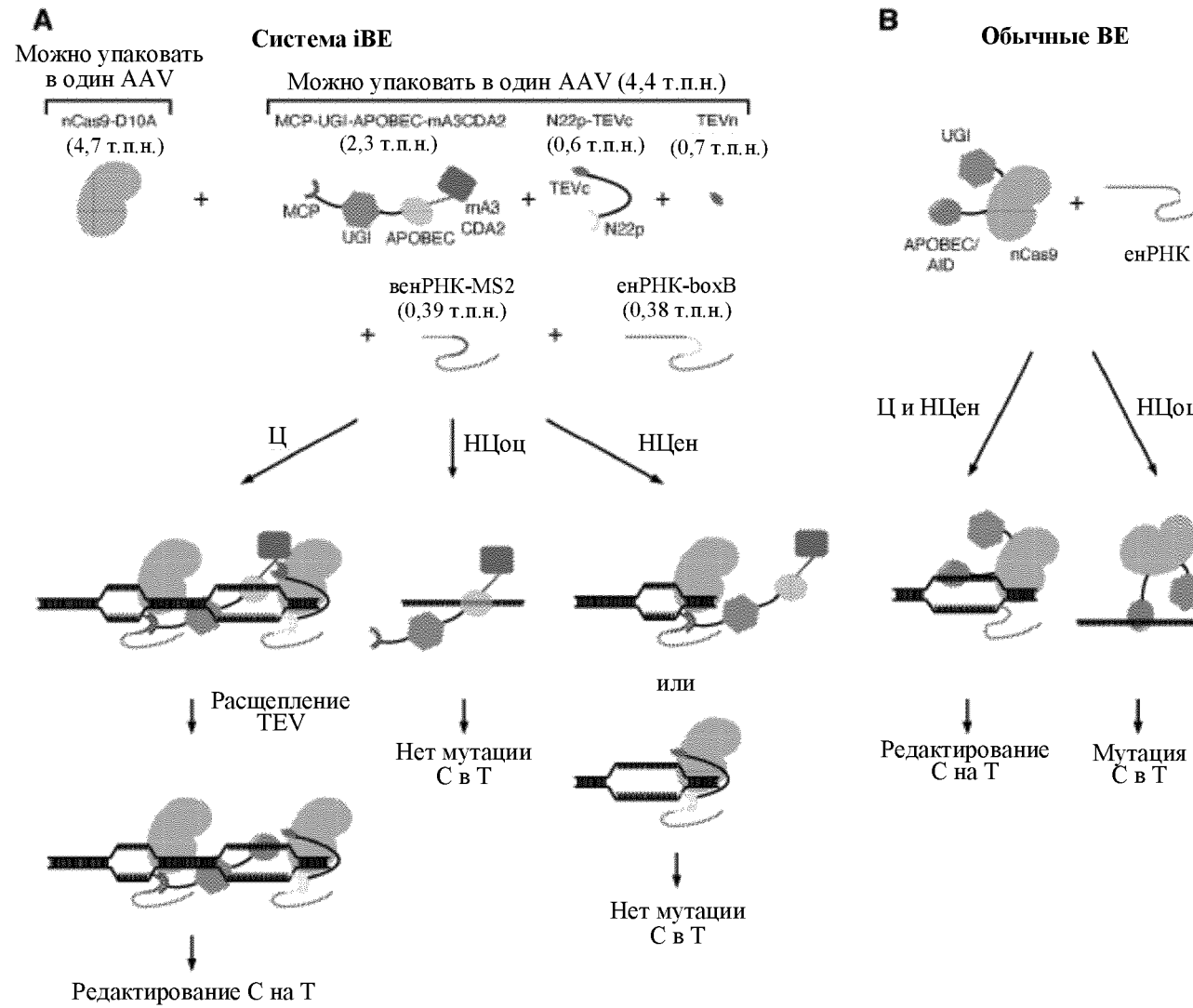
Фигура 22



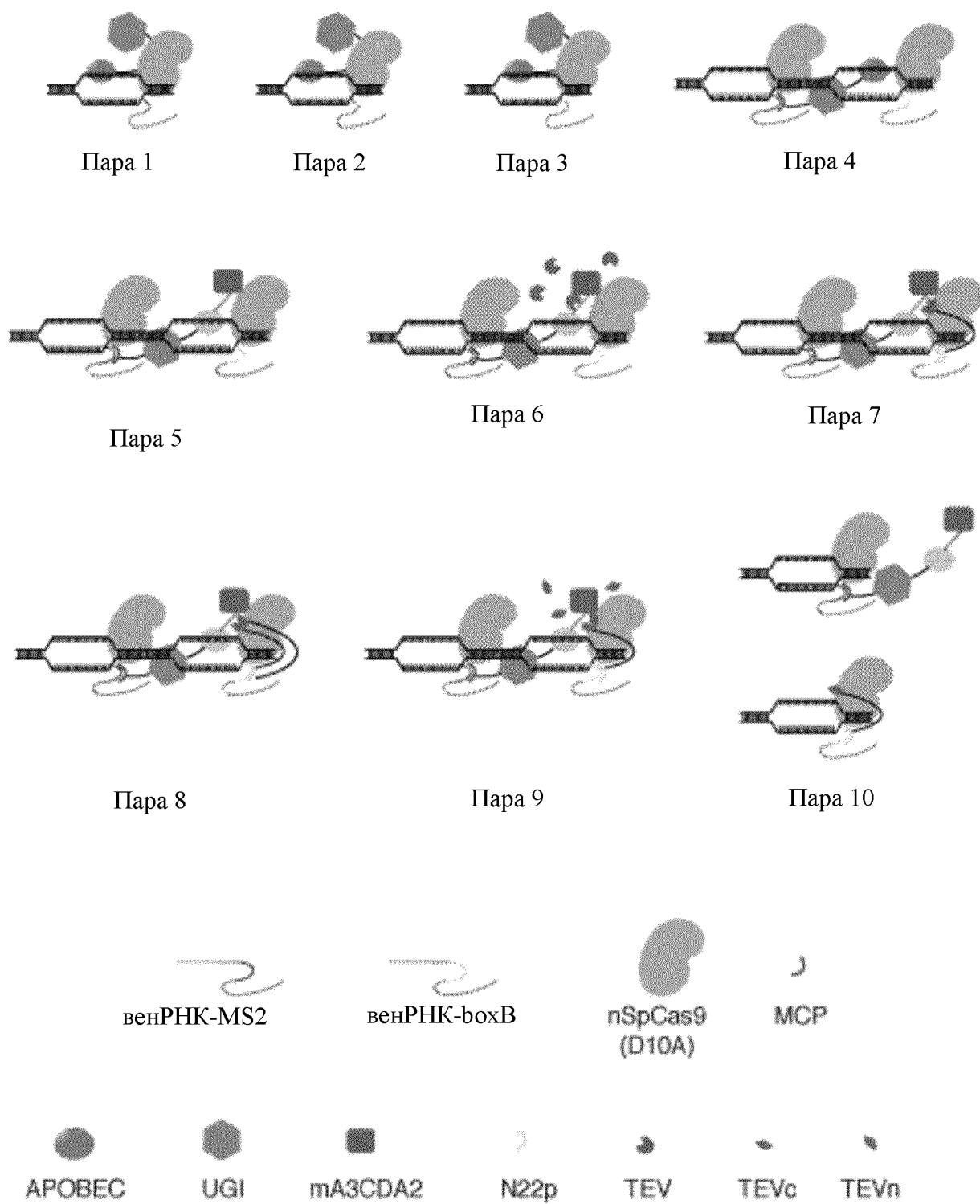
Фигура 23



Фигура 24

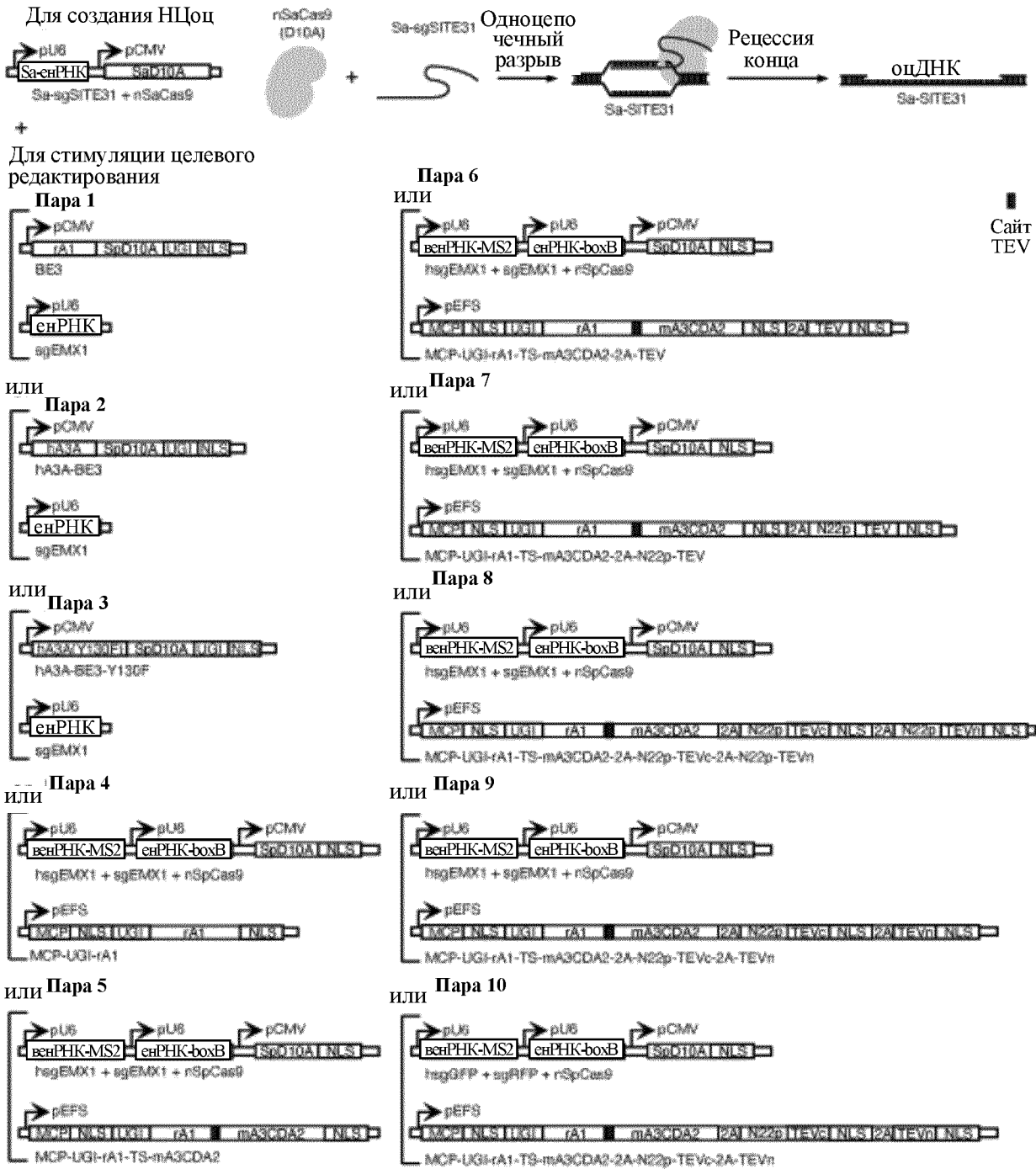


Фигура 25

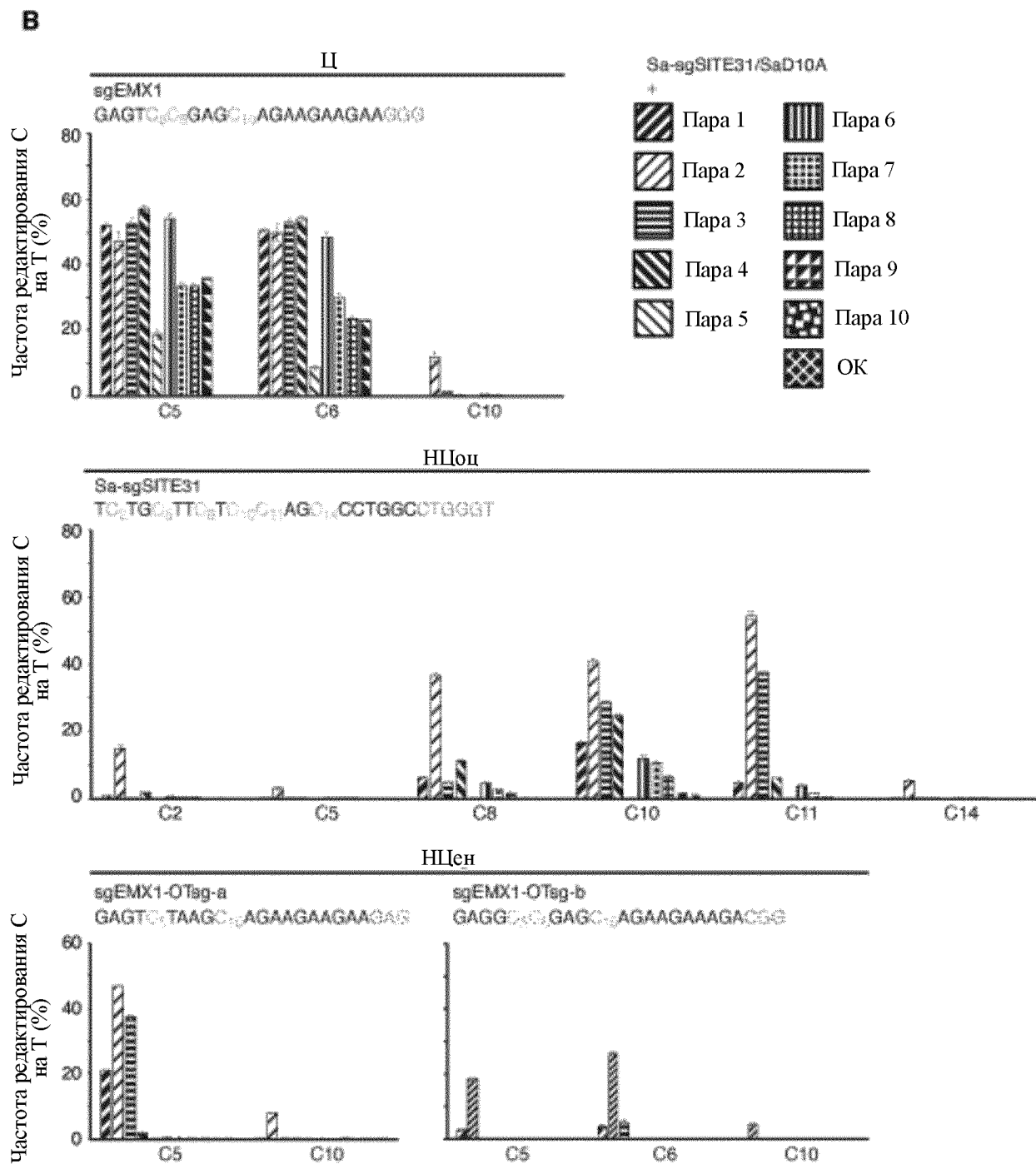


Фигура 26А

A

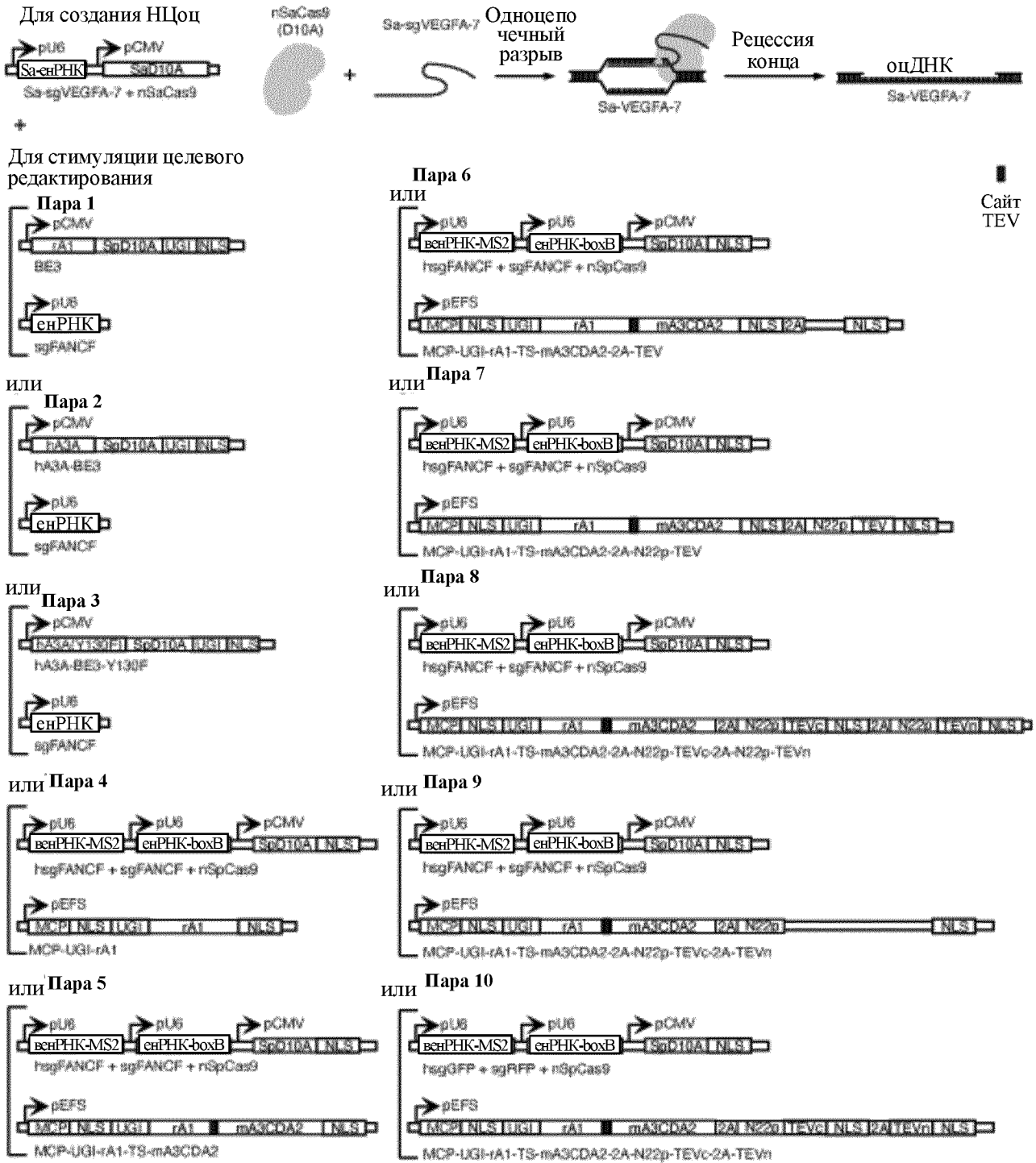


Фигура 26В

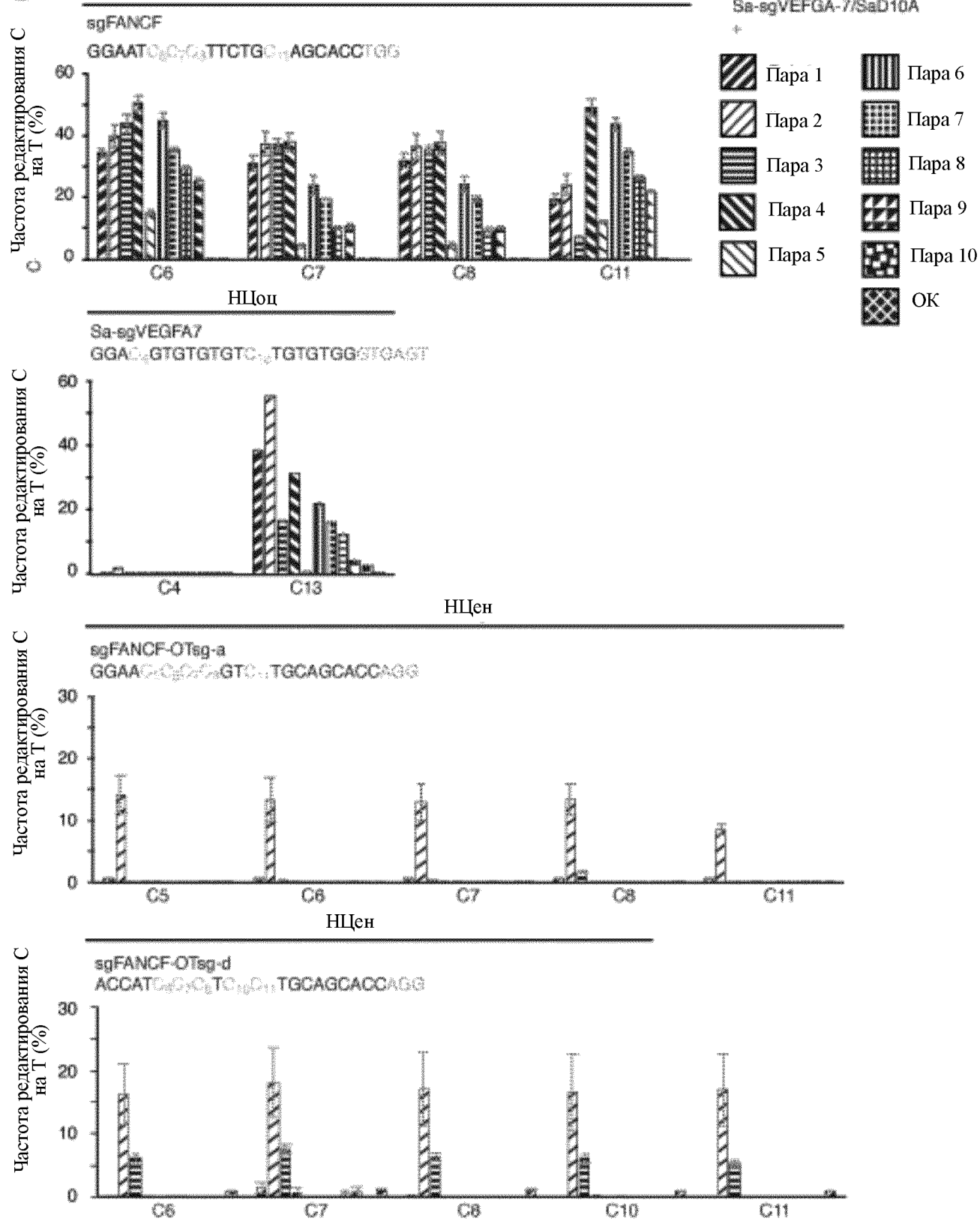


Фигура 27А

А



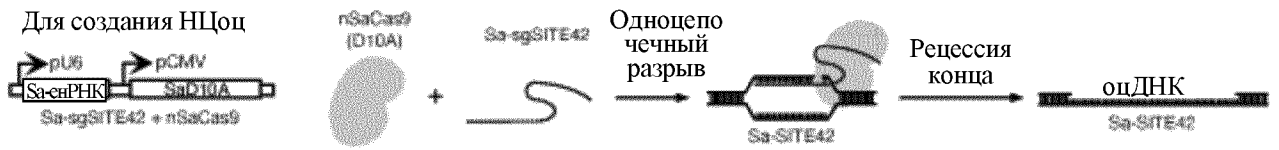
B



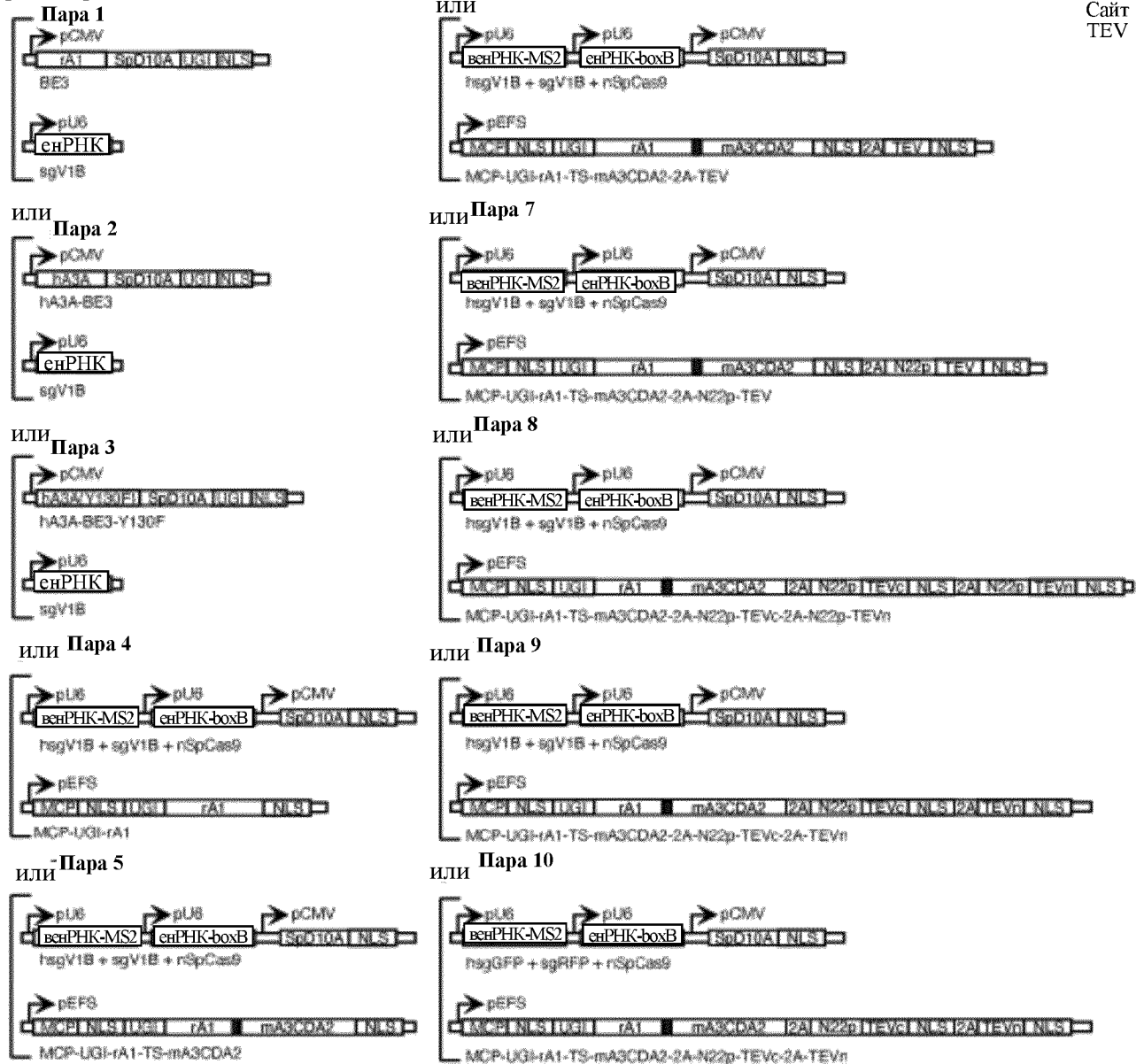
Фигура 28А

A

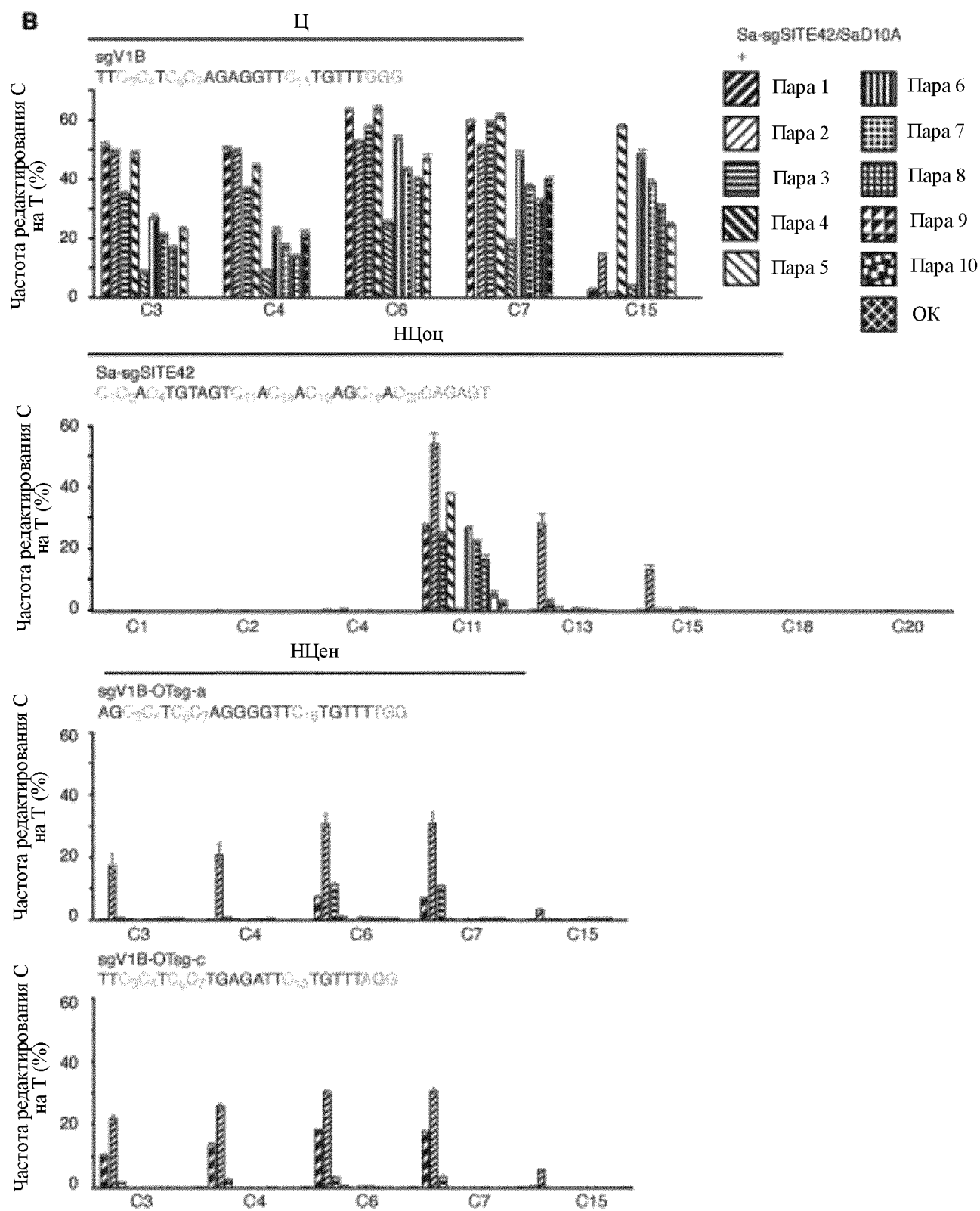
Для создания НЦоц



Для стимуляции целевого редактирования



Фигура 28В



Фигура 29

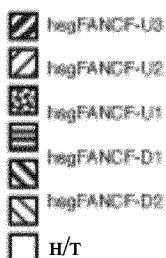
A



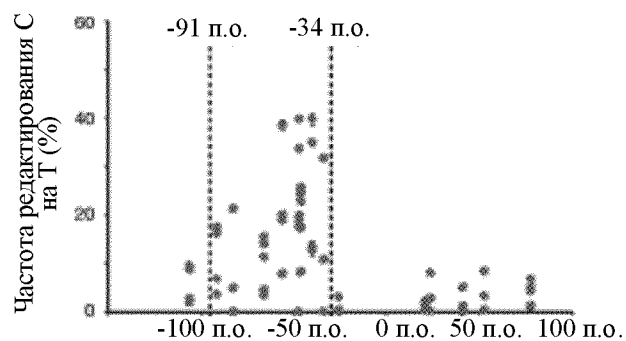
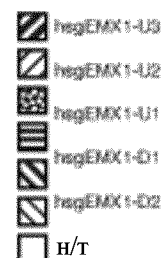
Частота редактирования С
на Т (%)



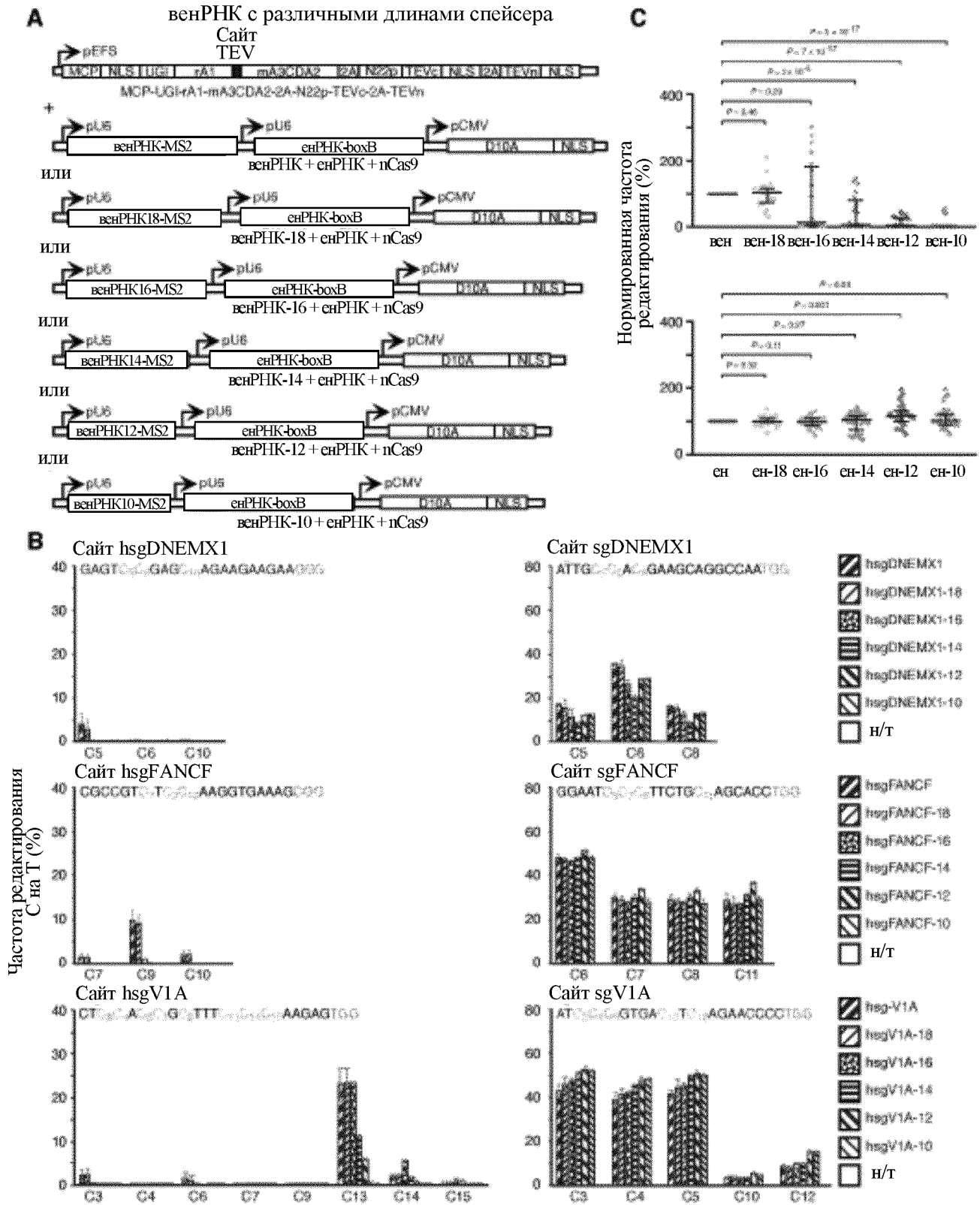
Частота редактирования С
на Т (%)



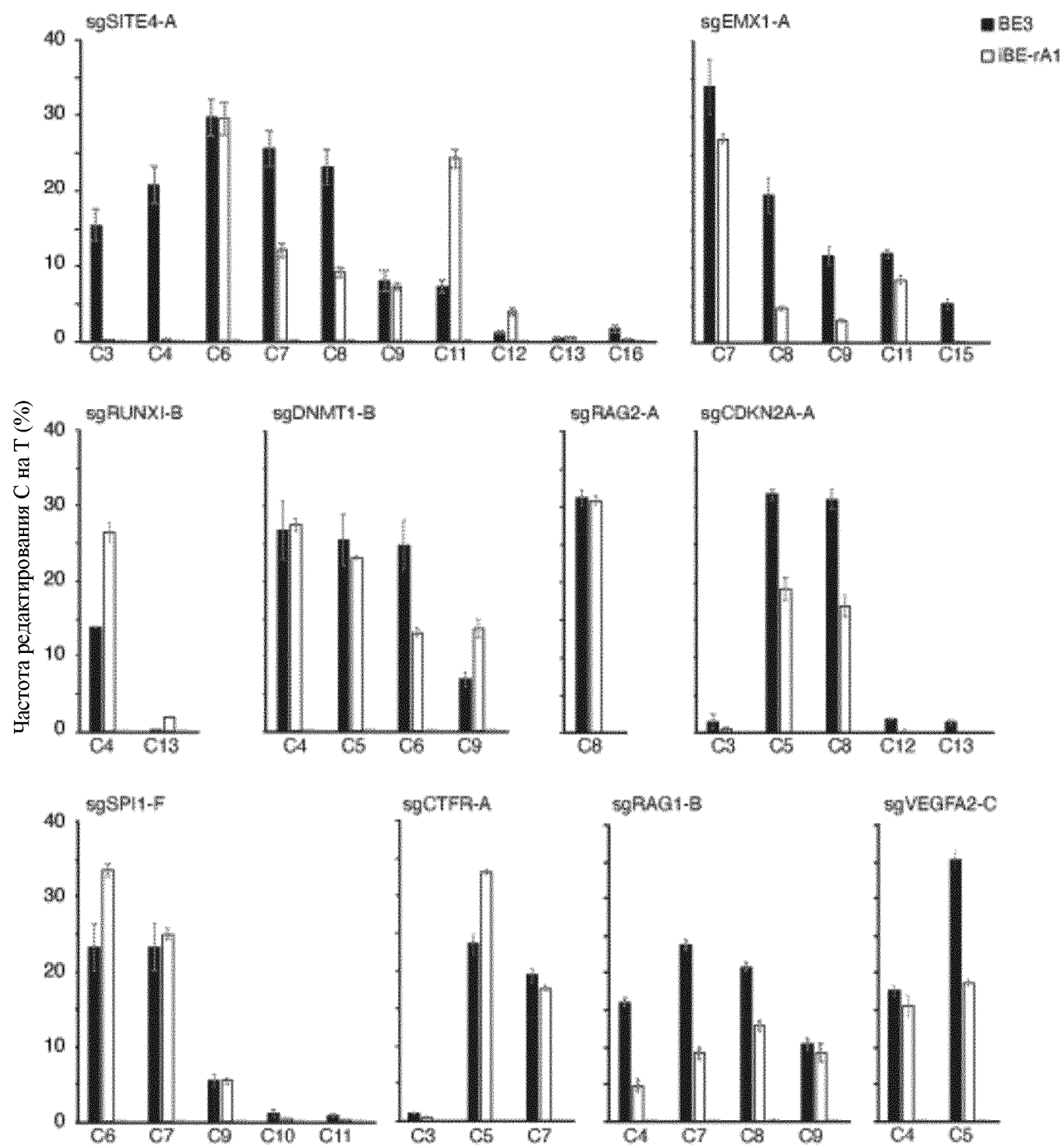
Частота редактирования С
на Т (%)



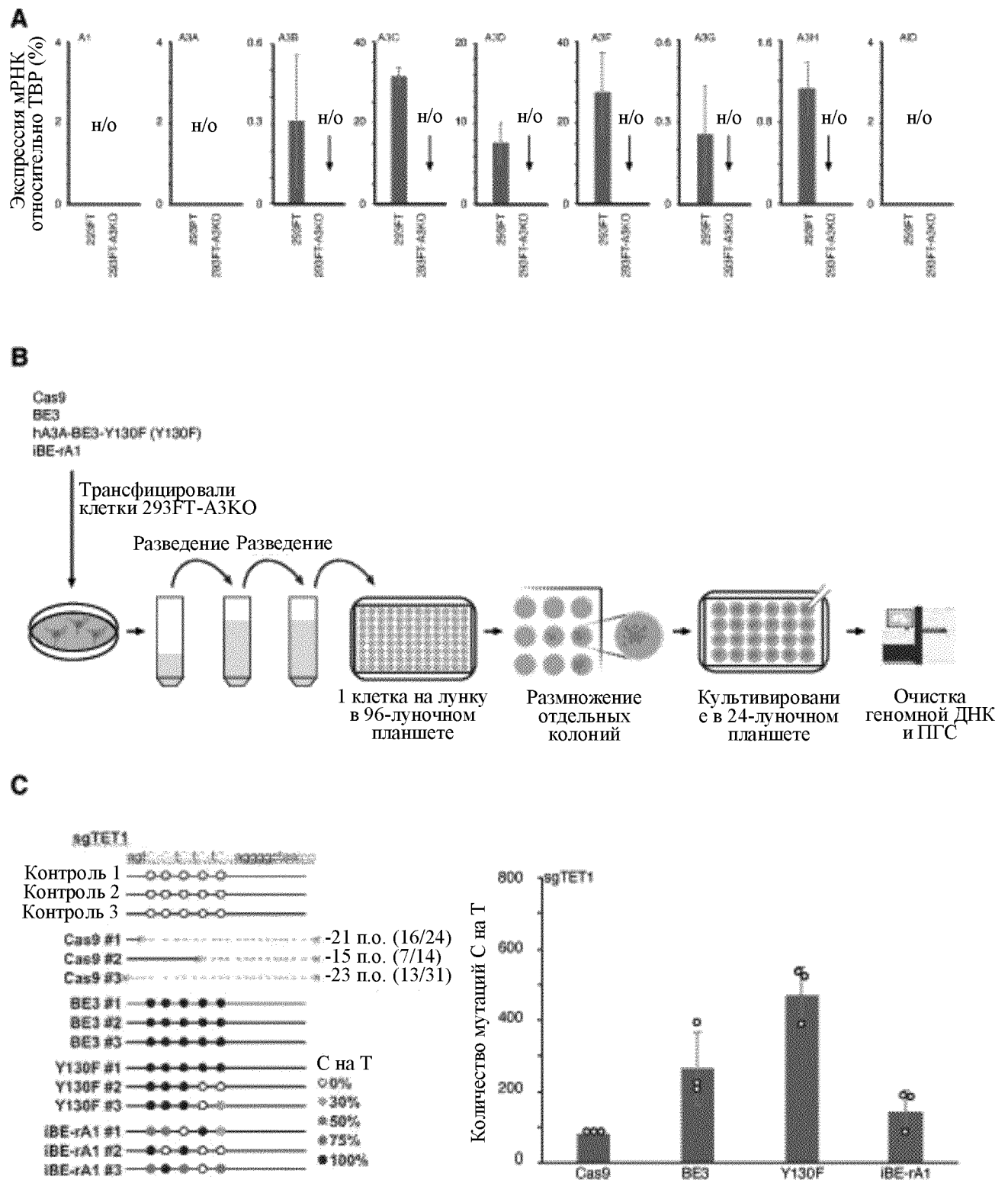
Фигура 30



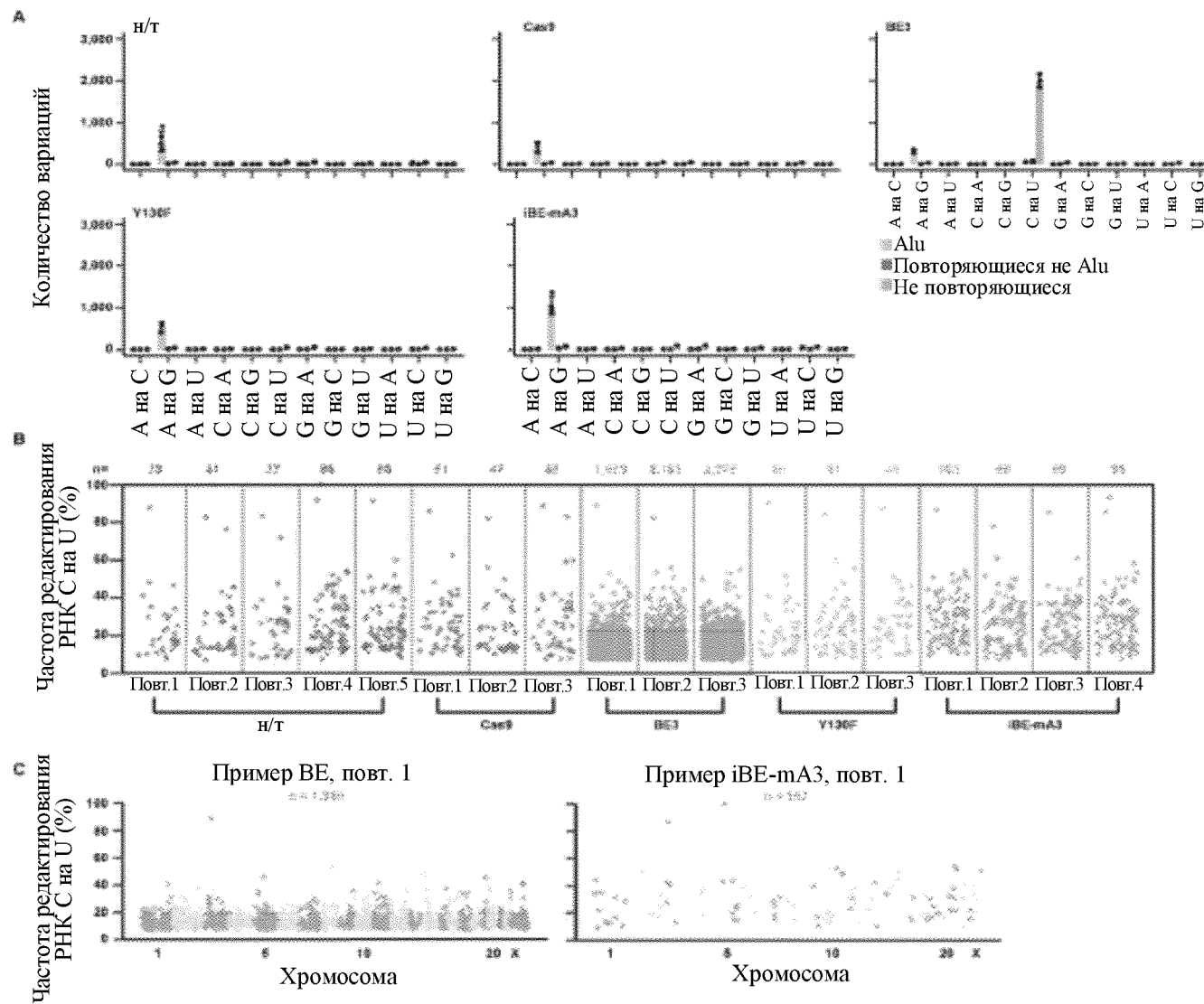
Фигура 31



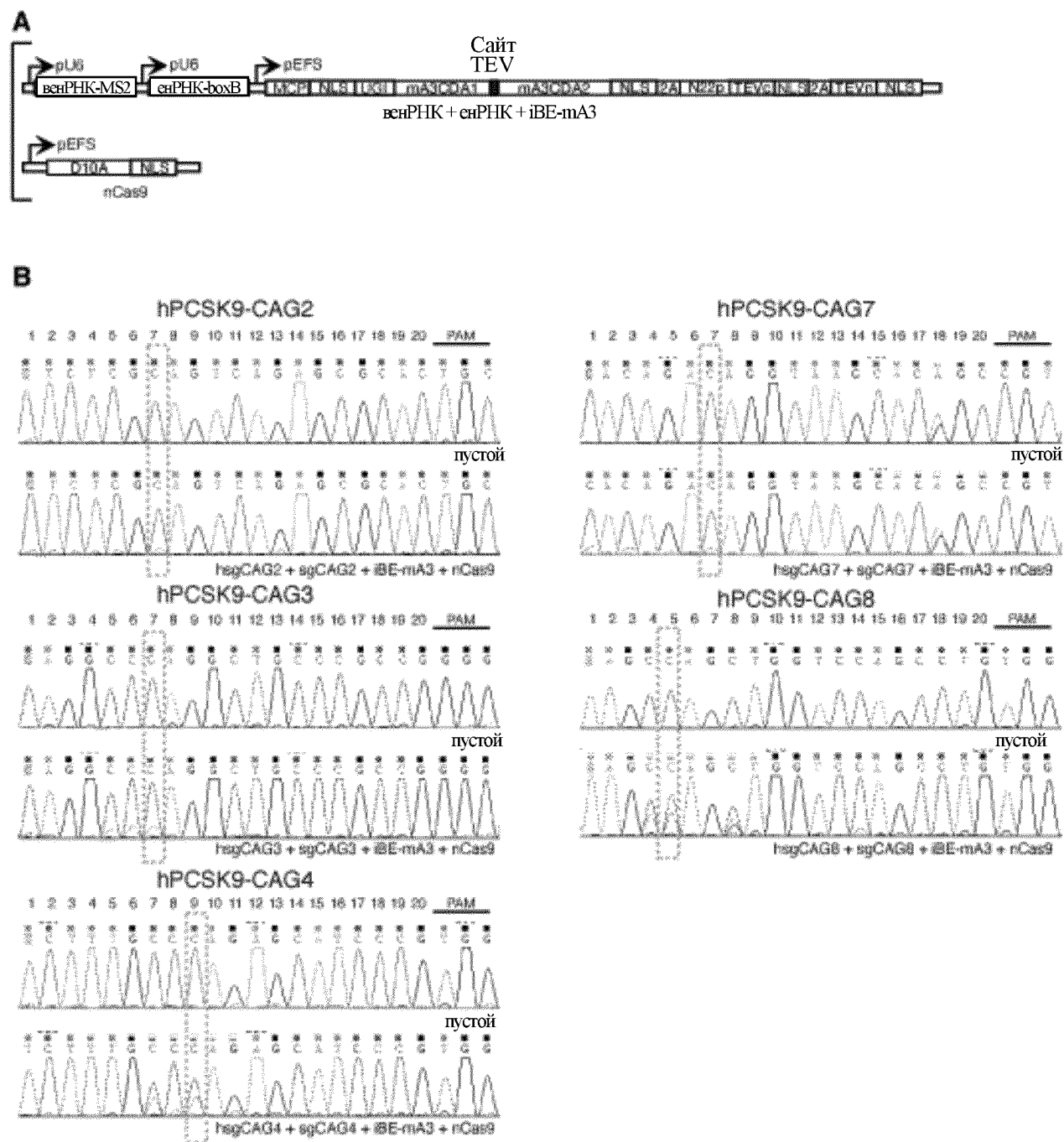
Фигура 32



Фигура 33

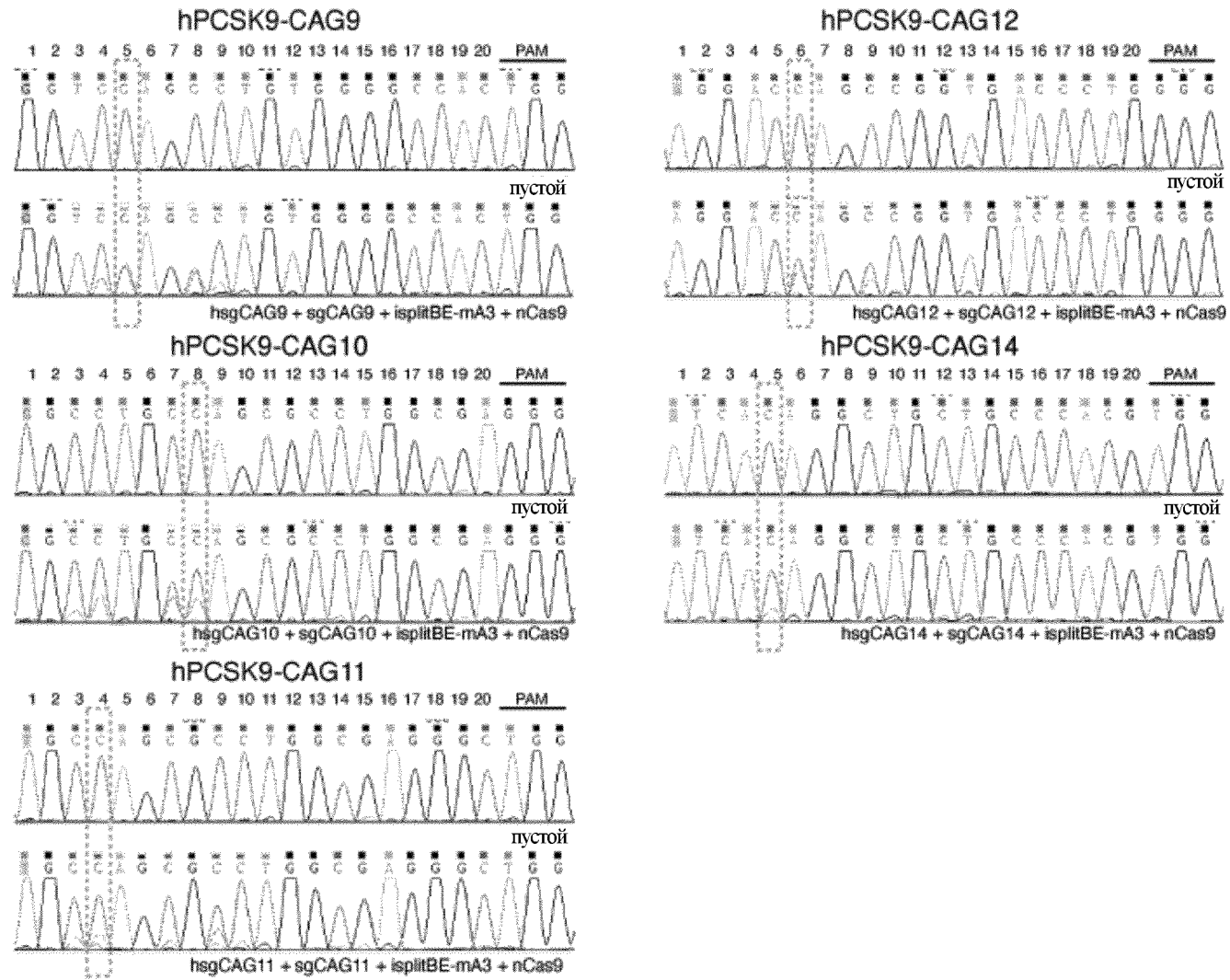


Фигура 34А-В



Фигура 34С

С



Фигура 34D

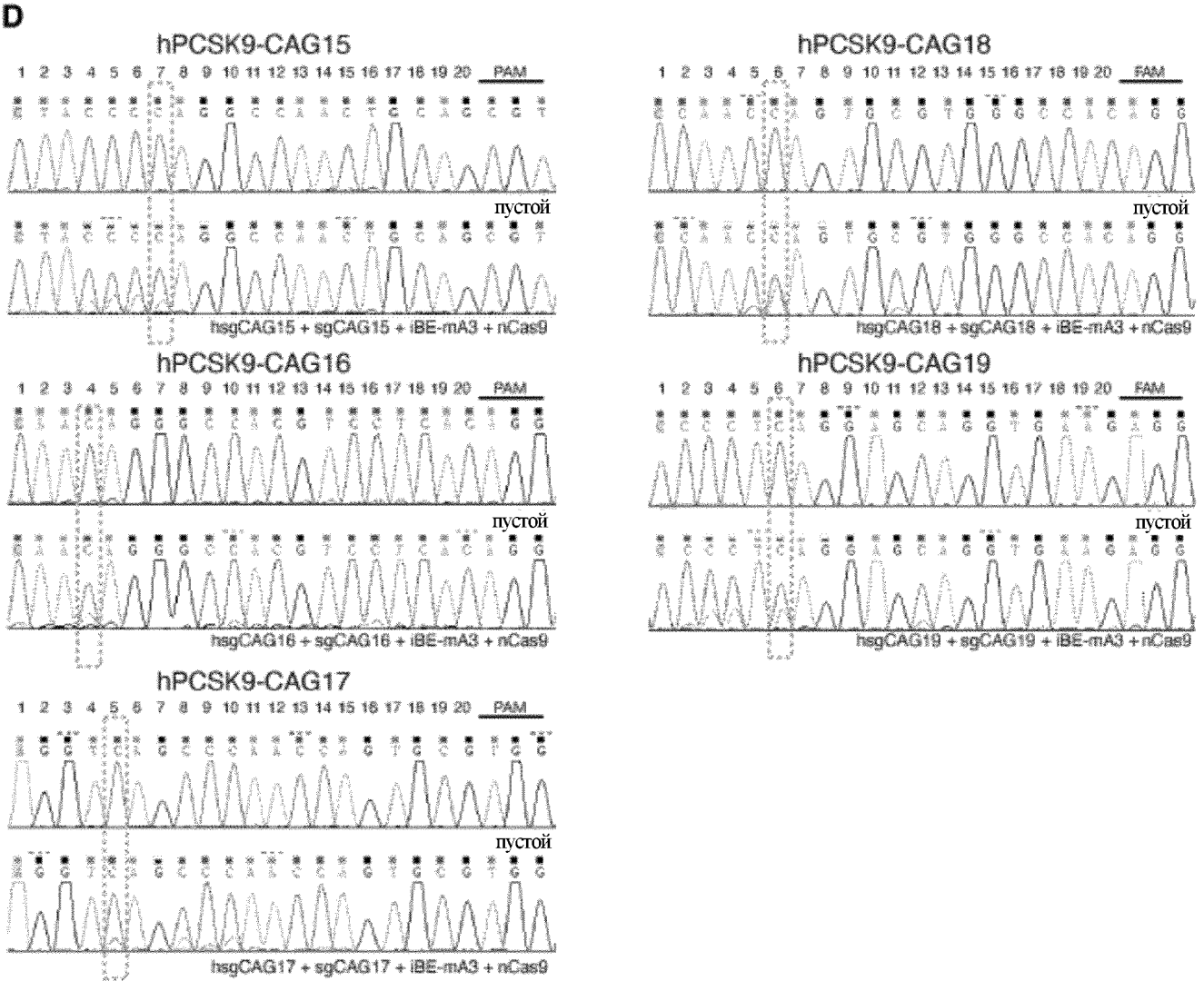




Diagram illustrating the structure of the ABEMax expression vector. The vector contains a pCMV promoter driving the expression of a gene. The gene is flanked by NLS (Nuclear Localization Signal) and a stop codon (X). The gene is composed of two exons, each flanked by a loxP site (L). The gene is flanked by a stop codon (X) and a loxP site (L). The vector is labeled ABEMax.

Schematic diagram of the mASCD A2-ABEMax expression vector. The vector contains a pCMV promoter, followed by the mASCD A2 gene, a nuclear localization signal (NLS), a tacA promoter, a tacA gene, an rCas9 gene, and another NLS. The entire construct is labeled mASCD A2-ABEMax.

pCMV

mΔ3CDA2 2A INLS YcdA YcdAm mCag9 INLS

mΔ3CDA2-2A-ABEMax

ANF2

Figure 1 displays four Sanger sequencing chromatograms of the 3' region of the hMLH1 gene, showing the effect of a frameshift mutation (Δ3CDA2) on the stop codon. The chromatograms are labeled as follows:

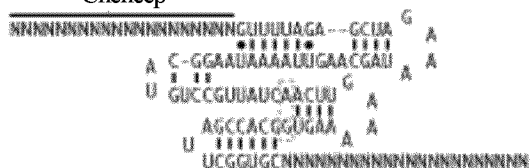
- пустой** (Empty): Shows the wild-type sequence with a stop codon (TGA) at position 20.
- ABE1Δx**: Shows the wild-type sequence with a stop codon (TGA) at position 20.
- mΔ3CDA2-ABE1Δx**: Shows the mutant sequence with a frameshift mutation (Δ3CDA2) that shifts the stop codon to position 21.
- mΔ3CDA2-2A-ABE1Δx**: Shows the mutant sequence with a frameshift mutation (Δ3CDA2) that shifts the stop codon to position 21.

The chromatograms show the sequence of the 3' region of the hMLH1 gene, with positions 1 through 20 indicated above the peaks. The stop codon (TGA) is highlighted in the original image.

FANCF

при РНК

Спейсер



Матрица ОТ

Сам
связывания
праймера

упрнРНК

Спейсер



Матрица ОТ

Сайт
связывания
праймера

The diagram shows a linear representation of the pCMV-OTax1-PE2 expression vector. From left to right, the components are: a pCMV promoter (indicated by a curved arrow), an NLS (Nuclear Localization Signal) in a box, a pA (polyA) signal, an O-Tax1 gene in a box, another NLS in a box, and a PE2 gene in a box. The entire construct is flanked by restriction enzyme sites (indicated by small vertical lines).

Одноцепоч.разрыв
sgSITE3

+

при РНК



ИЛИ

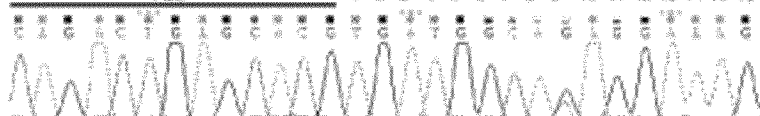
упрнРНК



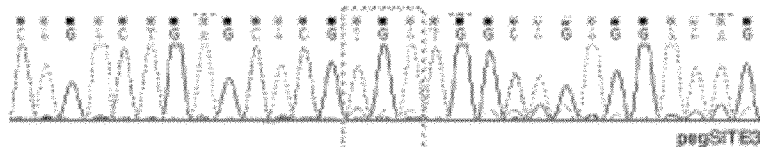
SITE3

FBS

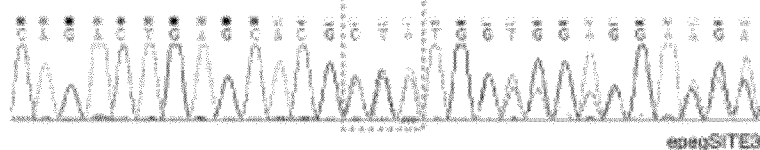
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16



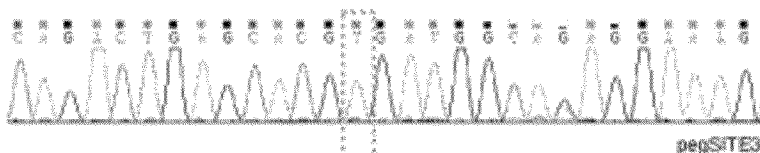
Пустой



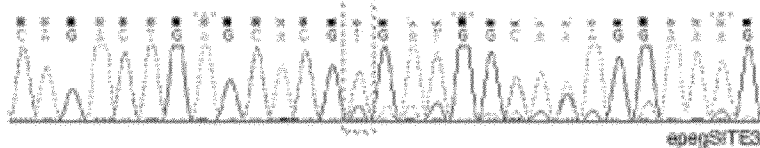
Вставка СТТ



Вставка СТТ

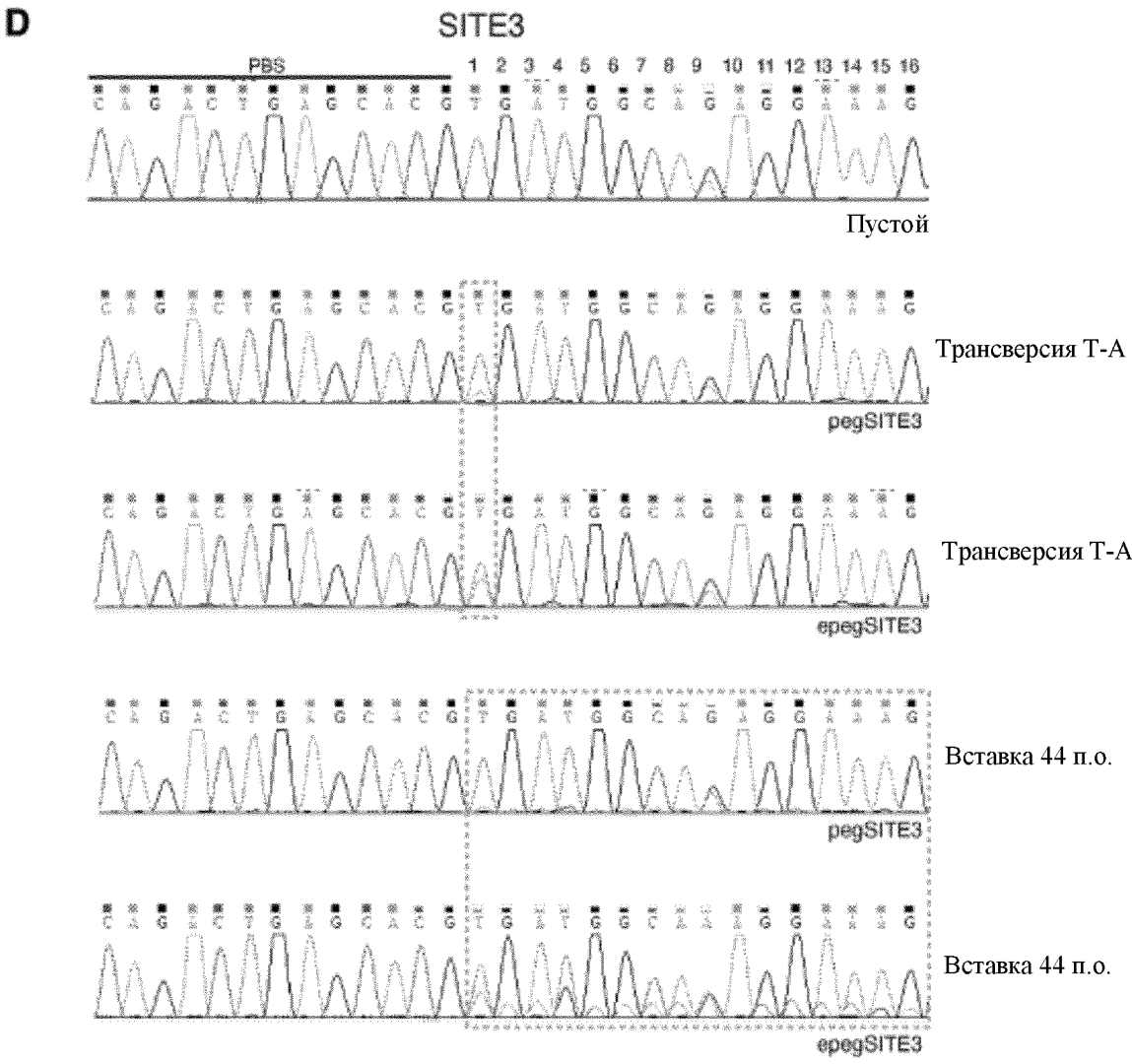


Делеция Т



Делеция Т

Фигура 36D

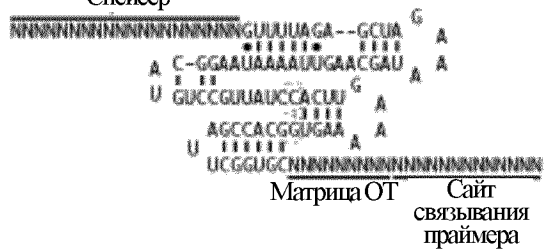


F

Спейсер



Спейсер



прнРНК



Одноцепоч.разрыв
sgSITE3

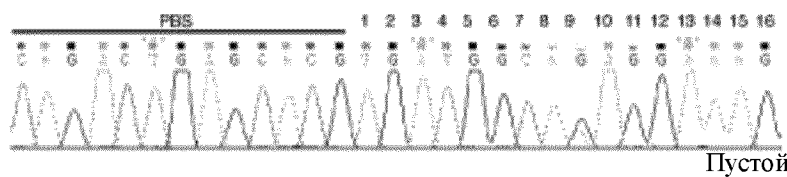
+



ИЛИ



SITE3

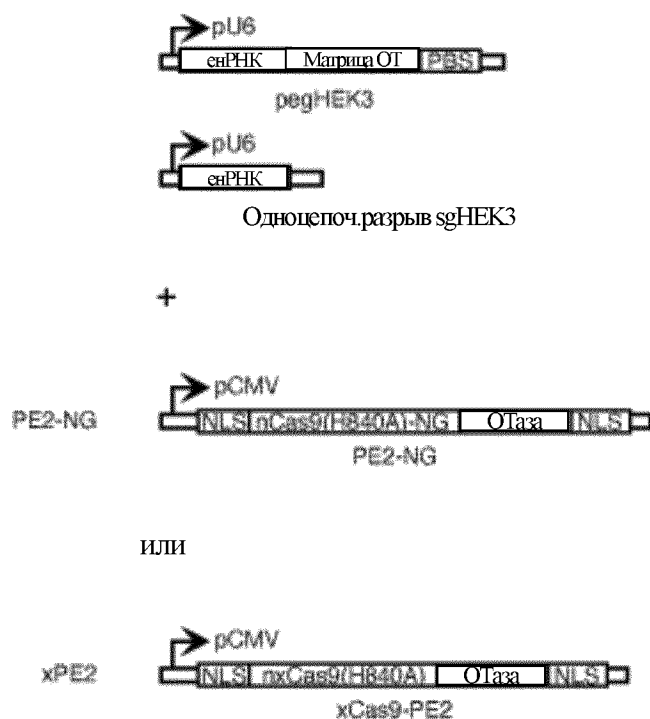


Вставка СТТ

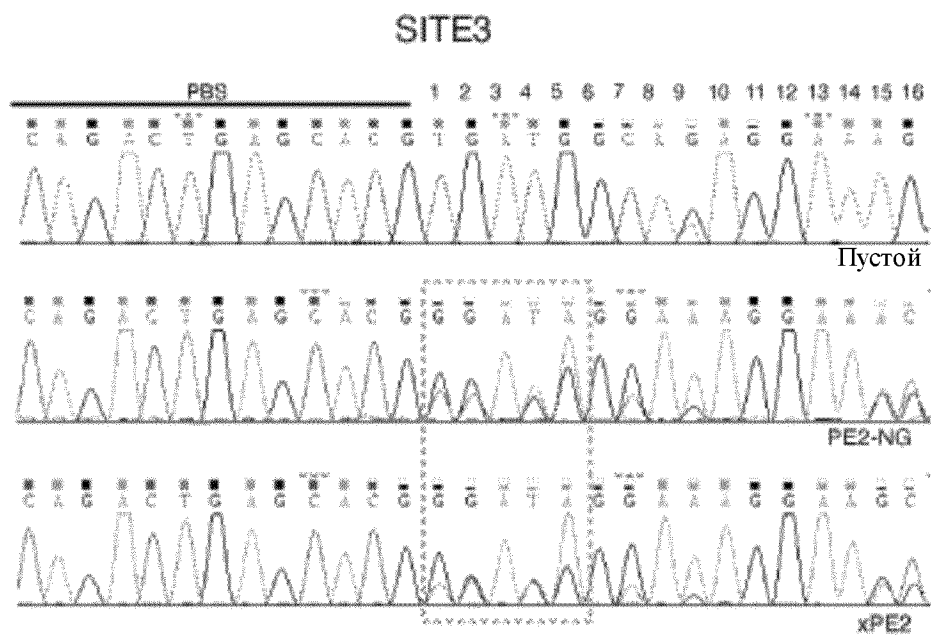
Вставка СТТ

Фигура 37

A



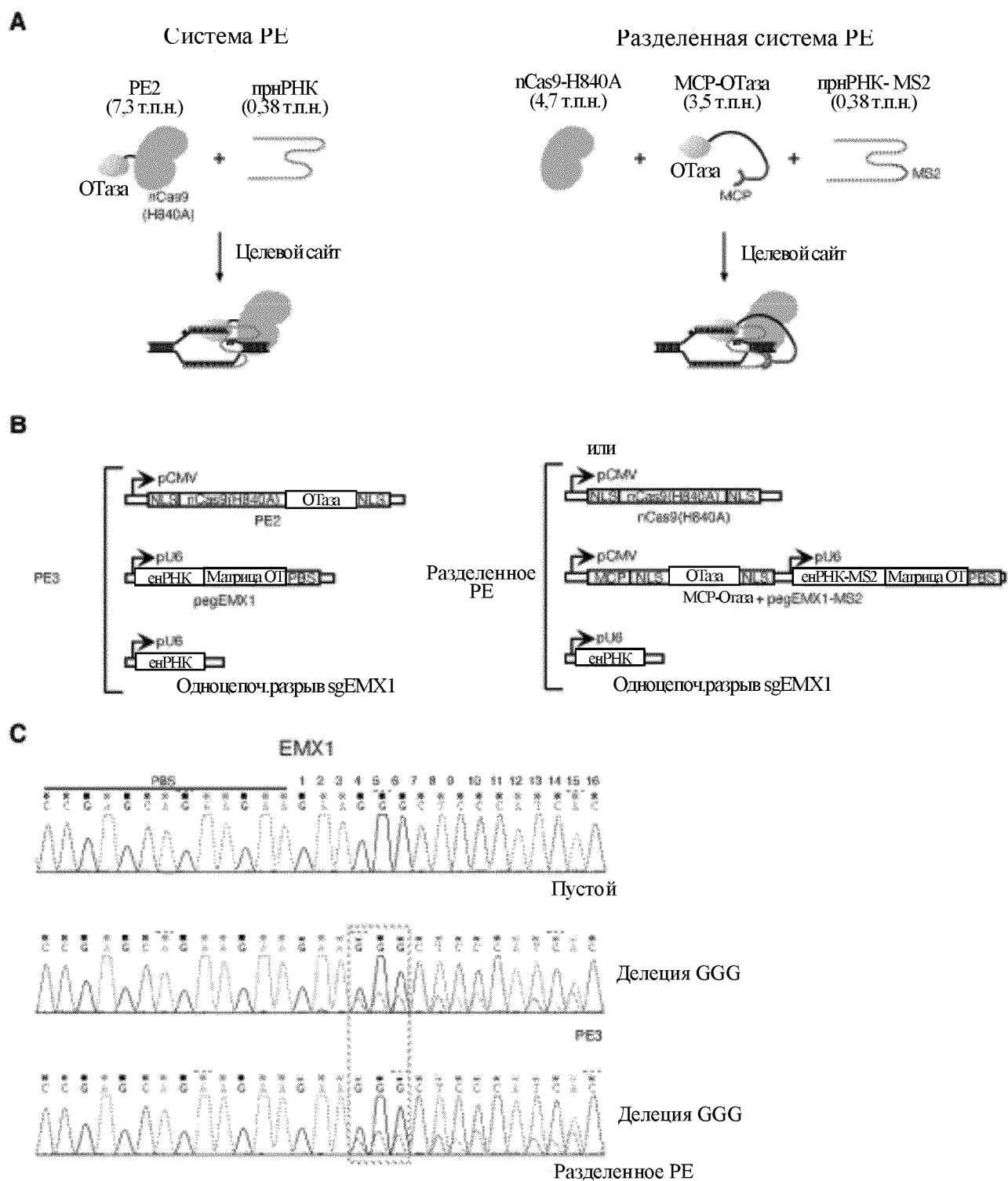
B



Делеция TGATG

Делеция TGATG

Фигура 38



Фигура 39А

Часть 1

	(1)	1	10	20
<i>Mus musculus</i> A3(282-355)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Mus spicilegius</i> A3(248-321)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Oricetulus longicaudatus</i> A3(249-322)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Mus tataricolor</i> A3(248-321)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Mus caroli</i> A3(260-333)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Mus pahari</i> A3(263-336)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Mus shortridgei</i> A3(233-306)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Mus setulosus</i> A3(229-302)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Grammomys surdaster</i> A3(270-344)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Rattus norvegicus</i> A3(256-328)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Mastomys coucha</i> A3(258-331)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Oricetulus griseus</i> A3B(235-307)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Peromyscus leucopus</i> A3(266-338)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Mesocricetus auratus</i> A3(258-340)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Microtus ochrogaster</i> A3B(266-338)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Nannospalax galii</i> A3(231-302)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Meriones unguiculatus</i> A3(233-305)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Dipodomys ordii</i> A3(256-330)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Jaquius jaculus</i> A3(303-374)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Chinchilla lanigera</i> A3H(86-161)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Heterocephalus glaber</i> A3(277-350)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Octodon degus</i> A3(256-329)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Uroditellus parryi</i> A3(256-330)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Actus nancymae</i> A3H(75-146)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Cebus capucinus imitator</i> A3H(55-126)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Saimiri boliviensis boliviensis</i> A3H(56-125)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Homo sapiens</i> A3H(49-123)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Homo sapiens</i> ARP10(48-123)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Pen paniscus</i> A3H(49-123)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Symphalangus syndactylus</i> A3H(49-123)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Macaca mulatta</i> A3H(49-123)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Theropithecus gelada</i> A3H(54-126)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Mandrillus leucophaeus</i> A3H(49-123)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Bos grunniens</i> A3(74-148)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Bubalus bubalis</i> A3(74-148)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Odocoileus virginianus texanus</i> A3H(209-283)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Sus scrofa</i> A3(51-125)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Ceratotherium maximosimum</i> A3B(232-306)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Equus caballus</i> A3H(79-153)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Enhydra lutris kanyoni</i> A3B(243-316)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Leptonychotes weddellii</i> A3H(50-123)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Ursus arctos horribilis</i> A3F(552-626)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Panthera leo bleyenberghi</i> A3H(50-124)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Panthera tigris sumatrae</i> A3H(50-124)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
<i>Tupaia belangeri</i> A3(46-120)	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	
Консенсус	(1)	SEKGGKQHAIEILFLDKIRSME---	LSQ	

Фигура 39В

Часть 2

	(27)	27	40	52
<i>Mus musculus</i> A3(282-355) (24)	VTITCYLTWS	PCPNC	AWOLAA	FKRDR
<i>Mus spicilegus</i> A3(248-321) (24)	VTITCYLTWS	PCPNC	AWOLAA	FKRDR
<i>Crictulus longicaudatus</i> A3(249-322) (24)	VTITCYLTWS	PCPNC	AWRLAA	FKRDR
<i>Mus terricolor</i> A3(248-321) (24)	VTITCYLTWS	PCPNC	AWOLAA	FKRDR
<i>Mus caroli</i> A3(260-333) (24)	VTITCYLTWS	PCPNC	AWOLAA	FKRDR
<i>Mus pahari</i> A3(263-336) (24)	MRITCYLTWS	PCPNC	AWOLAA	FOKDR
<i>Mus shortridgei</i> A3(233-306) (24)	MRITCYLTWS	PCPNC	AWOLAA	FOKDR
<i>Mus setulosus</i> A3(229-302) (24)	VRITCYLTWS	PCPNC	AWOLET	FKRDR
<i>Grammomys surdaster</i> A3(276-344) (25)	VRITCYLTWS	PCPNC	AROLAA	FKRDR
<i>Rattus norvegicus</i> A3(256-328) (23)	VRITCYLTWS	PCPNC	AROLAA	FKRDR
<i>Mastomys coucha</i> A3(258-331) (24)	VRITCYLTWS	PCPNC	AWOLAA	FKRDR
<i>Crictulus griseus</i> A3B(235-307) (23)	VOITCYLTWS	PCPNC	AOELAA	FKSDH
<i>Peromyscus leucopus</i> A3(266-338) (23)	ARITCYLTWS	PCPNC	AQKLA	AFKDR
<i>Mesocricetus auratus</i> A3(268-340) (23)	VRITCYLTWS	PCPNC	AOELAA	FKRDR
<i>Microtus odyrogaster</i> A3B(266-338) (23)	ERITCYLTWS	PCPNC	AOELAA	FKRDR
<i>Nannospalax galii</i> A3(231-302) (23)	VOITCYLTWS	PCPNC	AOELAA	FKRDR
<i>Meriones unguiculatus</i> A3(233-305) (23)	AOITCYLTWS	PCPNC	AOELAT	FKRDR
<i>Dipodomys ordi</i> A3(256-330) (26)	YOITCYLTWS	PCLDC	AFKLA	KLKDR
<i>Jaquius jaculus</i> A3(303-374) (23)	ALITCYLTWS	PCLDC	SOHLA	ALKDR
<i>Chinchilla lanigera</i> A3H(86-161) (27)	YOITCFLTWS	PCPSC	AOELAS	FKRAH
<i>Heterocephalus glaber</i> A3(277-350) (25)	YOVTICFLTWS	PCPSC	AOELV	SFKRAH
<i>Octodon degus</i> A3(256-329) (25)	YOITCFLTWS	PCPSC	AOELAS	FKSTH
<i>Uroctellus parryi</i> A3(256-330) (26)	YEVTICYLTWS	PCPDC	AOELV	ALTRSH
<i>Actus nancymae</i> A3H(75-146) (23)	YEVTICYLTWS	PCPSC	AQKLA	AFTKAQ
<i>Cebus capucinus imitator</i> A3H(55-126) (23)	YEVTICYLTWS	PCPSC	AOELV	AFKAQ
<i>Saimiri boliviensis boliviensis</i> A3H(56-125) (21)	YDVTCYLTWS	PCPSC	AOELAA	FAKAQ
<i>Homo sapiens</i> A3H(49-123) (26)	YOVTICYLTWS	PCSSC	AWELV	DFIKAH
<i>Homo sapiens</i> ARP10(48-123) (27)	YOVTICYLTWS	PCSSC	AWELV	DFIKAH
<i>Pen paniscus</i> A3H(49-123) (26)	YOVTICYLTWS	PCSSC	AWKLVD	DFIOAH
<i>Symphalangus syndactylus</i> A3H(49-123) (26)	YOVTICYLTWS	PCPSC	AWELV	DFIKAH
<i>Macaca mulatta</i> A3H(49-123) (26)	YOVTICYLTWS	PCPSC	AGELV	DFIKAH
<i>Theropithecus gelade</i> A3H(54-128) (26)	YOVTICYLTWS	PCPSC	AGELV	DFIKAH
<i>Mandrillus leucophaeus</i> A3H(49-123) (26)	YOVTICYLTWS	PCPSC	ARELV	DFIKAH
<i>Bos grunniens</i> A3(74-148) (26)	YKIIICYITWS	PCPNC	ANELV	NFEITN
<i>Bubalus bubalis</i> A3(74-148) (26)	YKIIICYITWS	PCPNC	ASELV	NFEITN
<i>Odocoileus virginianus texanus</i> A3H(209-283) (26)	YKIIICYITWS	PCPNC	ASELV	NFEITN
<i>Sus scrofa</i> A3(51-125) (26)	YRIICYVTWS	PCPNC	AKELV	DFISER
<i>Caprotherium sinum sinum</i> A3B(232-306) (26)	YEITCYITWS	PCPNC	ALELV	AFTIRDY
<i>Equus caballus</i> A3H(79-153) (26)	YEITCYVTWS	PCATC	ACKLIK	FIRKE
<i>Shihya lutris kanyoni</i> A3B(243-316) (25)	FEITCYLTWS	PCPNC	AKELV	AFVDRH
<i>Leptonychotes weddellii</i> A3H(50-123) (25)	FEITCYVTWS	PCPNC	AKELV	AFVDRH
<i>Ursus arctos horribilis</i> A3F(552-626) (26)	FEITCYVTWS	PCPNC	AEELV	AFVDRH
<i>Panthera leo bleyenberghi</i> A3H(50-124) (26)	FEIICYITWS	PCPNC	AEELV	AFVDRH
<i>Panthera tigris sumatrae</i> A3H(50-124) (26)	FEIICYITWS	PCPNC	AEELV	AFVDRH
<i>Tupaia belangeri</i> A3(46-120) (26)	YOLTCYLTWS	PCPSC	AOELV	TFMAES
Консенсус (27)	YITCYLTWS	PCP	CAELAA	FKDR

Фигура 39С

Часть 3

	(53)	53	50	77	SEQ ID NO:
<i>Mus musculus</i> A3(282-355) (50)		PDLILHIYTSRLYFHWKRPFQKGLC			48
<i>Mus spicilegus</i> A3(248-321) (50)		PDLIPBIYTSRLYFHWKRPFQKGLC			49
<i>Crictulus longicaudatus</i> A3(249-322) (50)		PDLILHIYTSRLYFHWKRPFQKGLC			50
<i>Mus ferricolor</i> A3(248-321) (50)		PDLILBIYTSRLYFHWKRPFQKGLC			51
<i>Mus caroli</i> A3(260-333) (50)		PDLILHIYTSRLYFHWKRPFQKGLC			52
<i>Mus pahari</i> A3(263-336) (50)		PDLILBIYTSRLYFHWRRIFQKGLC			53
<i>Mus shortridgei</i> A3(233-306) (50)		PDLILHIYTSRLYFHWRRIFQKGLC			54
<i>Mus setulosus</i> A3(229-302) (50)		PDLILHIYTSRLYFHWKRAFQEGLC			55
<i>Grammomys surdaster</i> A3(270-344) (51)		PGLILRIYTSRLYFYWRRKFQKGLC			56
<i>Rattus norvegicus</i> A3(256-328) (49)		PDLILRIYTSRLYFYWRKKFQKGLC			57
<i>Nastomys couchi</i> A3(258-331) (50)		PDLILRIYASRLYFHWRRAFQKGLC			58
<i>Crictulus griseus</i> A3B(235-307) (49)		PDLVLRIYTSRLYFHWRRKYQEGLC			59
<i>Peromyscus leucopus</i> A3(266-338) (49)		PDLVLEVYTSRLYFHWRRKYQEGLC			60
<i>Neosciurus auratus</i> A3(268-340) (49)		PDLVLRIYTSRLYFHWRRKYQEGLC			61
<i>Microtus ochrogaster</i> A3B(266-338) (49)		PGLVLRIYASRLYFHWRRKYQEGLC			62
<i>Nannospalax galii</i> A3(231-302) (49)		PDLVLRIYASRLYFHWKRKFQKGL-			63
<i>Meriones unguiculatus</i> A3(233-305) (49)		PDLVLEVYASRLYFHWRRKYQEGLC			64
<i>Dipodomys ordii</i> A3(256-330) (52)		PDLTLRIYTSRLYFHWIRKFQKGL-			65
<i>Jaquius jaculus</i> A3(303-374) (49)		PGLTLRIYTSRLYFHWVKKFQEGLC			66
<i>Chinchilla lanigera</i> A3H(86-161) (53)		PHLRFOIFVSRLYFHWKRISYQAGL-			67
<i>Heterocephalus glaber</i> A3(277-350) (51)		PHLRLOIFARLFFHWRRSYQEGLC			68
<i>Octodon degus</i> A3(256-329) (51)		PRVHLOIFVSRLYFHWKRISYQEGLC			69
<i>Urodellus parryi</i> A3(256-330) (52)		PHVRLRLFTSRLYFHWTFWSTQEGLC			70
<i>Actus nancymazei</i> A3H(75-146) (49)		VHLNLRIFASRLYTHWRRSYQKGL-			71
<i>Cebus capucinus imitator</i> A3H(55-126) (49)		DHLNLRIFASRLYTHWRRRYKEGL-			72
<i>Saimiri boliviensis boliviensis</i> A3H(56-125) (47)		DHLNLRIFASRLYTHWRRSYQKGL-			73
<i>Homo sapiens</i> A3H(49-123) (52)		DHLNLGIFASRLYTHWCKPQKGL-			74
<i>Homo sapiens</i> ARP10(49-123) (53)		DHLNLGIFASRLYTHWCKPQKGL-			75
<i>Pan paniscus</i> A3H(49-123) (52)		DHLNLRIFASRLYTHWCKPQKGL-			76
<i>Symphalangus syndactylus</i> A3H(49-123) (52)		DHLNLGIFASRLYTHWCHROEGLC			77
<i>Macaca mulatta</i> A3H(49-123) (52)		RHLNLRIFASRLYTHWRPHYQEGLC			78
<i>Theropithecus galadei</i> A3H(54-128) (52)		HHLNLRIFASRLYTHWRPHYQEGLC			79
<i>Mandrillus leucophaeus</i> A3H(49-123) (52)		RHLNLRIFASRLYTHWRPHYQEGLC			80
<i>Bos grunniens</i> A3(74-148) (52)		NHLNLEIFASRLYTHWIKPFKMGL-			81
<i>Bubalus bubalis</i> A3(74-148) (52)		DHLNLOIFASRLYTHWIKPFKRGGL-			82
<i>Odocoileus virginianus texanus</i> A3H(209-283) (52)		DHLNLOIFASRLYTHWENPFQKGL-			83
<i>Sus scrofa</i> A3(51-125) (52)		HHLNLOIFASRLYTHWVRCYQKGL-			84
<i>Capreolus sinensis sinensis</i> A3B(232-306) (52)		PHLSLOIFASRLYTHWRRRSIQGL-			85
<i>Equus caballus</i> A3H(79-153) (52)		PHLSLRIEVSRLYTHWFRQKQGL-			86
<i>Enhydra lutris kenyoni</i> A3B(243-316) (51)		PHLSLELFASRLYTHWRWKYQEGLC			87
<i>Leptomys weddellii</i> A3H(50-123) (51)		RHLSLRLFASRLYTHWLRNKKGL-			88
<i>Ursus arctos horribilis</i> A3F(552-626) (52)		PHVRLRLFASRLYTHWLRNKKYQEGLC			89
<i>Panthera leo bleyenberghi</i> A3H(50-124) (52)		PHLSLRIFASRLYTHWRRWKYQKGL-			90
<i>Panthera tigris sumatrae</i> A3H(50-124) (52)		PHLSLRIFASRLYTHWRRWKYQKGL-			91
<i>Tupaia belangeri</i> A3(46-120) (52)		RHLNLQVFSRLYTHWDRDFQKGL-			92
Консенсус (53)		PHL LRIFASRLYFHWRR FQ GL			

Фигура 40А

Часть 1

	(1)	1	10	20	30	41
Homo sapiens A3B(29-138)	(1)	GRSYTNLCYEYVAIRKGRSNLLWDTGVFRGQV	-----			
Gorilla A3B(29-138)	(1)	GRSYTNLCYEYVAIRKGRSNLLWNTGVFRGQM	-----			
Pan paniscus A3B(29-138)	(1)	GRSYTNLCYEYVAIRKGRSNLLWDTGVFRGQM	-----			
Pan troglodytes A3B(29-138)	(1)	GRSYTNLCYEYVAIRKGRSNLLWDTGVFRGQM	-----			
Gorilla A3F(30-127)	(1)	-RNTVNLCEYEVATR-GESRPFLDAKIFRGQV	-----			
Pan troglodytes A3F(30-137)	(1)	-RNTVNLCEYEVATR-GESRPFLDTKIFRGQV	-----			
Homo sapiens A3F(30-137)	(1)	-RNTVNLCEYEVATR-GESRPFLDAKIFRGQV	-----			
Macaca leonina A3F(30-137)	(1)	-RNTVNLCEYEVATR-GESMPFWDTKIFRGQV	-----			
Macaca nemestrina A3F(30-137)	(1)	-RNTVNLCEYEVATR-GESMPFWDTKIFRGQV	-----			
Rhinopithecus roxellana A3F(30-137)	(1)	-RNTVNLCEYEVATR-GESMPFWDAKIFRGQV	-----			
Mandrillus leucophaeus A3F(30-130)	(1)	-RNTVNLCEYEVATR-GESMPFWDTKIFRGQV	-----			
Macaca mulatta A3F(30-137)	(1)	-RNTVNLCEYEVATR-GESMPFWDTKIFRGQV	-----			
Theropithecus gelada A3F(30-137)	(1)	-RNTVNLCEYEVATR-GESMPFWDTKIFRGQV	-----			
Cercocebus atys A3B(29-138)	(1)	GRSYTNLCYEYVAIRKDESKLPNDTGVFRGQV	-----			
Macaca fascicularis A3B(29-138)	(1)	GRSYTNLCYEYVAIRKDESKLPNDTGVFRGQV	-----			
Macaca mulatta A3B(29-138)	(1)	GRSYTNLCYEYVAIRKDESKLPNDTGVFRGQV	-----			
Macaca leonina A3B(29-138)	(1)	GRSYTNLCYEYVAIRKDESKLPNDTGVFRGQV	-----			
Mandrillus leucophaeus A3B(29-138)	(1)	GRSYTNLCYEYVAIRKDESKLPNDTGVFRGQV	-----			
Macaca nemestrina A3B(29-138)	(1)	GRSYTNLCYEYVAIRKDESKLPNDTGVFRGQV	-----			
Rhinopithecus bieti A3F(29-138)	(1)	GRSYTNLCYEYVAIRKDESKLPNDTGVFRGQV	-----			
Rhinopithecus roxellana A3B(29-138)	(1)	GRSYTNLCYEYVAIRKDESKLPNDTGVFRGQV	-----			
Chlorocebus sabaeus A3B(29-138)	(1)	GRSYTNLCYEYVAIRKDESKLPNDTGVFRGQM	-----			
Nomascus leucogenys A3B(30-138)	(1)	-RSTVNLCEYEVAKRKDESKLPNDTGVFRGQM	-----			
Cercocebus atys A3F(29-138)	(1)	GRSYTNLCYEYVAIRKDESKLPNDTGVFRGQV	-----			
Papio anubis A3F(29-138)	(1)	GRSYTNLCYEYVAIRKDESKLPNDTGVFRGQV	-----			
Chlorocebus aethiops A3D(29-150)	(1)	GRSYTNLCYEYVAIRKDESKLPNDTGVFRGQV	-----			
Chlorocebus sabaeus A3D(29-134)	(1)	GRSYTNLCYEYVAIRKDESKLPNDTGVFRGQV	-----			
Chlorocebus sabaeus A3F(29-150)	(1)	GRSYTNLCYEYVAIRKDESKLPNDTGVFRGQV	-----			
Erythrocebus patas A3D(29-150)	(1)	GRSYTNLCYEYVAIRKDESKLPNDTGVFRGQV	-----			
Macaca fascicularis A3D(29-159)	(1)	GRSYTNLCYEYVAIRKDESKLPNDTGVFRGQV	-----			
Macaca fascicularis A3F(29-138)	(1)	GRSYTNLCYEYVAIRKDESKLPNDTGVFRGQV	-----			
Macaca nemestrina A3D(29-138)	(1)	GRSYTNLCYEYVAIRKDESKLPNDTGVFRGQV	-----			
Macaca leonina A3D(29-138)	(1)	GRSYTNLCYEYVAIRKDESKLPNDTGVFRGQV	-----			
Macaca mulatta A3D(29-138)	(1)	GRSYTNLCYEYVAIRKDESKLPNDTGVFRGQV	-----			
Gorilla A3D(29-150)	(1)	GRSYTNLCYEYVAIRKGRSNLLWNTGVFRGQV	-----			
Pan paniscus A3D(29-150)	(1)	GRSYTNLCYEYVAIRKGRSNLLWNTGVFRGQV	-----			
Pan troglodytes A3D(29-150)	(1)	GRSYTNLCYEYVAIRKGRSNLLWNTGVFRGQV	-----			
Homo sapiens A3D(29-150)	(1)	GRSYTNLCYEYVAIRKGRSNLLWNTGVFRGQV	-----			
Nomascus leucogenys A3D(29-150)	(1)	GRSYTNLCYEYVAIRKDESKLPNDTGVFRGQV	-----			
Saimiri boliviensis A3C(29-138)	(1)	GKSYTNLCYEYVAIRKDESKLPNDTGVFRGQV	-----			
Saimiri boliviensis A3F(29-138)	(1)	GKSYTNLCYEYVAIRKDESKLPNDTGVFRGQV	-----			
Ptilocolobus tephrosceles A3F(36-145)	(1)	GRSYTNLCYEYVAIRKDESKLPNDTGVFRGQV	-----			
Colobus angolensis palliatus A3F(29-138)	(1)	GRSYTNLCYEYVAIRKDESKLPNDTGVFRGQV	-----			
Pongo abelii A3F(30-150)	(1)	-RSTVNLCEYEVAKRKDESKLPNDTGVFRGQV	-----			
Консенсус	(1)	GRSYTNLCYEYVAIRKDESKLPNDTGVFRGQV	-----			

Фигура 40В

Часть 2

	(42)	42	50	60	70	82
Homo sapiens A38(29-138)	(32)	-----	YFQPQYHAEMCFLSWFCGNQLPAYKCFQIT			
Gorilla A38(29-138)	(32)	-----	YSQPEHHAEMCFLSWFCGNQLPAYKCFQIT			
Pan paniscus A38(29-138)	(32)	-----	YSQPEHHAEMMYFLSWFCGNQLPAYKCFQIT			
Pan troglodytes A38(29-138)	(32)	-----	YSQPEHHAEMCFLSWFCGNQLPAYKCFQIT			
Gorilla A3F(30-127)	(30)	-----	YFQPQYHAEMCFLSWFCGNQLPAYKCFQIW			
Pan troglodytes A3F(30-137)	(30)	-----	YFQPQYHAEMCFLSWFCGNQLPAYKCFQIT			
Homo sapiens A3F(30-137)	(30)	-----	YSQPEHHAEMCFLSWFCGNQLPAYKCFQIT			
Macaca leonina A3F(30-137)	(30)	-----	CFEPQYHAEMCFLSRFCGNQLPAYKRFQIT			
Macaca nemestrina A3F(30-137)	(30)	-----	CFEPQYHAEMCFLSRFCGNQLPAYKRFQIT			
Rhinopithecus roxellana A3F(30-137)	(30)	-----	YFEPQYHAEMCFLSWFCGNQLPAYKRFQIT			
Mandrillus leucophaeus A3F(30-138)	(30)	-----	YFQPQYHAEMCFLSWFCGNQLPAYKRFQIT			
Macaca mulatta A3F(30-137)	(30)	-----	YSKPEHHAEMCFLSRFCGNQLPAYKRFQIT			
Theropithecus gelada A3F(30-137)	(30)	-----	YFQPQYHAEMCFLSRFCGNQLPAYKRFQIT			
Cercocebus atys A38(29-138)	(32)	-----	YSQPEHHAEMCFLSRFCGNQLPAYKRFQIT			
Macaca fascicularis A38(29-138)	(32)	-----	YSKPEHHAEMCFLSRFCGNQLPAYKRFQIW			
Macaca mulatta A38(29-138)	(32)	-----	YSKPEHHAEMCFLSRFCGNQLPAYKRFQIT			
Macaca leonina A38(29-138)	(32)	-----	YSKPEHHAEMCFLSRFCGNQLPAYKRFQIT			
Mandrillus leucophaeus A38(29-138)	(32)	-----	YSQPEHHAEMCFLSRFCGNQLPAYKRFQIT			
Macaca nemestrina A38(29-138)	(32)	-----	YSKPEHHAEMCFLSRFCGNQLPAYKRFQIW			
Rhinopithecus bieti A3F(29-138)	(32)	-----	YSQPEHHAEMMYFLSWFCGNQLPAYKRFQIT			
Rhinopithecus roxellana A38(29-138)	(32)	-----	YSQPEHHAEMMYFLSWFCGNQLPAYKRFQIT			
Chlorocebus sabaeus A38(29-138)	(32)	-----	YSKPEHHAEMCFLSWFCGNQLPAYKRFQIW			
Nomascus leucogenys A38(30-138)	(31)	-----	YFQPQYHAEMCFLSWFCGNQLPAYKRFQIT			
Cercocebus atys A3F(29-138)	(32)	-----	YFQPQYHAEMCFLSRFCGNQLPAYKRFQIT			
Papio anubis A3F(29-138)	(32)	-----	YFQPQYHAEMCFLSRFCGNQLPAYKRFQIT			
Chlorocebus aethiops A3D(29-150)	(40)	-----	RYEVYFQPQYHAEMMYFLSWFCGNQLPAYKRFQIW			
Chlorocebus sabaeus A3D(29-134)	(32)	-----	QYHAEMMYFLSWFCGNQLPAYKRFQIT			
Chlorocebus sabaeus A3F(29-150)	(40)	-----	RQEVYFQPQYHAEMMYFLSWFCGNQLPAYKRFQIT			
Erythrocebus patas A3D(29-150)	(40)	-----	RYEVYFQPQYHAEMCFLSWFCGNQLPAYKRFQIW			
Macaca fascicularis A3D(29-159)	(42)	ELSNWECRKRVYFQPQYHAEMMYFLSWFCGNQLPAYKRFQIT				
Macaca fascicularis A3F(29-138)	(32)	-----	YFQPQYHAEMMYFLSWFCGNQLPAYKRFQIT			
Macaca nemestrina A3D(29-138)	(32)	-----	YFQPQYHAEMCFLSWFCGNQLPAYKRFQIT			
Macaca leonina A3D(29-138)	(32)	-----	YFQPQYHAEMCFLSWFCGNQLPAYKRFQIW			
Macaca mulatta A3D(29-138)	(32)	-----	YFQPQYHAEMCFLSWFCGNQLPAYKRFQIT			
Gorilla A3D(29-150)	(39)	-----	HRQEVYFQFENHAEMCFLSWFCGNQLPAYKRFQIT			
Pan paniscus A3D(29-150)	(39)	-----	HRQEVYFQFENHAEMCFLSWFCGNQLPAYKRFQIW			
Pan troglodytes A3D(29-150)	(39)	-----	HRQEVYFQFENHAEMCFLSWFCGNQLPAYKRFQIT			
Homo sapiens A3D(29-150)	(39)	-----	HRQEVYFQFENHAEMCFLSWFCGNQLPAYKRFQIT			
Nomascus leucogenys A3D(29-150)	(39)	-----	HRQEVYFQFENHAEMCFLSWFCGNQLPAYKRFQIW			
Saimiri boliviensis A3C(29-138)	(33)	-----	FNPENHAEMMYFLSWFCGNQLPAYKRFQIT			
Saimiri boliviensis A3F(29-138)	(33)	-----	FNPENHAEMMYFLSWFCGNQLPAYKRFQIT			
Ptilocolobus tephroceles A3F(36-145)	(32)	-----	YFEPQYHAEMCFLSWFCGNQLPAYKCFQIT			
Colobus angolensis palliatus A3F(29-138)	(32)	-----	YFEPQYHAEMCFLSWFCGNQLPAYKCFQIW			
Pongo abelii A3F(30-150)	(38)	-----	HRREVYFQPQYHAEMCFLSWFCGNQLPAYKRFQIT			
Консенсус	(42)		YFQPQYHAEMCFLSWFCGNQLPAYKRFQIT			

Фигура 40С

Часть 3

	(83)	83	90	100	110	123
<i>Homo sapiens</i> A3B(29-138)	(62)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Gorilla</i> A3B(29-138)	(62)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Pan paniscus</i> A3B(29-138)	(62)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Pan troglodytes</i> A3B(29-138)	(62)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Gorilla</i> A3F(30-127)	(60)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWE--				
<i>Pan troglodytes</i> A3F(30-137)	(60)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Homo sapiens</i> A3F(30-137)	(60)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Macaca leonina</i> A3F(30-137)	(60)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Macaca nemestrina</i> A3F(30-137)	(60)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Rhinopithecus roxellana</i> A3F(30-137)	(60)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Mandrillus leucophaeus</i> A3F(30-138)	(60)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Macaca mulatta</i> A3F(30-137)	(60)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Theropithecus gelada</i> A3F(30-137)	(60)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Cercocebus atys</i> A3B(29-138)	(62)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Macaca fascicularis</i> A3B(29-138)	(62)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Macaca mulatta</i> A3B(29-138)	(62)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Macaca leonina</i> A3B(29-138)	(62)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Mandrillus leucophaeus</i> A3B(29-138)	(62)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Macaca nemestrina</i> A3B(29-138)	(62)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Rhinopithecus bieti</i> A3F(29-138)	(62)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Rhinopithecus roxellana</i> A3B(29-138)	(62)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Chlorocebus sabaeus</i> A3B(29-138)	(62)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Nomascus leucogenys</i> A3B(30-138)	(61)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Cercocebus atys</i> A3F(29-138)	(62)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Papio anubis</i> A3F(29-138)	(62)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Chlorocebus aethiops</i> A3D(29-150)	(74)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Chlorocebus sabaeus</i> A3D(29-134)	(58)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Chlorocebus sabaeus</i> A3F(29-150)	(74)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Erythrocebus patas</i> A3D(29-150)	(74)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Macaca fascicularis</i> A3D(29-159)	(83)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Macaca fascicularis</i> A3F(29-138)	(62)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Macaca nemestrina</i> A3D(29-138)	(62)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Macaca leonina</i> A3D(29-138)	(62)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Macaca mulatta</i> A3D(29-138)	(62)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Горилла</i> A3D(29-150)	(74)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Pan paniscus</i> A3D(29-150)	(74)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Pan troglodytes</i> A3D(29-150)	(74)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Homo sapiens</i> A3D(29-150)	(74)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Nomascus leucogenys</i> A3D(29-150)	(74)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Saimiri boliviensis</i> A3C(29-138)	(62)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Saimiri boliviensis</i> A3F(29-138)	(62)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Ptilocolobus tephrosceles</i> A3F(36-145)	(62)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Colobus angolensis palliatus</i> A3F(29-138)	(62)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
<i>Pongo abelii</i> A3F(30-150)	(73)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				
Консенсус	(83)	WYVSWNEFCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERN				

Фигура 40D

Часть 4

	(124)	124	131	SEQ ID NO:
Homo sapiens A38(29-138) (103)	YRRALCRL			93
Gorilla A38(29-138) (103)	YRRALCRL			94
Pan paniscus A38(29-138) (103)	YRRALCRL			95
Pan troglodytes A38(29-138) (103)	YRRALCRL			96
Gorilla A3F(30-127) (99)	-----			97
Pan troglodytes A3F(30-137) (101)	YRRALCRL			98
Homo sapiens A3F(30-137) (101)	YRRALCRL			99
Macaca leonina A3F(30-137) (101)	YRRALCRL			100
Macaca nemestrina A3F(30-137) (101)	YRRALCRL			101
Rhinopithecus roxellana A3F(30-137) (101)	YRRALCRL			102
Mandrillus leucophaeus A3F(30-138) (101)	Y-----			103
Macaca mulatta A3F(30-137) (101)	YRRALCRL			104
Theropithecus gelada A3F(30-137) (101)	NRRLALRL			105
Cercocebus atys A38(29-138) (103)	NRRLALRL			106
Macaca fascicularis A38(29-138) (103)	NRRLALRL			107
Macaca mulatta A38(29-138) (103)	NRRLALRL			108
Macaca leonina A38(29-138) (103)	NRRLALRL			109
Mandrillus leucophaeus A38(29-138) (103)	NRRLALRL			110
Macaca nemestrina A38(29-138) (103)	NRRLALRL			111
Rhinopithecus bieti A3F(29-138) (103)	NRRLALRL			112
Rhinopithecus roxellana A38(29-138) (103)	NRRLALRL			113
Chlorocebus sabaeus A38(29-138) (103)	YRRALCRL			114
Nomascus leucogenys A38(30-138) (102)	NRRLALRL			115
Cercocebus atys A3F(29-138) (103)	XRRALRL			116
Papio anubis A3F(29-138) (103)	NRRLALRL			117
Chlorocebus aethiops A3D(29-150) (115)	NRRLALRL			118
Chlorocebus sabaeus A3D(29-134) (99)	NRRLALRL			119
Chlorocebus sabaeus A3F(29-150) (115)	NRRLALRL			120
Erythrocebus patas A3D(29-150) (115)	NRRLALRL			121
Macaca fascicularis A3D(29-150) (124)	NRRLALRL			122
Macaca fascicularis A3F(29-138) (103)	NRRLALRL			123
Macaca nemestrina A3D(29-138) (103)	NRRLALRL			124
Macaca leonina A3D(29-138) (103)	NRRLALRL			125
Macaca mulatta A3D(29-138) (103)	NRRLALRL			126
Gorilla A3D(29-150) (115)	NRRLALRL			127
Pan paniscus A3D(29-150) (115)	NRRLALRL			128
Pan troglodytes A3D(29-150) (115)	NRRLALRL			129
Homo sapiens A3D(29-150) (115)	NRRLALRL			130
Nomascus leucogenys A3D(29-150) (115)	NRRLALRL			131
Saimiri boliviensis A3C(29-138) (103)	NRRLALRL			132
Saimiri boliviensis A3F(29-138) (103)	NRRLALRL			133
Ptilocobus tephrosceles A3F(36-145) (103)	NRRLALRL			134
Colobus angolensis palliatus A3F(29-138) (103)	YRRALCRL			135
Pongo abelii A3F(30-150) (114)	YRGALRL			136
Консенсус (124)	NRRLALRL			