(19) Евразийское патентное ведомство

(21) 202391646 (13) A1

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ

(43) Дата публикации заявки
2023.10.03

(22) Дата подачи заявки
2021.12.02

(51) Int. Cl. ***C12N 15/113*** (2010.01)
C12N 9/22 (2006.01)

(54) СКОНСТРУИРОВАННЫЕ СИСТЕМЫ CRISPR КЛАССА 2, ТИПА V

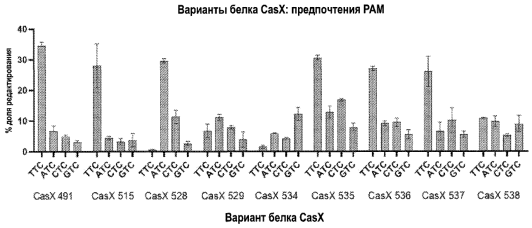
(31) **63/121,196; 63/162,346; 63/208,855**
(32) **2020.12.03; 2021.03.17; 2021.06.09**
(33) **US**
(86) **PCT/US2021/061673**
(87) **WO 2022/120095 2022.06.09**

(71) Заявитель:
СКРАЙБ ТЕРАПЬЮТИКС ИНК. (US)

(72) Изобретатель:
Хиггинс Шон, Колин Изабел, Денни Сара, Стаал Бретт Т., Оукс Бенджамин, Сидор Ангус, Макхиджа Сурадж, Виджаякумар Гаятхри (US)

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

(57) В настоящем документе предложены сконструированные нуклеазы класса 2, типа V и гидовые РНК, пригодные для редактирования целевых нуклеиновых кислот. Также предложены способы получения и применения таких вариантов для модификации нуклеиновых кислот.



202391646 A1

202391646 A1

ОПИСАНИЕ ИЗОБРЕТЕНИЯ

2420-578331EA/085

СКОНСТРУИРОВАННЫЕ СИСТЕМЫ CRISPR КЛАССА 2, ТИПА V

ПЕРЕКРЕСТНЫЕ ССЫЛКИ НА СМЕЖНЫЕ ЗАЯВКИ

[0001] Настоящая заявка испрашивает приоритет по предварительным заявкам на патент США №№ 63/121,196 от 3 декабря 2020 г.; 63/162,346 от 17 марта 2021 г. и 63/208,855 от 9 июня 2021 г., содержание каждой из которых в полном объеме включено в настоящий документ путем ссылки.

ВКЛЮЧЕНИЕ ПЕРЕЧНЯ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ПУТЕМ ССЫЛКИ

[0002] Содержание текстового файла, поданного в электронном виде, полностью включено в настоящий документ путем ссылки: Экземпляр списка последовательностей в машиночитаемом формате (имя файла: SCRB_031_03WO_SeqList_ST25, дата записи: 1 декабря 2021 г., размер файла 5,61 Мб).

ПРЕДПОСЫЛКИ СОЗДАНИЯ ИЗОБРЕТЕНИЯ

[0003] Системы CRISPR-Cas бактерий и архей обеспечивают форму приобретенного иммунитета против фага и вирусов. Интенсивные исследования, проводимые в последнее десятилетие, позволили определить биохимию этих систем. Системы CRISPR-Cas состоят из белков Cas, которые участвуют в обнаружении, нацеливании и расщеплении чужеродной ДНК или РНК, и матрицы CRISPR, которая включает в себя прямые повторы, фланкирующие короткие спейсерные последовательности, направляющие белки Cas к их мишеням. Системы CRISPR-Cas класса 2 представляют собой оптимизированные версии, в которых один белок Cas, связанный с РНК, отвечает за связывание и расщепление целевой последовательности. Программируемый характер этих минимальных систем облегчает их применение в качестве универсальной технологии, которая совершает революцию в сфере манипулирования геномом.

[0004] На сегодняшний день было обнаружено лишь несколько систем CRISPR/Cas класса 2, которые широко применяются. Из них тип V уникален тем, что в нем используется единый унифицированный домен RuvC-подобной эндонуклеазы (RuvC), который распознает 5’ PAM-последовательности, которые отличаются от 3’ PAM-последовательностей, распознаваемых Cas9, и он формирует смещенное расщепление в целевой нуклеиновой кислоте с 5-, 7- или 10-нуклеотидными 5’-липкими концами (Yang et al., PAM-dependent target DNA recognition and cleavage by C2c1 CRISPR-Cas endonuclease. Cell 167:1814 (2016)). Однако Cas типа V дикого типа и гидовые последовательности имеют низкую эффективность редактирования. Таким образом, в данной области существует потребность в дополнительных системах CRISPR/Cas класса 2, типа V (например, в комбинациях белок Cas плюс гидовая РНК), которые были оптимизированы и/или обеспечивают улучшение по сравнению с системами более ранних поколений для использования в различных терапевтических, диагностических и исследовательских областях применения.

ИЗЛОЖЕНИЕ СУЩНОСТИ ИЗОБРЕТЕНИЯ

[0005] Настоящее описание относится к гидовым рибонуклеиновым кислотам (гРНК), сконструированным белкам CRISPR класса 2, типа V и системам из сконструированных белков CRISPR класса 2, типа V и гидовых рибонуклеиновых кислот (гРНК), используемым для модификации целевой нуклеиновой кислоты гена в эукариотических клетках. В некоторых вариантах осуществления в настоящем описании предложены сконструированные белки класса 2, типа V, содержащие одну или более модификаций относительно домена референсного белка CasX и демонстрирующие одну или более улучшенных характеристик по сравнению с референсным белком CasX с SEQ ID NO: 2. В других вариантах осуществления в настоящем описании предложены сконструированные варианты последовательности варианта белка CasX, такого как CasX 491 (SEQ ID NO: 336) или CasX 515 (SEQ ID NO: 416), причем белок класса 2, типа V содержит по меньшей мере одну модификацию относительно домена варианта CasX и демонстрирует одну или более улучшенных характеристик по сравнению с вариантом белка CasX. В некоторых вариантах осуществления вариант класса 2, типа V способен образовывать комплекс с гидовой рибонуклеиновой кислотой (гРНК), причем комплекс может связывать целевую нуклеиновую кислоту и расщеплять ее, при этом целевая нуклеиновая кислота содержит нецелевую цепь и целевую цепь.

[0006] В некоторых вариантах осуществления в настоящем описании предложены гидовые рибонуклеиновые кислоты (гРНК), включая одиночные гидовые композиции, способные связывать вариант белка класса 2, типа V, причем гРНК содержит по меньшей мере одну модификацию в области по сравнению с гРНК с SEQ ID NO: 2238 или SEQ ID NO: 2239. В некоторых вариантах осуществления модифицированные области каркаса гРНК включают в себя: (a) удлиненную стеблевую петлю; (b) каркасную стеблевую петлю; (c) триплекс и (d) псевдоузел. В некоторых случаях каркасная удлиненная петля варианта гРНК дополнительно содержит модификацию пузырька. В других случаях каркас гРНК дополнительно содержит модификацию области триплексной петли. В других случаях каркас варианта гРНК дополнительно содержит гетерологичную РНК в удлиненном стебле, включая последовательности шпилек.

[0007] В некоторых вариантах осуществления в настоящем описании предложены пары для редактирования генов, содержащие сконструированные белки класса 2, типа V и варианты гРНК по любому из вариантов осуществления, описанных в настоящем документе, причем пара для редактирования генов демонстрирует по меньшей мере одну улучшенную характеристику по сравнению с парой для редактирования генов, содержащей референсный белок CasX с SEQ ID NO:1, SEQ ID NO:2 или SEQ ID NO:3 и гРНК с SEQ ID NO: 4 или SEQ ID NO: 5. В конкретном варианте осуществления сконструированный белок класса 2, типа V содержит последовательность, выбранную из группы, состоящей из последовательностей SEQ ID NO: 247-592 или 1147-1231, как указано в таблице 3, или последовательность, имеющую идентичность последовательности с ней по меньшей мере примерно 85%, или по меньшей мере примерно 90%, или по меньшей мере примерно 95%, или по меньшей мере примерно 95%, или по меньшей мере примерно 96%, или по меньшей

мере примерно 97%, или по меньшей мере примерно 98%, или по меньшей мере примерно 99%, а гРНК представляет собой последовательность, выбранную из группы, состоящей из последовательностей SEQ ID NO: 2101-2332 и 2353-2398, как указано в таблице 2, или последовательность, имеющую идентичность последовательности с ней по меньшей мере примерно 85%, по меньшей мере примерно 90%, или по меньшей мере примерно 95%, или по меньшей мере примерно 95%, или по меньшей мере примерно 96%, или по меньшей мере примерно 97%, или по меньшей мере примерно 98%, или по меньшей мере примерно 99%. В конкретном варианте осуществления сконструированный белок класса 2, типа V содержит последовательность, выбранную из группы, состоящей из последовательностей SEQ ID NO: 270-592 или 1147-1231, или последовательность, имеющую идентичность последовательности с ней по меньшей мере примерно 85%, или по меньшей мере примерно 90%, или по меньшей мере примерно 95%, или по меньшей мере примерно 95%, или по меньшей мере примерно 96%, или по меньшей мере примерно 97%, или по меньшей мере примерно 98%, или по меньшей мере примерно 99%, а гРНК представляет собой последовательность, выбранную из группы, состоящей из последовательностей SEQ ID NOS: 2238-2332 и 2353-2398, или последовательность, имеющую идентичность последовательности с ней по меньшей мере примерно 85%, по меньшей мере примерно 90%, или по меньшей мере примерно 95%, или по меньшей мере примерно 95%, или по меньшей мере примерно 96%, или по меньшей мере примерно 97%, или по меньшей мере примерно 98%, или по меньшей мере примерно 99%. В конкретном варианте осуществления сконструированный белок класса 2, типа V содержит последовательность, выбранную из группы, состоящей из последовательностей SEQ ID NO: 415-592 или 1147-1231, или последовательность, имеющую идентичность последовательности с ней по меньшей мере примерно 85%, или по меньшей мере примерно 90%, или по меньшей мере примерно 95%, или по меньшей мере примерно 95%, или по меньшей мере примерно 96%, или по меньшей мере примерно 97%, или по меньшей мере примерно 98%, или по меньшей мере примерно 99%, а гРНК представляет собой последовательность, выбранную из группы, состоящей из последовательностей SEQ ID NO: 2281-2332 и 2353-2398, или последовательность, имеющую идентичность последовательности с ней по меньшей мере примерно 85%, по меньшей мере примерно 90%, или по меньшей мере примерно 95%, или по меньшей мере примерно 95%, или по меньшей мере примерно 96%, или по меньшей мере примерно 97%, или по меньшей мере примерно 98%, или по меньшей мере примерно 99%.

[0008] В некоторых вариантах осуществления в настоящем описании предложены полинуклеотиды и векторы, кодирующие сконструированные варианты белков класса 2, типа V, варианты гРНК и пары для редактирования генов, описанные в настоящем документе. В некоторых вариантах осуществления векторы представляют собой вирусные векторы, такие как аденоассоциированные вирусные (AAV) векторы. В других вариантах осуществления векторы представляют собой частицы для доставки CasX, называемые XDP, которые содержат RNP с парами для редактирования генов.

[0009] В некоторых вариантах осуществления в настоящем описании предложены клетки, содержащие полинуклеотиды, векторы, сконструированные белки класса 2, типа V и гРНК, описанные в настоящем документе. В других вариантах осуществления в настоящем описании предложены клетки, содержащие целевую нуклеиновую кислоту, отредактированную способами редактирования по вариантам осуществления, описанным в настоящем документе.

[0010] В некоторых вариантах осуществления в настоящем описании предложены наборы, содержащие полинуклеотиды, векторы, сконструированные белки класса 2, типа V, гРНК и пары для редактирования генов, описанные в настоящем документе.

[0011] В некоторых вариантах осуществления в настоящем описании предложены способы редактирования целевой нуклеиновой кислоты, включающие в себя приведение целевой нуклеиновой кислоты в контакт с белком класса 2, типа V и вариантом гРНК, описанными в настоящем документе, причем приведение в контакт приводит к редактированию или модификации целевой нуклеиновой кислоты.

[0012] В некоторых вариантах осуществления в настоящем описании предложены способы редактирования целевой нуклеиновой кислоты в популяции клеток, включающие в себя приведение клеток в контакт с одной или более парами для редактирования генов, описанными в настоящем документе, причем приведение в контакт приводит к редактированию или модификации целевой нуклеиновой кислоты в популяции клеток.

[0013] В других вариантах осуществления в описании предложены способы лечения нуждающегося в этом субъекта, включающие введение пары для редактирования генов или векторов, содержащих или кодирующих пары для редактирования генов по любому из вариантов осуществления, описанных в настоящем документе.

[0014] В другом аспекте в настоящем документе предложены пары для редактирования генов, композиции, содержащие пары для редактирования генов, или векторы, содержащие или кодирующие пары для редактирования генов, для применения в качестве лекарственного средства.

[0015] В другом аспекте в настоящем документе предложены пары для редактирования генов, композиции, содержащие пары для редактирования генов, или векторы, содержащие или кодирующие пары для редактирования генов, для применения в способе лечения, причем способ включает редактирование или модификацию целевой нуклеиновой кислоты; необязательно при этом редактирование происходит у субъекта, имеющего мутацию в аллеле гена, причем мутация вызывает заболевание или расстройство у субъекта, при этом предпочтительно редактирование изменяет мутацию в аллель дикого типа гена или приводит к нокауту или нокдауну аллеля гена, вызывающего заболевание или расстройство у субъекта.

ВКЛЮЧЕНИЕ ПУТЕМ ССЫЛКИ

[0016] Все публикации, патенты и заявки на патенты, упомянутые в настоящем описании, включены в настоящий документ путем ссылки в той степени, как если бы каждая отдельная публикация, патент или заявка на патент была конкретно и

индивидуально обозначена как включенная путем ссылки. Содержание WO 2020/247882, WO 2020/247883 и WO 2021/113772, которые раскрывают варианты CasX и варианты гРНК, а также способы их доставки включены в настоящий документ путем ссылки в полном объеме.

КРАТКОЕ ОПИСАНИЕ ГРАФИЧЕСКИХ МАТЕРИАЛОВ

[0017] Новые признаки изобретения подробно описаны в приложенной формуле изобретения. Для лучшего понимания признаков и преимуществ настоящего изобретения рекомендуется изучить приведенное ниже подробное описание, в котором изложены наглядные примеры осуществления, использующие принципы изобретения, а также прилагаемые фигуры, из которых:

[0018] На ФИГ. 1 представлен график результатов анализа по количественному определению активных фракций RNP, образованных огРНК 174 (SEQ ID NO: 2238) и вариантами CasX 119, 457, 488 и 491, как описано в примере 8. Последовательности, соответствующие огРНК и вариантам CasX, представлены в таблицах 2 и 3 соответственно. Эквимолярные количества RNP и мишень совместно инкубировали и определяли количество расщепленной мишени в указанных временных точках. Для каждой временной точки показаны среднее и стандартное отклонение для трех независимых повторов. Показана аппроксимация объединенных повторов двухфазной моделью. «2» относится к референсному белку CasX с SEQ ID NO: 2.

[0019] На ФИГ. 2 представлено количественное определение активных фракций RNP, образованных CasX2 (референсный белок CasX с SEQ ID NO: 2) и модифицированными огРНК, как описано в примере 8. Эквимолярные количества RNP и мишень совместно инкубировали и определяли количество расщепленной мишени в указанных временных точках. Для каждой временной точки показаны среднее и стандартное отклонение для трех независимых повторов. Показана аппроксимация объединенных повторов двухфазной моделью.

[0020] На ФИГ. 3 представлено количественное определение активных фракций RNP, образованных CasX 491 и модифицированными огРНК в условиях ограничения гида, как описано в примере 8. Эквимолярные количества RNP и мишень совместно инкубировали и определяли количество расщепленной мишени в указанных временных точках. Показана аппроксимация данных двухфазной моделью.

[0021] На ФИГ. 4 представлено количественное определение скоростей расщепления RNP, образованного огРНК174 и вариантами CasX, как описано в примере 8. Целевую ДНК инкубировали с 20-кратным избытком указанного RNP и определяли количество расщепленной мишени в указанных временных точках. Среднее и стандартное отклонение трех независимых повторов показаны для каждой временной точки, за исключением 488 и 491, где показан один повтор. Показана однофазная аппроксимация объединенных повторов.

[0022] На ФИГ. 5 представлено количественное определение скоростей расщепления RNP, образованного CasX2 и вариантами огРНК, как описано в примере 8.

Целевую ДНК инкубировали с 20-кратным избытком указанного RNP и определяли количество расщепленной мишени в указанных временных точках. Для каждой временной точки показаны среднее и стандартное отклонение для трех независимых повторов. Показана однофазная аппроксимация объединенных повторов.

[0023] На ФИГ. 6 представлено количественное определение начальных скоростей для RNP, образованного CasX2 и вариантами огРНК, как описано в примере 8. Первые две временные точки из предыдущего эксперимента по расщеплению соответствовали аппроксимации линейной моделью для определения начальной скорости расщепления.

[0024] На ФИГ. 7 представлено количественное определение скоростей расщепления RNP, образованного CasX491 и вариантами огРНК, как описано в примере 8. Целевую ДНК инкубировали с 20-кратным избытком указанного RNP при 10 °C, и количество расщепленной мишени определяли в указанных временных точках. Показана однофазная аппроксимация временных точек.

[0025] На ФИГ. 8 представлено количественное определение компетентных фракций RNP из вариантов 515 и 526 CasX в комплексе с вариантом 174 гРНК по сравнению с RNP из референсного CasX 2 в комплексе с гРНК 2 с применением эквимолярных количеств указанного RNP и комплементарной мишени, как описано в примере 8. Показана аппроксимация двухфазной моделью для каждого периода времени или набора комбинированных повторов.

[0026] На ФИГ. 9 представлено количественное определение скоростей расщепления RNP из вариантов 515 и 526 CasX в комплексе с вариантом 174 гРНК по сравнению с RNP из референсного CasX 2 в комплексе с гРНК 2 с применением 20-кратного избытка указанного RNP, как описано в примере 8.

[0027] На ФИГ. 10A представлено количественное определение скоростей расщепления вариантов CasX на PAM TTC, как описано в примере 5. Целевые ДНК-субстраты с идентичными спейсерами и указанной PAM-последовательностью инкубировали с 20-кратным избытком указанного RNP при 37 °C и определяли количество расщепленной мишени в указанных временных точках. Показана однофазная аппроксимация одиночного повтора.

[0028] На ФИГ. 10B представлено количественное определение скоростей расщепления вариантов CasX на PAM CTC, как описано в примере 5. Целевые ДНК-субстраты с идентичными спейсерами и указанной PAM-последовательностью инкубировали с 20-кратным избытком указанного RNP при 37 °C и определяли количество расщепленной мишени в указанных временных точках. Показана однофазная аппроксимация одиночного повтора.

[0029] На ФИГ. 10C представлено количественное определение скоростей расщепления вариантов CasX на PAM GTC, как описано в примере 5. Целевые ДНК-субстраты с идентичными спейсерами и указанной PAM-последовательностью инкубировали с 20-кратным избытком указанного RNP при 37 °C и определяли количество расщепленной мишени в указанных временных точках. Показана однофазная

аппроксимация одиночного повтора.

[0030] На ФИГ. 10D представлено количественное определение скоростей расщепления вариантов CasX на PAM ATC, как описано в примере 5. Целевые ДНК-субстраты с идентичными спейсерами и указанной PAM-последовательностью инкубировали с 20-кратным избытком указанного RNP при 37 °C и определяли количество расщепленной мишени в указанных временных точках. Показана однофазная аппроксимация одиночного повтора.

[0031] На ФИГ. 11A представлено количественное определение скоростей расщепления RNP из варианта 491 CasX и гида 174 на PAM NTC, как описано в примере 5. Временные точки были взяты в пределах 2 минут, и расщепленную фракцию графически изображали для каждой мишени и временной точки, но для ясности показаны только первые две минуты временного периода.

[0032] На ФИГ. 11B представлено количественное определение скоростей расщепления RNP из варианта 491 CasX и гида 174 на PAM NTT, как описано в примере 5. Временные точки были взяты в пределах 10 минут, и расщепленную фракцию графически изображали для каждой мишени и временной точки.

[0033] На ФИГ. 12A представлено количественное определение расщепления с помощью RNP, образованного огРНК174 и вариантом 515 CasX, с использованием спейсеров длиной 18, 19 или 20 нуклеотидов, как описано в примере 9. Целевую ДНК инкубировали с 20-кратным избытком указанного RNP и определяли количество расщепленной мишени в указанных временных точках. Для каждой временной точки показаны среднее и стандартное отклонение для трех независимых повторов. Показана однофазная аппроксимация объединенных повторов.

[0034] На ФИГ. 12B представлено количественное определение расщепления с помощью RNP, образованного огРНК174 и вариантом 526 CasX, с использованием спейсеров длиной 18, 19 или 20 нуклеотидов, как описано в примере 9. Целевую ДНК инкубировали с 20-кратным избытком указанного RNP и определяли количество расщепленной мишени в указанных временных точках. Для каждой временной точки показаны среднее и стандартное отклонение для трех независимых повторов. Показана однофазная аппроксимация объединенных повторов.

[0035] На ФИГ. 13 представлено схематическое изображение примера белка CasX и каркасной ДНК-последовательности для упаковки в аденоассоциированный вирус (AAV). Сегмент ДНК между инвертированными концевыми повторами (ITR) в AAV, состоящий из ДНК, кодирующей CasX, и ее промотора, и ДНК, кодирующей каркас, и ее промотора, упаковывают в капсид AAV во время получения AAV.

[0036] На ФИГ. 14 представлены результаты анализа редактирования, в котором сравниваются каркасы гРНК 229-237 с каркасом 174 в мышиных нейрональных клетках-предшественниках (mNPC), выделенных из трансгенных мышей Ai9-tdtomato, как описано в примере 21. Клетки нуклеофицировали указанными дозами плазмид p59, кодирующих CasX 491, каркас и спейсер 11.30 (5’ AAGGGGCUCCGCACCACGCC 3’, SEQ ID NO: 17),

нацеленных на mRHO. Редактирование в локусе mRHO оценивали через 5 дней после трансфекции методом NGS, и редактирование с помощью конструктов с каркасами 230, 231, 234 и 235, как продемонстрировано, было больше по сравнению с конструктами с каркасом 174 при обеих дозах.

[0037] На ФИГ. 15 представлены результаты анализа редактирования со сравнением каркасов гРНК 229-237 с каркасом 174 в клетках mNPC, как описано в примере 21. Клетки нуклеофицировали указанными дозами плазмид p59, кодирующих CasX 491, каркас и спейсер 12.7 (5' CUGCAUUCUAGUUGUGGUUU 3', SEQ ID NO: 1146), нацеленных на повторяющиеся элементы, предотвращающие экспрессию флуоресцентного белка tdTomato. Редактирование оценивали через 5 дней после трансфекции методом FACS для количественного определения доли tdTomato-положительных клеток. Клетки, нуклеофицированные каркасами 231-235, демонстрировали приблизительно на 35% большее редактирование по сравнению с конструктами с каркасом 174 в высокой дозе и приблизительно на 25% большее редактирование при низкой дозе.

[0038] На ФИГ. 16 представлена схема, иллюстрирующая пример способа получения белка CasX и вариантов гидовой РНК описания с использованием глубокой мутационной эволюции (DME). В некоторых примерах осуществления DME создает и проверяет почти каждую возможную мутацию, вставку и делецию в биомолекуле и их комбинации/мультипликации, а также обеспечивает почти полную и непредвзятую оценку ландшафта пригодности биомолекулы и путей в пространстве последовательностей к требуемым результатам. Как описано в настоящем документе, DME можно применить как к белку CasX, так и к гидовой РНК.

[0039] На ФИГ. 17A представлена CryoEM-структура RNP-комплекса *Deltaproteobacteria* из белка CasX : огРНК (PDB id: 6YN2), включающего две стеблевые петли, псевдоузел и триплекс, как описано в примере 13.

[0040] На ФИГ. 17B представлена вторичная структура огРНК с SEQ ID NO: 4, идентифицированная по структуре, показанной в (A), с использованием инструмента RNAPDBee 2.0 (rnapdbee.cs.put.poznan.pl/, с использованием инструментов 3DNA/DSSR и с использованием средства визуализации VARNA). Указаны области РНК. Остатки, которые не были очевидны по файлу кристаллической структуры PDB, указаны буквами обычного текста (т. е. не обведенные) и не включены в нумерацию остатков.

[0041] На ФИГ. 18 представлена схема областей и доменов гидовой РНК, используемой для конструирования библиотеки каркасов, как описано в примере 13.

[0042] На ФИГ. 19 представлена секторная диаграмма относительного распределения и конструкции библиотеки каркасов с несистематическими (двойными и одиночными) мутациями и нацеленными мутациями (на триплекс, пузырек каркасного стебля, псевдоузел и удлиненный стебель и петлю), как описано в примере 13.

[0043] На ФИГ. 20 представлена схема мутагенеза триплекса, выполненного с возможностью специфического встраивания перемежающихся триплекс-образующих пар оснований в триплекс, как описано в примере 13. Сплошные линии обозначают пару

Уотсона - Крика в триплексе; нуклеотид третьей цепи обозначен пунктирной линией, представляющей неканоническое взаимодействие с пурином дуплекса. В библиотеке каждую из 5 указанных позиций заменяли на все возможные триплексные мотивы (G:GC, T:AT, G:GC)=243 последовательности. Последовательность ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCANNNAUCAAAG (SEQ ID NO: 1022).

[0044] На ФИГ. 21 представлена гистограмма с результатами значений обогащения референсных гидовых каркасов 174 и 175 в каждом скрининге, как описано в примере 13.

[0045] На ФИГ. 22 представлены графики рассеяния, показывающие $\log_2$ обогащения для каждой измеренной однонуклеотидной замены, делеции или вставки, измеренные в каждом из двух независимых скринингов мутантных библиотек для гидовых каркасов 174 и 175, как описано в примере 13.

[0046] На ФИГ. 23 представлены тепловые карты для одиночных мутантов в гидовых каркасах 174 и 175, показывающие конкретные пригодные для мутации области в каркасе по последовательностям, как описано в примере 13. Желтые тона отражают значения с обогащением, сходным с референсными каркасами; красные тона указывают на увеличение обогащения и, таким образом, на активность относительно референсного каркаса; синие тона указывают на потерю активности относительно каркаса дикого типа; белый цвет указывает на отсутствующие данные (или замену, которая приводит к последовательности дикого типа).

[0047] На ФИГ. 24 представлен график рассеяния, который сравнивает $\log_2$ обогащения однонуклеотидными мутациями референсных гидовых каркасов 174 и 175, описанных в примере 13. Показаны только мутации в положениях, которые были аналогичны в каркасах 174-175. Результаты позволяют предположить, что, в целом, гидовый каркас 174 обладает большей толерантностью к изменениям, чем 175.

[0048] На ФИГ. 25 представлена гистограмма, демонстрирующая среднее значение (и 95%-й доверительный интервал) для $\log_2$ обогащения для набора каркасов, в которых тасовались пары в псевдоузлах, так что каждый новый псевдоузел имеет тот же состав пар оснований, но расположенных в другом порядке в стебле, описанном в примере 13. Каждый столбик представляет набор каркасов с указанным положением пары G:A (или A:G) (см. схему справа). Протестировали 291 стебель псевдоузла; числа над столбцами указывают на число стеблей с парой G:A (или A:G) в каждом положении.

[0049] На ФИГ. 26 представлено схематическое изображение последовательности псевдоузла, показанного на ФИГ. 55 и 56, в порядке от 5’ к 3’, с последовательностями двух цепей, разделенных подчеркиванием.

[0050] На ФИГ. 27 представлена гистограмма, демонстрирующая среднее значение (и 95%-й доверительный интервал) для $\log_2$ обогащения для каркасов, поделенное на прогнозируемую стабильность вторичной структуры области стебля псевдоузла, как описано в примере 13. Каркасы с очень стабильными стеблями (например, $\Delta G < -7$ ккал/моль) имели в среднем высокие значения обогащения, а каркасы с

дестабилизированными стеблями ($\Delta G \geq -5$ ккал/моль) имели в среднем низкие значения обогащения.

[0051] На ФИГ. 28 представлена тепловая карта всех двойных мутантов в положениях 7 и 29 в каркасе 175, как описано в примере 13. Последовательность псевдоузла представлена справа в порядке от 5' до 3'.

[0052] На ФИГ. 29A представлены результаты редактирования клеток ARPE-19, нуклеофицированных с использованием сконструированной гидовой 235 в сравнении с 174, со спейсером 11.1, нацеленным на сайт P23 локуса Rho (с CasX 491), демонстрирующие улучшенную активность варианта 235, с повышенной специфической активностью в отношении экзогенного Rho дикого типа без нецелевого расщепления мутантного Rho-репортерного гена (с ненацеливающим спейсером), как описано в примере 21.

[0053] На ФИГ. 29B представлена гистограмма, демонстрирующая кратность изменения уровней редактирования p59.491.235.11.1, нормализованную к эталонным уровням p59.491.174.11.1 (взятым за 1,0) в клетках ARPE-1, нуклеофицированных 1000 нг каждой плазмиды, как описано в примере 21.

[0054] На ФИГ. 30 представлены результаты анализа редактирования, со сравнением Cas-нуклеаз 2 119, 491, 515, 527, 528, 529, 530 и 531 в специальной клеточной линии HEK293, PASS_V1.01, как описано в примере 17. Клетки липофицировали 2 мкг плазмиды p67, кодирующей указанный белок Cas. Через пять дней извлекали геномную ДНК клеток. Проводили ПЦР-амплификацию и секвенирование нового поколения для выделения и количественного определения доли отредактированных клеток в специально разработанных сайтах редактирования на мишенях. Для каждой пробы редактирование оценивали в целевых сайтах (индивидуальных точках), состоящих из следующих PAM-последовательностей: индивидуальные сайты 48 TTC, 14 ATC, 22 CTC, 11 GTC, и процентную долю редактирования нормализовали к контролю - несущей среде. Клетки, липофицированные любой нуклеазой, демонстрировали более высокое среднее редактирование в целевых сайтах PAM TTC (горизонтальная полоса), чем в случае нуклеазы дикого типа Cas 2, за исключением Cas 528. Относительное предпочтение любой данной нуклеазы к четырем разным PAM-последовательностям также представлено на скрипичных графиках. В частности, Cas-нуклеазы 527, 528 и 529 демонстрируют по существу иные предпочтения PAM, чем Cas 2-нуклеазы дикого типа.

[0055] На ФИГ. 31 представлены результаты анализа редактирования, где сравнивается улучшенная Cas-нуклеаза 491 с улучшенными нуклеазами 532 и 533 в специально разработанной клеточной линии HEK293, PASS_V1.01, как описано в примере 18. Клетки липофицировали в двух повторностях 2 мкг плазмиды p67, кодирующей указанный белок Cas и ген резистентности к пуромицину, и выращивали в условиях пуромициновой селекции. Через три дня извлекали геномную ДНК клеток. Проводили ПЦР-амплификацию и секвенирование нового поколения для выделения и количественного определения доли отредактированных клеток в специально разработанных сайтах редактирования на мишенях. Для каждой пробы редактирование оценивали в целевых

сайтах, состоящих из следующих PAM-последовательностей: индивидуальные сайты 48 TTC, 14 ATC, 22 CTC, 11 GTC, и долю редактирования нормализовали к контролю - несущей среде. Клетки, липофицированные Cas 532 или 533, демонстрировали в среднем большее редактирование, чем Cas 491, в случае каждой из PAM-последовательностей, за исключением Cas 533 в целевых сайтах TTC PAM. Планки погрешностей представляют стандартную ошибку среднего для n=2 биологических проб.

[0056] На ФИГ. 32 представлен график анализа выживаемости для определения селективной строгости отбора CcdB с разными спейсерами при нацеливании белком CasX 515 и каркасом 174, как описано в примере 14.

[0057] На ФИГ. 33A-E представлены тепловые карты вариантов CasX 515, демонстрирующие нейтральное или улучшенное биохимическое расщепление для каждого мутанта в целевом сайте TTC PAM в виде среднего значения для трех спейсеров, как описано в примере 14. На фигурах представлены результаты по всей длине последовательности CasX 515.

[0058] На ФИГ. 34A-E представлены тепловые карты вариантов CasX 515, демонстрирующие нейтральное или улучшенное биохимическое расщепление для каждого мутанта в целевом сайте CTC PAM в виде среднего значения для трех биологических повторностей одного спейсера, как описано в примере 14. На фигурах представлены результаты по всей длине последовательности CasX 515.

[0059] На ФИГ. 35A-E представлены тепловые карты вариантов CasX 515, демонстрирующие нейтральное или улучшенное биохимическое расщепление для каждого мутанта в целевом сайте CTC PAM в виде среднего значения для трех биологических повторностей одного спейсера, как описано в примере 14. На фигурах представлены результаты по всей длине последовательности CasX 515.

[0060] На ФИГ. 36A-E представлены тепловые карты вариантов CasX 515, демонстрирующие нейтральное или улучшенное биохимическое расщепление для каждого мутанта в целевом сайте ATC PAM в виде среднего значения для трех биологических повторностей одного спейсера, как описано в примере 14. На фигурах представлены результаты по всей длине последовательности CasX 515.

[0061] На ФИГ. 37A представлен график, показывающий влияние длины спейсера на способность редактировать целевую нуклеиновую кислоту с помощью RNP в клетках Jurkat, как описано в примере 15, для спейсера 15.3.

[0062] На ФИГ. 37B представлен график, показывающий влияние длины спейсера на способность редактировать целевую нуклеиновую кислоту с помощью RNP в клетках Jurkat, как описано в примере 15, для спейсера 15.5.

[0063] На ФИГ. 38 представлена гистограмма выбранных вариантов белка CasX и их эффективности редактирования при четырех разных PAM-последовательностях (TTC, ATC, CTC и GTC) для дублирующихся проб, как описано в примере 16. Данные представлены в виде процентной доли редактирования +/- станд. откл. (SD).

[0064] На ФИГ. 39 представлена гистограмма, демонстрирующая среднюю

эффективность редактирования относительно CasX 491 выбранных CasX-нуклеаз на 48 разных целевых сайтах TTC PAM, как описано в примере 19. Планками погрешностей показана распространяемая стандартная ошибка среднего по двум экспериментам. Звездочками указано значимое различие между CasX 527 и CasX 491 (p=0,0000635 с помощью двухстороннего t-критерия Уэлча).

[0065] На ФИГ. 40 представлена схема, основанная на опубликованной CryoEM-структуре гомологичного референсного CasX 1 (SEQ ID NO: 1; идентификационный номер Protein Data Bank: 6NY2), с указанием физических местоположений PAM-последовательности целевой ДНК, взаимодействующей с PAM петли, домена NTSB и положения аминокислоты 26, как описано в примере 19.

[0066] На ФИГ. 41 представлен скрипичный график для выбранных вариантов белка CasX и их эффективности редактирования на 48 целевых сайтах TTC PAM, как описано в примере 19.

[0067] На ФИГ. 42 представлена гистограмма выбранных вариантов белка CasX и их эффективности редактирования относительно CasX 491 в 48 целевых сайтах TTC PAM, как описано в примере 19. Данные представлены в виде средней относительной эффективности редактирования, где 1,0 соответствует редактированию с помощью CasX 491. Серая пунктирная линия иллюстрирует эффективность редактирования CasX 119. Ошибка представляет собой распространяемую станд. ош. среднего для дублирующихся проб.

[0068] На ФИГ. 43 представлена гистограмма, демонстрирующая среднюю эффективность редактирования относительно CasX 491, а также среднее соотношение специфичности выбранных нуклеаз CasX, как описано в примере 20.

[0069] На ФИГ. 44 представлена блок-схема, иллюстрирующая качественную взаимосвязь между протестированными комбинациями мутаций и их влиянием как на активность, так и на специфичность полученных вариантов CasX, как описано в примере 21.

[0070] На ФИГ. 45A представлены результаты анализов AAV-опосредованного редактирования, где сравниваются каркас 235 и каркас 174 гРНК и гиды 11.30 и 11.31 на эндогенном локусе экзона 1 Rho мыши в клетках mNPC в диапазоне MOI, как описано в примере 21.

[0071] На ФИГ. 45B представлены результаты редактирования в виде кратности изменения уровней редактирования для каркаса 235 относительно гида 174 (взятого на 1,0) со спейсером 11.30 на клетках, инфицированных 5,0e+ 5 MOI, как описано в примере 21.

[0072] На ФИГ. 46 представлено схематическое изображение, демонстрирующее модификации, внесенные в удлиненную стеблевую петлю в варианте 175 гРНК, которые были встроены в вариант 235 гРНК. Удлиненная стеблевая петля огРНК 175: SEQ ID NO: 1285; удлиненная стеблевая петля огРНК 325: SEQ ID NO: 1286.

[0073] На ФИГ. 47 представлена схема варианта 235 гРНК с указанием модификаций в триплексе, пузырьке каркасного стебля и удлиненной стеблевой петле относительно вариантов 174 и 175 гРНК. Псевдоузел и триплексная петля: SEQ ID NO: 1287; каркасный

стебель и удлиненный стебель: SEQ ID NO: 1288.

[0074] На ФИГ. 48 представлено схематическое изображение, иллюстрирующее положения оснований внутри шпильки MS2, как описано в примере 23. Последовательность MS2 на фигуре: SEQ ID NO: 1289.

[0075] На ФИГ. 49 представлен график процентной доли редактирования в локусе tdTomato, измеренной с помощью флуоресценции tdTomato, в случае XDP, в которую упакованы указанные варианты каркаса, с использованием каркасов 188 и 251 гРНК в качестве базовых вариантов, как описано в примере 23. Использовали две версии MS2 (MS2 353 и MS2 дикого типа).

[0076] На ФИГ. 50 представлены улучшения значений EC50, определенные с помощью NanoSight, для редактирования в локусе tdTomato в NPC, относительно титров, в случае XDP, в которые упакованы указанные варианты каркаса гРНК, с помощью каркасов 188 и 251 в качестве базовых контролей, как описано в примере 23. Использовали две версии MS2: MS2 353 и MS2 дикого типа (WT).

[0077] На ФИГ. 51 представлена корреляция между аффинностью шпильки MS2 (K_D) и EC50 для XDP, в которые упакованы изображенные варианты каркаса гРНК, как описано в примере 23.

[0078] На ФИГ. 52 представлена корреляция между аффинностью шпильки MS2 (K_D) и титром для XDP, в которые упакованы изображенные варианты каркаса гРНК, как описано в примере 23.

ПОДРОБНОЕ ОПИСАНИЕ ИЗОБРЕТЕНИЯ

[0079] Хотя в настоящем документе показаны и описаны предпочтительные варианты осуществления настоящего изобретения, специалистам в данной области будет очевидно, что такие варианты осуществления предложены только в качестве примера. Специалистам в данной области будут понятны многочисленные вариации, изменения и замены без выхода за рамки изобретения. Следует понимать, что при реализации изобретения на практике можно использовать различные альтернативы вариантов осуществления изобретения, описанных в настоящем документе. Предполагается, что следующие пункты формулы изобретения определяют объем изобретения, при этом способы и структуры в рамках пунктов формулы, а также их эквиваленты охвачены формулой изобретения.

[0080] Если не указано иное, все технические и научные термины, применяемые в настоящем документе, имеют общепринятое значение, понятное любому рядовому специалисту в данной области, к которой относится настоящее изобретение. В настоящем документе описаны способы и материалы, хотя для проверки или анализа настоящих вариантов осуществления можно применять приемлемые способы и материалы, аналогичные или эквивалентные описанным ниже. В случае противоречий приоритет будет иметь настоящее описание патента, включая определения. Кроме того, материалы, способы и примеры приведены только для иллюстрации и не имеют ограничительного характера. Специалистам в данной области будут понятны многочисленные вариации, изменения и

замены без выхода за рамки изобретения.

Определения

[0081] В настоящем документе термины «полинуклеотид» и «нуклеиновая кислота» применяются взаимозаменяемо и относятся к полимерной форме нуклеотидов любой длины - либо рибонуклеотидов, либо дезоксирибонуклеотидов. Таким образом, термины «полинуклеотид» и «нуклеиновая кислота» охватывают одноцепочечную ДНК; двухцепочечную ДНК; многоцепочечную ДНК; одноцепочечную РНК; двухцепочечную РНК; многоцепочечную РНК; геномную ДНК; кДНК; гибриды ДНК-РНК и полимер, содержащий пуриновые и пиримидиновые основания или другие природные, химически или биохимически модифицированные, неприродные или производные нуклеотидные основания.

[0082] Термины «гибридизуемый» или «комплементарный» применяются взаимозаменяемо для обозначения того, что нуклеиновая кислота (например, РНК, ДНК) содержит последовательность нуклеотидов, которая позволяет ей нековалентно связываться, т. е. образовывать пары оснований Уотсона - Крика и/или пары оснований G/U, «отжигаться», «гибридизоваться» с другой нуклеиновой кислотой специфичным для последовательности антипараллельным способом (т. е. нуклеиновая кислота специфически связывается с комплементарной нуклеиновой кислотой) в соответствующих условиях температуры и ионной силы раствора *in vitro* и/или *in vivo*. Следует понимать, что для обеспечения специфической гибридизации последовательность полинуклеотида необязательно должна быть на 100% комплементарна последовательности целевой нуклеиновой кислоты; она может иметь идентичность последовательности по меньшей мере примерно 70%, по меньшей мере примерно 80%, или по меньшей мере примерно 90%, или по меньшей мере примерно 95%, и все равно гибридизоваться с последовательностью целевой нуклеиновой кислоты. Более того, полинуклеотид может гибридизоваться на одном или более сегментах, так что промежуточные или прилегающие сегменты не участвуют в событии гибридизации (например, структура петли или шпилька, «утолщение», «пузырек» и т. п.).

[0083] Термин «ген» для целей настоящего описания включает в себя область ДНК, кодирующую генный продукт (например, белок, РНК), а также все области ДНК, которые регулируют продукцию генного продукта, независимо от того, являются ли такие регуляторные последовательности смежными с кодирующими и/или транскрибируемыми последовательностями. Соответственно, ген может включать в себя последовательности вспомогательных элементов, включая, без ограничений, промоторные последовательности, терминаторы, последовательности регуляции трансляции, такие как сайты связывания рибосомы и внутренние сайты входа для рибосомы, энхансеры, сайленсеры, изоляторы, пограничные элементы, точки начала репликации, сайты присоединения матрицы и области управления локусом. Кодирующие последовательности кодируют генный продукт при транскрипции или транскрипции и трансляции; кодирующие последовательности описания могут содержать фрагменты и не обязательно содержат полноразмерную открытую рамку

считывания. Ген может включать в себя как транскрибируемую цепь, так и комплементарную цепь, содержащую антикодоны.

[0084] Термин «расположенный ниже» относится к нуклеотидной последовательности, которая расположена в 3'-направлении к референсной нуклеотидной последовательности. В определенных вариантах осуществления расположенные ниже нуклеотидные последовательности относятся к последовательностям, которые следуют за начальной точкой транскрипции. Например, кодон инициации трансляции гена расположен ниже начального сайта транскрипции.

[0085] Термин «расположенный выше» относится к нуклеотидной последовательности, которая расположена в 5'-направлении к референсной нуклеотидной последовательности. В определенных вариантах осуществления расположенными выше нуклеотидными последовательностями называются последовательности, которые расположены с 5'-стороны от кодирующей области или начальной точки транскрипции. Например, большинство промоторов расположены выше от начального сайта транскрипции.

[0086] Термин «смежный» по отношению к полинуклеотидным или аминокислотным последовательностям относится к последовательностям, которые находятся рядом друг с другом или смежны друг с другом в полинуклеотиде или полипептиде. Специалисту в данной области будет понятно, что две последовательности могут считаться смежными друг с другом и все же включать в себя ограниченное количество промежуточной последовательности, например, 1, 2, 3, 4, 5, 6, 7, 8, 9 или 10 нуклеотидов или аминокислот.

[0087] Термин «вспомогательный элемент» применяется в настоящем документе взаимозаменяемо с термином «вспомогательная последовательность» и предназначен для обозначения, *помимо прочего*, сигналов полиаденилирования (сигнал поли(A)), энхансерных элементов, интронов, элементов посттранскрипционной регуляции (PTRE), сигналов ядерной локализации (NLS), деаминаз, ингибиторов ДНК-гликозилазы, дополнительных промоторов, факторов, стимулирующих CRISPR-опосредованную направляемую гомологией репарацию (например, цис- или транс-), активаторов или репрессоров транскрипции, саморасщепляющихся последовательностей и доменов слияния, например, домена слияния, слитого с белком CRISPR. Следует понимать, что выбор соответствующего вспомогательного элемента или элементов будет зависеть от кодируемого компонента, который нужно экспрессировать (например, белка или РНК), или от того, содержит ли нуклеиновая кислота множество компонентов, которые требуют разных полимераз, или они не предназначены для экспрессии в виде слитного белка.

[0088] Термин «промотор» относится к ДНК-последовательности, которая содержит начальный сайт транскрипции и дополнительные последовательности для облегчения связывания полимеразы и осуществления транскрипции. Примеры эукариотических промоторов включают в себя такие элементы, как TATA-бокс, и/или элемент распознавания B (BRE), и они способствуют или стимулируют транскрипцию и экспрессию связанной

транскрибируемой полинуклеотидной последовательности и/или гена (или трансгена). Промотор может быть получен синтетически или может быть получен из известной или природной промоторной последовательности или другой промоторной последовательности. Промотор может быть проксимальным или дистальным по отношению к транскрибируемому гену. Промотор также может включать химерный промотор, содержащий комбинацию из двух или более гетерологичных последовательностей для придания определенных свойств. Промотор настоящего описания может включать варианты промоторных последовательностей, которые аналогичны по составу, но не идентичны другой (-им) промоторной (-ым) последовательности (-ям), известной или представленной в настоящем документе. Промотор может быть классифицирован по критериям, связанным с характером экспрессии связанной кодирующей или транскрибируемой последовательности или гена, функционально связанного с промотором, таким как конститутивный, связанный с развитием, тканеспецифический, индуцибельный и т. д. Промотор также может быть классифицирован по его силе. В контексте промотора «сила» относится к скорости транскрипции гена, контролируемого промотором. «Сильный» промотор означает, что скорость транскрипции высока, а «слабый» промотор означает, что скорость транскрипции относительно низкая.

[0089] Промотор описания может представлять собой промотор полимеразы II (Pol II). Полимераза II транскрибирует все гены, кодирующие белки, и многие некодирующие гены. Типичный промотор Pol II включает в себя основной промотор, который представляет собой последовательность из примерно 100 пар оснований, окружающих начальный сайт транскрипции, и служит в качестве платформы связывания для полимеразы Pol II и связанных общих факторов транскрипции. Промотор может содержать один или более элементов основного промотора, таких как TATA-бокс, BRE, инициатор (INR), элемент мотив-десять (MTE), нижележащий элемент основного промотора (DPE), нижележащий основной элемент (DCE), хотя в данной области известны основные промоторы, не содержащие этих элементов.

[0090] Промотор описания может представлять собой промотор полимеразы III (Pol III). Pol III транскрибирует ДНК для синтеза малых рибосомных РНК, таких как 5S рРНК, тРНК и другие малые РНК. В типичных промоторах Pol III используются внутренние управляющие последовательности (последовательности в транскрибируемой части гена) для поддержки транскрипции, хотя также иногда используются вышележащие элементы, такие как TATA-бокс. Все промоторы Pol III предусмотрены в рамках объема настоящего описания.

[0091] Термин «энхансер» относится к регуляторным ДНК-последовательностям, которые при связывании со специфическими белками, называемыми транскрипционными факторами, регулируют экспрессию связанного гена. Энхансеры могут быть расположены в интроне гена или на 5’- или 3’-стороне от кодирующей последовательности гена. Энхансеры могут быть проксимальными к гену (т. е. располагаться в пределах нескольких десятков или сотен пар нуклеотидов (п. о.) от промотора) или могут быть расположены

дистально по отношению к гену (т. е. на расстоянии тысячи п. о., сотни тысяч п. о. или даже миллионов п. о. от промотора). Один ген может регулироваться более чем одним энхансером, и все они предусмотрены в рамках объема настоящего описания.

[0092] В настоящем документе термин «посттранскрипционный регуляторный элемент (PRE)», такой как PRE гепатита, относится к ДНК-последовательности, которая при транскрибировании создает третичную структуру, способную демонстрировать посттранскрипционную активность для усиления или стимулирования экспрессии ассоциированного функционально связанного с ней гена.

[0093] В настоящем документе термин «посттранскрипционный регуляторный элемент (PTRE)», такой как PTRE гепатита, относится к ДНК-последовательности, которая при транскрибировании создает третичную структуру, способную демонстрировать посттранскрипционную активность для усиления или стимулирования экспрессии ассоциированного функционально связанного с ней гена.

[0094] В настоящем документе термин «рекомбинантный» означает, что конкретная нуклеиновая кислота (ДНК или РНК) представляет собой продукт различных комбинаций стадий клонирования, рестрикции и/или лигирования, что приводит к получению конструкта, имеющего структурную кодирующую или некодирующую последовательность, отличную от эндогенных нуклеиновых кислот, встречающихся в природных системах. По существу, ДНК-последовательности, кодирующие структурную кодирующую последовательность, могут быть собраны из фрагментов кДНК и коротких олигонуклеотидных линкеров или из ряда синтетических олигонуклеотидов для получения синтетической нуклеиновой кислоты, которая способна экспрессироваться из рекомбинантной единицы транскрипции, содержащейся в клетке или в бесклеточной системе транскрипции и трансляции. Такие последовательности могут быть представлены в форме открытой рамки считывания, без прерываний внутренними нетранслируемыми последовательностями или интронами, которые обычно присутствуют в эукариотических генах. Геномную ДНК, содержащую соответствующие последовательности, также можно применять при создании рекомбинантного гена или единицы транскрипции. Последовательности нетранслируемой ДНК могут присутствовать на 5’- или 3’-стороне от открытой рамки считывания, где такие последовательности не препятствуют манипуляциям или экспрессии кодирующих областей, и с их действием может быть фактически связано модулирование продукции требуемого продукта с помощью различных механизмов (см. «энхансеры» и «промоторы» выше).

[0095] Термин «рекомбинантный полинуклеотид» или «рекомбинантная нуклеиновая кислота» относится к таким, которые не встречаются в природе, например, получены посредством искусственного комбинирования двух отличных по иным параметрам отдельных сегментов последовательности посредством вмешательства человека. Такую искусственную комбинацию часто получают либо средствами химического синтеза, либо посредством искусственного манипулирования отдельными сегментами нуклеиновых кислот, например, методами генной инженерии. Эта операция

обычно выполняется для замены кодона на избыточный кодон, кодирующий ту же или консервативную аминокислоту, при этом, как правило, вводят или удаляют сайт распознавания последовательности. Альтернативно эту операцию выполняют для объединения сегментов нуклеиновых кислот с требуемыми функциями для получения требуемой комбинации функций. Такую искусственную комбинацию часто получают либо средствами химического синтеза, либо посредством искусственного манипулирования отдельными сегментами нуклеиновых кислот, например, методами генной инженерии.

[0096] Аналогичным образом, термин «рекомбинантный полипептид» или «рекомбинантный белок» относится к полипептиду или белку, который не встречается в природе, например, получен посредством искусственного комбинирования двух отличных по иным параметрам отдельных сегментов аминокислотной последовательности посредством вмешательства человека. Таким образом, например, белок, который содержит гетерологичную аминокислотную последовательность, является рекомбинантным.

[0097] В настоящем документе термин «приведение в контакт» означает установление физического соединения между двумя или более объектами. Например, приведение целевой нуклеиновой кислоты в контакт с гидовой нуклеиновой кислотой означает, что обеспечена возможность общей физической связи целевой нуклеиновой кислоты и гидовой нуклеиновой кислоты; например, они могут гибридизоваться, если последовательности обладают сходством последовательностей.

[0098] Термины «константа диссоциации» или «K_d» применяются взаимозаменяемо и означают аффинность между лигандом «L» и белком «P»; т. е. то, насколько прочно лиганд связывается с конкретным белком. Ее можно рассчитать с применением формулы K_d=[L] [P]/[LP], где [P], [L] и [LP] представляют собой молярные концентрации белка, лиганда и комплекса соответственно.

[0099] В описании предложены системы и способы, используемые для редактирования последовательности целевой нуклеиновой кислоты. В настоящем документе термин «редактирование» применяется взаимозаменяемо с термином «модификация» и включает в себя, без ограничений, расщепление, никирование, делецию, нокин, нокаут и т. п.

[0100] Под «расщеплением» подразумевается разрыв ковалентной основной цепи молекулы целевой нуклеиновой кислоты (например, РНК, ДНК). Расщепление можно инициировать различными способами, включая, без ограничений, ферментативный или химический гидролиз фосфодиэфирной связи. Возможны как одноцепочечное расщепление, так и двухцепочечной расщепление, и двухцепочечное расщепление может происходить в результате двух отдельных событий одноцепочечного расщепления.

[0101] Термин «нокаут» относится к устранению гена или экспрессии гена. Например, ген может быть нокаутирован посредством либо делеции, либо добавления нуклеотидной последовательности, что приводит к нарушению рамки считывания. В качестве другого примера, ген может быть нокаутирован посредством замены части гена на нерелевантную последовательность. В настоящем документе термин «нокаут» относится к

снижению экспрессии гена или его генного (-ых) продукта (-ов). В результате нокаута гена активность или функция белка могут быть ослаблены или уровни белка могут быть снижены или устранены.

[0102] В настоящем документе термин «направляемая гомологией репарация» (HDR) относится к форме репарации ДНК, которая происходит во время восстановления двухцепочечных разрывов в клетках. Этот процесс требует гомологии нуклеотидной последовательности, и в нем применяется донорная матрица для репарации или нокаута целевой ДНК, и он приводит к передаче генетической информации от донора к мишени. Направляемая гомологией репарация может приводить к изменению последовательности целевой последовательности посредством вставки, делеции или мутации, если донорная матрица отличается от целевой ДНК-последовательности и если часть или вся последовательность донорной матрицы встроена в целевую ДНК.

[0103] В настоящем документе термин «негомологичное соединение концов» (NHEJ) относится к репарации двухцепочечных разрывов в ДНК посредством прямого лигирования концов разрыва друг с другом без необходимости в гомологичной матрице (в отличие от направляемой гомологией репарации, которая требует гомологичной последовательности, направляющей репарацию). NHEJ часто приводит к потере (делеции) нуклеотидной последовательности рядом с сайтом двухцепочечного разрыва.

[0104] В настоящем документе термин «соединение концов, обусловленное микрогомологией» (MMEJ) относится к мутагенному механизму репарации DSB, который всегда ассоциируется с делециями, фланкирующими сайты разрывов, без необходимости в гомологичной матрице (в отличие от направляемой гомологией репарации, которая требует того, чтобы гомологичная последовательность направляла репарацию). MMEJ часто приводит к потере (делеции) нуклеотидной последовательности рядом с сайтом двухцепочечного разрыва. Полинуклеотид или полипептид имеет определенную процентную долю «сходства последовательности» или «идентичности последовательности» с другим полинуклеотидом или полипептидом, и это означает, что при выравнивании процентные доли оснований или аминокислот двух сравниваемых последовательностей одинаковы, и они находятся в том же относительном положении. Сходство последовательности (иногда называемое процентным сходством, процентной идентичностью или гомологией) может быть определено несколькими разными способами. Для определения сходства последовательностей последовательности можно выровнять с применением способов и компьютерных программ, известных в данной области, включая BLAST, доступ к которой во всемирной сети осуществляется по адресу ncbi.nlm.nih.gov/BLAST. Процент комплементарности между конкретными участками нуклеотидных последовательностей в пределах нуклеиновых кислот можно определять с применением любого удобного способа. Примеры способов включают программы BLAST (basic local alignment search tools) и программы PowerBLAST (Altschul et al., J. Mol. Biol., 1990, 215, 403-410; Zhang and Madden, Genome Res., 1997, 7, 649-656) или с применением программ Gap (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer

Group, University Research Park, г. Мэдисон, штат Висконсин, США), например, с применением заданных по умолчанию параметров, в которых применяется алгоритм Смита - Ватермана (Adv. Appl. Math., 1981, 2, 482-489).

[0105] Термины «полипептид» и «белок» применяются в настоящем документе взаимозаменяемо и относятся к полимерной форме аминокислот любой длины, которая может включать в себя кодированные и некодированные аминокислоты, химически или биохимически модифицированные или дериватизированные аминокислоты, а также полипептиды, имеющие модифицированные пептидные основные цепи. Термин включает в себя слитные белки, включая, без ограничений, слитные белки с гетерологичной аминокислотной последовательностью.

[0106] «Вектор» или «экспрессионный вектор» представляет собой репликон, такой как плазмида, фаг, вирус или космида, к которому можно присоединить другой сегмент ДНК, т. е. «вставку», так чтобы вызвать репликацию или экспрессию присоединенного сегмента в клетке.

[0107] Термин «встречающийся в природе», или «немодифицированный», или «дикого типа», используемый в настоящем документе применительно к нуклеиновой кислоте, полипептиду, клетке или организму, относится к нуклеиновой кислоте, полипептиду, клетке или организму, встречающемуся в природе.

[0108] В настоящем документе термин «мутация» относится к вставке, делеции, замене, дублированию или инверсии одной или более аминокислот или нуклеотидов по сравнению с референсной аминокислотной последовательностью или аминокислотной последовательностью дикого типа или с референсной нуклеотидной последовательностью или нуклеотидной последовательностью дикого типа.

[0109] В настоящем документе термин «выделенный» предназначен для описания полинуклеотида, полипептида или клетки, находящейся в среде, отличной от той, в которой полинуклеотид, полипептид или клетка встречается в природных условиях. Выделенная генетически модифицированная клетка-хозяин может присутствовать в смешанной популяции генетически модифицированных клеток-хозяев.

[0110] Термин «клетка-хозяин», используемый в настоящем документе, обозначает эукариотическую клетку, прокариотическую клетку или клетку из многоклеточного организма (например, клеточную линию), причем эукариотические или прокариотические клетки применяются в качестве реципиентов нуклеиновой кислоты (например, экспрессионного вектора), и включают в себя потомство исходной клетки, которая была генетически модифицирована нуклеиновой кислотой. Следует понимать, что потомство одной клетки не обязательно может быть полностью идентичным по морфологии, или геному, или общему комплементу ДНК исходному родителю вследствие естественной, случайной или преднамеренной мутации. «Рекомбинантная клетка-хозяин» (также называемая «генетически модифицированной клеткой-хозяином») представляет собой клетку-хозяина, в которую была введена гетерологичная нуклеиновая кислота, например, экспрессионный вектор.

[0111] В настоящем документе термин «тропизм» относится к преимущественному проникновению вирусоподобной частицы (XDP, иногда также называемой в настоящем документе XDP) в определенный (-ые) тип (-ы) клеток или тканей и/или преимущественному взаимодействию с клеточной поверхностью, что облегчает проникновение в определенные типы клеток или тканей, необязательно и предпочтительно с последующим экспрессией (например, транскрипцией и необязательно трансляцией) последовательностей, перенесенных XDP в клетку.

[0112] Термины «псевдотип» или «псевдотипирование», используемые в настоящем документе, относятся к белкам вирусной оболочки, которые были замещены белками оболочки другого вируса, обладающими предпочтительными характеристиками. Например, ВИЧ можно псевдотипировать оболочечным G-белком вируса везикулярного стоматита (VSV-G) (описанным, помимо прочего, в настоящем документе ниже), что позволяет ВИЧ заражать более широкий диапазон клеток, так как оболочечные белки ВИЧ нацеливают вирус главным образом на CD4+-презентирующие клетки.

[0113] В настоящем документе термин «фактор тропизма» относится к компонентам, интегрированным в поверхность XDP, что обеспечивает тропизм к определенному типу клеток или тканей. Не имеющими ограничительного характера примерами факторов тропизма являются гликопротеины, фрагменты антител (например, scFv, нанотела, линейные антитела и т. д.), рецепторы и лиганды к маркерам клеток-мишеней.

[0114] Термин «маркер клеток-мишеней» относится к молекуле, экспрессируемой клеткой-мишенью, включая, без ограничений, рецепторы клеточной поверхности, цитокиновые рецепторы, антигены, ассоциированные с опухолью антигены, гликопротеины, олигонуклеотиды, субстраты ферментов, антигенные детерминанты или сайты связывания, которые могут присутствовать в на поверхности ткани-мишени или клетки-мишени и которые могут служить в качестве лигандов для фрагмента антитела или фактора тропизма.

[0115] Термин «консервативная аминокислотная замена» относится к взаимозаменяемости в белках аминокислотных остатков, имеющих сходные боковые цепи. Например, группа аминокислот, имеющих алифатические боковые цепи, состоит из глицина, аланина, валина, лейцина и изолейцина; группа аминокислот, имеющих алифатически-гидроксильные боковые цепи, состоит из серина и треонина; группа аминокислот, имеющих амидосодержащие боковые цепи, состоит из аспарагина и глутамина; группа аминокислот, имеющих ароматические боковые цепи, состоит из фенилаланина, тирозина и триптофана; группа аминокислот, имеющих основные боковые цепи, состоит из лизина, аргинина и гистидина; а группа аминокислот, имеющих серосодержащие боковые цепи, состоит из цистеина и метионина. Примерами групп консервативных аминокислотных замен являются: валин-лейцин-изолейцин, фенилаланин-тирозин, лизин-аргинин, аланин-валин и аспарагин-глутамин.

[0116] В настоящем документе термин «антитело» охватывает различные структуры антител, включая, без ограничений, моноклональные антитела, поликлональные антитела,

мультиспецифические антитела (например, биспецифические антитела), нанотела, однодоменные антитела, такие как антитела VHH, и фрагменты антител, при условии, что они проявляют требуемую антигенсвязывающую активность или иммунологическую активность. Антитела представляют собой большое семейство молекул, которые включают в себя несколько типов молекул, таких как IgD, IgG, IgA, IgM и IgE.

[0117] Термин «фрагмент антитела» относится к молекуле, отличной от интактного антитела, которая содержит часть интактного антитела и которая связывается с антигеном, с которым связывается интактное антитело. Примеры фрагментов антител включают в себя, без ограничений, Fv, Fab, Fab', Fab'-SH, F(ab')2, диатела, одноцепочечные диатела, линейные антитела, однодоменное антитело, однодоменное верблюжье антитело, антительные молекулы с одноцепочечным вариабельным фрагментом (scFv) и мультиспецифические антитела, образованные из фрагментов антител.

[0118] В настоящем документе термины «лечение» или «терапия» применяются взаимозаменяемо и относятся к подходу для получения благоприятных или требуемых результатов, включая, без ограничений, терапевтический благоприятный эффект и/или профилактический благоприятный эффект. Под терапевтическим благоприятным эффектом подразумевается устранение или облегчение базового расстройства или заболевания, подлежащего лечению. Терапевтический благоприятный эффект также может быть достигнут посредством устранения или облегчения одного или более симптомов или улучшения одного или более клинических параметров, связанных с базовым заболеванием, так что у субъекта наблюдается улучшение, несмотря на то что субъект все еще может страдать базовым расстройством.

[0119] Термины «терапевтически эффективное количество» и «терапевтически эффективная доза», используемые в настоящем документе, относятся к количеству лекарственного средства или биологического препарата, отдельно или в составе композиции, которое способно оказывать какой-либо обнаруживаемый благоприятный эффект в отношении любого симптома, аспекта, измеренного параметра или характеристики болезненного состояния или заболевания при введении одной дозы или повторяющихся доз субъекту, такому как человек или экспериментальное животное. Такой эффект не должен быть абсолютным, чтобы быть признанным благоприятным.

[0120] В настоящем документе термин «введение» означает способ предоставления дозы соединения (например, композиции описания) или композиции (например, фармацевтической композиции) субъекту.

[0121] «Субъект» представляет собой млекопитающее. К млекопитающим относятся, без ограничений, одомашненные животные, приматы, не являющиеся человеком, люди, собаки, кролики, мыши, крысы и другие грызуны.

[0122] Все публикации, патенты и заявки на патенты, упомянутые в настоящем описании, включены в настоящий документ путем ссылки в той степени, как если бы каждая отдельная публикация, патент или заявка на патент была конкретно и индивидуально обозначена как включенная путем ссылки.

I. Общие способы

[0123] При практическом применении настоящего изобретения, если не указано иное, используются традиционные методики иммунологии, биохимии, химии, молекулярной биологии, микробиологии, клеточной биологии, геномики и рекомбинантной ДНК, которые можно найти в таких стандартных справочных пособиях, как: Molecular Cloning: A Laboratory Manual, 3rd Ed. (Sambrook et al., Harbor Laboratory Press 2001); Short Protocols in Molecular Biology, 4th Ed. (Ausubel et al. eds., John Wiley & Sons 1999); Protein Methods (Bollag et al., John Wiley & Sons 1996); Nonviral Vectors for Gene Therapy (Wagner et al. eds., Academic Press 1999); Viral Vectors (Kaplift & Loewy eds., Academic Press 1995); Immunology Methods Manual (I. Lefkovits ed., Academic Press 1997) и Cell and Tissue Culture: Laboratory Procedures in Biotechnology (Doyle & Griffiths, John Wiley & Sons 1998), содержимое которых включено в настоящий документ путем ссылки.

[0124] Если приведен диапазон значений, следует понимать, что в него входят конечные точки и каждое промежуточное значение до десятой доли нижнего предела, если контекстом явно не предусмотрено иное, между верхним и нижним пределами этого диапазона, а также любое другое указанное или промежуточное значение в этом указанном диапазоне. Верхний и нижний пределы этих меньших диапазонов могут быть независимо включены в меньшие диапазоны и также включены при наличии любого специально исключенного предела в указанном диапазоне. Если указанный диапазон включает один или оба предела, то диапазоны, исключающие любой из двух или оба таких предела сразу, также включены.

[0125] Если не указано иное, все технические и научные термины, используемые в настоящем документе, имеют общепринятое значение, понятное любому специалисту в области, к которой относится настоящее изобретение. Все публикации, упомянутые в настоящем документе, включены в настоящий документ путем ссылки для раскрытия и описания способов и/или материалов, в связи с которыми эти публикации были процитированы.

[0126] Необходимо отметить, что в настоящем документе и в прилагаемой формуле изобретения формы единственного числа также включают объекты во множественном числе, если из контекста явно не следует иное.

[0127] Следует понимать, что определенные признаки описания, которые для ясности приведены в контексте разных вариантов осуществления, также могут быть представлены в комбинации в одном варианте осуществления. В других случаях различные признаки описания, которые для краткости описаны в контексте одного варианта осуществления, также могут быть представлены отдельно или в любой приемлемой подкомбинации. Подразумевается, что все комбинации вариантов осуществления, относящихся к описанию, конкретно включены в настоящее описание и раскрыты в настоящем документе так, как если бы каждая комбинация была четко раскрыта в отдельности. Кроме того, все подкомбинации различных вариантов осуществления и их элементы также конкретно включены в настоящее описание и описаны в настоящем

документе так, как если бы каждая такая подкомбинация была четко раскрыта в отдельности.

II. Системы для редактирования генов и пары для редактирования генов

[0128] В первом аспекте в настоящем описании предложены системы, содержащие белок-нуклеазу CRISPR класса 2, типа V и одну или более гидовых нуклеиновых кислот (например, гРНК) для применения при модификации или редактировании целевой нуклеиновой кислоты гена, включающего кодирующие и некодирующие области. Как правило, с применением программируемых систем и способов, предложенных в настоящем документе, можно нацеливаться на любую часть гена. В настоящем документе термин «система», например системы, содержащие белок-нуклеазу CRISPR и одну или более гРНК описания в качестве пар для редактирования генов, а также системы, содержащие нуклеиновые кислоты, кодирующие белки-нуклеазы CRISPR и гРНК, и системы, содержащие векторы, которые содержат нуклеиновые кислоты, или системы, содержащие белок-нуклеазу CRISPR и одну или более гРНК описания, применяются взаимозаменяемо с термином «композиция».

[0129] В некоторых вариантах осуществления в описании предложены системы, специально предназначенные для модификации целевой нуклеиновой кислоты гена в эукариотических клетках; *in vitro*, *ex vivo* или in vivo у субъекта. Как правило, с применением программируемых систем и способов, предложенных в настоящем документе, можно нацеливаться на любую часть гена. В некоторых вариантах осуществления нуклеаза CRISPR представляет собой нуклеазу класса 2, типа V. Несмотря на то что элементы нуклеаз CRISPR-Cas класса 2, типа V имеют различия, они также имеют некоторые общие характеристики, которые отличают их от систем Cas9. Во-первых, нуклеазы типа V обладают одиночным РНК-направляемым эффектором, содержащим домен RuvC, но не содержащим домен HNH, и они распознают ТС-мотив PAM на 5'-стороне от целевой области на нецелевой цепи, что отличается от систем Cas9, которые используют богатый G-нуклеотидами PAM на 3'-стороне от целевых последовательностей. Нуклеазы типа V создают ступенчатые двухцепочечные разрывы дистально от PAM-последовательности, в отличие от Cas9, которые создают тупой конец в проксимальном сайте вблизи PAM. Кроме того, нуклеазы типа V разрушают транс-активные оцДНК при активации целевой дцДНК или связывании с цис-активной оцДНК. В некоторых вариантах осуществления в описании предложена нуклеаза класса 2, типа V, выбранная из группы, состоящей из Cas12a, Cas12b, Cas12c, Cas12d (CasY), Cas12j, Cas12k, C2c4, C2c8, C2c5, C2c10, C2c9, CasZ и CasX. В некоторых вариантах осуществления в описании предложены системы, содержащие один или более вариантов белка CasX и один или более вариантов гидовых нуклеиновых кислот (гРНК) в качестве системы CasX : гРНК.

[0130] В настоящем документе предложены системы, содержащие белок класса 2, типа V и вариант гРНК, называемые в настоящем документе парой для редактирования генов. В некоторых вариантах осуществления вариант белка класса 2, типа V представляет собой вариант CasX, такой как, без ограничений, последовательность SEQ ID NO: 416.

Термины «вариант белка CasX» и «вариант CasX» используются в настоящем документе взаимозаменяемо. В некоторых вариантах осуществления гРНК представляет собой вариант другой гРНК, такой как, без ограничений, последовательности SEQ ID NO: 2238 и 2239. гРНК и белок CasX могут связываться друг с другом посредством нековалентных взаимодействий с образованием комплекса - пары для редактирования генов, называемого рибонуклеопротеиновым (RNP) комплексом. В некоторых вариантах осуществления применение предварительно сформированных RNP-комплексов CasX : гРНК дает преимущества при доставке в клетку компонентов системы или последовательности целевой нуклеиновой кислоты для редактирования целевой нуклеиновой кислоты. В RNP гРНК может обеспечивать специфичность RNP-комплекса посредством включения нацеливающей последовательности (или «спейсера»), имеющего нуклеотидную последовательность, которая комплементарна последовательности целевой нуклеиновой кислоты. В RNP белок CasX предварительно сформированного комплекса CasX : гРНК обеспечивает сайт-специфическую активность и направляется к целевому сайту (и дополнительно стабилизируется в целевом сайте) в последовательности целевой нуклеиновой кислоты, подлежащей модификации, посредством связи с гРНК. Вариант белка CasX из RNP-комплекса обеспечивает сайт-специфические активности комплекса, такие как связывание, расщепление или никирование целевой последовательности белком CasX. В настоящем документе предложены системы и клетки, содержащие варианты белка CasX, варианты гРНК и пары для редактирования генов CasX : гРНК из любой комбинации вариантов осуществления варианта CasX и варианта гРНК, описанных в настоящем документе, а также способы доставки, включающие в себя CasX : гРНК. Каждый из этих компонентов и их применение при редактировании целевой нуклеиновой кислоты гена описаны в настоящем документе ниже.

[0131] В некоторых вариантах осуществления в описании предложены системы пар для редактирования генов, содержащие вариант белка CasX, выбранный из любого из вариантов белков CasX из таблицы 3 (SEQ ID NO: 247-592 или 1147-1231, или последовательность, имеющую идентичность последовательности с ней по меньшей мере примерно 85%, или по меньшей мере примерно 90%, или по меньшей мере примерно 95%, или по меньшей мере примерно 95%, или по меньшей мере примерно 96%, или по меньшей мере примерно 97%, или по меньшей мере примерно 98%, или по меньшей мере примерно 99%, и гРНК представляет собой вариант гРНК, описанный в настоящем документе (например, SEQ ID NO: 2101-2332 и 2353-2398, представленные в таблице 2), или варианты последовательности, имеющие идентичность последовательности с ней по меньшей мере 60%, или по меньшей мере 70%, по меньшей мере примерно 80%, или по меньшей мере примерно 90%, или по меньшей мере примерно 95%, причем гРНК содержит нацеливающую последовательность, комплементарную целевой нуклеиновой кислоте. В некоторых вариантах осуществления в описании предложены системы пар для редактирования генов, содержащие вариант белка CasX, выбранный из любого из вариантов белков CasX из таблицы 3 (SEQ ID NO: 270-592 и 1147-1231), а гРНК

представляет собой вариант гРНК, описанный в настоящем документе (например, SEQ ID NO: 2238-2332 и 2353-2398), причем гРНК содержит нацеливающую последовательность, комплементарную целевой нуклеиновой кислоте. В некоторых вариантах осуществления в описании предложены системы пар для редактирования генов, содержащие вариант белка CasX, выбранный из любого из вариантов белков CasX из таблицы 3 (SEQ ID NO: 415-592 и 1147-1231), а гРНК представляет собой вариант гРНК, описанный в настоящем документе (например, SEQ ID NO: 2281-2332 и 2353-2398), причем гРНК содержит нацеливающую последовательность, комплементарную целевой нуклеиновой кислоте. В других вариантах осуществления в описании предложены системы из пары для редактирования генов, содержащей вариант белка CasX, первый вариант гРНК, описанный в настоящем документе; например, SEQ ID NO: 2101-2332 или 2353-2398, указанные в таблице 2) с нацеливающей последовательностью, и второй вариант гРНК, причем второй вариант гРНК имеет нацеливающую последовательность, комплементарную другой или перекрывающейся части целевой нуклеиновой кислоты по сравнению с нацеливающей последовательностью первой гРНК. В других вариантах осуществления в описании предложены системы из пары для редактирования генов, содержащей вариант белка CasX, первый вариант гРНК, описанный в настоящем документе; например, SEQ ID NO: 2101-2332 или 2353-2398) с нацеливающей последовательностью, и второй вариант гРНК, причем второй вариант гРНК имеет нацеливающую последовательность, комплементарную другой или перекрывающейся части целевой нуклеиновой кислоты по сравнению с нацеливающей последовательностью первой гРНК. В других вариантах осуществления в описании предложены системы из пары для редактирования генов, содержащей вариант белка CasX, первый вариант гРНК, описанный в настоящем документе; например, SEQ ID NO: 2281-2332 или 2353-2398) с нацеливающей последовательностью, и второй вариант гРНК, причем второй вариант гРНК имеет нацеливающую последовательность, комплементарную другой или перекрывающейся части целевой нуклеиновой кислоты по сравнению с нацеливающей последовательностью первой гРНК. В некоторых вариантах осуществления пар для редактирования генов описания CasX : гРНК вариант белка CasX выбран из группы, состоящей из вариантов белков CasX 515, 528, 529, 534-539, 668, 672 и 678 таблицы 3 (SEQ ID NO: 416, 428, 434-439, 567, 570 и 576), а вариант огРНК выбран из группы, состоящей из вариантов 229-237 гРНК таблицы 2 (SEQ ID NO: 2286-2294). В конкретном варианте осуществления пара для редактирования генов содержит вариант белка CasX, выбранный из любого из вариантов белков CasX 668 (SEQ ID NO: 567), 672 (SEQ ID NO: 570) или 676 (SEQ ID NO: 574) и варианта 235 гРНК (SEQ ID NO: 2292).

[0132] В некоторых вариантах осуществления пара для редактирования генов способна связываться друг с другом с образованием рибонуклеопротеинового комплекса (RNP). В других вариантах осуществления пара для редактирования генов связана в рибонуклеопротеиновый комплекс (RNP). В некоторых вариантах осуществления RNP пары для редактирования генов способен связывать и расщеплять двойную цепь целевой нуклеиновой кислоты, включая кодирующую последовательность, комплемент

кодирующей последовательности, некодирующую последовательность и регуляторные элементы. В некоторых вариантах осуществления RNP пары для редактирования генов способен связываться с целевой нуклеиновой кислотой и образовывать один или более одноцепочечных ников в целевой нуклеиновой кислоте. В некоторых вариантах осуществления RNP пары для редактирования генов способен связывать целевую нуклеиновую кислоту, но не способен расщеплять целевую нуклеиновую кислоту.

[0133] В некоторых вариантах осуществления пара для редактирования генов имеет одну или более улучшенных характеристик по сравнению с референсной парой для редактирования генов, содержащей референсный белок CasX с SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3 и референсную гРНК с SEQ ID NO: 5 или SEQ ID NO: 4. В других вариантах осуществления вариант пары для редактирования генов из варианта CasX и варианта гРНК имеет одну или более улучшенных характеристик по сравнению с парой для редактирования генов, содержащей вариант CasX, из которого получен вариант (например, CasX 515, SEQ ID NO: 416), и вариант гРНК, производным которого является данный вариант (например, каркас гРНК 174 (SEQ ID NO: 2238) или 175 (SEQ ID NO: 2239). В вышеуказанных вариантах осуществления одну или более улучшенных характеристик можно проанализировать в анализе *in vitro* в сопоставимых условиях для пары для редактирования генов и референсного CasX и референсной гРНК. Примеры улучшенных характеристик, как описано в настоящем документе, в некоторых вариантах осуществления могут включать в себя стабильность RNP-комплекса CasX : гРНК, повышенную аффинность связывания между CasX и гРНК, улучшенную кинетику образования RNP-комплекса, более высокий процент RNP, компетентных в отношении расщепления, повышенную аффинность связывания RNP с целевой нуклеиновой кислотой, раскручивание целевой нуклеиновой кислоты, повышенную активность редактирования, повышенную эффективность редактирования, повышенную специфичность редактирования для целевой нуклеиновой кислоты, сниженное нецелевое редактирование или расщепление, повышенную активность нуклеазы, повышенную загрузку целевой нуклеиновой кислоты для двухцепочечного расщепления, сниженную загрузку целевой нуклеиновой кислоты для одноцепочечного никирования, повышенное связывание нецелевой цепи ДНК или повышенную устойчивость к нуклеазной активности. В вышеуказанных вариантах осуществления кратность улучшения составляет по меньшей мере примерно 2 раза, по меньшей мере примерно 5 раза, по меньшей мере примерно 10 раз, по меньшей мере примерно 50 раз, по меньшей мере примерно 100 раз, по меньшей мере примерно 500 раз, по меньшей мере примерно 1000 раз, по меньшей мере примерно 5000 раз, по меньшей мере примерно 10 000 раз или по меньшей мере примерно 100 000 раз по сравнению с характеристикой пары референсного белка CasX и референсной гРНК или с характеристикой варианта CasX и варианта гРНК, из которых получена пара для редактирования генов. В других случаях одна или более улучшенных характеристик может иметь кратность улучшения примерно 1,1-100,00 раза, примерно 1,1-10,00 раза, примерно 1,1-1000 раз, примерно 1,1-500 раз, примерно 1,1-100 раз, примерно 1,1-50 раз, примерно

1,1-20 раз, примерно 10-100,00 раза, примерно 10-10,00 раза, примерно 10-1000 раз, примерно 10-500 раз, примерно 10-100 раз, примерно 10-50 раз, примерно 10-20 раз, примерно 2-70 раз, примерно 2-50 раз, примерно 2-30 раз, примерно 2-20 раз, примерно 2-10 раз, примерно 5-50 раз, примерно 5-30 раз, примерно 5-10 раз, примерно 100-100,00 раза, примерно 100-10,00 раза, примерно 100-1000 раз, примерно 100-500 раз, примерно 500-100,00 раза, примерно 500-10,00 раза, примерно 500-1000 раз, примерно 500-750 раз, примерно 1000-100,00 раза, примерно 10 000-100,00 раза, примерно 20-500 раз, примерно 20-250 раз, примерно 20-200 раз, примерно 20-100 раз, примерно 20-50 раз, примерно 50-10 000 раз, примерно 50-1000 раз, примерно 50-500 раз, примерно 50-200 раз или примерно 50-100 раз относительно референсной пары для редактирования генов или характеристики варианта CasX и варианта гРНК, из которых получена пара для редактирования генов. В других случаях одна или более из улучшенных характеристик может иметь кратность улучшения примерно 1,1 раза, 1,2 раза, 1,3 раза, 1,4 раза, 1,5 раза, 1,6 раза, 1,7 раза, 1,8 раза, 1,9 раза, 2 раза, 3 раза, 4 раза, 5 раз, 6 раз, 7 раз, 8 раз, 9 раз, 10 раз, 11 раз, 12 раз, 13 раз, 14 раз, 15 раз, 16 раз, 17 раз, 18 раз, 19 раз, 20 раз, 25 раз, 30 раз, 40 раз, 45 раз, 50 раз, 55 раз, 60 раз, 70 раз, 80 раз, 90 раз, 100 раз, 110 раз, 120 раз, 130 раз, 140 раз, 150 раз, 160 раз, 170 раз, 180 раз, 190 раз, 200 раз, 210 раз, 220 раз, 230 раз, 240 раз, 250 раз, 260 раз, 270 раз, 280 раз, 290 раз, 300 раз, 310 раз, 320 раз, 330 раз, 340 раз, 350 раз, 360 раз, 370 раз, 380 раз, 390 раз, 400 раз, 425 раз, 450 раз, 475 раз или 500 раз или более относительно референсной пары для редактирования генов или относительно характеристики варианта CasX и варианта гРНК, из которых получена пара для редактирования генов.

[0134] В некоторых вариантах осуществления, в которых пара для редактирования генов содержит как вариант CasX, так и вариант гРНК, описанные в настоящем документе, одна или более характеристик пары для редактирования генов улучшаются за пределы того, что может быть достигнуто посредством изменения только белка CasX или только гРНК. В некоторых вариантах осуществления вариант белка CasX и вариант гРНК действуют аддитивно в плане улучшения одной или более характеристик пары для редактирования генов. В некоторых вариантах осуществления вариант белка CasX и вариант гРНК действуют синергетически в плане улучшения одной или более характеристик пары для редактирования генов. В вышеуказанных вариантах осуществления кратность улучшения составляет по меньшей мере примерно 2 раза, по меньшей мере примерно 5 раза, по меньшей мере примерно 10 раз, по меньшей мере примерно 50 раз, по меньшей мере примерно 100 раз, по меньшей мере примерно 500 раз, по меньшей мере примерно 1000 раз, по меньшей мере примерно 5000 раз, по меньшей мере примерно 10 000 раз или по меньшей мере примерно 100 000 раз по сравнению с характеристикой пары референсного белка CasX и референсной гРНК или с характеристикой варианта CasX и варианта гРНК, из которых получена пара для редактирования генов.

[0135] В некоторых вариантах осуществления в описании предложены композиции пар для редактирования генов по любому из вариантов осуществления, описанных в настоящем документе, для применения в качестве лекарственного средства для лечения

субъекта, имеющего заболевание.

[0136] В других вариантах осуществления системы описания содержат один или более вариантов белка CasX, одну или более гидовых нуклеиновых кислот (гРНК) и одну или более нуклеиновых кислот донорной матрицы, содержащей нуклеиновую кислоту, которая кодирует часть гена, причем нуклеиновая кислота донорной матрицы содержит последовательность дикого типа для коррекции мутации или содержит делецию, вставку или мутацию одного или более нуклеотидов по сравнению с геномной нуклеотидной последовательностью дикого типа для нокдауна или нокаута гена.

[0137] В других вариантах осуществления в описании предложены векторы, кодирующие или содержащие вариант CasX, вариант гРНК и необязательно донорные матрицы для получения и/или доставки систем CasX : гРНК. Также в настоящем документе предложены способы получения вариантов белка CasX и вариантов гРНК, а также способы применения вариантов CasX и вариантов гРНК, включая способы редактирования генов и способы лечения. Компоненты в виде вариантов белка CasX и вариантов гРНК систем CasX : гРНК и их признаки, а также способы доставки и способы применения систем более полно описаны ниже.

[0138] Донорные матрицы систем CasX : гРНК разработаны в зависимости от того, используются ли они для коррекции мутаций в целевом гене или вставки трансгена в другой локус в геноме («нокин») или используются для нарушения экспрессии генного продукта, являющегося аномальным; например, донорная матрица содержит одну или более мутаций, которые снижают экспрессию генного продукта или делают белок дисфункциональным («нокдаун» или «нокаут»). В некоторых вариантах осуществления донорная матрица представляет собой одноцепочечную ДНК-матрицу или одноцепочечную РНК-матрицу. В других вариантах осуществления донорная матрица представляет собой двухцепочечную ДНК-матрицу. В некоторых вариантах осуществления системы CasX : гРНК, используемые при редактировании целевой нуклеиновой кислоты, содержат донорную матрицу, содержащую полностью или по меньшей мере частично открытую рамку считывания гена в целевой нуклеиновой кислоте для вставки корректирующей последовательности дикого типа для коррекции дефектного белка. В других случаях донорная матрица содержит полностью или частично ген дикого типа для вставки в другой локус генома для экспрессии генного продукта. В других случаях часть гена может быть вставлена выше (с 5’-стороны) от мутации в целевой нуклеиновой кислоте, причем часть гена в донорной матрице распространяется до С-конца гена или 3'-конца последовательности, имеющей мутацию, что приводит при ее вставке в целевую нуклеиновую кислоту к экспрессии продукта функционального гена.

[0139] В некоторых вариантах осуществления последовательность донорной матрицы содержит негомологичную последовательность, фланкированную двумя областями гомологии с 5’- и 3’-стороны от сайтов разрыва целевой нуклеиновой кислоты (т. е. гомологичные плечи), что облегчает вставку негомологичной последовательности в целевую область, которая может быть опосредована направляемой гомологией репарацией

(HDR) или независимой от гомологии нацеленной интеграцией (HITI). Экзогенная донорная матрица, вставленная методом HITI, может иметь любую длину, например быть относительно короткой последовательностью длиной от 10 до 50 нуклеотидов или более длинной последовательностью длиной более 50-1000 нуклеотидов. Отсутствие гомологии может иметь, например, идентичность последовательности не более чем 20-50% и/или отсутствие специфической гибридизации в условиях сниженной жесткости. В других случаях отсутствие гомологии может дополнительно включать критерий идентичности не более 5, 6, 7, 8 или 9 п. о. В таких случаях применение гомологичных плеч облегчает вставку негомологичной последовательности в сайт (-ы) разрыва, вводимую нуклеазой. В некоторых вариантах осуществления донорный матричный полинуклеотид содержит по меньшей мере примерно 10, по меньшей мере примерно 50, по меньшей мере примерно 100, или по меньшей мере примерно 200, или по меньшей мере примерно 300, или по меньшей мере примерно 400, или по меньшей мере примерно 500, или по меньшей мере примерно 600, или по меньшей мере примерно 700, или по меньшей мере примерно 800, или по меньшей мере примерно 900, или по меньшей мере примерно 1000, или по меньшей мере примерно 10 000, или по меньшей мере примерно 15 000 нуклеотидов. В других вариантах осуществления донорная матрица содержит от по меньшей мере примерно 10 до примерно 15 000 нуклеотидов, или от по меньшей мере примерно 100 до примерно 10 000 нуклеотидов, или от по меньшей мере примерно 400 до примерно 8000 нуклеотидов, или от по меньшей мере примерно 600 до примерно 5000 нуклеотидов, или от по меньшей мере примерно 1000 до примерно 2000 нуклеотидов. Последовательность донорной матрицы может содержать определенные различия в последовательности по сравнению с геномной последовательностью; например сайты рестрикции, нуклеотидные полиморфизмы, маркеры селекции (например, гены резистентности к лекарственному средству, флуоресцентные белки, ферменты и т. д.) и т. д., которые можно применять для оценки успешной вставки донорной нуклеиновой кислоты в сайт расщепления или, в некоторых случаях, применять для других целей (например, чтобы обозначить экспрессию в целевом геномном локусе). Альтернативно эти различия в последовательности могут включать фланкирующие последовательности рекомбинации, такие как FLP, последовательности loxP или т. п., которые могут быть активированы позднее для удаления маркерной последовательности.

III. Гидовые нуклеиновые кислоты систем редактирования генов

[0140] В другом аспекте описание относится к специально разработанным гидовым рибонуклеиновым кислотам (гРНК), содержащим нацеливающую последовательность (также называемую в настоящем документе спейсером), комплементарную последовательности целевой нуклеиновой кислоты гена (и, следовательно, способную гибридизоваться с ней), которые могут находиться в комплексе с нуклеазой CRISPR для геномного редактирования целевой нуклеиновой кислоты в клетке. Предполагается, что в некоторых вариантах осуществления в системах для модификации целевой нуклеиновой кислоты обеспечивается доставка множества гРНК. Например, пару гРНК с

последовательностями, нацеливающимися на разные или перекрывающиеся области последовательности целевой нуклеиновой кислоты, можно применять, когда каждая из них находится в комплексе с нуклеазой CRISPR, чтобы связываться и выполнить расщепление в двух разных или перекрывающихся сайтах внутри гена, которые затем подвергают редактированию методами негомологичного соединения концов (NHEJ), направляемой гомологией репарации (HDR), независимой от гомологии нацеленной интеграции (HITI), соединения концов, обусловленного микрогомологией (MMEJ), отжига одиночных цепей (SSA) или эксцизионной репарации оснований (BER).

[0141] В некоторых вариантах осуществления в описании предложены гРНК, используемые в системах, которые используются в геномном редактировании гена в эукариотической клетке. В конкретном варианте осуществления гРНК систем способна образовывать комплекс с нуклеазой CRISPR; рибонуклеопротеиновый (RNP) комплекс, более подробно описанный ниже.

a. Референсная гРНК и варианты гРНК

[0142] В настоящем документе термин «референсная гРНК» относится к гидовой нуклеиновой кислоте для CRISPR, содержащей последовательность дикого типа встречающейся в природе гРНК. В некоторых вариантах осуществления референсная гРНК описания может быть подвергнута одному или более способам мутагенеза, таким как способы мутагенеза, описанные в настоящем документе в примерах (например, в примере 13, а также в PCT/US20/36506 и WO2020247883A2, включенных в настоящий документ путем ссылки), которые могут включать в себя глубокую мутационную эволюцию (DME), глубокое мутационное сканирование (DMS), ПЦР сниженной точности, кассетный мутагенез, случайный мутагенез, ПЦР со ступенчатой достройкой, перетасовку генов или перестановку доменов, с целью создания одного или более вариантов гидовой нуклеиновой кислоты (называются в настоящем документе «вариантом гРНК») с улучшенными или измененными свойствами относительно референсной гРНК. Варианты гРНК также включают в себя варианты, содержащие одну или более экзогенных последовательностей, например слитых с 5’- или 3’-концом или вставленных внутрь. Активность референсных гРНК или варианта, из которого он получен, можно применять в качестве эталона, с которым сравнивают активность вариантов гРНК, измеряя таким образом улучшения функции или других характеристик вариантов гРНК. В других вариантах осуществления референсную гРНК или вариант гРНК можно подвергать одной или более произвольным специфически нацеленным мутациям для получения варианта гРНК; например, логически сконструированного варианта.

[0143] гРНК описания содержат два сегмента: нацеливающую последовательность и белок-связывающий сегмент. Нацеливающий сегмент гРНК включает нуклеотидную последовательность (называемую взаимозаменяемо гидовой последовательностью, спейсером, нацеливателем или нацеливающей последовательностью), которая комплементарна (и, следовательно, гибридизуется с) конкретной последовательности (целевому сайту) в последовательности целевой нуклеиновой кислоты (например, целевой

оцРНК, целевой оцДНК, цепи двухцепочечной целевой ДНК и т. д.), как более подробно описано ниже. Нацеливающая последовательность гРНК способна связываться с последовательностью целевой нуклеиновой кислоты, в том числе с кодирующей последовательностью, комплементом кодирующей последовательности, с некодирующей последовательностью и регуляторными элементами. Белок-связывающий сегмент (или «активатор», или «белок-связывающая последовательность») взаимодействует с (например, связывается с) белком CasX в комплекс, образуя RNP (более подробно описан ниже). Белок-связывающий сегмент альтернативно называется в настоящем документе «каркасом», и он состоит из нескольких областей, более подробно описанных ниже.

[0144] В случае двойной гидовой РНК (дгРНК) каждый из нацеливающего и активаторного участка имеет дуплекс-образующий сегмент, где дуплекс-образующий сегмент нацеливателя и дуплекс-образующий сегмент активатора обладают комплементарностью друг к другу и гибридизуются друг с другом с образованием двухцепочечного дуплекса (дцРНК-дуплекс для гРНК). Если гРНК представляет собой гРНК, то термин «нацеливатель» или «нацеливающая РНК» применяется в настоящем документе для обозначения крРНК-подобной молекулы (крРНК: «CRISPR-РНК») двойной гидовой РНК CasX (и, следовательно, одинарной гидовой РНК CasX, если «активатор» и «нацеливатель» связаны друг с другом, например, посредством промежуточных нуклеотидов). крРНК имеет 5’-область, которая отжигается с тракрРНК, за которой следуют нуклеотиды нацеливающей последовательности. Таким образом, например, гидовая РНК (дгРНК или огРНК) содержит гидовую последовательность и дуплекс-образующий сегмент крРНК, который также может называться крРНК-повтором. Соответствующая тракрРНК-подобная молекула (активирующий участок) содержит дуплекс-образующий участок нуклеотидов, который формирует другую половину дцРНК-дуплекса белок-связывающего сегмента гидовой РНК. Таким образом, нацеливатель и активатор как соответствующая пара гибридизуются с образованием двойной гидовой РНК, называемой в настоящем документе «двухмолекулярной гРНК», «дгРНК», «бимолекулярной гидовой РНК» или «двухмолекулярной гидовой РНК». Сайт-специфическое связывание и/или расщепление последовательности целевой нуклеиновой кислоты (например, геномной ДНК) белком CasX может происходить в одном или более местоположениях (например, в последовательности целевой нуклеиновой кислоты), определяемом комплементарностью связывания оснований между нацеливающей последовательностью гРНК и последовательностью целевой нуклеиновой кислоты. Таким образом, например, гРНК описания имеет комплементарность последовательности и, следовательно, может гибридизоваться с целевой нуклеиновой кислотой, которая смежна с последовательностью, комплементарной PAM-мотиву TC или PAM-последовательности, такой как ATC, CTC, GTC или TTC. Так как нацеливающая последовательность гидовой последовательности гибридизуется с последовательностью целевой нуклеиновой кислоты, нацеливатель может быть модифицирован пользователем для гибридизации с конкретной последовательностью целевой нуклеиновой кислоты, при условии, что будет учитываться

местоположение PAM-последовательности. Таким образом, в некоторых случаях последовательность нацеливателя может представлять собой комплемент к неприродной последовательности. В других случаях последовательность нацеливателя может представлять собой встречающуюся в природе последовательность, полученную из комплемента к последовательности гена, подлежащего редактированию. В других вариантах осуществления активатор и нацеливатель в гРНК ковалентно связаны друг с другом (а не гибридизуются друг с другом) и представляют собой одну молекулу, которая в настоящем документе называется «одномолекулярной гРНК», «одиночной гидовой НК», «одиночной гидовой РНК», «одномолекулярной гидовой РНК», «мономолекулярной гидовой РНК» или «огРНК». В некоторых вариантах осуществления огРНК включает «активатор» или «нацеливатель» и, таким образом, может представлять собой «активатор-РНК» и «нацеливатель-РНК» соответственно. В некоторых вариантах осуществления гРНК представляет собой молекулу рибонуклеиновой кислоты («гРНК»), а в других вариантах осуществления гРНК представляет собой химеру и содержит как ДНК, так и РНК. В настоящем документе термин «гРНК» охватывает встречающиеся в природе молекулы, а также варианты последовательностей (например, неприродные модифицированные нуклеотиды).

[0145] В совокупности, собранные гРНК описания содержат четыре различных области или домена: триплекс РНК, каркасный стебель, удлиненный стебель и нацеливающую последовательность, которая в вариантах осуществления описания специфична для целевой нуклеиновой кислоты и расположена на 3'-конце гРНК. Триплекс РНК, каркасный стебель и удлиненный стебель в совокупности называются «каркасом» гРНК (гРНК-каркас). Каркасы гРНК описания могут содержать РНК или РНК и ДНК. Каркасы гРНК могут содержать урацилы (U), и один или более урацилов могут быть заменены на тимины (T).

b. Триплекс РНК

[0146] В некоторых вариантах осуществления гидовых РНК, предложенных в настоящем документе, гРНК содержит триплекс РНК, который в некоторых случаях содержит последовательность стеблевой петли UUU--N_X(~ 4-15)--UUU (SEQ ID NO: 241), которая заканчивается AAAG после 2 промежуточных стеблевых петель (каркасной стеблевой петли и удлиненной стеблевой петли), образуя псевдоузел, который также может переходить из триплексного в дуплексный псевдоузел. Последовательность триплекса UU-UUU-AAA образуется как связующее звено между нацеливающей последовательностью, каркасным стеблем и удлиненным стеблем. В примерах гРНК сначала кодируется область UUU-петля-UUU, а затем каркасная стеблевая петля, после этого - удлиненная стеблевая петля, которая связана четверной петлей, а затем AAAG закрывает триплекс, после чего переходит в нацеливающую последовательность.

c. Каркасная стеблевая петля

[0147] В некоторых вариантах осуществления гРНК описания за триплексной областью следует каркасная стеблевая петля. Каркасная стеблевая петля представляет

собой область гРНК, которая связывается с белком CasX (таким как референсный CasX или вариант белка), когда образуется RNP. В некоторых вариантах осуществления каркасная стеблевая петля представляет собой достаточно короткую и стабильную стеблевую петлю и повышает общую стабильность гРНК. В некоторых случаях каркасная стеблевая петля не позволяет вносить много изменений, и для нее необходима некоторая форма РНК-пузырька. В некоторых вариантах осуществления каркасный стебель необходим для функционирования гРНК. Хотя он, возможно, аналогичен связующему стеблю гида Cas9 в роли критической стеблевой петли, каркасный стебель гРНК в некоторых вариантах осуществления имеет обязательное утолщение (РНК-пузырек), которое отличается от многих других стеблевых петель, обнаруживаемых в системах CRISPR/Cas. В некоторых вариантах осуществления присутствие этого утолщения консервативно для гРНК, которые взаимодействуют с разными белками CasX. Пример последовательности каркасной стеблевой петли гРНК содержит последовательность CCAGCGACUAUGUCGUAUGG (SEQ ID NO: 242).

d. Удлиненная стеблевая петля

[0148] В некоторых вариантах осуществления гРНК описания за каркасной стеблевой петлей следует удлиненная стеблевая петля. В некоторых вариантах осуществления удлиненный стебель содержит синтетический слитный конструкт из тракр- и крРНК, который по большей части не связывается белком CasX. В некоторых вариантах осуществления удлиненная стеблевая петля может иметь высокую гибкость. В некоторых вариантах осуществления одиночная гидовая гРНК образована с четырехпетлевым линкером GAAA или линкером GAGAAA между тракр- и крРНК в удлиненной стеблевой петле. В некоторых случаях нацеливатель и активатор в огРНК связаны друг с другом промежуточными нуклеотидами, и линкер может иметь длину от 3 до 20 нуклеотидов. В некоторых вариантах осуществления огРНК описания удлиненный стебель представляет собой большую петлю 32 п. о., которая находится за пределами белка CasX в рибонуклеопротеиновом комплексе. Пример последовательности удлиненной стеблевой петли референсной гРНК содержит последовательность GCGCUUAUUUAUCGGAGAGAAAUCCGAUAAAUAAGAAGC (SEQ ID NO: 15).

e. Нацеливающая последовательность

[0149] В некоторых вариантах осуществления гРНК описания за удлиненной стеблевой петлей следует область, которая образует часть триплекса, а затем нацеливающая последовательность (или «спейсер») на 3’-конце гРНК. Нацеливающая последовательность нацеливает цельный рибонуклеопротеиновый комплекс CasX на конкретную область последовательности целевой нуклеиновой кислоты гена, подлежащего модификации. Таким образом, например, нацеливающие последовательности гРНК описания имеют комплементарность последовательностям и, следовательно, могут гибридизоваться с частью гена в целевой нуклеиновой кислоте в эукариотической клетке (например, эукариотической хромосоме, хромосомной последовательности и т. д.) в качестве компонента RNP, если на расстоянии 1 нуклеотида с 5’-стороны от последовательности

нецелевой цепи, комплементарной целевой последовательности, находится PAM-мотив TC или любая из PAM-последовательностей TTC, ATC, GTC или CTC. Нацеливающую последовательность гРНК можно модифицировать таким образом, чтобы гРНК могла нацеливаться на требуемую последовательность любой требуемой последовательности целевой нуклеиновой кислоты, при условии, что учитывается местоположение PAM-последовательности. В некоторых вариантах осуществления каркас гРНК находится с 5'-стороны нацеливающей последовательности, при этом нацеливающая последовательность находится на 3'-конце гРНК. В некоторых вариантах осуществления последовательность PAM-мотива, распознаваемая нуклеазой RNP, представляет собой TC. В других вариантах осуществления PAM-последовательность, распознаваемая нуклеазой RNP, представляет собой NTC; т. е. ATC, CTC, GTC или TTC.

[0150] В некоторых вариантах осуществления в описании предложена гРНК, в которой нацеливающая последовательность гРНК комплементарна последовательности целевой нуклеиновой кислоты гена, подлежащего модификации. В некоторых вариантах осуществления нацеливающая последовательность гРНК комплементарна последовательности целевой нуклеиновой кислоты гена, содержащей одну или более мутаций по сравнению с последовательностью гена дикого типа для целей редактирования последовательности, содержащей мутации системами CasX : гРНК описания. В таких случаях модификация, выполняемая системой CasX : гРНК, может либо корректировать, либо компенсировать мутацию, или может осуществлять нокдаун или нокаут экспрессии продукта мутантного гена. В других вариантах осуществления нацеливающая последовательность гРНК комплементарна последовательности целевой нуклеиновой кислоты гена дикого типа для целей редактирования последовательности и введения мутации системами CasX : гРНК описания для нокаута или нокаута гена. В некоторых вариантах осуществления нацеливающая последовательность гРНК выполнена специфичной к экзону гена целевой нуклеиновой кислоты. В других вариантах осуществления нацеливающая последовательность гРНК выполнена специфичной к интрону гена целевой нуклеиновой кислоты. В других вариантах осуществления нацеливающая последовательность гРНК выполнена специфичной к месту соединения интрона-экзона гена целевой нуклеиновой кислоты. В других вариантах осуществления нацеливающая последовательность гРНК выполнена специфичной к регуляторному элементу гена целевой нуклеиновой кислоты. В некоторых вариантах осуществления нацеливающая последовательность гРНК выполнена комплементарной к последовательности, содержащей один или более однонуклеотидных полиморфизмов (SNP) в гене целевой нуклеиновой кислоты. Обе из SNP, которые находятся в пределах кодирующей последовательности или в пределах некодирующих последовательностей, входят в объем настоящего описания. В других вариантах осуществления нацеливающая последовательность гРНК выполнена комплементарной последовательности межгенной области гена целевой нуклеиновой кислоты.

[0151] В некоторых вариантах осуществления нацеливающая последовательность

является специфичной к регуляторному элементу, который регулирует экспрессию генного продукта. Такие регуляторные элементы включают, без ограничений, промоторные области, энхансерные области, межгенные области, 5'-нетранслируемые области (5'-UTR), 3'-нетранслируемые области (3'-UTR), консервативные элементы и области, содержащие цис-регуляторные элементы. Предполагается, что промоторная область охватывает нуклеотиды, находящиеся в пределах 5 т. п. о. от точки начала кодирующей последовательности или, в случае энхансерных элементов или консервативных элементов гена, находящиеся в пределах тысяч п. о., сотен тысяч п. о. или даже миллионов п. о. от кодирующей последовательности гена целевой нуклеиновой кислоты. В вышеизложенном мишенями являются молекулы, в которых кодирующий ген мишени предназначен для нокаута или нокдауна таким образом, чтобы генный продукт не экспрессировался или экспрессировался на более низком уровне в клетке.

[0152] В некоторых вариантах осуществления нацеливающая последовательность гРНК содержит от 14 до 35 последовательных нуклеотидов. В некоторых вариантах осуществления нацеливающая последовательность гРНК содержит от 10 до 30 последовательных нуклеотидов. В некоторых вариантах осуществления нацеливающая последовательность имеет 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 или 30 последовательных нуклеотидов. В некоторых вариантах осуществления нацеливающая последовательность гРНК состоит из 20 последовательных нуклеотидов. В некоторых вариантах осуществления нацеливающая последовательность состоит из 19 последовательных нуклеотидов. В некоторых вариантах осуществления нацеливающая последовательность состоит из 18 последовательных нуклеотидов. В некоторых вариантах осуществления нацеливающая последовательность состоит из 17 последовательных нуклеотидов. В некоторых вариантах осуществления нацеливающая последовательность состоит из 16 последовательных нуклеотидов. В некоторых вариантах осуществления нацеливающая последовательность состоит из 15 последовательных нуклеотидов. В некоторых вариантах осуществления нацеливающая последовательность имеет 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 или 30 последовательных нуклеотидов, и нацеливающая последовательность может содержать от 0 до 5, от 0 до 4, от 0 до 3 или от 0 до 2 несовпадений относительно последовательности целевой нуклеиновой кислоты и сохранять достаточную специфичность связывания таким образом, чтобы RNP, содержащий гРНК, имеющую нацеливающую последовательность, был способен образовывать комплементарную связь по отношению к целевой нуклеиновой кислоте.

[0153] В некоторых вариантах осуществления система CasX : гРНК содержит первую гРНК и дополнительно содержит вторую (и необязательно третью, четвертую, пятую или более) гРНК, причем вторая гРНК или дополнительная гРНК имеет нацеливающую последовательность, комплементарную другой или перекрывающейся части последовательности целевой нуклеиновой кислоты по сравнению с нацеливающей последовательностью первой гРНК, так что нацеливание происходит на множество точек в целевой нуклеиновой кислоте, и, например, в целевую нуклеиновую кислоту CasX вводит

множество разрывов. Следует понимать, что в таких случаях вторая или дополнительная гРНК находится в комплексе с дополнительной копией белка CasX. За счет выбора нацеливающих последовательностей гРНК определенные области последовательности целевой нуклеиновой кислот, охватывающие мутацию, могут быть модифицированы или отредактированы с использованием систем CasX : гРНК, описанных в настоящем документе, включая облегчение вставки донорной матрицы или вырезание ДНК между сайтами расщепления в случаях, например, когда происходят мутантные повторы или когда удаление экзона, содержащего мутации, все равно приводит к экспрессии продукта функционального гена.

f. Каркасы гРНК

[0154] Остальные области гРНК, за исключением области нацеливающей последовательности, называются в настоящем документе каркасом. В некоторых вариантах осуществления каркасы гРНК получены из встречающихся в природе последовательностей, описанных ниже в качестве референсной гРНК. В других вариантах осуществления каркасы гРНК представляют собой варианты других вариантов гРНК, в которые вводят мутации, вставки, делеции или замены доменов для придания гРНК желательных свойств.

[0155] В некоторых вариантах осуществления референсная гРНК содержит последовательность, выделенную или полученную из *Deltaproteobacteria*. В некоторых вариантах осуществления последовательность представляет собой последовательность тракрРНК CasX. Пример последовательностей референсных тракрРНК, выделенных или полученных из *Deltaproteobacteria,* могут включать в себя: ACAUCUGGCGCGUUUAUUCCAUUACUUUGGAGCCAGUCCCAGCGACUAUGUCGUA UGGACGAAGCGCUUAUUUAUCGGAGA (SEQ ID NO: 6) и ACAUCUGGCGCGUUUAUUCCAUUACUUUGGAGCCAGUCCCAGCGACUAUGUCGUA UGGACGAAGCGCUUAUUUAUCGG (SEQ ID NO: 7). Примеры последовательностей крРНК, выделенных или полученных из *Deltaproteobacteria,* могут включать в себя последовательность CCGAUAAGUAAAACGCAUCAAAG (SEQ ID NO: 243).

[0156] В некоторых вариантах осуществления референсная гидовая РНК содержит последовательность, выделенную или полученную из *Planctomycetes*. В некоторых вариантах осуществления последовательность представляет собой последовательность тракрРНК. Примеры последовательностей референсных тракрРНК, выделенных или полученных из *Planctomycetes,* могут включать в себя: UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAU GGGUAAAGCGCUUAUUUAUCGGAGA (SEQ ID NO: 8) и

UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACUAUGU CGUAUGGGUAAAGCGCUUAUUUAUCGG (SEQ ID NO: 9). Примеры последовательностей крРНК, выделенные или полученные из *Planctomycetes,* могут содержать последовательность UCUCCGAUAAAUAAGAAGCAUCAAAG (SEQ ID NO: 244).

[0157] В некоторых вариантах осуществления референсная гРНК содержит

последовательность, выделенную или полученную из *Candidatus Sungbacteria*. Примеры последовательностей референсных тракрРНК CasX, выделенных или полученных из *Candidatus Sungbacteria,* могут включать в себя: GUUUACACACUCCCUCUCAUAGGGU (SEQ ID NO: 10), GUUUACACACUCCCUCUCAUGAGGU (SEQ ID NO: 11), UUUUACAUACCCCCUCUCAUGGGAU (SEQ ID NO: 12) и GUUUACACACUCCCUCUCAUGGGGG (SEQ ID NO: 13).

[0158] В таблице 1 представлены тракр-, кр- и каркасные последовательности референсной гРНК. В некоторых вариантах осуществления в описании предложены последовательности вариантов гРНК, в которых гРНК имеет каркас, содержащий последовательность, имеющую по меньшей мере одну модификацию нуклеотидов относительно референсной последовательности гРНК, имеющей последовательность по любой из SEQ ID NO: 4-16 из таблицы 1. Следует понимать, что в тех вариантах осуществления, в которых вектор содержит ДНК, кодирующую последовательность для гРНК, или где гРНК представляет собой химеру РНК и ДНК, тиминовые (T) основания могут быть заменены на урациловые (U) основания в любом из вариантов осуществления последовательности гРНК, описанных в настоящем документе.

Таблица 1. Тракр-, кр- и каркасные последовательности референсной гРНК

SEQ ID NO:	Нуклеотидная последовательность
4	ACAUCUGGCGCGUUUAUUCCAUUACUUUGGAGCCAGUCCCAGCGACUAUGUCGUAUGGACGAAGCGCUUAUUUAUCGGAGAGAAACCGAUAAGUAAAACGCAUCAAAG
5	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAAUCCGAUAAAUAAGAAGCAUCAAAG
6	ACAUCUGGCGCGUUUAUUCCAUUACUUUGGAGCCAGUCCCAGCGACUAUGUCGUAUGGACGAAGCGCUUAUUUAUCGGAGA
7	ACAUCUGGCGCGUUUAUUCCAUUACUUUGGAGCCAGUCCCAGCGACUAUGUCGUAUGGACGAAGCGCUUAUUUAUCGG
8	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGA
9	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGG
10	GUUUACACACUCCCUCUCAUAGGGU
11	GUUUACACACUCCCUCUCAUGAGGU
12	UUUUACAUACCCCCUCUCAUGGGAU
13	GUUUACACACUCCCUCUCAUGGGGG
14	CCAGCGACUAUGUCGUAUGG
15	GCGCUUAUUUAUCGGAGAGAAAUCCGAUAAAUAAGAAGC
16	GGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGA

g. Варианты гРНК

[0159] В другом аспекте описание относится к вариантам гРНК, которые содержат одну или более модификаций относительно референсного каркаса гРНК или получены из другого варианта гРНК. В настоящем документе термин «каркас» относится ко всем частям гРНК, необходимым для функционирования гРНК, за исключением нацеливающей последовательности.

[0160] В некоторых вариантах осуществления вариант гРНК содержит одну или более нуклеотидных замен, вставок, делеций, перестановок или заменных областей относительно референсной последовательности гРНК описания. В некоторых вариантах осуществления мутация может осуществляться в любой области каркаса референсной гРНК с получением варианта гРНК. В некоторых вариантах осуществления каркас варианта последовательности гРНК на по меньшей мере 20%, по меньшей мере 30%, по меньшей мере 40%, по меньшей мере 50%, по меньшей мере 60% или по меньшей мере 70%, по меньшей мере 80%, по меньшей мере 85%, по меньшей мере примерно 90%, по меньшей мере примерно 95%, по меньшей мере примерно 96%, по меньшей мере примерно 97%, по меньшей мере примерно 98% или по меньшей мере примерно 99% идентичен последовательности с SEQ ID NO: 4 или SEQ ID NO: 5. В других вариантах осуществления вариант гРНК содержит одну или более нуклеотидных замен, вставок, делеций, перестановок или заменных областей относительно последовательности варианта гРНК описания. В некоторых вариантах осуществления каркас варианта последовательности гРНК имеет по меньшей мере 50%, по меньшей мере 60% или по меньшей мере 70%, по меньшей мере 80%, по меньшей мере 85%, по меньшей мере примерно 90%, по меньшей мере примерно 95%, по меньшей мере примерно 96%, по меньшей мере примерно 97%, по меньшей мере примерно 98% или по меньшей мере примерно 99% идентичность последовательности с SEQ ID NO: 2238 или SEQ ID NO: 2239.

[0161] В некоторых вариантах осуществления вариант гРНК содержит одно или более изменений нуклеотидов в пределах одной или более областей каркаса референсной гРНК, которые улучшают характеристику референсной гРНК. В других вариантах осуществления вариант гРНК содержит одно или более изменений нуклеотидов в пределах одной или более областей варианта каркаса гРНК, из которого он получен, которые улучшают характеристику относительно этой гРНК. Примеры областей включают триплекс РНК, псевдоузел, каркасную стеблевую петлю и удлиненную стеблевую петлю. В некоторых случаях вариант каркасного стебля дополнительно содержит пузырек. В других случаях вариант каркаса дополнительно содержит область петли триплекса. В других случаях вариант каркаса дополнительно содержит 5'-неструктурированную область. В некоторых вариантах осуществления каркас варианта гРНК содержит каркасную стеблевую петлю, которая имеет по меньшей мере 60% идентичность последовательности, по меньшей мере 70% идентичность последовательности, по меньшей мере 80% идентичность последовательности, по меньшей мере 90% идентичность последовательности, по меньшей мере 95% идентичность последовательности или по меньшей мере 99% идентичность последовательности с SEQ ID NO: 14. В других вариантах осуществления вариант гРНК

содержит каркасную стеблевую петлю, имеющую последовательность CCAGCGACUAUGUCGUAGUGG (SEQ ID NO: 245). В других вариантах осуществления в описании предложен каркас гРНК, содержащий относительно SEQ ID NO: 5 одно или более из замены C18G, вставки G55, делеции U1 и модифицированной удлиненной стеблевой петли, в которой исходная 6-нуклеотидная петля и 13 пар наиболее проксимальных к петле оснований (всего 32 нуклеотидов) заменены на шпильку Uvsx (4-нуклеотидная петля и 5 проксимальных к петле оснований; всего 14 нуклеотидов), и дистальное от петли основание удлиненного стебля преобразовано в стебель с полным спариванием оснований, смежный с новой шпилькой Uvsx, посредством делеции A99 и замены G65U. В вышеуказанном варианте осуществления каркас гРНК представляет собой вариант 174 гРНК и содержит последовательность
ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAG
UGGGUAAAGCUCCCUCUUCGGAGGGAGCAUCAAAG (SEQ ID NO: 2238).

[0162] Все варианты гРНК, которые имеют одну или более улучшенных характеристик или добавление одной или более новых функций при сравнении варианта гРНК с референсной гРНК или вариантом гРНК, подвергнутым мутагенезу для получения нового варианта гРНК, описанного в настоящем документе, находятся в рамках объема описания. Типичный пример такого варианта гРНК представляет собой гид 235 (SEQ ID NO: 2292), конструкция которого описана в разделе «Примеры». В некоторых вариантах осуществления вариант гРНК добавляет новую функцию в RNP, содержащий вариант гРНК. В некоторых вариантах осуществления вариант гРНК имеет улучшенную характеристику, выбранную из: повышенной стабильности; повышенной транскрипции гРНК; повышенной стойкости к нуклеазной активности; повышенной скорости свертывания гРНК; сниженного образования побочных продуктов при свертывании; повышенного продуктивного свертывания; повышенной аффинности связывания с белком CasX; повышенной аффинности связывания с целевой нуклеиновой кислотой при нахождении в комплексе с белком CasX; повышенного редактирования гена при нахождении в комплексе с белком CasX; пповышенной специфичности редактирования целевой нуклеиновой кислоты при нахождении в комплексе с белком CasX; сниженного нецелевого редактирования при нахождении в комплексе с белком CasX и улучшенной способности использовать больший спектр из одной или более PAM-последовательностей, включая ATC, CTC, GTC или TTC, при редактировании целевой нуклеиновой кислоты при нахождении в комплексе с белком CasX, и любой их комбинации. В некоторых случаях одна или более из улучшенных характеристик варианта гРНК имеет кратность повышения от по меньшей мере примерно 1,1 до примерно 100 000 раз относительно референсной гРНК с SEQ ID NO: 4 или SEQ ID NO: 5 или с вариантом 174 или 175 гРНК. В других случаях одна или более улучшенных характеристик варианта гРНК имеет кратность повышения по меньшей мере примерно 1,1, по меньшей мере примерно 10, по меньшей мере примерно 100, по меньшей мере примерно 1000, по меньшей мере примерно 10 000, по меньшей мере примерно 100 000 или более раз относительно референсной гРНК с SEQ ID NO: 4 или SEQ

ID NO: 5 или с вариантом 174 или 175 гРНК. В других случаях одна или более из улучшенных характеристик варианта гРНК имеет кратность повышения примерно 1,1-100,00 раза, примерно 1,1-10,00 раза, примерно 1,1-1000 раз, примерно 1,1-500 раз, примерно 1,1-100 раз, примерно 1,1-50 раз, примерно 1,1-20 раз, примерно 10-100,00 раза, примерно 10-10,00 раза, примерно 10-1000 раз, примерно 10-500 раз, примерно 10-100 раз, примерно 10-50 раз, примерно 10-20 раз, примерно 2-70 раз, примерно 2-50 раз, примерно 2-30 раз, примерно 2-20 раз, примерно 2-10 раз, примерно 5-50 раз, примерно 5-30 раз, примерно 5-10 раз, примерно 100-100,00 раза, примерно 100-10,00 раза, примерно 100-1000 раз, примерно 100-500 раз, примерно 500-100,00 раза, примерно 500-10,00 раза, примерно 500-1000 раз, примерно 500-750 раз, примерно 1000-100,00 раза, примерно 10 000-100,00 раза, примерно 20-500 раз, примерно 20-250 раз, примерно 20-200 раз, примерно 20-100 раз, примерно 20-50 раз, примерно 50-10 000 раз, примерно 50-1000 раз, примерно 50-500 раз, примерно 50-200 раз или примерно 50-100 раз относительно референсной гРНК с SEQ ID NO: 4 или SEQ ID NO: 5 или с вариантом 174 или 175 гРНК. В других случаях одна или более улучшенных характеристик варианта гРНК имеет кратность повышения 1,1 раза, 1,2 раза, 1,3 раза, 1,4 раза, 1,5 раза, 1,6 раза, 1,7 раза, 1,8 раза, 1,9 раза, 2 раза, 3 раза, 4 раза, 5 раз, 6 раз, 7 раз, 8 раз, 9 раз, 10 раз, 11 раз, 12 раз, 13 раз, 14 раз, 15 раз, 16 раз, 17 раз, 18 раз, 19 раз, 20 раз, 25 раз, 30 раз, 40 раз, 45 раз, 50 раз, 55 раз, 60 раз, 70 раз, 80 раз, 90 раз, 100 раз, 110 раз, 120 раз, 130 раз, 140 раз, 150 раз, 160 раз, 170 раз, 180 раз, 190 раз, 200 раз, 210 раз, 220 раз, 230 раз, 240 раз, 250 раз, 260 раз, 270 раз, 280 раз, 290 раз, 300 раз, 310 раз, 320 раз, 330 раз, 340 раз, 350 раз, 360 раз, 370 раз, 380 раз, 390 раз, 400 раз, 425 раз, 450 раз, 475 раз или 500 раз относительно референсной гРНК с SEQ ID NO: 4 или SEQ ID NO: 5 или с вариантом 174 или 175 гРНК.

[0163] В некоторых вариантах осуществления новый вариант гРНК может быть создан посредством воздействия на референсную гРНК или вариант гРНК одним или более способами мутагенеза, такими как способы мутагенеза, описанные в настоящем документе в разделе «Примеры» ниже, которые могут включать в себя глубокую мутационную эволюцию (DME), глубокое мутационное сканирование (DMS), ПЦР сниженной точности, кассетный мутагенез, случайный мутагенез, ПЦР со ступенчатой достройкой, перетасовку генов или перестановку доменов для получения вариантов гРНК описания. Активность референсных гРНК или варианта гРНК, подлежащего мутагенезу, можно использовать в качестве эталона, с которым сравнивают активность вариантов гРНК, измеряя таким образом улучшения функции вариантов гРНК. В других вариантах осуществления референсная гРНК или гРНК может быть подвергнута одной или более намеренным, нацеленным мутациям, заменам или перестановкам доменов для получения варианта гРНК, например, логически сконструированного варианта. Примеры вариантов гРНК, полученных такими способами, описаны в разделе «Примеры», а типичные последовательности каркасов гРНК представлены в таблице 2.

[0164] В некоторых вариантах осуществления вариант гРНК содержит одну или более модификаций по сравнению с каркасной последовательностью референсной гРНК

или варианта гРНК, причем одна или более модификаций выбраны из: по меньшей мере одной нуклеотидной замены в области гРНК; по меньшей мере одной делеции нуклеотида в области гРНК; по меньшей мере одной вставки нуклеотида в области гРНК; полной или частичной замены области гРНК; полной или частичной делеции области гРНК или любой комбинации вышеизложенного. В некоторых случаях модификация представляет собой замену от 1 до 15 последовательных или непоследовательных нуклеотидов в одной или более областях гРНК. В других случаях модификация представляет собой делецию от 1 до 10 последовательных или непоследовательных нуклеотидов в гРНК в одной или более областях. В других случаях модификация представляет собой вставку от 1 до 10 последовательных или непоследовательных нуклеотидов в гРНК в одной или более областях. В других случаях модификация представляет собой замену каркасной стеблевой петли или удлиненной стеблевой петли на последовательность стеблевой петли РНК из гетерологичного источника РНК с проксимальными 5'- и 3'-концами. В некоторых случаях вариант гРНК описания содержит две или более модификации в одной области относительно референсной гРНК или варианта гРНК. В других случаях вариант гРНК описания содержит модификации в двух или более областях. В других случаях вариант гРНК содержит любую комбинацию вышеуказанных модификаций, описанных в данном параграфе.

[0165] В некоторых вариантах осуществления к варианту последовательности гРНК добавляют 5'-G относительно исходной гРНК для экспрессии *in vivo*, так как транскрипция с промотора U6 более эффективна и более единообразна по отношению к начальному сайту, если нуклеотид +1 представляет собой G. В других вариантах осуществления добавляют два 5'-G с получением последовательности варианта гРНК для транскрипции *in vitro* для повышения эффективности получения, так как T7-полимераза имеет сильное предпочтение к G в положении +1 и пурину в положении +2. В некоторых случаях основания 5'-G добавляют к референсным каркасам из таблицы 1. В других случаях основания 5'-G добавляют к вариантам каркасов из таблицы 2.

[0166] В таблице 2 представлены примеры вариантов последовательностей каркаса гРНК. В некоторых вариантах осуществления вариант каркаса гРНК содержит любую из последовательностей SEQ ID NO: 2101-2332 или 2353-2398, указанных в таблице 2, или последовательность, имеющую идентичность последовательности с ней по меньшей мере примерно 50%, по меньшей мере примерно 60%, по меньшей мере примерно 70%, по меньшей мере примерно 80%, по меньшей мере примерно 90%, по меньшей мере примерно 95%, по меньшей мере примерно 95%, по меньшей мере примерно 96%, по меньшей мере примерно 97%, по меньшей мере примерно 98%, по меньшей мере примерно 99%. В некоторых вариантах осуществления вариант каркаса гРНК содержит любую из последовательностей SEQ ID NO: 2238-2332 или 2353-2398 или последовательность, имеющую идентичность последовательности с ней по меньшей мере примерно 50%, по меньшей мере примерно 60%, по меньшей мере примерно 70%, по меньшей мере примерно 80%, по меньшей мере примерно 90%, по меньшей мере примерно 95%, по меньшей мере

примерно 95%, по меньшей мере примерно 96%, по меньшей мере примерно 97%, по меньшей мере примерно 98%, по меньшей мере примерно 99%. В некоторых вариантах осуществления вариант каркаса гРНК содержит любую из последовательностей SEQ ID NO: 2281-2332 или 2353-2398 или последовательность, имеющую идентичность последовательности с ней по меньшей мере примерно 50%, по меньшей мере примерно 60%, по меньшей мере примерно 70%, по меньшей мере примерно 80%, по меньшей мере примерно 90%, по меньшей мере примерно 95%, по меньшей мере примерно 95%, по меньшей мере примерно 96%, по меньшей мере примерно 97%, по меньшей мере примерно 98%, по меньшей мере примерно 99%. Следует понимать, что в тех вариантах осуществления, в которых вектор содержит ДНК, кодирующую последовательность для гРНК, или где гРНК представляет собой химеру РНК и ДНК, тиминовые (T) основания могут быть заменены на урациловые (U) основания в любом из вариантов осуществления последовательности гРНК, описанных в настоящем документе.

Таблица 2. Примеры вариантов последовательностей каркаса гРНК

SEQ ID NO:	**Название**	**НУКЛЕОТИДНАЯ ПОСЛЕДОВАТЕЛЬНОСТЬ ИЛИ ОПИСАНИЕ МОДИФИКАЦИИ**
2101	НД	стабильная репликация фага
2102	НД	Слипающаяся петля_b1
2103	НД	Слипающаяся петля_a
2104	НД	32: шпилька uvsX
2105	НД	PP7
2106	НД	64: тройн. мут., усеченный удлиненный стебель
2107	НД	гиперстабильная четверная петля
2108	НД	C18G
2109	НД	U17G
2110	НД	петля CUUCGG
2111	НД	MS2
2112	НД	-1, A2G, -78, G77U
2113	НД	Q-бета
2114	НД	шпилька 45, 44
2115	НД	U1A
2116	НД	A14C, U17G
2117	НД	петля CUUCGG, модифицированная
2118	НД	Слипающаяся петля_b2
2119	НД	-76 : 78, -83 : 87
2120	НД	-4
2121	НД	усеченный удлиненный стебель
2122	НД	C55
2123	НД	тройн. мут.
2124	НД	-76 : 78
2125	НД	-1 : 5
2126	НД	-83 : 87
2127	НД	=+G28, A82U, -84,
2128	НД	=+51U
2129	НД	-1 : 4, +G5A, +G86,
2130	НД	=+A94

2131	НД	=+G72
2132	НД	укороч. перед, модиф. петля CUUCGG. удлинен. удлин.
2133	НД	A14C
2134	НД	-1 : 3, +G3
2135	НД	=+C45, +U46
2136	НД	модифицированная петля CUUCGG, улучш. начало
2137	НД	-93 : 94
2138	НД	=+U45
2139	НД	-69, -94
2140	НД	-94
2141	НД	модифицированная CUUCGG, минус U в 1-м триплексе
2142	НД	-1 : 4, +C4, A14C, U17G, +G72, -76 : 78, -83 : 87
2143	НД	U1C, -73
2144	НД	Каркас uuCG, стебель uuCG. Перестановка стебля, укорочение t
2145	НД	Каркас uuCG, стебель uuCG. Перестановка стебля
2146	НД	=+G60
2147	НД	без стебля Каркас uuCG
2148	НД	без стебля Каркас uuCG, улучш. начало
2149	НД	Каркас uuCG, стебель uuCG, улучш. начало
2150	НД	Псевдоузлы
2151	НД	Каркас uuCG, стебель uuCG
2152	НД	Каркас uuCG, стебель uuCG, без начала
2153	НД	Каркас uuCG
2154	НД	=+GCUC36
2155	НД	G-квадруплексная теломерная корзина+концы
2156	НД	G-квадруплекс M3q
2157	НД	G-квадруплексная теломерная корзина, без концов
2158	НД	шпилька 45, 44 (старая версия)
2159	НД	Сарцин-рициновая петля
2160	НД	uvsX, C18G
2161	НД	усеченная стеблевая петля, C18G, тройн. мут. (U10C)
2162	НД	корот. фаг. репл., C18G
2163	НД	петля фаг. репл., C18G
2164	НД	=+G18, наложенные на 64
2165	НД	усеченная стеблевая петля, C18G, -1 A2G
2166	НД	петля фаг. репл., C18G, тройн. мут. (U10C)
2167	НД	корот. фаг. репл., C18G, тройн. мут. (U10C)
2168	НД	uvsX, тройн. мут. (U10C)
2169	НД	усеченная стеблевая петля
2170	НД	=+A17, наложенные на 64
2171	НД	геномный рибозим 3’ HDV
2172	НД	петля фаг. репл., тройн. мут. (U10C)
2173	НД	-79 : 80
2174	НД	корот. фаг. репл., тройн. мут. (U10C)
2175	НД	экстраусеченная стеблевая петля
2176	НД	U17G, C18G
2177	НД	корот. фаг. репл.
2178	НД	uvsX, C18G, -1 A2G
2179	НД	uvsX, C18G, тройн. мут. (U10C), -1 A2G, HDV -99 G65U
2180	НД	Антигеномный рибозим 3’ HDV

2181	НД	uvsX, C18G, тройн. мут. (U10C), -1 A2G, HDV AA(98:99)C
2182	НД	3’ рибозим HDV (Lior Nissim, Timothy Lu)
2183	НД	TAC(1:3)GA, наложенный на 64
2184	НД	uvsX, -1 A2G
2185	НД	усеченная стеблевая петля, C18G, тройн. мут. (U10C), -1 A2G, HDV -99 G65U
2186	НД	корот. фаг. репл., C18G, тройн. мут. (U10C), -1 A2G, HDV -99 G65U
2187	НД	вирусный рибозим Hammerhead дикого типа 3’ sTRSV
2188	НД	корот. фаг. репл., C18G, -1 A2G
2189	НД	корот. фаг. репл., C18G, тройн. мут. (U10C), -1 A2G, 3’ геномный HDV
2190	НД	петля фаг. репл., C18G, тройн. мут. (U10C), -1 A2G, HDV -99 G65U
2191	НД	3’ рибозим HDV (Owen Ryan, Jamie Cate)
2192	НД	петля фаг. репл., C18G, -1 A2G
2193	НД	0,14
2194	НД	-78, G77U
2195	НД	НД
2196	НД	корот. фаг. репл., -1 A2G
2197	НД	усеченная стеблевая петля, C18G, тройн. мут. (U10C), -1 A2G
2198	НД	-1, A2G
2199	НД	усеченная стеблевая петля, тройн. мут. (U10C), -1 A2G
2200	НД	uvsX, C18G, тройн. мут. (U10C), -1 A2G
2201	НД	петля фаг. репл., -1 A2G
2202	НД	петля фаг. репл., тройн. мут. (U10C), -1 A2G
2203	НД	петля фаг. репл., C18G, тройн. мут. (U10C), -1 A2G
2204	НД	усеченная стеблевая петля, C18G
2205	НД	uvsX, тройн. мут. (U10C), -1 A2G
2206	НД	усеченная стеблевая петля, -1 A2G
2207	НД	корот. фаг. репл., тройн. мут. (U10C), -1 A2G
2208	НД	Рибозим 5’ HDV (Owen Ryan, Jamie Cate)
2209	НД	Геномный рибозим 5’ HDV
2210	НД	усеченная стеблевая петля, C18G, тройн. мут. (U10C), -1 A2G, HDV AA(98:99)C
2211	НД	Рибозим 5’ env25 pistol (с добавленной петлей CUUCGG)
2212	НД	антигеномный рибозим 5’ HDV
2213	НД	Рибозим 3’ Hammerhead (Lior Nissim, Timothy Lu), гидовый каркас, scar
2214	НД	=+A27, наложенные на 64
2215	НД	Рибозим 5’ Hammerhead (Lior Nissim, Timothy Lu), меньший scar
2216	НД	петля фаг. репл., C18G, тройн. мут. (U10C), -1 A2G, HDV AA(98:99)C
2217	НД	-27, наложенные на 64
2218	НД	3’ Hatchet
2219	НД	Рибозим 3’ Hammerhead (Lior Nissim, Timothy Lu)
2220	НД	5’ Hatchet
2221	НД	Рибозим 5’ HDV (Lior Nissim, Timothy Lu)
2222	НД	Рибозим 5’ Hammerhead (Lior Nissim, Timothy Lu)
2223	НД	Минимальный рибозим Hammerhead 3’ HH15
2224	НД	5’-рекрутирующий мотив RBMX
2225	НД	Рибозим 3’ Hammerhead (Lior Nissim, Timothy Lu), меньший scar
2226	НД	Рибозим 3’ env25 pistol (с добавленной петлей CUUCGG)

2227	НД	3’ Env-9 Twister
2228	НД	=+AUUAUCUCAUUACU25
2229	НД	5’ Env-9 Twister
2230	НД	3’ Twisted Sister 1
2231	НД	без стебля
2232	НД	Минимальный рибозим Hammerhead 5’ HH15
2233	НД	Рибозим 5’ Hammerhead (Lior Nissim, Timothy Lu), гидовый каркас, scar
2234	НД	5’ Twisted Sister 1
2235	НД	Вирусный рибозим Hammerhead дикого типа 5’ sTRSV
2236	НД	148: =+G55, наложенные на 64
2237	НД	158: 103+148(+G55) -99, G65U
2238	174	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAG CAUCAAAG
2239	175	ACUGGCGCCUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCC GUAAGAAGCAUCAAAG
2240	176	GCUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAG CAUCAAAG
2241	177	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCUCCCUCUUCGGAGGGAGC AUCAAAG
2242	181	ACUGGCGCCUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCC GUAAGAAGCAUCAAAG
2243	182	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCC GUAAGAAGCAUCAAAG
2244	183	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCGCUUACGGACUUCGGUC CGUAAGAAGCAUCAAAG
2245	184	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUUGGGUAAAGCUCCCUCUUCGGAGGGAG CAUCAAAG
2246	185	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUUGGGUAAAGCGCUUACGGACUUCGGUC CGUAAGAAGCAUCAAAG
2247	186	ACUGGCGCCUUUAUCAUCAUUACUUUGAGAGCCAUCACCAG CGACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUC CGUAAGAAGCAUCAAAG
2248	187	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCGCCCUCUUCGGAGGGAA GCAUCAAAG
2249	188	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCACAUGAGGAUCACCC AUGUGAGCAUCAAAG
2250	189	ACUGGCACUUUUACCUGAUUACUUUGAGAGCCAACACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAG CAUCAAAG

2251	190	ACUGGCACUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAG CAUCAAAG
2252	191	ACUGGCCCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAG CAUCAAAG
2253	192	ACUGGCGCUUUUACCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAG CAUCAAAG
2254	193	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAACACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAG CAUCAAAG
2255	195	ACUGGCACCUUUACCUGAUUACUUUGAGAGCCAACACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCC GUAAGAAGCAUCAAAG
2256	196	ACUGGCACCUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCC GUAAGAAGCAUCAAAG
2257	197	ACUGGCCCCUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCC GUAAGAAGCAUCAAAG
2258	198	ACUGGCGCCUUUAUCUGAUUACUUUGAGAGCCAACACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCC GUAAGAAGCAUCAAAG
2259	199	GCUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAG CAUCAAAG
2260	200	GACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCA GCGACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGG AGCAUCAAAG
2261	201	ACUGGCGCCUUUAUCUGAUUACUUUGGAGAGCCAUCACCAG CGACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGA GCAUCAAAG
2262	202	ACUGGCGCAUUUAUCUGAUUACUUUGUGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAG CAUCAAAG
2263	203	ACUGGCGCCUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAG CAUCAAAG
2264	204	ACUGGCGCUUUUAUCUGAUUACUUUGGAGAGCCAUCACCA GCGACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGG AGCAUCAAAG
2265	205	ACUGGCGCAUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAG CAUCAAAG
2266	206	ACUGGCGCUUUUAUCUGAUUACUUUGUGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAG CAUCAAAG
2267	207	ACUGGCGCUUUUAUUCUGAUUACUUUGAGAGCCAUCACCA GCGACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGG AGCAUCAAAG
2268	208	ACGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCG

		ACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGC AUCAAAG
2269	209	ACUGGCGCUUUUAUAUGAUUACUUUGAGAGCCAUCACCAG CGACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGA GCAUCAAAG
2270	210	ACUGGCGCUUUUAUCUUGAUUACUUUGAGAGCCAUCACCA GCGACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGG AGCAUCAAAG
2271	211	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAGCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAG CAUCAAAG
2272	212	ACUGGCGCUGUUAUCUGAUUACUUCGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAG CAUCGAAG
2273	213	ACUGGCGCUCUUAUCUGAUUACUUCGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAG CAUCGAAG
2274	214	ACUGGCGCUUGUAUCUGAUUACUCUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAG CAUCAGAG
2275	215	ACUGGCGCUUCUAUCUGAUUACUCUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAG CAUCAGAG
2276	216	ACUGGCGCUUUGAUCUGAUUACCUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAG CAUCAAGG
2277	217	ACUGGCGCUUUCAUCUGAUUACCUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAG CAUCAAGG
2278	218	ACUGGCGCUGUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAG CAUCAAAG
2279	219	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAG CAUCGAAG
2280	220	ACUGGCGCUUUUAUCUGAUUACUUCGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAG CAUCAAAG
2281	221	ACUGGCACUUCUAUCUGAUUACUCUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCCGCUUACGGACUUCGGUC CGUAAGAGGCAUCAGAG
2282	222	ACUGGCACUUCUAUCUGAUUACUCUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAG CAUCAGAG
2283	223	ACUGGCACCUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCCGCUUACGGACUUCGGUC CGUAAGAGGCAUCAAAG
2284	224	ACUGGCACUUGUAUCUGAUUACUCUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCCGCUUACGGACUUCGGUC CGUAAGAGGCAUCAGAG
2285	225	ACUGGCACUUGUAUCUGAUUACUCUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAG

		CAUCAGAG
2332	226	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGC GUCAAUGACGCUGACGGUACAGGCCAGACAAUUAUUGUCU GGUAUAGUGCAGCAUCAAAG
2286	229	ACUGGCACUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCC GUAAGAAGCAUCAAAG
2287	230	ACUGGCACUUCUAUCUGAUUACUCUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCC GUAAGAAGCAUCAGAG
2288	231	ACUGGCGCUUCUAUCUGAUUACUCUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCCGCUUACGGACUUCGGUC CGUAAGAGGCAUCAGAG
2289	232	ACUGGCACUUCUAUCUGAUUACUCUGAGCGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCCGCUUACGGACUUCGGUC CGUAAGAGGCAUCAGAG
2290	233	ACUGGCGCUUCUAUCUGAUUACUCUGAGCGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCCGCUUACGGACUUCGGUC CGUAAGAGGCAUCAGAG
2291	234	ACUGGCGCUUCUAUCUGAUUACUCUGAGCGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCCUUACGGACUUCGGUC CGUAAGGAGCAUCAGAG
2292	235	ACUGGCGCUUCUAUCUGAUUACUCUGAGCGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCCGCUUACGGACUUCGGU CCGUAAGAGGCAUCAGAG
2293	236	ACGGGACUUUCUAUCUGAUUACUCUGAAGUCCCUCACCAGC GACUAUGUCGUAUGGGUAAAGCCGCUUACGGACUUCGGUC CGUAAGAGGCAUCAGAG
2294	237	ACCUGUAGUUCUAUCUGAUUACUCUGACUACAGUCACCAGC GACUAUGUCGUAUGGGUAAAGCCGCUUACGGACUUCGGUC CGUAAGAGGCAUCAGAG
2295	238	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACGGUGGGCGCAGC UUCGGCUGACGGUACACCGUGCAGCAUCAAAG
2296	239	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACGGUGGGCGCAGC UUCGGCUGACGGUACACCGGUGGGCGCAGCUUCGGCUGACG GUACACCGUGCAGCAUCAAAG
2297	240	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACGGUGGGCGCAGC UUCGGCUGACGGUACACCGGUGGGCGCAGCUUCGGCUGACG GUACACCGGUGGGCGCAGCUUCGGCUGACGGUACACCGUGC AGCAUCAAAG
2298	241	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACGGUGGGCGCAGC UUCGGCUGACGGUACACCGGUGGGCGCAGCUUCGGCUGACG GUACACCGGUGGGCGCAGCUUCGGCUGACGGUACACCGGUG GGCGCAGCUUCGGCUGACGGUACACCGUGCAGCAUCAAAG
2299	242	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACGGUGGGCGCAGC

		UUCGGCUGACGGUACACCGGUGGGCGCAGCUUCGGCUGACG GUACACCGGUGGGCGCAGCUUCGGCUGACGGUACACCGGUG GGCGCAGCUUCGGCUGACGGUACACCGGUGGGCGCAGCUUC GGCUGACGGUACACCGUGCAGCAUCAAAG
2300	243	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACCUAGCGGAGGCU AGGUGCAGCAUCAAAG
2301	244	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACCUCGGCUUGCUG AAGCGCGCACGGCAAGAGGCGAGGUGCAGCAUCAAAG
2302	245	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACCUCUCUCGACGCA GGACUCGGCUUGCUGAAGCGCGCACGGCAAGAGGCGAGGG GCGGCGACUGGUGAGUACGCCAAAAAUUUUGACUAGCGGA GGCUAGAAGGAGAGAGGUGCAGCAUCAAAG
2303	246	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACGGUGCCCGUCUG UUGUGUCGAGAGACGCCAAAAAUUUUGACUAGCGGAGGCU AGAAGGAGAGAGAUGGGUGCCGUGCAGCAUCAAAG
2304	247	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACAUGGAGAGGAGA UGUGCAGCAUCAAAG
2305	248	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACAUGGAGAUGUGC AGCAUCAAAG
2306	249	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUUGGGCGCAGCGUCAAU GACGCUGACGGUACAAGCAUCAAAG
2307	250	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGC GUCAAUGACGCUGACGGUACAGGCCACAUGAGGAUCACCCA UGUGGUAUAGUGCAGCAUCAAAG
2308	251	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGC UCAUGAGGAUCACCCAUGAGCUGACGGUACAGGCCACAUGA GGAUCACCCAUGUGGUAUAGUGCAGCAUCAAAG
2309	252	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGC GUCAAUGACGCUGACGGUACAGGCCACAUGGCAGUCGUAAC GACGCGGGUGGUAUAGUGCAGCAUCAAAG
2310	253	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGC AAACAUGGCAGUCCUAAGGACGCGGGUUUUGCUGACGGUA CAGGCCACAUGGCAGUCGUAACGACGCGGGUGGUAUAGUG CAGCAUCAAAG
2311	254	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGA CAUGGCAGUCGUAACGACGCGGGUCUGACGGUACAGGCCAC AUGAGGAUCACCCAUGUGGUAUAGUGCAGCAUCAAAG
2312	255	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAAGGAGUUUAU AUGGAAACCCUUAGUGCAGCAUCAAAG

2313	256	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCAGGAAGCACUAUGGG CGCAGCGUCAAUGACGCUGACGGUACAGGCCAGACAAUUAU UGUCUGGUAUAGUGCAGCAGCAGAACAAUUUGCUGAGGGC UAUUGAGGCGCAACAGCAUCUGUUGCAACUCACAGUCUGG GGCAUCAAGCAGCUCCAGGCAAGAAUCCUGAGCAUCAAAG
2314	257	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACGCCCUGAAGAAG GGCGUGCAGCAUCAAAG
2315	258	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACGGCUCGUGUAGC UCAUUAGCUCCGAGCCGUGCAGCAUCAAAG
2316	259	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACCCGUGUGCAUCC GCAGUGUCGGAUCCACGGGUGCAGCAUCAAAG
2317	260	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACGGAAUCCAUUGC ACUCCGGAUUUCACUAGGUGCAGCAUCAAAG
2318	261	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACAUGCAUGUCUAA GACAGCAUGUGCAGCAUCAAAG
2319	262	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACAAAACAUAAGGA AAACCUAUGUUGUGCAGCAUCAAAG
2320	263	ACUGGCGCUUCUAUCUGAUUACUCUGAGCGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCCGCUUACGGACUAUGGG CGCAGCGUCAAUGACGCUGACGGUACAGGCCAGACAAUUAU UGUCUGGUAUAGUCCGUAAGAGGCAUCAGAG
2321	264	ACUGGCGCUUCUAUCUGAUUACUCUGAGCGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCCGCUUACGGGUGGGCGC AGCGUCAAUGACGCUGACGGUACAGGCCAGACAAUUAUUG UCUGGUACCCGUAAGAGGCAUCAGAG
2322	265	ACUGGCGCUUCUAUCUGAUUACUCUGAGCGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCCGCUUACGGUAUGGGCG CAGCGUCAAUGACGCUGACGGUACAGGCCACAUGAGGAUCA CCCAUGUGGUAUACCGUAAGAGGCAUCAGAG
2323	266	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUAUGGGCGCAGCG UCAAUGACGCUGACGGUACAGGCCACAUGAGGAUCACCCAU GUGGUAUAGGGAGCAUCAAAG
2324	267	ACUGGCGCUUCUAUCUGAUUACUCUGAGCGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCCGCUUACGGUAUGGGCG CAGCUCAUGAGGAUCACCCAUGAGCUGACGGUACAGGCCAC AUGAGGAUCACCCAUGUGGUAUACCGUAAGAGGCAUCAGA G
2325	268	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUAUGGGCGCAGCU CAUGAGGAUCACCCAUGAGCUGACGGUACAGGCCACAUGAG GAUCACCCAUGUGGUAUAGGGAGCAUCAAAG
2326	269	ACUGGCGCUUCUAUCUGAUUACUCUGAGCGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCCGCUUACGGUAUGGGCG CAGCGUCAAUGACGCUGACGGUACAGGCCACAUGGCAGUCG

		UAACGACGCGGGUGGUAUACCGUAAGAGGCAUCAGAG
2327	270	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUAUGGGCGCAGCG UCAAUGACGCUGACGGUACAGGCCACAUGGCAGUCGUAACG ACGCGGGUGGUAUAGGGAGCAUCAAAG
2328	271	ACUGGCGCUUCUAUCUGAUUACUCUGAGCGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCCGCUUACGGUAUGGGCG CAGCAAACAUGGCAGUCCUAAGGACGCGGGUUUUGCUGAC GGUACAGGCCACAUGGCAGUCGUAACGACGCGGGUGGUAU ACCGUAAGAGGCAUCAGAG
2329	272	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUAUGGGCGCAGCA AACAUGGCAGUCCUAAGGACGCGGGUUUUGCUGACGGUAC AGGCCACAUGGCAGUCGUAACGACGCGGGUGGUAUAGGGA GCAUCAAAG
2330	273	ACUGGCGCUUCUAUCUGAUUACUCUGAGCGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCCGCUUACGGUAUGGGCG CAGACAUGGCAGUCGUAACGACGCGGGUCUGACGGUACAG GCCACAUGAGGAUCACCCAUGUGGUAUACCGUAAGAGGCA UCAGAG
2331	274	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUAUGGGCGCAGAC AUGGCAGUCGUAACGACGCGGGUCUGACGGUACAGGCCACA UGAGGAUCACCCAUGUGGUAUAGGGAGCAUCAAAG
2353	275	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGC ACCUGAGGAUCACCCAGGUGCUGACGGUACAGGCCACCUGA GGAUCACCCAGGUGGUAUAGUGCAGCAUCAAAG
2354	276	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGC GCAUGAGGAUCACCCAUGCGCUGACGGUACAGGCCGCAUGA GGAUCACCCAUGCGGUAUAGUGCAGCAUCAAAG
2355	277	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGC GCCUGAGGAUCACCCAGGCGCUGACGGUACAGGCCGCCUGA GGAUCACCCAGGCGGUAUAGUGCAGCAUCAAAG
2356	278	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGC GCCUGAGCAUCAGCCAGGCGCUGACGGUACAGGCCGCCUGA GCAUCAGCCAGGCGGUAUAGUGCAGCAUCAAAG
2357	279	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGC ACAUGAGCAUCAGCCAUGUGCUGACGGUACAGGCCACAUGA GCAUCAGCCAUGUGGUAUAGUGCAGCAUCAAAG
2358	280	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGC ACAUGAGUAUCAACCAUGUGCUGACGGUACAGGCCACAUG AGUAUCAACCAUGUGGUAUAGUGCAGCAUCAAAG
2359	281	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGC ACAUGAGAAUCAGCCAUGUGCUGACGGUACAGGCCACAUG AGAAUCAGCCAUGUGGUAUAGUGCAGCAUCAAAG

2360	282	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGC CCUUGAGGAUCACCCAUGUGCUGACGGUACAGGCCCCUUGA GGAUCACCCAUGUGGUAUAGUGCAGCAUCAAAG
2361	283	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGC ACUUGAGGAUCACCCAUGUGCUGACGGUACAGGCCACUUGA GGAUCACCCAUGUGGUAUAGUGCAGCAUCAAAG
2362	284	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGC ACCUGAGGAUCACCCAUGUGCUGACGGUACAGGCCACCUGA GGAUCACCCAUGUGGUAUAGUGCAGCAUCAAAG
2363	285	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGC ACAUGAGGAUCACCUAUGUGCUGACGGUACAGGCCACAUG AGGAUCACCUAUGUGGUAUAGUGCAGCAUCAAAG
2364	286	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGC ACAUUAGGAUCACCAAUGUGCUGACGGUACAGGCCACAUU AGGAUCACCAAUGUGGUAUAGUGCAGCAUCAAAG
2365	287	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGC ACAUUAGGAUCACCGAUGUGCUGACGGUACAGGCCACAUU AGGAUCACCGAUGUGGUAUAGUGCAGCAUCAAAG
2366	288	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGC ACAUUAGGAUCACCUAUGUGCUGACGGUACAGGCCACAUU AGGAUCACCUAUGUGGUAUAGUGCAGCAUCAAAG
2367	289	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGC ACAUGAGGAUUACCCAUGUGCUGACGGUACAGGCCACAUG AGGAUUACCCAUGUGGUAUAGUGCAGCAUCAAAG
2368	290	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGC ACAUGAGGAUAACCCAUGUGCUGACGGUACAGGCCACAUG AGGAUAACCCAUGUGGUAUAGUGCAGCAUCAAAG
2369	291	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGC ACAUGAGGAUGACCCAUGUGCUGACGGUACAGGCCACAUG AGGAUGACCCAUGUGGUAUAGUGCAGCAUCAAAG
2370	292	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGC ACAUGAGGACCACCCAUGUGCUGACGGUACAGGCCACAUGA GGACCACCCAUGUGGUAUAGUGCAGCAUCAAAG
2371	293	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGC AGAUGAGGAUCACCCAUGGGCUGACGGUACAGGCCAGAUG AGGAUCACCCAUGGGGUAUAGUGCAGCAUCAAAG
2372	294	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGC ACAUGGGGAUCACCCAUGUGCUGACGGUACAGGCCACAUGG GGAUCACCCAUGUGGUAUAGUGCAGCAUCAAAG

2373	295	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGC ACAUGAGGAUCACCCAUGUGCUGACGGUACAGGCCACAUGA GGAUCACCCAUGUGGUAUAGUGCAGCAUCAAAG
2374	296	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCACCUGAGGAUCACCC AGGUGAGCAUCAAAG
2375	297	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCGCAUGAGGAUCACCC AUGCGAGCAUCAAAG
2376	298	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCGCCUGAGGAUCACCC AGGCGAGCAUCAAAG
2377	299	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCGCCUGAGCAUCAGCC AGGCGAGCAUCAAAG
2378	300	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCACAUGAGCAUCAGCC AUGUGAGCAUCAAAG
2379	301	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCACAUGAGUAUCAACC AUGUGAGCAUCAAAG
2380	302	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCACAUGAGAAUCAGCC AUGUGAGCAUCAAAG
2381	303	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUUGAGGAUCACCC AUGUGAGCAUCAAAG
2382	304	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCACUUGAGGAUCACCC AUGUGAGCAUCAAAG
2383	305	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCACCUGAGGAUCACCC AUGUGAGCAUCAAAG
2384	306	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCACAUGAGGAUCACCU AUGUGAGCAUCAAAG
2385	307	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCACAUUAGGAUCACCA AUGUGAGCAUCAAAG
2386	308	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCACAUUAGGAUCACCG AUGUGAGCAUCAAAG
2387	309	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCACAUUAGGAUCACCU AUGUGAGCAUCAAAG
2388	310	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCACAUGAGGAUUACCC AUGUGAGCAUCAAAG
2389	311	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCACAUGAGGAUAACCC AUGUGAGCAUCAAAG

2390	312	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCACAUGAGGAUGACCC AUGUGAGCAUCAAAG
2391	313	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCACAUGAGGACCACCC AUGUGAGCAUCAAAG
2392	314	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCAGAUGAGGAUCACCC AUGGGAGCAUCAAAG
2393	315	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCACAUGGGGAUCACCC AUGUGAGCAUCAAAG
2394	317	ACUGGCGCUUCUAUCUGAUUACUCUGAGCGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCACAUGAGGAUCACCC AUGUGAGCAUCAGAG
2395	318	ACUGGCGCUUCUAUCUGAUUACUCUGAGCGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGC GUCAAUGACGCUGACGGUACAGGCCACAUGAGGAUCACCCA UGUGGUAUAGUGCAGCAUCAGAG
2396	319	ACUGGCGCUUCUAUCUGAUUACUCUGAGCGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGC UCAUGAGGAUCACCCAUGAGCUGACGGUACAGGCCACAUGA GGAUCACCCAUGUGGUAUAGUGCAGCAUCAGAG
2397	320	ACUGGCGCUUCUAUCUGAUUACUCUGAGCGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGCGCAGA CAUGGCAGUCGUAACGACGCGGGUCUGACGGUACAGGCCAC AUGAGGAUCACCCAUGUGGUAUAGUGCAGCAUCAGAG
2398	321	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUGCACUAUGGGGCCACA UGAGGAUCACCCAUGUGGUGUACAGCGCAGCGUCAAUGAC GCUGACGAUAGUGCAGCAUCAAAG

[0167] В некоторых вариантах осуществления вариант огРНК содержит одно или более дополнительных модификаций в последовательности SEQ ID NO:2238, SEQ ID NO:2239, SEQ ID NO:2240, SEQ ID NO:2241, SEQ ID NO:2243, SEQ ID NO:2256, SEQ ID NO:2274, SEQ ID NO:2275, SEQ ID NO:2279, SEQ ID NO:2281, SEQ ID NO: 2285, SEQ ID NO: 2289, SEQ ID NO: 2292 или SEQ ID NO: 2308 из таблицы 2.

[0168] В некоторых вариантах осуществления варианты гРНК описания вариант гРНК содержит по меньшей мере одну модификацию по сравнению с каркасом референсного гида с SEQ ID NO:5, причем по меньшей мере одна модификация выбрана из одного или более из: (a) замены C18G в петле триплекса; (b) вставки G55 в пузырьке стебля; (c) делеции U1; (d) модификации удлиненной стеблевой петли, в которой (i) петля из 6 нукл. и 13 проксимальных к петле пар оснований заменены шпилькой Uvsx и (ii) делеция A99 и замена G65U приводят к образованию дистального от петли основания, которое является полностью спаренным.

[0169] В некоторых вариантах осуществления вариант гРНК содержит экзогенную стеблевую петлю, имеющую длинную некодирующую РНК (днкРНК). В настоящем документе термин «днкРНК» относится к некодирующей РНК длиной более чем примерно

200 п. о. В некоторых вариантах осуществления 5’- и 3’-концы экзогенной стеблевой петли представляют собой спаренные основания; т. е. взаимодействуют с образованием области дуплексной РНК. В некоторых вариантах осуществления 5’- и 3’-концы экзогенной стеблевой петли представляют собой спаренные основания, и одна или более областей между 5’- и 3’-концами экзогенной стеблевой петли не являются спаренными основаниями, образуя петлю.

[0170] В некоторых вариантах осуществления в описании предложены варианты гРНК с нуклеотидными модификациями относительно референсной гРНК, включающими: (a) замену 1-15 последовательных или непоследовательных нуклеотидов в варианте гРНК в одной или более областях; (b) делецию от 1-10 последовательных или непоследовательных нуклеотидов в варианте гРНК в одной или более областях; (c) вставку 1-10 последовательных или непоследовательных нуклеотидов в варианте гРНК в одной или более областях; (d) замену каркасной стеблевой петли или удлиненной стеблевой петли на последовательность стеблевой петли РНК из гетерологичного источника РНК с проксимальными 5’- и 3’-концами; или любую комбинацию пп. (a)-(d). Любые из замен, вставок и делеций, описанных в настоящем документе, можно комбинировать для создания варианта гРНК описания. Например, вариант гРНК может содержать по меньшей мере одну замену и по меньшей мере одну делецию относительно референсной гРНК, по меньшей мере одну замену и по меньшей мере одну вставку относительно референсной гРНК, по меньшей мере одну вставку и по меньшей мере одну делецию относительно референсной гРНК или по меньшей мере одну замену, одну вставку и одну делецию относительно референсной гРНК.

[0171] В некоторых вариантах осуществления вариант огРНК описания содержит одну или более модификаций последовательности ранее созданного варианта, причем ранее созданный вариант сам по себе служит в качестве последовательности, подлежащей модификации. В некоторых случаях одну или модификаций вводят в область псевдоузла каркаса. В других случаях одну или модификаций вводят в триплексную область каркаса. В других случаях одну или модификаций вводят в область пузырька каркаса. В других случаях одну или модификаций вводят в область удлиненного стебля каркаса. В других случаях одну из модификаций вводят в две или более из вышеуказанных областей. Такие модификации могут включать в себя вставку, делецию или замену одного или более нуклеотидов в вышеуказанных областях или любую их комбинацию. Примеры способов создания и оценки модификаций описаны в примере 15.

[0172] В некоторых вариантах осуществления вариант огРНК содержит одну или более модификаций последовательности SEQ ID NO: 2238, SEQ ID NO: 2239, SEQ ID NO: 2240, SEQ ID NO: 2241, SEQ ID NO:2241, SEQ ID NO:2274, SEQ ID NO:2275, SEQ ID NO: 2279 или SEQ ID NO: 2285, SEQ ID NO: 2289, SEQ ID NO: 2292 или SEQ ID NO: 2308.

[0173] В примерах осуществления вариант гРНК содержит одну или более модификаций относительно варианта каркаса 174 гРНК (SEQ ID NO:2238), причем полученный вариант гРНК демонстрирует улучшенную функциональную характеристику

по сравнению с исходным 174 при оценке в анализе *in vitro* или in vivo в сопоставимых условиях. В других примерах осуществления вариант гРНК содержит одну или более модификаций относительно варианта каркаса 175 гРНК (SEQ ID NO:2239), причем полученный вариант гРНК демонстрирует улучшенную функциональную характеристику по сравнению с исходным 175 при оценке в анализе *in vitro* или in vivo в сопоставимых условиях. Например, варианты с модификациями в триплексной петле варианта 175 гРНК демонстрируют высокое обогащение относительно каркаса 175, в особенности мутации на C15 или C17. Дополнительно изменения в любом члене прогнозируемой пары в стебле псевдоузла между G7 и A29 дают высокую степень обогащения относительно каркаса 175 при превращении A29 в C или T с образованием канонической пары Уотсона - Крика (G7:C29) и образованием вторым из них неустойчивой пары GU (G7:U29), и обе они, как можно ожидать, повысят стабильность спирали относительно пары G:A. Кроме того, вставка C в положении 54 в каркас гида 175 приводит к получению обогащенной модификации.

[0174] В некоторых примерах осуществления в описании предложены варианты гРНК, содержащие одну или более модификаций варианта каркаса 174 гРНК (SEQ ID NO: 2238), выбранных из группы, состоящей из модификаций в таблице 19, причем полученный вариант гРНК демонстрирует улучшенную функциональную характеристику по сравнению с исходным 174 при оценке *in vitro* или in vivo в сопоставимых условиях. В некоторых вариантах осуществления улучшенная функциональная характеристика представляет собой одно или более функциональных свойств, выбранных из группы, состоящей из повышенной активности редактирования, повышенной стабильности стебля псевдоузла, повышенной стабильности триплексной области, повышенной стабильности каркасного стебля, повышенной стабильности удлиненного стебля, сниженного уровня нецелевого свертывания и повышенной аффинности связывания с белком CRISPR класса 2, типа V. В вышеуказанных вариантах осуществления гРНК, содержащая одну или более модификаций варианта каркаса 174 гРНК, выбранных из группы, состоящей из модификаций в таблице 16 (с присоединенной целевой последовательностью и в комплексе с белком CRISPR класса 2, типа V), демонстрирует улучшенную оценку обогащения ($\log_2$) по меньшей мере примерно 2,0, по меньшей мере примерно 2,5, по меньшей мере примерно 3 или по меньшей мере примерно 3,5 по сравнению с оценкой каркаса гРНК с SEQ ID NO: 2238 в анализе *in vitro*.

[0175] В некоторых примерах осуществления в описании предложены варианты гРНК, содержащие одну или более модификаций варианта каркаса 175 гРНК (SEQ ID NO: 2239), выбранных из группы, состоящей из модификаций в таблице 20, причем полученный вариант гРНК демонстрирует улучшенную функциональную характеристику по сравнению с исходным 175 при оценке *in vitro* или in vivo в сопоставимых условиях. В некоторых вариантах осуществления улучшенная функциональная характеристика представляет собой одно или более функциональных свойств, выбранных из группы, состоящей из повышенной активности редактирования, повышенной стабильности стебля псевдоузла, повышенной

стабильности триплексной области, повышенной стабильности каркасного стебля, повышенной стабильности удлиненного стебля, сниженного уровня нецелевого свертывания и повышенной аффинности связывания с белком CRISPR класса 2, типа V. В вышеуказанных вариантах осуществления гРНК, содержащая одну или более модификаций варианта каркаса 175 гРНК, выбранных из группы, состоящей из модификаций в таблице 16 (с присоединенной целевой последовательностью и в комплексе с белком CRISPR класса 2, типа V), демонстрирует улучшенную оценку обогащения (log_2) по меньшей мере примерно 1,2, по меньшей мере примерно 1,5, по меньшей мере примерно 2,0, по меньшей мере примерно 2,5, по меньшей мере примерно 3 или по меньшей мере примерно 3,5 по сравнению с оценкой каркаса гРНК с SEQ ID NO: 2239 в анализе *in vitro*.

[0176] В конкретном варианте осуществления одна или более модификаций варианта каркаса 174 гРНК выбраны из группы, состоящей из положений нуклеотидов U11, U24, A29, U65, C66, C68, A69, U76, G77, A79 и A87. В конкретном варианте осуществления модификации варианта каркаса 174 гРНК представляют собой U11C, U24C, A29C, U65C, C66G, C68U, вставку ACGGA в положении 69, вставку UCCGU в положении 76, G77A, вставку GA в положении 79, A87G. В другом конкретном варианте осуществления модификации варианта каркаса 175 гРНК выбраны из группы, состоящей из положений нуклеотидов C9, U11, C17, U24, A29, G54, C65, A89 и A96. В конкретном варианте осуществления модификации варианта каркаса 174 гРНК представляют собой C9U, U11C, C17G, U24C, A29C, вставку G в положении 54, вставку C в положении 65, A89G и A96G.

[0177] В примерах осуществления вариант гРНК содержит одну или более модификаций относительно варианта каркаса 215 гРНК (SEQ ID NO:2275), причем полученный вариант гРНК демонстрирует улучшенную функциональную характеристику по сравнению с исходным 215 при оценке в анализе *in vitro* или in vivo в сопоставимых условиях.

[0178] В примерах осуществления вариант гРНК содержит одну или более модификаций относительно варианта каркаса 221 гРНК (SEQ ID NO: 2281), причем полученный вариант гРНК демонстрирует улучшенную функциональную характеристику по сравнению с исходным 221 при оценке *in vitro* или in vivo в сопоставимых условиях.

[0179] В примерах осуществления вариант гРНК содержит одну или более модификаций относительно варианта каркаса 225 гРНК (SEQ ID NO: 2285), причем полученный вариант гРНК демонстрирует улучшенную функциональную характеристику по сравнению с исходным 225 при оценке в анализе *in vitro* или in vivo в сопоставимых условиях.

[0180] В примерах осуществления вариант гРНК содержит одну или более модификаций относительно варианта каркаса 235 гРНК (SEQ ID NO: 2292), причем полученный вариант гРНК демонстрирует улучшенную функциональную характеристику по сравнению с исходным 225 при оценке *in vitro* или in vivo в сопоставимых условиях.

[0181] В примерах осуществления вариант гРНК содержит одну или более модификаций относительно варианта каркаса 251 гРНК (SEQ ID NO: 2308), причем

полученный вариант гРНК демонстрирует улучшенную функциональную характеристику по сравнению с исходным 251 при оценке в анализе *in vitro* или in vivo в сопоставимых условиях.

[0182] В вышеуказанных вариантах осуществления улучшенная функциональная характеристика включает в себя, без ограничений, одно или более из повышенной стабильности, повышенной транскрипции гРНК, повышенной стойкости к нуклеазной активности, повышенной скорости свертывания гРНК, сниженного образования побочных продуктов во время свертывания, повышенной продуктивности свертывания, повышенной аффинности связывания с белком CasX, повышенной аффинности связывания с целевой нуклеиновой кислотой при нахождении в комплексе с белком CasX, повышенного редактирования гена при нахождении в комплексе с белком CasX, повышенной специфичности редактирования при нахождении в комплексе с белком CasX, сниженного нецелевого редактирования при нахождении в комплексе с белком CasX и улучшенной способности к использованию большего спектра из одной или более PAM-последовательностей, включая ATC, CTC, GTC или TTC, при модификации целевой нуклеиновой кислоты при нахождении в комплексе с белком CasX. В некоторых случаях одна или более из улучшенных характеристик варианта гРНК имеет кратность повышения от по меньшей мере примерно 1,1 до примерно 100 000 раз относительно гРНК, из которой он получен. В других случаях одна или более улучшенных характеристик варианта гРНК имеет кратность улучшения по меньшей мере примерно 1,1, по меньшей мере примерно 10, по меньшей мере примерно 100, по меньшей мере примерно 1000, по меньшей мере примерно 10 000, по меньшей мере примерно 100 000 или более раз относительно гРНК, из которого он получен. В других случаях одна или более из улучшенных характеристик варианта гРНК имеет кратность улучшения примерно 1,1-100,00 раза, примерно 1,1-10,00 раза, примерно 1,1-1000 раз, примерно 1,1-500 раз, примерно 1,1-100 раз, примерно 1,1-50 раз, примерно 1,1-20 раз, примерно 10-100,00 раза, примерно 10-10,00 раза, примерно 10-1000 раз, примерно 10-500 раз, примерно 10-100 раз, примерно 10-50 раз, примерно 10-20 раз, примерно 2-70 раз, примерно 2-50 раз, примерно 2-30 раз, примерно 2-20 раз, примерно 2-10 раз, примерно 5-50 раз, примерно 5-30 раз, примерно 5-10 раз, примерно 100-100,00 раза, примерно 100-10,00 раза, примерно 100-1000 раз, примерно 100-500 раз, примерно 500-100,00 раза, примерно 500-10,00 раза, примерно 500-1000 раз, примерно 500-750 раз, примерно 1000-100,00 раза, примерно 10 000-100,00 раза, примерно 20-500 раз, примерно 20-250 раз, примерно 20-200 раз, примерно 20-100 раз, примерно 20-50 раз, примерно 50-10 000 раз, примерно 50-1000 раз, примерно 50-500 раз, примерно 50-200 раз или примерно 50-100 раз относительно гРНК, из которой он получен. В других случаях одна или более улучшенных характеристик варианта гРНК имеет кратность улучшения 1,1 раза, 1,2 раза, 1,3 раза, 1,4 раза, 1,5 раза, 1,6 раза, 1,7 раза, 1,8 раза, 1,9 раза, 2 раза, 3 раза, 4 раза, 5 раз, 6 раз, 7 раз, 8 раз, 9 раз, 10 раз, 11 раз, 12 раз, 13 раз, 14 раз, 15 раз, 16 раз, 17 раз, 18 раз, 19 раз, 20 раз, 25 раз, 30 раз, 40 раз, 45 раз, 50 раз, 55 раз, 60 раз, 70 раз, 80 раз, 90 раз, 100 раз, 110 раз, 120 раз, 130 раз, 140 раз, 150 раз, 160 раз, 170 раз, 180 раз, 190 раз, 200 раз, 210 раз,

220 раз, 230 раз, 240 раз, 250 раз, 260 раз, 270 раз, 280 раз, 290 раз, 300 раз, 310 раз, 320 раз, 330 раз, 340 раз, 350 раз, 360 раз, 370 раз, 380 раз, 390 раз, 400 раз, 425 раз, 450 раз, 475 раз или 500 раз относительно гРНК, из которой он получен.

[0183] В некоторых вариантах осуществления вариант гРНК содержит экзогенную удлиненную стеблевую петлю с такими отличиями от референсной гРНК, которые описаны ниже. В некоторых вариантах осуществления экзогенная удлиненная стеблевая петля имеет низкую идентичность или отсутствие идентичности с областями референсной стеблевой петли, описанными в настоящем документе (например, SEQ ID NO:15). В некоторых вариантах осуществления экзогенная стеблевая петля содержит по меньшей мере 10 п. о., по меньшей мере 20 п. о., по меньшей мере 30 п. о., по меньшей мере 40 п. о., по меньшей мере 50 п. о., по меньшей мере 60 п. о., по меньшей мере 70 п. о., по меньшей мере 80 п. о., по меньшей мере 90 п. о., по меньшей мере 100 п. о., по меньшей мере 200 п. о., по меньшей мере 300 п. о., по меньшей мере 400 п. о., по меньшей мере 500 п. о., по меньшей мере 600 п. о., по меньшей мере 700 п. о., по меньшей мере 800 п. о., по меньшей мере 900 п. о., по меньшей мере 1000 п. о., по меньшей мере 2000 п. о., по меньшей мере 3000 п. о., по меньшей мере 4000 п. о., по меньшей мере 5000 п. о., по меньшей мере 6000 п. о., по меньшей мере 7000 п. о., по меньшей мере 8000 п. о., по меньшей мере 9000 п. о., по меньшей мере 10 000 п. о., по меньшей мере 12 000 п. о., по меньшей мере 15 000 п. о. или по меньшей мере 20 000 п. о. В некоторых вариантах осуществления вариант гРНК содержит область удлиненной стеблевой петли, содержащую по меньшей мере 10, по меньшей мере 100, по меньшей мере 500, по меньшей мере 1000 или по меньшей мере 10 000 нуклеотидов. В некоторых вариантах осуществления гетерологичная стеблевая петля повышает стабильность гРНК. В некоторых вариантах осуществления гетерологичная стеблевая петля РНК способна связывать белок, структуру РНК, ДНК-последовательность или малую молекулу. В некоторых вариантах осуществления область экзогенной стеблевой петли, заменяющая стеблевую петлю, содержит стеблевую петлю РНК или шпильку, в которой полученная гРНК обладает повышенной стабильностью и, в зависимости от выбора петли, может взаимодействовать с определенными клеточными белками или РНК. Такие экзогенные удлиненные стеблевые петли могут содержать, например, термостабильную РНК, такую как шпилька MS2 (ACAUGAGGAUCACCCAUGU (SEQ ID NO: 1137)), шпилька Qβ (UGCAUGUCUAAGACAGCA (SEQ ID NO: 32)), шпилька II U1 (AAUCCAUUGCACUCCGGAUU (SEQ ID NO: 33)), Uvsx (CCUCUUCGGAGG (SEQ ID NO: 34)), шпилька PP7 (AGGAGUUUCUAUGGAAACCCU (SEQ ID NO: 35)), петля фаговой репликации (AGGUGGGACGACCUCUCGGUCGUCCUAUCU (SEQ ID NO: 36)), слипающаяся петля_a (UGCUCGCUCCGUUCGAGCA (SEQ ID NO: 37)), слипающаяся петля_b1 (UGCUCGACGCGUCCUCGAGCA (SEQ ID NO: 38)), слипающаяся петля_b2 (UGCUCGUUUGCGGCUACGAGCA (SEQ ID NO: 39)), G-квадруплекс M3q (AGGGAGGGAGGGAGAGG (SEQ ID NO: 40)), G-квадруплексная теломерная корзина (GGUUAGGGUUAGGGUUAGG (SEQ ID NO: 41)), сарцин-рициновая петля (CUGCUCAGUACGAGAGGAACCGCAG (SEQ ID NO: 42)) или псевдоузлы

(UACACUGGGAUCGCUGAAUUAGAGAUCGGCGUCCUUUCAUUCUAUAUACUUUGG AGUUUUAAAAUGUCUCUAAGUACA (SEQ ID NO: 43)). В некоторых вариантах осуществления одна из вышеуказанных шпилечных последовательностей встроена в стеблевую петлю для содействия встраиванию гРНК (и связанного CasX в комплексе RNP) в отпочковывающуюся XDP (более подробно описан ниже), когда соответствующий лиганд встроен в Gag-полипротеин XDP.

[0184] В некоторых вариантах осуществления вариант гРНК содержит партнера по концевому слиянию. Термин «вариант гРНК» включает в себя варианты, которые включают в себя экзогенные последовательности, такие как концевые слияния или внутренние вставки. Примеры концевых слияний могут включать в себя слияние гРНК с саморасщепляющимся рибозимом или белок-связывающим мотивом. В настоящем документе термин «рибозим» относится к РНК или ее сегменту, обладающим одной или более каталитическими активностями, аналогичными белковому ферменту. Примеры каталитических активностей рибозима могут включать в себя, например, расщепление и/или лигирование РНК, расщепление и/или лигирование ДНК или образование пептидной связи. В некоторых вариантах осуществления такие слияния могут либо улучшать свертывание каркаса, либо рекрутировать механизм репарации ДНК. Например, гРНК в некоторых вариантах осуществления может быть слита с антигеномным рибозимом вируса гепатита дельта (HDV), геномным рибозимом HDV, рибозимом Hatchet (из метагеномных данных), pistol-рибозимом env25 (характерный представитель Aliistipes putredinis), минимальным рибозимом Hammerhead HH15, рибозимом вируса кольцевой пятнистости табака (TRSV), вирусным рибозимом Hammerhead дикого типа (и целесообразными вариантами) или мотивом для рекрутирования Twisted Sister 1 или RBMX. Рибозимы Hammerhead представляют собой мотивы РНК, которые катализируют обратимые реакции расщепления и лигирования в конкретном сайте в молекуле РНК. Рибозимы Hammerhead включают в себя рибозимы hammerhead типа I, типа II и типа III. Рибозимы HDV, pistol и hatchet обладают саморасщепляющейся активностью. Варианты гРНК, содержащие один или более рибозимов, могут обеспечивать расширенную функцию гРНК по сравнению с референсной гРНК. Например, гРНК, содержащие саморасщепляющиеся рибозимы, в некоторых вариантах осуществления могут быть транскрибированы и процессированы в зрелые гРНК в составе полицистронных транскриптов. Такие слияния могут происходить либо на 5’-, либо на 3’-конце гРНК. В некоторых вариантах осуществления вариант гРНК содержит слияние как на 5’-, так и на 3’-конце, причем каждое слияние независимо представляет собой слияние, описанное в настоящем документе.

[0185] В вариантах осуществления вариантов гРНК вариант гРНК дополнительно содержит спейсерную область (или нацеливающую последовательность), расположенную на 3’-конце гРНК, способную гибридизоваться с целевой нуклеиновой кислотой, которая содержит от по меньшей мере 14 до примерно 35 нуклеотидов, причем спейсер выполнен с последовательностью, которая комплементарна целевой нуклеиновой кислоте. В некоторых вариантах осуществления кодируемый вариант гРНК содержит нацеливающую

последовательность из по меньшей мере 10-20 нуклеотидов, комплементарную целевой нуклеиновой кислоте. В некоторых вариантах осуществления нацеливающая последовательность имеет 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 или 30 нуклеотидов. В некоторых вариантах осуществления кодируемый вариант гРНК содержит нацеливающую последовательность, имеющую 20 нуклеотидов. В некоторых вариантах осуществления нацеливающая последовательность имеет 25 нуклеотидов. В некоторых вариантах осуществления нацеливающая последовательность имеет 24 нуклеотида. В некоторых вариантах осуществления нацеливающая последовательность имеет 23 нуклеотида. В некоторых вариантах осуществления нацеливающая последовательность имеет 22 нуклеотида. В некоторых вариантах осуществления нацеливающая последовательность имеет 21 нуклеотид. В некоторых вариантах осуществления нацеливающая последовательность имеет 20 нуклеотидов. В некоторых вариантах осуществления нацеливающая последовательность имеет 19 нуклеотидов. В некоторых вариантах осуществления нацеливающая последовательность имеет 18 нуклеотидов. В некоторых вариантах осуществления нацеливающая последовательность имеет 17 нуклеотидов. В некоторых вариантах осуществления нацеливающая последовательность имеет 16 нуклеотидов. В некоторых вариантах осуществления нацеливающая последовательность имеет 15 нуклеотидов. В некоторых вариантах осуществления нацеливающая последовательность имеет 14 нуклеотидов.

h. Образование комплекса с белком класса 2, типа V

[0186] В некоторых вариантах осуществления при экспрессии вариант гРНК образует комплекс RNP с белком класса 2, типа V, включая варианты белка CasX, содержащие любую из последовательностей SEQ ID NO: 247-592 или 1147-1231 из таблицы 3 или последовательность, которая имеет идентичность с ней по меньшей мере примерно 50%, по меньшей мере примерно 60%, по меньшей мере примерно 70%, по меньшей мере примерно 80%, по меньшей мере примерно 85%, по меньшей мере примерно 90%, по меньшей мере примерно 91%, по меньшей мере примерно 92%, по меньшей мере примерно 93%, по меньшей мере примерно 94%, по меньшей мере примерно 95%, по меньшей мере примерно 96%, по меньшей мере примерно 97%, по меньшей мере примерно 98% или по меньшей мере примерно 99%. В некоторых вариантах осуществления при экспрессии вариант гРНК образует комплекс RNP с вариантом белка CasX, содержащим любую из последовательностей SEQ ID NO: 270-592 или 1147-1231 или последовательность, которая имеет идентичность с ней по меньшей мере примерно 50%, по меньшей мере примерно 60%, по меньшей мере примерно 70%, по меньшей мере примерно 80%, по меньшей мере примерно 85%, по меньшей мере примерно 90%, по меньшей мере примерно 91%, по меньшей мере примерно 92%, по меньшей мере примерно 93%, по меньшей мере примерно 94%, по меньшей мере примерно 95%, по меньшей мере примерно 96%, по меньшей мере примерно 97%, по меньшей мере примерно 98% или по меньшей мере примерно 99%. В некоторых вариантах осуществления при экспрессии вариант гРНК образует комплекс RNP с вариантом белка CasX, содержащим любую из последовательностей SEQ ID NO: 415-592

или 1147-1231 или последовательность, которая имеет идентичность с ней по меньшей мере примерно 50%, по меньшей мере примерно 60%, по меньшей мере примерно 70%, по меньшей мере примерно 80%, по меньшей мере примерно 85%, по меньшей мере примерно 90%, по меньшей мере примерно 91%, по меньшей мере примерно 92%, по меньшей мере примерно 93%, по меньшей мере примерно 94%, по меньшей мере примерно 95%, по меньшей мере примерно 96%, по меньшей мере примерно 97%, по меньшей мере примерно 98% или по меньшей мере примерно 99%.

[0187] В некоторых вариантах осуществления вариант гРНК имеет улучшенную способность образовывать комплекс с вариантом белка CasX по сравнению с референсной гРНК, улучшая таким образом ее способность образовывать компетентный в отношении расщепления рибонуклеопротеиновый (RNP) комплекс с белком CasX, как описано в разделе «Примеры». Улучшение образования рибонуклеопротеинового комплекса может в некоторых вариантах осуществления повышать эффективность, с которой собираются функциональные RNP. В некоторых вариантах осуществления более 90%, более 93%, более 95%, более 96%, более 97%, более 98% или более 99% RNP, содержащих вариант гРНК и ее нацеливающую последовательность, обладают компетентностью в отношении генетического редактирования целевой нуклеиновой кислоты.

[0188] Примеры нуклеотидных изменений, которые могут улучшать способность вариантов гРНК образовывать комплекс с белком CasX, могут в некоторых вариантах осуществления включать замену каркасного стебля на термостабильную стеблевую петлю. Без ограничений, накладываемых какой-либо теорией, замена каркасного стебля на термостабильную стеблевую петлю может повышать общую стабильность связывания варианта гРНК с белком CasX. Альтернативно или в дополнение удаление большого участка стеблевой петли может изменять кинетику сворачивания варианта гРНК и облегчать и ускорять структурную сборку функционально свернутой гРНК, например, посредством уменьшения степени «запутывания» вариант гРНК на самой себе. В некоторых вариантах осуществления выбор последовательности каркасной стеблевой петли может различаться для разных нацеливающих последовательностей, используемых для гРНК. В некоторых вариантах осуществления последовательность каркаса может быть адаптирована к нацеливающей последовательности и, следовательно, к целевой последовательности. Можно применять биохимические анализы для оценки аффинности связывания белка CasX с вариантом гРНК с образованием RNP, в том числе анализы из раздела «Примеры». Например, специалист в данной области может измерять изменения количества флуоресцентно-меченой гРНК, которая связана с иммобилизованным белком CasX, в ответ на увеличение концентраций дополнительной немеченой гРНК-«холодного конкурента». Альтернативно или в дополнение флуоресцентный сигнал можно отслеживать или можно наблюдать, как он изменяется, при протекании разных количеств флуоресцентно-меченой гРНК над иммобилизованным белком CasX. Альтернативно возможность образования RNP можно оценивать с применением анализов расщепления *in vitro* относительно определенной последовательности целевой нуклеиновой кислоты, как

описано в разделе «Примеры».

Химически модифицированные гРНК

[0189] В некоторых вариантах осуществления в описании предложены химически модифицированные гРНК. В некоторых вариантах осуществления в настоящем описании предложена химически модифицированная гРНК, которая имеет функциональность гидовой НК и имеет сниженную восприимчивость к расщеплению нуклеазой. гРНК, которая содержит любой нуклеотид, отличный от четырех канонических рибонуклеотидов A, C, G и U, или дезоксинуклеотид, представляет собой химически модифицированную гРНК. В некоторых случаях химически модифицированная гРНК содержит любую основную цепь или межнуклеотидную связь, отличную от естественной фосфодиэфирной связи. В определенных вариантах осуществления сохраненная функциональность включает в себя способность модифицированной гРНК связываться с CasX по любому из вариантов осуществления, описанных в настоящем документе. В определенных вариантах осуществления сохраненная функциональность включает в себя способность модифицированной гРНК связываться с последовательностью целевой нуклеиновой кислоты. В определенных вариантах осуществления сохраненная функциональность включает в себя нацеливание белка CasX или способность предварительно образованного RNP-комплекса связываться с последовательностью целевой нуклеиновой кислоты. В определенных вариантах осуществления сохраненная функциональность включает в себя способность к никированию целевого полинуклеотида посредством CasX-гРНК. В определенных вариантах осуществления сохраненная функциональность включает в себя способность расщеплять последовательность целевой нуклеиновой кислоты посредством CasX-гРНК. В определенных вариантах осуществления сохраненная функциональность представляет собой любую другую известную функцию гРНК в рекомбинантной системе с химерным белком CasX по вариантам осуществления описания.

[0190] В некоторых вариантах осуществления в описании предложена химически модифицированная гРНК, в которой в гРНК встроена модификация нуклеотидного сахара, выбранная из группы, состоящей из 2′-O-C_{1-4} алкила, такого как 2′-O-метил (2′-OMe), 2′-дезокси (2′-H), 2′-O-C_{1-3} алкил-O-C_{1-3} алкила, такого как 2′-метоксиэтил («2′-MOE»), 2′-фтора («2′-F»), 2′-амино («2′-NH_2»), 2′-арабинозил («2′-арабино») нуклеотида, 2′-F-арабинозил («2′-F-арабино») нуклеотида, нуклеотида 2′-запертой нуклеиновой кислоты («LNA»), нуклеотида 2′-незапертой нуклеиновой кислоты («ULNA»), сахара в L-форме («L-сахар») и 4′-тиорибозилнуклеотида. В других вариантах осуществления модификация межнуклеотидной связи, которая встроена в гидовую РНК, выбрана из группы, состоящей из: фосфоротиоата («P(S)» (P(S)), фосфонокарбоксилата ($P(CH_2)_nCOOR$), например, фосфоноацетата «PACE» ($P(CH_2COO^-)$), тиофосфонокарбоксилата ($(S)P(CH_2)_nCOOR$), например, тиофосфоноацетата «тиоPACE» ($(S)P(CH_2)_nCOO^-$)), алкилфосфоната ($P(C_{1-3}$алкил), например метилфосфоната -$P(CH_3)$), боранофосфоната ($P(BH_3)$) и фосфородитиоата ($P(S)_2$).

[0191] В определенных вариантах осуществления в описании предложена

химически модифицированная гРНК, в которой в гРНК встроена модификация нуклеотидного основания («основания»), выбранная из группы, состоящей из: 2-тиоурацила («2-тиоU»), 2-тиоцитозина («2-тиоC»), 4-тиоурацила («4-тиоU»), 6-тиогуанина («6-тиоG»), 2-аминоаденина («2-аминоA»), 2-аминопурина, псевдоурацила, гипоксантина, 7-деазагуанина, 7-деаза-8-азагуанина, 7-деазааденина, 7-деаза-8-азааденина, 5-метилцитозина («5-метилC»), 5-метилурацила («5-метилU»), 5-гидроксиметилцитозина, 5-гидроксиметилурацила, 5,6-дегидроурацила, 5-пропинилцитозина, 5-пропинилурацила, 5-этинилцитозина, 5-этинилурацила, 5-аллилурацила («5-аллилU»), 5-аллилцитозина («5-аллилC»), 5-аминоаллилурацила («5-аминоаллилU»), 5-аминоаллил-цитозина («5-аминоаллилC»), абазического нуклеотида, Z-основания, P-основания, неструктурированной нуклеиновой кислоты («UNA»), изогуанина («isoG»), изоцитозина («isoC»), 5-метил-2-пиримидина, x(A, G, C, T) и y(A, G, C, T).

[0192] В других вариантах осуществления в описании предложена химически модифицированная гРНК, в которую введены одна или более изотопных модификаций в нуклеотидный сахар, нуклеотидное основание, фосфодиэфирную связь и/или нуклеотидфосфаты, включая нуклеотиды, содержащие один или более атомов ^{15}N, ^{13}C, ^{14}C, дейтерия, ^{3}H, ^{32}P, ^{125}I, ^{131}I или других атомов или элементов, используемых в качестве маркерных веществ.

[0193] В некоторых вариантах осуществления «концевая» модификация, встроенная в гРНК, выбрана из группы, состоящей из: PEG (полиэтиленгликоля), углеводородных линкеров (включая: замещенные гетероатомом (O, S, N) углеводородные спейсеры; галогензамещенные углеводородные спейсеры; кето-, карбоксил-, амидо-, тионил-, карбамоил-, тионокарбамоил-содержащие углеводородные спейсеры), сперминовых линкеров, красителей, включая флуоресцентные красители (например, флуоресцеины, родамины, цианины), присоединенные к линкерам, таким как, например, 6-флуоресцеингексил, гасителей (например, дабцила, BHQ) и других меток (например, биотина, дигоксигенина, акридина, стрептавидина, авидина, пептидов и/или белков). В некоторых вариантах осуществления «концевая» модификация включает в себя конъюгирование (или лигирование) гРНК с другой молекулой, содержащей олигонуклеотид из дезоксинуклеотидов и/или рибонуклеотидов, пептид, белок, сахар, олигосахарид, стероид, липид, фолиевую кислоту, витамин и/или другую молекулу. В некоторых вариантах осуществления в описании предложена химически модифицированная гРНК, в которой «концевая» модификация (описанная выше) расположена внутри последовательности гРНК с помощью линкера, такого как, например, 2-(4-бутиламидофлуоресцеин)пропан-1,3-диол бис(фосфодиэфирный) линкер, который встроен в качестве фосфодиэфирной связи, и он может быть встроен в любое место между двумя нуклеотидами в гРНК.

[0194] В некоторых вариантах осуществления в описании предложена химически модифицированная гРНК, имеющая концевую модификацию, содержащую концевую функциональную группу, такую как амин, тиол (или сульфгидрил), гидроксил, карбоксил,

карбонил, тионил, тиокарбонил, карбамоил, тиокарбамоил, фосфорил, алкен, алкин, галоген или терминированный функциональной группой линкер, который впоследствии может быть конъюгирован с требуемым фрагментом, выбранным из группы, состоящей из флуоресцентного красителя, нефлуоресцентной метки, маркера (например, ^{14}C, биотина, авидина, стрептавидина или фрагмента, содержащего изотопную метку, такую как ^{15}N, ^{13}C, дейтерий, ^{3}H, ^{32}P, ^{125}I и т. п.), олигонуклеотида (содержащего дезоксинуклеотиды и/или рибонуклеотиды, включая аптамер), аминокислоты, пептида, белка, сахара, олигосахарида, стероида, липида, фолиевой кислоты и витамина. При конъюгации используют стандартные химические реакции, известные в данной области, включая, без ограничений, присоединение посредством N-гидроксисукцинимида, изотиоцианата, DCC (или DCI) и/или любого другого стандартного способа, как описано в «Bioconjugate Techniques», Greg T. Hermanson, Publisher Eslsevier Science, 3rd ed. (2013), содержание которой полностью включено в настоящий документ путем ссылки.

IV. Белки CRISPR класса 2, типа V для модификации целевой нуклеиновой кислоты

[0195] В настоящем описании предложены системы, содержащие нуклеазу CRISPR, которая может применяться при редактировании генома эукариотических клеток. В некоторых вариантах осуществления нуклеаза CRISPR, применяемая в системах редактирования генома, представляет собой нуклеазу класса 2, типа V. Хотя члены систем CRISPR-Cas класса 2, типа V имеют различия, они обладают некоторыми общими характеристиками, которые отличают их от систем Cas9. Во-первых, нуклеазы класса 2, типа V содержат единственный РНК-направляемый содержащий домен RuvC эффектор, не содержат домена HNH и распознают ТС-мотив PAM на 5'-стороне от целевой области на нецелевой цепи, и этим отличаются от систем Cas9, которые используют богатый G-нуклеотидами PAM-мотив на 3'-стороне от целевых последовательностей. Нуклеазы типа V создают ступенчатые двухцепочечные разрывы дистально от PAM-последовательности, в отличие от Cas9, которые создают тупой конец в проксимальном сайте вблизи PAM. Кроме того, нуклеазы типа V разрушают транс-активные оцДНК при активации целевой дцДНК или связывании с цис-активной оцДНК. В некоторых вариантах осуществления нуклеазы типа V вариантов осуществления распознают PAM-мотив 5'-TC и образуют ступенчатые концы, расщепленные исключительно доменом RuvC. В некоторых вариантах осуществления нуклеаза типа V выбрана из группы, состоящей из Cas12a, Cas12b, Cas12c, Cas12d (CasY), Cas12j, Cas12k, C2c4, C2c8, C2c5, C2c10, C2c9, CasZ и CasX. В некоторых вариантах осуществления в настоящем описании предложены системы, содержащие вариант белка CasX и один или более вариантов гРНК (система CasX : гРНК), которые специально разработаны для модификации последовательности целевой нуклеиновой кислоты в эукариотических клетках.

[0196] Термин «белок CasX», используемый в настоящем документе, означает семейство белков и охватывает все природные белки CasX, белки, которые на по меньшей мере 50% идентичны природным белкам CasX, а также варианты CasX, обладающие одной

или более улучшенными характеристиками относительно природного референсного белка CasX или другого варианта CasX, из которого он получен.

[0197] Белки CasX описания содержат по меньшей мере один из следующих доменов: домен, связывающийся с нецелевой цепью (NTSB), домен загрузки целевой цепи (TSL), домен спиральный I (который дополнительно делится на поддомены спиральные I-I и I-II), домен спиральный II, олигонуклеотид-связывающий домен (OBD, который дополнительно делится на поддомены OBD-I и OBD-II) и ДНК-расщепляющий домен RuvC (который дополнительно делится на поддомены RuvC-I и II). Домен RuvC может быть модифицирован или удален в каталитически неактивном варианте CasX, как более подробно описано ниже.

[0198] В некоторых вариантах осуществления белок CasX может связывать и/или модифицировать (например, никировать, катализировать двухцепочечный разрыв, метилировать, деметилировать и т. д.) целевую нуклеиновую кислоту в конкретной последовательности, на которую нацелена ассоциированная гРНК, гибридизирующаяся с последовательностью в пределах последовательности целевой нуклеиновой кислоты.

a. Референсные белки CasX

[0199] В описании предложены природные белки CasX (называемые в настоящем документе «референсным белком CasX»), которые подвергались последующей модификации для создания вариантов CasX описания. Например, референсные белки CasX могут быть выделены из природных прокариотов, таких как виды *Deltaproteobacteria*, *Planctomycetes* или *Candidatus Sungbacteria*. Референсный белок CasX представляет собой эндонуклеазу CRISPR/Cas типа II, относящуюся к семейству белков CasX (взаимозаменяемо называемому Cas12e), которые взаимодействуют с гидовой РНК с образованием рибонуклеопротеинового комплекса (RNP).

[0200] В некоторых случаях референсный белок CasX выделяют или получают из *Deltaproteobacter,* и он имеет последовательность:

1 MEKRINKIRK KLSADNATKP VSRSGPMKTL LVRVMTDDLK KRLEKRRKKP EVMPQVISNN

61 AANNLRMLLD DYTKMKEAIL QVYWQEFKDD HVGLMCKFAQ PASKKIDQNK LKPEMDEKGN

121 LTTAGFACSQ CGQPLFVYKL EQVSEKGKAY TNYFGRCNVA EHEKLILLAQ LKPEKDSDEA

181 VTYSLGKFGQ RALDFYSIHV TKESTHPVKP LAQIAGNRYA SGPVGKALSD ACMGTIASFL

241 SKYQDIIIEH QKVVKGNQKR LESLRELAGK ENLEYPSVTL PPQPHTKEGV DAYNEVIARV

301 RMWVNLNLWQ KLKLSRDDAK PLLRLKGFPS FPVVERRENE VDWWNTINEV KKLIDAKRDM

361 GRVFWSGVTA EKRNTILEGY NYLPNENDHK KREGSLENPK KPAKRQFGDL LLYLEKKYAG

421 DWGKVFDEAW ERIDKKIAGL TSHIEREEAR NAEDAQSKAV LTDWLRAKAS FVLERLKEMD

481 EKEFYACEIQ LQKWYGDLRG NPFAVEAENR VVDISGFSIG SDGHSIQYRN LLAWKYLENG

541 KREFYLLMNY GKKGRIRFTD GTDIKKSGKW QGLLYGGGKA KVIDLTFDPD DEQLIILPLA

601 FGTRQGREFI WNDLLSLETG LIKLANGRVI EKTIYNKKIG RDEPALFVAL TFERREVVDP

661 SNIKPVNLIG VDRGENIPAV IALTDPEGCP LPEFKDSSGG PTDILRIGEG YKEKQRAIQA

721 AKEVEQRRAG GYSRKFASKS RNLADDMVRN SARDLFYHAV THDAVLVFEN LSRGFGRQGK

781 RTFMTERQYT KMEDWLTAKL AYEGLTSKTY LSKTLAQYTS KTCSNCGFTI TTADYDGMLV

841 RLKKTSDGWA TTLNNKELKA EGQITYYNRY KRQTVEKELS AELDRLSEES GNNDISKWTK

901 GRRDEALFLL KKRFSHRPVQ EQFVCLDCGH EVHADEQAAL NIARSWLFLN SNSTEFKSYK

961 SGKQPFVGAW QAFYKRRLKE VWKPNA (SEQ ID NO: 1).

[0201] В некоторых случаях референсный белок CasX выделяют или получают из *Planctomycetes*, и он имеет последовательность:

1 MQEIKRINKI RRRLVKDSNT KKAGKTGPMK TLLVRVMTPD LRERLENLRK KPENIPQPIS

61 NTSRANLNKL LTDYTEMKKA ILHVYWEEFQ KDPVGLMSRV AQPAPKNIDQ RKLIPVKDGN

121 ERLTSSGFAC SQCCQPLYVY KLEQVNDKGK PHTNYFGRCN VSEHERLILL SPHKPEANDE

181 LVTYSLGKFG QRALDFYSIH VTRESNHPVK PLEQIGGNSC ASGPVGKALS DACMGAVASF

241 LTKYQDIILE HQKVIKKNEK RLANLKDIAS ANGLAFPKIT LPPQPHTKEG IEAYNNVVAQ

301 IVIWVNLNLW QKLKIGRDEA KPLQRLKGFP SFPLVERQAN EVDWWDMVCN VKKLINEKKE

361 DGKVFWQNLA GYKRQEALLP YLSSEEDRKK GKKFARYQFG DLLLHLEKKH GEDWGKVYDE

421 AWERIDKKVE GLSKHIKLEE ERRSEDAQSK AALTDWLRAK ASFVIEGLKE ADKDEFCRCE

481 LKLQKWYGDL RGKPFAIEAE NSILDISGFS KQYNCAFIWQ KDGVKKLNLY LIINYFKGGK

541 LRFKKIKPEA FEANRFYTVI NKKSGEIVPM EVNFNFDDPN LIILPLAFGK

RQGREFIWND
601 LLSLETGSLK LANGRVIEKT LYNRRTRQDE PALFVALTFE RREVLDSSNI KPMNLIGIDR
661 GENIPAVIAL TDPEGCPLSR FKDSLGNPTH ILRIGESYKE KQRTIQAAKE VEQRRAGGYS
721 RKYASKAKNL ADDMVRNTAR DLLYYAVTQD AMLIFENLSR GFGRQGKRTF MAERQYTRME
781 DWLTAKLAYE GLPSKTYLSK TLAQYTSKTC SNCGFTITSA DYDRVLEKLK KTATGWMTTI
841 NGKELKVEGQ ITYYNRYKRQ NVVKDLSVEL DRLSEESVNN DISSWTKGRS GEALSLLKKR
901 FSHRPVQEKF VCLNCGFETH ADEQAALNIA RSWLFLRSQE YKKYQTNKTT GNTDKRAFVE
961 TWQSFYRKKL KEVWKPAV (SEQ ID NO: 2).

[0202] В некоторых случаях референсный белок CasX выделяют или получают из *Candidatus Sungbacteria,* и он имеет последовательность

1 MDNANKPSTK SLVNTTRISD HFGVTPGQVT RVFSFGIIPT KRQYAIIERW FAAVEAARER
61 LYGMLYAHFQ ENPPAYLKEK FSYETFFKGR PVLNGLRDID PTIMTSAVFT ALRHKAEGAM
121 AAFHTNHRRL FEEARKKMRE YAECLKANEA LLRGAADIDW DKIVNALRTR LNTCLAPEYD
181 AVIADFGALC AFRALIAETN ALKGAYNHAL NQMLPALVKV DEPEEAEESP RLRFFNGRIN
241 DLPKFPVAER ETPPDTETII RQLEDMARVI PDTAEILGYI HRIRHKAARR KPGSAVPLPQ
301 RVALYCAIRM ERNPEEDPST VAGHFLGEID RVCEKRRQGL VRTPFDSQIR ARYMDIISFR
361 ATLAHPDRWT EIQFLRSNAA SRRVRAETIS APFEGFSWTS NRTNPAPQYG MALAKDANAP
421 ADAPELCICL SPSSAAFSVR EKGGDLIYMR PTGGRRGKDN PGKEITWVPG SFDEYPASGV
481 ALKLRLYFGR SQARRMLTNK TWGLLSDNPR VFAANAELVG KKRNPQDRWK LFFHMVISGP
541 PPVEYLDFSS DVRSRARTVI GINRGEVNPL AYAVVSVEDG QVLEEGLLGK KEYIDQLIET
601 RRRISEYQSR EQTPPRDLRQ RVRHLQDTVL GSARAKIHSL IAFWKGILAI ERLDDQFHGR
661 EQKIIPKKTY LANKTGFMNA LSFSGAVRVD KKGNPWGGMI EIYPGGISRT CTQCGTVWLA

721 RRPKNPGHRD AMVVIPDIVD DAAATGFDNV DCDAGTVDYG ELFTLSREWV RLTPRYSRVM

781 RGTLGDLERA IRQGDDRKSR QMLELALEPQ PQWGQFFCHR CGFNGQSDVL AATNLARRAI

841 SLIRRLPDTD TPPTP (SEQ ID NO: 3).

b. Варианты белков CasX класса 2, типа V

[0203] В настоящем описании предложены варианты белков CasX класса 2, типа V от референсного белка CasX или варианты, полученные из других вариантов CasX (см., например, ФИГ. 44) (взаимозаменяемо называемые в настоящем документе «вариант CasX класса 2, типа V», «вариант CasX» или «вариант белка CasX»), причем варианты CasX класса 2, типа V содержат по меньшей мере одну модификацию в по меньшей мере одном домене относительно референсного белка CasX, включая, без ограничений, последовательности SEQ ID NO:1-3 или по меньшей мере одну модификацию относительно другого варианта CasX. Любое изменение в аминокислотной последовательности референсного белка CasX или другого варианта белка CasX, которое приводит к получению улучшенной характеристики белка CasX, считается вариантом белка CasX описания. Например, варианты CasX могут содержать одну или более аминокислотных замен, вставок, делеций или переставленных доменов или любые их комбинации относительно референсной последовательности белка CasX.

[0204] Варианты CasX описания имеют одну или более улучшенных характеристик по сравнению с референсным белком CasX SEQ ID NO:1, SEQ ID NO:2 или SEQ ID NO:3 или вариантом, из которого они получены; например, CasX 491 (SEQ ID NO: 336) или CasX 515 (SEQ ID NO: 416). Примеры улучшенных характеристик вариантов осуществления варианта CasX включают в себя, без ограничений, улучшенное свертывание варианта, повышенную аффинность связывания с гРНК, повышенную аффинность связывания с целевой нуклеиновой кислотой, улучшенную способность использовать более широкий спектр PAM-последовательностей при редактировании и/или связывании целевой нуклеиновой кислоты, улучшенное раскручивание целевой ДНК, повышенную активность редактирования, улучшенную эффективность редактирования, улучшенную специфичность редактирования целевой нуклеиновой кислоты, сниженное нецелевое редактирование или расщепление, повышенную процентную долю эукариотического генома, которую можно эффективно редактировать, повышенную активность нуклеазы, повышенную загрузку целевой цепи для двухцепочечного расщепления, сниженную загрузку целевой цепи для одноцепочечного никирования, повышенное связывание с нецелевой цепью ДНК, улучшенную стабильность белка, улучшенную стабильность комплекса белок : гРНК (RNP) и улучшенные характеристики слияния. В вышеуказанных вариантах осуществления одна или более улучшенных характеристик варианта CasX имеет кратность улучшения от по меньшей мере примерно 1,1 до примерно 100 000 раз относительно референсного белка CasX с SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3, или CasX 491 (SEQ ID NO: 336) или CasX 515 (SEQ ID NO: 416), при анализе сопоставимым

образом. В других вариантах осуществления улучшение составляет по меньшей мере примерно 1,1 раза, по меньшей мере примерно 2 раза, по меньшей мере примерно 5 раз, по меньшей мере примерно 10 раз, по меньшей мере примерно 50 раз, по меньшей мере примерно 100 раз, по меньшей мере примерно 500 раз, по меньшей мере примерно 1000 раз, по меньшей мере примерно 5000 раз, по меньшей мере примерно 10 000 раз или по меньшей мере примерно 100 000 раз по сравнению с референсным белком CasX с SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3, или CasX 491, или CasX 515, при анализе сопоставимым образом. В других случаях одна или более улучшенных характеристик RNP из варианта CasX и варианта гРНК имеет кратность улучшения по меньшей мере примерно 1,1, по меньшей мере примерно 10, по меньшей мере примерно 100, по меньшей мере примерно 1000, по меньшей мере примерно 10 000, по меньшей мере примерно 100 000 или более раз относительно RNP с референсным белком CasX с SEQ ID NO:1, SEQ ID NO:2 или SEQ ID NO:3 и гРНК из таблицы 1, или CasX 491, или CasX 515 с гРНК 174. В других случаях одна или более улучшенных характеристик RNP из варианта CasX и варианта гРНК имеет кратность улучшения от примерно 1,1 до 100,00 раза, от примерно 1,1 до 10,00 раза, от примерно 1,1 до 1000 раз, от примерно 1,1 до 500 раз, от примерно 1,1 до 100 раз, от примерно 1,1 до 50 раз, от примерно 1,1 до 20 раз, от примерно 10 до 100,00 раза, от примерно 10 до 10,00 раза, от примерно 10 до 1000 раз, от примерно 10 до 500 раз, от примерно 10 до 100 раз, от примерно 10 до 50 раз, от примерно 10 до 20 раз, от примерно 2 до 70 раз, от примерно 2 до 50 раз, от примерно 2 до 30 раз, от примерно 2 до 20 раз, от примерно 2 до 10 раз, от примерно 5 до 50 раз, от примерно 5 до 30 раз, от примерно 5 до 10 раз, от примерно 100 до 100,00 раза, от примерно 100 до 10,00 раза, от примерно 100 до 1000 раз, от примерно 100 до 500 раз, от примерно 500 до 100,00 раза, от примерно 500 до 10,00 раза, от примерно 500 до 1000 раз, от примерно 500 до 750 раз, от примерно 10 000 до 100,00 раза, от примерно 1000 до 100,00 раза, от примерно 20 до 500 раз, от примерно 20 до 250 раз, от примерно 20 до 200 раз, от примерно 20 до 100 раз, от примерно 20 до 50 раз, от примерно 50 до 10 000 раз, от примерно 50 до 1000 раз, от примерно 50 до 500 раз, от примерно 50 до 200 раз или от примерно 50 до 100 раз относительно RNP с референсным белком CasX с SEQ ID NO:1, SEQ ID NO:2 или SEQ ID NO:3 и гРНК из таблицы 1, или CasX 491, или CasX 515 с гРНК 174, при анализе сопоставимым образом. В других случаях одна или более улучшенных характеристик RNP с вариантом CasX и вариантом гРНК имеет кратность улучшения примерно 1,1 раза, 1,2 раза, 1,3 раза, 1,4 раза, 1,5 раза, 1,6 раза, 1,7 раза, 1,8 раза, 1,9 раза, 2 раза, 3 раза, 4 раза, 5 раз, 6 раз, 7 раз, 8 раз, 9 раз, 10 раз, 11 раз, 12 раз, 13 раз, 14 раз, 15 раз, 16 раз, 17 раз, 18 раз, 19 раз, 20 раз, 25 раз, 30 раз, 40 раз, 45 раз, 50 раз, 55 раз, 60 раз, 70 раз, 80 раз, 90 раз, 100 раз, 110 раз, 120 раз, 130 раз, 140 раз, 150 раз, 160 раз, 170 раз, 180 раз, 190 раз, 200 раз, 210 раз, 220 раз, 230 раз, 240 раз, 250 раз, 260 раз, 270 раз, 280 раз, 290 раз, 300 раз, 310 раз, 320 раз, 330 раз, 340 раз, 350 раз, 360 раз, 370 раз, 380 раз, 390 раз, 400 раз, 425 раз, 450 раз, 475 раз или 500 раз относительно RNP с референсным белком CasX с SEQ ID NO:1, SEQ ID NO:2 или SEQ ID NO:3 и гРНК из таблицы 1, или CasX 491, или CasX 515 с гРНК 174, при анализе сопоставимым образом.

[0205] В некоторых вариантах осуществления модификация варианта CasX представляет собой мутацию одной или более аминокислот референсного CasX. В других вариантах осуществления модификация включает в себя вставку или замену части или всего домена из другого белка CasX. В конкретном варианте осуществления варианты CasX с SEQ ID NO: 415-592 и 1147-1231 имеют домен NTSB и спиральный 1B с SEQ ID NO: 1, а другие домены получены из SEQ ID NO: 2, в дополнение к индивидуальным модификациям в выбранных доменах, описанным в настоящем документе. Мутации могут быть введены в любом одном или более доменах референсного белка CasX или в варианте CasX с получением варианта CasX и могут включать в себя, например, полную или частичную делецию одного или более доменов или одну или более аминокислотных замен, делеций или вставок в любом домене референсного белка CasX или варианта CasX, из которого они получены. Домены белков CasX включают в себя домен связывания с нецелевой цепью (NTSB), домен загрузки целевой цепи (TSL), домен спиральный I, домен спиральный II, олигонуклеотид-связывающий домен (OBD) и домен расщепления ДНК RuvC. Без ограничений, накладываемых теорией или механизмом, домен NTSB в CasX обеспечивает связывание с нецелевой цепью нуклеиновой кислоты и может способствовать раскручиванию нецелевой и целевой цепей. Предполагается, что домен NTSB отвечает за раскручивание или захват нецелевой цепи нуклеиновой кислоты в нераскрученном состоянии. Пример домена NTSB содержит аминокислоты 100-190 из SEQ ID NO: 1 или аминокислоты 102-191 из SEQ ID NO: 2. В некоторых вариантах осуществления домен NTSB референсного белка CasX содержит четырехцепочечный бета-лист. В некоторых вариантах осуществления функцией TSL является размещение или захват целевой цепи в свернутое состоянии, в результате чего расщепляемый фосфат целевой цепи основной цепи ДНК попадает в активный сайт RuvC. Пример TSL содержит аминокислоты 824-933 из SEQ ID NO: 1 или аминокислоты 811-920 из SEQ ID NO: 2. Без ограничений, накладываемых какой-либо теорией, считается, что в некоторых случаях домен спиральный I может способствовать связыванию смежного с протоспейсером мотива (PAM). В некоторых вариантах осуществления домен спиральный I референсного белка CasX содержит одну или более альфа-спиралей. Примеры доменов спиральный I_I и I-II содержат аминокислоты 56-99 и 191-331 из SEQ ID NO: 1 соответственно или аминокислоты 58-101 и 192-332 из SEQ ID NO: 2 соответственно. Домен спиральный II отвечает за связывание с каркасной стеблевой петлей гРНК, а также связанной ДНК. Пример домена спирального II содержит аминокислоты 332-508 из SEQ ID NO: 1 или аминокислоты 333-500 из SEQ ID NO: 2. OBD в основном связывает РНК-триплекс каркаса гидовой РНК. OBD может также отвечать за связывание со смежным с протоспейсером мотивом (PAM). Примеры доменов OBD I и II содержат аминокислоты 1-55 и 509-659 из SEQ ID NO: 1 соответственно или аминокислоты 1-57 и 501-646 из SEQ ID NO: 2 соответственно. RuvC имеет активный сайт мотива DED, который отвечает за расщепление обеих цепей ДНК (по одной, наиболее вероятно, сначала нецелевую цепь на 11-14 нуклеотидах (нт) в целевой последовательности, а затем целевую цепь на 2-4 нуклеотидах после целевой последовательности, что приводит к смещенному

разрезу). В частности, в CasX домен RuvC является уникальным в том смысле, что он также отвечает за связывание каркасной стеблевой петли гидовой РНК, которая имеет критическое значение для функции CasX. Примеры доменов RuvC I и II включают в себя аминокислоты 660-823 и 934-986 из SEQ ID NO: 1 соответственно или аминокислоты 647-810 и 921-978 из SEQ ID NO: 2 соответственно, хотя варианты CasX могут содержать мутации в положениях I658 и A708 относительно SEQ ID NO: 2 или мутации CasX 515, описанные ниже.

[0206] В некоторых вариантах осуществления вариант белка CasX содержит по меньшей мере одну модификацию в по меньшей мере 1 домене, в по меньшей мере каждом из 2 доменов, в по меньшей мере каждом из 3 доменов, в по меньшей мере каждом из 4 доменов или в по меньшей мере каждом из 5 доменов референсного белка CasX, включающего последовательности с SEQ ID NO: 1-3. В некоторых вариантах осуществления вариант белка CasX содержит две или более модификации в по меньшей мере одном домене референсного белка CasX. В некоторых вариантах осуществления вариант белка CasX содержит по меньшей мере две модификации в по меньшей мере одном домене референсного белка CasX, по меньшей мере три модификации в по меньшей мере одном домене референсного белка CasX или по меньшей мере четыре или более модификаций в по меньшей мере одном домене референсного белка CasX. В некоторых вариантах осуществления, в которых вариант CasX содержит две или более модификации по сравнению с референсным белком CasX, и каждая модификация осуществляется в домене, независимо выбранном из группы, состоящей из NTSB, TSL, домена спиральный I, домена спиральный II, OBD и домена расщепления ДНК RuvC. В некоторых вариантах осуществления, в которых вариант CasX содержит две или более модификации по сравнению с референсным белком CasX, модификация выполняется в двух или более доменах. В некоторых вариантах осуществления по меньшей мере одна модификация варианта белка CasX включает делецию по меньшей мере участка одного домена референсного белка CasX с SEQ ID NO: 1-3. В некоторых вариантах осуществления делеция находится в домене NTSB, домене TSL, домене спиральный I, домене спиральный II, OBD или домене расщепления ДНК RuvC.

[0207] В некоторых случаях варианты CasX описания содержат модификации в структурных областях, которые могут охватывать один или более доменов. В некоторых вариантах осуществления вариант CasX содержит по меньшей мере одну модификацию области несмежных аминокислотных остатков варианта CasX, которые образуют канал, в котором происходит образование комплекса гРНК : целевая нуклеиновая кислота с вариантом CasX. В других вариантах осуществления вариант CasX содержит по меньшей мере одну модификацию области несмежных аминокислотных остатков варианта CasX, которые образуют поверхность взаимодействия, связывающуюся с гРНК. В других вариантах осуществления вариант CasX содержит по меньшей мере одну модификацию области несмежных аминокислотных остатков варианта CasX, которые образуют канал, связывающийся с нецелевой цепью ДНК. В других вариантах осуществления вариант CasX

содержит по меньшей мере одну модификацию области несмежных аминокислотных остатков варианта CasX, которые образуют поверхность взаимодействия, связывающуюся со смежным с протоспейсером мотивом (PAM) целевой нуклеиновой кислоты. В других вариантах осуществления вариант CasX содержит по меньшей мере одну модификацию области несмежных поверхностных аминокислотных остатков варианта CasX. В других вариантах осуществления вариант CasX содержит по меньшей мере одну модификацию области несмежных аминокислотных остатков, которые образуют ядро посредством гидрофобной упаковки в домене варианта CasX. В вышеуказанных вариантах осуществления параграфа модификации области могут включать в себя одно или более из делеции, вставки или замены одной или более аминокислот в области; или от 2 до 15 аминокислотных остатков области варианта CasX замещены заряженными аминокислотами; или от 2 до 15 аминокислотных остатков области варианта CasX замещены полярными аминокислотами; или от 2 до 15 аминокислотных остатков в области варианта CasX замещены аминокислотами, которые образуют стопку или имеют аффинность к основаниям ДНК или РНК.

[0208] В других вариантах осуществления в описании предложены варианты CasX, причем варианты CasX содержат по меньшей мере одну модификацию относительно другого варианта CasX; например, варианты CasX 515 и 527 представляют собой варианты варианта CasX 491, а варианты CasX 668 и 672 представляют собой варианты CasX 535 (см. ФИГ. 44). В некоторых вариантах осуществления по меньшей мере одна модификация выбрана из группы, состоящей из вставки, делеции или замены аминокислоты. Предполагается, что все варианты, которые улучшают одну или более функций или характеристик варианта белка CasX по сравнению с референсным белком CasX или вариантом, из которого он получен, описанным в настоящем документе, находятся в рамках объема описания. Как описано в разделе «Примеры», вариант CasX может быть подвергнут мутации с образованием другого варианта CasX. В конкретном варианте осуществления в описании, в примере 14, предложены варианты белка CasX 515 (SEQ ID NO: 416), созданные посредством введения модификаций в кодирующую последовательность, что приводит к аминокислотным заменам, делециям или вставкам в одном или более положениях в одном или более доменах.

[0209] Приемлемые способы мутагенеза для создания вариантов белка CasX описания могут включать в себя, например, глубокую мутационную эволюцию (DME), глубокое мутационное сканирование (DMS), подверженную ошибкам ПЦР, кассетный мутагенез, случайный мутагенез, ПЦР со ступенчатой достройкой, перетасовку генов или перестановку доменов (описаны в PCT/US20/36506 и WO2020247883A2, включенных в настоящий документ путем ссылки). В некоторых вариантах осуществления варианты CasX разработаны, например, посредством выбора множества требуемых мутаций в выявленном варианте CasX, например, с использованием анализов, описанных в разделе «Примеры». В определенных вариантах осуществления активность референсного CasX или варианта белка CasX до мутагенеза используется в качестве эталона, с которым сравнивают

активность одного или более полученных вариантов CasX, измеряя таким образом улучшения функции у новых вариантов CasX.

[0210] В некоторых вариантах осуществления вариантов CasX, описанных в настоящем документе, по меньшей мере одна модификация содержит: (a) замену от 1 до 100 последовательных или непоследовательных аминокислот в варианте CasX по сравнению с референсным CasX с SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, вариантом 491 CasX (SEQ ID NO: 336) и вариантом CasX 515 (SEQ ID NO: 416); (b) делецию от 1 до 100 последовательных или непоследовательных аминокислот в варианте CasX по сравнению с референсным CasX или вариантом, из которого он получен; (c) вставку от 1 до 100 последовательных или непоследовательных аминокислот в CasX по сравнению с референсным CasX или вариантом, из которого он получен; или (d) любую комбинацию пп. (a)-(c). В некоторых вариантах осуществления по меньшей мере одна модификация включает: (a) замену 1-10 последовательных или непоследовательных аминокислот в варианте CasX по сравнению с референсным CasX с SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3 или вариантом, из которого он получен; (b) делецию от 1 до 5 последовательных или непоследовательных аминокислот в варианте CasX по сравнению с референсным CasX или вариантом, из которого он получен; (c) вставку от 1 до 5 последовательных или непоследовательных аминокислот в CasX по сравнению с референсным CasX или вариантом, из которого он получен; или (d) любую комбинацию пп. (a)-(c).

[0211] В некоторых вариантах осуществления вариант белка CasX содержит или состоит из последовательности, которая имеет по меньшей мере 1, по меньшей мере 2, по меньшей мере 3, по меньшей мере 4, по меньшей мере 5, по меньшей мере 6, по меньшей мере 7, по меньшей мере 8, по меньшей мере 9, по меньшей мере 10, по меньшей мере 20, по меньшей мере 30, по меньшей мере 40, по меньшей мере 50, по меньшей мере 60, по меньшей мере 70, по меньшей мере 80, по меньшей мере 90 или по меньшей мере 100 изменений относительно последовательности SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, CasX 491 или CasX 515. В некоторых вариантах осуществления вариант белка CasX содержит одну более замен относительно CasX 491 или SEQ ID NO: 336. В некоторых вариантах осуществления вариант белка CasX содержит одну более замен относительно CasX 515 или SEQ ID NO: 416. Эти изменения могут представлять собой аминокислотные вставки, делеции, замены или любые их комбинации. Изменения могут присутствовать в одном домене, или в любом домене, или в любой комбинации доменов варианта CasX. В заменах, описанных в настоящем документе, любая аминокислота может быть замещена на любую другую аминокислоту. Замещение может представлять собой консервативную замену (например, основную аминокислоту замещают другой основной аминокислотой). Замещение может представлять собой неконсервативную замену (например, основную аминокислоту замещают на кислотную аминокислоту, или наоборот). Например, пролин в референсном белке CasX может быть замещен на любое из аргинина, гистидина, лизина, аспарагиновой кислоты, глутаминовой кислоты, серина, треонина, аспарагина, глутамина, цистеина, глицина, аланина, изолейцина, лейцина, метионина, фенилаланина, триптофана,

тирозина или валина для создания варианта белка CasX описания.

[0212] Можно комбинировать любые сочетания вариантов осуществления с заменой, вставкой и делецией, описанных в настоящем документе, для создания варианта белка CasX описания. Например, вариант белка CasX может содержать по меньшей мере одну замену и по меньшей мере одну делецию относительно последовательности референсного белка CasX или последовательности CasX 491 или CasX 515, по меньшей мере одну замену и по меньшей мере одну вставку относительно последовательности референсного белка CasX или последовательности CasX 491 или CasX 515, по меньшей мере одну вставку и по меньшей мере одну делецию относительно последовательности референсного белка CasX или последовательности CasX 491 или CasX 515, или по меньшей мере одну замену, одну вставку и одну делецию относительно последовательности референсного белка CasX или последовательности CasX 491 или CasX 515.

[0213] В некоторых вариантах осуществления вариант белка CasX содержит от 400 до 2000 аминокислот, от 500 до 1500 аминокислот, от 700 до 1200 аминокислот, от 800 до 1100 аминокислот или от 900 до 1000 аминокислот.

[0214] В некоторых вариантах осуществления вариант белка CasX содержит последовательность с SEQ ID NO: 247-592 или 1147-1231, как указано в таблице 3. В некоторых вариантах осуществления вариант белка CasX состоит из последовательности с SEQ ID NO: 247-592 или 1147-1231, как указано в таблице 3. В других вариантах осуществления вариант белка CasX содержит последовательность, которая на по меньшей мере 60% идентична, на по меньшей мере 65% идентична, на по меньшей мере 70% идентична, на по меньшей мере 75% идентична, на по меньшей мере 80% идентична, на по меньшей мере 81% идентична, на по меньшей мере 82% идентична, на по меньшей мере 83% идентична, на по меньшей мере 84% идентична, на по меньшей мере 85% идентична, на по меньшей мере 86% идентична, на по меньшей мере 86% идентична, на по меньшей мере 87% идентична, на по меньшей мере 88% идентична, на по меньшей мере 89% идентична, на по меньшей мере 89% идентична, на по меньшей мере 90% идентична, на по меньшей мере 91% идентична, на по меньшей мере 92% идентична, на по меньшей мере 93% идентична, на по меньшей мере 94% идентична, на по меньшей мере на 95% идентична, на по меньшей мере 96% идентична, по меньшей мере на 97% идентична, на по меньшей мере 98% идентична, на по меньшей мере 99% идентична, на по меньшей мере 99,5% идентична последовательности с SEQ ID NO: 247-592 и 1147-1231, как указано в таблице 3. В некоторых вариантах осуществления вариант белка CasX содержит или состоит из последовательности с SEQ ID NO: 270-592 или 1147-1231. В других вариантах осуществления вариант белка CasX содержит последовательность, которая на по меньшей мере 60% идентична, на по меньшей мере 65% идентична, на по меньшей мере 70% идентична, на по меньшей мере 75% идентична, на по меньшей мере 80% идентична, на по меньшей мере 81% идентична, на по меньшей мере 82% идентична, на по меньшей мере 83% идентична, на по меньшей мере 84% идентична, на по меньшей мере 85% идентична, на по меньшей мере 86% идентична, на по меньшей мере 86% идентична, на по меньшей

мере 87% идентична, на по меньшей мере 88% идентична, на по меньшей мере 89% идентична, на по меньшей мере 89% идентична, на по меньшей мере 90% идентична, на по меньшей мере 91% идентична, на по меньшей мере 92% идентична, на по меньшей мере 93% идентична, на по меньшей мере 94% идентична, на по меньшей мере на 95% идентична, на по меньшей мере 96% идентична, по меньшей мере на 97% идентична, на по меньшей мере 98% идентична, на по меньшей мере 99% идентична, на по меньшей мере 99,5% идентична последовательности с SEQ ID NO: 270-592 или 1147-1231. В некоторых вариантах осуществления вариант белка CasX содержит или состоит из последовательности с SEQ ID NO: 415-592 или 1147-1231. В других вариантах осуществления вариант белка CasX содержит последовательность, которая на по меньшей мере 60% идентична, на по меньшей мере 65% идентична, на по меньшей мере 70% идентична, на по меньшей мере 75% идентична, на по меньшей мере 80% идентична, на по меньшей мере 81% идентична, на по меньшей мере 82% идентична, на по меньшей мере 83% идентична, на по меньшей мере 84% идентична, на по меньшей мере 85% идентична, на по меньшей мере 86% идентична, на по меньшей мере 86% идентична, на по меньшей мере 87% идентична, на по меньшей мере 88% идентична, на по меньшей мере 89% идентична, на по меньшей мере 89% идентична, на по меньшей мере 90% идентична, на по меньшей мере 91% идентична, на по меньшей мере 92% идентична, на по меньшей мере 93% идентична, на по меньшей мере 94% идентична, на по меньшей мере на 95% идентична, на по меньшей мере 96% идентична, по меньшей мере на 97% идентична, на по меньшей мере 98% идентична, на по меньшей мере 99% идентична, на по меньшей мере 99,5% идентична последовательности с SEQ ID NO: 416-592 или 1147-1231. (НД=не описано или не предоставлено иным образом).

Таблица 3. Последовательности вариантов CasX

SEQ ID NO	Вариант №	Описание варианта
247	НД	Домены TSL, спиральный I, спиральный II, OBD и RuvC из SEQ ID NO: 2 и домен NTSB из SEQ ID NO: 1
248	НД	Домены NTSB, спиральный I, спиральный II, OBD и RuvC из SEQ ID NO: 2 и домен TSL из SEQ ID NO: 1
249	НД	Домены TSL, спиральный I, спиральный II, OBD и RuvC из SEQ ID NO: 1 и домен NTSB из SEQ ID NO: 2
250	НД	Домены NTSB, спиральный I, спиральный II, OBD и RuvC из SEQ ID NO: 1 и домен TSL из SEQ ID NO: 2
251	НД	Домены NTSB, TSL, спиральный I, спиральный II и OBD из SEQ ID NO:2 и экзогенный домен RuvC или его часть из второго белка CasX.
252	НД	Без описания
253	НД	Домены NTSB, TSL, спиральный II, OBD и RuvC из SEQ ID NO: 2 и домен спиральный I из SEQ ID NO: 1

254	НД	Домены NTSB, TSL, спиральный I, OBD и RuvC из SEQ ID NO: 2 и домен спиральный II из SEQ ID NO: 1
255	НД	Домены NTSB, TSL, спиральный I, спиральный II и RuvC из первого белка CasX и экзогенный OBD или его часть из второго белка CasX
256	НД	Без описания
257	НД	Без описания
258	НД	Замена L379R, замена C477K, замена A708K, делеция P в положении 793 и замена T620P в SEQ ID NO: 2
259	НД	Замена M771A в SEQ ID NO: 2
260	НД	Замена L379R, замена A708K, делеция P в положении 793 и замена D732N в SEQ ID NO: 2
261	НД	Замена W782Q в SEQ ID NO: 2
262	НД	Замена M771Q в SEQ ID NO: 2
263	НД	Замена R458I и замена A739V в SEQ ID NO: 2
264	НД	L379R, замена A708K, делеция P в положении 793 и замена M771N в SEQ ID NO: 2
265	НД	Замена L379R, замена A708K, делеция P в положении 793 и замена A739T в SEQ ID NO: 2
266	НД	Замена L379R, замена C477K, замена A708K, делеция P в положении 793 и замена D489S в SEQ ID NO: 2.
267	НД	Замена L379R, замена C477K, замена A708K, делеция P в положении 793 и замена D732N в SEQ ID NO: 2.
268	НД	Замена V711K в SEQ ID NO: 2
269	НД	Замена L379R, замена C477K, замена A708K, делеция P в положении 793 и замена Y797L в SEQ ID NO: 2.
270	119	НД
283	429	НД
284	430	НД
285	431	НД
286	432	НД
287	433	НД
288	434	НД
289	435	НД
290	436	НД

291	437	НД
292	438	НД
293	439	НД
294	440	НД
295	441	НД
296	442	НД
297	443	НД
298	444	НД
299	445	НД
300	446	НД
301	447	НД
302	448	НД
303	449	НД
304	450	НД
305	451	НД
306	452	НД
307	453	НД
308	454	НД
309	455	НД
310	456	НД
311	457	НД
312	458	НД
313	459	НД
314	460	НД
315	278	НД
316	279	НД
317	280	НД
318	285	НД
319	286	НД
320	287	НД

321	288	НД
322	290	НД
323	291	НД
324	293	НД
325	300	НД
326	492	НД
327	493	НД
328	387	НД
329	395	НД
330	485	НД
331	486	НД
332	487	НД
333	488	НД
334	489	НД
335	490	НД
336	491	НД
337	494	НД
411	328	НД
412	388	НД
413	389	НД
414	390	НД
415	514	НД
416	515	НД
417	516	НД
418	517	НД
419	518	НД
420	519	НД
421	520	НД
422	522	НД
423	523	НД

424	524	НД
425	525	НД
426	526	НД
427	527	НД
428	528	НД
429	529	НД
430	530	НД
431	531	НД
432	532	НД
433	533	НД
434	534	НД
435	535	НД
436	536	НД
437	537	НД
438	538	НД
439	539	НД
440	540	НД
441	541	НД
442	542	НД
443	543	НД
444	544	НД
445	545	НД
446	546	НД
447	547	НД
448	548	НД
449	550	НД
450	551	НД
451	552	НД
452	553	НД
453	554	НД

454	555	НД
455	556	НД
456	557	НД
457	558	НД
458	559	НД
459	560	НД
460	561	НД
461	562	НД
462	563	НД
563	564	НД
464	565	НД
465	566	НД
466	567	НД
467	568	НД
468	569	НД
469	570	НД
470	571	НД
471	572	НД
472	573	НД
473	574	НД
474	575	НД
475	576	НД
476	577	НД
477	578	НД
478	579	НД
479	580	НД
480	581	НД
481	582	НД
482	583	НД
483	584	НД

484	585	НД
485	586	НД
486	587	НД
487	588	НД
488	589	НД
489	590	НД
490	591	НД
491	592	НД
492	593	НД
493	594	НД
494	595	НД
495	596	НД
496	597	НД
497	598	НД
498	599	НД
499	600	НД
500	601	НД
501	602	НД
502	603	НД
503	604	НД
504	605	НД
505	606	НД
506	607	НД
507	608	НД
508	609	НД
509	610	НД
510	611	НД
511	612	НД
512	613	НД
513	614	НД

514	615	НД
515	616	НД
516	617	НД
517	618	НД
518	619	НД
519	620	НД
520	621	НД
521	622	НД
522	623	НД
523	624	НД
524	625	НД
525	626	НД
526	627	НД
527	628	НД
528	629	НД
529	630	НД
530	631	НД
531	632	НД
532	633	НД
533	634	НД
534	635	НД
535	636	НД
536	637	НД
537	638	НД
538	639	НД
539	640	НД
540	641	НД
541	642	НД
542	643	НД
543	644	НД

544	645	НД
545	646	НД
546	647	НД
547	648	НД
548	649	НД
549	650	НД
550	651	НД
551	652	НД
552	653	НД
553	654	НД
554	655	НД
555	656	НД
556	657	НД
557	658	НД
558	659	НД
559	660	НД
560	661	НД
561	662	НД
562	663	НД
563	664	НД
564	665	НД
565	666	НД
566	667	НД
567	668	НД
568	669	НД
569	671	НД
570	672	НД
571	673	НД
572	674	НД
573	675	НД

574	676	НД
575	677	НД
576	678	НД
577	679	НД
578	680	НД
579	681	НД
580	682	НД
581	683	НД
582	684	НД
583	685	НД
584	686	НД
585	687	НД
586	688	НД
587	689	НД
588	690	НД
589	691	НД
590	692	НД
591	693	НД
592	694	НД
1147	701	НД
1148	702	НД
1149	703	НД
1150	704	НД
1151	705	НД
1152	706	НД
1153	707	НД
1154	708	НД
1155	709	НД
1156	710	НД
1157	711	НД

1158	712	НД
1159	713	НД
1160	714	НД
1161	715	НД
1162	716	НД
1163	717	НД
1164	718	НД
1165	719	НД
1166	720	НД
1167	721	НД
1168	722	НД
1169	723	НД
1170	724	НД
1171	725	НД
1172	726	НД
1173	727	НД
1174	728	НД
1175	729	НД
1176	730	НД
1177	731	НД
1178	732	НД
1179	733	НД
1180	734	НД
1181	735	НД
1182	736	НД
1183	737	НД
1184	738	НД
1185	739	НД
1186	740	НД
1187	741	НД

1188	742	НД
1189	743	НД
1190	744	НД
1191	745	НД
1192	746	НД
1193	747	НД
1194	748	НД
1195	749	НД
1196	750	НД
1197	751	НД
1198	752	НД
1199	753	НД
1200	754	НД
1201	755	НД
1202	756	НД
1203	757	НД
1204	758	НД
1205	759	НД
1206	760	НД
1207	761	НД
1208	762	НД
1209	763	НД
1210	764	НД
1211	765	НД
1212	766	НД
1213	767	НД
1214	768	НД
1215	769	НД
1216	770	НД
1217	777	НД

1218	778	НД
1219	779	НД
1220	780	НД
1221	781	НД
1222	782	НД
1223	783	НД
1224	784	НД
1225	785	НД
1226	786	НД
1227	787	НД
1228	788	НД
1229	789	НД
1230	790	НД
1231	791	НД

c. Варианты белков CasX с доменами из множества исходных белков

[0215] В определенных вариантах осуществления в описании предложен химерный белок CasX, содержащий белковые домены из двух или более разных белков CasX, таких как два или более природных белков CasX, или две или более последовательностей вариантов белков CasX, описанных в настоящем документе. В настоящем документе термин «химерный белок CasX» относится к CasX, содержащему по меньшей мере два домена, выделенных или полученных из разных источников, таких как два природных белка, которые в некоторых вариантах осуществления могут быть выделены из разных биологических видов. Например, в некоторых вариантах осуществления химерный белок CasX содержит первый домен из первого белка CasX и второй домена из второго, другого белка CasX. В некоторых вариантах осуществления первый домен может быть выбран из группы, состоящей из доменов NTSB, TSL, спирального I, спирального II, OBD и RuvC. В некоторых вариантах осуществления второй домен выбран из группы, состоящей из NTSB, TSL, спирального I, спирального II, OBD и RuvC, причем второй домен отличается от вышеуказанного первого домена. В конкретном варианте осуществления варианты CasX 514-791 (SEQ ID NO: 415-592 и 1147-1231) имеют домен NTSB и спиральный 1B, полученные из последовательности SEQ ID NO: 1, а другие домены получены из SEQ ID NO: 2, и следует понимать, что эти варианты имеют дополнительные аминокислотные изменения в выбранных местоположениях.

d. Аффинность белка к гРНК

[0216] В некоторых вариантах осуществления вариант белка CasX имеет

улучшенную аффинность к гРНК относительно референсного белка CasX, что приводит к образованию рибонуклеопротеинового комплекса (RNP). Повышенная аффинность варианта белка CasX к гРНК может, например, приводить к более низкой K_d образования комплекса RNP, что в некоторых случаях может приводить к образованию более стабильного рибонуклеопротеинового комплекса. В некоторых вариантах осуществления повышенная аффинность варианта белка CasX к гРНК приводит к повышению стабильности рибонуклеопротеинового комплекса при доставке в человеческие клетки. Эта повышенная стабильность может влиять на функциональность и пригодность комплекса в клетках субъекта, а также приводить к улучшенным фармакокинетическим свойствам в крови при доставке субъекту. В некоторых вариантах осуществления повышенная аффинность варианта белка CasX и полученная повышенная стабильность рибонуклеопротеинового комплекса позволяют доставлять более низкую дозу варианта белка CasX субъекту или к клетке с сохранением требуемой активности, например редактирования гена in vivo или *in vitro*. В некоторых вариантах осуществления более высокая аффинность (более плотное связывание) варианта белка CasX с гРНК обеспечивает большее число событий редактирования, когда и вариант белка CasX, и гРНК остаются в комплексе RNP. Увеличение числа событий редактирования можно оценивать с применением анализов редактирования, таких как анализы редактирования tdTom, описанные в настоящем документе. В некоторых вариантах осуществления K_d связи варианта белка CasX с гРНК увеличивается относительно референсного белка CasX с коэффициентом по меньшей мере примерно 1,1, по меньшей мере примерно 1,2, по меньшей мере примерно 1,3, по меньшей мере примерно 1,4, по меньшей мере примерно 1,5, по меньшей мере примерно 1,6, по меньшей мере примерно 1,7, по меньшей мере примерно 1,8, по меньшей мере примерно 1,9, по меньшей мере примерно 2, по меньшей мере примерно 3, по меньшей мере примерно 4, по меньшей мере примерно 5, по меньшей мере примерно 6, по меньшей мере примерно 7, по меньшей мере примерно 8, по меньшей мере примерно 9, по меньшей мере примерно 10, по меньшей мере примерно 15, по меньшей мере примерно 20, по меньшей мере примерно 25, по меньшей мере примерно 30, по меньшей мере примерно 35, по меньшей мере примерно 40, по меньшей мере примерно 45, по меньшей мере примерно 50, по меньшей мере примерно 60, по меньшей мере примерно 70, по меньшей мере примерно 80, по меньшей мере примерно 90 или по меньшей мере примерно 100. В некоторых вариантах осуществления вариант CasX имеет кратность повышения аффинности связывания с гРНК от примерно 1,1 до примерно 10 раз по сравнению с референсным белком CasX с SEQ ID NO: 2.

[0217] В некоторых вариантах осуществления повышение аффинности варианта белка CasX к гРНК приводит к повышению стабильности рибонуклеопротеинового комплекса при доставке в клетки млекопитающих, в том числе при доставке in vivo субъекту. Эта повышенная стабильность может влиять на функциональность и пригодность комплекса в клетках субъекта, а также приводить к улучшенным фармакокинетическим свойствам в крови при доставке субъекту. В некоторых вариантах осуществления

повышенная аффинность варианта белка CasX и полученная повышенная стабильность рибонуклеопротеинового комплекса позволяют доставлять более низкую дозу варианта белка CasX субъекту или в клетки с сохранением требуемой активности; например, редактирования генов in vivo или *in vitro*. Улучшенная способность к образованию RNP и поддержанию их в стабильной форме можно оценивать с помощью анализов, таких как анализы расщепления *in vitro*, описанные в разделе «Примеры» в настоящем документе. В некоторых вариантах осуществления RNP, содержащий варианты CasX описания, способен достигать константы скорости $k_{расщепл.}$ в комплексе в виде RNP, которая в по меньшей мере 2 раза, по меньшей мере 5 раз или по меньшей мере 10 раз выше по сравнению с таковой для RNP, содержащего референсный CasX с SEQ ID NO: 1-3.

[0218] В некоторых вариантах осуществления более высокая аффинность (более плотное связывание) варианта белка CasX с гРНК обеспечивает большее число событий редактирования, когда и вариант белка CasX, и гРНК остаются в комплексе RNP. Увеличение числа событий редактирования можно оценивать с применением анализов редактирования, таких как анализы, описанные в настоящем документе.

[0219] Без ограничений, накладываемых какой-либо теорией, в некоторых вариантах осуществления аминокислотные изменения в домене спиральный I могут повышать аффинность связывания варианта белка CasX с нацеливающей последовательностью гРНК, тогда как изменения в домене спиральный II могут повышать аффинность связывания варианта белка CasX с каркасной стеблевой петлей гРНК, а изменения в олигонуклеотид-связывающем домене (OBD) повышают аффинность связывания варианта белка CasX с триплексом гРНК.

[0220] Способы измерения аффинности связывания белка CasX с гРНК включают способы *in vitro* с применением очищенного белка CasX и гРНК. Аффинность связывания для референсного CasX и вариантов белков может быть измерена посредством флуоресцентной поляризации, если гРНК или белок CasX помечен флуорофором. Альтернативно или в дополнение к этому аффинность связывания может быть измерена посредством интерферометрии биослоев, анализа методом сдвига электрофоретической подвижности (EMSA) или связывания на фильтрах. Дополнительные стандартные методики количественного определения абсолютных аффинностей связывающих РНК белков, таких как референсный CasX и варианты белков описания для конкретных гРНК, таких как референсные гРНК и их варианты, включают, без ограничений, изотермическую калориметрию (ITC) и поверхностный плазмонный резонанс (SPR), а также способы из раздела «Примеры».

[0221] Каталитически неактивный В некоторых вариантах осуществления вариант белка CasX содержит одну или более модификаций в области несмежных остатков, которые образуют канал, в котором происходит образование комплекса гРНК : целевая нуклеиновая кислота. В некоторых вариантах осуществления вариант белка CasX содержит одну или более модификаций, содержащих область несмежных остатков, образующих поверхность взаимодействия, которая связывается с гРНК. Например, в некоторых вариантах

осуществления референсного белка CasX все из доменов спиральный I, спиральный II и OBD контактируют или находятся поблизости от комплекса гРНК : целевая нуклеиновая кислота, и одна или более модификаций несмежных остатков в любом из этих доменов могут улучшать функцию варианта белка CasX.

[0222] В некоторых вариантах осуществления вариант белка CasX содержит одну или более модификаций в области несмежных остатков, которые образуют канал, который связывается с нецелевой цепью ДНК. Например, вариант белка CasX может содержать одну или более модификаций несмежных остатков домена NTSB. В некоторых вариантах осуществления вариант белка CasX содержит одну или более модификаций в области несмежных остатков, которые образуют поверхность взаимодействия, которая связывается с PAM. Например, вариант белка CasX может содержать одну или более модификаций несмежных остатков в домене спиральный I или OBD. В некоторых вариантах осуществления вариант белка CasX содержит одну или более модификаций, содержащих область несмежных поверхностных остатков. В настоящем документе термин «поверхностные остатки» относится к аминокислотам на поверхности белка CasX или аминокислотам, у которых по меньшей мере часть аминокислоты, такая как основная цепь или часть боковой цепи, находится на поверхности белка. Поверхностные остатки клеточных белков, таких как CasX, которые открыты для воздействия водной внутриклеточной среды, часто выбирают из положительно заряженных гидрофильных аминокислот, например аргинина, аспарагина, аспартата, глутамина, глутамата, гистидина, лизина, серина и треонина. Таким образом, например, в некоторых вариантах осуществления вариантов, предложенных в настоящем документе, область поверхностных остатков содержит одну или более вставок, делеций или замен по сравнению с референсным белком CasX. В некоторых вариантах осуществления один или более положительно заряженных остатков замещены одним или более другими положительно заряженными остатками, или отрицательно заряженными остатками, или незаряженными остатками, или любыми их комбинациями. В некоторых вариантах осуществления один или более аминокислотных остатков для замены находятся вблизи связанной нуклеиновой кислоты, например остатки в домене RuvC или домене спиральный I, которые контактируют с целевой нуклеиновой кислотой, или остатки в доменах OBD или спиральный II, которые связывают гРНК, могут быть замещены одной или более положительно заряженными или полярными аминокислотами.

[0223] В некоторых вариантах осуществления вариант белка CasX содержит одну или более модификаций в области несмежных остатков, которые образуют ядро посредством гидрофобной упаковки в домене референсного белка CasX. Без ограничений, накладываемых какой-либо теорией, области, которые образуют ядро посредством гидрофобной упаковки, богаты гидрофобными аминокислотами, такими как валин, изолейцин, лейцин, метионин, фенилаланин, триптофан и цистеин. Например, в некоторых референсных белках CasX домены RuvC содержат гидрофобный карман, смежный с активным сайтом. В некоторых вариантах осуществления от 2 до 15 остатков в области

являются заряженными, полярными или вступают в межплоскостное взаимодействие оснований. Заряженные аминокислоты (иногда называемые в настоящем документе остатками) могут включать в себя, например, аргинин, лизин, аспарагиновую кислоту и глутаминовую кислоту, и боковые цепи этих аминокислот могут образовывать солевые мостики при условии наличия партнера по мостику (см. ФИГ. 14). Полярные аминокислоты могут включать в себя, например, глутамин, аспарагин, гистидин, серин, треонин, тирозин и цистеин. Полярные аминокислоты в некоторых вариантах осуществления могут образовывать водородные связи в качестве доноров или акцепторов протонов в зависимости от характера их боковых цепей. В настоящем документе термин «межплоскостное взаимодействие оснований» включает в себя взаимодействие ароматических боковых цепей аминокислотного остатка (такого как триптофан, тирозин, фенилаланин или гистидин) со стопкой нуклеотидных оснований в нуклеиновой кислоте. В рамках объема описания предусмотрена любая модификация в области несмежных аминокислот, которые находятся в непосредственной близости с образованием функциональной части варианта белка CasX.

e. Варианты белков CasX с доменами из множества исходных белков

[0224] В определенных вариантах осуществления в описании предложен химерный вариант белка CasX, содержащий белковые домены из двух или более разных белков CasX, таких как два или более природных белков CasX, или две или более последовательностей вариантов белков CasX, описанных в настоящем документе. В настоящем документе термин «химерный белок CasX» относится к CasX, содержащему по меньшей мере два домена, выделенных или полученных из разных источников, таких как два природных белка, которые в некоторых вариантах осуществления могут быть выделены из разных биологических видов. Например, в некоторых вариантах осуществления химерный белок CasX содержит первый домен из первого белка CasX и второй домена из второго, другого белка CasX. В некоторых вариантах осуществления первый домен может быть выбран из группы, состоящей из доменов NTSB, TSL, спирального I, спирального II, OBD и RuvC. В некоторых вариантах осуществления второй домен выбран из группы, состоящей из NTSB, TSL, спирального I, спирального II, OBD и RuvC, причем второй домен отличается от вышеуказанного первого домена. Например, химерный белок CasX может содержать домены NTSB, TSL, спиральный I, спиральный II, OBD из белка CasX с SEQ ID NO: 2 и домен RuvC из белка CasX с SEQ ID NO: 1, или наоборот. В качестве дополнительного примера химерный белок CasX может содержать домены NTSB, TSL, спиральный II, OBD и RuvC из белка CasX с SEQ ID NO: 2 и домен спиральный I из белка CasX с SEQ ID NO: 1, или наоборот. Таким образом, в определенных вариантах осуществления химерный белок CasX может содержать домены NTSB, TSL, спиральный II, OBD и RuvC из первого белка CasX и домен спиральный I из второго белка CasX. В некоторых вариантах осуществления химерных белков CasX домены первого белка CasX получены из последовательностей SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3, а домены второго белка CasX получены из последовательностей SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3, и первый и второй

белки CasX не являются одинаковыми. В некоторых вариантах осуществления домены первого белка CasX содержат последовательности, полученные из SEQ ID NO: 1, а домены второго белка CasX содержат последовательности, полученные из SEQ ID NO: 2. В некоторых вариантах осуществления домены первого белка CasX содержат последовательности, полученные из SEQ ID NO: 1, а домены второго белка CasX содержат последовательности, полученные из SEQ ID NO: 3. В некоторых вариантах осуществления домены первого белка CasX содержат последовательности, полученные из SEQ ID NO: 2, а домены второго белка CasX содержат последовательности, полученные из SEQ ID NO: 3. В качестве примера вышеизложенного, химерный домен RuvC содержит аминокислоты 660-823 из SEQ ID NO: 1 и аминокислоты 921-978 из SEQ ID NO: 2. В качестве альтернативного примера вышеизложенного химерный домен RuvC содержит аминокислоты 647-810 из SEQ ID NO: 2 и аминокислоты 934-986 из SEQ ID NO: 1. В некоторых вариантах осуществления по меньшей мере один химерный домен содержит химерный домен спиральный I, причем химерный домен спиральный I содержит аминокислоты 56-99 из SEQ ID NO: 1 и аминокислоты 192-332 из SEQ ID NO: 2. В некоторых вариантах осуществления химерный вариант CasX дополнительно модифицирован, включая варианты CasX, выбранные из группы, состоящей из последовательностей SEQ ID NO: 270, SEQ ID NO: 328, SEQ ID NO: 336, SEQ ID NO: 780, SEQ ID NO: 412, SEQ ID NO: 413, SEQ ID NO: 414, SEQ ID NO: 416, SEQ ID NO: 435, SEQ ID NO: 329, SEQ ID NO: 781, SEQ ID NO: 330, SEQ ID NO: 782, SEQ ID NO: 331, SEQ ID NO: 783, SEQ ID NO: 332, SEQ ID NO: 784, SEQ ID NO: 333, SEQ ID NO: 785, SEQ ID NO: 334, SEQ ID NO: 786, SEQ ID NO: 335, SEQ ID NO: 567, SEQ ID NO: 570, SEQ ID NO: 574, SEQ ID NO: 787 и SEQ ID NO: 788. В некоторых вариантах осуществления одна или более дополнительных модификаций содержат вставку, замену или делецию, как описано в настоящем документе.

[0225] В случае разделенных или несмежных доменов, таких как спиральный I, RuvC и OBD, участок несмежного домена может быть заменен на соответствующий участок из любого другого источника. Например, домен спиральный I-I (иногда называемый спиральный I-a) в SEQ ID NO: 2 можно заменить на соответствующую последовательность домена спиральный I-I из SEQ ID NO: 1 и т. п. Последовательности доменов из референсных белков CasX и их координаты представлены в таблице 4. Типичные примеры химерных белков CasX включают в себя варианты CasX 472-483, 485-491 и 515, последовательности которых приведены в таблице 3.

Таблица 4. Координаты доменов в референсных белках CasX

Доменное имя	Координаты в SEQ ID NO: 1	Координаты в SEQ ID NO: 2
OBD-I	1-55	1-57
спиральный I-I	56-99	58-101
NTSB	100-190	102-191
спиральный I-II	191-331	192-332
спиральный II	332-508	333-500
OBD-II	509-659	501-646
RuvC-I	660-823	647-810
TSL	824-933	811-920

RuvC-II	934-986	921-978

* Домены OBD I и II, спиральные I-I и I-II и RuvC I и II также называются в настоящем документе OBD a и b, спиральный I a и b и RuvC a и b.

[0226] Примеры последовательностей доменов представлены ниже в таблице 5.

Таблица 5. Примеры последовательностей доменов

Deltaproteobacter sp. (референсный CasX с SEQ ID NO: 1)		
SEQ ID	Домен	Последовательность
2333	OBD-I	EKRINKIRKKLSADNATKPVSRSGPMKTLLVRVMTDDLKKRLEKR RKKPEVMPQ
2334	спираль-ный I-I	VISNNAANNLRMLLDDYTKMKEAILQVYWQEFKDDHVGLMCKFA
2335	NTSB	QPASKKIDQNKLKPEMDEKGNLTTAGFACSQCGQPLFVYKLEQVS EKGKAYTNYFGRCNVAEHEKLILLAQLKPEKDSDEAVTYSLGKFG Q
2336	спираль-ный I-II	RALDFYSIHVTKESTHPVKPLAQIAGNRYASGPVGKALSDACMGTI ASFLSKYQDIIIEHQKVVKGNQKRLESLRELAGKENLEYPSVTLPPQ PHTKEGVDAYNEVIARVRMWVNLNLWQ KLKLSRDDAKPLLRLKGFPSF
2337	спираль-ный II	PVVERRENEVDWWNTINEVKKLIDAKRDMGRVFWSGVTAEKRNTI LEGYNYLPNENDHKKREGSLENPKKPAKRQFGDLLLYLEKKYAGD WGKVFDEAWERIDKKIAGLTSHIEREEARNAEDAQSKAVLTDWLR AKASFVLERLKEMDEKEFYACEIQLQKWYGDLRG NPFAVEAE
2338	OBD-II	NRVVDISGFSIGSDGHSIQYRNLLAWKYLENGKREFYLLMNYGKK GRIRFTDGTDIKKSGKWQGLLYGGGKAKVIDLTFDPDDEQLIILPLA FGTRQGREFIWNDLLSLETGLIKLANGRVIEKTIYNKKIG RDEPALFVALTFERREVVD
2339	RuvC-I	PSNIKPVNLIGVDRGENIPAVIALTDPEGCPLPEFKDSSGGPTDILRIG EGYKEKQRAIQAAKEVEQRRAGGYSRKFASKSRNLADDMVRNSA RDLFYHAVTHDAVLVFENLSRGFGRQGKRTFMTERQYTKMEDWL TAKLAYEGLTSKTYLSKTLAQYTSKTC
2340	TSL	SNCGFTITTADYDGMLVRLKKTSDGWATTLNNKELKAEGQITYYN RYKRQTVEKELSAELDRLSEESGNNDISKWTKGRRDEALFLLKKRF SHRPVQEQFVCLDCGHEVH
2341	RuvC-II	ADEQAALNIARSWLFLN

		SNSTEFKSYKSGKQPFVGAWQAFYKRRLKEVWKPNA
Planctomycetes sp. (референсный CasX с SEQ ID NO: 2)		
SEQ ID	Домен	Последовательность
2342	OBD-I	QEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLEN LRKKPENIPQ
2343	спиральный I-II	PISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVA
2344	NTSB	QPAPKNIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVND KGKPHTNYFGRCNVSEHERLILLSPHKPEANDELVTYSLGKFGQ
2345	спиральный I-II	RALDFYSIHVTRESNHPVKPLEQIGGNSCASGPVGKALSDACMGAV ASFLTKYQDIILEHQKVIKKNEKRLANLKDIASANGLAFPKITLPPQP HTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRDEAKPLQRLKGFPS F
2346	спиральный II	PLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAGYKRQ EALLPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDE AWERIDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIE GLKEADKDEFCRCELKLQKWYGDLRGKPFAIEAE
2347	OBD-II	NSILDISGFSKQYNCAFIWQKDGVKKLNLYLIINYFKGGKLRFKKIK PEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLIILPLAFGKRQGR EFIWNDLLSLETGSLK LANGRVIEKTLYNRRTRQDEPALFVALTFERREVLD
2348	RuvC-I	SSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPTHILRIG ESYKEKQRTIQAAKEVEQRRAGGYSRKYASKAKNLADDMVRNTA RDLLYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWL TAKLAYEGLPSKTYLSKTLAQYTSKTC
2349	TSL	SNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQITYYNR YKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETH
2350	RuvC-II	ADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDKRAFVETWQSF YRKKLKEVWKPAV

[0227] Дополнительный пример последовательности домена спиральный II представлен в SEQ ID NO: 2351 и дополнительный пример последовательности домена RuvC представлен в SEQ ID NO: 2352.

[0228] В других вариантах осуществления вариант белка CasX содержит

последовательность с SEQ ID NO: 247-592 или 1147-1231, как указано в таблице 3, и дополнительно содержит один или более NLS, описанных в настоящем документе, на N-конце и/или C-конце или рядом с обоими. В других вариантах осуществления вариант белка CasX содержит последовательность с SEQ ID NO: 270-592 и 1147-1231 и дополнительно содержит один или более NLS, описанных в настоящем документе, на N-конце и/или C-конце или рядом с обоими. В других вариантах осуществления вариант белка CasX содержит последовательность с SEQ ID NO: 415-592 или 1147-1231 и дополнительно содержит один или более NLS, описанных в настоящем документе, на N-конце и/или C-конце или рядом с обоими. Следует понимать, что в некоторых случаях N-концевой метионин вариантов CasX, указанный в таблицах, удаляется из экспрессированного варианта CasX во время посттрансляционной модификации. Специалисту в данной области будет понятно, что NLS рядом с N- или C-концом белка может находиться в пределах 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20 или 20 аминокислот от N- или C-конца.

f. Варианты CasX, полученные из других вариантов CasX

[0229] В дополнительных итерациях создания вариантов белка вариант белка можно использовать для получения дополнительных вариантов CasX описания. Например, а также как представлено на ФИГ. 44, CasX 119 (SEQ ID NO: 270), CasX 491 (SEQ ID NO: 336) и CasX 515 (SEQ ID NO: 416) представляют собой примеры вариантов белка, которые были модифицированы для создания дополнительных вариантов CasX описания, имеющих улучшения или дополнительные свойства относительно референсных вариантов CasX или вариантов CasX, производными которых они являются. CasX 119 содержит замену L379R, замену A708K и делецию P в положении 793 в SEQ ID NO: 2. CasX 491 содержит перестановку доменов NTSB и спиральный 1B в SEQ ID NO: 1. CasX 515 был получен из CasX 491 посредством вставки P в положении 793 (относительно SEQ ID NO: 2), и его использовали для создания вариантов CasX, описанных в примерах 13 и 14. Например, CasX 668 имеет вставку R в положении 26 и замену G223S относительно CasX 515. CasX 672 имеет замены L169K и G223S относительно CasX 515. CasX 676 имеет замены L169K и G223S и вставку R в положении 26 относительно CasX 515.

[0230] Примеры способов, используемых для получения и оценки вариантов CasX, полученных из других вариантов CasX, описаны в примерах, которые были получены посредством внесения модификаций в кодирующую последовательность, что приводит к аминокислотным заменам, делециям или вставкам в одном или более положениях в одном или более доменах варианта CasX. В частности, в примере 14 и примере 15 описаны способы, используемые для создания вариантов CasX 515 (SEQ ID NO: 416), которые затем анализировали для определения тех положений в последовательности, которые при модификации с помощью аминокислотной вставки, делеции или замены приводили к обогащению или улучшению в анализах. В некоторых случаях результаты анализов использовали для получения тепловых карт, представленных на ФИГ. 34-36, которые дают качественные и количественные данные для модифицированного этими способами заданного аминокислотного положения. Для целей настоящего описания

последовательности доменов CasX 515 представлены в таблице 4 и включают в себя домен OBD-I, имеющий последовательность SEQ ID NO: 2342, домен OBD-II, имеющий последовательность SEQ ID NO: 2347, домен NTSB, имеющий последовательность SEQ ID NO: 2335, домен спиральный I-I, имеющий последовательность SEQ ID NO: 2343, домен спиральный I-II, имеющий последовательность SEQ ID NO: 2336, домен спиральный II, имеющий последовательность SEQ ID NO: 2351, домен RuvC-I, имеющий последовательность SEQ ID NO: 2352, домен RuvC-II, имеющий последовательность SEQ ID NO: 2350, и домен TSL, имеющий последовательность SEQ ID NO: 2349. В способах описания отдельные положения в доменах CasX 515 были подвергнуты модификации, проанализированы, и полученные положения и примеры модификаций, приводящие к последующему обогащению или улучшению, представлены в зависимости от их положения в каждом домене или поддомене. В некоторых случаях такие положения описаны в таблицах 21-24 раздела «Примеры». В некоторых вариантах осуществления в описании предложены варианты CasX, полученные из CasX 515, содержащие одну или более модификаций (т. е. вставку, делецию или замену) в одном или более аминокислотных положениях в домене NTSB относительно SEQ ID NO: 2335, выбранных из группы, состоящей из P2, S4, Q9, E15, G20, G33, L41, Y51, F55, L68, A70, E75, K88 и G90, причем модификация приводит к получению улучшенной характеристики относительно CasX 515. В конкретном варианте осуществления одна или более модификаций в одном или более аминокислотных положениях в домене NTSB относительно SEQ ID NO: 2335 выбраны из группы, состоящей из ^G2, ^I4, ^L4, Q9P, E15S, G20D, [S30], G33T, L41A, Y51T, F55V, L68D, L68E, L68K, A70Y, A70S, E75A, E75D, E75P, K88Q и G90Q (где «^» обозначает вставку, а «[]» обозначает делецию в этом положении). В некоторых вариантах осуществления в описании предложены варианты CasX, полученные из CasX 515, содержащие одну или более модификаций в одном или более аминокислотных положениях в домене спиральный I-II относительно SEQ ID NO: 2336, выбранных из группы, состоящей из I24, A25, Y29 G32, G44, S48, S51, Q54, I56, V63, S73, L74, K97, V100, M112, L116, G137, F138 и S140, причем модификация приводит к получению улучшенной характеристики относительно CasX 515. В конкретном варианте осуществления одна или более модификаций в одном или более аминокислотных положениях в домене спиральный I-II выбраны из группы, состоящей из ^T24, ^C25, Y29F, G32Y, G32N, G32H, G32S, G32T, G32A, G32V, [G32], G32S, G32T, G44L, G44H, S48H, S48T, S51T, Q54H, I56T, V63T, S73H, L74Y, K97G, K97S, K97D, K97E, V100L, M112T, M112W, M112R, M112K, L116K, G137R, G137K, G137N, ^Q138 и S140Q. В некоторых вариантах осуществления в описании предложены варианты CasX, полученные из CasX 515, содержащие одну или более модификаций в одном или более аминокислотных положениях в домене спиральный II относительно SEQ ID NO: 2351, выбранных из группы, состоящей из L2, V3, E4, R5, Q6, A7, E9, V10, D11, W12, W13, D14, M15, V16, C17, N18, V19, K20, L22, I23, E25, K26, K31, Q35, L37, A38, K41,R 42, Q43, E44, L46, K57, Y65, G68, L70, L71, L72, E75, G79, D81, W82, K84, V85, Y86, D87, I93, K95, K96, E98, L100, K102, I104, K105, E109, R110, D114, K118,

A120, L121, W124, L125, R126, A127, A129, I133, E134, G135, L136, E138, D140, K141, D142, E143, F144, C145, C147, E148, L149, K150, L151, Q152, K153, L158, E166 и A167, причем модификация приводит к получению улучшенной характеристики относительно CasX 515. В конкретном варианте осуществления одна или более модификаций в одном или более аминокислотных положениях в домене спиральный II выбраны из группы, состоящей из ^A2, ^H2, [L2]+[V3], V3E, V3Q, V3F, [V3], ^D3, V3P, E4P, [E4], E4D, E4L, E4R, R5N, Q6V, ^Q6, ^G7, ^H9, ^A9, VD10, ^T10, [V10], ^F10, ^D11, [D11], D11S, [W12], W12T, W12H, ^P12, ^Q13, ^G12, ^R13, W13P, W13D, ^D13, W13L, ^P14, ^D14, [D14]+[M15], [M15], ^T16, ^P17, N18I, V19N, V19H, K20D, L22D, I23S, E25C, E25P, ^G25, K26T, K27E, K31L, K31Y, Q35D, Q35P, ^S37, [L37]+[A38], K41L, ^R42, [Q43]+[E44], L46N, K57Q, Y65T, G68M, L70V, L71C, L72D, L72N, L72W, L72Y, E75F, E75L, E75Y, G79P, ^E79, ^T81, ^R81, ^W81, ^Y81, ^W82, ^Y82, W82G, W82R, K84D, K84H, K84P, K84T, V85L, V85A, ^L85, Y86C, D87G, D87M, D87P, I93C, K95T, K96R, E98G, L100A, K102H, I104T, I104S, I104Q, K105D, ^K109, E109L, R110D, [R110], D114E, ^D114, K118P, A120R, L121T, W124L, L125C, R126D, A127E, A127L, A129T, A129K, I133E, ^C133, ^S134, ^G134, ^R135, G135P, L136K, L136D, L136S, L136H, [E138], D140R, ^D140, ^P141, ^D142, [E143]+[F144], ^Q143, F144K, [F144], [F144]+[C145], C145R, ^G145, C145K, C147D, ^V148, E148D, ^H149, L149R, K150R, L151H, Q152C, K153P, L158S, E166L и ^F167. В некоторых вариантах осуществления в описании предложены варианты CasX, полученные из CasX 515, содержащие одну или более модификаций в одном или более аминокислотных положениях в домене RuvC-I относительно SEQ ID NO: 2352, выбранных из группы, состоящей из I4, K5, P6, M7, N8, L9, V12, G49, K63, K80, N83, R90, M125 и L146, причем модификация приводит к получению улучшенной характеристики относительно CasX 515. В конкретном варианте осуществления одна или более модификаций в одном или более аминокислотных положениях в домене RuvC-I выбраны из группы, состоящей из ^I4, ^S5, ^T6, ^N6, ^R7, ^K7, ^H8, ^S8, V12L, G49W, G49R, S51R, S51K, K62S, K62T, K62E, V65A, K80E, N83G, R90H, R90G, M125S, M125A, L137Y, ^P137, [L141], L141R, L141D, ^Q142, ^R143, ^N143, E144N, ^P146, L146F, P147A, K149Q, T150V, ^R152, ^H153, T155Q, ^H155, ^R155, ^L156, [L156], ^W156, ^A157, ^F157, A157S, Q158K, [Y159], T160Y, T160F, ^I161, S161P, T163P, ^N163, C164K и C164M. В некоторых вариантах осуществления в описании предложены варианты CasX, полученные из CasX 515, содержащие одну или более модификаций в одном или более аминокислотных положениях в домене OBD-I относительно SEQ ID NO: 2342, выбранных из группы, состоящей из I4, K5, P6, M7, N8, L9, V12, G49, K63, K80, N83, R90, M125 и L146, причем модификация приводит к получению улучшенной характеристики относительно CasX 515. В конкретном варианте осуществления одна или более модификаций в одном или более аминокислотных положениях в домене OBD-I выбраны из группы, состоящей из ^G3, I3G, I3E, ^G4, K4G, K4P, K4S, K4W, K4W, R5P, ^P5, ^G5, R5S, ^S5, R5A, R5P, R5G, R5L, I6A, I6L, ^G6, N7Q, N7L, N7S, K8G, K15F, D16W, ^F16, ^F18, ^P27, M28P, M28H, V33T, R34P, M36Y, R41P, L47P, ^P48, E52P, ^P55, [P55]+[Q56], Q56S, Q56P, ^D56, ^T56 и Q56P. В некоторых вариантах осуществления в описании

предложены варианты CasX, полученные из CasX 515, содержащие одну или более модификаций в одном или более аминокислотных положениях в домене OBD-II относительно SEQ ID NO: 2347, выбранных из группы, состоящей из I4, K5, P6, M7, N8, L9, V12, G49, K63, K80, N83, R90, M125 и L146, причем модификация приводит к получению улучшенной характеристики относительно CasX 515. В конкретном варианте осуществления одна или более модификаций в одном или более аминокислотных положениях в домене OBD-I выбраны из группы, состоящей из [S2], I3R, I3K, [I3]+[L4], [L4], K11T, ^P24, K37G, R42E, ^S53, ^R58, [K63], M70T, I82T, Q92I, Q92F, Q92V, Q92A, ^A93, K110Q, R115Q, L121T, ^A124, ^R141, ^D143, ^A143, ^W144 и ^A145. В некоторых вариантах осуществления в описании предложены варианты CasX, полученные из CasX 515, содержащие одну или более модификаций в одном или более аминокислотных положениях в домене TSL относительно SEQ ID NO: 2349, выбранных из группы, состоящей из S1, N2, C3, G4, F5, I7, K18, V58, S67, T76, G78, S80, G81, E82, S85, V96 и E98, причем модификация приводит к получению улучшенной характеристики относительно CasX 515. В конкретном варианте осуществления одна или более модификаций в одном или более аминокислотных положениях в домене OBD-I выбраны из группы, состоящей из ^M1, [N2], ^V2, C3S, ^G4, ^W4, F5P, ^W7, K18G, V58D, ^A67, T76E, T76D, T76N, G78D, [S80], [G81], ^E82, ^N82, S85I, V96C, V96T и E98D. Следует понимать, что комбинации любых из указанных в параграфе модификаций можно аналогичным образом вводить в варианты CasX описания, получая вариант CasX с улучшенными характеристиками. Например, в одном варианте осуществления в описании предложен вариант 535 CasX (SEQ ID NO: 435), который имеет одну мутацию G223S относительно CasX 515. В другом варианте осуществления в описании предложен вариант 668 CasX (SEQ ID NO: 567), который имеет вставку R в положении 26 и замену G223S относительно CasX 515. В другом варианте осуществления в описании предложен вариант 672 CasX (SEQ ID NO:570), который имеет замены L169K и G223S относительно CasX 515. В другом варианте осуществления в описании предложен CasX 676 (SEQ ID NO: 574), который имеет замены L169K и G223S и вставку R в положении 26 относительно CasX 515. Варианты CasX с улучшенными характеристиками относительно CasX 515 включают варианты из таблицы 3.

[0231] Примеры характеристик, которые могут быть улучшены в вариантах белков CasX относительно таких же характеристик в референсных белках CasX или относительно варианта CasX, из которого они были получены, включают в себя, без ограничений, улучшенное свертывание варианта, повышенную аффинность связывания с гРНК, повышенную аффинность связывания с целевой нуклеиновой кислотой, улучшенную способность использовать более широкий спектр PAM-последовательностей при редактировании и/или связывании целевой нуклеиновой кислоты, улучшенное раскручивание целевой ДНК, повышенную активность редактирования, улучшенную эффективность редактирования, улучшенную специфичность редактирования, улучшенную специфичность редактирования целевой нуклеиновой кислоты, сниженное нецелевое редактирование или расщепление, повышенную долю эукариотического генома,

который можно эффективно отредактировать, повышенную активность нуклеазы, повышенную загрузку целевой цепи для двухцепочечного расщепления, сниженную загрузку целевой цепи для одноцепочечного никирования, повышенное связывание с нецелевой цепью ДНК, улучшенную стабильность белка, улучшенную стабильность комплекса белок : гРНК (RNP) и улучшенные характеристики слияния. В конкретном варианте осуществления, как описано в разделе «Примеры», такие улучшенные характеристики могут включать в себя, без ограничений, улучшенную активность расщепления в целевых нуклеиновых кислотах, имеющих PAM-последовательности TTC, ATC и CTC, повышенную специфичность расщепления последовательности целевой нуклеиновой кислоты и сниженное нецелевое расщепление целевой нуклеиновой кислоты.

Таблица 6. Последовательности доменов CasX 515

Домен	SEQ ID NO	Аминокислотная последовательность
ODB-I	2342	QEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDL RERLENLRKKPENIPQ
Спиральный I-I	2343	PISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGL MSRVA
NTSB	2335	QPASKKIDQNKLKPEMDEKGNLTTAGFACSQCGQPLFVY KLEQVSEKGKAYTNYFGRCNVAEHEKLILLAQLKPEKDS DEAVTYSLGKFGQ
Спиральный I-II	2336	RALDFYSIHVTKESTHPVKPLAQIAGNRYASGPVGKALSD ACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESLRELAGK ENLEYPSVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLW QKLKLSRDDAKPLLRLKGFPSF
Спиральный II	2351	PLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNL AGYKRQEALRPYLSSEEDRKKGKKFARYQLGDLLLHLEK KHGEDWGKVYDEAWERIDKKVEGLSKHIKLEEERRSEDA QSKAALTDWLRAKASFVIEGLKEADKDEFCRCELKLQKW YGDLRGKPFAIEAE
OBD-II	2347	NSILDISGFSKQYNCAFIWQKDGVKKLNLYLIINYFKGGKL RFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYN RRTRQDEPALFVALTFERREVLD
RuvC-I	2352	SSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNP THILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAK NLADDMVRNTARDLLYYAVTQDAMLIFENLSRGFGRQG KRTFMAERQYTRMEDWLTAKLAYEGLPSKTYLSKTLAQ YTSKTC
TSL	2349	SNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKG RSGEALSLLKKRFSHRPVQEKFVCLNCGFETH

RuvC-II	2350	ADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDKRAFV ETWQSFYRKKLKEVWKPAV

[0232] Варианты CasX по вариантам осуществления, описанным в настоящем документе, обладают способностью образовывать комплекс RNP с гРНК, описанной в настоящем документе. В некоторых вариантах осуществления RNP, содержащий вариант белка CasX и гРНК описания, в концентрации 20 пМ или менее способен расщеплять двухцепочечную ДНК-мишень с эффективностью по меньшей мере 80%. В некоторых вариантах осуществления RNP в концентрации 20 пМ или менее способен расщеплять двухцепочечную ДНК-мишень с эффективностью по меньшей мере 40%, по меньшей мере 50%, по меньшей мере 60%, по меньшей мере 70%, по меньшей мере 80%, по меньшей мере 85%, по меньшей мере 90% или по меньшей мере 95%. В некоторых вариантах осуществления RNP в концентрации 50 пМ или менее, 40 пМ или менее, 30 пМ или менее, 20 пМ или менее, 10 пМ или менее или 5 пМ или менее способен расщеплять двухцепочечную ДНК-мишень с эффективностью по меньшей мере 40%, по меньшей мере 50%, по меньшей мере 60%, по меньшей мере 70%, по меньшей мере 80%, по меньшей мере 85%, по меньшей мере 90% или по меньшей мере 95%. Эти улучшенные характеристики более подробно описаны ниже.

g. Стабильность белка

[0233] В некоторых вариантах осуществления в описании предложен вариант белка CasX с улучшенной стабильностью относительно референсного белка CasX. В некоторых вариантах осуществления улучшенная стабильность варианта белка CasX приводит к экспрессии более устойчивого белка, что повышает эффективность редактирования. В некоторых вариантах осуществления улучшенная стабильность варианта белка CasX приводит к более высокой доле белка CasX, который остается свернутым в функциональную конформацию, и повышает эффективность редактирования или улучшает пригодность для очистки для целей производства. В настоящем документе термин «функциональная конформация» относится к белку CasX, который находится в конформации, в которой белок способен связываться с гРНК и целевой нуклеиновой кислотой. В вариантах осуществления, в которых вариант CasX не несет одной или более мутаций, делающих его каталитически неактивным, вариант CasX способен расщеплять, никировать или иным образом модифицировать целевую нуклеиновую кислоту при нахождении в комплексе с гРНК с нацеливающей последовательностью, способной гибридизоваться с целевой нуклеиновой кислотой. Функциональной конформацией CasX называется «компетентная в отношении расщепления» конформация. В некоторых примерах осуществления, включая те варианты осуществления, в которых вариант белка CasX имеет более высокую долю белка CasX, остающегося свернутым в функциональную конформацию, для использования в редактировании генов требуется более низкая концентрация варианта CasX, например, по сравнению с референсным белком CasX. Таким образом, в некоторых вариантах осуществления вариант CasX с улучшенной стабильностью обладает улучшенной эффективностью по сравнению с референсным CasX

в одном или более контекстах редактирования генов.

[0234] В некоторых вариантах осуществления в описании предложен вариант белка CasX, имеющий улучшенную стабильность RNP-комплекса вариант белка CasX : гРНК относительно комплекса референсный белок CasX : гРНК, так что RNP остается в функциональной форме. Улучшения стабильности могут включать в себя повышенную термостабильность, устойчивость к протеолитической деградации, улучшенные фармакокинетические свойства, стабильность в диапазоне значений pH, солевых условий и тоничности. В некоторых вариантах осуществления улучшенная стабильность комплекса может приводить к улучшенной эффективности редактирования. В некоторых вариантах осуществления RNP из варианта CasX и варианта гРНК демонстрирует по меньшей мере 2-кратную, по меньшей мере 3-кратную или по меньшей мере 4-кратную процентную долю компетентного в отношении расщепления RNP по сравнению с RNP из референсного белка CasX с SEQ ID NO: 1-3 и гРНК с SEQ ID NO: 4 или 5 из таблицы 1. Примеры данных о повышенной доле компетентного в отношении расщепления RNP представлены в разделе «Примеры».

[0235] В некоторых вариантах осуществления улучшенная стабильность варианта белка CasX включает в себя улучшенную кинетику сворачивания варианта белка CasX относительно референсного белка CasX. В некоторых вариантах осуществления кинетика сворачивания варианта белка CasX имеет кратность улучшения относительно референсного белка CasX по меньшей мере примерно 5, по меньшей мере примерно 10, по меньшей мере примерно 50, по меньшей мере примерно 100, по меньшей мере примерно 500, по меньшей мере примерно 1000, по меньшей мере примерно 2000, по меньшей мере примерно 3000, по меньшей мере примерно 4000, по меньшей мере примерно 5000 или по меньшей мере примерно 10 000 раз. В некоторых вариантах осуществления кинетика сворачивания варианта белка CasX улучшена по сравнению с референсным белком CasX на по меньшей мере примерно 1 кДж/моль, по меньшей мере примерно 5 кДж/моль, по меньшей мере примерно 10 кДж/моль, по меньшей мере примерно 20 кДж/моль, по меньшей мере примерно 30 кДж/моль, по меньшей мере примерно 40 кДж/моль, по меньшей мере примерно 50 кДж/моль, по меньшей мере примерно 60 кДж/моль, по меньшей мере примерно 70 кДж/моль, по меньшей мере примерно 80 кДж/моль, по меньшей мере примерно 90 кДж/моль, по меньшей мере примерно 100 кДж/моль, по меньшей мере примерно 150 кДж/моль, по меньшей мере примерно 200 кДж/моль, по меньшей мере примерно 250 кДж/моль, по меньшей мере примерно 300 кДж/моль, по меньшей мере примерно 350 кДж/моль, по меньшей мере примерно 400 кДж/моль, по меньшей мере примерно 450 кДж/моль или по меньшей мере примерно 500 кДж/моль.

[0236] Примеры аминокислотных изменений, которые могут повышать стабильность варианта белка CasX относительно референсного белка CasX, могут включать в себя, без ограничений, изменения аминокислот, которые увеличивают число водородных связей в варианте белка CasX, увеличивают число дисульфидных мостиков в варианте белка CasX, увеличивают число солевых мостиков в варианте белка CasX, усиливают

взаимодействия между частями варианта белка CasX, увеличивают заглубленный гидрофобный участок CasX или любые их комбинации.

h. Аффинность белка к гРНК

[0237] В некоторых вариантах осуществления вариант белка CasX имеет улучшенную аффинность к гРНК относительно референсного белка CasX или другого варианта CasX, из которого он получен, что приводит к образованию рибонуклеопротеинового комплекса (RNP). Повышенная аффинность варианта белка CasX к гРНК может, например, приводить к более низкой K_d образования комплекса RNP, что в некоторых случаях может приводить к образованию более стабильного RNP-комплекса. В некоторых вариантах осуществления повышенная аффинность варианта белка CasX к гРНК приводит к получению повышенной стабильности RNP-комплекса при доставке в человеческие клетки. Эта повышенная стабильность может влиять на функциональность и пригодность комплекса в клетках субъекта, а также приводить к улучшенным фармакокинетическим свойствам в крови при доставке субъекту. В некоторых вариантах осуществления повышенная аффинность варианта белка CasX и полученная повышенная стабильность RNP-комплекса позволяют доставлять более низкую дозу варианта белка CasX субъекту или в клетки с сохранением требуемой активности, например редактирования гена in vivo или *in vitro*.

[0238] В некоторых вариантах осуществления более высокая аффинность (более плотное связывание) варианта белка CasX с гРНК обеспечивает большее число событий редактирования, когда и вариант белка CasX, и гРНК остаются в комплексе RNP. Повышенное число событий редактирования можно оценивать с применением анализов редактирования, описанных в настоящем документе.

[0239] В некоторых вариантах осуществления K_d связи варианта белка CasX с гРНК повышен относительно референсного белка CasX или другого варианта CasX, из которого он получен, с коэффициентом по меньшей мере примерно 1,1, по меньшей мере примерно 1,2, по меньшей мере примерно 1,3, по меньшей мере примерно 1,4, по меньшей мере примерно 1,5, по меньшей мере примерно 1,6, по меньшей мере примерно 1,7, по меньшей мере примерно 1,8, по меньшей мере примерно 1,9, по меньшей мере примерно 2, по меньшей мере примерно 3, по меньшей мере примерно 4, по меньшей мере примерно 5, по меньшей мере примерно 6, по меньшей мере примерно 7, по меньшей мере примерно 8, по меньшей мере примерно 9, по меньшей мере примерно 10, по меньшей мере примерно 15, по меньшей мере примерно 20, по меньшей мере примерно 25, по меньшей мере примерно 30, по меньшей мере примерно 35, по меньшей мере примерно 40, по меньшей мере примерно 45, по меньшей мере примерно 50, по меньшей мере примерно 60, по меньшей мере примерно 70, по меньшей мере примерно 80, по меньшей мере примерно 90 или по меньшей мере примерно 100. В некоторых вариантах осуществления вариант CasX имеет кратность повышения аффинности связывания с гРНК от примерно 1,1 до примерно 10 раз по сравнению с референсным белком CasX с SEQ ID NO: 2.

[0240] Без ограничений, накладываемых какой-либо теорией, в некоторых вариантах

осуществления аминокислотные изменения в домене спиральный I могут повышать аффинность связывания варианта белка CasX с нацеливающей последовательностью гРНК, тогда как изменения в домене спиральный II могут повышать аффинность связывания варианта белка CasX с каркасной стеблевой петлей гРНК, а изменения в олигонуклеотид-связывающем домене (OBD) повышают аффинность связывания варианта белка CasX с триплексом гРНК.

[0241] Способы измерения аффинности связывания белка CasX с гРНК включают способы *in vitro* с применением очищенного белка CasX и гРНК. Аффинность связывания для референсного CasX и вариантов белков может быть измерена посредством флуоресцентной поляризации, если гРНК или белок CasX помечен флуорофором. Альтернативно или в дополнение к этому аффинность связывания может быть измерена посредством интерферометрии биослоев, анализа методом сдвига электрофоретической подвижности (EMSA) или связывания на фильтрах. Дополнительные стандартные методики количественного определения абсолютных аффинностей связывающих РНК белков, таких как референсный CasX и варианты белков описания для конкретных гРНК, таких как референсные гРНК и их варианты, включают, без ограничений, изотермическую калориметрию (ITC) и поверхностный плазмонный резонанс (SPR), а также способы из раздела «Примеры».

i. Аффинность к целевой нуклеиновой кислоте

[0242] В некоторых вариантах осуществления вариант белка CasX имеет повышенную аффинность связывания с целевой нуклеиновой кислотой относительно аффинности референсного белка CasX с последовательностью целевой нуклеиновой кислоты или другого варианта CasX, из которого он получен. Варианты CasX с более высокой аффинностью к своей целевой нуклеиновой кислоте могут в некоторых вариантах осуществления расщеплять последовательность целевой нуклеиновой кислоты быстрее, чем референсный белок CasX, который не имеет повышенной аффинности к целевой нуклеиновой кислоте.

[0243] В некоторых вариантах осуществления улучшенная аффинность к целевой нуклеиновой кислоте включает в себя улучшенную аффинность к целевой последовательности или последовательности протоспейсера целевой нуклеиновой кислоты, улучшенную аффинность к PAM-последовательности, улучшенную способность поиска целевой последовательности в ДНК или любые их комбинации. Без ограничений, накладываемых какой-либо теорией, считается, что системы CRISPR/Cas, такие как CasX, могут находить свои целевые последовательности посредством одномерной диффузии вдоль молекулы ДНК. Считается, что процесс включает в себя (1) связывание рибонуклеопротеина с молекулой ДНК с последующей (2) остановкой на целевой последовательности, и на любую из стадий может в некоторых вариантах осуществления повлиять улучшенная аффинность белков CasX к последовательности целевой нуклеиновой кислоты, в результате чего улучшается функция варианта белка CasX по сравнению с референсным белком CasX.

[0244] В некоторых вариантах осуществления вариант белка CasX с улучшенной аффинностью к целевой нуклеиновой кислоте имеет повышенную аффинность или способность использовать специфические PAM-последовательности, отличные от канонической PAM TTC, распознаваемой референсным белком CasX SEQ ID NO: 2, включая PAM-последовательности, выбранные из группы, состоящей из TTC, ATC, GTC и CTC, в результате чего увеличивается объем целевой нуклеиновой кислоты, которая может быть отредактирована, по сравнению с нуклеазами CasX дикого типа или нуклеазами CasX 199 или 491. Без ограничений, накладываемых какой-либо теорией, существует возможность того, что эти варианты белка могут в целом сильнее взаимодействовать с ДНК и могут иметь улучшенную способность к осуществлению доступа и редактированию последовательностей в целевой нуклеиновой кислоте из-за способности использовать дополнительные PAM-последовательности помимо тех, которые используют референсный CasX дикого типа или нуклеазы CasX 199 или 491, обеспечивая таким образом более эффективный процесс поиска целевой последовательности для белка CasX. Более высокая общая аффинность к ДНК также в некоторых вариантах осуществления может увеличивать частоту, с которой белок CasX может эффективно начать и завершить стадию связывания и раскручивания, облегчая таким образом внедрение в целевую цепь, образование R-петли и, в конечном итоге, расщепление последовательности целевой нуклеиновой кислоты.

[0245] Без ограничений, накладываемых какой-либо теорией, существует возможность того, что аминокислотные изменения в домене NTSB, которые повышают эффективность раскручивания или захвата нецелевой цепи нуклеиновой кислоты в нераскрученном состоянии, могут увеличивать аффинность вариантов белков CasX к целевой нуклеиновой кислоте. Альтернативно или в дополнение аминокислотные изменения в домене NTSB, которые увеличивают способность домена NTSB стабилизировать ДНК во время раскручивания, могут увеличивать аффинность вариантов белков CasX к целевой нуклеиновой кислоте. Альтернативно или в дополнение аминокислотные изменения в OBD могут увеличивать аффинность связывания варианта белка CasX со смежным с протоспейсером мотивом (PAM), увеличивая таким образом аффинность варианта белка CasX к целевой нуклеиновой кислоте. Альтернативно или в дополнение аминокислотные изменения в доменах спиральный I и/или II, RuvC и TSL, которые увеличивают аффинность варианта белка CasX к целевой цепи нуклеиновой кислоты, могут увеличивать аффинность варианта белка CasX к целевой нуклеиновой кислоте.

[0246] В некоторых вариантах осуществления аффинность связывания варианта белка CasX описания к молекуле целевой нуклеиновой кислоты повышена относительно референсного белка CasX или другого варианта CasX, от которого он получен, с коэффициентом по меньшей мере примерно 1,1, по меньшей мере примерно 1,2, по меньшей мере примерно 1,3, по меньшей мере примерно 1,4, по меньшей мере примерно 1,5, по меньшей мере примерно 1,6, по меньшей мере примерно 1,7, по меньшей мере примерно 1,8, по меньшей мере примерно 1,9, по меньшей мере примерно 2, по меньшей

мере примерно 3, по меньшей мере примерно 4, по меньшей мере примерно 5, по меньшей мере примерно 6, по меньшей мере примерно 7, по меньшей мере примерно 8, по меньшей мере примерно 9, по меньшей мере примерно 10, по меньшей мере примерно 15, по меньшей мере примерно 20, по меньшей мере примерно 25, по меньшей мере примерно 30, по меньшей мере примерно 35, по меньшей мере примерно 40, по меньшей мере примерно 45, по меньшей мере примерно 50, по меньшей мере примерно 60, по меньшей мере примерно 70, по меньшей мере примерно 80, по меньшей мере примерно 90 или по меньшей мере примерно 100. В некоторых вариантах осуществления вариант белка CasX имеет кратность повышения аффинности связывания с целевой нуклеиновой кислотой от примерно 1,1 до примерно 100 раз по сравнению с референсным белком с SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3 или с вариантами CasX 491 и 515.

[0247] В некоторых вариантах осуществления вариант белка CasX имеет повышенную аффинность связывания с нецелевой цепью целевой нуклеиновой кислоты. В настоящем документе термин «нецелевая цепь» относится к цепи ДНК последовательности целевой нуклеиновой кислоты, которая не образует пары оснований Уотсона - Крика с нацеливающей последовательностью в гРНК и является комплементарной целевой цепи нуклеиновой кислоты. В некоторых вариантах осуществления вариант белка CasX имеет кратность повышения аффинности связывания с нецелевой цепью целевой нуклеиновой кислоты от примерно 1,1 до примерно 100 раз по сравнению с референсным белком с SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3 или с вариантами CasX с SEQ ID NO: 270 или SEQ ID NO: 336.

[0248] Способы измерения аффинности белка CasX (например, референсного или варианта) к молекуле целевой и/или нецелевой нуклеиновой кислоты могут включать в себя анализы методом сдвига электрофоретической подвижности (EMSA), связывание на фильтрах, изотермическую калориметрию (ITC), а также поверхностный плазмонный резонанс (SPR), флуоресцентную поляризацию и интерферометрию биослоев (BLI). Дополнительные способы измерения аффинности белка CasX к мишени включают в себя биохимические анализы *in vitro*, которые измеряют число событий расщепления ДНК с течением времени.

j. Улучшенная специфичность к целевому сайту

[0249] В некоторых вариантах осуществления вариант белка CasX имеет улучшенную специфичность к последовательности целевой нуклеиновой кислоты относительно референсного белка CasX или другого варианта CasX, из которого он получен. В настоящем документе термин «специфичность» иногда называется «специфичностью к мишени» и означает уровень расщепления рибонуклеопротеиновым комплексом CRISPR/Cas нецелевых последовательностей, которые являются аналогичными, но не идентичными последовательности целевой нуклеиновой кислоты; например, RNP с вариантом CasX с более высокой степенью специфичности будет демонстрировать меньшее нецелевое расщепление последовательностей относительно референсного белка CasX. Специфичность и уменьшение потенциально вредных

нецелевых эффектов системы CRISPR/Cas могут быть очень важными для достижения приемлемого терапевтического индекса при применении у субъектов-млекопитающих.

[0250] В некоторых вариантах осуществления вариант белка CasX имеет улучшенную специфичность к целевому сайту в целевой последовательности, который комплементарен нацеливающей последовательности гРНК. Как описано *выше*, коррелятом улучшенной специфичности является уменьшение нецелевого редактирования. В некоторых вариантах осуществления вариант белка CasX демонстрирует сниженное нецелевое редактирование или расщепление целевого сайта в целевой последовательности, которая не является на 100% комплементарной нацеливающей последовательности гРНК в комплексе с вариантом CasX в виде RNP. Без ограничений, накладываемых какой-либо теорией, существует возможность того, что аминокислотные изменения в спиральных доменах I и II, которые увеличивают специфичность варианта белка CasX к цепи целевой нуклеиновой кислоты, могут повышать специфичность варианта белка CasX к целевой нуклеиновой кислоте в целом. В некоторых вариантах осуществления аминокислотные изменения, которые повышают специфичность вариантов белков CasX к целевой нуклеиновой кислоте, также могут приводить к получению сниженной аффинности вариантов белков CasX к ДНК.

[0251] Способы тестирования специфичности белка CasX (такого как вариант или референсный белок) к мишени могут включать направление и циркуляризацию для регистрации *in vitro* эффектов расщепления посредством секвенирования (CIRCLE-seq) или аналогичные способы. Вкратце, в методиках CIRCLE-seq геномную ДНК расщепляют и циркуляризируют посредством лигирования адаптеров петля-стебель, которые никируют в областях петля-стебель с открытием 4-нуклеотидных палиндромных липких концов. После этого осуществляется внутримолекулярное лигирование и разложение оставшейся линейной ДНК. Затем циклические молекулы ДНК, содержащие сайт расщепления CasX, линеаризуют с помощью CasX, и адаптеры адаптера лигируют к открытым концам, после чего проводят высокопроизводительное секвенирование для создания прочтений спаренных концов, которые содержат информацию о нецелевом сайте. Дополнительные анализы, которые можно применять для обнаружения нецелевых событий, и, следовательно, специфичности белка CasX, включают анализы, применяемые для обнаружения и количественного определения инделов (вставок и делеций), образованных в этих выбранных нецелевых сайтах, например, анализы нуклеазами обнаружения несовпадений и секвенирование нового поколения (NGS). Примеры анализов обнаружения несовпадений включают нуклеазные анализы, в которых геномную ДНК из клеток, обработанных CasX и огРНК, подвергают ПЦР амплификации, денатурированию и регибридизируют с образованием гетеродуплексной ДНК, содержащей одну цепь дикого типа и одну цепь с инделом. Несовпадения распознаются и расщепляются нуклеазами обнаружения несовпадений, такими как нуклеаза Surveyor или T7 эндонуклеаза I. Способы оценки специфичности вариантов CasX вместе с вспомогательными данными, демонстрирующими улучшенную специфичность вариантов осуществления вариантов

CasX, описаны в разделе «Примеры».

k. Протоспейсер и PAM-последовательности

[0252] В настоящем документе протоспейсер определяется как ДНК-последовательность, комплементарная нацеливающей последовательности гидовой РНК, и ДНК, комплементарная этой последовательности, называемые целевой цепью и нецелевой цепью соответственно. В настоящем документе PAM представляет собой нуклеотидную последовательность, проксимальную к протоспейсеру, которая в сочетании с нацеливающей последовательностью гРНК помогает ориентировать и позиционировать CasX для потенциального расщепления цепи (-ей) протоспейсера.

[0253] PAM-последовательности могут быть вырожденными, и конкретные конструкты RNP могут иметь разные предпочтительные и допустимые PAM-последовательности, которые поддерживают разные эффективности расщепления. В соответствии с соглашением, если не указано иное, описание относится как к PAM, так и к последовательности протоспейсера и их направленности в соответствии с ориентацией нецелевой цепи. Это не подразумевает, что PAM-последовательность нецелевой цепи, а не целевая цепь является определяющей для расщепления или механистически вовлечена в распознавание мишени. Например, если речь идет о PAM TTC, то она фактически может быть комплементарна последовательности GAA, которая необходима для расщепления мишени, или может представлять собой некоторую комбинацию нуклеотидов из обеих цепей. В случае белков CasX, описанных в настоящем документе, PAM находится с 5'-стороны от протоспейсера, и один нуклеотид отделяет PAM от первого нуклеотида протоспейсера. Таким образом, в случае референсного CasX TTC PAM следует понимать как последовательность, соответствующую формуле 5'-…NN**TTC**N(протоспейсер)NNNNNN…3' (SEQ ID NO: 19), где N представляет собой любой нуклеотид ДНК, а «(протоспейсер)» представляет собой ДНК-последовательность, идентичную нацеливающей последовательности гидовой РНК. В случае варианта CasX с расширенным распознаванием PAM следует понимать, что PAM TTC, CTC, GTC или ATC означают последовательности, соответствующие формулам: 5'-…NNTTCN(протоспейсер)NNNNNN…3' (SEQ ID NO: 19); 5'-…NNCTCN(протоспейсер)NNNNNN…3' (SEQ ID NO: 20); 5'-…NNGTCN(протоспейсер)NNNNNN…3' (SEQ ID NO: 21); или 5'-…NNATCN(протоспейсер)NNNNNN…3' (SEQ ID NO: 22). Альтернативно TTC PAM следует понимать как последовательность, соответствующую формуле 5'-…NNNTCN(протоспейсер)NNNNNN…3' (SEQ ID NO: 23).

[0254] Кроме того, варианты белков CasX описания обладают улучшенной способностью эффективно редактировать и/или связывать целевую нуклеиновую кислоту при нахождении в комплексе с гРНК в виде RNP с использованием PAM-мотива TC, включая PAM-последовательности, выбранные из TTC, ATC, GTC или CTC (в ориентации от 5' к 3'), по сравнению с RNP из референсного белка CasX и референсной гРНК, или с RNP из другого варианта CasX, из которого он получен, например, CasX 491, и гРНК 174.

В вышеизложенном PAM-последовательность расположена на расстоянии по меньшей мере 1 нуклеотида с 5'-стороны от нецелевой цепи протоспейсера, имеющей идентичность с нацеливающей последовательностью гРНК, в системе анализа по сравнению с эффективностью редактирования и/или связывания RNP, содержащего референсный белок CasX и референсную гРНК, в сопоставимой системе анализа. В одном варианте осуществления RNP из варианта CasX и варианта гРНК демонстрирует более высокую эффективность редактирования и/или связывания целевой последовательности в целевой нуклеиновой кислоте по сравнению с RNP, содержащим референсный белок CasX и референсную гРНК (или с RNP из другого варианта CasX, из которого он получен, например, CasX 491, и гРНК 174), в сопоставимой системе анализа, причем PAM-последовательность целевой ДНК представляет собой TTC. В другом варианте осуществления RNP из варианта CasX и варианта гРНК демонстрирует более высокую эффективность редактирования и/или связывания целевой последовательности в целевой нуклеиновой кислоте по сравнению с RNP, содержащим референсный белок CasX и референсную гРНК (или с RNP из другого варианта CasX, из которого он получен, например, CasX 491, и гРНК 174), в сопоставимой системе анализа, причем PAM-последовательность целевой ДНК представляет собой ATC. В конкретном варианте осуществления вышеизложенного, в котором вариант CasX демонстрирует улучшенное редактирование с PAM ATC, вариант CasX представляет собой 528 (SEQ ID NO: 428). В другом варианте осуществления RNP из варианта CasX и варианта гРНК демонстрирует более высокую эффективность редактирования и/или связывания целевой последовательности в целевой нуклеиновой кислоте по сравнению с RNP, содержащим референсный белок CasX и референсную гРНК (или с RNP из другого варианта CasX, от которого он получен, например, CasX 491, и гРНК 174), в сопоставимой системе анализа, причем PAM-последовательность целевой ДНК представляет собой CTC. В другом варианте осуществления RNP из варианта CasX и варианта гРНК демонстрирует более высокую эффективность редактирования и/или связывания целевой последовательности в целевой нуклеиновой кислоте по сравнению с RNP, содержащим референсный белок CasX и референсную гРНК (или с RNP из другого варианта CasX, из которого он получен, и гРНК 174), в сопоставимой системе анализа, причем PAM-последовательность целевой ДНК представляет собой GTC. В вышеизложенных вариантах осуществления повышенная эффективность редактирования и/или аффинность связывания с одной или более PAM-последовательностями по меньшей мере в 1,5 раза, по меньшей мере в 2 раза, по меньшей мере в 4 раза, по меньшей мере в 10 раз, по меньшей мере в 20 раз, по меньшей мере в 30 раз или по меньшей мере в 40 раз или более раз больше, чем эффективность редактирования и/или аффинность связывания для RNP с любым из белков CasX с SEQ ID NO: 1-3 и гРНК из таблицы 1 для PAM-последовательностей. Примеры анализов, демонстрирующие улучшенное редактирование, описаны в настоящем документе в разделе «Примеры» (см., например, ФИГ. 41). В некоторых вариантах осуществления белок CasX может связывать и/или модифицировать (например, расщеплять, никировать, метилировать, деметилировать

и т. д.) целевую нуклеиновую кислоту и/или полипептид, ассоциированный с целевой нуклеиновой кислотой (например, метилировать или ацетилировать гистонный хвост). В некоторых вариантах осуществления белок CasX является каталитически неактивным (dCasX), но сохраняет способность связывать целевую нуклеиновую кислоту.

l. Раскручивание ДНК

[0255] В некоторых вариантах осуществления вариант белка CasX имеет улучшенную способность к раскручиванию ДНК относительно референсного белка CasX. Ранее было показано, что плохое раскручивание дцДНК нарушает или препятствует способности белков системы CRISPR/Cas AnaCas9 или Cas14s расщеплять ДНК. Таким образом, без ограничений, накладываемых какой-либо теорией, вероятно, что повышенная активность по расщеплению ДНК у некоторых вариантов белков CasX описания по меньшей мере частично обусловлена улучшенной способностью находить и раскручивать дцДНК в целевом сайте. Способы измерения способности белков CasX (таких как вариант или референсный белок) к раскручиванию ДНК включают, без ограничений, анализы *in vitro*, которые оценивают повышение скоростей для дцДНК-мишеней при флуоресцентной поляризации или интерферометрии биослоев.

[0256] Без ограничений, накладываемых какой-либо теорией, считается, что изменения аминокислот в домене NTSB могут давать варианты белков CasX с повышенными характеристиками раскручивания ДНК. Альтернативно или в дополнение аминокислотные изменения в OBD или областях спирального домена, которые взаимодействуют с PAM, могут также давать варианты белков CasX с повышенными характеристиками раскручивания ДНК.

[0257] Способы измерения способности белков CasX (таких как вариант или референсный белок) к раскручиванию ДНК включают, без ограничений, анализы *in vitro*, которые оценивают повышение скоростей для дцДНК-мишеней при флуоресцентной поляризации или интерферометрии биослоев.

m. Каталитическая активность

[0258] Рибонуклеопротеиновый комплекс описанных в настоящем документе систем CasX : гРНК содержит вариант CasX в комплексе с вариантом гРНК, который связывается с целевой нуклеиновой кислотой и в некоторых случаях расщепляет целевую нуклеиновую кислоту. В некоторых вариантах осуществления вариант белка CasX имеет улучшенную каталитическую активность относительно референсного белка CasX или другого варианта CasX, из которого он получен. Без ограничений, накладываемых какой-либо теорией, считается, что в некоторых случаях расщепление целевой цепи может быть ограничивающим фактором для Cas12-подобных молекул при создании разрыва в дцДНК. В некоторых вариантах осуществления варианты белков CasX улучшают изгибание целевой цепи ДНК и расщепление этой цепи, что приводит к улучшению общей эффективности расщепления дцДНК рибонуклеопротеиновым комплексом CasX.

[0259] В некоторых вариантах осуществления вариант белка CasX имеет повышенную нуклеазную активность относительно референсного белка CasX или другого

варианта CasX, из которого он получен. Варианты с повышенной нуклеазной активностью могут быть созданы, например, посредством аминокислотных изменений в нуклеазном домене RuvC. В некоторых вариантах осуществления вариант CasX содержит нуклеазный домен RuvC, обладающий никазной активностью. В вышеизложенном никаза CasX системы CasX : гРНК создает одноцепочечный разрыв в пределах 10-18 нуклеотидов в сторону 3' от сайта PAM в нецелевой цепи. В других вариантах осуществления вариант CasX содержит нуклеазный домен RuvC, обладающий активностью по расщеплению двух цепей. В вышеизложенном CasX системы CasX : гРНК создает двухцепочечный разрыв в пределах 18-26 нуклеотидов в сторону 5' от сайта PAM на целевой цепи и 10-18 нуклеотидов в сторону 3' на нецелевой цепи. Нуклеазную активность можно анализировать различными способами, включая приведенные в разделе «Примеры». В некоторых вариантах осуществления вариант CasX имеет константу $k_{расщепл.}$, которая в по меньшей мере 2 раза, или в по меньшей мере 3 раза, или в по меньшей мере 4 раза, или в по меньшей мере 5 раз, или в по меньшей мере 6 раз, или в по меньшей мере 7 раз, или в по меньшей мере 8 раз, или в по меньшей мере 9 раз, или в по меньшей мере 10 раз выше по сравнению с референсным CasX.

[0260] В некоторых вариантах осуществления вариант белка CasX имеет улучшенную характеристику образования RNP с гРНК, которая приводит к более высокой процентной доли компетентных в отношении расщепления RNP по сравнению с RNP референсного белка CasX SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3 и гРНК, описанной в примерах. Под фразой «компетентный в отношении расщепления» подразумевается, что образованный RNP обладает способностью расщеплять целевую нуклеиновую кислоту. В некоторых вариантах осуществления RNP из варианта CasX и гРНК демонстрирует по меньшей мере 2-кратную, или по меньшей мере 3-кратную, или по меньшей мере 4-кратную, или по меньшей мере 5-кратную, или по меньшей мере 10-кратную скорость расщепления по сравнению с RNP из референсного белка CasX с SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3 и гРНК из таблицы 2. В вышеуказанном варианте осуществления улучшенный уровень компетентности может быть продемонстрирован в анализе *in vitro*, таком как описанный в разделе «Примеры».

[0261] В некоторых вариантах осуществления вариант белка CasX имеет повышенную загрузку целевой цепи при двухцепочечном расщеплении по сравнению с референсным CasX. Варианты с повышенной активностью загрузки целевой цепи могут быть созданы, например, посредством аминокислотных изменений в домене TLS. Без ограничений, накладываемых какой-либо теорией, аминокислотные изменения в домене TSL могут приводить к получению вариантов CasX с улучшенной каталитической активностью. Альтернативно или в дополнение аминокислотные изменения вокруг канала связывания для дуплекса РНК : ДНК также могут улучшать каталитическую активность варианта белка CasX.

[0262] В некоторых вариантах осуществления вариант белка CasX имеет повышенную побочную активность расщепления по сравнению с референсным белком

CasX. В настоящем документе «побочная активность расщепления» относится к дополнительному нецелевому расщеплению нуклеиновых кислот после распознавания и расщепления последовательности целевой нуклеиновой кислоты. В некоторых вариантах осуществления вариант белка CasX имеет сниженную побочную активность расщепления по сравнению с референсным белком CasX.

[0263] Примеры способов характеризации каталитической активности белков CasX могут включать, без ограничений, анализы расщепления *in vitro*, включая описанные в разделе «Примеры». В некоторых вариантах осуществления кинетику расщепления цепи можно оценивать посредством электрофореза продуктов ДНК на агарозных гелях.

n. Аффинность к целевой РНК

[0264] В некоторых вариантах осуществления рибонуклеопротеиновый комплекс, содержащий референсный белок CasX или его вариант, связывается с целевой РНК и расщепляет целевую нуклеиновую кислоту. В некоторых вариантах осуществления варианты референсного белка CasX повышают специфичность варианта белка CasX к целевой РНК и повышают активность варианта белка CasX в отношении целевой РНК по сравнению с референсным белком CasX. Например, варианты белка CasX могут демонстрировать повышенную аффинность связывания с целевыми РНК или повышенное расщепление целевых РНК по сравнению с референсными белками CasX. В некоторых вариантах осуществления рибонуклеопротеиновый комплекс, содержащий вариант белка CasX, связывается с целевой РНК и/или расщепляет целевую РНК. В некоторых вариантах осуществления вариант CasX имеет аффинность связывания с целевой нуклеиновой кислотой, повышенную в от по меньшей мере примерно двух до примерно 10 раз по сравнению с референсным белком с SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3 или с вариантом CasX с SEQ ID NO: 270 или SEQ ID NO: 336.

o. Каталитически неактивные варианты CasX

[0265] В некоторых вариантах осуществления, например, в тех вариантах осуществления, которые охватывают области применения, где расщепление последовательности целевой нуклеиновой кислоты не является требуемым результатом, улучшение каталитической активности варианта белка CasX включает изменение, уменьшение или устранение каталитической активности варианта белка CasX. В некоторых вариантах в описании предложены каталитически неактивные варианты белков CasX, которые, хотя они и способны связывать целевую нуклеиновую кислоту при нахождении в комплексе с гРНК, имеющей нацеливающую последовательность, комплементарную целевой нуклеиновой кислоте, не способны расщеплять целевую нуклеиновую кислоту. Примеры каталитически неактивных белков CasX содержат одну или более мутаций в активном сайте домена RuvC белка CasX. В некоторых вариантах осуществления каталитически неактивный вариант белка CasX содержит замены в остатках 672, 769 и/или 935 относительно SEQ ID NO: 1. В одном варианте осуществления каталитически неактивный вариант белка CasX содержит замены D672A, E769A и/или D935A относительно референсного белка CasX с SEQ ID NO: 1. В других вариантах осуществления

каталитически неактивный вариант белка CasX содержит замены в аминокислотах 659, 756 и/или 922 относительно референсного белка CasX с SEQ ID NO: 2. В некоторых вариантах осуществления каталитически неактивный вариант белка CasX содержит замены D659A, E756A и/или D922A относительно референсного белка CasX с SEQ ID NO: 2. В некоторых вариантах осуществления каталитически неактивные варианты белков CasX 527, 668 и 676 содержат модификации D660A, E757A и D922A для устранения эндонуклеазной активности. В дополнительных вариантах осуществления каталитически неактивный белок CasX содержит делеции всего домена RuvC белка CasX или его части. Следует понимать, что те же вышеуказанные замены могут быть аналогичным образом введены в варианты CasX описания, что приводит к получению каталитически неактивного варианта CasX (dCasX). В одном варианте осуществления весь домен RuvC или его часть удаляют из варианта CasX, что приводит к получению варианта dCasX. Каталитически неактивные варианты белка dCasX в некоторых вариантах осуществления могут использоваться для редактирования оснований или эпигенетических модификаций. Обладая более высокой аффинностью к ДНК, в некоторых вариантах осуществления каталитически неактивные варианты белка dCasX могут, по сравнению с каталитически активным CasX, находить свою целевую нуклеиновую кислоту быстрее, оставаться связанными с целевой нуклеиновой кислотой дольше, связывать целевую нуклеиновую кислоту стабильнее или обладать комбинацией этих характеристик, в результате чего эти функции каталитически неактивного белка CasX улучшаются по сравнению с вариантом CasX, который сохраняет свою способность к расщеплению. Примеры последовательностей вариантов dCasX описаны как SEQ ID NO: 44-62 и 1232-1235, как указано в таблице 7. В некоторых вариантах осуществления вариант dCasX по меньшей мере на 80% идентичен, по меньшей мере на 85% идентичен, по меньшей мере на 90% идентичен, по меньшей мере на 95% идентичен, по меньшей мере на 96% идентичен, по меньшей мере на 97% идентичен, по меньшей мере на 98% или по меньшей мере на 99% идентичен последовательности SEQ ID NO: 44-62 или 1232-1235 и сохраняет функциональные свойства варианта белка dCasX. В некоторых вариантах осуществления вариант dCasX содержит последовательность SEQ ID NO: 44-62 или 1232-1235.

Таблица 7. Каталитически неактивные варианты белков CasX

Конструкт	SEQ ID NO
CAS100	44
CAS098	45
CAS085	46
CAS087	47
CAS086	48
CAS083	49
CAS082	50
CAS069	51
CAS068	52
CAS070	53

CAS071	54
CAS072	55
CAS073	56
CAS074	57
CAS075	58
CAS076	59
CAS077	60
CAS078	61
CAS081	62
CAS096	1232
CAS401	1233
CAS142 (dCasX527)	1234
CAS402 (dCasX676)	1235

p. Слитные белки CasX

[0266] В некоторых вариантах осуществления в описании предложены варианты белка CasX, содержащие гетерологичный белок, слитый с CasX, в том числе с вариантом CasX по любому из вариантов осуществления, описанных в настоящем документе. К ним относятся варианты CasX, содержащие N-концевые, C-концевые или внутренние слияния CasX с гетерологичным белком или его доменом.

[0267] В некоторых вариантах осуществления слитный белок CasX содержит любой из вариантов SEQ ID NO: 247-592 или 1147-1231 или последовательности из таблицы 3, слитые с одним или более белками или их доменами, которые имеют другую интересующую активность, что приводит к получению слитного белка. В некоторых вариантах осуществления слитный белок CasX содержит любой из вариантов SEQ ID NO: 270-592 или 1147-1231, слитый с одним или более белками или их доменами, которые имеют другую интересующую активность. В некоторых вариантах осуществления слитный белок CasX содержит любой из вариантов SEQ ID NO: 415-592 или 1147-1231, слитые с одним или более белками или их доменами, которые имеют другую интересующую активность. Например, в некоторых вариантах осуществления вариант белка CasX слит с белком (или его доменом), который ингибирует транскрипцию, модифицирует целевую нуклеиновую кислоту или модифицирует полипептид, ассоциированный с нуклеиновой кислотой (например, модификация гистона).

[0268] В некоторых вариантах осуществления гетерологичный полипептид (или гетерологичную аминокислоту, например, остаток цистеина или неприродная аминокислота) можно вставить в одно или более положений в белке CasX с созданием слитного белка CasX. В других вариантах осуществления остаток цистеина может быть вставлен в одно или более положений в белке CasX с последующей конъюгацией гетерологичного полипептида, описанного ниже. В некоторых альтернативных вариантах осуществления гетерологичный полипептид или гетерологичную аминокислоту можно добавлять на N- или C-конце референсного или варианта белка CasX. В других вариантах осуществления гетерологичный полипептид или гетерологичную аминокислоту можно

вставлять внутрь последовательности белка CasX.

[0269] В некоторых вариантах осуществления слитный вариант белка CasX сохраняет РНК-направляемую специфичную к последовательности активность по связыванию с целевой нуклеиновой кислотой и расщеплению. В некоторых случаях слитный вариант белка CasX имеет (сохраняет) 50% или более активности (например, активность по расщеплению и/или связыванию) относительно соответствующего варианта белка CasX, который не имеет вставки гетерологичного белка. В некоторых случаях слитный вариант белка CasX сохраняет по меньшей мере примерно 60%, или по меньшей мере примерно 70%, по меньшей мере примерно 80%, или по меньшей мере примерно 90%, или по меньшей мере примерно 92%, или по меньшей мере примерно 95%, или по меньшей мере примерно 98%, или примерно 100% активности (например, активность по расщеплению и/или связыванию) относительно соответствующего белка CasX, который не имеет вставки гетерологичного белка.

[0270] В некоторых случаях референсный CasX или слитный вариант белка CasX сохраняет (имеет) активность по связыванию с целевой нуклеиновой кислотой относительно активности белка CasX без вставленной гетерологичной аминокислоты или гетерологичного полипептида. В некоторых случаях референсный CasX или слитный вариант белка CasX сохраняет по меньшей мере примерно 60%, или по меньшей мере примерно 70%, по меньшей мере примерно 80% или по меньшей мере примерно 90%, или по меньшей мере примерно 92%, или по меньшей мере примерно 95%, или по меньшей мере примерно 98%, или примерно 100% активности связывания относительно соответствующего белка CasX, который не имеет вставки гетерологичного белка.

[0271] В некоторых случаях слитный вариант белка CasX сохраняет (имеет) активность по связыванию и/или расщеплению целевой нуклеиновой кислоты относительно активности исходного белка CasX без вставленной гетерологичной аминокислоты или гетерологичного полипептида. Например, в некоторых случаях слитный вариант белка CasX имеет (сохраняет) 50% или более активности связывания и/или расщепления относительно соответствующего исходного белка CasX (белка CasX, который не имеет вставки). Например, в некоторых случаях слитный вариант белка CasX имеет (сохраняет) 60% или более (70% или более, 80% или более, 90% или более, 92% или более, 95% или более, 98% или более, или 100%) активности по связыванию и/или расщеплению относительно соответствующего исходного белка CasX (белка CasX, который не имеет вставки). Способы измерения активности расщепления и/или связывания белка CasX и/или слитного белка CasX известны специалисту в данной области, и можно применять любой удобный способ.

[0272] Для включения в слитный вариант белка CasX описания подходят разнообразные гетерологичные полипептиды. В некоторых случаях партнер по слиянию может модулировать транскрипцию (например, ингибировать транскрипцию, увеличивать транскрипцию) целевой нуклеиновой кислоты. Например, в некоторых случаях партнер по слиянию представляет собой белок (или домен из белка), который ингибирует

транскрипцию (например, репрессор транскрипции, белок, который функционирует посредством рекрутирования белков-ингибиторов транскрипции, модификации целевой нуклеиновой кислоты, такой как метилирование, рекрутирования модификатора ДНК, модуляции гистонов, связанных с целевой нуклеиновой кислотой, рекрутирования модификатора гистонов, например, модифицирующего ацетилирование и/или метилирование гистонов, и т. п.). В некоторых случаях партнер по слиянию представляет собой белок (или домен из белка), который увеличивает транскрипцию (например, активатор транскрипции, белок, который действует посредством рекрутирования белков-активаторов транскрипции, модификации целевой нуклеиновой кислоты, такой как деметилирование, рекрутирования модификатора ДНК, модуляции гистонов, связанных с целевой нуклеиновой кислотой, рекрутирования модификатора гистонов, например, модифицирующего ацетилирование и/или метилирование гистонов, и т. п.).

[0273] В некоторых случаях партнер по слиянию обладает ферментативной активностью, которая модифицирует последовательность целевой нуклеиновой кислоты; например, нуклеазной активностью, метилтрансферазной активностью, деметилазной активностью, активностью репарации ДНК, активностью повреждения ДНК, активностью деаминирования, дисмутазной активностью, активностью алкилирования, активностью депуринизации, оксидазной активностью, активность образования пиримидинового димера, интеразной активностью, транспозазной активностью, лигазной активностью, геликазной активностью, фотолиазной активностью или гликозилазной активностью. В некоторых вариантах осуществления вариант CasX содержит любую из SEQ ID NO: 247-592 или 1147-1231, и полипептид с активностью метилтрансферазы, активностью деметилазы, активностью ацетилтрансферазы, активностью деацетилазы, активностью киназы, активностью фосфатазы, активностью убиквитинлигазы, деубиквитинирующей активностью, аденилирующей активностью, деаденилирующей активностью, SUMOилирующей активностью, деSUMOилирующей активностью, рибозилирующей активностью, дерибозилирующей активностью, миристоилирующей активностью или демиристоилирующей активностью. В некоторых вариантах осуществления вариант CasX содержит любую из SEQ ID NO: 270-592 или 1147-1231 и полипептид, описанный *выше*. В некоторых вариантах осуществления вариант CasX содержит любую из SEQ ID NO: 415-592 или 1147-1231 и полипептид, описанный *выше*.

[0274] Примеры белков (или их фрагментов), которые можно применять в качестве партнера по слиянию для усиления транскрипции, включают, без ограничений: активаторы транскрипции, такие как VP16, VP64, VP48, VP160, субдомен p65 (например, из NFkB) и активационный домен EDLL и/или активационный домен TAL (например, для активности в растениях); гистоновые лизинметилтрансферазы, такие как SET1A, SET1B, MLL1-5, ASH1, SYMD2, NSD1 и т. п.; гистоновые лизиндеметилазы, такие как JHDM2a/b, UTX, JMJD3 и т. п.; гистоновые ацетилтрансферазы, такие как GCN5, PCAF, CBP, p300, TAF1, TIP60/PLIP, MOZ/MYST3, MORF/MYST4, SRC1, ACTR, P160, CLOCK и т. п.; и ДНК-деметилазы, такие как Ten-Eleven Translocation (TET) диоксигеназа 1 (TET1CD), TET1,

DME, DML1, DML2, ROS1 и т. п.

[0275] Примеры белков (или их фрагментов), которые можно применять в качестве партнера по слиянию для снижения транскрипции, включают, без ограничений: репрессоры транскрипции, такие как Kruppel-ассоциированный бокс (KRAB или SKD); репрессионный домен KOX1; домен взаимодействия mSIN3 (SID) из Mad; репрессорный домен ERF (ERD), репрессорный домен SRDX (например, для репрессии в растениях) и т. п.; гистоновые лизинметилтрансферазы, такие как Pr-SET7/8, SUV4- 20H1, RIZ1 и т. п.; гистоновые лизиндеметилазы, такие как JMJD2A/JHDM3A, JMJD2B, JMJD2C/GASC1, JMJD2D, JARID1A/RBP2, JARID1B/PLU-1, JARID 1C/SMCX, JARID1D/SMCY и т. п.; гистоновые лизиндеацетилазы, такие как HDAC1, HDAC2, HDAC3, HDAC8, HDAC4, HDAC5, HDAC7, HDAC9, SIRT1, SIRT2, HDAC11 и т. п.; ДНК-метилазы, такие как HhaI ДНК m5c-метилтрансфераза (M.HhaI), (M.HhaI), ДНК-метилтрансфераза 1 (DNMT1), ДНК-метилтрансфераза 3a (DNMT3a), ДНК-метилтрансфераза 3b (DNMT3b), METI, DRM3 (растения), ZMET2, CMT1, CMT2 (растения) и т. п.; и периферические рекрутирующие элементы, такие как ламин A, ламин B и т. п.

[0276] В некоторых случаях партнер по слиянию с вариантом CasX обладает ферментативной активностью, которая модифицирует целевую нуклеиновую кислоту (например, оцРНК, дцРНК, оцДНК, дцДНК). Примеры ферментативной активности, которую может обеспечивать партнер по слиянию, включают в себя, без ограничений: нуклеазную активность, например, обеспечиваемую рестрикционным ферментом (например, нуклеазой FokI), метилтрансферазную активность, например, обеспечиваемую метилтрансферазой (например, Hhal ДНК m5c-метилтрансферазой (M.Hhal), ДНК-метилтрансферазой 1 (DNMT1), ДНК-метилтрансферазой 3a (DNMT3a), ДНК-метилтрансферазой 3b (DNMT3b), METI, DRM3 (растения), ZMET2, CMT1, CMT2 (растения) и т. п.); деметилазную активность, например, обеспечиваемую деметилазой (например, Ten-Eleven Translocation (TET) dioxygenase 1 (TET 1 CD), TET1, DME, DML1, DML2, ROS1 и т. п.), активность репарации ДНК, активность повреждения ДНК, деаминирующую активность, например, обеспечиваемую дезаминазой (например, ферментом цитозиндезаминазой, например, белком APOBEC, таким как редактирующий мРНК аполипопротеина B крысиный фермент, каталитический полипептид 1 {APOBEC1}), дисмутазную активность, алкилирующую активность, активность депуринизации, оксидазную активность, активность образования пиримидинового димера, интегразную активность, например, обеспечиваемую интегразой и/или резольвазой (например, Gin-инвертазой, такой как гиперактивный мутант Gin-инвертазы, GinH106Y; интегразой вируса иммунодефицита человека типа 1 (IN); Tn3-резольвазой; и т. п.), транспозазную активность, рекомбиназную активность, такую как обеспечиваемая рекомбиназой (например, каталитическим доменом рекомбиназы Gin), полимеразную активность, лигазную активность, геликазную активность, фотолиазную активность и гликозилазную активность).

[0277] В некоторых случаях вариант белка CasX настоящего описания слит с

полипептидом, выбранным из домена для усиления транскрипции (например, домена VP16, домена VP64), домена для ослабления транскрипции (например, домена KRAB, например, из белка Kox1), основного каталитического домена гистоновой ацетилтрансферазы (например, гистоновой ацетилтрансферазы p300), белка/домена, который обеспечивает обнаруживаемый сигнал (например, флуоресцентного белка, такого как GFP), нуклеазного домена (например, нуклеазы FokI) или редактирующего основания фермента (например, цитидиндезаминазы, такой как APOBEC1).

[0278] В некоторых вариантах осуществления вариант CasX содержит любую из SEQ ID NO: 247-592 или 1147-1231, или любую из SEQ ID NO: 270-592 или 1147-1231, или любую из SEQ ID NO: 415-592 или 1147-1231, или последовательность из таблицы 3, слитую с полипептидом, выбранным из группы, состоящей из домена для уменьшения транскрипции, домена с ферментативной активностью, базового каталитического домена ацетилтрансферазы гистонов, белка/домена, который обеспечивает обнаруживаемый сигнал, нуклеазного домена и редактора оснований. В некоторых вариантах осуществления вариант CasX содержит любую из SEQ ID NO: 247-592 или 1147-1231, слитую с полипептидом, выбранным из группы, состоящей из домена для уменьшения транскрипции, домена с ферментативной активностью, базового каталитического домена ацетилтрансферазы гистонов, белка/домена, который обеспечивает обнаруживаемый сигнал, нуклеазного домена и редактора оснований. В некоторых вариантах осуществления вариант CasX содержит любую из SEQ ID NO: 270-592 или 1147-1231, слитый с полипептидом, описанным *выше*. В некоторых вариантах осуществления вариант CasX содержит любую из SEQ ID NO: 415-592 или 1147-1231, слитый с полипептидом, описанным *выше*. В некоторых вариантах осуществления вариант CasX содержит любую из SEQ ID NO: 760-789, слитую с полипептидом, выбранным из группы, состоящей из домена для уменьшения транскрипции, домена с ферментативной активностью, базового каталитического домена ацетилтрансферазы гистонов, белка/домена, который обеспечивает обнаруживаемый сигнал, нуклеазного домена и редактора оснований. В некоторых вариантах осуществления вариант CasX содержит любую из SEQ ID NO: 411-592, слитую с полипептидом, выбранным из группы, состоящей из домена для уменьшения транскрипции, домена с ферментативной активностью, базового каталитического домена ацетилтрансферазы гистонов, белка/домена, который обеспечивает обнаруживаемый сигнал, нуклеазного домена и редактора оснований.

[0279] В некоторых случаях референсный белок CasX или вариант CasX настоящего описания слит с редактором оснований. Редакторы оснований включают в себя те, которые могут изменять гуанин, аденин, цитозин, тимин или урацил на нуклеозиде или нуклеотиде. Редакторы оснований включают в себя, без ограничений, аденозиндезаминазу, цитозиндезаминазу (например, APOBEC1) и гуаниноксидазу. Соответственно, любой из вариантов CasX, предложенных в настоящем документе, может содержать редактор оснований (т. е. быть слит с ним); например, вариант CasX описания может быть слит с аденозиндезаминазой, цитозиндезаминазой или гуаниноксидазой. В примерах

осуществления вариант CasX описания, который содержит любую из SEQ ID NO: 247-592 или 1147-1231, слит с аденозиндезаминазой, цитозиндезаминазой или гуаниноксидазой. В дополнительных примерах осуществления вариант CasX описания, который содержит любую из SEQ ID NO: 270-592 или 1147-1231, слит с аденозиндезаминазой, цитозиндезаминазой или гуаниноксидазой. В дополнительных примерах осуществления вариант CasX описания, который содержит любую из SEQ ID NO: 415-592 или 1147-1231, слит с аденозиндезаминазой, цитозиндезаминазой или гуаниноксидазой.

[0280] В некоторых случаях партнер по слиянию с вариантом CasX обладает ферментативной активностью, которая модифицирует белок, ассоциированный с целевой нуклеиновой кислотой (например, оцРНК, дцРНК, оцДНК, дцДНК) (например, гистон, РНК-связывающий белок, ДНК-связывающий белок и т. п.). Примеры ферментативной активности (которая модифицируют белок, ассоциированный с целевой нуклеиновой кислотой), которую может обеспечивать партнер по слиянию с вариантом CasX, включают в себя, без ограничений: метилтрансферазную активность, например, обеспечиваемую гистоновой метилтрансферазой (HMT) (например, гомологом 1 супрессора мозаичности 3-9 (SUV39H1, также известен как KMT1A), эухроматической гистоновой лизинметилтрансферазой 2 (G9A, также известна как KMT1C и EHMT2), SUV39H2, ESET/SETDB 1 и т. п., SET1A, SET1B, MLL1-5, ASH1, SMYD2, NSD1, DOT-1-подобная лизинметилтрансфераза гистонов (DOT1L), Pr-SET7/8, лизинметилтрансфераза 5B (SUV4-20H1), энхансер субъединицы zeste 2 репрессорного polycomb-комплекса 2 (EZH2), PR/SET домен 2 (RIZ1), деметилазную активность, например, обеспечиваемую гистоновой деметилазой (например, лизиндеметилазой 1A (KDM1A также известна как LSD1), JHDM2a/b, JMJD2A/JHDM3A, JMJD2B, JMJD2C/GASC1, JMJD2D, JARID1A/RBP2, JARID1B/PLU-1, JARID1C/SMCX, JARID1D/SMCY, UTX, JMJD3 и т. п.), ацетилтрансферазную активность, например, обеспечиваемую гистоновой ацетилазой-трансферазой (например, каталитическое ядро/фрагмент ацетилтрансферазы p300, GCN5, PCAF, CBP, TAF1, TIP60/PLIP, MOZ/MYST3, MORF/MYST4, HB01/MYST2, HMOF/MYST1, SRC1, ACTR, P160, CLOCK и т. п.), деацетилазную активность, например, обеспечиваемую гистоновой деацетилазой (например, HDAC1, HDAC2, HDAC3, HDAC8, HDAC4, HDAC5, HDAC7, HDAC9, SIRT1, SIRT2, HDAC11 и т. п.), киназную активность, фосфатазную активность, активность убиквитинлигазы, деубиквитинирующую активность, аденилирующую активность, деаденилирующую активность, SUMOилирующую активность, деSUMOилирующую активность, рибозилирующую активность, дерибозилирующую активность, миристоилирующую активность или демиристоилирующую активность.

[0281] Дополнительные примеры приемлемых партнеров по слиянию с вариантом CasX представляют собой (i) домен дестабилизации дигидрофолатредуктазы (DHFR) (например, для получения химически контролируемого направляемого РНК субъекта полипептида) и (ii) пептида транзита к хлоропластам.

[0282] В некоторых вариантах осуществления вариант CasX содержит любую из

SEQ ID NO: 247-592 или 1147-1231, или любую из SEQ ID NO: 270-592 или 1147-1231, или любую из SEQ ID NO: 415-592 или 1147-1231, или последовательность из таблицы 3 и пептид транзита к хлоропластам, включая, без ограничений: MASMISSSAVTTVSRASRGQSAAMAPFGGLKSMTGFPVRKVNTDITSITSNGGR VKCMQVWPPIGKKKFETLSYLPPLTRDSRA (SEQ ID NO: 338); MASMISSSAVTTVSRASRGQSAAMAPFGGLKSMTGFPVRKVNTDITSITSNGGRVKS (SEQ ID NO: 339); MASSMLSSATMVASPAQATMVAPFNGLKSSAAFPATRKANNDITSITSNGGRVNCMQV WPPIEKKKFETLSYLPDLTDSGGRVNC (SEQ ID NO: 340); MAQVSRICNGVQNPSLISNLSKSSQRKSPLSVSLKTQQHPRAYPISSSWGLKKSGMTLIG SELRPLKVMSSVSTAC (SEQ ID NO: 341); MAQVSRICNGVWNPSLISNLSKSSQRKSPLSVSLKTQQHPRAYPISSSWGLKKSGMTLIG SELRPLKVMSSVSTAC (SEQ ID NO: 342); MAQINNMAQGIQTLNPNSNFHKPQVPKSSSFLVFGSKKLKNSANSMLVLKKDSIFMQLF CSFRISASVATAC (SEQ ID NO: 343); MAALVTSQLATSGTVLSVTDRFRRPGFQGLRPRNPADAALGMRTVGASAAPKQSRKPH RFDRRCLSMVV (SEQ ID NO: 344); MAALTTSQLATSATGFGIADRSAPSSLLRHGFQGLKPRSPAGGDATSLSVTTSARATPKQ QRSVQRGSRRFPSVVVC (SEQ ID NO: 345); MASSVLSSAAVATRSNVAQANMVAPFTGLKSAASFPVSRKQNLDITSIASNGGRVQC (SEQ ID NO: 346); MESLAATSVFAPSRVAVPAARALVRAGTVVPTRRTSSTSGTSGVKCSAAVTPQASPVIS RSAAAA (SEQ ID NO: 347) и MGAAATSMQSLKFSNRLVPPSRRLSPVPNNVTCNNLPKSAAPVRTVKCCASSWNSTING AAATTNGASAASS (SEQ ID NO: 348).

[0283] В некоторых случаях вариант белка CasX настоящего описания может включать пептид избегания эндосом. В некоторых случаях полипептид избегания эндосом содержит аминокислотную последовательность GLFXALLXLLXSLWXLLLXA (SEQ ID NO: 349), причем каждый X независимо выбран из лизина, гистидина и аргинина. В некоторых случаях полипептид избегания эндосом содержит аминокислотную последовательность GLFHALLHLLHSLWHLLLHA (SEQ ID NO: 350) или HHHHHHHHH (SEQ ID NO: 351). В некоторых вариантах осуществления вариант CasX содержит последовательность с любой из SEQ ID NO: 247-592 или 1147-1231, или любую из SEQ ID NO: 270-592 или 1147-1231, или любую из SEQ ID NO: 415-592 или 1147-1231, или последовательность из таблицы 3 и полипептид избегания эндосом.

[0284] Не имеющие ограничительного характера примеры приемлемых партнеров по слиянию для варианта белка CasX для применения при нацеливании на целевые нуклеиновые кислоты оцРНК включают в себя (без ограничений): факторы сплайсинга (например, домены RS); компоненты трансляции белка (например, факторы инициации трансляции, элонгации и/или высвобождения; например, эукариотический фактор

инициации трансляции 4 гамма {eIF4G}); РНК-метилазы; ферменты редактирования РНК (например, РНК-дезаминазы, например, аденозиндезаминаза, действующая на РНК (ADAR), в том числе ферменты редактирования A в I и/или C в U); геликазы; РНК-связывающие белки и т. п. Следует понимать, что гетерологичный полипептид может включать в себя весь белок или в некоторых случаях может включать в себя фрагмент белка (например, функциональный домен).

[0285] В некоторых вариантах осуществления вариант CasX по любой из SEQ ID NO: 247-592 или 1147-1231, или любую из SEQ ID NO: 270-592 или 1147-1231, или любую из SEQ ID NO: 415-592 или 1147-1231, или последовательность из таблицы 3 содержит партнера по слиянию - любой домен, способный взаимодействовать с оцРНК (к которому для целей настоящего описания относятся внутримолекулярные и/или межмолекулярные вторичные структуры, например, двухцепочечные дуплексы РНК, такие как шпильки, стеблевые петли и т. д.), как временно, так и необратимо, прямо или косвенно, включая, без ограничений, эффекторный домен, выбранный из группы, включающей в себя: эндонуклеазы (например, РНКаза III, домен CRR22 DYW, домены Dicer и PIN (N-конец PilT) из белков, таких как SMG5 и SMG6); белки и белковые домены, ответственные за стимуляцию расщепления РНК (например, специфический фактор расщепления и полиаденилирования {CPSF}, фактор стимуляции расщепления {CstF}, CFIm и CFIIm); экзонуклеазы (например, хроматинсвязывающая экзонуклеаза XRN1 (XRN-1) или экзонуклеаза T); деаденилазы (например, ДНК 5'-аденозин монофосфат гидролаза {HNT3}); белки и белковые домены, ответственные за нонсенс-опосредованное разрушение РНК (например, UPF1 РНК-геликаза и АТФаза {UPF1}, UPF2, UPF3, UPF3b, RNP SI, содержащий РНК-связывающий мотив белок 8A {Y14}, протоонкоген DEK {DEK}, РНК-процессирующий белок REF2 {REF2}, матричный белок 1 с серин-аргининовыми повторами {SRm160}); белки и белковые домены, ответственные за стабилизацию РНК (например, поли(A)-связывающий цитоплазматический белок 1, {PABP}); белки и белковые домены, ответственные за подавление трансляции (например, каталитический RISC-компонент argonaute 2 {Ago2} и Ago4); белки и белковые домены, ответственные за стимуляцию трансляции (например, Staufen); белки и белковые домены, ответственные за (например, способные к) модуляцию трансляции (например, факторы трансляции, такие как факторы инициации, факторы элонгации, факторы высвобождения и т. д., например, eIF4G); белки и белковые домены, ответственные за полиаденилирование РНК (например, поли(A) полимеразы (PAP1), белок, содержащий ассоциированный с PAP домен;поли(A)-РНК-полимеразу gld-2 {GLD-2} и Star-PAP); белки и белковые домены, ответственные за полиуридинилирование РНК (например, концевая уридилилтрансфераза {CID1} и концевая уридилаттрансфераза); белки и белковые домены, ответственные за локализацию РНК (например, связывающий мРНК инсулиноподобного фактора роста 2 белок 1 {IMP1}, z-ДНК связывающий белок 1 {ZBP1}, She2p, She3p и Bicaudal-D); белки и белковые домены, ответственные за задержку РНК в ядре (например, Rrp6); белки и белковые домены, ответственные за экспорт РНК из ядра (например, фактор 1 ядерного экспорта РНК {TAP},

фактор 1 ядерного экспорта РНК {NXF1}, комплекс THO {THO}, TREX, REF и фактор экспорта Aly/REF {Aly}); белки и белковые домены, ответственные за подавление сплайсинга РНК (например, связывающий полипиримидиновый тракт белок 1 {PTB}, KH РНК-связывающий домен-содержащий, ассоциированный с сигнальной трансдукцией белок 1 Sam68} и гетерогенный ядерный рибонуклеопротеин A1 {hnRNP A1}); белки и белковые домены, ответственные за стимуляцию сплайсинга РНК (например, домены, богатые серином/аргинином (SR)); белки и белковые домены, ответственные за снижение эффективности транскрипции (например, РНК-связывающий белок FUS {FUS (TLS)}); а также белки и белковые домены, ответственные за стимуляцию транскрипции (например, циклин-зависимая киназа 7 (CDK7) и Tat ВИЧ). Альтернативно эффекторный домен может быть выбран из группы, содержащей эндонуклеазы; белки и белковые домены, способные стимулировать расщепление РНК; экзонуклеазы; деаденилазы; белки и белковые домены, имеющие нонсенс-опосредованную активность по распаду РНК; белки и белковые домены, способные стабилизировать РНК; белки и белковые домены, способные подавлять трансляцию; белки и белковые домены, способные стимулировать трансляцию; белки и белковые домены, способные модулировать трансляцию (например, факторы трансляции, такие как факторы инициации, факторы элонгации, факторы высвобождения и т. д., например, eIF4G); белки и белковые домены, способные к полиаденилированию РНК; белки и белковые домены, способные к полиуридинилированию РНК; белки и белковые домены, обладающие активностью по локализации РНК; белки и белковые домены, способные к задерживанию РНК в ядре; белки и белковые домены, обладающие активностью ядерного экспорта РНК; белки и белковые домены, способные подавлять сплайсинг РНК; белки и белковые домены, способные стимулировать сплайсинг РНК; белки и белковые домены, способные снижать эффективность транскрипции; и белки и белковые домены, способные стимулировать транскрипцию. Другой приемлемый гетерологичный полипептид представляет собой РНК-связывающий домен PUF, который более подробно описан в WO2012068627, полностью включенной в настоящий документ путем ссылки.

[0286] Некоторые факторы сплайсинга РНК, которые можно применять (полностью или в виде его фрагментов) в качестве партнера по слиянию с вариантом CasX, имеют модульную организацию с отдельными специфичными для последовательности РНК-связывающими модулями и эффекторными доменами сплайсинга. Например, члены семейства богатых серином/аргинином (SR) белков содержат N-концевые мотивы распознавания РНК (RRM), которые связываются с экзонными энхансерами сплайсинга (ESE) в пре-мРНК и С-концевыми доменами RS, которые способствуют включению экзона. В качестве другого примера, hnRNP-белок hnRNP A1 связывается с экзонными сайленсерами сплайсинга (ESS) своими доменами RRM и ингибирует включение экзона с помощью С-концевого богатого глицином домена. Некоторые факторы сплайсинга могут регулировать альтернативное использование сайта сплайсинга (ss) посредством связывания с регуляторными последовательностями между двумя альтернативными сайтами.

Например, ASF/SF2 может распознавать ESE и способствовать использованию проксимальных к интрону сайтов, тогда как hnRNP A1 может связываться с ESS и сдвигать сплайсинг к использованию дистальных к интрону сайтов. Одним из вариантов применения таких факторов является создание ESF, которые модулируют альтернативный сплайсинг эндогенных генов, в частности генов, связанных с заболеванием. Например, пре-мРНК BCL2-подобного белка 1 (Bcl-x) дает две сплайсинговые изоформы с двумя альтернативными 5'-сайтами сплайсинга для кодирования белков с противоположными функциями. Длинная сплайсинговая изоформа Bcl-xL представляет собой мощный ингибитор апоптоза, экспрессируемый в долгоживущих постмитотических клетках, и она имеет повышенную экспрессию во многих раковых клетках, защищая клетки против апоптотических сигналов. Короткая изоформа Bcl-xS представляет собой проапоптотическую изоформу и экспрессируется с высокими уровнями в клетках с высокой скоростью оборота (например, в развивающихся лимфоцитах). Соотношение двух сплайсинговых изоформ Bcl-x регулируется множеством cc-элементов, которые расположены либо в основной области экзона, либо в удлинительной области экзона (т. е. между двумя альтернативными 5'-сайтами сплайсинга). См. дополнительные примеры в WO2010075303, полностью включенной в настоящий документ путем ссылки.

[0287] Дополнительные приемлемые партнеры по слиянию для использования с вариантом CasX включают в себя, без ограничений, белки (или их фрагменты), которые представляют собой граничные элементы (например, CTCF), белки и их фрагменты, которые обеспечивают рекрутирование периферии (например, ламин A, ламин B и т. д.) и стыковочные элементы белков (например, FKBP/FRB, Pill/Abyl и т. д.).

[0288] Дополнительно или альтернативно вариант белка CasX настоящего описания может быть слит с полипептидным доменом проникновения, который способствует поглощению клеткой. В данной области известен ряд доменов проникновения, и их можно использовать в неинтегрирующихся полипептидах настоящего описания, включая пептиды, пептидомиметики и непептидные носители. Например, в WO2017/106569 и US20180363009A1, полностью включенных в настоящий документ путем ссылки, описано слияние белка Cas с одной или более последовательностями сигнала ядерной локализации (NLS) для облегчения поглощения клеткой. В других вариантах осуществления пептид проникновения может быть получен из третьей альфа-спирали транскрипционного фактора Antennapaedia *Drosophila melanogaster*, называемого пенетратином, который содержит аминокислотную последовательность RQIKIWFQNRRMKWKK (SEQ ID NO: 398). В качестве другого примера, пептид проникновения содержит аминокислотную последовательность основной области белка tat ВИЧ-1, которая может включать в себя, например, аминокислоты 49-57 природного белка tat. Другие домены проникновения включают полиаргининовые мотивы, например область аминокислот 34-56 белка rev ВИЧ -1, нона-аргинин, окта-аргинин и т. п. Сайт, в котором происходит слияние, может быть выбран с учетом оптимизации биологической активности, секреции или связывающих характеристик полипептида. Оптимальный сайт определяют с помощью стандартных

экспериментов.

[0289] В некоторых случаях гетерологичный полипептид (партнер по слиянию) для использования с вариантом CasX обеспечивает субклеточную локализацию; т. е. гетерологичный полипептид содержит последовательность субклеточной локализации (например, сигнал ядерной локализации (NLS) для нацеливания на ядро, последовательность для удержания слитного белка вне ядра, например последовательность экспорта из ядра (NES), последовательность для удержания слитного белка в цитоплазме, сигнал митохондриальной локализации для нацеливания на митохондрию, сигнал локализации в хлоропластах для нацеливания на хлоропласт, сигнал удерживания в эндоплазматическом ретикулуме (ER) и т. п.). В некоторых вариантах осуществления рассматриваемый РНК-направляемый полипептид или условно активный РНК-направляемый полипептид и/или рассматриваемый слитный белок CasX не включает в себя NLS, так что белок не нацелен на ядро, что может обеспечивать преимущество; например, когда целевая нуклеиновая кислота представляет собой РНК, которая присутствует в цитозоле. В некоторых вариантах осуществления партнер по слиянию может обеспечивать метку (т. е. гетерологичный полипептид представляет собой детектируемую метку) для простоты отслеживания и/или очистки (например, флуоресцентный белок, например, зеленый флуоресцентный белок (GFP), желтый флуоресцентный белок (YFP), красный флуоресцентный белок (RFP), голубой флуоресцентный белок (CFP), mCherry, tdTomato и т. п.; гистидиновую метку, например метку 6XHis; гемагглютининовую (HA) метку; метку FLAG; метку Myc и т. п.). В некоторых вариантах осуществления вариант CasX содержит любую из SEQ ID NO: XX-XX и последовательность субклеточной локализации или метку.

[0290] В некоторых случаях референсный или вариант белка CasX включает в себя сигнал ядерной локализации (NLS) (слит с ним). Не имеющие ограничительного характера примеры NLS, приемлемых для применения с вариантом CasX, включают в себя последовательности, на по меньшей мере примерно 80%, по меньшей мере примерно 90%, или по меньшей мере примерно 95% идентичные или идентичные последовательностям, полученным из: NLS большого Т-антигена вируса SV40, имеющего аминокислотную последовательность PKKKRKV (SEQ ID NO: 352); NLS из нуклеоплазмина (например, нуклеоплазмина двухкомпонентного NLS с последовательностью KRPAATKKAGQAKKKK (SEQ ID NO: 353); NLS, c-myc, имеющий аминокислотную последовательность PAAKRVKLD (SEQ ID NO: 354)) или RQRRNELKRSP (SEQ ID NO: 355); NLS hRNPAl M9, имеющий последовательность NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY (SEQ ID NO: 356); последовательность RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV (SEQ ID NO: 357) домена IBB из импортина-альфа; последовательность VSRKRPRP (SEQ ID NO: 358) и PPKKARED (SEQ ID NO: 359) из Т-белка миомы; последовательность PQPKKKPL (SEQ ID NO: 360) из человеческого p53; последовательность SALIKKKKKMAP (SEQ ID NO: 361) из мышиного c-abl IV; последовательности DRLRR (SEQ ID NO: 362) и PKQKKRK (SEQ ID NO: 363) из вируса гриппа NS1;

последовательность RKLKKKIKKL (SEQ ID NO: 364) из дельта-антигена вируса гепатита; последовательность REKKKFLKRR (SEQ ID NO: 365) из мышиного белка Mxl; последовательность KRKGDEVDGVDEVAKKKSKK (SEQ ID NO: 366) из поли(АДФ-рибоза) полимеразы человека; последовательность RKCLQAGMNLEARKTKK (SEQ ID NO: 367) из рецепторов стероидных гормонов (человеческих), глюкокортикоидных; последовательность PRPRKIPR (SEQ ID NO: 368) из P-белка вируса болезни Борна (BDV-P1); последовательность PPRKKRTVV (SEQ ID NO: 369) из неструктурного белка вируса гепатита C (HCV-NS5A);последовательность NLSKKKKRKREK (SEQ ID NO: 370) из LEF1; последовательность RRPSRPFRKP (SEQ ID NO: 371) из ORF57 обезьян; последовательность KRPRSPSS (SEQ ID NO: 372) из EBV LANA; последовательность KRGINDRNFWRGENERKTR (SEQ ID NO: 373) из белка гриппа A; последовательность PRPPKMARYDN (SEQ ID NO: 374) из РНК-геликазы A человека (RHA); последовательность KRSFSKAF (SEQ ID NO: 375) из ядрышковой РНК-геликазы II; последовательность KLKIKRPVK (SEQ ID NO: 376) из TUS-белка; последовательность PKKKRKVPPPPAAKRVKLD (SEQ ID NO: 377), ассоциированная с импортином-альфа; последовательность PKTRRRPRRSQRKRPPT (SEQ ID NO: 378) из белка Rex в HTLV-1; последовательность SRRRKANPTKLSENAKKLAKEVEN (SEQ ID NO: 379) из белка EGL-13 Caenorhabditis elegans и последовательности KTRRRPRRSQRKRPPT (SEQ ID NO: 380), RRKKRRPRRKKRR (SEQ ID NO: 381), PKKKSRKPKKKSRK (SEQ ID NO: 382), HKKKHPDASVNFSEFSK (SEQ ID NO: 383), QRPGPYDRPQRPGPYDRP (SEQ ID NO: 384), LSPSLSPLLSPSLSPL (SEQ ID NO: 385), RGKGGKGLGKGGAKRHRK (SEQ ID NO: 386), PKRGRGRPKRGRGR (SEQ ID NO: 387), PKKKRKVPPPPAAKRVKLD (SEQ ID NO: 388), PKKKRKVPPPPKKKRKV (SEQ ID NO: 389), PAKRARRGYKC (SEQ ID NO: 63), KLGPRKATGRW (SEQ ID NO: 64), PRRKREE (SEQ ID NO: 65), PYRGRKE (SEQ ID NO:66), PLRKRPRR (SEQ ID NO: 67), PLRKRPRRGSPLRKRPRR (SEQ ID NO:68), PAAKRVKLDGGKRTADGSEFESPKKKRKV (SEQ ID NO: 69), PAAKRVKLDGGKRTADGSEFESPKKKRKVGIHGVPAA (SEQ ID NO: 70), PAAKRVKLDGGKRTADGSEFESPKKKRKVAEAAAKEAAAKEAAAKA (SEQ ID NO: 71), PAAKRVKLDGGKRTADGSEFESPKKKRKVPG (SEQ ID NO: 72), KRKGSPERGERKRHW (SEQ ID NO: 73), KRTADSQHSTPPKTKRKVEFEPKKKRKV (SEQ ID NO: 74) и PKKKRKVGGSKRTADSQHSTPPKTKRKVEFEPKKKRKV (SEQ ID NO: 75). В некоторых вариантах осуществления один или более NLS связаны с CRISPR-белком или со смежным NLS линкерным пептидом, причем линкерный пептид выбран из группы, состоящей из RS, (G)n (SEQ ID NO: 1023), (GS)n (SEQ ID NO: 1024), (GSGGS)n (SEQ ID NO: 399), (GGSGGS)n (SEQ ID NO: 400), (GGGS)n (SEQ ID NO: 401), GGSG (SEQ ID NO: 402), GGSGG (SEQ ID NO: 403), GSGSG (SEQ ID NO: 404), GSGGG (SEQ ID NO: 405), GGGSG (SEQ ID NO: 406), GSSSG (SEQ ID NO: 407), GPGP (SEQ ID NO: 408), GGP, PPP, PPAPPA (SEQ ID NO: 409), PPPG (SEQ ID NO: 24), PPPGPPP (SEQ ID NO: 410), PPP(GGGS)n (SEQ ID NO: 25), (GGGS)nPPP (SEQ ID NO: 26), AEAAAKEAAAKEAAAKA (SEQ ID NO: 1025) и TPPKTKRKVEFE (SEQ ID NO: 27), где n составляет от 1 до 5. Как правило, NLS (или

множество NLS) обладают достаточной силой, чтобы обеспечивать накопление слитного варианта белка CasX в ядре эукариотической клетки. Обнаружение накопления в ядре может осуществляться любой приемлемой методикой. Например, обнаруживаемый маркер может быть слит со слитным вариантом белка CasX, так что местоположение внутри клетки может быть визуализировано. Также можно выделить клеточные ядра из клеток и проанализировать их содержание любым приемлемым способом обнаружения белка, например, методом иммуногистохимии, вестерн-блоттинга или анализом ферментативной активности. Накопление в ядре также может быть определено опосредованно.

[0291] В описании предусмотрена сборка множества NLS в различных конфигурациях для соединения с белком CRISPR. В некоторых вариантах осуществления 1, 2, 3, 4 или более NLS связаны линкерными пептидами с N-концом белка CRISPR. В других вариантах осуществления 1, 2, 3, 4 или более NLS связаны линкерными пептидами с C-концом белка CRISPR. В некоторых вариантах осуществления NLS, связанный с N-концом белка CRISPR, идентичен NLS, связанному с C-концом. В других вариантах осуществления NLS, связанный с N-концом белка CRISPR, отличается от NLS, связанного с C-концом. В некоторых вариантах осуществления NLS, связанный с N-концом белка CRISPR, выбран из группы, состоящей из N-концевых последовательностей, представленных в таблице 8. В некоторых вариантах осуществления NLS, связанный с C-концом белка CRISPR, выбран из группы, состоящей из C-концевых последовательностей, представленных в таблице 8. Обнаружение накопления в ядре может осуществляться любой приемлемой методикой. Например, обнаруживаемый маркер может быть слит с референсным или слитным вариантом белка CasX, так что местоположение внутри клетки может быть визуализировано. Также можно выделить клеточные ядра из клеток и проанализировать их содержание любым приемлемым способом обнаружения белка, например, методом иммуногистохимии, вестерн-блоттинга или анализом ферментативной активности. Накопление в ядре также может быть определено опосредованно.

Таблица 8. Последовательности NLS

N-концевые последовательности	**SEQ ID NO**	**C-концевые последовательности**	**SEQ ID NO**
PKKKRKVGGSPKKKRKVSRQEIK RINKIRRRLVKDSNTKKAGKTGP	219	TLESPAAKRVKLDGGSPAAK RVKLDGGSPAAKRVKLDGGS PAAKRVKLDGGSPAAKRVKL DGGSPAAKRVKLDTLESKRP AATKKAGQAKKKKGGSKRP AATKKAGQAKKKKGGSKRP AATKKAGQAKKKKGGSKRP AATKKAGQAKKKK	1236
PKKKRKVGGSPKKKRKVGGSPK KKRKVGGSPKKKRKVSRQEIKRI NKIRRRLVKDSNTKKAGKTGP	220	TLESKRPAATKKAGQAKKKK TLESKRPAATKKAGQAKKKK GGSKRPAATKKAGQAKKKK GGSKRPAATKKAGQAKKKK GGSKRPAATKKAGQAKKKK GGSKRPAATKKAGQAKKKK GGSKRPAATKKAGQAKKKK	1237

PKKKRKVGGSPKKKRKVGGSPKKKRKVGGSPKKKRKVGGSPKKKRKVGGSPKKKRKVSRQEIKRINKIRRRLVKDSNTKKAGKTGP	221	TLESKRPAATKKAGQAKKKKGGSKRPAATKKAGQAKKKKTLESPKKKRKVGGSPKKKRKVGGSPKKKRKVGGSPKKKRKV	1238
PAAKRVKLDGGSPAAKRVKLDSRQEIKRINKIRRRLVKDSNTKKAGKTGP	222	TLEGGSPKKKRKVTLESPKKKRKVGGSPKKKRKVGGSPKKKRKVGGSPKKKRKV	1239
PAAKRVKLDGGSPAAKRVKLDGGSPAAKRVKLDGGSPAAKRVKLDSRQEIKRINKIRRRLVKDSNTKKAGKTGP	223	TLEGGSPKKKRKVTLESPAAKRVKLDGGSPAAKRVKLDGGSPAAKRVKLDGGSPAAKRVKLD	1240
PAAKRVKLDGGSPAAKRVKLDGGSPAAKRVKLDGGSPAAKRVKLDDGGSPAAKRVKLDGGSPAAKRVKLDSRQEIKRINKIRRRLVKDSNTKKAGKTGP	224	TLEGGSPKKKRKVTLESPAAKRVKLDGGSPAAKRVKLDGGSPAAKRVKLDGGSPAAKRVKLDGGSPAAKRVKLD	1241
KRPAATKKAGQAKKKKSRDISRQEIKRINKIRRRLVKDSNTKKAGKTGP	225	TLEGGSPKKKRKVTLESKRPAATKKAGQAKKKK	1242
KRPAATKKAGQAKKKKSRQEIKRINKIRRRLVKDSNTKKAGKTGP	226	TLEGGSPKKKRKVTLESKRPAATKKAGQAKKKKGGSKRPAATKKAGQAKKKK	1243
KRPAATKKAGQAKKKKGGSKRPAATKKAGQAKKKKSRDISRQEIKRINKIRRRLVKDSNTKKAGKTGP	227	TLEGGSPKKKRKVTLEGGSPKKKRKV	1244
KRPAATKKAGQAKKKKGGSKRPAATKKAGQAKKKKGGSKRPAATKKAGQAKKKKGGSKRPAATKKAGQAKKKKSRDISRQEIKRINKIRRRLVKDSNTKKAGKTGP	228	TLEGGSPKKKRKVTLEGGSPKKKRKV	1244
KRPAATKKAGQAKKKKGGSKRPAATKKAGQAKKKKGGSKRPAATKKAGQAKKKKGGSKRPAATKKAGQAKKKKGGSKRPAATKKAGQAKKKKSRDISRQEIKRINKIRRRLVKDSNTKKAGKTGP	229	TLEGGSPKKKRKVTLEGGSPKKKRKV	1244
PKKKRKVGGSPKKKRKVGGSPKKKRKVGGSPKKKRKVSRDISRQEIKRINKIRRRLVKDSNTKKAGKTGP	230	TLEGGSPKKKRKVTLEGGSPKKKRKV	1244
PKKKRKVGGSPKKKRKVGGSPKKKRKVGGSPKKKRKVSRDISRQEIKRINKIRRRLVKDSNTKKAGKTGP	230	TLEGGSPKKKRKVTLEGGSPKKKRKV	1244
PAAKRVKLDGGSPAAKRVKLDGGSPAAKRVKLDSRDISRQEIKRINKIRRRLVKDSNTKKAGKTGP	231	TLEVGPKRTADSQHSTPPKTKRKVEFEPKKKRKVTLEGGSPKKKRKV	1245
PAAKRVKLDGGSPAAKRVKLDGGSPAAKRVKL	232	TLEVGGGSGGGSKRTADSQHSTPPKTKRKVEFEPKKKRKVT	1246

DGGSPAAKRVKLDGGSPAAKRVKLDSRDISRQEIKRINKIRRRLVKDSNTKKAGKTGP		LEGGSPKKKRKV	
KRPAATKKAGQAKKKKSRDISRQEIKRINKIRRRLVKDSNTKKAGKTGP	225	TLEVAEAAAKEAAAKEAAAKAKRTADSQHSTPPKTKRKVEFEPKKKRKVTLEGGSPKKKRKV	1247
KRPAATKKAGQAKKKKGGSKRPAATKKAGQAKKKKSRDISRQEIKRINKIRRRLVKDSNTKKAGKTGP	227	TLEVGPPKKKRKVGGSKRTADSQHSTPPKTKRKVEFEPKKKRKVTLEGGSPKKKRKV	1248
PAAKRVKLDGGKRTADGSEFESPKKKRKVGGSSRDISRQEIKRINKIRRRLVKDSNTKKAGKTGP	233	TLEVGPAEAAAKEAAAKEAAAKAPAAKRVKLDTLEGGSPKKKRKV	1249
PAAKRVKLDGGKRTADGSEFESPKKKRKVPPPPGSRDISRQEIKRINKIRRRLVKDSNTKKAGKTGP	234	TLEVGPGGGSGGGSGGGSPAAKRVKLDTLEVGPKRTADSQHSTPPKTKRKVEFEPKKKRKV	1250
PAAKRVKLDGGKRTADGSEFESPKKKRKVGIHGVPAAPGSRDISRQEIKRINKIRRRLVKDSNTKKAGKTGP	235	TLEVGPPKKKRKVPPPPAAKRVKLDTLEVGGGSGGGSKRTADSQHSTPPKTKRKVEFEPKKKRKV	1251
PAAKRVKLDGGKRTADGSEFESPKKKRKVGGGSGGGSPGSRDISRQEIKRINKIRRRLVKDSNTKKAGKTGP	236	TLEVGPPAAKRVKLDTLEVAEAAAKEAAAKEAAAKAKRTADSQHSTPPKTKRKVEFEPKKKRKV	1252
PAAKRVKLDGGKRTADGSEFESPKKKRKVPGGGSGGGSPGSRDISRQEIKRINKIRRRLVKDSNTKKAGKTGP	239	TLEVGPKRTADSQHSTPPKTKRKVEFEPKKKRKVTLEVGPPKKKRKVGGSKRTADSQHSTPPKTKRKVEFEPKKKRKV	1253
PAAKRVKLDGGKRTADGSEFESPKKKRKVAEAAAKEAAAKEAAAKAPGSRDISRQEIKRINKIRRRLVKDSNTKKAGKTGP	983	TLEVGGGSGGGSKRTADSQHSTPPKTKRKVEFEPKKKRKVTLEVGPAEAAAKEAAAKEAAAKAPAAKRVKLD	1254
PAAKRVKLDGGKRTADGSEFESPKKKRKVPGSRDISRQEIKRINKIRRRLVKDSNTKKAGKTGP	984	GSKRPAATKKAGQAKKKKTLEVGPGGGSGGGSGGGSPAAKRVKLD	1255
PAAKRVKLDGGSPKKKRKVGGSSRDISRQEIKRINKIRRRLVKDSNTKKAGKTGP	985	GSKRPAATKKAGQAKKKKTLEVGPPKKKRKVPPPPAAKRVKLD	1256
PAAKRVKLDPPPPKKKRKVPGSRDISRQEIKRINKIRRRLVKDSNTKKAGKTGP	986	GSKRPAATKKAGQAKKKKTLEVGPPAAKRVKLD	1257
PAAKRVKLDPGRSRDISRQEIKRINKIRRRLVKDSNTKKAGKTGP	987	GSPKKKRKVTLEVGPKRTADSQHSTPPKTKRKVEFEPKKKRKV	1258
PKKKRKVSRDISRQEIKRINKIRRRLVKDSNTKKAGKTGP	988	GSKRPAATKKAGQAKKKKTLEVGGGSGGGSKRTADSQHSTPPKTKRKVEFEPKKKRKV	1259
PAAKRVKLDGGKRTADGSEFESPKKKRKVGGSSRDISRQEIKRINKIRRRLVKDSNTKKAGKTGP	233	GSKRPAATKKAGQAKKKKGSKRPAATKKAGQAKKKK	1260

PAAKRVKLDGGKRTADGSEFESPKKKRKVGGGSGGGSPGSRDISRQEIKRINKIRRRLVKDSNTKKAGKTGP	236	GSKRPAATKKAGQAKKKKGSKRPAATKKAGQAKKKK	1261
PKKKRKVSRQEIKRINKIRRRLVKDSNTKKAGKTGP	989	GSKRPAATKKAGQAKKKKGSKRPAATKKAGQAKKKK	1262
PKKKRKVGGSPKKKRKVGGSPKKKRKVGGSPKKKRKVSRQEIKRINKIRRRLVKDSNTKKAGKTGP	220	GSPKKKRKVGSPKKKRKV	1263
PKKKRKVGGSPKKKRKVGGSPKKKRKVGGSPKKKRKVGGSPKKKRKVSRQEIKRINKIRRRLVKDSNTKKAGKTGP	221	GGGSGGGSKRTADSQHSTPPKTKRKVEFEPKKKRKVGSKRPAATKKAGQAKKKK	1264
PAAKRVKLDSRQEIKRINKIRRRLVKDSNTKKAGKTGP	990	GPPKKKRKVGGSKRTADSQHSTPPKTKRKVEFEPKKKRKVGSKRPAATKKAGQAKKKK	1265
PAAKRVKLDGGSPAAKRVKLDSRQEIKRINKIRRRLVKDSNTKKAGKTGP	222	TGGGPGGGAAAGSGSPKKKRKVGSGSGSKRPAATKKAGQAKKKK	1266
PAAKRVKLDGGSPAAKRVKLDGGSPAAKRVKLDGGSPAAKRVKLDSRQEIKRINKIRRRLVKDSNTKKAGKTGP	223	GPKRTADSQHSTPPKTKRKVEFEPKKKRKVGSKRPAATKKAGQAKKKK	1267
PAAKRVKLDGGSPAAKRVKLDGGSPAAKRVKLDGGSPAAKRVKLDGGSPAAKRVKLDSRQEIKRINKIRRRLVKDSNTKKAGKTGP	224	AEAAAKEAAAKEAAAKAKRTADSQHSTPPKTKRKVEFEPKKKRKVGSPKKKRKV	1268
KRPAATKKAGQAKKKKSRQEIKRINKIRRRLVKDSNTKKAGKTGP	226	GPPKKKRKVPPPPAAKRVKLDGGGSGGGSKRTADSQHSTPPKTKRKVEFEPKKKRKV	1269
TSPKKKRKVALEYPYDVPDYA	991	GSPAAKRVKLDGGSPAAKRVKLDGGSPAAKRVKLDGGSPAAKRVKLDGGSPAAKRVKLDGGSPAAKRVKLDGPPKKKRKVGGSKRTADSQHSTPPKTKRKVEFEPKKKRKV	1270
TLESKRPAATKKAGQAKKKKAPGEYPYDVPDYA	992	GSPAAKRVKLGGSPAAKRVKLGGSPKKKRKVGGSPKKKRKVTGGGPGGGAAAGSGSPKKKRKVGSGS	1271
GSKRPAATKKAGQAKKKKYPYDVPDYA	993	GSKRPAATKKAGQAKKKKGGSKRPAATKKAGQAKKKKGPKRTADSQHSTPPKTKRKVEFEPKKKRKV	1272
TLESKRPAATKKAGQAKKKKGGSKRPAATKKAGQAKKKKAPGEYPYDVPDYATSPKKKRKVALEYPYDVPDYA	1274	GSKRPAATKKAGQAKKKKGGSKRPAATKKAGQAKKKKAEAAAKEAAAKEAAAKAKRTADSQHSTPPKTKRKVEFEPKKKRKV	1273
TLESKRPAATKKAGQAKKKKGG	1275	GPPKKKRKVPPPPAAKRVKL	1018

SKRPAATKKAGQAKKKKGGSKRPAATKKAGQAKKKKGGSKRPAATKKAGQAKKKKTSPKKKRKVALEYPYDVPDYA		D	
TLESKRPAATKKAGQAKKKKGGSKRPAATKKAGQAKKKKGGSKRPAATKKAGQAKKKKGGSKRPAATKKAGQAKKKKGGSKRPAATKKAGQAKKKKTSPKKKRKVALEYPYDVPDYA	1276	GSPAAKRVKLDGGSPAAKRVKLDGGSPAAKRVKLDGGSPAAKRVKLDGGSPAAKRVKLDGGSPAAKRVKLD	1019
TLESPKKKRKVGGSPKKKRKVGGSPKKKRKVGGSPKKKRKVTLESKRPAATKKAGQAKKKKAPGEYPYDVPDYA	1277	GSPAAKRVKLGGSPAAKRVKLGGSPKKKRKVGGSPKKKRKV	1020
TLESPKKKRKVGGSPKKKRKVGGSPKKKRKVGGSPKKKRKVGSKRPAATKKAGQAKKKKYPYDVPDYA	1278	GSKRPAATKKAGQAKKKKGGSKRPAATKKAGQAKKKK	1021
TLESPAAKRVKLDGGSPAAKRVKLDGGSPAAKRVKLDGGSPAAKRVKLDTLESKRPAATKKAGQAKKKKGGSKRPAATKKAGQAKKKKAPGEYPYDVPDYA	1279	GSKRPAATKKAGQAKKKKGGSKRPAATKKAGQAKKKK	1021

[0292] В некоторых вариантах осуществления вариант CasX содержит любую из SEQ ID NO: 247-592 или 1147-1231, или любую из SEQ ID NO: 270-592 или 1147-1231, или любую из SEQ ID NO: 415-592 или 1147-1231, или последовательность из таблицы 3, слитую с одним или более NLS с любой из SEQ ID NO: 63-75, 219-236, 239, 352-389, 983-1021, 1237-1278, или любой из последовательностей из таблицы 8. В некоторых вариантах осуществления один или более NLS слиты с N-концом варианта CasX или находятся рядом с ним. В некоторых вариантах осуществления один или более NLS слиты с C-концом варианта CasX или находятся рядом с ним. В некоторых вариантах осуществления один или более NLS слиты как с N-, так и с C-концом варианта CasX. В некоторых вариантах осуществления NLS связан с другим NLS с помощью линкера.

[0293] В некоторых случаях референсный или слитный вариант белка CasX включает в себя «домен трансдукции белка», или PTD (также известен как CPP - проникающий в клетку пептид), который относится к белку, полинуклеотиду, углеводу или органическому или неорганическому соединению, которое облегчает прохождение через липидный бислой, мицеллу, клеточную мембрану, мембрану органеллы или везикулярную мембрану. PTD, присоединенный к другой молекуле, которая может находиться в диапазоне от малой полярной молекулы до крупной макромолекулы и/или наночастицы, способствует прохождению молекулы через мембрану, например, из внеклеточного пространства во внутриклеточное пространство или из цитозоля внутрь органеллы. В некоторых вариантах осуществления PTD ковалентно связан с аминоконцом референсного или слитного варианта белка CasX. В некоторых вариантах осуществления PTD ковалентно

связан с карбоксильным концом референсного или слитного варианта белка CasX. В некоторых случаях PTD вставлен внутрь последовательности референсного или слитного варианта белка CasX в приемлемом для вставки месте. В некоторых случаях референсный или слитный вариант белка CasX включает в себя (конъюгирован с, слит с) один или более PTD (например, два или более, три или более, четыре или более PTD). В некоторых случаях PTD включает один или более сигналов ядерной локализации (NLS). Примеры PTD включают в себя, без ограничений, домен трансдукции пептида HIV TAT, содержащий YGRKKRRQRRR (SEQ ID NO: 390), RKKRRQRR (SEQ ID NO: 391); YARAAARQARA (SEQ ID NO: 392); THRLPRRRRRR (SEQ ID NO: 393) и GGRRARRRRRR (SEQ ID NO: 394); полиаргининовую последовательность, содержащую ряд аргининов, достаточный для прямого входа в клетку (например, 3, 4, 5, 6, 7, 8, 9, 10 или 10-50 аргининов, SEQ ID NO: 1026); домен VP22 (Zender et al. (2002) Cancer Gene Ther. 9(6):489-96); домен трансдукции белка Drosophila Antennapedia (Noguchi et al. (2003) Diabetes 52(7): 1732-1737); усеченный пептид кальцитонина человека (Trehin et al. (2004) Pharm. Research 21: 1248-1256); полилизин (Wender et al. (2000) Proc. Natl. Acad. Sci. USA 97: 13003-13008); RRQRRTSKLMKR (SEQ ID NO: 395); транспортан GWTLNSAGYLLGKINLKALAALAKKIL (SEQ ID NO: 396); KALAWEAKLAKALAKALAKHLAKALAKALKCEA (SEQ ID NO: 397) и RQIKIWFQNRRMKWKK (SEQ ID NO: 398). В некоторых вариантах осуществления PTD представляет собой активируемый CPP (ACPP) (Aguilera et al. (2009) Integr Biol (Camb) June; 1(5-6); 371-381). ACPP содержат поликатионный CPP (например, Arg9 или «R9»), соединенный с помощью расщепляемого линкера с соответствующим полианионом (например, Glu9 или «E9»), что снижает чистый заряд почти до нуля и таким образом ингибирует адгезию и поглощение в клетки. После расщепления линкера полианион высвобождается, локально демаскируя полиаргинин и присущую ему адгезионную способность, таким образом «активируя» ACPP для прохождения через мембрану. В некоторых вариантах осуществления вариант CasX содержит любую из SEQ ID NO: 247-592 или 1147-1231, или любую из SEQ ID NO: 270-592 или 1147-1231, или любую из SEQ ID NO: 415-592 или 1147-1231, или последовательность из таблицы 3, и PTD.

[0294] В некоторых вариантах осуществления слитный вариант белка CasX может включать в себя белок CasX, который связан с вставленной внутрь гетерологичной аминокислотой или гетерологичным полипептидом (гетерологичной аминокислотной последовательностью) с помощью линкерного полипептида (например, одного или более линкерных полипептидов). В некоторых вариантах осуществления референсный или слитный вариант белка CasX может быть связан на C-конце и/или N-конце с гетерологичным полипептидом (партнером по слиянию) с помощью линкерного полипептида (например, одного или более линкерных полипептидов). Линкерный полипептид может содержать любую из разнообразных аминокислотных последовательностей. Белки могут быть соединены спейсерным пептидом, как правило, гибким, хотя не исключаются другие химические связи. Приемлемые линкеры включают

полипептиды длиной от 4 аминокислот до 40 аминокислот или длиной от 4 аминокислот до 25 аминокислот. Эти линкеры по существу получают с применением синтетических кодирующих линкер олигонуклеотидов для соединения белков. Могут применяться пептидные линкеры с определенной степенью гибкости. Линкерные пептиды могут иметь практически любую аминокислотную последовательность, учитывая, что предпочтительные линкеры будут иметь последовательность, которая приводит к получению по существу гибкого пептида. При создании гибкого пептида полезно применять небольшие аминокислоты, такие как глицин и аланин. Создание таких последовательностей является обычным для специалистов в данной области. Различные линкеры доступны в продаже и считаются приемлемыми для применения. Примеры линкерных полипептидов включают в себя глициновые полимеры (G)n, глицин-сериновый полимер (включая, например, (GS)n (SEQ ID NO: 1024), (GSGGS)n (SEQ ID NO: 399), (GGSGGS)n (SEQ ID NO: 400) и (GGGS)n (SEQ ID NO: 401), где n представляет собой целое число, по меньшей мере единицу), глицин-аланиновые полимеры, аланин-сериновые полимеры, глицин-пролиновые полимеры, пролиновые полимеры и пролин-аланиновые полимеры. Примеры линкеров могут включать в себя аминокислотные последовательности, включая, без ограничений, RS, (G)n, (GS)n (SEQ ID NO: 1024), (GSGGS)n (SEQ ID NO: 399), (GGSGGS)n (SEQ ID NO: 400), (GGGS)n (SEQ ID NO: 401), GGSG (SEQ ID NO: 402), GGSGG (SEQ ID NO: 403), GSGSG (SEQ ID NO: 404), GSGGG (SEQ ID NO: 405), GGGSG (SEQ ID NO: 406), GSSSG (SEQ ID NO: 407), GPGP (SEQ ID NO:408), GGP, PPP, PPAPPA (SEQ ID NO: 409), PPPG (SEQ ID NO: 24), PPPGPPP (SEQ ID NO: 410), PPP(GGGS)n (SEQ ID NO: 25), (GGGS)nPPP (SEQ ID NO: 26), AEAAAKEAAAKEAAAKA (SEQ ID NO: 1025) и TPPKTKRKVEFE (SEQ ID NO: 27), где n равно от 1 до 5. Специалисту в данной области будет понятно, что конструкция пептида, конъюгированного с любыми описанными выше элементами, может включать линкеры, которые являются полностью или частично гибкими, так что линкер может включать гибкий линкер, а также один или более участков, которые придают ему менее гибкую структуру.

V. Способы получения вариантов белка CasX и вариантов гРНК

[0295] Варианты белка CasX и варианты гРНК, описанные в настоящем документе, могут быть сконструированы различными способами. Такие способы могут включать в себя, например, глубокую мутационную эволюцию (DME), описанную ниже и в разделе «Примеры», а также в заявках PCT/US20/36506 и WO2020247883A2, включенных в настоящий документ путем ссылки.

a. Глубокая мутационная эволюция (DME)

[0296] В некоторых вариантах осуществления DME используют для идентификации вариантов белка CasX и каркасов огРНК с улучшенной функцией. В некоторых вариантах осуществления способ DME включает в себя построение и тестирование полного набора мутаций исходной биомолекулы для получения библиотеки вариантов биомолекулы; например, библиотеки вариантов белка CasX или вариантов каркасов огРНК. DME может

охватывать получение всех возможных замен, а также всех возможных небольших вставок и всех возможных делеций аминокислот (в случае белков) или нуклеотидов (в случае РНК или ДНК) в исходной биомолекуле. Схема, иллюстрирующая способы DME, показана на ФИГ. 16. В некоторых вариантах осуществления DME содержит подмножество всех таких возможных замен, вставок и делеций. В определенных вариантах осуществления DME конструируют одну или более библиотек вариантов, оценивают их на функциональные изменения, и эту информацию используют для конструирования одной или более дополнительных библиотек. Такие итеративное конструирование и оценка вариантов могут приводить, например, к идентификации мутационных тем, приводящих к определенным функциональным результатам, таким как области белка или РНК, которые при введении определенных мутаций дают одну или более улучшенных функций. Наслаивание таких идентифицированных мутаций может дополнительно улучшать функцию, например, посредством аддитивных или синергетических взаимодействий. DME включает в себя составление библиотеки, конструирование библиотеки и скрининг библиотеки. В некоторых вариантах осуществления проводят множество раундов составления, конструирования и скрининга.

b. Составление библиотеки

[0297] Способы DME позволяют получать варианты биомолекул, которые представляют собой полимеры из многих мономеров. В некоторых вариантах осуществления биомолекула содержит молекулу белка или рибонуклеиновой кислоты (РНК), причем мономерные звенья представляют собой аминокислоты или рибонуклеотиды соответственно. Фундаментальные единицы мутации биомолекулы включают в себя любые из следующих: (1) замену одного мономера на другой мономера с другой идентичностью (замены); (2) вставку одного или более дополнительных мономеров в биомолекулу (вставки) или (3) удаление одного или более мономеров из биомолекулы (делеции). Библиотеки DME, содержащие замены, вставки и делеции, отдельно или в комбинации, в любом одном или более мономерах в пределах любой биомолекулы, описанной в настоящем документе, считаются входящими в объем изобретения.

[0298] В некоторых вариантах осуществления DME используют для построения и проверки полного набора мутаций в биомолекуле, включая все возможные замены, а также небольшие вставки и делеции аминокислот (в случае белков) или нуклеотидов (в случае РНК). Конструирование и функциональное считывание этих мутаций может быть достигнуто с помощью различных установленных способов молекулярной биологии. В некоторых вариантах осуществления библиотека содержит подмножество всех возможных модификаций в мономерах. Например, в некоторых вариантах осуществления библиотека в совокупности представляет единичную модификацию одного мономера для по меньшей мере 10% всех позиций мономеров в биомолекуле, причем каждая единичная модификация выбрана из группы, состоящей из замены, единичной вставки и единичной делеции. В некоторых вариантах осуществления библиотека в совокупности представляет одну модификацию одного мономера для по меньшей мере 5%, по меньшей мере 10%, по

меньшей мере 20%, по меньшей мере 30%, по меньшей мере 40%, по меньшей мере 50%, по меньшей мере 60%, по меньшей мере 70%, по меньшей мере 80%, по меньшей мере 90%, по меньшей мере 95% или до 100% всех позиций мономеров в исходной биомолекуле. В определенных вариантах осуществления для определенной процентной доли всех положений мономеров в исходной биомолекуле библиотека в совокупности представляет каждую возможную единичную модификацию одного мономера, такую как все возможные замены на 19 других природных аминокислот (для белка) или 3 других природных рибонуклеотидов (для РНК), вставку каждой из 20 природных аминокислот (для белка) или 4 природных рибонуклеотидов (для РНК), или делецию мономера. В других дополнительных вариантах осуществления вставка в каждом положении независимо составляет больше одного мономера, например, вставка двух или более, трех или более, или четырех или более мономеров, или вставка от одного до четырех, от двух до четырех или от одного до трех мономеров. В некоторых вариантах осуществления делеция в положении независимо составляет больше одного мономера, например, делеция двух или более, трех или более, или четырех или более мономеров, или делеция от одного до четырех, от двух до четырех или от одного до трех мономеров. Примеры таких библиотек вариантов CasX и вариантов гРНК описаны в примерах 14 и 15 соответственно.

[0299] В некоторых вариантах осуществления биомолекула представляет собой белок, а индивидуальные мономеры представляют собой аминокислоты. В тех вариантах осуществления, в которых биомолекула представляет собой белок, число возможных DME-мутаций в каждом положении мономера (аминокислоты) в белке составляет 19 аминокислотных замен, 20 аминокислотных вставок и 1 аминокислотную делецию, что дает в совокупности 40 возможных мутаций на аминокислоту в белке.

[0300] В некоторых вариантах осуществления DME-библиотека вариантов белка CasX, содержащая вставки, представляет собой библиотеку 1-аминокислотных вставок, библиотеку 2-аминокислотных вставок, библиотеку 3-аминокислотных вставок, библиотеку 4-аминокислотных вставок, библиотеку 5-аминокислотных вставок, библиотеку 6-аминокислотных вставок, библиотеку 7-аминокислотных вставок, библиотеку 8-аминокислотных вставок, библиотеку 9-аминокислотных вставок или библиотеку 10-аминокислотных вставок. В некоторых вариантах осуществления DME-библиотека вариантов белков CasX, содержащих вставки, содержит 1-4-аминокислотные вставки.

[0301] В некоторых вариантах осуществления биомолекула представляет собой РНК. В тех вариантах осуществления, в которых биомолекула представляет собой РНК, число возможных DME-мутаций в каждом положении мономера (рибонуклеотида) в РНК включает в себя 3 нуклеотидных замены, 4 нуклеотидных вставки и 1 нуклеотидную делецию, что дает в совокупности 8 возможных мутаций на нуклеотид.

[0302] В некоторых вариантах осуществления составление DME-библиотеки включает в себя подсчет всех возможных мутаций для каждого из одного или более целевых мономеров в биомолекуле. В настоящем документе термин «целевой мономер» относится

к мономеру в полимере биомолекулы, на который нацеливают DME, с заменами, вставками и делециями, описанными в настоящем документе. Например, целевой мономер может представлять собой аминокислоту в указанном положении в белке или нуклеотид в указанном положении в РНК. Биомолекула может иметь по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 100 или более целевых мономеров, которые подвергают систематическим мутациям для получения DME-библиотеки вариантов биомолекул. В некоторых вариантах осуществления каждый мономер в биомолекуле представляет собой целевой мономер. Например, при DME белка, в котором имеются две целевые аминокислоты, составление DME-библиотеки включает в себя подсчет 40 возможных DME-мутаций в каждой из двух целевых аминокислот. В дополнительном варианте осуществления при DME РНК, в которой имеются четыре целевых нуклеотида, составление DME-библиотеки включает в себя подсчет 8 возможных DME-мутаций для каждого из четырех целевых нуклеотидов. В некоторых вариантах осуществления каждый целевой мономер биомолекулы независимо выбирают случайным образом или выбирают по преднамеренной схеме. Таким образом, в некоторых вариантах осуществления DME-библиотека содержит случайные варианты или варианты, которые были разработаны, или варианты, содержащие случайные мутации и разработанные мутации в одной биомолекуле, или любые их комбинации.

[0303] В некоторых вариантах осуществления способов DME DME-мутации встраивают в двухцепочечную ДНК, кодирующую биомолекулу. Эту ДНК можно поддерживать и реплицировать в стандартном клонирующем векторе, например, бактериальной плазмиде, называемой в настоящем документе целевой плазмидой. Пример целевой плазмиды содержит ДНК-последовательность, кодирующую исходную биомолекулу, которая будет подвергаться DME, бактериальную точку начала репликации и приемлемую экспрессионную кассету резистентности к антибиотикам. В некоторых вариантах осуществления кассета резистентности к антибиотикам придает резистентность к канамицину, ампициллину, спектиномицину, блеомицину, стрептомицину, эритромицину, тетрациклину или хлорамфениколу. В некоторых вариантах осуществления кассета резистентности к антибиотикам придает резистентность к канамицину.

[0304] Библиотеку, содержащую указанные варианты, можно конструировать различными способами. В определенных вариантах осуществления для конструирования библиотеки используют рекомбинационный инжиниринг плазмиды. В таких способах можно использовать ДНК-олигонуклеотиды, кодирующие одну или более мутаций для встраивания указанных мутаций в плазмиду, кодирующую референсную биомолекулу. Для вариантов биомолекул со множеством мутаций в некоторых вариантах осуществления используют более одного олигонуклеотида. В некоторых вариантах осуществления ДНК-олигонуклеотиды кодируют одну или более мутаций, причем область мутации фланкирована 10-100 нуклеотидами, гомологичными целевой плазмиде, как с 5’-стороны, так и с 3’-стороны от мутации. Такие олигонуклеотиды в некоторых вариантах осуществления могут быть коммерчески синтезированы и использованы при ПЦР-

амплификации. Ниже представлен пример шаблона олигонуклеотида, кодирующего мутацию:

5 ’- $(N)_{10-100}$ - мутация - $(N')_{10-100}$-3’

[0305] В этой примерной схеме олигонуклеотида N представляют последовательности, идентичные целевой плазмиде и называемые в настоящем документе гомологичными плечами. При нацеливании на конкретный мономер в биомолекуле для его мутации эти гомологичные плечи непосредственно фланкируют ДНК, кодирующую мономер, в целевой плазмиде. В некоторых примерах осуществления, в которых биомолекула, подвергаемая DME, представляет собой белок, используют 40 разных олигонуклеотидов с одним и тем же набором гомологичных плеч для кодирования перечисленных 40 разных аминокислотных мутаций для каждого аминокислотного остатка в белке, который выбран для DME. Когда мутация затрагивает одну аминокислоту, область, кодирующая требуемую мутацию или мутации, содержит три нуклеотида, кодирующих аминокислоту (для замен или одиночных вставок) или нуль нуклеотидов (для делеций). В некоторых вариантах осуществления олигонуклеотид кодирует вставку более чем одной аминокислоты. В примере, в котором олигонуклеотид кодирует вставку из X аминокислот, область, кодирующая требуемую мутацию, содержит 3*X нуклеотидов, кодирующих X аминокислот. В некоторых вариантах осуществления область мутации кодирует более чем одну мутацию, например мутации двух или более мономеров биомолекулы, которые находятся в непосредственной близости (например, рядом друг с другом, или в пределах 1, 2, 3, 4, 5, 6, 7, 8, 9 или 10 или более мономеров друг от друга).

[0306] В некоторых примерах осуществления, в которых биомолекула, подвергаемая DME, представляет собой РНК, 8 разных олигонуклеотидов с использованием одного и того же набора гомологичных плеч кодируют 8 разных однонуклеотидных мутаций для каждого нуклеотида в РНК, которая выбрана для DME. Если мутация затрагивает один рибонуклеотид, область олигонуклеотида, кодирующая мутации, может состоять из следующих нуклеотидных последовательностей: один нуклеотид, указывающий нуклеотид (в случае замен или вставок) или нуль нуклеотидов (в случае делеций). В некоторых вариантах осуществления олигонуклеотиды синтезируют в виде одноцепочечных ДНК-олигонуклеотидов. В некоторых вариантах осуществления в пул объединяют все олигонуклеотиды, нацеленные на конкретную аминокислоту или нуклеотид биомолекулы, подвергаемой DME. В некоторых вариантах осуществления в пул объединяют все олигонуклеотиды, нацеленные на биомолекулу, подвергаемую DME. Нет никаких ограничений, накладываемых на тип или число мутаций, которые можно одновременно создавать в DME-библиотеке.

с. Скрининг библиотеки

[0307] В рамках изобретений предусмотрен любой подходящий способ скрининга или отбора в DME-библиотеке. Для оценки больших библиотек с тысячами индивидуальных мутаций можно использовать высокопроизводительные способы. В некоторых вариантах осуществления производительность скрининга библиотеки или

селекционного анализа имеет пропускную способность в миллионы индивидуальных ячеек. В некоторых вариантах осуществления предпочтительными являются анализы с использованием живых клеток, так как в живых клетках фенотип и генотип физически связаны по своей природе, поскольку находятся внутри одного в того же липидного бислоя. Живые клетки также можно использовать для прямой амплификации субпопуляций общей библиотеки. В других вариантах осуществления в способах DME используют анализы меньшего масштаба, например, для скрининга сфокусированной библиотеки, созданной посредством множества циклов мутации и оценки. Примеры способов скрининга библиотек описаны в примерах 14 и 15.

[0308] В некоторых вариантах осуществления DME-библиотеки, которые были подвергнуты скринингу или отбору на высокофункциональные варианты, подвергают дополнительной характеризации. В некоторых вариантах осуществления дополнительная характеризация DME-библиотеки включает в себя индивидуальный анализ DME-вариантов посредством секвенирования, например секвенирования по методу Сэнгера, для идентификации конкретной мутации или мутаций, которые привели к высокофункциональному варианту. Индивидуальные мутантные варианты биомолекулы могут быть выделены с помощью стандартных методик молекулярной биологии для последующего анализа функции. В некоторых вариантах осуществления дополнительная характеризация DME-библиотеки включает в себя высокопроизводительное секвенирование как библиотеки I, так и одной или более библиотек высокофункциональных вариантов. Этот подход в некоторых вариантах осуществления может обеспечивать быструю идентификацию мутаций, которые сверхпредставлены в одной или более библиотеках высокофункциональных вариантов по сравнению с наивной DME-библиотекой. Без ограничений, накладываемых какой-либо теорией, мутации, которые сверхпредставлены в одной или более библиотеках высокофункциональных вариантов, вероятнее всего, будут ответственны за активность высокофункциональных вариантов. В некоторых вариантах осуществления дополнительная характеризация DME-библиотеки включает в себя как секвенирование индивидуальных вариантов, так и высокопроизводительное секвенирование как наивной библиотеки, так и одной или более библиотек высокофункциональных вариантов.

[0309] Высокопроизводительное секвенирование может обеспечивать с высокой пропускной способностью данные, показывающие функциональный эффект элементов библиотеки. В вариантах осуществления, в которых одна или более библиотек представляют любую возможную мутацию каждой позиции мономера, такое высокопроизводительное секвенирование позволяет оценивать функциональный эффект каждой возможной DME-мутации. Такое секвенирование также можно использовать для оценки одной или более высокофункциональных субпопуляций данной библиотеки, что в некоторых вариантах осуществления может приводить к идентификации мутаций, которые приводят к получению улучшенной функции. Глубокое мутационное сканирование

[0310] В некоторых вариантах осуществления для идентификации вариантов белка

CasX с улучшенной функцией используется глубокое мутационное сканирование (DMS). При глубоком мутационном сканировании оценивают пластичность белка в той мере, в которой она относится к функции. В способах DMS каждую аминокислоту белка заменяют на каждую вторую аминокислоту и оценивают абсолютную функцию белка. Например, каждую аминокислоту в белке CasX можно заменить на каждую вторую аминокислоту, и мутированные белки CasX можно проанализировать на способность связываться с ДНК или расщеплять ДНК. Примеры анализов, таких как анализ CRISPRi или анализы расщепления на основе бактерий, которые можно использовать для характеризации наборов DMS-вариантов белка CasX, описаны в Oakes et al. (2016) «Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch» Nat Biotechnol 34(6):646-51 и Liu et al. (2019) «CasX enzymes comprise a distinct family of RNA-guided genome editors» Nature doi.org/10.1038/s41586-019-0908; содержание которых включено в настоящий документ путем ссылки.

[0311] В некоторых вариантах осуществления DMS используется для идентификации белков CasX с улучшенной активностью связывания с ДНК. В некоторых вариантах осуществления активность связывания с ДНК анализируют с использованием анализа CRISPRi. В не имеющем ограничительного характера примере осуществления анализа CRISPRi клетки, экспрессирующие флуоресцентный белок, такой как зеленый флуоресцентный белок (GFP) или красный флуоресцентный белок (RFP), анализируют с использованием FACS для идентификации вариантов CasX, способных подавлять экспрессию флуоресцентного белка в зависимости от огРНК. В этом примере каталитически неактивный белок CasX (dCasX) используют для получения набора DMS-мутантов для анализа. Белок CasX дикого типа связывается со своей узнаваемой огРНК и образует комплекс белок-РНК. Комплекс связывается с конкретными ДНК-мишенями путем спаривания оснований Уотсона - Крика между огРНК и ДНК-мишенью, в данном случае - ДНК-последовательностью, кодирующей флуоресцентный белок. В случае CasX дикого типа ДНК будет расщепляться из-за нуклеазной активности белка CasX. Однако без ограничений, накладываемых какой-либо теорией, вероятно, что dCasX все же способен образовывать комплекс с огРНК и связываться со специфической ДНК-мишенью. Когда происходит нацеливание dCasX на область кодирования белка, dCasX блокирует РНК-полимеразу II и инициацию создания и/или удлинения транскрипта, что приводит к снижению экспрессии флуоресцентного белка, которое может быть обнаружено с помощью FACS.

[0312] В некоторых вариантах осуществления DMS используется для идентификации белков CasX с улучшенной активностью расщепления ДНК. Способы анализа эффективности расщепления ДНК вариантами белка CasX будут очевидны специалисту в данной области. Например, белки CasX в комплексе с огРНК со спейсером, комплементарным конкретной последовательности целевой нуклеиновой кислоты, можно использовать для расщепления целевой последовательности ДНК *in vitro* или in vivo в клетках приемлемого типа, и анализируют частоту вставок и делеций в сайте расщепления.

Без ограничений, накладываемых какой-либо теорией, расщепление или никирование посредством CasX генерирует двухцепочечные разрывы в ДНК, последующая репарация которых с помощью негомологичного соединения концов (NHEJ) приводит к образованию небольших вставок или делеций (инделов) в местах двухцепочечных разрывов. Частоту инделов в сайте расщепления CasX можно измерять с использованием высокопроизводительного секвенирования или секвенирования целевой последовательности по методу Сэнгера. Альтернативно или в дополнение частота образования инделов посредством расщепления CasX целевой последовательности может быть измерена с использованием анализов несоответствий, таких как анализы несовпадений с использованием T7 эндонуклеазы I (T7EI) или нуклеазы Surveyor.

[0313] В некоторых вариантах осуществления после DMS получают карту генотипов DMS-мутантов в соединении с итоговым фенотипом (например, тепловую карту) и используют ее для определения основных принципов белка. Все возможные мутации характеризуют на основе того, приводят они к функциональным или нефункциональным белковым продуктам, создавая функциональный профиль данного белка.

d. ПЦР сниженной точности

[0314] В некоторых вариантах осуществления ПЦР сниженной точности используют для создания вариантов белка CasX или каркасов огРНК с улучшенной функцией. Полимеразы, которые реплицируют ДНК, имеют разные уровни точности. Одним из способов введения случайных мутаций в ген является полимераза сниженной точности, которая встраивает неправильные нуклеотиды в диапазоне частотностей. Такую частотность можно модулировать в зависимости от требуемого результата. В некоторых вариантах осуществления выбирают полимеразу и условия полимеразной активности, которые приводят к частоте изменений нуклеотидов, обеспечивающей среднее n, равное 1-4 аминокислотным изменениям в последовательности белка. Типичная полимераза сниженной точности входит в набор Agilent GeneMorphII. Набор GeneMorphII можно использовать для амплификации ДНК-последовательности, кодирующей белок CasX дикого типа (например, белок SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3) в соответствии с протоколом производителя, подвергая таким образом белок бессистемному случайному мутагенезу и создавая разнообразную популяцию вариантов белка CasX. Затем эту разнообразную популяцию вариантов белка CasX можно анализировать с использованием анализов, описанных выше для DMS, наблюдая, как изменения в генотипе соотносятся с изменениями фенотипа.

e. Кассетный мутагенез

[0315] В некоторых вариантах осуществления для получения варианта белка CasX или вариантов каркаса огРНК с улучшенной функцией используют кассетный мутагенез. В кассетном мутагенезе используют уникальные сайты рестрикционных ферментов, которые заменяют вырожденными нуклеотидами, создавая малые области высокого разнообразия в выбранных областях интересующего гена, такого как белок CasX или каркас огРНК. В примере протокола кассетного мутагенеза с помощью рестрикционных ферментов

выполняют расщепление рядом с последовательностью, выбранной для мутагенеза, на молекуле ДНК, кодирующей белок CasX или каркас огРНК, содержащейся в приемлемом векторе. На этой стадии удаляют последовательность, выбранную для мутагенеза, и все, что находится между сайтами рестрикции. Затем вместо последовательности, удаленной посредством рестрикции, лигируют синтетические двухцепочечные молекулы ДНК, содержащие требуемую мутацию и концы, комплементарные концам рестрикции, и трансформируют лигированным вектором приемлемые клетки, такие как *E. coli*. В некоторых вариантах осуществления кассетный мутагенез можно использовать для получения одной или более конкретных мутаций в белке CasX или каркасе огРНК. В некоторых вариантах осуществления кассетный мутагенез можно использовать для получения библиотеки вариантов белка CasX или вариантов каркасов огРНК, которые можно повергнуть скринингу или отбору на улучшенную функцию с использованием способов, описанных в настоящем документе. Например, при использовании кассетного мутагенеза для получения вариантов CasX части домена, связывающегося с нецелевой нитью (NTSB), можно заменить последовательностью вырожденных нуклеотидов. Последовательности вырожденных нуклеотидов могут быть преимущественно локализованы в областях белка CasX, например, в областях NTSB, представляющих интерес из-за их высокомобильных элементов или прямых контактов с ДНК. Затем библиотеки вариантов белков CasX, полученные посредством кассетного мутагенеза, могут быть подвергнуты скринингу с использованием анализов, описанных в настоящем документе применительно к DME, DMS и ПЦР сниженной точности, и варианты можно подвергнуть отбору на улучшенную функцию.

f. Случайный мутагенез

[0316] В некоторых вариантах осуществления для получения вариантов белка CasX или вариантов каркасов огРНК с улучшенной функцией используют случайный мутагенез. Случайный мутагенез является бессистемным способом изменения ДНК. Примеры способов случайного мутагенеза известны специалисту в данной области и включают в себя воздействие химических веществ, УФ-излучения, рентгеновского излучения или использование нестабильных клеточных линий. Разные мутагенные агенты дают разные типы мутаций, и обычный специалист сможет выбрать подходящий агент для получения мутаций требуемого типа. Например, этилметансульфонат (EMS) и N-этил-N-нитрозомочевину (ENU) можно использовать для создания единичных изменений пар оснований, в то время как рентгеновское излучение часто приводит к делециям и макроскопическим хромосомным перестройкам. Воздействие УФ-излучения приводит к образованию димеров между соседними пиримидинами в ДНК, что может приводить к точечным мутациям, делециям и перестановкам. Для введения мутаций, например, в плазмиду, содержащую белок CasX или каркас огРНК, также можно использовать клеточные линии сниженной точности. Популяцию молекул ДНК, кодирующих белок CasX (например, белок SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3), или каркас огРНК можно подвергать воздействию мутагена для получения набора вариантов белка CasX или

вариантов каркасов огРНК, и эти наборы могут быть проанализированы на улучшенную функцию с использованием любого из анализов, описанных в настоящем документе.

g. Способ ступенчатой достройки (StEP)

[0317] В некоторых вариантах осуществления для создания вариантов белка CasX или вариантов каркасов огРНК с улучшенной функцией используют способ ступенчатой достройки (StEP). Способ ступенчатой достройки представляет собой специализированный протокол ПЦР, который позволяет размножить множество вариантов белка во время ПЦР-реакции. Во время StEP используют полимеразу с низкой процессивностью (например, полимеразу Taq или Vent) для создания коротких праймеров из двух или более разных матричных цепей со значительным уровнем сходства последовательностей. Затем короткие праймеры удлиняют в течение коротких временных интервалов, что позволяет перетасовать матричные цепи. Этот способ также можно использовать в качестве средства стекинга DME-вариантов. Примеры протоколов StEP описаны в Zhao, H. et al. (1998) «Molecular evolution by staggered extension process (StEP) *in vitro* recombination» Nature Biotechnology 16: 258-261, содержание которой полностью включено в настоящий документ путем ссылки. StEP можно использовать для создания коллекций вариантов белка CasX или вариантов каркасов огРНК, и эти коллекции могут быть проанализированы на улучшенную функцию с использованием любого из анализов, описанных в настоящем документе.

h. Перестановка генов

[0318] В некоторых вариантах осуществления для получения вариантов белка CasX или вариантов каркасов огРНК с улучшенной функцией используют перестановку генов. В некоторых вариантах осуществления перестановка генов используется для комбинирования (иногда называемого в настоящем документе «стекинг») вариантов, полученных другими способами, описанными в настоящем документе, такими как рекомбинационный инжиниринг плазмиды. В примере протокола перетасовки генов используют ДНКазу, например ДНКазу I, для сдвига позиции набора исходных генов по длине на 50-100 пар оснований (п.о.). В некоторых вариантах осуществления эти исходные гены содержат варианты белка CasX с улучшенной функцией, созданные и выделенные с использованием способов, описанных в настоящем документе. В некоторых вариантах осуществления эти исходные гены содержат варианты каркаса огРНК с улучшенной функцией, созданные и выделенные с использованием способов, описанных в настоящем документе. После ДНКазной фрагментации следует полимеразная цепная реакция (ПЦР) без праймеров. Фрагменты ДНК с достаточно перекрывающейся гомологичной последовательностью будут отжигаться друг с другом, а затем удлиняться ДНК-полимеразой. Если отожжены разные фрагменты, содержащие разные мутации, результатом будет новый вариант, объединяющий эти две мутации. В некоторых вариантах осуществления после ПЦР без праймеров следует ПЦР-удлинение и очистка переставленных молекул ДНК, которые достигли размера исходных генов (например, последовательности, кодирующей белок CasX или каркас огРНК). Затем эти гены можно амплифицировать в другой ПЦР, например, посредством добавления праймеров для ПЦР,

комплементарных 5’- и 3’-концам гена, подвергающегося перестановке. В некоторых вариантах осуществления праймеры могут иметь дополнительные последовательности, добавленные к их 5’-концам, такие как последовательности сайтов распознавания рестрикционными ферментами, необходимые для лигирования в клонирующий вектор.

i. Обмен доменами

[0319] В некоторых вариантах осуществления для получения вариантов белка CasX или вариантов каркасов огРНК с улучшенной функцией используют обмен доменами. Для получения вариантов белков CasX в целях смешивания и сопоставления частей с другими белками и молекулами CRISPR можно использовать обмен сконструированными доменами. Например, белки CRISPR имеют консервативные домены RuvC, поэтому домен CasX RuvC можно заменить на домен других белков CRISPR, а полученный белок проанализировать на улучшенное расщепление ДНК с использованием анализов, описанных в настоящем документе. В случае огРНК каркасный стебель, удлиненный стебель или петли можно заменять структурами, обнаруженными в других РНК, например, каркасный стебель и удлиненный стебель огРНК можно обменивать на термостабильные стеблевые петли из других РНК, а полученный вариант анализировать на улучшенную функцию с использованием анализов, описанных в настоящем документе. В некоторых вариантах осуществления обмен доменами можно использовать для вставки новых доменов в белок CasX или огРНК. В некоторых примерах осуществления, где обмен доменами применяется к белку, вставленный домен содержит весь второй белок.

j. Получение вариантов CasX и гРНК

[0320] Вариант белка CasX настоящего описания может продуцироваться *in vitro* эукариотическими клетками или прокариотическими клетками, трансформированными кодирующими векторами (описаны ниже) с использованием стандартных методик клонирования и молекулярной биологии или так, как описано в примерах. Конкретная последовательность и способ получения будут определяться удобством, экономичностью, требуемой чистотой и т. п. В некоторых вариантах осуществления сначала получают конструкт, содержащий ДНК-последовательность, кодирующую вариант CasX. Примеры способов получения таких конструктов описаны в разделе «Примеры». В некоторых вариантах осуществления нуклеотидная последовательность, кодирующая белок CasX, кодон-оптимизирована для предполагаемой клетки-хозяина. Затем конструкт применяют для создания экспрессионного вектора, приемлемого для трансформации клетки-хозяина, такой как прокариотическая или эукариотическая клетка-хозяин, для экспрессии и извлечения белка. При необходимости клетка-хозяин представляет собой *E. coli*. В других вариантах осуществления клетка-хозяин представляет собой эукариотическую клетку. Эукариотическая клетка-хозяин может быть выбрана из клеток-фибробластов почки новорожденного хомяка (ВНК), клеток эмбриональной почки человека 293 (HEK293), клеток эмбриональной почки человека 293T (HEK293T), клеток NS0, клеток SP2/0, клеток миеломы YO, мышиных клеток миеломы P3X63, клеток PER, клеток PER.C6, клеток гибридомы, клеток NIH3T3, клеток, происходящих от CV-1 (обезьяньих), с генетическим

материалом SV40 (COS), клеток HeLa, клеток яичника китайского хомячка (CHO), или дрожжевых клеток, или других эукариотических клеток, известных в данной области, приемлемых для получения рекомбинантных продуктов.

[0321] При необходимости в последовательность во время синтеза или во время экспрессии могут быть введены различные группы, которые позволяют связываться с другими молекулами или с поверхностью. Таким образом, можно использовать цистеины для получения тиоэфиров, гистидины для связывания с комплексом иона металла, карбоксильные группы для формирования амидов или сложных эфиров, аминогруппы для формирования амидов и т. п. Вариант белка CasX описания также можно выделять и очищать в соответствии с традиционными способами рекомбинантного синтеза. Лизат можно получать из экспрессирующего хозяина и очищать с использованием высокоэффективной жидкостной хроматографии (ВЭЖХ), эксклюзионной хроматографии, гель-электрофореза, аффинной хроматографии или другой методики очистки. В большинстве случаев используемые композиции будут содержать 80% или более по массе требуемого продукта, более типично 90% или более по массе, предпочтительно 95% или более по массе, и для терапевтических целей обычно 99,5% или более по массе относительно загрязняющих веществ, связанных со способом получения продукта и его очисткой.

[0322] В случае получения гРНК настоящего описания рекомбинантные векторы экспрессии, кодирующие гРНК, можно транскрибировать *in vitro*, например, с использованием промоторных регуляторных последовательностей T7 и полимеразы T7 для получения гРНК, которую затем можно извлекать традиционными способами; например, очисткой посредством гель-электрофореза, как описано в примерах. После синтеза гРНК можно использовать в паре для редактирования генов для непосредственного контакта с целевой нуклеиновой кислотой или ее можно вводить в клетку любым из хорошо известных способов введения нуклеиновых кислот в клетки (например, микроинъекция, электропорация, трансфекция и т. д.).

VI. Полинуклеотиды и векторы

[0323] В другом аспекте настоящее описание относится к полинуклеотидам, кодирующим нуклеазы класса 2, типа V и гРНК, которые можно использовать при редактировании целевой нуклеиновой кислоты в клетке. В некоторых вариантах осуществления в описании предложены полинуклеотиды, кодирующие белки CasX и полинуклеотиды гРНК по любому из вариантов осуществления системы CasX : гРНК, описанных в настоящем документе.

[0324] В некоторых вариантах осуществления в описании предложена полинуклеотидная последовательность, кодирующая варианты CasX по любому из вариантов осуществления, описанных в настоящем документе, включая варианты белка CasX с SEQ ID NO: 247-592 или 1147-1231, описанные в таблице 3, или последовательности, имеющие идентичность последовательности по меньшей мере примерно 50%, по меньшей мере примерно 60%, по меньшей мере примерно 70%, по меньшей мере примерно 80%, по

меньшей мере примерно 90%, по меньшей мере примерно 95%, по меньшей мере примерно 96%, по меньшей мере примерно 97%, по меньшей мере примерно 98%, по меньшей мере примерно 99% с последовательностью SEQ ID NO: 247-592 и 1147-1231 из таблицы 3. В некоторых вариантах осуществления в описании предложена полинуклеотидная последовательность, кодирующая вариант CasX, с любым из SEQ ID NO: 270-592 или 1147-1231, или последовательность, имеющая идентичность с ней по меньшей мере примерно 50%, по меньшей мере примерно 60%, по меньшей мере примерно 70%, по меньшей мере примерно 80%, по меньшей мере примерно 90%, по меньшей мере примерно 95%, по меньшей мере примерно 96%, по меньшей мере примерно 97%, по меньшей мере примерно 98% или по меньшей мере примерно 99%. В некоторых вариантах осуществления в описании предложена полинуклеотидная последовательность, кодирующая вариант CasX, с любым из SEQ ID NO: 415-592 и 1147-1231, или последовательность, идентичную последовательности с ней по меньшей мере примерно 50%, по меньшей мере примерно 60%, по меньшей мере примерно 70%, по меньшей мере примерно 80% или по меньшей мере примерно 90%, или по меньшей мере примерно 95%, или по меньшей мере примерно 96%, или по меньшей мере примерно 97%, или по меньшей мере примерно 98%, или по меньшей мере примерно 99%. В некоторых вариантах осуществления в описании предложена выделенная полинуклеотидная последовательность, кодирующая последовательность варианта гРНК по любому из вариантов осуществления, описанных в настоящем документе, включая последовательности SEQ ID NO: 2101-2332 и 2353-2398 из таблицы 2 вместе с нацеливающими последовательностями, способными гибридизоваться с целевой нуклеиновой кислотой, подлежащей модификации. В некоторых вариантах осуществления в описании предложена выделенная полинуклеотидная последовательность, кодирующая вариант последовательности гРНК с SEQ ID NO: 2238-2332 или 2353-2398 вместе с нацеливающими последовательностями, способными гибридизоваться с целевой нуклеиновой кислотой, подлежащей модификации. В некоторых вариантах осуществления в описании предложена выделенная полинуклеотидная последовательность, кодирующая вариант последовательности гРНК с SEQ ID NO: 2281-2332 и 2353-2398 вместе с нацеливающими последовательностями, способными гибридизоваться с целевой нуклеиновой кислотой, подлежащей модификации.

[0325] В некоторых вариантах осуществления в описании предложены донорные матричные полинуклеотиды, кодирующие части или весь ген, подлежащий модификации. В некоторых вариантах осуществления донорная матрица предназначена для редактирования генов в сочетании с системой CasX : гРНК и содержит по меньшей мере часть гена, подлежащего модификации. В других вариантах осуществления донорная последовательность содержит последовательность, которая кодирует по меньшей мере часть экзона гена, подлежащего модификации. В других вариантах осуществления донорная матрица содержит последовательность, которая кодирует по меньшей мере часть интрона гена, подлежащего модификации. В других вариантах осуществления донорная матрица содержит последовательность, которая кодирует по меньшей мере часть области

соединения интрон-экзон гена, подлежащего модификации. В других вариантах осуществления донорная матрица содержит последовательность, которая кодирует по меньшей мере часть межгенной области гена, подлежащего модификации. В других вариантах осуществления донорная матрица содержит последовательность, которая кодирует по меньшей мере часть регуляторного элемента гена, подлежащего модификации. В некоторых случаях донорная матрица представляет собой последовательность дикого типа, которая кодирует по меньшей мере часть гена, подлежащего модификации. В других случаях последовательность донорной матрицы содержит одну или более мутаций относительно гена дикого типа, подлежащего нокдауну или нокауту. В таких случаях донорная матрица будет иметь по меньшей мере от 1 до 5 или более мутаций относительно последовательности дикого типа. В вышеуказанных вариантах осуществления донорная матрица содержит по меньшей мере 10 нуклеотидов, по меньшей мере 100 нуклеотидов, по меньшей мере 200 нуклеотидов, по меньшей мере 300 нуклеотидов, по меньшей мере 400 нуклеотидов, по меньшей мере 500 нуклеотидов, по меньшей мере 600 нуклеотидов, по меньшей мере 700 нуклеотидов, по меньшей мере 800 нуклеотидов, по меньшей мере 900 нуклеотидов, по меньшей мере 1000 нуклеотидов, по меньшей мере 2000 нуклеотидов, по меньшей мере 3000 нуклеотидов, по меньшей мере 4000 нуклеотидов, по меньшей мере 5000 нуклеотидов, по меньшей мере 6000 нуклеотидов, по меньшей мере 7000 нуклеотидов, по меньшей мере 8000 нуклеотидов, по меньшей мере 9000 нуклеотидов, по меньшей мере 10 000 нуклеотидов, по меньшей мере 12 000 нуклеотидов или по меньшей мере 15 000 нуклеотидов. В некоторых вариантах осуществления донорная матрица содержит от по меньшей мере примерно 10 до примерно 15 000 нуклеотидов. В некоторых вариантах осуществления донорная матрица представляет собой одноцепочечную ДНК-матрицу. В других вариантах осуществления донорная матрица представляет собой одноцепочечную РНК-матрицу. В других вариантах осуществления донорная матрица представляет собой двухцепочечную ДНК-матрицу. В некоторых вариантах осуществления донорная матрица может быть представлена в виде оголенной нуклеиновой кислоты в системах для редактирования гена, и ее не нужно встраивать в вектор. В других вариантах осуществления донорная матрица может быть встроена в вектор для облегчения ее доставки в клетку; например, в вирусный вектор.

[0326] В других аспектах описание относится к способам получения полинуклеотидных последовательностей, кодирующих варианты CasX или гРНК по любому из вариантов осуществления, описанных в настоящем документе, включая их гомологичные варианты, а также способам экспрессии белков, в которых с полинуклеотидных последовательностей экспрессируются белки или транскрибируется РНК. В целом, способы включают получение полинуклеотидной последовательности, кодирующей варианты CasX или гРНК по любому из вариантов осуществления, описанных в настоящем документе, и встраивание кодирующего гена в экспрессионный вектор, подходящий для клетки-хозяина. Для получения полинуклеотидов и экспрессионных векторов настоящего описания можно применять стандартные рекомбинантные методики

молекулярной биологии. Для получения кодируемого референсного CasX, вариантов CasX или гРНК по любому из вариантов осуществления, описанных в настоящем документе, способы включают трансформирование соответствующей клетки-хозяина экспрессионным вектором, содержащим кодирующий полинуклеотид, и культивирование клетки-хозяина в условиях, приводящих или способствующих получению референсного CasX, вариантов CasX или гРНК по любому из вариантов осуществления, описанных в настоящем документе, подлежащих экспрессированию или транскрибированию в трансформированной клетке-хозяине, с получением таким образом вариантов CasX или гРНК, которые извлекают способами, описанными в настоящем документе, или с помощью стандартных способов очистки, известных в данной области или описанных в разделе «Примеры».

[0327] В соответствии с описанием последовательности нуклеиновых кислот, которые кодируют варианты CasX или гРНК по любому из вариантов осуществления, описанных в настоящем документе (или их комплемент), применяются для создания рекомбинантных молекул ДНК, которые направляют экспрессию в соответствующих клетках-хозяевах. Несколько стратегий клонирования подходят для реализации настоящего описания, многие из которых применяются для создания конструкта, содержащего ген, кодирующий композицию настоящего описания, или его комплемент. В некоторых вариантах осуществления стратегия клонирования применяется для создания гена, который кодирует конструкт, содержащий нуклеотиды, кодирующие варианты CasX, или гРНК, которая применяется для трансформации клетки-хозяина и экспрессии композиции.

[0328] В некоторых подходах сначала получают конструкт, содержащий ДНК-последовательность, кодирующую вариант CasX или гРНК. Примеры способов получения таких конструктов описаны в разделе «Примеры». Затем конструкт применяют для создания экспрессионного вектора, приемлемого для трансформации клетки-хозяина, такой как прокариотическая или эукариотическая клетка-хозяин, для экспрессии и извлечения белкового конструкта, в случае CasX или гРНК. При необходимости клетка-хозяин представляет собой *E. coli*. В других вариантах осуществления клетка-хозяин представляет собой эукариотическую клетку. Эукариотическая клетка-хозяин может быть выбрана из клеток-фибробластов почки новорожденного хомяка (BHK), клеток эмбриональной почки человека 293 (HEK293), клеток эмбриональной почки человека 293T (HEK293T), клеток NS0, клеток SP2/0, клеток миеломы YO, мышиных клеток миеломы P3X63, клеток PER, клеток PER.C6, клеток гибридомы, клеток NIH3T3, клеток, происходящих от CV-1 (обезьяньих), с генетическим материалом SV40 (COS), клеток HeLa, клеток яичника китайского хомячка (CHO), или дрожжевых клеток, или других эукариотических клеток, известных в данной области, приемлемых для получения рекомбинантных продуктов. Примеры способов создания экспрессионных векторов, трансформации клеток-хозяев и экспрессии и извлечения вариантов CasX или гРНК описаны в разделе «Примеры».

[0329] Ген, кодирующий вариант CasX или конструкт гРНК, может быть получен в одну или более стадий, либо полностью синтетически, либо посредством синтеза в

комбинации с ферментативными процессами, такими как опосредованное рестрикционным ферментом клонирование, ПЦР с перекрывающимися праймерами, включая способы, более подробно описанные в разделе «Примеры». Способы, описанные в настоящем документе, можно применять, например, для лигирования к последовательностям полинуклеотидов, кодирующих гены различных компонентов (например, CasX и гРНК) требуемой последовательности. Гены, кодирующие полипептидные композиции, собирают из олигонуклеотидов с применением стандартных методик синтеза генов.

[0330] В некоторых вариантах осуществления нуклеотидная последовательность, кодирующая белок CasX, кодон-оптимизирована для предполагаемой клетки-хозяина. Этот тип оптимизации может предполагать мутацию кодирующей нуклеотидной последовательности для имитации предпочтительных кодонов предполагаемого организма или клетки-хозяина при кодировании того же белка CasX. Таким образом, кодоны можно изменять, но кодируемый белок или гРНК остается неизменным. Например, если предполагаемая клетка-мишень для белка CasX представляла собой человеческую клетку, то можно было применять человеческую кодон-оптимизированную кодирующую CasX нуклеотидную последовательность. В качестве другого не имеющего ограничительного характера примера, если предполагаемая клетка-хозяин представляла собой клетку мыши, то можно было создать мышиную кодон-оптимизированную кодирующую CasX нуклеотидную последовательность. Конструирование гена может быть выполнено с применением алгоритмов, которые оптимизируют применение кодонов и аминокислотный состав, делая их подходящими для клетки-хозяина, используемой при получении референсного CasX или вариантов CasX. В одном способе описания создают и затем собирают библиотеку полинуклеотидов, кодирующих компоненты конструктов, описанных выше. Затем полученные гены собирают и готовые гены применяют для трансформации клетки-хозяина, продуцируют и извлекают варианты CasX или композиции гРНК для оценки их свойств, как описано в настоящем документе.

[0331] Описание предусматривает применение плазмидных экспрессионных векторов, содержащих репликационные и управляющие последовательности, которые совместимы с, распознаются клеткой-хозяином и функционально связаны с геном, кодирующим полипептид, для управляемой экспрессии полипептида или транскрипции РНК. Такие последовательности векторов хорошо известны для разнообразных бактерий, дрожжей и вирусов. Используемые экспрессионные векторы, подходящие для применения, включают в себя, например, сегменты хромосомных, нехромосомных и синтетических ДНК-последовательностей. Термин «экспрессионный вектор» относится к конструкту ДНК, содержащему ДНК-последовательность, которая функционально связана с приемлемой управляющей последовательностью, и способна осуществлять экспрессию ДНК, кодирующей полипептид, в приемлемом хозяине. Требования заключаются в том, чтобы векторы были способны к репликации и жизнеспособны в выбранной клетке-хозяине. При необходимости можно применять низкокопийные или высококопийные векторы. Управляющие последовательности вектора включают промотор для

осуществления транскрипции, необязательную операторную последовательность для контроля такой транскрипции, последовательность, кодирующую приемлемые сайты связывания мРНК с рибосомой и последовательности, которые контролируют прекращение транскрипции и трансляции. В некоторых вариантах осуществления нуклеотидная последовательность, кодирующая гРНК, функционально связана с управляющим элементом, например, элементом управления транскрипцией, таким как промотор. В некоторых вариантах осуществления нуклеотидная последовательность, кодирующая белок CasX, функционально связана с управляющим элементом, например, элементом управления транскрипцией, таким как промотор. В других случаях нуклеотидная последовательность, кодирующая CasX и гРНК, связана и функционально связана с одним элементом управления. Промотор может представлять собой любую ДНК-последовательность, которая демонстрирует транскрипционную активность в выбранной клетке-хозяине и может быть получена из генов, кодирующих белки, либо гомологичные, либо гетерологичные для клетки-хозяина. Примеры регуляторных элементов включают промотор транскрипции, элемент-энхансер транскрипции, сигнал терминации транскрипции, внутренний сайт входа рибосомы (IRES) или пептид P2A для обеспечения трансляции множества генов с одного транскрипта, последовательности полиаденилирования для активации последующей терминации транскрипции, последовательности для оптимизации инициации трансляции и последовательности терминации трансляции. В некоторых случаях промотор представляет собой конститутивно активный промотор. В некоторых случаях промотор представляет собой регулируемый промотор. В одном варианте осуществления промотор представляет собой индуцибельный промотор. В некоторых случаях промотор представляет собой тканеспецифичный промотор. В некоторых случаях промотор представляет собой специфичный для типа клеток промотор. В некоторых случаях элемент управления транскрипцией (например, промотор) является функциональным в типе клеток-мишеней или популяции клеток-мишеней. Например, в некоторых случаях элемент управления транскрипцией может быть функциональным в эукариотических клетках, например, пакующих клетках для вирусных или XDP-векторов, гематопоэтических стволовых клетках (HSC), гематопоэтических клетках-предшественниках (HPC), клетках CD34+, мезенхимальных стволовых клетках (MSC), эмбриональных стволовых (ES) клетках, индуцированных плюрипотентных стволовых клетках (iPSC), общих миелоидных клетках-предшественниках, клетках-проэритробластах и клетках-эритробластах.

[0332] Не имеющие ограничительного характера примеры промоторов pol II включают в себя, без ограничений, EF-1альфа, основной промотор EF-1альфа, Jens Tornoe (JeT), промоторы из цитомегаловируса (CMV), немедленный-ранний промотор CMV (CMVIE), энхансер CMV, тимидинкиназу вируса простого герпеса (HSV), ранний и поздний промотор обезьяньего вируса 40 (SV40), энхансер SV40, промотор с длинными концевыми повторами (LTR) ретровируса, промотор металлотионеина-I мыши, главный поздний промотор аденовируса (Ad MLP), полноразмерный промотор CMV, минимальный

промотор CMV, промотор Œ≤-актина цыпленка (CBA), гибридный CBA (CBh), промотор Œ≤-актина цыпленка с энхансером цитомегаловируса (CB7), слитый промотор бета-актина цыпленка и акцепторный сайт сплайса бета-глобина кролика (CAG), промотор вируса саркомы Рауса (RSV), промотор ВИЧ-Ltr, промотор hPGK, промотор TK HSV, промотор 7SK, промотор Mini-TK, промотор синапсина I человека (SYN), который придает нейрон-специфическую экспрессию, промотор бета-актина, суперкоровый промотор 1 (SCP1), промотор Mecp2 для селективной экспрессии в нейронах, минимальный промотор IL-2, энхансер/промотор (единый) вируса саркомы Рауса, промотор с длинными концевыми повторами (LTR) вируса некроза селезенки, промотор TBG, промотор из гена тироксинсвязывающего глобулина человека (специфический для печени), промотор PGK, промотор убиквитина C человека (UBC), промотор UCOE (промотор HNRPA2B1-CBX3), синтетический промотор CAG, промотор гистона H2, промотор гистона H3, промотор малой ядерной РНК U1a1 (226 нукл.), промотор малой ядерной РНК U1a1 (226 нукл.), промотор малой ядерной РНК U1b2 (246 нукл.) 26, промотор GUSB, промотор CBh, промотор родопсина (Rho), склонный к сайленсингу промотор вируса некроза селезенки (SFFV), промотор H1 человека (H1), промотор POL1, минимальный энхансер/промотор TTR, промотор b-кинезина, промотор с длинными терминальными повторами (LTR) вируса опухоли молочной железы мыши, промотор эукариотического фактора 4A инициации человека (EIF4A1), промотор ROSA26, промотор глицеральдегид-3-фосфатдегидрогеназы (GAPDH), промоторы тРНК, а также усеченные версии и варианты последовательности вышеизложенного. В конкретном варианте осуществления промотор pol II представляет собой EF-1альфа, причем промотор усиливает эффективность трансфекции, транскрипцию или экспрессию нуклеазы CRISPR, долю экспрессионно-положительных клонов и число копий эписомального вектора в долгосрочной культуре.

[0333] Не имеющие ограничительного характера примеры промоторов pol III включают в себя, без ограничений, U6, мини-U6, U6-усеченные промоторы, 7SK, варианты H1, BiH1 (двусторонний промотор H1), BiU6, Bi7SK, BiH1 (двусторонние промоторы U6, 7SK и H1), U6 гориллы, U6 резуса, 7SK человека, H1 человека и варианты их последовательностей. В вышеуказанном варианте осуществления промотор pol III усиливает транскрипцию гРНК.

[0334] Выбор соответствующего вектора и промотора знаком рядовому специалисту в данной области, так как связан с контролем экспрессии, например, для модификации гена. Экспрессионный вектор также может содержать сайт связывания рибосомы для инициации трансляции и терминатор транскрипции. Экспрессионный вектор также может включать подходящие последовательности для усиления экспрессии. Экспрессионный вектор также может включать нуклеотидные последовательности, кодирующие белковые метки (например, метка 6xHis, гемагглютининовая метка, флуоресцентный белок и т. д.), которые могут быть слиты с белком CasX, что приводит к получению химерного белка CasX, который применяют для очистки или обнаружения.

[0335] Рекомбинантные экспрессионные векторы описания могут также содержать

элементы, которые способствуют устойчивой экспрессии белков CasX и гРНК описания. Например, рекомбинантные экспрессионные векторы могут включать одно или более из сигнала полиаденилирования (поли(А)), интронной последовательности или посттранскрипционного регуляторного элемента, такого как посттранскрипционный регуляторный элемент вируса гепатита лесного сурка (WPRE). Типичные последовательности поли(А) включают сигнал поли(А) hGH (короткий), сигнал поли(А) HSV TK, синтетические сигналы полиаденилирования, сигнал поли(А) SV40, сигнал поли(А) β-глобина и т. п. Специалист в данной области сможет выбрать приемлемые элементы для включения в рекомбинантные экспрессионные векторы, описанные в настоящем документе.

[0336] В некоторых вариантах осуществления в настоящем документе предложены один или более рекомбинантных экспрессионных векторов, содержащих одно или более из: (i) нуклеотидной последовательности нуклеиновой кислоты донорной матрицы, где донорная матрица содержит нуклеотидную последовательность, имеющую гомологию с последовательностью целевого локуса целевой нуклеиновой кислоты (например, целевого генома); (ii) нуклеотидной последовательности, кодирующей гРНК, которая гибридизуется с целевой последовательностью локуса целевого генома (например, выполненную в виде одинарной или двойной гидовой РНК), функционально связанной с промотором, который способен функционировать в клетке-мишени, такой как эукариотическая клетка; и (iii) нуклеотидной последовательности, кодирующей белок CasX, функционально связанный с промотором, который способен функционировать в клетке-мишени, такой как эукариотическая клетка. В некоторых вариантах осуществления последовательности, кодирующие донорную матрицу, гРНК и белок CasX, находятся в разных рекомбинантных экспрессионных векторах, а в других вариантах осуществления одна или более полинуклеотидных последовательностей (донорной матрицы, CasX и гРНК) находятся в одном рекомбинантном экспрессионном векторе.

[0337] Полинуклеотидную (-ые) последовательность (-и) вводят в вектор различными процедурами. В целом, ДНК вставляют в соответствующий (-ие) сайт (-ы) рестрикционной эндонуклеазы с применением методик, известных в данной области. Компоненты вектора, как правило, включают в себя, без ограничений, одно или более из сигнальной последовательности, точки начала репликации, одного или более маркерных генов, энхансерного элемента, промотора и последовательности терминации транскрипции. При конструировании приемлемых векторов, содержащих один или более из этих компонентов, используют стандартные методики лигирования, известные специалистам в данной области. Такие методики хорошо известны в данной области и хорошо описаны в научной и патентной литературе. Различные векторы являются общедоступными. Вектор может, например, быть в форме плазмиды, космиды, вирусной частицы или фага, который можно удобно подвергать процедурам получения рекомбинантной ДНК, и выбор вектора часто зависит от клетки-хозяина, в которую его необходимо ввести. Таким образом, вектор может представлять собой автономно реплицирующийся вектор, т. е. вектор, который

существует как внехромосомный объект, репликация которого не зависит от хромосомной репликации, например, плазмиду. Альтернативно вектор может представлять собой вектор, который при введении в клетку-хозяина интегрируется в геном клетки-хозяина и реплицируется вместе с хромосомой (-ами), в которую он интегрирован. После введения в приемлемую клетку-хозяина экспрессию белка, участвующего в процессинге антигена, презентации антигена, распознавания антигена и/или реакции на антиген, можно определять с применением любого анализа нуклеиновых кислот или белков, известного в данной области. Например, наличие транскрибируемой мРНК референсного CasX или вариантов CasX может быть обнаружено и/или количественно определено обычными гибридизационными анализами (например, анализом нозерн-блот), процедурами амплификации (например, ОТ-ПЦР), серийным анализом генной экспрессии (SAGE) (патент США № 5,695,937) и технологиями на основе матриц (см., например, патенты США № 5,405,783, 5,412,087 и 5,445,934) с применением зондов, комплементарных любой области полинуклеотида.

[0338] Полинуклеотиды и рекомбинантные экспрессионные векторы могут быть доставлены в целевые клетки-хозяева различными способами. Такие способы включают в себя, без ограничений, вирусную инфекцию, трансфекцию, липофекцию, электропорацию, преципитацию фосфатом кальция, опосредованную полиэтиленимином (PEI) трансфекцию, опосредованную DEAE-декстраном трансфекцию, микроинъекцию, опосредованную липосомами трансфекцию, технологию генной пушки, нуклеофекцию, прямое добавление проникающих в клетки белков CasX, которые слиты или рекрутируют донорную ДНК, сжатие клеток, преципитацию фосфатом кальция, прямую микроинъекцию, опосредованную наночастицами доставку нуклеиновых кислот, а также применение коммерческих наборов реагентов TransMessenger® от Qiagen, наборов для трансфекции StemfectTM RNA Transfection Kit от Stemgent и TransIT®-mRNA Transfection Kit от Mirus Bio LLC, нуклеофекция Lonza, электропорация Maxagen и т. п.

[0339] Последовательность рекомбинантного экспрессионного вектора может быть упакована в вирус или вирусоподобную частицу (также называемую в настоящем документе «частица» или «вирион») для последующей инфекции и трансформации клетки *ex vivo*, *in vitro* или *in vivo*. Такие частицы или вирионы, как правило, включают белки, которые заключают в капсид или упаковывают векторный геном. Приемлемые экспрессионные векторы могут включать вирусные экспрессионные векторы на основе вируса осповакцины; полиовируса; аденовируса; ретровирусный вектор (например, вирус лейкоза мыши), вирус некроза селезенки и векторы, полученные из ретровирусов, таких как вирус саркомы Рауса, вирус саркомы Харви, вирус птичьего лейкоза, лентивирус, вирус иммунодефицита человека, вирус миелопролиферативной саркомы и вирус опухоли молочной железы; и т. п. В некоторых вариантах осуществления рекомбинантный экспрессионный вектор настоящего описания представляет собой рекомбинантный аденоассоциированный вирусный (AAV) вектор. В некоторых вариантах осуществления рекомбинантный экспрессионный вектор настоящего описания представляет собой

рекомбинантный лентивирусный вектор. В некоторых вариантах осуществления рекомбинантный экспрессионный вектор настоящего описания представляет собой рекомбинантный ретровирусный вектор.

[0340] В некоторых вариантах осуществления рекомбинантный экспрессионный вектор настоящего описания представляет собой рекомбинантный аденоассоциированный вирусный (AAV) вектор. В некоторых вариантах осуществления рекомбинантный экспрессионный вектор настоящего описания представляет собой рекомбинантный лентивирусный вектор. В некоторых вариантах осуществления рекомбинантный экспрессионный вектор настоящего описания представляет собой рекомбинантный ретровирусный вектор.

[0341] AAV представляет собой небольшой (20 нм) непатогенный вирус, который можно использовать для лечения заболеваний человека в ситуациях, когда вирусный вектор используется для доставки в клетку, такую как эукариотическая клетка, in vivo или *ex vivo* с целью подготовки клеток для введения субъекту. Создают конструкт, например конструкт, кодирующий любой из белков CasX и/или вариантов осуществления CasX и гРНК, как описано в настоящем документе, и фланкируют последовательностями инвертированных концевых повторов AAV (ITR), таким образом обеспечивая упаковку AAV-вектора в вирусную частицу AAV.

[0342] Вектор «AAV» может относиться к самому природному вирусу дикого типа или к его производным. Термин охватывает все подтипы, серотипы и псевдотипы, и как природные, так и рекомбинантные формы, за исключением случаев с другими требованиями. В настоящем документе термин «серотип» относится к AAV, который идентифицируется и отличается от других AAV на основе реактивности капсидного белка с определенной антисывороткой, например существует множество известных серотипов AAV приматов. В некоторых вариантах осуществления AAV-вектор выбран из AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAV 44.9, AAV-Rh74 (AAV, полученный от макака-резуса) и AAVRh10 и модифицированных капсидов этих серотипов. Например, серотип AAV-2 означает AAV, который содержит капсидные белки, кодируемые геном Cap AAV-2, и геном содержит последовательности 5'- и 3'-ITR из того же серотипа AAV-2. Псевдотипом AAV называется AAV, который содержит капсидные белки из одного серотипа, а вирусный геном, включающий 5'-3' ITR, - от второго серотипа. Ожидается, что псевдотипичный rAAV будет иметь свойства связывания с клеточной поверхностью, характерные для серотипа капсида, и генетические свойства, соответствующие серотипу ITR. Псевдотипичный рекомбинантный AAV (rAAV) получают с применением стандартных методик, описанных в данной области. В настоящем документе, например, термин «rAAV1» может применяться для обозначения AAV, имеющего как капсидные белки, так и 5'-3' ITR из одного серотипа, или он может относиться к AAV, имеющему капсидные белки из серотипа 1 и 5'-3' ITR из другого серотипа AAV, например AAV серотипа 2. В каждом примере, представленном в настоящем документе, в описании конструкции и продукции вектора приводится серотип

капсида и последовательностей 5’-3’ ITR.

[0343] Термин «вирус AAV» или «вирусная частица AAV» относится к вирусной частице, состоящей из по меньшей мере одного капсидного белка AAV (предпочтительно всех капсидных белков AAV дикого типа) и инкапсулированного полинуклеотида. Если частица дополнительно содержит гетерологичный полинуклеотид (*т. е.* полинуклеотид, отличный от генома AAV дикого типа, который должен быть доставлен в клетку млекопитающего), она, как правило, называется «rAAV». Пример гетерологичного полинуклеотида представляет собой полинуклеотид, содержащий белок CasX и/или orРНК, и необязательно донорную матрицу по любому из вариантов осуществления, описанных в настоящем документе.

[0344] Под «инвертированными концевыми повторами аденоассоциированного вируса», или «ITR AAV», понимаются области, известные в данной области, находящиеся на каждом конце генома AAV, которые в совокупности функционируют как цис-активные в качестве точки начала репликации ДНК и в качестве сигналов паковки вируса. ITR AAV вместе с кодирующей областью rep AAV обеспечивают эффективное вырезание и извлечение, а также интеграцию нуклеотидной последовательности, расположенной между двумя фланкирующими ITR, в геном клетки млекопитающего.

[0345] Нуклеотидные последовательности областей ITR AAV известны. См., например, Kotin, R.M. (1994) Human Gene Therapy 5:793-801; Berns, K. I. «Parvoviridae and their Replication» in Fundamental Virology, 2nd Edition, (B. N. Fields and D. M. Knipe, eds.). В настоящем документе ITR AAV не обязательно должен иметь изображенную нуклеотидную последовательность дикого типа, она может быть изменена, например посредством вставки, делеции или замены нуклеотидов. Кроме того, ITR AAV могут быть получены из любого из нескольких серотипов AAV, включая, без ограничений, AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV-Rh74 и AAVRh10, и модифицированных капсидов этих серотипов. Более того, 5’ и 3’-ITR, которые фланкируют выбранную нуклеотидную последовательность в AAV-векторе, не обязательно должны быть идентичными или полученными из одного и того же серотипа или изолята AAV, при условии, что они функционируют так, как ожидается, *т. е.* обеспечивают вырезание и выделение интересующей последовательности из генома или вектора клетки-хозяина и обеспечивают интеграцию гетерологичной последовательности в геном клетки-реципиента, когда в клетке присутствуют генные продукты Rep AAV. В данной области известно применение серотипов AAV для интеграции гетерологичных последовательностей в клетку-хозяина (см., например, WO2018195555A1 и US20180258424A1, включенные в настоящий документ путем ссылки).

[0346] Под термином «кодирующая область AAV Rep» подразумевается область генома AAV, кодирующая репликационные белки Rep 78, Rep 68, Rep 52 и Rep 40. Было показано, что продукты экспрессии Rep обладают многими функциями, включая распознавание, связывание и никирование точки начала репликации ДНК AAV, ДНК-геликазной активностью и модуляцией транскрипции с промоторов AAV (или других

гетерологичных промоторов). Продукты экспрессии Rep в совокупности требуются для репликации генома AAV. Под термином «область кодирования AAV Cap» подразумевается область генома AAV, кодирующая капсидные белки VP1, VP2 и VP3 или их функциональные гомологи. Эти продукты экспрессии гена Cap обеспечивают функции упаковки, которые в совокупности необходимы для упаковки вирусного генома.

[0347] В некоторых вариантах осуществления капсиды AAV, используемые для доставки кодирующих последовательностей для CasX и гРНК и необязательно донорных матриц DMPK в клетку-хозяина, могут быть получены из любого из нескольких серотипов AAV, включая, без ограничений, AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAV 44.9, AAV-Rh74 (AAV из макака-резуса) и AAVRh10, а AAV ITR получают из AAV серотипа 2. В конкретном варианте осуществления AAV1, AAV7, AAV6, AAV8 или AAV9 используют для доставки CasX, гРНК и необязательно донорных матричных нуклеотидов в мышечную клетку-хозяина.

[0348] Для получения вирусных частиц rAAV экспрессионный AAV-вектор вводят в приемлемую клетку-хозяина с применением известных методик, таких как трансфекция. Для формирования частиц вируса обычно применяют пакующие клетки; к таким клеткам относятся клетки HEK293 (и другие клетки, известные в данной области), которые пакуют аденовирус. Ряд методик трансфекции общеизвестен в данной области; см., например, Sambrook et al. (1989) Molecular Cloning, a laboratory manual, Cold Spring Harbor Laboratories, New York. Особенно подходящие способы трансфекции включают сопреципитацию с фосфатом кальция, прямую микроинъекцию в культивируемые клетки, электропорацию, опосредованный липосомами перенос гена, опосредованную липидами трансдукцию и доставку нуклеиновой кислоты с применением бомбардировки микрочастицами.

[0349] В преимуществе конструктов rAAV настоящего описания меньший размер нуклеаз CRISPR класса 2, типа V; например, варианты CasX по вариантам осуществления, позволяет включать все необходимые компоненты для редактирования и вспомогательные компоненты для экспрессии в трансген таким образом, чтобы одна частица rAAV могла доставлять и трансдуцировать эти компоненты в клетку-мишень в форме, которая приводит к экспрессии нуклеазы CRISPR и гРНК, способным эффективно модифицировать целевую нуклеиновую кислоту клетки-мишени. Типичная схема такого конструкта представлена на ФИГ. 13. Этим она заметно отличается от других систем CRISPR, таких как Cas9, где обычно используется система с двумя частицами для доставки необходимых компонентов для редактирования в целевую клетку. Таким образом, в некоторых вариантах осуществления систем rAAV в описании предложены: i) первая плазмида, содержащая ITR, последовательности, кодирующие вариант CasX, последовательности, кодирующие одну или более гРНК, первый промотор, функционально связанный с CasX, и второй промотор, функционально связанный с гРНК, и необязательно один или более энхансерных элементов; ii) вторая плазмида, содержащая гены rep и cap; и iii) третья плазмида, содержащая хелперные гены, причем при трансфекции соответствующей пакующей клетки клетка способна продуцировать rAAV, обладающий способностью доставлять в клетку-

мишень в одной частице последовательности, способные экспрессировать нуклеазу CasX и гРНК, имеющие способность редактировать целевую нуклеиновую кислоту клетки-мишени. В некоторых вариантах осуществления систем rAAV последовательность, кодирующая белок CRISPR, и последовательность, кодирующая по меньшей мере первую гРНК, имеют длину менее чем примерно 3100, менее чем примерно 3090, менее чем примерно 3080, менее чем примерно 3070, менее чем примерно 3060, менее чем примерно 3050 или менее чем примерно 3040 нуклеотидов в длину, так что последовательности, кодирующие первый и второй промотор, и необязательно один или более энхансерных элементов могут иметь в общей сложности длину по меньшей мере примерно 1300, по меньшей мере примерно 1350, по меньшей мере примерно 1360, по меньшей мере примерно 1370, по меньшей мере примерно 1380, по меньшей мере примерно 1390 нуклеотидов, по меньшей мере примерно 1400, по меньшей мере примерно 1500, по меньшей мере примерно 1600 нуклеотидов, по меньшей мере 1650, по меньшей мере примерно 1700, по меньшей мере примерно 1750, по меньшей мере примерно 1800, по меньшей мере примерно 1850 или по меньшей мере примерно 1900 нуклеотидов. В некоторых вариантах осуществления систем rAAV последовательность, кодирующая первый промотор и по меньшей мере один вспомогательный элемент, имеет в общей сложности длину более чем по меньшей мере примерно 1300, по меньшей мере примерно 1350, по меньшей мере примерно 1360, по меньшей мере примерно 1370, по меньшей мере примерно 1380, по меньшей мере примерно 1390, по меньшей мере примерно 1400, по меньшей мере примерно 1500, по меньшей мере примерно 1600 нуклеотидов, по меньшей мере 1650, по меньшей мере примерно 1700, по меньшей мере примерно 1750, по меньшей мере примерно 1800, по меньшей мере примерно 1850 или по меньшей мере примерно 1900 нуклеотидов. В некоторых вариантах осуществления систем rAAV последовательность, кодирующая первый и второй промоторы и по меньшей мере один вспомогательный элемент, имеет в общей сложности длину более чем по меньшей мере примерно 1300, по меньшей мере примерно 1350, по меньшей мере примерно 1360, по меньшей мере примерно 1370, по меньшей мере примерно 1380, по меньшей мере примерно 1390, по меньшей мере примерно 1400, по меньшей мере примерно 1500, по меньшей мере примерно 1600 нуклеотидов, по меньшей мере 1650, по меньшей мере примерно 1700, по меньшей мере примерно 1750, по меньшей мере примерно 1800, по меньшей мере примерно 1850 или по меньшей мере примерно 1900 нуклеотидов.

[0350] В некоторых вариантах осуществления клетки-хозяева, трансфицированные вышеописанными экспрессионными векторами AAV, становятся способными обеспечивать хелперные функции AAV для репликации и заключения в капсид нуклеотидных последовательностей, фланкированных ITR AAV, с получением вирусных частиц rAAV. Хелперные функции AAV обычно представляют собой кодирующие последовательности, полученные из AAV, которые могут быть экспрессированы с получением генных продуктов AAV, которые, в свою очередь, функционируют как транс-активные, обеспечивая продуктивную репликацию AAV. В настоящем документе хелперные функции AAV применяются для дополнения необходимых функций AAV,

которые отсутствуют в экспрессионных векторах AAV. Таким образом, хелперные функции AAV включают одну или обе из основных ORF (открытых рамок считывания) AAV, которые кодируют кодирующие области rep и cap, или их функциональные гомологи. Вспомогательные функции могут быть введены в клетки-хозяева и затем экспрессированы способами, известными специалистам в данной области. Обычно вспомогательные функции обеспечиваются посредством инфицирования клеток-хозяев неродственным хелперным вирусом. В некоторых вариантах осуществления вспомогательные функции обеспечиваются с применением вспомогательного функционального вектора. В зависимости от используемой системы хозяин/вектор в экспрессионном векторе можно применять любое число приемлемых элементов управления транскрипцией и трансляцией, включая конститутивные и индуцибельные промоторы, транскрипционные энхансерные элементы, терминаторы транскрипции и т. д. В некоторых вариантах осуществления в описании предложены клетки-хозяева, содержащие AAV-векторы по вариантам осуществления, описанным в настоящем документе.

[0351] В других вариантах осуществления приемлемые векторы могут включать вирусоподобные частицы (VLP). Вирусоподобные частицы (VLP) представляют собой частицы, которые очень похожи на вирусы, но не содержат вирусного генетического материала и, следовательно, не являются инфекционными. В некоторых вариантах осуществления VLP содержат полинуклеотид, кодирующий интересующий трансген, например любой из белка CasX и/или вариантов осуществления гРНК, и необязательно донорные матричные полинуклеотиды, описанные в настоящем документе, упакованные с одним или более вирусными структурными белками.

[0352] В других вариантах осуществления в описании предложены частицы для доставки CasX (XDP), полученные *in vitro*, которые содержат RNP-комплекс CasX : гРНК и необязательно донорную матрицу. Для создания XDP можно применять комбинации структурных белков из разных вирусов, включая компоненты из семейств вирусов, включая Parvoviridae (например, аденоассоциированный вирус), Retroviridae (например, альфа-ретровирус, бета-ретровирус, гамма-ретровирус, дельта-ретровирус, эпсилон-ретровирус или лентивирус), Flaviviridae (например, вирус гепатита С), Paramyxoviridae (например, вирус Нипах) и бактериофаги (например, Qβ, AP205). В некоторых вариантах осуществления в описании предложены системы XDP, разработанные с применением компонентов ретровируса, включая лентивирусы (такие как ВИЧ) и альфа-ретровирус, бета-ретровирус, гамма-ретровирус, дельта-ретровирус, эпсилон-ретровирус, в которых отдельные плазмиды, содержащие полинуклеотиды, кодирующие различные компоненты, вводят в пакующую клетку, которая, в свою очередь, продуцирует XDP. В некоторых вариантах осуществления в описании предложена XDP, содержащая один или более компонентов из i) протеазы, ii) сайта расщепления протеазой, iii) одного или более компонентов полипротеина Gag, выбранных из матриксного белка (MA), нуклеокапсидного белка (NC), капсидного белка (CA), пептида p1, пептида p6, пептида P2A, пептида P2B, пептида P10, пептида p12, пептида PP21/24, пептида P12/P3/P8 и пептида P20; v) CasX; vi)

гРНК и vi) нацеливающих гликопротеинов или фрагментов антител, причем полученная XDP-частица заключает в капсид RNP-комплекс CasX : гРНК. Полинуклеотиды, кодирующие Gag, CasX и гРНК, могут дополнительно содержать парные компоненты, предназначенные для содействия перемещению компонентов из ядра клетки-хозяина и облегчения привлечения комплекса CasX : гРНК в отпочковывающуюся XDP. Не имеющие ограничительного характера примеры таких компонентов включают в себя шпилечную РНК, такую как шпилька MS2, шпилька PP7, шпилька Qβ и U1-шпилька II, встроенные в гРНК в качестве партнеров по связыванию, которые имеют аффинность связывания с упаковочным рекрутером-белком оболочки MS2, белком оболочки PP7, белком оболочки Qβ и частицей распознавания сигнала U1A соответственно, которые слиты с полипротеином Gag. Было обнаружено, что встраивание партнера по связыванию, вставленного в гидовую РНК, и упаковочного рекрутера в нуклеиновую кислоту, содержащую полипептид Gag, облегчает упаковку частиц XDP, отчасти из-за аффинности CasX к гРНК, что приводит к образованию RNP, так что как гРНК, так и CasX связаны с Gag в процессе инкапсуляции в XDP, увеличению доли XDP, содержащих RNP, по сравнению с конструктом, не содержащим партнера по связыванию и упаковочного рекрутера. В других вариантах осуществления гРНК может содержать rev-отвечающий элемент (RRE) или его участки, которые имеют аффинность связывания с Rev, который может быть связан с полипротеином Gag. В других вариантах осуществления гРНК может содержать один или более RRE и одну или более последовательностей шпильки MS2. RRE может быть выбран из группы, состоящей из стебля IIB rev-отвечающего элемента (RRE), стебля II-V RRE, стебля II RRE, rev-связывающего элемента (RBE) стебля IIB и полноразмерного RRE. В вышеуказанном варианте осуществления компоненты включают в себя последовательности UGGGCGCAGCGUCAAUGACGCUGACGGUACA (стебель IIB, SEQ ID NO: 1280), GCACUAUGGGCGCAGCGUCAAUGACGCUGACGGUACAGGCCAGACAAUUAUUGUC UGGUAUAGUGC (стебель II, SEQ ID NO: 1281), CAGGAAGCACUAUGGGCGCAGCGUCAAUGACGCUGACGGUACAGGCCAGACAAUU AUUGUCUGGUAUAGUGCAGCAGCAGAACAAUUUGCUGAGGGCUAUUGAGGCGCA ACAGCAUCUGUUGCAACUCACAGUCUGGGGCAUCAAGCAGCUCCAGGCAAGAAUC CUG (стебель II-V, SEQ ID NO: 1282), GCUGACGGUACAGGC (RBE, SEQ ID NO: 1284) и AGGAGCUUUGUUCCUUGGGUUCUUGGGAGCAGCAGGAAGCACUAUGGGCGCAGC GUCAAUGACGCUGACGGUACAGGCCAGACAAUUAUUGUCUGGUAUAGUGCAGCA GCAGAACAAUUUGCUGAGGGCUAUUGAGGCGCAACAGCAUCUGUUGCAACUCAC AGUCUGGGGCAUCAAGCAGCUCCAGGCAAGAAUCCUGGCUGUGGAAAGAUACCU AAAGGAUCAACAGCUCCU (полноразмерный RRE, SEQ ID NO: 1283). В других вариантах осуществления гРНК может содержать один или более RRE и одну или более последовательностей шпильки MS2. В конкретном варианте осуществления гРНК содержит вариант шпильки MS2, который оптимизирован для повышения аффинности связывания с белком оболочки MS2, улучшая таким образом встраивание гРНК и

ассоциированного CasX в отпочковывающийся XDP. В некоторых вариантах осуществления варианты гРНК, содержащие варианты шпильки MS2 и RRE, включают в себя варианты 275-315 гРНК (SEQ ID NO: 2353-2393), содержащие последовательности MS2, представленные в таблице 36. В некоторых вариантах осуществления в описании предложены варианты гРНК, содержащие один или более вариантов последовательностей шпильки MS2, причем вариант демонстрирует K_D для своего лиганда - белка оболочки MS2 менее 100 нМ, менее 50 нМ, менее 35 нМ, менее 10 нМ, менее 3 нМ или менее 2 нМ, и XDP, содержащая вариант гРНК, демонстрирует улучшенную активность редактирования в отношении целевой нуклеиновой кислоты в клеточном анализе *in vitro*, причем EC_{50} составляет менее 10^8, или менее 10^7, или менее 10^6 частиц для достижения редактирования в 50% клеток. Нацеливающие гликопротеины или фрагменты антител на поверхности, которые обеспечивает тропизм XDP к клетке-мишени, причем при введении и входе в клетку-мишень молекула RNP свободно транспортируется в ядро клетки. Гликопротеин оболочки может быть получен из любых оболочечных вирусов, которые, как известно в данной области, придают XDP тропизм, включая, без ограничений, группу, состоящую из вируса аргентинской геморрагической лихорадки, вируса австралийских летучих мышей, вируса множественного ядерного полиэдроза Autographa californica, вируса птичьего лейкоза, эндогенного вируса бабуина, вируса боливийской геморрагической лихорадки, вируса болезни Борна, вируса Бреда, вируса Буньямвера, вируса Чандипура, вируса Чикунгунья, вируса конго-крымской геморрагической лихорадки, вируса лихорадки Денге, вируса Дювенажа, вируса восточного энцефалита лошадей, вируса геморрагической лихорадки Эбола, вируса Эбола Заира, кишечного аденовируса, эфемеровируса, вируса Эпштейна - Барр (EBV), вируса 1 европейских летучих мышей, вируса 2 европейских летучих мышей, продукта синтетического слияния Fug и gP, вируса лейкоза обезьян-гиббонов, хантавируса, вируса Хендра, вируса гепатита A, вируса гепатита B, вируса гепатита C, вируса гепатита D, вируса гепатита E, вируса гепатита G (вируса GB C), вируса простого герпеса типа 1, вируса простого герпеса типа 2, цитомегаловируса человека (HHV5), пенящего вируса человека, герпесвируса человека (HHV), герпесвируса человека типа 7, герпесвируса человека типа 6, герпесвируса человека типа 8, вируса иммунодефицита человека 1 (ВИЧ-1), метапневмовируса человека, Т-лимфотропного вируса 1 человека, вируса гриппа A, гриппа B, гриппа C, вируса японского энцефалита, вируса герпеса, ассоциированного с саркомой Капоши (HHV8), вируса Кьясанурской лесной болезни, вируса Ла-Кросс, вируса лагосских летучих мышей, вируса лихорадки Ласса, вируса лимфоцитарного хориоменингита (LCMV), вируса Мачупо, вируса геморрагической лихорадки Марбурга, вируса кори, коронавируса, связанного с ближневосточным респираторным синдромом, вируса Мокола, вируса мышиного лейкоза Молони, вируса оспы обезьян, вируса опухоли молочной железы мыши, вируса эпидемического паротита, мышиного гамма-герпесвируса, вируса болезни Ньюкасла, вируса Нипах, вируса Нипах, норовируса, вируса омской геморрагической лихорадки, вируса папилломы, парвовируса, вируса псевдобешенства, вируса Quaranfil, вируса

бешенства, эндогенного ретровируса кошачьих RD114, респираторно-синцитиального вируса (RSV), вируса лихорадки долины Рифт, вируса реки Росс, ротавируса, вируса саркомы Рауса, вируса краснухи, вируса геморрагической лихорадки, ассоциированной с вирусом Сабиа, SARS-ассоциированного коронавируса (SARS-CoV), вируса Сендай, вируса Такарибе, тоготовируса, вируса, вызывающего клещевой энцефалит, вируса ветряной оспы (HHV3), вируса ветряной оспы (HHV3), вируса натуральной оспы, вируса малой оспы, вируса венесуэльского энцефалита лошадей, вируса венесуэльской геморрагической лихорадки, вируса везикулярного стоматита (VSV), VSV-G, везикуловируса, вируса Западного Нила, вируса западного энцефалита лошадей и вируса Зика.

[0353] В других вариантах осуществления в описании предложена XDP согласно вышеизложенному, дополнительно содержащая один или более компонентов полипротеина pol (например, протеазу) и необязательно второй CasX или донорную матрицу. В описании предусмотрено множество конфигураций расположения закодированных компонентов, включая дубликаты некоторых закодированных компонентов. Вышеизложенное обеспечивает преимущества по сравнению с другими векторами в данной области тем, что вирусная трансдукция в делящиеся и неделящиеся клетки является эффективной и что XDP доставляет эффективный и короткий RNP, который избегает механизмов иммунного надзора субъекта, которые в противном случае обнаружили бы чужеродный белок. Не имеющие ограничительного характера примеры систем XDP описаны в PCT/US20/63488 и WO2021113772A1, включенных в настоящий документ путем ссылки. В некоторых вариантах осуществления в описании предложены клетки-хозяева, содержащие полинуклеотиды или векторы, кодирующие любые из вышеуказанных вариантов осуществления XDP.

[0354] После получения и извлечения XDP, содержащей RNP CasX : гРНК по любому из вариантов осуществления, описанных в настоящем документе, XDP можно применять в способах редактирования клеток-мишеней субъектов посредством введения такого XDP, как более подробно описано ниже.

[0355] Векторы также можно доставлять невирусной доставкой, при которой вектор или векторы, кодирующие варианты CasX и гРНК, готовят в наночастицах, причем предусмотренные наночастицы включают в себя, без ограничений, наносферы, липосомы, липидные наночастицы, квантовые точки, частицы полиэтиленгликоля, гидрогели и мицеллы. Липидные наночастицы по существу состоят из ионизируемого катионного липида и трех или более дополнительных компонентов, таких как холестерин, DOPE, полимолочная-ко-гликолевая кислота и полиэтиленгликоль (PEG)-содержащий липид. В некоторых вариантах осуществления варианты CasX по вариантам осуществления, описанным в настоящем документе, приготовлены в липидной наночастице. В некоторых вариантах осуществления липидная наночастица содержит гРНК по вариантам осуществления, описанным в настоящем документе. В некоторых вариантах осуществления липидная наночастица содержит RNP варианта CasX в комплексе с гРНК. В некоторых

вариантах осуществления система содержит липидную наночастицу, содержащую нуклеиновые кислоты, кодирующие варианты CasX и гРНК, и необязательно нуклеиновую кислоту донорной матрицы. В некоторых вариантах осуществления компоненты системы CasX : гРНК готовят в отдельных липидных наночастицах для доставки в клетки или для введения нуждающемуся в этом субъекту.

VII. Способы модификации целевой нуклеиновой кислоты

[0356] Белки CRISPR, гиды, нуклеиновые кислоты и их варианты, предложенные в настоящем документе, а также векторы, кодирующие такие компоненты, можно использовать в различных областях применения, включая терапевтические, диагностические и исследовательские.

[0357] В некоторых вариантах осуществления для осуществления способов описания для редактирования генов и модификации целевой нуклеиновой кислоты в клетке в настоящем документе предложены программируемые пары для редактирования из варианта CasX класса 2, типа V и варианта гРНК (CasX : гРНК). Программируемый характер пар, предложенных в настоящем документе, позволяет точно нацеливать их на достижение требуемой модификации в одной или более заданных интересующих областях в целевой нуклеиновой кислоте гена. Для модификации последовательности целевой нуклеиновой кислоты в клетке с применением систем, предложенных в настоящем документе, можно использовать различные стратегии и способы. В настоящем документе термин «модификация» включает, без ограничений, расщепление, никирование, редактирование, делецию, нокаут, нокдаун, мутирование, коррекцию, пропуск экзона и т. п. Как описано в настоящем документе, вариант CasX, вводящий расщепление двухцепочечных молекул в целевую нуклеиновую кислоту, создает двухцепочечный разрыв в пределах 18-26 нуклеотидов с 5’-стороны от сайта PAM на целевой цепи и 10-18 нуклеотидов с 3’-стороны на нецелевой цепи. Полученная в результате этого модификация может приводить к случайным вставкам или делециям (инделам) или к замене, дублированию, сдвигу рамки или инверсии одного или более нуклеотидов в этих областях с помощью репарационных механизмов соединения негомологичных концов ДНК (NHEJ).

[0358] В некоторых вариантах осуществления в описании предложены способы модификации целевой нуклеиновой кислоты в клетке, причем способ включает приведение целевой нуклеиновой кислоты клетки в контакт с: i) редактирующей парой из белка CRISPR класса 2, типа V и гРНК (CasX : гРНК), содержащей вариант CasX и вариант гРНК по любому из вариантов осуществления, описанных в настоящем документе; ii) редактирующей парой CasX : гРНК в сочетании с донорной матрицей по любому из вариантов осуществления, описанных в настоящем документе; iii) нуклеиновой кислоты, кодирующей редактирующую пару из CasX и гРНК и необязательно содержащей донорную матрицу; iv) вектора, содержащего нуклеиновую кислоту по п. (iii), выше; v) XDP, содержащей редактирующую пару CasX : гРНК по любому из вариантов осуществления, описанных в настоящем документе; или vi) комбинациями двух или более из пп. (i)-(v), причем приведение целевой нуклеиновой кислоты в контакт с парой для редактирования

генов из белка CasX и гРНК и необязательно донорной матрицей приводит к модификации целевой нуклеиновой кислоты.

В некоторых случаях модификация приводит к коррекции или компенсации мутации в клетке, создавая таким образом отредактированную клетку, так что в ней может происходить экспрессия продукта функционального гена. В других вариантах осуществления способа модификация включает в себя подавление или устранение экспрессии продукта гена посредством нокдауна или нокаута гена.

[0359] В некоторых вариантах осуществления способа модификации последовательности целевой нуклеиновой кислоты в клетке, причем способ включает приведение целевой нуклеиновой кислоты в контакт с парой для редактирования CasX : гРНК, при этом пара для редактирования содержит вариант CasX, выбранный из группы, состоящей из SEQ ID NO: 247-592 и 1147-1231, указанных в таблице 3, или вариант CasX, выбранный из группы, состоящей из SEQ ID NO: 270-592 и 1147-1231, или вариант CasX, выбранный из группы, состоящей из SEQ ID NO: 415-592 и 1147-1231, или последовательность варианта, которая на по меньшей мере 60% идентична, на по меньшей мере 70% идентична, на по меньшей мере 80% идентична, на по меньшей мере 81% идентична, на по меньшей мере 82% идентична, на по меньшей мере 83% идентична, на по меньшей мере 84% идентична, на по меньшей мере 85% идентична, на по меньшей мере 86% идентична, на по меньшей мере 86% идентична, на по меньшей мере 87% идентична, на по меньшей мере 88% идентична, на по меньшей мере 89% идентична, на по меньшей мере 89% идентична, на по меньшей мере 90% идентична, на по меньшей мере 91% идентична, на по меньшей мере 92% идентична, на по меньшей мере 93% идентична, на по меньшей мере 94% идентична, на по меньшей мере 95% идентична, на по меньшей мере 96% идентична, на по меньшей мере 97% идентична, на по меньшей мере 98% идентична, на по меньшей мере 99% идентична, или на по меньшей мере 99,5% идентична ей, и каркас гРНК содержит последовательность, выбранную из группы, состоящей из SEQ ID NO: 2101-2332 и 2353-2398, указанных в таблице 2, причем каркас гРНК содержит последовательность, выбранную из группы, состоящей из SEQ ID NO: 2238-2332 и 2353-2398, при этом каркас РНК содержит последовательность, выбранную из группы, состоящей из SEQ ID NO: 2281-2332 и 2353-2398, или последовательности, которая на по меньшей мере 65% идентична, на по меньшей мере 70% идентична, на по меньшей мере 75% идентична, на по меньшей мере 80% идентична, на по меньшей мере 81% идентична, на по меньшей мере 82% идентична, на по меньшей мере 83% идентична, на по меньшей мере 84% идентична, на по меньшей мере 85% идентична, на по меньшей мере 86% идентична, на по меньшей мере 86% идентична, на по меньшей мере 87% идентична, на по меньшей мере 88% идентична, на по меньшей мере 89% идентична, на по меньшей мере 89% идентична, на по меньшей мере 90% идентична, на по меньшей мере 91% идентична, на по меньшей мере 92% идентична, на по меньшей мере 93% идентична, на по меньшей мере 94% идентична, на по меньшей мере 95% идентична, на по меньшей мере 96% идентична, на по меньшей мере 97% идентична, на по меньшей мере 98% идентична, на по

меньшей мере 99% идентична, на по меньшей мере 99,5% идентична с ней, и гРНК содержит нацеливающую последовательность, комплементарную целевой нуклеиновой кислоте и способную гибридизоваться с целевой нуклеиновой кислотой.

[0360] В некоторых вариантах осуществления пара для редактирования генов CasX : гРНК способна связываться друг с другом в рибонуклеопротеиновый комплекс (RNP). В некоторых вариантах осуществления пара для редактирования генов CasX : гРНК связана друг с другом в рибонуклеопротеиновый комплекс (RNP). В некоторых вариантах осуществления RNP способен связывать целевую нуклеиновую кислоту и создавать двухцепочечный разрыв в целевой нуклеиновой кислоте, что приводит к устойчивому инделу или мутации в целевой нуклеиновой кислоте. В других вариантах осуществления RNP способен связывать целевую нуклеиновую кислоту и создавать один или более одноцепочечных ников в целевой нуклеиновой кислоте, что приводит к устойчивому инделу или мутации в целевой нуклеиновой кислоте. В других вариантах осуществления RNP способен связывать целевую нуклеиновую кислоту, но не способен расщеплять целевую нуклеиновую кислоту; т. е. содержит вариант dCasX. В некоторых вариантах осуществления способа вариант белка CasX может быть представлен в клетках в виде полипептида, который может быть необязательно слит с полипептидным доменом, который повышает растворимость продукта. Домен может быть связан с полипептидом посредством определенного сайта расщепления протеазой; например, последовательностью TEV, которая расщепляется протеазой TEV. Линкер может также включать в себя одну или более гибких последовательностей; например, от 1 до 10 глициновых остатков. В некоторых вариантах осуществления расщепление слитного белка осуществляют в буфере, который сохраняет растворимость продукта; например, в присутствии от 0,5 до 2 М мочевины, в присутствии полипептидов и/или полинуклеотидов, которые повышают растворимость, и т. п. Интересующие домены могут включать в себя эндосомолитические домены; например домен HA вируса гриппа; и другие полипептиды, которые способствуют продукции, например домен IF2, домен GST, домен GRPE и т. п. Полипептид может быть приготовлен так, чтобы обеспечивать улучшенную стабильность. Например, пептиды могут быть пегилированы, причем полиэтиленоксигруппа обеспечивает увеличенную продолжительность жизни в кровотоке.

[0361] В других вариантах осуществления способа модификации последовательности целевой нуклеиновой кислоты в клетке способ включает приведение последовательности целевой нуклеиновой кислоты в контакт с множеством RNP с первой и второй или множеством гРНК, нацеленных на разные или перекрывающиеся части гена, причем белок CasX вводит множество разрывов в целевую нуклеиновую кислоту, что приводит к устойчивым инделам или мутациям в целевой нуклеиновой кислоте, как описано в настоящем документе, или вырезание вставочной последовательности между разрывами с соответствующей модуляцией экспрессии или изменением функции генного продукта, создавая таким образом модифицированную клетку.

[0362] В некоторых вариантах осуществления способ модификации целевой

нуклеиновой кислоты включает приведение целевой нуклеиновой кислоты в контакт с парой для редактирования генов CasX : гРНК, описанной в настоящем документе, и донорной матрицей. Таким образом, в некоторых случаях способ, предложенный в настоящем документе, включает приведение целевой нуклеиновой кислоты в контакт с донорным полинуклеотидом (например, посредством введения донорного полинуклеотида в клетку), причем донорный полинуклеотид, часть донорного полинуклеотида, копия донорного полинуклеотида или часть копии донорного полинуклеотида встраивается в целевую нуклеиновую кислоту. Например, экзогенная донорная матрица может содержать корректирующую последовательность, подлежащую встраиванию, фланкированную вышележащей последовательностью и нижележащей последовательностью, которая вводится в последовательность целевой нуклеиновой кислоты в клетке. В других случаях донорная матрица может содержать одну или более одиночных замен оснований, вставок, делеций, инверсий или перестроений по отношению к геномной последовательности, при условии, что существует достаточная гомология с последовательностью целевой нуклеиновой кислоты для обеспечения ее интеграции в целевую нуклеиновую кислоту, что может приводить к мутации сдвига рамки или другой мутации или замене этой части последовательности целевой нуклеиновой кислоты с соответствующим нокдауном или нокаутом дефектного гена в клетке. Последовательности, расположенные выше и ниже относительно сайта (-ов) расщепления, имеют сходство последовательности с одной из сторон сайта встраивания в целевой нуклеиновой кислоте (т. е. гомологичные плечи), облегчающие вставку. В других случаях экзогенную донорную матрицу вставляют между концами, образующимися в результате расщепления при помощи CasX, с помощью механизмов независимой от гомологии нацеленной интеграции (HITI). Экзогенная последовательность, вставленная посредством HITI, может иметь любую длину, например, относительно короткая последовательность длиной от 10 до 50 нуклеотидов или более длинная последовательность длиной примерно 50-1000 нуклеотидов. Отсутствие гомологии может иметь, например, идентичность последовательности не более чем 20-50% и/или отсутствие специфической гибридизации в условиях сниженной жесткости. В других случаях отсутствие гомологии может дополнительно включать критерий идентичности не более 5, 6, 7, 8 или 9 п. о. В некоторых вариантах осуществления донорный матричный полинуклеотид содержит по меньшей мере примерно 10, по меньшей мере примерно 50, по меньшей мере примерно 100, или по меньшей мере примерно 200, или по меньшей мере примерно 300, или по меньшей мере примерно 400, или по меньшей мере примерно 500, или по меньшей мере примерно 600, или по меньшей мере примерно 700, или по меньшей мере примерно 800, или по меньшей мере примерно 900, или по меньшей мере примерно 1000, или по меньшей мере примерно 10 000, или по меньшей мере примерно 15 000 нуклеотидов. В других вариантах осуществления донорная матрица содержит от по меньшей мере примерно 10 до примерно 15 000 нуклеотидов, или от по меньшей мере примерно 100 до примерно 10 000 нуклеотидов, или от по меньшей мере примерно 400 до примерно 8000 нуклеотидов, или от по меньшей мере примерно 600 до примерно 5000

нуклеотидов, или от по меньшей мере примерно 1000 до примерно 2000 нуклеотидов. В некоторых вариантах осуществления донорная матрица представляет собой одноцепочечную ДНК-матрицу или одноцепочечную РНК-матрицу. В других вариантах осуществления донорная матрица представляет собой двухцепочечную ДНК-матрицу. Последовательность донорной матрицы может содержать определенные различия в последовательности по сравнению с геномной последовательностью, например сайты рестрикции, нуклеотидные полиморфизмы, маркеры селекции (например, гены резистентности к лекарственному средству, флуоресцентные белки, ферменты и т. д.) и т. д., которые можно применять для оценки успешной вставки донорной нуклеиновой кислоты в сайт расщепления или, в некоторых случаях, применять для других целей (например, чтобы обозначить экспрессию в целевом геномном локусе). Альтернативно эти различия последовательности могут включать в себя фланкирующие рекомбинационные последовательности, такие как FLP, последовательности loxP или т. п.

[0363] В некоторых вариантах осуществления в описании предложены способы модификации последовательности целевой нуклеиновой кислоты клетки, включающие приведение целевой нуклеиновой кислоты указанной клетки в контакт с одним или более полинуклеотидами по любому из вариантов осуществления, описанных в настоящем документе, причем полинуклеотид (-ы) кодирует (-ют) пару для редактирования генов CasX : гРНК, при этом гРНК содержит нацеливающую последовательность, комплементарную и, следовательно, способную гибридизоваться с последовательностью целевой нуклеиновой кислоты, и причем приведение в контакт приводит к модификации целевой нуклеиновой кислоты. Способы введения нуклеиновой кислоты (например, нуклеиновой кислоты, содержащей донорную полинуклеотидную последовательность, одну или более нуклеиновых кислот, кодирующих вариант белка CasX и вариант гРНК, описанные в настоящем документе), в клетку известны в данной области, и можно использовать любой удобный способ. Приемлемые способы включают вирусную инфекцию, трансфекцию, липофекцию, электропорацию, преципитацию фосфатом кальция, опосредованную полиэтиленимином (PEI) трансфекцию, опосредованную DEAE-декстраном трансфекцию, опосредованную липосомами трансфекцию, технологию генной пушки, нуклеофекцию, электропорацию, прямое добавление проникающих в клетку белков CasX, которые слиты или рекрутируют донорную ДНК, сжатие клеток, преципитацию фосфатом кальция, прямую микроинъекцию и опосредованную наночастицами доставку нуклеиновой кислоты. Нуклеиновые кислоты могут быть представлены в клетках с применением хорошо разработанных методик трансфекции и коммерческих реагентов TransMessenger® от Qiagen, наборов для трансфекции Stemfect™ RNA Transfection Kit от Stemgent и TransIT®-mRNA Transfection Kit от Mirus Bio LLC, нуклеофекции Lonza, электропорации Maxagen и т. п. Нуклеиновая кислота, содержащая нуклеотидную последовательность, кодирующую вариант белка CasX, в некоторых случаях представляет собой РНК. Таким образом, в некоторых вариантах осуществления вариант белка CasX может быть введен в клетки в виде РНК. Способы введения РНК в клетки известны в данной области и могут включать,

например, прямую инъекцию, трансфекцию или любой другой способ, используемый для введения ДНК.

[0364] В других вариантах осуществления в описании предложены способы модификации последовательности целевой нуклеиновой кислоты клетки, включающие приведение указанной клетки в контакт с вектором по любому из вариантов осуществления, описанных в настоящем документе, содержащим нуклеиновую кислоту, кодирующую пару для редактирования генов CasX : гРНК, содержащую вариант белка CasX и вариант гРНК по любому из вариантов осуществления, описанных в настоящем документе, и необязательно донорную матрицу, причем гРНК содержит нацеливающую последовательность, комплементарную последовательности целевой нуклеиновой кислоты и, следовательно, способную гибридизоваться с ней, при этом приведение в контакт приводит к модификации целевой нуклеиновой кислоты. Введение рекомбинантных векторов экспрессии в клетки может происходить в любой приемлемой культуральной среде и в любых приемлемых условиях культивирования, которые способствуют выживанию клеток. Введение рекомбинантных векторов экспрессии в клетку-мишень можно осуществлять *in vivo*, *in vitro* или *ex vivo*.

[0365] В некоторых вариантах осуществления векторы могут быть введены непосредственно в целевую клетку-хозяина. Например, клетки можно приводить в контакт с векторами, содержащими нуклеиновые кислоты субъекта (например, рекомбинантные экспрессионные векторы, содержащие последовательность донорной матрицы и нуклеиновую кислоту, кодирующую вариант CasX и вариант гРНК), так чтобы векторы поглощались клетками. Способы приведения клеток в контакт с полинуклеотидными векторами, которые представляют собой плазмиды, включают электропорацию, трансфекцию хлоридом кальция, микроинъекцию и липофекцию, хорошо известные в данной области. Для доставки вирусного вектора клетки можно приводить в контакт с вирусными частицами, содержащими рассматриваемые вирусные экспрессионные векторы; например, векторы представляют собой вирусные частицы, такие как AAV или VLP, которые содержат полинуклеотиды, кодирующие компоненты CasX : гРНК. Для невирусной доставки векторы или компоненты CasX : гРНК также можно готовить для доставки в липидных наночастицах, причем липидные наночастицы включают в себя, без ограничений, наносферы, липосомы, квантовые точки, полиэтиленгликолевые частицы, гидрогели и мицеллы.

[0366] В некоторых вариантах осуществления редактирование целевой нуклеиновой кислоты происходит *in vitro*, внутри клетки, например в системе клеточной культуры. В некоторых вариантах осуществления редактирование происходит in vivo внутри клетки субъекта, например в клетке животного. В некоторых вариантах осуществления клетка представляет собой эукариотическую клетку. К примерам эукариотических клеток могут относиться клетки, выбранные из группы, состоящей из клетки мыши, клетки крысы, клетки свиньи, клетки собаки и клетки примата, не являющегося человеком. В некоторых вариантах осуществления клетка представляет собой клетку человека. Не имеющие

ограничительного характера примеры клеток включают в себя эмбриональную стволовую клетку, индуцированную плюрипотентную стволовую клетку, зародышевую клетку, фибробласт, олигодендроцит, глиальную клетку, гематопоэтическую стволовую клетку, нейрональную клетку-предшественника, нейрон, мышечную клетку, костную клетку, гепатоцит, клетку поджелудочной железы, клетку сетчатки, раковую клетку, Т-клетку, В-клетку, NK-клетку, эмбриональный кардиомиоцит, миофибробласт, мезенхимальную стволовую клетку, аутотрансплантированный размноженный кардиомиоцит, адипоцит, тотипотентную клетку, плюрипотентную клетку, стволовую клетку крови, миобласт, взрослую стволовую клетку, клетку костного мозга, мезенхимальную клетку, паренхиматозную клетку, эпителиальную клетку, эндотелиальную клетку, мезотелиальную клетку, фибробласты, остеобласты, хондроциты, экзогенную клетку, эндогенную клетку, стволовую клетку, гематопоэтическую стволовую клетку, костномозговую клетку-предшественника, клетку миокарда, скелетную клетку, эмбриональную клетку, недифференцированную клетку, мультипотентную клетку-предшественника, унипотентную клетку-предшественника, моноцит, сердечный миобласт, скелетный миобласт, макрофаг, капиллярную эндотелиальную клетку, ксеногенную клетку, аллогенную клетку или постнатальную стволовую клетку. В альтернативных вариантах осуществления клетка представляет собой прокариотическую клетку.

[0367] В некоторых вариантах осуществления способов модификации целевой нуклеиновой кислоты в клетке *in vitro* или *ex vivo* для индукции расщепления или любой требуемой модификации целевой нуклеиновой кислоты вариант гРНК и/или вариант белка CasX настоящего описания и необязательно последовательность донорной матрицы, независимо от того, вводят ли их в виде нуклеиновых кислот или полипептидов, комплексов RNP, векторов или XDP, вводят в клетки в течение периода от примерно 30 минут до примерно 24 часов или по меньшей мере примерно 1 часа, 1,5 часа, 2 часов, 2,5 часа, 3 часа, 3,5 часа, 4 часа, 5 часов, 6 часов, 7 часов, 8 часов, 12 часов, 16 часов, 18 часов, 20 часов, или в течение любого периода от примерно 30 минут до примерно 24 часов, и введение можно повторять с частотой от примерно раза в день до примерно раза в 4 дня, например, раз в 1,5 дня, раз в 2 дня, раз в 3 дня, или с любой частотой от примерно раза в день до примерно раза в четыре дня. Агент (-ы) может (могут) вводиться в клетки субъекта один или более раз, например, один раз, два, три или более трех раз, и клеткам позволяют инкубироваться с агентом (-ами) в течение некоторого периода времени после каждого события приведения в контакт; например, от 30 минут до примерно 24 часов. В случае способов на основе *in vitro* после периода инкубации с CasX и гРНК (и необязательно донорной матрицей) среду заменяют свежей средой, и клетки культивируют дополнительно.

[0368] В некоторых вариантах осуществления способ включает введение субъекту терапевтически эффективной дозы популяции клеток, модифицированной для коррекции или компенсации мутации гена. В некоторых вариантах осуществления введение модифицированных клеток приводит к экспрессии продукта гена дикого типа или продукта

функционального гена у субъекта. В некоторых вариантах осуществления способа доза общего количества клеток находится в диапазоне от 10^4 или примерно этого значения до 10^9 или примерно этого значения клеток/килограмм (кг) массы тела, например, от 10^5 до 10^6 клеток/кг массы тела, например составляет 1×10^5 клеток/кг или примерно это значение, $1,5 \times 10^5$ клеток/кг, 2×10^5 клеток/кг или 1×10^6 клеток/кг массы тела. Например, в некоторых вариантах осуществления клетки вводят в количестве или в пределах определенного диапазона ошибки от 10^4 или примерно этого значения до 10^9 клеток/килограмм (кг) массы тела или примерно этого значения, например, от 10^5 до 10^6 клеток/кг массы тела, например, составляет 1×10^5 клеток/кг или примерно это значение, $1,5 \times 10^5$ клеток/кг, 2×10^5 клеток/кг или 1×10^6 клеток/кг массы тела. В одном варианте осуществления клетки являются аутологичными по отношению к субъекту, которому должны быть введены клетки. В другом варианте осуществления клетки являются аллогенными по отношению к субъекту, которому должны быть введены клетки. В некоторых случаях субъект выбран из группы, состоящей из мыши, крысы, свиньи и примата, не являющегося человеком. В других случаях субъект представляет собой человека.

VIII. Терапевтические способы

[0369] В другом аспекте настоящее описание относится к способам лечения заболевания или расстройства у нуждающегося в этом субъекта. Для разработки систем для применения в способах лечения субъекта с заболеванием или расстройством, связанным с генетической мутацией, применяли ряд терапевтических стратегий. В некоторых вариантах осуществления модификация целевой нуклеиновой кислоты происходит у субъекта, имеющего мутацию в аллеле гена, причем мутация вызывает заболевание или расстройство у субъекта. В некоторых вариантах осуществления модификация целевой нуклеиновой кислоты изменяет мутацию на аллель дикого типа гена или приводит к экспрессии продукта функционального гена. В некоторых вариантах осуществления модификация целевой нуклеиновой кислоты приводит к нокдауну или нокауту экспрессии аллеля гена, вызывающего заболевание или расстройство у субъекта.

[0370] В некоторых вариантах осуществления способ включает введение субъекту терапевтически эффективной дозы системы, содержащей пару для редактирования генов из варианта нуклеазы CRISPR класса 2, типа V и варианта гидовой РНК, описанных в настоящем документе. В некоторых вариантах осуществления способ лечения включает введение субъекту терапевтически эффективной дозы: i) системы CasX : гРНК, содержащей первый вариант белка CasX и первый вариант гРНК (с нацеливающей последовательностью, комплементарной подлежащей модификации целевой нуклеиновой кислоте) по любому из вариантов осуществления, описанных в настоящем документе; ii) системы CasX : гРНК, содержащей первый белок CasX и первую гРНК с нацеливающей последовательностью, комплементарной целевой нуклеиновой кислоте, и донорной матрицей; iii) нуклеиновой кислотой, кодирующей систему CasX : гРНК, из п. (i) или п. (ii); iv) вектором, содержащим нуклеиновую кислоту по п. (iii), который может представлять

собой AAV по любому из вариантов осуществления, описанных в настоящем документе; v) XDP, содержащей систему CasX : гРНК из п. (i) или п. (ii); или vi) комбинаций двух или более из пп. (i)-(v), причем 1) ген клеток субъекта, на который нацелена первая гРНК, модифицируется (например, осуществляется нокаут или нокдаун) белком CasX (и необязательно донорной матрицей); или 2) ген клеток субъекта, на который нацелена первая гРНК, корректируется или модифицируется белком CasX (и необязательно донорной матрицей) таким образом, чтобы можно было экспрессировать продукт функционального гена. В некоторых вариантах осуществления способ лечения дополнительно включает введение второй или множества гРНК или нуклеиновой кислоты, кодирующей вторую или множество гРНК, причем вторая или множество гРНК имеют нацеливающие последовательности, комплементарные другой или перекрывающейся части последовательности целевой нуклеиновой кислоты в сравнении с первой гРНК. Следует понимать, что в вышеизложенном каждая другая гРНК спарена с белком CasX. В вариантах осуществления, в которых в клетку вводят две или более пары для редактирования генов (например, содержащие две гРНК, имеющие два или более разных спейсеров, которые комплементарны разным последовательностям в пределах той же или другой целевой нуклеиновой кислоты), одновременно могут быть введены или одновременно могут быть доставлены пары для редактирования генов (например, в виде двух RNP и/или векторов). Альтернативно они могут быть введены последовательно, например, сначала вводят первую пару для редактирования генов, а затем вторую пару для редактирования генов, или наоборот.

[0371] В некоторых вариантах осуществления способ лечения включает введение терапевтически эффективной дозы AAV-вектора, кодирующего систему CasX : гРНК, и его вводят субъекту в дозе по меньшей мере примерно 1×10^5 векторных геномов/кг (вг/кг), по меньшей мере примерно 1×10^6 вг/кг, по меньшей мере примерно 1×10^7 вг/кг, по меньшей мере примерно 1×10^8 вг/кг, по меньшей мере примерно 1×10^9 вг/кг, по меньшей мере примерно 1×10^{10} вг/кг, по меньшей мере примерно 1×10^{11} вг/кг, по меньшей мере примерно 1×10^{12} вг/кг, по меньшей мере примерно 1×10^{13} вг/кг, по меньшей мере примерно 1×10^{14} вг/кг, по меньшей мере примерно 1×10^{15} вг/кг или по меньшей мере примерно 1×10^{16} вг/кг. В других вариантах осуществления способа AAV-вектор вводят субъекту в дозе от по меньшей мере примерно 1×10^5 вг/кг до примерно 1×10^{16} вг/кг, от по меньшей мере примерно 1×10^6 вг/кг до примерно 1×10^{15} вг/кг или от по меньшей мере примерно 1×10^7 вг/кг до примерно 1×10^{14} вг/кг. В вышеизложенном AAV-вектор выбран из AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV-Rh74 или AAVRh10. В других вариантах осуществления способ лечения включает введение субъекту терапевтически эффективной дозы XDP, содержащей RNP системы CasX : гРНК. В одном варианте осуществления XDP вводят субъекту в дозе по меньшей мере примерно 1×10^5 частиц/кг, по меньшей мере примерно 1×10^6 частиц/кг, по меньшей мере примерно 1×10^7 частиц/кг, по меньшей мере примерно 1×10^8 частиц/кг, по меньшей мере примерно 1×10^9 частиц/кг, по меньшей мере примерно 1×10^{10} частиц/кг, по меньшей мере примерно 1×10^{11}

частиц/кг, по меньшей мере примерно 1×10^{12} частиц/кг, по меньшей мере примерно 1×10^{13} частиц/кг, по меньшей мере примерно 1×10^{14} частиц/кг, по меньшей мере примерно 1×10^{15} частиц/кг, по меньшей мере примерно 1×10^{16} частиц/кг. В другом варианте осуществления XDP вводят субъекту в дозе от по меньшей мере примерно 1×10^{5} частиц/кг до примерно 1×10^{16} частиц/кг, или от по меньшей мере примерно 1×10^{6} частиц/кг до примерно 1×10^{15} частиц/кг, или от по меньшей мере примерно 1×10^{7} частиц/кг до примерно 1×10^{14} частиц/кг. Вектор или XDP можно вводить путем введения, выбранным из группы, состоящей из интрапаренхиматозного, внутривенного, внутриартериального, внутримышечного, подкожного, интрацеребровентрикулярного, интрацистернального, интратекального, интракраниального, интравитреального, субретинального, интракапсулярного и внутрибрюшинного путей или их комбинаций, причем способ введения представляет собой инъекцию, трансфузию или имплантацию. Введение может осуществляться один раз, два раза или может осуществляться множество раз по схеме: еженедельно, раз в две недели, ежемесячно, раз в четыре месяца, раз в шесть месяцев, раз в год или раз в 2 или 3 года. В некоторых случаях субъект выбран из группы, состоящей из мыши, крысы, свиньи и примата, не являющегося человеком. В других случаях субъект представляет собой человека.

[0372] В некоторых вариантах осуществления способа модификация включает в себя введение одноцепочечного разрыва в целевую нуклеиновую кислоту целевых клеток субъекта. В других случаях модификация включает в себя введение двухцепочечного разрыва в целевую нуклеиновую кислоту целевых клеток субъекта. В некоторых вариантах осуществления модификация вводит одну или более мутаций в целевой нуклеиновой кислоте, например вставку, делецию, замену, дублирование или инверсию одного или более нуклеотидов в гене, причем экспрессия генного продукта в модифицированных клетках субъекта снижается на по меньшей мере примерно 10%, по меньшей мере примерно 20%, по меньшей мере примерно 30%, по меньшей мере примерно 40%, по меньшей мере примерно 50%, по меньшей мере примерно 60%, по меньшей мере примерно 70%, по меньшей мере примерно 80% или по меньшей мере примерно 90% по сравнению с клеткой, которая не была модифицирована. В некоторых случаях ген модифицированных клеток субъекта модифицирован так, что по меньшей мере 70%, по меньшей мере 75%, по меньшей мере 80%, по меньшей мере 85%, по меньшей мере 90% или по меньшей мере 95% модифицированных клеток не экспрессируют обнаруживаемый уровень генного продукта. В некоторых вариантах осуществления введение терапевтически эффективного количества системы CasX : гРНК для нокдауна или нокаута экспрессии генного продукта субъекту с заболеванием приводит к предотвращению или облегчению основного заболевания, так что у субъекта наблюдается улучшение, несмотря на то что субъект все еще может страдать основным заболеванием. В других вариантах осуществления ген может быть модифицирован репарационными механизмами NHEJ хозяина или применяется в комбинации с донорной матрицей, которую вставляют с помощью механизмов HDR или HITI с целью удаления, коррекции или компенсации мутации в клетках субъекта, так что

экспрессия генного продукта дикого типа или продукта функционального гена в модифицированных клетках повышена по меньшей мере на примерно 50%, по меньшей мере на примерно 60%, по меньшей мере на примерно 70%, по меньшей мере на примерно 80%, по меньшей мере на примерно 90% или по меньшей мере на примерно 95% по сравнению с клеткой, которая не была модифицирована. В некоторых вариантах осуществления введение терапевтически эффективного количества CasX-гРНК приводит к улучшению по меньшей мере одного клинически значимого параметра заболевания.

[0373] В некоторых случаях нуклеиновая кислота настоящего описания (например, рекомбинантный вектор экспрессии настоящего описания), или вариант CasX, или вариант гРНК могут быть покрыты липидами в виде организованной структуры, такой как мицелла, липосома или липидная наночастица. Когда организованная структура образует комплекс с ДНК, она называется липоплексом. Существует три типа липидов: анионные (отрицательно заряженные), нейтральные и катионные (положительно заряженные). Липоплексы, в которых используются катионные липиды, обладают доказанной пригодностью для переноса генов. Катионные липиды из-за их положительного заряда естественным образом образуют комплекс с отрицательно заряженной ДНК. Также вследствие своего заряда они взаимодействуют с клеточной мембраной. Далее происходит эндоцитоз липоплекса, и ДНК высвобождается в цитоплазму. Катионные липиды также обеспечивают защиту от деградации ДНК клеткой.

[0374] В некоторых случаях нуклеиновая кислота описания (например, экспрессионный вектор) включает в себя место вставки интересующей гидовой последовательности. Например, нуклеиновая кислота может включать в себя место вставки интересующей гидовой последовательности, где место вставки непосредственно смежно с нуклеотидной последовательностью, кодирующей часть варианта гРНК (например, каркасную область), которая не изменяется при изменении гидовой последовательности для гибридизации с требуемой целевой последовательностью. Таким образом, в некоторых случаях экспрессионный вектор включает в себя нуклеотидную последовательность, кодирующую гРНК, за исключением того, что часть, кодирующая часть спейсерной последовательности гРНК, представляет собой последовательность вставки (место вставки). Место вставки представляет собой любую нуклеотидную последовательность, используемую для введения спейсера в требуемую последовательность. Специалистам в данной области известны «места вставки» для применения с различными технологиями, и можно использовать любое удобное место вставки. Место вставки может быть предназначено для любого способа манипулирования последовательностями нуклеиновых кислот. Например, в некоторых случаях место вставки представляет собой сайт множественного клонирования (MCS) (например, сайт, содержащий одну или более последовательностей распознавания рестрикционным ферментом), сайт независимого от лигирования клонирования, сайт для клонирования на основе рекомбинации (например, рекомбинации на основе сайтов ATT), нуклеотидную последовательность, распознаваемую технологией на основе CRISPR/Cas (например, Cas9), и т. п.

IX. Клетки

[0375] В других дополнительных вариантах осуществления в настоящем документе предложены клетки, содержащие компоненты любой из систем CasX : гРНК, описанных в настоящем документе. В некоторых вариантах осуществления клетки содержат любой из вариантов осуществления варианта гРНК, описанных в настоящем документе, и дополнительно содержат спейсер, комплементарный целевой нуклеиновой кислоте. В некоторых вариантах осуществления клетки дополнительно содержат вариант CasX, описанный в настоящем документе (например, последовательности из таблиц 3 и 7). В других вариантах осуществления клетки содержат RNP по любому из вариантов осуществления CasX : гРНК, описанных в настоящем документе. В других вариантах осуществления в описании предложены клетки, содержащие векторы, кодирующие системы CasX : гРНК по любому из вариантов осуществления, описанных в настоящем документе. В других вариантах осуществления клетки содержат целевую нуклеиновую кислоту, которая была отредактирована вариантами осуществления CasX : гРНК, описанными в настоящем документе; либо для коррекции мутации (нокин), либо для нокдауна или нокаута дефектного гена.

[0376] В некоторых вариантах осуществления клетка представляет собой модифицированную клетку (например, генетически модифицированную клетку), содержащую нуклеиновую кислоту, имеющую нуклеотидную последовательность, кодирующую вариант белка CasX описания. В некоторых вариантах осуществления генетически модифицированная клетка генетически модифицирована мРНК, содержащей нуклеотидную последовательность, кодирующую вариант белка CasX. В некоторых вариантах осуществления клетка генетически модифицирована рекомбинантным вектором экспрессии, содержащим: a) нуклеотидную последовательность, кодирующую вариант белка CasX настоящего описания; и b) нуклеотидную последовательность, кодирующую гРНК описания, и необязательно содержащим нуклеотидную последовательность, содержащую донорную матрицу. В некоторых случаях такие клетки используют для получения индивидуальных компонентов или RNP систем CasX : гРНК для применения при редактировании целевой нуклеиновой кислоты. В других случаях клетки, которые были генетически модифицированы таким способом, можно вводить субъекту для таких целей, как генная терапия; например, для лечения заболевания или состояния, вызванного генетической мутацией или дефектом.

[0377] Клетка, которая может служить реципиентом для варианта белка CasX и/или гРНК настоящего описания, и/или нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую вариант белка CasX и/или вариант гРНК, может представлять собой любую из множества клеток, включая, например, клетки *in vitro*; клетки *in vivo*; клетки *ex vivo*; первичные клетки; клетки иммортализованной клеточной линии; раковые клетки; клетки животных; клетки растений; клетки водорослей; клетки грибов и т. д. Клетка может быть реципиентом RNP CasX настоящего описания. Клетка может представлять собой реципиента одного компонента системы CasX настоящего описания.

Клетка может представлять собой реципиента вектора, кодирующего CasX, гРНК и необязательно донорную матрицу систем CasX : гРНК по любому из вариантов осуществления, описанных в настоящем документе.

[0378] Не имеющие ограничительного характера примеры клеток, которые могут служить клетками-хозяевами для продукции систем CasX : гРНК, описанных в настоящем документе, включают в себя прокариотические клетки (например, *E coli*) и эукариотические клетки (например, клетки-фибробласты почки новорожденного хомяка (BHK), клетки эмбриональной почки человека 293 (HEK293), клетки эмбриональной почки человека 293T (HEK293T), клетки NS0, клетки SP2/0, клетки миеломы YO, мышиные клетки миеломы P3X63, клетки PER, клетки PER.C6, клетки гибридомы, клетки NIH3T3, клетки, происходящие от CV-1 (обезьяньих), с генетическим материалом SV40 (COS), клетки HeLa, клетки яичника китайского хомячка (CHO) или дрожжевые клетки, или другие эукариотические клетки, известные в данной области, приемлемые для получения рекомбинантных продуктов.

[0379] В некоторых вариантах осуществления в описании предложены популяции клеток, модифицированных для введения субъекту с целью лечения заболевания или расстройства. Такие клетки могут быть аутологичными по отношению к субъекту, которому вводят указанную (-ые) клетку (-и). В других вариантах осуществления клетки являются аллогенными по отношению к субъекту, которому вводят указанную (-ые) клетку (-и). Клетка может представлять собой клетку животного или полученную из клетки животного. Клетка может представлять собой клетку млекопитающего или полученную из клетки млекопитающего. Клетка может представлять собой клетку грызуна или полученную из клетки грызуна, такого как крыса или мышь. Клетка может представлять собой клетку примата, не являющегося человеком, или полученную из клетки примата, не являющегося человеком. Клетка может представлять собой человеческую клетку или полученную из человеческой клетки. Приемлемые клетки могут включать в себя в некоторых вариантах осуществления стволовую клетку (например, эмбриональную стволовую (ES) клетку, индуцированную плюрипотентную стволовую (iPS) клетку); зародышевую клетку (например, ооцит, сперматозоид, оогоний, сперматогоний и т. д.); соматическую клетку, например, фибробласт, олигодендроцит, глиальную клетку, гематопоэтическую стволовую клетку, нейрональную клетку-предшественника, нейрон, мышечную клетку, костную клетку, гепатоцит, клетку поджелудочной железы, клетку сетчатки, раковую клетку, Т-клетку, В-клетку, эмбриональный кардиомиоцит, миофибробласт, мезенхимальную стволовую клетку, аутотрансплантированный размноженный кардиомиоцит, адипоцит, тотипотентную клетку, плюрипотентную клетку, стволовую клетку крови, миобласт, взрослую стволовую клетку, клетку костного мозга, мезенхимальную клетку, паренхиматозную клетку, эпителиальную клетку, эндотелиальную клетку, мезотелиальную клетку, фибробласты, остеобласты, хондроциты, экзогенную клетку, эндогенную клетку, стволовую клетку, гематопоэтическую стволовую клетку, костномозговую клетку-предшественник, клетку миокарда, скелетную клетку,

эмбриональную клетку, недифференцированную клетку, мультипотентную клетку-предшественник, унипотентную клетку-предшественник, моноцит, сердечный миобласт, скелетный миобласт, макрофаг, капиллярную эндотелиальную клетку, ксеногенную клетку, аллогенную клетку и постнатальную стволовую клетку. В некоторых вариантах осуществления клетка представляет собой иммунную клетку. В некоторых случаях иммунная клетка представляет собой Т-клетку, В-клетку, моноцит, натуральную киллерную клетку, дендритную клетку или макрофаг. В некоторых случаях иммунная клетка представляет собой цитотоксическую Т-клетку. В некоторых случаях иммунная клетка представляет собой хелперную Т-клетку. В некоторых случаях иммунная клетка представляет собой регуляторную Т-клетку (Treg). В некоторых случаях клетка экспрессирует химерный антигенный рецептор (Car-T). В некоторых вариантах осуществления клетка представляет собой стволовую клетку. К стволовым клеткам могут относиться, например, взрослые стволовые клетки. Взрослые стволовые клетки также могут называться соматическими стволовыми клетками. В некоторых вариантах осуществления стволовая клетка представляет собой гематопоэтическую стволовую клетку (HSC), нейрональную стволовую клетку или мезенхимальную стволовую клетку. В других вариантах осуществления стволовая клетка представляет собой мезенхимальную стволовую клетку (MSC). MSC, первоначально полученные из эмбриональной мезодермы и выделенные из костного мозга взрослого человека, могут дифференцироваться с образованием мышц, кости, хряща, жира, стромы мозга и сухожилия. Способы выделения MSC известны в данной области; и для получения MSC может использоваться любой известный способ.

X. Производимые наборы и изделия

[0380] В другом аспекте в настоящем документе предложены наборы, содержащие белок CasX одну или множество гРНК по любому из вариантов осуществления описания и приемлемый контейнер (например, пробирку, флакон или планшет). В некоторых вариантах осуществления набор содержит вариант гРНК описания или референсную гРНК с SEQ ID NO: 5 или SEQ ID NO: 4. Примеры вариантов гРНК, которые могут быть включены, включают в себя последовательность по любой из SEQ ID NO: 2238-XX, указанных в таблице 2.

[0381] В некоторых вариантах осуществления набор содержит вариант белка CasX описания (например, последовательность из таблицы 3 и 7) или референсный белок CasX с SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3. В примерах осуществления набор описания содержит вариант CasX по любой из SEQ ID NO: 247-592 и 1147-1231. В дополнительных примерах осуществления набор описания содержит вариант CasX по любой из SEQ ID NO: 270-592 и 1147-1231. В дополнительных примерах осуществления набор описания содержит вариант CasX по любой из SEQ ID NO: 415-592 и 1147-1231.

[0382] В некоторых вариантах осуществления набор содержит гРНК или вектор, кодирующий гРНК, причем гРНК содержит последовательность, выбранную из группы, состоящей из SEQ ID NO: 2101-2332 и 2353-2398. В некоторых вариантах осуществления

гРНК содержит последовательность, выбранную из группы, состоящей из SEQ ID NO: 2238-2332 и 2353-2398. В некоторых вариантах осуществления гРНК содержит последовательность, выбранную из группы, состоящей из SEQ ID NO: 2281-2332 и 2353-2398. В некоторых вариантах осуществления гРНК содержит последовательность, выбранную из группы, состоящей из SEQ ID NO: 2236, 2237, 2238, 2241, 2244, 2248, 2249 и 2259-2280. В некоторых вариантах осуществления гРНК содержит последовательность, выбранную из группы, состоящей из любой из последовательностей, представленных в таблице 2.

[0383] В определенных вариантах осуществления в настоящем документе предложены наборы, содержащие пару для редактирования из белка CasX и гРНК, содержащую вариант белка CasX из таблицы 3 и 7 и вариант гРНК, описанный в настоящем документе (например, последовательность из таблицы 2). В примерах осуществления набор описания содержит пару для редактирования из CasX и гРНК, причем вариант CasX содержит любую из SEQ ID NO: 247-592 или 1147-1231. В дополнительных примерах осуществления набор описания содержит пару для редактирования из CasX и гРНК, причем вариант CasX содержит любую из SEQ ID NO: 270-592 и 1147-1231. В дополнительных примерах осуществления набор описания содержит пару для редактирования из CasX и гРНК, причем вариант CasX содержит любую из SEQ ID NO: 415-592 и 1147-1231. В некоторых вариантах осуществления гРНК пары для редактирования генов содержит любую из SEQ ID NO: 2101-2332 или 2353-2398. В некоторых вариантах осуществления гРНК из пары для редактирования генов содержит любую из SEQ ID NO: 2238-2332 или 2353-2398. В некоторых вариантах осуществления гРНК из пары для редактирования генов содержит любую из SEQ ID NO: 2281-2332 или 2353-2398. В некоторых вариантах осуществления гРНК пары для редактирования генов содержит любую из SEQ ID NO: 2236, 2237, 2238, 2241, 2244, 2248, 2249 или 2259-2280.

[0384] В некоторых вариантах осуществления набор дополнительно содержит буфер, ингибитор нуклеазы, ингибитор протеазы, липосому, терапевтический агент, метку, реагент для визуализации метки или любую комбинацию вышеизложенного. В некоторых вариантах осуществления набор дополнительно содержит фармацевтически приемлемый носитель, разбавитель или эксципиент.

[0385] В некоторых вариантах осуществления набор содержит соответствующие контрольные композиции для применения при редактировании генов и инструкции по применению.

[0386] В некоторых вариантах осуществления набор содержит вектор, содержащий последовательность, кодирующую вариант белка CasX описания, вариант гРНК описания, необязательно донорную матрицу или их комбинацию.

[0387] В настоящем описании представлены многочисленные примеры конфигураций, способов, параметров и т. п. Однако следует понимать, что такое описание не предназначено для ограничения объема настоящего описания, а, напротив, представлено в качестве описания примеров осуществления. Варианты осуществления настоящего

объекта изобретения, описанные выше, могут быть полезными отдельно или в комбинации с одним или более другими аспектами или вариантами осуществления. Не ограничивая приведенное выше описание, ниже приведены некоторые не имеющие ограничительного характера варианты осуществления настоящего описания. Как будет очевидно специалистам в данной области после прочтения настоящего описания, каждый из индивидуально пронумерованных вариантов осуществления может использоваться или комбинироваться с любым из предшествующих или последующих индивидуально пронумерованных вариантов осуществления. Это предполагает поддержку всех таких комбинаций вариантов осуществления, не ограничиваясь комбинациями вариантов осуществления, явно представленными ниже:

ВАРИАНТЫ ОСУЩЕСТВЛЕНИЯ В ВИДЕ НУМЕРОВАННОГО СПИСКА

[0388] Изобретение может быть определено путем ссылки на следующие иллюстративные варианты осуществления, перечисленные ниже в виде нумерованного списка.

Набор I

[0389] Вариант осуществления 1. Вариант референсного белка CasX (вариант CasX), причем:

a. вариант CasX содержит по меньшей мере одну модификацию в референсном белке CasX; и

b. вариант CasX демонстрирует по меньшей мере одну улучшенную характеристику по сравнению с референсным белком CasX, причем необязательно вариант содержит последовательность, выбранную из представленных в таблицах 3 и 8.

[0390] Вариант осуществления 2. Вариант CasX по варианту осуществления 1, причем улучшенная характеристика варианта CasX выбрана из группы, состоящей из: улучшенного свертывания варианта CasX; улучшенной аффинности связывания с гидовой нуклеиновой кислотой (гНК); улучшенной аффинности связывания с целевой ДНК; улучшенной способности использовать более широкий спектр из одной или более PAM-последовательностей, включая ATC, CTC, GTC или TTC, при редактировании целевой ДНК; улучшенного раскручивания целевой ДНК; повышенной активности редактирования; улучшенной эффективности редактирования; улучшенной специфичности редактирования; повышенной нуклеазной активности; повышенной загрузки целевой цепи для двухцепочечного расщепления; сниженной загрузки целевой цепи для одноцепочечного никирования; сниженного нецелевого расщепления; улучшенного связывания с нецелевой цепью ДНК; улучшенной стабильности белка; улучшенной растворимости белка; улучшенной стабильности комплекса белок : гРНК (RNP); улучшенной растворимости комплекса белок : гРНК; улучшенного выхода белка; улучшенной экспрессии белка; улучшенной характеристики слияния или их комбинации.

[0391] Вариант осуществления 3. Вариант Cas X по варианту осуществления 1 или 2, причем по меньшей мере одна модификация содержит:

a. по меньшей мере одну замену аминокислоты в домене варианта CasX;

b. по меньшей мере одну делецию аминокислоты в домене варианта CasX;

c. по меньшей мере одну вставку аминокислоты в домене варианта CasX

d. полную или частичную замену доменом из другого CasX;

e. полную или частичную делецию домена варианта CasX или

f. или любую комбинацию пп. (a)-(d).

[0392] Вариант осуществления 4. Вариант CasX по любому из вариантов осуществления 1-3, причем референсный белок CasX содержит последовательность SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 270 или SEQ ID NO: 336.

[0393] Вариант осуществления 5. Вариант CasX по любому из вариантов осуществления 1-4, причем по меньшей мере одна модификация находится в домене, выбранном из:

a. домена связывания с нецелевой цепью (NTSB);

b. домена загрузки целевой цепи (TSL);

c. домена спиральный I;

d. домена спиральный II;

e. олигонуклеотид-связывающего домена (OBD) или

f. расщепляющего ДНК домена RuvC.

[0394] Вариант осуществления 6. Вариант CasX по варианту осуществления 5, содержащий по меньшей мере одну модификацию в домене NTSB.

[0395] Вариант осуществления 7. Вариант CasX по варианту осуществления 5, содержащий по меньшей мере одну модификацию в домене TSL.

[0396] Вариант осуществления 8. Вариант CasX по варианту осуществления 5, содержащий по меньшей мере одну модификацию в домене спиральный I.

[0397] Вариант осуществления 9. Вариант CasX по любому из вариантов осуществления 5-8, содержащий по меньшей мере одну модификацию в домене спиральный II.

[0398] Вариант осуществления 10. Вариант CasX по варианту осуществления 5, содержащий по меньшей мере одну модификацию в домене OBD.

[0399] Вариант осуществления 11. Вариант CasX по варианту осуществления 5, содержащий по меньшей мере одну модификацию в расщепляющем ДНК домене RuvC.

[0400] Вариант осуществления 12. Вариант CasX по любому из вариантов осуществления 5-11, причем модификация приводит к получению улучшенной способности редактировать целевую ДНК.

[0401] Вариант осуществления 13. Вариант CasX по любому из вариантов осуществления 1-12, причем вариант CasX способен образовывать рибонуклеопротеиновый комплекс (RNP) с гидовой нуклеиновой кислотой (гНК).

[0402] Вариант осуществления 14. Вариант CasX по любому из вариантов осуществления 1-13, причем по меньшей мере одна модификация содержит:

a. замену от 1 до 100 последовательных или непоследовательных аминокислот в варианте CasX;

b. делецию от 1 до 100 последовательных или непоследовательных аминокислот в варианте CasX;

c. вставку от 1 до 100 последовательных или непоследовательных аминокислот в CasX или

d. любую комбинацию пп. (a)-(c).

[0403] Вариант осуществления 15. Вариант CasX по варианту осуществления 14, причем по меньшей мере одна модификация содержит:

a. замену 5-10 последовательных или непоследовательных аминокислот в варианте CasX;

b. делецию 1-5 последовательных или непоследовательных аминокислот в варианте CasX;

c. вставку 1-5 последовательных или непоследовательных аминокислот в CasX или

d. любую комбинацию пп. (a)-(c).

[0404] Вариант осуществления 16. Вариант CasX по любому из вариантов осуществления 1-15, причем вариант CasX содержит две или более модификаций в одном домене.

[0405] Вариант осуществления 17. Вариант CasX по любому из вариантов осуществления 1-16, причем вариант CasX содержит модификации в двух или более доменах.

[0406] Вариант осуществления 18. Вариант CasX по любому из вариантов осуществления 1-15, содержащий по меньшей мере одну модификацию области несмежных аминокислотных остатков варианта CasX, которые образуют канал, в котором происходит образование комплекса гРНК : целевая ДНК с вариантом CasX.

[0407] Вариант осуществления 19. Вариант CasX по любому из вариантов осуществления 1-15, содержащий по меньшей мере одну модификацию области несмежных аминокислотных остатков варианта CasX, которые образуют поверхность взаимодействия, связывающуюся с гНК.

[0408] Вариант осуществления 20. Вариант CasX по любому из вариантов осуществления 1-15, содержащий по меньшей мере одну модификацию области несмежных аминокислотных остатков варианта CasX, которые образуют канал, связывающийся с нецелевой цепью ДНК.

[0409] Вариант осуществления 21. Вариант CasX по любому из вариантов осуществления 1-15, содержащий по меньшей мере одну модификацию области несмежных аминокислотных остатков варианта CasX, которые образуют поверхность взаимодействия, связывающуюся со смежным с протоспейсером мотивом (PAM) целевой ДНК.

[0410] Вариант осуществления 22. Вариант CasX по любому из вариантов осуществления 1-15, содержащий по меньшей мере одну модификацию области несмежных поверхностных аминокислотных остатков варианта CasX.

[0411] Вариант осуществления 23. Вариант CasX по любому из вариантов осуществления 1-15, содержащий по меньшей мере одну модификацию области несмежных

аминокислотных остатков, которые образуют сердцевину посредством гидрофобной упаковки в домене варианта CasX.

[0412] Вариант осуществления 24. Вариант CasX по любому из вариантов осуществления 18-23, в котором модификация представляет собой одно или более из делеции, вставки или замены одной или более аминокислот области.

[0413] Вариант осуществления 25. Вариант CasX по любому из вариантов осуществления 18-23, в котором от 2 до 15 аминокислотных остатков области варианта CasX замещены заряженными аминокислотами.

[0414] Вариант осуществления 26. Вариант CasX по любому из вариантов осуществления 18-23, в котором от 2 до 15 аминокислотных остатков области варианта CasX замещены полярными аминокислотами.

[0415] Вариант осуществления 27. Вариант CasX по любому из вариантов осуществления 18-23, в котором от 2 до 15 аминокислотных остатков области варианта CasX замещены аминокислотами, которые образуют стопку с основаниями ДНК или РНК.

[0416] Вариант осуществления 28. Вариант CasX по любому из вариантов осуществления 1-5, имеющий последовательность, выбранную из группы, состоящей из последовательностей в таблице 3, или последовательность, имеющую идентичность последовательности с ней по меньшей мере примерно 50%, по меньшей мере примерно 60%, по меньшей мере примерно 70%, по меньшей мере примерно 80%, по меньшей мере примерно 90%, или по меньшей мере примерно 95%, или по меньшей мере примерно 96%, или по меньшей мере примерно 97%, или по меньшей мере примерно 98%, или по меньшей мере примерно 99%.

[0417] Вариант осуществления 29. Вариант CasX по любому из вариантов осуществления 1-5, дополнительно содержащий замену NTSB и/или домена спиральный 1b из другого CasX.

[0418] Вариант осуществления 30. Вариант CasX по варианту осуществления 29, в котором замещенный NTSB и/или домен спиральный 1b получен из референсного CasX с SEQ ID NO: 1.

[0419] Вариант осуществления 31. Вариант CasX по любому из вариантов осуществления 1-30, дополнительно содержащий один или более сигналов ядерной локализации (NLS).

[0420] Вариант осуществления 32. Вариант CasX по варианту осуществления 31, в котором один или более NLS выбраны из группы последовательностей, состоящей из PKKKRKV (SEQ ID NO: 352), KRPAATKKAGQAKKKK (SEQ ID NO: 353), PAAKRVKLD (SEQ ID NO: 354), RQRRNELKRSP (SEQ ID NO: 355), NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY (SEQ ID NO: 356), RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV (SEQ ID NO: 357), VSRKRPRP (SEQ ID NO: 358), PPKKARED (SEQ ID NO: 35(), PQPKKKPL (SEQ ID NO: 360), SALIKKKKKMAP (SEQ ID NO: 361), DRLRR (SEQ ID NO: 362), PKQKKRK (SEQ ID NO: 363), RKLKKKIKKL (SEQ ID NO: 364), REKKKFLKRR (SEQ ID NO: 365),

KRKGDEVDGVDEVAKKKSKK (SEQ ID NO: 366), RKCLQAGMNLEARKTKK (SEQ ID NO: 367), PRPRKIPR (SEQ ID NO: 368), PPRKKRTVV (SEQ ID NO: 369), NLSKKKKRKREK (SEQ ID NO: 370), RRPSRPFRKP (SEQ ID NO: 371), KRPRSPSS (SEQ ID NO: 372), KRGINDRNFWRGENERKTR (SEQ ID NO: 373), PRPPKMARYDN (SEQ ID NO: 374), KRSFSKAF (SEQ ID NO: 375), KLKIKRPVK (SEQ ID NO: 376), PKKKRKVPPPPAAKRVKLD (SEQ ID NO: 377), PKTRRRPRRSQRKRPPT (SEQ ID NO: 378), SRRRKANPTKLSENAKKLAKEVEN (SEQ ID NO: 379), KTRRRPRRSQRKRPPT (SEQ ID NO: 380), RRKKRRPRRKKRR (SEQ ID NO: 381), PKKKSRKPKKKSRK (SEQ ID NO: 382), HKKKHPDASVNFSEFSK (SEQ ID NO: 383), QRPGPYDRPQRPGPYDRP (SEQ ID NO: 384), LSPSLSPLLSPSLSPL (SEQ ID NO: 385), RGKGGKGLGKGGAKRHRK (SEQ ID NO: 386), PKRGRGRPKRGRGR (SEQ ID NO: 387) и PKKKRKVPPPPKKKRKV (SEQ ID NO: 389).

[0421] Вариант осуществления 33. Вариант CasX по варианту осуществления 31 или варианту осуществления 32, в котором один или более NLS расположены на C-конце белка CasX или рядом с ним.

[0422] Вариант осуществления 34. Вариант CasX по варианту осуществления 31 или варианту осуществления 32, в котором один или более NLS расположены на N-конце белка CasX или рядом с ним.

[0423] Вариант осуществления 35. Вариант CasX по варианту осуществления 31 или варианту осуществления 32, содержащий по меньшей мере два NLS, причем по меньшей мере два NLS расположены на N-конце белка CasX или рядом с ним и на C-конце белка CasX или рядом с ним.

[0424] Вариант осуществления 36. Вариант CasX по любому из вариантов осуществления 2-35, причем кратность улучшения одной или более из улучшенных характеристик варианта CasX составляет от по меньшей мере примерно 1,1 до примерно 100 раз или более относительно референсного белка CasX с SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3.

[0425] Вариант осуществления 37. Вариант CasX по любому из вариантов осуществления 2-35, причем кратность улучшения одной или более из улучшенных характеристик варианта CasX составляет по меньшей мере примерно 1,1, по меньшей мере примерно 2, по меньшей мере примерно 10, по меньшей мере примерно 100 раз или более относительно референсного белка CasX с SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 270 или SEQ ID NO: 336.

[0426] Вариант осуществления 38. Вариант CasX по любому из вариантов осуществления 2-37, причем улучшенная характеристика включает в себя эффективность редактирования, и при этом вариант CasX имеет кратность улучшения эффективности редактирования от 1,1 до 100 раз по сравнению с референсным белком CasX с SEQ ID NO: 270 или SEQ ID NO: 336.

[0427] Вариант осуществления 39. Вариант CasX по любому из вариантов осуществления 1-38, причем RNP, содержащий вариант CasX, демонстрирует более

высокие эффективность редактирования и/или связывание целевой последовательности в целевой ДНК, когда любая из PAM-последовательностей TTC, ATC, GTC или CTC расположена на расстоянии 1 нуклеотида от 5'-конца нецелевой цепи протоспейсера, имеющей идентичность с нацеливающей последовательностью гРНК, в системе клеточного анализа по сравнению с эффективностью редактирования и/или связыванием RNP, содержащего референсный белок CasX, в сопоставимой системе анализа.

[0428] Вариант осуществления 40. Вариант CasX по варианту осуществления 39, причем PAM-последовательность представляет собой TTC.

[0429] Вариант осуществления 41. Вариант CasX по варианту осуществления 39, причем PAM-последовательность представляет собой ATC.

[0430] Вариант осуществления 42. Вариант CasX по варианту осуществления 39, причем PAM-последовательность представляет собой CTC.

[0431] Вариант осуществления 43. Вариант CasX по варианту осуществления 39, причем PAM-последовательность представляет собой GTC.

[0432] Вариант осуществления 44. Вариант CasX по любому из вариантов осуществления 39, причем улучшенные эффективность редактирования и/или связывание с целевой ДНК RNP, содержащего вариант CasX, имеют кратность улучшения от по меньшей мере примерно 1,1 до примерно 100 раз относительно RNP, содержащего референсный CasX.

[0433] Вариант осуществления 45. Вариант CasX по любому из вариантов осуществления 1-44, содержащий от 400 до 2000 аминокислот.

[0434] Вариант осуществления 46. Вариант CasX по любому из вариантов осуществления 1-45, причем вариант белка CasX содержит нуклеазный домен, обладающий никазной активностью.

[0435] Вариант осуществления 47. Вариант CasX по любому из вариантов осуществления 1-45, причем вариант белка CasX содержит нуклеазный домен, обладающий активностью двухцепочечного расщепления.

[0436] Вариант осуществления 48. Вариант CasX по любому из вариантов осуществления 1-45, причем белок CasX представляет собой каталитически неактивный белок CasX (dCasX), и при этом dCasX и гНК сохраняют способность связываться с целевой ДНК.

[0437] Вариант осуществления 49. Вариант CasX по варианту осуществления 48, причем dCasX содержит мутацию по остаткам:

a. D672, и/или E769, и/или D935, соответствующим белку CasX с SEQ ID NO: 1; или

b. D659, и/или E756, и/или D922, соответствующим белку CasX с SEQ ID NO: 2.

[0438] Вариант осуществления 50. Вариант CasX по варианту осуществления 49, в котором мутация представляет собой замену остатка на аланин.

[0439] Вариант осуществления 51. Вариант CasX по любому из вариантов осуществления 1-50, содержащий первый домен из первого белка CasX и второй домен из второго белка CasX, отличного от первого белка CasX.

[0440] Вариант осуществления 52. Вариант CasX по варианту осуществления 51, в котором первый домен выбран из группы, состоящей из доменов NTSB, TSL, спирального I, спирального II, OBD и RuvC.

[0441] Вариант осуществления 53. Вариант CasX по варианту осуществления 51, в котором второй домен выбран из группы, состоящей из доменов NTSB, TSL, спирального I, спирального II, OBD и RuvC.

[0442] Вариант осуществления 54. Вариант CasX по любому из вариантов осуществления 51-53, в котором первый и второй домены не являются одним и тем же доменом.

[0443] Вариант осуществления 55. Вариант CasX по любому из вариантов осуществления 1-50, содержащий по меньшей мере один химерный домен, который содержит первую часть из первого белка CasX и вторую часть из второго белка CasX, отличного от первого белка CasX.

[0444] Вариант осуществления 56. Вариант CasX по варианту осуществления 55, в котором по меньшей мере один химерный домен выбран из группы, состоящей из доменов NTSB, TSL, спирального I, спирального II, OBD и RuvC.

[0445] Вариант осуществления 57. Вариант CasX по варианту осуществления 56, в котором по меньшей мере один химерный домен содержит химерный домен RuvC.

[0446] Вариант осуществления 58. Вариант CasX по любому из вариантов осуществления 1-57, содержащий гетерологичный белок или его домен, слитый с CasX.

[0447] Вариант осуществления 59. Вариант CasX по варианту осуществления 58, в котором гетерологичный белок или его домен представляет собой редактор оснований.

[0448] Вариант осуществления 60. Вариант CasX по варианту осуществления 59, в котором редактор оснований представляет собой аденозиндезаминазу, цитозиндезаминазу или гуанинoксидазу.

[0449] Вариант осуществления 61. Вариант каркаса референсной гидовой нуклеиновой кислоты (вариант гНК), способный связывать референсный белок CasX или вариант CasX, причем:

a. вариант гНК содержит по меньшей мере одну модификацию по сравнению с каркасной последовательностью референсной гидовой нуклеиновой кислоты и

b. вариант гНК демонстрирует одну или более улучшенных характеристик по сравнению с каркасом референсной гидовой нуклеиновой кислоты.

[0450] Вариант осуществления 62. Вариант гНК по варианту осуществления 61, причем одна или более улучшенных характеристик выбраны из группы, состоящей из: улучшенной стабильности; улучшенной растворимости; улучшенной транскрипции гНК; улучшенной стойкости к нуклеазной активности; повышенной скорости свертывания гНК; сниженного образования побочных продуктов при свертывании; повышенного продуктивного свертывания; улучшенной аффинности связывания с белком CasX; улучшенной аффинности связывания с целевой ДНК при нахождении в комплексе с белком CasX; улучшенного редактирования генов при нахождении в комплексе с белком CasX;

улучшенной специфичности редактирования при нахождении в комплексе с белком CasX и улучшенной способности использовать больший спектр из одной или более PAM-последовательностей, включая ATC, CTC, GTC или TTC, во время редактирования целевой ДНК при нахождении в комплексе с белком CasX.

[0451] Вариант осуществления 63. Вариант гНК по варианту осуществления 61 или 62, в котором каркас референсного гида содержит последовательность, выбранную из группы, состоящей из последовательностей SEQ ID NO: 4-16 или SEQ ID NO: 2238 или SEQ ID NO: 2239.

[0452] Вариант осуществления 64. Вариант гНК по любому из вариантов осуществления 61-63, в котором по меньшей мере одна модификация содержит:

a. по меньшей мере одну замену нуклеотида в области варианта гНК;

b. по меньшей мере одну делецию нуклеотида в области варианта гНК;

c. по меньшей мере одну вставку нуклеотида в области варианта гНК;

d. полную или частичную замену области варианта гНК;

e. полную или частичную делецию области варианта гНК; или

f. или любую комбинацию пп. (a)-(d).

[0453] Вариант осуществления 65. Вариант гНК по варианту осуществления 64, в котором область варианта гНК выбрана из группы, состоящей из удлиненной стеблевой петли, каркасной стеблевой петли, триплекса и псевдоузла.

[0454] Вариант осуществления 66. Вариант гНК по варианту осуществления 65, в котором каркасный стебель дополнительно содержит пузырек.

[0455] Вариант осуществления 67. Вариант гНК по варианту осуществления 65 или варианту осуществления 66, в котором каркас дополнительно содержит триплексную петлевую область.

[0456] Вариант осуществления 68. Вариант гНК по любому из вариантов осуществления 65-67, в котором каркас дополнительно содержит 5'-неструктурированную область.

[0457] Вариант осуществления 69. Вариант гНК по любому из вариантов осуществления 64-68, в котором по меньшей мере одна модификация содержит:

a. замену 1-15 последовательных или не последовательных нуклеотидов в варианте гНК в одной или более областях;

b. делецию 1-10 последовательных или не последовательных нуклеотидов в варианте гНК в одной или более областях;

c. вставку 1-10 последовательных или не последовательных нуклеотидов в варианте гНК в одной или более областях;

d. замену каркасной стеблевой петли или удлиненной стеблевой петли на последовательность стеблевой петли РНК из гетерологичного РНК-источника с проксимальными 5’- и 3’-концами или

e. любую комбинацию пп. (a)-(d).

[0458] Вариант осуществления 70. Вариант гНК по любому из вариантов

осуществления 61-69, содержащий область удлиненной стеблевой петли, содержащую по меньшей мере 10, по меньшей мере 100, по меньшей мере 500, по меньшей мере 1000 или по меньшей мере 10 000 нуклеотидов.

[0459] Вариант осуществления 71. Вариант гНК по варианту осуществления 69, в котором гетерологичная последовательность стеблевой петли РНК повышает стабильность гНК.

[0460] Вариант осуществления 72. Вариант гНК по варианту осуществления 71, в котором гетерологичная стеблевая петля РНК способна связывать белок, структуру РНК, ДНК-последовательность или малую молекулу.

[0461] Вариант осуществления 73. Вариант гНК по варианту осуществления 71 или варианту осуществления 72, в котором гетерологичная последовательность стеблевой петли РНК выбрана из стеблевых петель MS2, Qβ, шпильки II U1, Uvsx или PP7.

[0462] Вариант осуществления 74. Вариант гНК по любому из вариантов осуществления 61-73, содержащий две или более модификаций в одной области.

[0463] Вариант осуществления 75. Вариант гНК по любому из вариантов осуществления 61-74, содержащий модификации в двух или более областях.

[0464] Вариант осуществления 76. Вариант гНК по любому из вариантов осуществления 61-75, дополнительно содержащий нацеливающую последовательность, причем нацеливающая последовательность комплементарна целевой ДНК-последовательности.

[0465] Вариант осуществления 77. Вариант гНК по варианту осуществления 76, в котором нацеливающая последовательность имеет 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34 или 35 нуклеотидов.

[0466] Вариант осуществления 78. Вариант гНК по любому из варианта осуществления 76 или варианта осуществления 77, в котором нацеливающая последовательность имеет 20 нуклеотидов.

[0467] Вариант осуществления 79. Вариант гНК по любому из вариантов осуществления 76-78, в котором гНК представляет собой одномолекулярную гНК, содержащую последовательность каркаса, связанную с нацеливающей последовательностью.

[0468] Вариант осуществления 80. Вариант гНК по любому из вариантов осуществления 61-79, причем кратность улучшения одной или более из улучшенных характеристик варианта CasX составляет от по меньшей мере примерно 1,1 до примерно 100 раз или более относительно референсной гНК с SEQ ID NO: 4 или SEQ ID NO: 5.

[0469] Вариант осуществления 81. Вариант гНК по любому из вариантов осуществления 61-79, причем кратность улучшения одной или более из улучшенных характеристик варианта гНК составляет по меньшей мере примерно 1,1, по меньшей мере примерно 2, по меньшей мере примерно 10 или по меньшей мере примерно 100 раз или более относительно референсной гНК с SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 2238, SEQ ID NO: 2239, вариантного каркаса 174 (таблица 2) или вариантного каркаса 175

(таблица 2).

[0470] Вариант осуществления 82. Вариант гНК по любому из вариантов осуществления 61-81, содержащий каркасную область, которая имеет идентичность последовательности по меньшей мере 60% с SEQ ID NO: 4 или SEQ ID NO: 5, за исключением области удлиненного стебля.

[0471] Вариант осуществления 83. Вариант гНК по любому из вариантов осуществления 61-81, содержащий каркасную стеблевую петлю, которая имеет идентичность последовательности по меньшей мере 60% с SEQ ID NO: 14.

[0472] Вариант осуществления 84. Вариант гНК по любому из вариантов осуществления 61-81, в котором каркас последовательности варианта гНК имеет идентичность последовательности по меньшей мере 20%, по меньшей мере 30%, по меньшей мере 40%, по меньшей мере 50%, по меньшей мере 60% или по меньшей мере 70%, по меньшей мере 80%, по меньшей мере 85%, по меньшей мере примерно 90%, по меньшей мере примерно 95%, по меньшей мере примерно 96%, по меньшей мере примерно 97%, по меньшей мере примерно 98% или по меньшей мере примерно 99% с SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 2238, SEQ ID NO: 2239, вариантного каркаса 174 (таблица 2) или вариантного каркаса 175 (таблица 2).

[0473] Вариант осуществления 85. Вариант гНК по любому из вариантов осуществления 61-81, в котором каркас последовательности варианта гНК содержит последовательность, выбранную из группы последовательностей SEQ ID NO: 2101-2285 или 4433-4437 или последовательности, имеющей идентичность с ней по меньшей мере примерно 80%, по меньшей мере примерно 90%, по меньшей мере примерно 95%, по меньшей мере примерно 96%, по меньшей мере примерно 97%, по меньшей мере примерно 98%, по меньшей мере примерно 99%.

[0474] Вариант осуществления 86. Вариант гНК по варианту осуществления 85, в котором каркас последовательности варианта гНК состоит из последовательности, выбранной из группы последовательностей SEQ ID NO: 2101-2285 или 4433-4437.

[0475] Вариант осуществления 87. Вариант гНК по любому из вариантов осуществления 61-86, дополнительно содержащий один или более рибозимов.

[0476] Вариант осуществления 88. Вариант гНК по варианту осуществления 87, в котором один или более рибозимов независимо слиты с концом варианта гНК.

[0477] Вариант осуществления 89. Вариант гНК по варианту осуществления 87 или варианту осуществления 88, в котором по меньшей мере один из одного или более рибозимов представляет собой рибозим вируса гепатита дельта (HDV), рибозим hammerhead, пистолетный рибозим, рибозим hatchet или рибозим вируса кольцевой пятнистости табака (TRSV).

[0478] Вариант осуществления 90. Вариант гНК по любому из вариантов осуществления 61-89, дополнительно содержащий мотив связывания белка.

[0479] Вариант осуществления 91. Вариант гНК по любому из вариантов осуществления 61-90, дополнительно содержащий термостабильную стеблевую петлю.

[0480] Вариант осуществления 92. Вариант гНК по любому из вариантов осуществления 61-91, в котором гНК химически модифицирована.

[0481] Вариант осуществления 93. Вариант гНК по любому из вариантов осуществления 61-92, в котором гНК содержит первую область из первой гНК и вторую область из второй гНК, отличной от первой гНК.

[0482] Вариант осуществления 94. Вариант гНК по варианту осуществления 93, в котором первая область выбрана из группы, состоящей из триплексной области, каркасной стеблевой петли и удлиненной стеблевой петли.

[0483] Вариант осуществления 95. Вариант гНК по варианту осуществления 93 или варианту осуществления 94, в котором вторая область выбрана из группы, состоящей из триплексной области, каркасной стеблевой петли и удлиненной стеблевой петли.

[0484] Вариант осуществления 96. Вариант гНК по любому из вариантов осуществления 93-95, в котором первая и вторая области не являются одной и той же областью.

[0485] Вариант осуществления 97. Вариант гНК по любому из вариантов осуществления 93-95, в котором первая гНК содержит последовательность SEQ ID NO: 4, а вторая гНК содержит последовательность SEQ ID NO: 5.

[0486] Вариант осуществления 98. Вариант гНК по любому из вариантов осуществления 61-97, содержащий по меньшей мере одну химерную область, которая содержит первую часть из первой гНК и вторую часть из второй гНК.

[0487] Вариант осуществления 99. Вариант гНК по варианту осуществления 98, в котором по меньшей мере одна химерная область выбрана из группы, состоящей из триплексной области, каркасной стеблевой петли и удлиненной стеблевой петли.

[0488] Вариант осуществления 100. Вариант гНК по варианту осуществления 61, содержащий последовательность с любой из любой из SEQ ID NO: 2101-2285.

[0489] Вариант осуществления 101. Пара для редактирования генов, содержащая белок CasX и первую гНК.

[0490] Вариант осуществления 102. Пара для редактирования генов по варианту осуществления 101, в которой CasX и гНК способны связываться друг с другом в рибонуклеопротеиновом комплексе (RNP).

[0491] Вариант осуществления 103. Пара для редактирования генов по варианту осуществления 101, в которой CasX и гНК связаны друг с другом в рибонуклеопротеиновом комплексе (RNP).

[0492] Вариант осуществления 104. Пара для редактирования генов по любому из вариантов осуществления 101-103, в которой первая гНК содержит вариант гНК по любому из вариантов осуществления 76-100, нацеливающую последовательность, причем нацеливающая последовательность комплементарна целевой ДНК.

[0493] Вариант осуществления 105. Пара для редактирования генов по любому из вариантов осуществления 101-104, в которой CasX содержит вариант CasX по любому из вариантов осуществления 1-60.

[0494] Вариант осуществления 106. Пара для редактирования генов по любому из вариантов осуществления 101-105, содержащая:

a. вариант гНК по любому из вариантов осуществления 76-100 и

b. вариант CasX по любому из вариантов осуществления 1-60.

[0495] Вариант осуществления 107. Пара для редактирования генов по варианту осуществления 106, в которой пара для редактирования генов из варианта CasX и варианта гНК имеет одну или более улучшенных характеристик по сравнению с парой для редактирования генов, содержащей референсный белок CasX с SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 270 или SEQ ID NO: 336, и референсную гидовую нуклеиновую кислоту с SEQ ID NO: 4, 5, 2238 или 2239.

[0496] Вариант осуществления 108. Пара для редактирования генов по варианту осуществления 107, причем одна или более улучшенных характеристик включают в себя улучшенную стабильность комплекса CasX : гНК (RNP), улучшенную аффинность связывания между CasX и гНК, улучшенную кинетику образования комплекса RNP, более высокую процентную долю компетентных в отношении расщепления RNP, улучшенную аффинность связывания RNP с целевой ДНК, способность использовать расширенный спектр PAM-последовательностей, улучшенное раскручивание целевой ДНК, повышенную активность редактирования, улучшенную эффективность редактирования, улучшенную специфичность редактирования, повышенную нуклеазную активность, повышенную загрузку целевой цепи для двухцепочечного расщепления, сниженную загрузку целевой цепи для одноцепочечного никирования, сниженное нецелевое расщепление, улучшенное связывание нецелевой цепи ДНК или улучшенную стойкость к нуклеазной активности.

[0497] Вариант осуществления 109. Пара для редактирования генов по варианту осуществления 107 или варианту осуществления 108, причем кратность улучшения по меньшей мере одной или более из улучшенных характеристик составляет от по меньшей мере примерно 1,1 раза до примерно 100 раз или более относительно пары для редактирования генов из референсного белка CasX и референсной гидовой нуклеиновой кислоты.

[0498] Вариант осуществления 110. Пара для редактирования генов по варианту осуществления 107 или 108, причем одна или более из улучшенных характеристик варианта CasX имеют кратность улучшения по меньшей мере примерно 1,1 раза, по меньшей мере примерно 2 раза, по меньшей мере примерно 10 раз или по меньшей мере примерно 100 раз или более относительно пары для редактирования генов из референсного белка CasX и референсной гидовой нуклеиновой кислоты.

[0499] Вариант осуществления 111. Пара для редактирования генов по варианту осуществления 107 или варианту осуществления 108, причем улучшенная характеристика имеет кратность увеличения активности редактирования 4-9 раз по сравнению с референсной парой для редактирования с SEQ ID NO: 2 и SEQ ID NO: 5.

[0500] Вариант осуществления 112. Композиция, содержащая пару для редактирования генов по любому из вариантов осуществления 101-111, содержащая:

a. вторую пару для редактирования генов, содержащую вариант CasX по любому из вариантов осуществления 1-60; и

b. второй вариант гНК по любому из вариантов осуществления 61-100, причем второй вариант гНК имеет нацеливающую последовательность, комплементарную другой или перекрывающейся части целевой ДНК по сравнению с нацеливающей последовательностью первой гНК.

[0501] Вариант осуществления 113. Пара для редактирования генов по любому из вариантов осуществления 101-112, в которой RNP из варианта CasX и варианта гНК имеет более высокую процентную долю компетентных в отношении расщепления RNP по сравнению с RNP из референсного белка CasX и референсной гидовой нуклеиновой кислоты.

[0502] Вариант осуществления 114. Пара для редактирования генов по любому из вариантов осуществления 101-113, в которой RNP способен связывать целевую ДНК и расщеплять ее.

[0503] Вариант осуществления 115. Пара для редактирования генов по любому из вариантов осуществления 101-112, в которой RNP способен связывать целевую ДНК, но не способен расщеплять целевую ДНК.

[0504] Вариант осуществления 116. Пара для редактирования генов по любому из вариантов осуществления 101-112, в которой RNP способен связывать целевую ДНК и создавать один или более одноцепочечных ников в целевой ДНК.

[0505] Вариант осуществления 117. Вариант CasX, содержащий аминокислотную последовательность с любой из SEQ ID NO: 4416-4432.

[0506] Вариант осуществления 118. Вариант гНК, содержащий аминокислотную последовательность с любой из SEQ ID NO: 4433-4437.

Набор II

[0507] Вариант осуществления 1. Вариант референсного белка CasX (вариант CasX), причем:

a. вариант CasX содержит по меньшей мере одну модификацию в референсном белке CasX; и

b. вариант CasX демонстрирует по меньшей мере одну улучшенную характеристику по сравнению с референсным белком CasX, причем необязательно вариант содержит последовательность, выбранную из представленных в таблицах 3 и 8.

[0508] Вариант осуществления 2. Вариант CasX по варианту осуществления 1, причем улучшенная характеристика варианта CasX выбрана из группы, состоящей из: улучшенного свертывания варианта CasX; улучшенной аффинности связывания с гидовой нуклеиновой кислотой (гНК); улучшенной аффинности связывания с целевой ДНК; улучшенной способности использовать более широкий спектр из одной или более PAM-последовательностей, включая ATC, CTC, GTC или TTC, при редактировании целевой ДНК; улучшенного раскручивания целевой ДНК; повышенной активности редактирования; улучшенной эффективности редактирования; улучшенной специфичности редактирования;

повышенной нуклеазной активности; повышенной загрузки целевой цепи для двухцепочечного расщепления; сниженной загрузки целевой цепи для одноцепочечного никирования; сниженного нецелевого расщепления; улучшенного связывания с нецелевой цепью ДНК; улучшенной стабильности белка; улучшенной растворимости белка; улучшенной стабильности комплекса белок : гНК (RNP); улучшенной растворимости комплекса белок : гНК; улучшенного выхода белка; улучшенной экспрессии белка; улучшенной характеристики слияния или их комбинации.

[0509] Вариант осуществления 3. Вариант Cas X по варианту осуществления 1 или 2, причем по меньшей мере одна модификация содержит:

a. по меньшей мере одну замену аминокислоты в домене варианта CasX;

b. по меньшей мере одну делецию аминокислоты в домене варианта CasX;

c. по меньшей мере одну вставку аминокислоты в домене варианта CasX

d. полную или частичную замену доменом из другого CasX;

e. полную или частичную делецию домена варианта CasX или

f. или любую комбинацию пп. (a)-(d).

[0510] Вариант осуществления 4. Вариант CasX по любому из вариантов осуществления 1-3, причем референсный белок CasX содержит последовательность SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3.

[0511] Вариант осуществления 5. Вариант CasX по любому из вариантов осуществления 1-4, причем по меньшей мере одна модификация находится в домене, выбранном из:

a. домена связывания с нецелевой цепью (NTSB);

b. домена загрузки целевой цепи (TSL);

c. домена спиральный I;

d. домена спиральный II;

e. олигонуклеотид-связывающего домена (OBD) или

f. расщепляющего ДНК домена RuvC.

[0512] Вариант осуществления 6. Вариант CasX по варианту осуществления 5, содержащий по меньшей мере одну модификацию в домене NTSB.

[0513] Вариант осуществления 7. Вариант CasX по варианту осуществления 5, содержащий по меньшей мере одну модификацию в домене TSL.

[0514] Вариант осуществления 8. Вариант CasX по варианту осуществления 5, содержащий по меньшей мере одну модификацию в домене спиральный I.

[0515] Вариант осуществления 9. Вариант CasX по любому из вариантов осуществления 5-8, содержащий по меньшей мере одну модификацию в домене спиральный II.

[0516] Вариант осуществления 10. Вариант CasX по варианту осуществления 5, содержащий по меньшей мере одну модификацию в домене OBD.

[0517] Вариант осуществления 11. Вариант CasX по варианту осуществления 5, содержащий по меньшей мере одну модификацию в расщепляющем ДНК домене RuvC.

[0518] Вариант осуществления Вариант CasX по любому из вариантов осуществления 5-11, в котором модификация приводит к улучшенной способности редактировать целевую ДНК.

[0519] Вариант осуществления 13. Вариант CasX по любому из вариантов осуществления 1-12, причем вариант CasX способен образовывать рибонуклеопротеиновый комплекс (RNP) с гидовой нуклеиновой кислотой (гНК).

[0520] Вариант осуществления 14. Вариант CasX по любому из вариантов осуществления 1-13, причем по меньшей мере одна модификация содержит:

a. замену от 1 до 100 последовательных или непоследовательных аминокислот в варианте CasX;

b. делецию от 1 до 100 последовательных или непоследовательных аминокислот в варианте CasX;

c. вставку от 1 до 100 последовательных или непоследовательных аминокислот в CasX или

d. любую комбинацию пп. (a)-(c).

[0521] Вариант осуществления 15. Вариант CasX по варианту осуществления 14, причем по меньшей мере одна модификация содержит:

a. замену 5-10 последовательных или непоследовательных аминокислот в варианте CasX;

b. делецию 1-5 последовательных или непоследовательных аминокислот в варианте CasX;

c. вставку 1-5 последовательных или непоследовательных аминокислот в CasX или

d. любую комбинацию пп. (a)-(c).

[0522] Вариант осуществления 16. Вариант CasX по любому из вариантов осуществления 1-15, причем вариант CasX содержит две или более модификаций в одном домене.

[0523] Вариант осуществления 17. Вариант CasX по любому из вариантов осуществления 1-16, в котором вариант CasX содержит модификации в двух или более доменах.

[0524] Вариант осуществления 18. Вариант CasX по любому из вариантов осуществления 1-15, содержащий по меньшей мере одну модификацию области несмежных аминокислотных остатков варианта CasX, которые образуют канал, в котором происходит образование комплекса гРНК : целевая ДНК с вариантом CasX.

[0525] Вариант осуществления 19. Вариант CasX по любому из вариантов осуществления 1-15, содержащий по меньшей мере одну модификацию области несмежных аминокислотных остатков варианта CasX, которые образуют поверхность взаимодействия, связывающуюся с гНК.

[0526] Вариант осуществления 20. Вариант CasX по любому из вариантов осуществления 1-15, содержащий по меньшей мере одну модификацию области несмежных аминокислотных остатков варианта CasX, которые образуют канал, связывающийся с

нецелевой цепью ДНК.

[0527] Вариант осуществления 21. Вариант CasX по любому из вариантов осуществления 1-15, содержащий по меньшей мере одну модификацию области несмежных аминокислотных остатков варианта CasX, которые образуют поверхность взаимодействия, связывающуюся со смежным с протоспейсером мотивом (PAM) целевой ДНК.

[0528] Вариант осуществления 22. Вариант CasX по любому из вариантов осуществления 1-15, содержащий по меньшей мере одну модификацию области несмежных поверхностных аминокислотных остатков варианта CasX.

[0529] Вариант осуществления 23. Вариант CasX по любому из вариантов осуществления 1-15, содержащий по меньшей мере одну модификацию области несмежных аминокислотных остатков, которые образуют сердцевину посредством гидрофобной упаковки в домене варианта CasX.

[0530] Вариант осуществления 24. Вариант CasX по любому из вариантов осуществления 18-23, в котором модификация представляет собой одно или более из делеции, вставки или замены одной или более аминокислот области.

[0531] Вариант осуществления 25. Вариант CasX по любому из вариантов осуществления 18-23, в котором от 2 до 15 аминокислотных остатков области варианта CasX замещены заряженными аминокислотами.

[0532] Вариант осуществления 26. Вариант CasX по любому из вариантов осуществления 18-23, в котором от 2 до 15 аминокислотных остатков области варианта CasX замещены полярными аминокислотами.

[0533] Вариант осуществления 27. Вариант CasX по любому из вариантов осуществления 18-23, в котором от 2 до 15 аминокислотных остатков области варианта CasX замещены аминокислотами, которые образуют стопку с основаниями ДНК или РНК.

[0534] Вариант осуществления 28. Вариант CasX по любому из вариантов осуществления 1-5, имеющий последовательность, выбранную из группы, состоящей из последовательностей в таблице 3, или последовательность, имеющую идентичность последовательности с ней по меньшей мере примерно 50%, по меньшей мере примерно 60%, по меньшей мере примерно 70%, по меньшей мере примерно 80%, по меньшей мере примерно 90%, или по меньшей мере примерно 95%, или по меньшей мере примерно 96%, или по меньшей мере примерно 97%, или по меньшей мере примерно 98%, или по меньшей мере примерно 99%.

[0535] Вариант осуществления 29. Вариант CasX по любому из вариантов осуществления 1-5, дополнительно содержащий замену NTSB и/или домена спиральный 1b из другого CasX.

[0536] Вариант осуществления 30. Вариант CasX по варианту осуществления 29, в котором замещенный NTSB и/или домен спиральный 1b получен из референсного CasX с SEQ ID NO: 1.

[0537] Вариант осуществления 31. Вариант CasX по любому из вариантов осуществления 1-30, дополнительно содержащий один или более сигналов ядерной

локализации (NLS).

[0538] Вариант осуществления 32. Вариант CasX по варианту осуществления 31, в котором один или более NLS выбраны из группы последовательностей, состоящей из PKKKRKV (SEQ ID NO: 352), KRPAATKKAGQAKKKK (SEQ ID NO: 353), PAAKRVKLD (SEQ ID NO: 354), RQRRNELKRSP (SEQ ID NO: 355), NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY (SEQ ID NO: 356), RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV (SEQ ID NO: 357), VSRKRPRP (SEQ ID NO: 358), PPKKARED (SEQ ID NO: 35(), PQPKKKPL (SEQ ID NO: 360), SALIKKKKKMAP (SEQ ID NO: 361), DRLRR (SEQ ID NO: 362), PKQKKRK (SEQ ID NO: 363), RKLKKKIKKL (SEQ ID NO: 364), REKKKFLKRR (SEQ ID NO: 365), KRKGDEVDGVDEVAKKKSKK (SEQ ID NO: 366), RKCLQAGMNLEARKTKK (SEQ ID NO: 367), PRPRKIPR (SEQ ID NO: 368), PPRKKRTVV (SEQ ID NO: 369), NLSKKKKRKREK (SEQ ID NO: 370), RRPSRPFRKP (SEQ ID NO: 371), KRPRSPSS (SEQ ID NO: 372), KRGINDRNFWRGENERKTR (SEQ ID NO: 373), PRPPKMARYDN (SEQ ID NO: 374), KRSFSKAF (SEQ ID NO: 375), KLKIKRPVK (SEQ ID NO: 376), PKKKRKVPPPPAAKRVKLD (SEQ ID NO: 377), PKTRRRPRRSQRKRPPT (SEQ ID NO: 378), SRRRKANPTKLSENAKKLAKEVEN (SEQ ID NO: 379), KTRRRPRRSQRKRPPT (SEQ ID NO: 380), RRKKRRPRRKKRR (SEQ ID NO: 381), PKKKSRKPKKKSRK (SEQ ID NO: 382), HKKKHPDASVNFSEFSK (SEQ ID NO: 383), QRPGPYDRPQRPGPYDRP (SEQ ID NO: 384), LSPSLSPLLSPSLSPL (SEQ ID NO: 385), RGKGGKGLGKGGAKRHRK (SEQ ID NO: 386), PKRGRGRPKRGRGR (SEQ ID NO: 387), PKKKRKVPPPPKKKRKV (SEQ ID NO: 389), PAKRARRGYKC (SEQ ID NO: 4599), KLGPRKATGRW (SEQ ID NO: 4600), PRRKREE (SEQ ID NO: 4601), PYRGRKE (SEQ ID NO: 4602), PLRKRPRR (SEQ ID NO: 4603), PLRKRPRRGSPLRKRPRR (SEQ ID NO: 4604), PAAKRVKLDGGKRTADGSEFESPKKKRKV (SEQ ID NO: 4605), PAAKRVKLDGGKRTADGSEFESPKKKRKVGIHGVPAA (SEQ ID NO: 4606), PAAKRVKLDGGKRTADGSEFESPKKKRKVAEAAAKEAAAKEAAAKA (SEQ ID NO: 4607), PAAKRVKLDGGKRTADGSEFESPKKKRKVPG (SEQ ID NO: 4608), KRKGSPERGERKRHW (SEQ ID NO: 4609), KRTADSQHSTPPKTKRKVEFEPKKKRKV (SEQ ID NO: 4610) и PKKKRKVGGSKRTADSQHSTPPKTKRKVEFEPKKKRKV (SEQ ID NO: 4611).

[0539] Вариант осуществления 33. Вариант CasX по варианту осуществления 31 или варианту осуществления 32, в котором один или более NLS расположены на C-конце белка CasX или рядом с ним.

[0540] Вариант осуществления 34. Вариант CasX по варианту осуществления 31 или варианту осуществления 32, в котором один или более NLS расположены на N-конце белка CasX или рядом с ним.

[0541] Вариант осуществления 35. Вариант CasX по варианту осуществления 31 или варианту осуществления 32, содержащий по меньшей мере два NLS, причем по меньшей мере два NLS расположены на N-конце белка CasX или рядом с ним и на C-конце белка

CasX или рядом с ним.

[0542] Вариант осуществления 36. Вариант CasX по любому из вариантов осуществления 2-35, причем кратность улучшения одной или более из улучшенных характеристик варианта CasX составляет от по меньшей мере примерно 1,1 до примерно 100 раз или более относительно референсного белка CasX с SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3.

[0543] Вариант осуществления 37. Вариант CasX по любому из вариантов осуществления 2-35, причем кратность улучшения одной или более из улучшенных характеристик варианта CasX составляет по меньшей мере примерно 1,1, по меньшей мере примерно 2, по меньшей мере примерно 10, по меньшей мере примерно 100 раз или более относительно белка CasX с SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 270 или SEQ ID NO: 336.

[0544] Вариант осуществления 38. Вариант CasX по любому из вариантов осуществления 2-37, причем улучшенная характеристика включает в себя эффективность редактирования, и вариант CasX имеет кратность улучшения эффективности редактирования от 1,1 до 100 раз по сравнению с белком CasX с SEQ ID NO: 270 или SEQ ID NO: 336.

[0545] Вариант осуществления 39. Вариант CasX по любому из вариантов осуществления 1-38, причем RNP, содержащий вариант CasX, демонстрирует более высокие эффективность редактирования и/или связывание целевой последовательности в целевой ДНК, когда любая из PAM-последовательностей TTC, ATC, GTC или CTC расположена на расстоянии 1 нуклеотида от 5'-конца нецелевой цепи протоспейсера, имеющей идентичность с нацеливающей последовательностью гНК, в системе клеточного анализа по сравнению с эффективностью редактирования и/или связыванием RNP, содержащего референсный белок CasX, в сопоставимой системе анализа.

[0546] Вариант осуществления 40. Вариант CasX по варианту осуществления 39, причем PAM-последовательность представляет собой TTC.

[0547] Вариант осуществления 41. Вариант CasX по варианту осуществления 39, причем PAM-последовательность представляет собой ATC.

[0548] Вариант осуществления 42. Вариант CasX по варианту осуществления 39, причем PAM-последовательность представляет собой CTC.

[0549] Вариант осуществления 43. Вариант CasX по варианту осуществления 39, причем PAM-последовательность представляет собой GTC.

[0550] Вариант осуществления 44. Вариант CasX по любому из вариантов осуществления 39, причем улучшенные эффективность редактирования и/или связывание с целевой ДНК RNP, содержащего вариант CasX, имеют кратность улучшения от по меньшей мере примерно 1,1 до примерно 100 раз относительно RNP, содержащего референсный CasX.

[0551] Вариант осуществления 45. Вариант CasX по любому из вариантов осуществления 1-44, содержащий от 400 до 2000 аминокислот.

[0552] Вариант осуществления 46. Вариант CasX по любому из вариантов осуществления 1-45, причем вариант белка CasX содержит нуклеазный домен, обладающий никазной активностью.

[0553] Вариант осуществления 47. Вариант CasX по любому из вариантов осуществления 1-45, причем вариант белка CasX содержит нуклеазный домен, обладающий активностью двухцепочечного расщепления.

[0554] Вариант осуществления 48. Вариант CasX по любому из вариантов осуществления 1-45, причем белок CasX представляет собой каталитически неактивный белок CasX (dCasX), и при этом dCasX и гНК сохраняют способность связываться с целевой ДНК.

[0555] Вариант осуществления 49. Вариант CasX по варианту осуществления 48, причем dCasX содержит мутацию по остаткам:

[0556] a. D672, и/или E769, и/или D935, соответствующим белку CasX с SEQ ID NO: 1; или

[0557] b. D659, и/или E756, и/или D922, соответствующим белку CasX с SEQ ID NO: 2.

[0558] Вариант осуществления 50. Вариант CasX по варианту осуществления 49, в котором мутация представляет собой замену остатка на аланин.

[0559] Вариант осуществления 51. Вариант CasX по любому из вариантов осуществления 1-50, содержащий первый домен из первого белка CasX и второй домен из второго белка CasX, отличного от первого белка CasX.

[0560] Вариант осуществления 52. Вариант CasX по варианту осуществления 51, в котором первый домен выбран из группы, состоящей из доменов NTSB, TSL, спирального I, спирального II, OBD и RuvC.

[0561] Вариант осуществления 53. Вариант CasX по варианту осуществления 51, в котором второй домен выбран из группы, состоящей из доменов NTSB, TSL, спирального I, спирального II, OBD и RuvC.

[0562] Вариант осуществления 54. Вариант CasX по любому из вариантов осуществления 51-53, в котором первый и второй домены не являются одним и тем же доменом.

[0563] Вариант осуществления 55. Вариант CasX по любому из вариантов осуществления 1-50, содержащий по меньшей мере один химерный домен, который содержит первую часть из первого белка CasX и вторую часть из второго белка CasX, отличного от первого белка CasX.

[0564] Вариант осуществления 56. Вариант CasX по варианту осуществления 55, в котором по меньшей мере один химерный домен выбран из группы, состоящей из доменов NTSB, TSL, спирального I, спирального II, OBD и RuvC.

[0565] Вариант осуществления 57. Вариант CasX по варианту осуществления 56, в котором по меньшей мере один химерный домен содержит химерный домен RuvC.

[0566] Вариант осуществления 58. Вариант CasX по любому из вариантов

осуществления 1-57, содержащий гетерологичный белок или его домен, слитый с CasX.

[0567] Вариант осуществления 59. Вариант CasX по варианту осуществления 58, в котором гетерологичный белок или его домен представляет собой редактор оснований.

[0568] Вариант осуществления 60. Вариант CasX по варианту осуществления 59, в котором редактор оснований представляет собой аденозиндезаминазу, цитозиндезаминазу или гуаниноксидазу.

[0569] Вариант осуществления 61. Вариант каркаса референсной гидовой нуклеиновой кислоты (вариант гНК), способный связывать референсный белок CasX или вариант CasX, причем:

a. вариант гНК содержит по меньшей мере одну модификацию по сравнению с каркасной последовательностью референсной гидовой нуклеиновой кислоты и

b. вариант гНК демонстрирует одну или более улучшенных характеристик по сравнению с каркасом референсной гидовой нуклеиновой кислоты.

[0570] Вариант осуществления 62. Вариант гНК по варианту осуществления 61, причем одна или более улучшенных характеристик выбраны из группы, состоящей из: улучшенной стабильности; улучшенной растворимости; улучшенной транскрипции гНК; улучшенной стойкости к нуклеазной активности; повышенной скорости свертывания гНК; сниженного образования побочных продуктов при свертывании; повышенного продуктивного свертывания; улучшенной аффинности связывания с белком CasX; улучшенной аффинности связывания с целевой ДНК при нахождении в комплексе с белком CasX; улучшенного редактирования генов при нахождении в комплексе с белком CasX; улучшенной специфичности редактирования при нахождении в комплексе с белком CasX и улучшенной способности использовать больший спектр из одной или более PAM-последовательностей, включая ATC, CTC, GTC или TTC, во время редактирования целевой ДНК при нахождении в комплексе с белком CasX.

[0571] Вариант осуществления 63. Вариант гНК по варианту осуществления 61 или 62, в котором каркас референсного гида содержит последовательность, выбранную из группы, состоящей из последовательностей SEQ ID NO: 4-16.

[0572] Вариант осуществления 64. Вариант гНК по любому из вариантов осуществления 61-63, в котором по меньшей мере одна модификация содержит:

a. по меньшей мере одну замену нуклеотида в области варианта гНК;

b. по меньшей мере одну делецию нуклеотида в области варианта гНК;

c. по меньшей мере одну вставку нуклеотида в области варианта гНК;

d. полную или частичную замену области варианта гНК;

e. полную или частичную делецию области варианта гНК; или

f. или любую комбинацию пп. (a)-(d).

[0573] Вариант осуществления 65. Вариант гНК по варианту осуществления 64, в котором область варианта гНК выбрана из группы, состоящей из удлиненной стеблевой петли, каркасной стеблевой петли, триплекса и псевдоузла.

[0574] Вариант осуществления 66. Вариант гНК по варианту осуществления 65, в

котором каркасный стебель дополнительно содержит пузырек.

[0575] Вариант осуществления 67. Вариант гНК по варианту осуществления 65 или варианту осуществления 66, в котором каркас дополнительно содержит триплексную петлевую область.

[0576] Вариант осуществления 68. Вариант гНК по любому из вариантов осуществления 65-67, в котором каркас дополнительно содержит 5'-неструктурированную область.

[0577] Вариант осуществления 69. Вариант гНК по любому из вариантов осуществления 64-68, в котором по меньшей мере одна модификация содержит:

a. замену 1-15 последовательных или не последовательных нуклеотидов в варианте гНК в одной или более областях;

b. делецию 1-10 последовательных или не последовательных нуклеотидов в варианте гНК в одной или более областях;

c. вставку 1-10 последовательных или не последовательных нуклеотидов в варианте гНК в одной или более областях;

d. замену каркасной стеблевой петли или удлиненной стеблевой петли на последовательность стеблевой петли РНК из гетерологичного РНК-источника с проксимальными 5’- и 3’-концами или

e. любую комбинацию пп. (a)-(d).

[0578] Вариант осуществления 70. Вариант гНК по любому из вариантов осуществления 61-69, содержащий область удлиненной стеблевой петли, содержащую по меньшей мере 10, по меньшей мере 100, по меньшей мере 500, по меньшей мере 1000 или по меньшей мере 10 000 нуклеотидов.

[0579] Вариант осуществления 71. Вариант гНК по варианту осуществления 69, в котором гетерологичная последовательность стеблевой петли РНК повышает стабильность гНК.

[0580] Вариант осуществления 72. Вариант гНК по варианту осуществления 71, в котором гетерологичная стеблевая петля РНК способна связывать белок, структуру РНК, ДНК-последовательность или малую молекулу.

[0581] Вариант осуществления 73. Вариант гНК по варианту осуществления 71 или варианту осуществления 72, в котором гетерологичная последовательность стеблевой петли РНК выбрана из стеблевых петель MS2, Qβ, шпильки II U1, Uvsx или PP7.

[0582] Вариант осуществления 74. Вариант гНК по любому из вариантов осуществления 61-73, содержащий две или более модификаций в одной области.

[0583] Вариант осуществления 75. Вариант гНК по любому из вариантов осуществления 61-74, содержащий модификации в двух или более областях.

[0584] Вариант осуществления 76. Вариант гНК по любому из вариантов осуществления 61-75, дополнительно содержащий нацеливающую последовательность, причем нацеливающая последовательность комплементарна целевой ДНК-последовательности.

[0585] Вариант осуществления 77. Вариант гНК по варианту осуществления 76, в котором нацеливающая последовательность имеет 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34 или 35 нуклеотидов.

[0586] Вариант осуществления 78. Вариант гНК по любому из варианта осуществления 76 или варианта осуществления 77, в котором нацеливающая последовательность имеет 20 нуклеотидов.

[0587] Вариант осуществления 79. Вариант гНК по любому из вариантов осуществления 76-78, в котором гНК представляет собой одномолекулярную гНК, содержащую последовательность каркаса, связанную с нацеливающей последовательностью.

[0588] Вариант осуществления 80. Вариант гНК по любому из вариантов осуществления 61-79, причем кратность улучшения одной или более из улучшенных характеристик варианта CasX составляет от по меньшей мере примерно 1,1 до примерно 100 раз или более относительно референсной гНК с SEQ ID NO: 4 или SEQ ID NO: 5.

[0589] Вариант осуществления 81. Вариант гНК по любому из вариантов осуществления 61-79, причем кратность улучшения одной или более из улучшенных характеристик варианта гНК составляет по меньшей мере примерно 1,1, по меньшей мере примерно 2, по меньшей мере примерно 10 или по меньшей мере примерно 100 раз или более относительно референсной гНК с SEQ ID NO: 4, референсной гНК с SEQ ID NO: 5, вариантным каркасом с SEQ ID NO: 2238 или вариантным каркасом с SEQ ID NO: 2239.

[0590] Вариант осуществления 82. Вариант гНК по любому из вариантов осуществления 61-81, содержащий каркасную область, которая имеет идентичность последовательности по меньшей мере 60% с SEQ ID NO: 4 или SEQ ID NO: 5, за исключением области удлиненного стебля.

[0591] Вариант осуществления 83. Вариант гНК по любому из вариантов осуществления 61-81, содержащий каркасную стеблевую петлю, которая имеет идентичность последовательности по меньшей мере 60% с SEQ ID NO: 14.

[0592] Вариант осуществления 84. Вариант гНК по любому из вариантов осуществления 61-81, в котором каркас последовательности варианта гНК имеет идентичность последовательности по меньшей мере 20%, по меньшей мере 30%, по меньшей мере 40%, по меньшей мере 50%, по меньшей мере 60% или по меньшей мере 70%, по меньшей мере 80%, по меньшей мере 85%, по меньшей мере примерно 90%, по меньшей мере примерно 95%, по меньшей мере примерно 96%, по меньшей мере примерно 97%, по меньшей мере примерно 98% или по меньшей мере примерно 99% с SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 2238 или SEQ ID NO: 2239.

[0593] Вариант осуществления 85. Вариант гНК по любому из вариантов осуществления 61-81, в котором каркас последовательности варианта гНК содержит последовательность, выбранную из группы последовательностей SEQ ID NO: 2101-2280 или 4433-4446 или имеющую идентичность последовательности с ней по меньшей мере примерно 80%, по меньшей мере примерно 90%, по меньшей мере примерно 95%, по

меньшей мере примерно 96%, по меньшей мере примерно 97%, по меньшей мере примерно 98%, по меньшей мере примерно 99%.

[0594] Вариант осуществления 86. Вариант гНК по варианту осуществления 85, в котором каркас последовательности варианта гНК состоит из последовательности, выбранной из группы последовательностей SEQ ID NO: 2101-2280 или 4433-4446.

[0595] Вариант осуществления 87. Вариант гНК по любому из вариантов осуществления 61-86, дополнительно содержащий один или более рибозимов.

[0596] Вариант осуществления 88. Вариант гНК по варианту осуществления 87, в котором один или более рибозимов независимо слиты с концом варианта гНК.

[0597] Вариант осуществления 89. Вариант гНК по варианту осуществления 87 или варианту осуществления 88, в котором по меньшей мере один из одного или более рибозимов представляет собой рибозим вируса гепатита дельта (HDV), рибозим hammerhead, пистолетный рибозим, рибозим hatchet или рибозим вируса кольцевой пятнистости табака (TRSV).

[0598] Вариант осуществления 90. Вариант гНК по любому из вариантов осуществления 61-89, дополнительно содержащий мотив связывания белка.

[0599] Вариант осуществления 91. Вариант гНК по любому из вариантов осуществления 61-90, дополнительно содержащий термостабильную стеблевую петлю.

[0600] Вариант осуществления 92. Вариант гНК по любому из вариантов осуществления 61-91, в котором гНК химически модифицирована.

[0601] Вариант осуществления 93. Вариант гНК по любому из вариантов осуществления 61-92, в котором гНК содержит первую область из первой гНК и вторую область из второй гНК, отличной от первой гНК.

[0602] Вариант осуществления 94. Вариант гНК по варианту осуществления 93, в котором первая область выбрана из группы, состоящей из триплексной области, каркасной стеблевой петли и удлиненной стеблевой петли.

[0603] Вариант осуществления 95. Вариант гНК по варианту осуществления 93 или варианту осуществления 94, в котором вторая область выбрана из группы, состоящей из триплексной области, каркасной стеблевой петли и удлиненной стеблевой петли.

[0604] Вариант осуществления 96. Вариант гНК по любому из вариантов осуществления 93-95, в котором первая и вторая области не являются одной и той же областью.

[0605] Вариант осуществления 97. Вариант гНК по любому из вариантов осуществления 93-95, в котором первая гНК содержит последовательность SEQ ID NO: 4, а вторая гНК содержит последовательность SEQ ID NO: 5.

[0606] Вариант осуществления 98. Вариант гНК по любому из вариантов осуществления 61-97, содержащий по меньшей мере одну химерную область, которая содержит первую часть из первой гНК и вторую часть из второй гНК.

[0607] Вариант осуществления 99. Вариант гНК по варианту осуществления 98, в котором по меньшей мере одна химерная область выбрана из группы, состоящей из

триплексной области, каркасной стеблевой петли и удлиненной стеблевой петли.

[0608] Вариант осуществления 100. Вариант гНК по варианту осуществления 61, содержащий последовательность с любой из любой из SEQ ID NO: 2101-2280 или 4433-4446.

[0609] Вариант осуществления 101. Пара для редактирования генов, содержащая белок CasX и первую гНК.

[0610] Вариант осуществления 102. Пара для редактирования генов по варианту осуществления 101, в которой CasX и гНК способны связываться друг с другом в рибонуклеопротеиновом комплексе (RNP).

[0611] Вариант осуществления 103. Пара для редактирования генов по варианту осуществления 101, в которой CasX и гНК связаны друг с другом в рибонуклеопротеиновом комплексе (RNP).

[0612] Вариант осуществления 104. Пара для редактирования генов по любому из вариантов осуществления 101-103, в которой первая гНК содержит вариант гНК по любому из вариантов осуществления 76-100, нацеливающую последовательность, причем нацеливающая последовательность комплементарна целевой ДНК.

[0613] Вариант осуществления 105. Пара для редактирования генов по любому из вариантов осуществления 101-104, в которой CasX содержит вариант CasX по любому из вариантов осуществления 1-60.

[0614] Вариант осуществления 106. Пара для редактирования генов по любому из вариантов осуществления 101-105, содержащая:

a. вариант гНК по любому из вариантов осуществления 76-100 и

b. вариант CasX по любому из вариантов осуществления 1-60.

[0615] Вариант осуществления 107. Пара для редактирования генов по варианту осуществления 106, в которой пара для редактирования генов из варианта CasX и варианта гНК имеет одну или более улучшенных характеристик по сравнению с парой для редактирования генов, содержащей референсный белок CasX с SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3 и референсную гидовую нуклеиновую кислоту с SEQ ID NO: 4 или 5.

[0616] Вариант осуществления 108. Пара для редактирования генов по варианту осуществления 107, причем одна или более улучшенных характеристик включают в себя улучшенную стабильность комплекса CasX : гНК (RNP), улучшенную аффинность связывания между CasX и гНК, улучшенную кинетику образования комплекса RNP, более высокую процентную долю компетентных в отношении расщепления RNP, улучшенную аффинность связывания RNP с целевой ДНК, способность использовать расширенный спектр PAM-последовательностей, улучшенное раскручивание целевой ДНК, повышенную активность редактирования, улучшенную эффективность редактирования, улучшенную специфичность редактирования, повышенную нуклеазную активность, повышенную загрузку целевой цепи для двухцепочечного расщепления, сниженную загрузку целевой цепи для одноцепочечного никирования, сниженное нецелевое расщепление, улучшенное связывание нецелевой цепи ДНК или улучшенную стойкость к нуклеазной активности.

[0617] Вариант осуществления 109. Пара для редактирования генов по варианту осуществления 107 или варианту осуществления 108, причем кратность улучшения по меньшей мере одной или более из улучшенных характеристик составляет от по меньшей мере примерно 1,1 раза до примерно 100 раз или более относительно пары для редактирования генов из референсного белка CasX и референсной гидовой нуклеиновой кислоты.

[0618] Вариант осуществления 110. Пара для редактирования генов по варианту осуществления 107 или 108, причем одна или более из улучшенных характеристик варианта CasX имеют кратность улучшения по меньшей мере примерно 1,1 раза, по меньшей мере примерно 2 раза, по меньшей мере примерно 10 раз или по меньшей мере примерно 100 раз или более относительно пары для редактирования генов из референсного белка CasX и референсной гидовой нуклеиновой кислоты.

[0619] Вариант осуществления 111. Пара для редактирования генов по варианту осуществления 107 или варианту осуществления 108, причем улучшенная характеристика имеет кратность увеличения активности редактирования 4-9 раз по сравнению с референсной парой для редактирования с SEQ ID NO: 2 и SEQ ID NO: 5.

[0620] Вариант осуществления 112. Композиция, содержащая пару для редактирования генов по любому из вариантов осуществления 101-111, содержащая:

a. вторую пару для редактирования генов, содержащую вариант CasX по любому из вариантов осуществления 1-60; и

b. второй вариант гНК по любому из вариантов осуществления 61-100, причем второй вариант гНК имеет нацеливающую последовательность, комплементарную другой или перекрывающейся части целевой ДНК по сравнению с нацеливающей последовательностью первой гНК.

[0621] Вариант осуществления 113. Пара для редактирования генов по любому из вариантов осуществления 101-112, в которой RNP из варианта CasX и варианта гНК имеет более высокую процентную долю компетентных в отношении расщепления RNP по сравнению с RNP из референсного белка CasX и референсной гидовой нуклеиновой кислоты.

[0622] Вариант осуществления 114. Пара для редактирования генов по любому из вариантов осуществления 101-113, в которой RNP способен связывать целевую ДНК и расщеплять ее.

[0623] Вариант осуществления 115. Пара для редактирования генов по любому из вариантов осуществления 101-112, в которой RNP способен связывать целевую ДНК, но не способен расщеплять целевую ДНК.

[0624] Вариант осуществления 116. Пара для редактирования генов по любому из вариантов осуществления 101-112, в которой RNP способен связывать целевую ДНК и создавать один или более одноцепочечных ников в целевой ДНК.

[0625] Вариант осуществления 117. Вариант CasX, содержащий аминокислотную последовательность с любой из SEQ ID NO: 4416-4432 или 4597-4598.

[0626] Вариант осуществления 118. Вариант гНК, содержащий аминокислотную последовательность с любой из SEQ ID NO: 4433-4446.

Набор III

[0627] Вариант осуществления 1. Вариант референсного белка CasX (вариант CasX), причем:

a. вариант CasX содержит по меньшей мере одну модификацию в референсном белке CasX; и

b. вариант CasX демонстрирует по меньшей мере одну улучшенную характеристику по сравнению с референсным белком CasX,

причем необязательно вариант содержит последовательность, выбранную из группы, состоящей из SEQ ID NO: 89-101, 247-337, 411-592 и 760-982.

[0628] Вариант осуществления 2. Вариант CasX по варианту осуществления 1, причем улучшенная характеристика варианта CasX выбрана из группы, состоящей из: улучшенного свертывания варианта CasX; улучшенной аффинности связывания с гидовой нуклеиновой кислотой (гНК); улучшенной аффинности связывания с целевой нуклеиновой кислотой; улучшенной способности использовать более широкий спектр из одной или более PAM-последовательностей, включая ATC, CTC, GTC или TTC, при редактировании целевой нуклеиновой кислоты; улучшенного раскручивания целевой нуклеиновой кислоты; повышенной активности редактирования; улучшенной эффективности редактирования; улучшенной специфичности редактирования; повышенной нуклеазной активности; повышенной загрузки целевой цепи для двухцепочечного расщепления; сниженной загрузки целевой цепи для одноцепочечного никирования; сниженного нецелевого расщепления; улучшенного связывания нецелевой цепи нуклеиновой кислоты; улучшенной стабильности белка; улучшенной растворимости белка; улучшенной стабильности комплекса белок : гНК (RNP); улучшенной растворимости комплекса белок : гНК; улучшенного выхода белка; улучшенной экспрессии белка; улучшенной характеристики слияния или их комбинации.

[0629] Вариант осуществления 3. Вариант Cas X по варианту осуществления 1 или 2, причем по меньшей мере одна модификация содержит:

a. по меньшей мере одну замену аминокислоты в домене варианта CasX;

b. по меньшей мере одну делецию аминокислоты в домене варианта CasX;

c. по меньшей мере одну вставку аминокислоты в домене варианта CasX

d. полную или частичную замену доменом из другого CasX;

e. полную или частичную делецию домена варианта CasX или

f. или любую комбинацию пп. (a)-(d).

[0630] Вариант осуществления 4. Вариант CasX по любому из вариантов осуществления 1-3, причем референсный белок CasX содержит последовательность SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3.

[0631] Вариант осуществления 5. Вариант CasX по любому из вариантов осуществления 1-4, причем по меньшей мере одна модификация находится в домене,

выбранном из:

a. домена связывания с нецелевой цепью (NTSB);

b. домена загрузки целевой цепи (TSL);

c. домена спиральный I;

d. домена спиральный II;

e. олигонуклеотид-связывающего домена (OBD) или

f. расщепляющего ДНК домена RuvC.

[0632] Вариант осуществления 6. Вариант CasX по варианту осуществления 5, содержащий по меньшей мере одну модификацию в домене NTSB.

[0633] Вариант осуществления 7. Вариант CasX по варианту осуществления 5, содержащий по меньшей мере одну модификацию в домене TSL.

[0634] Вариант осуществления 8. Вариант CasX по варианту осуществления 5, содержащий по меньшей мере одну модификацию в домене спиральный I.

[0635] Вариант осуществления 9. Вариант CasX по любому из вариантов осуществления 5-8, содержащий по меньшей мере одну модификацию в домене спиральный II.

[0636] Вариант осуществления 10. Вариант CasX по варианту осуществления 5, содержащий по меньшей мере одну модификацию в домене OBD.

[0637] Вариант осуществления 11. Вариант CasX по варианту осуществления 5, содержащий по меньшей мере одну модификацию в расщепляющем ДНК домене RuvC.

[0638] Вариант осуществления 12. Вариант CasX по любому из вариантов осуществления 5-11, в котором модификация приводит к улучшенной способности редактировать целевую нуклеиновую кислоту.

[0639] Вариант осуществления 13. Вариант CasX по любому из вариантов осуществления 1-12, причем вариант CasX способен образовывать рибонуклеопротеиновый комплекс (RNP) с гидовой нуклеиновой кислотой (гНК).

[0640] Вариант осуществления 14. Вариант CasX по любому из вариантов осуществления 1-13, причем по меньшей мере одна модификация содержит:

a. замену от 1 до 100 последовательных или непоследовательных аминокислот в варианте CasX;

b. делецию от 1 до 100 последовательных или непоследовательных аминокислот в варианте CasX;

c. вставку от 1 до 100 последовательных или непоследовательных аминокислот в CasX или

d. любую комбинацию пп. (a)-(c).

[0641] Вариант осуществления 15. Вариант CasX по варианту осуществления 14, причем по меньшей мере одна модификация содержит:

a. замену 5-10 последовательных или непоследовательных аминокислот в варианте CasX;

b. делецию 1-5 последовательных или непоследовательных аминокислот в варианте

CasX;

c. вставку 1-5 последовательных или непоследовательных аминокислот в CasX или

d. любую комбинацию пп. (a)-(c).

[0642] Вариант осуществления 16. Вариант CasX по любому из вариантов осуществления 1-15, причем вариант CasX содержит две или более модификаций в одном домене.

[0643] Вариант осуществления 17. Вариант CasX по любому из вариантов осуществления 1-16, в котором вариант CasX содержит модификации в двух или более доменах.

[0644] Вариант осуществления 18. Вариант CasX по любому из вариантов осуществления 1-15, содержащий по меньшей мере одну модификацию области несмежных аминокислотных остатков варианта CasX, которые образуют канал, в котором происходит образование комплекса гНК : целевая нуклеиновая кислота с вариантом CasX.

[0645] Вариант осуществления 19. Вариант CasX по любому из вариантов осуществления 1-15, содержащий по меньшей мере одну модификацию области несмежных аминокислотных остатков варианта CasX, которые образуют поверхность взаимодействия, связывающуюся с гНК.

[0646] Вариант осуществления 20. Вариант CasX по любому из вариантов осуществления 1-15, содержащий по меньшей мере одну модификацию области несмежных аминокислотных остатков варианта CasX, которые образуют канал, связывающийся с нецелевой цепью ДНК.

[0647] Вариант осуществления 21. Вариант CasX по любому из вариантов осуществления 1-15, содержащий по меньшей мере одну модификацию области несмежных аминокислотных остатков варианта CasX, которые образуют поверхность взаимодействия, которая связывается со смежным с протоспейсером мотивом (PAM) целевой нуклеиновой кислоты.

[0648] Вариант осуществления 22. Вариант CasX по любому из вариантов осуществления 1-15, содержащий по меньшей мере одну модификацию области несмежных поверхностных аминокислотных остатков варианта CasX.

[0649] Вариант осуществления 23. Вариант CasX по любому из вариантов осуществления 1-15, содержащий по меньшей мере одну модификацию области несмежных аминокислотных остатков, которые образуют сердцевину посредством гидрофобной упаковки в домене варианта CasX.

[0650] Вариант осуществления 24. Вариант CasX по любому из вариантов осуществления 18-23, в котором модификация представляет собой одно или более из делеции, вставки или замены одной или более аминокислот области.

[0651] Вариант осуществления 25. Вариант CasX по любому из вариантов осуществления 18-23, в котором от 2 до 15 аминокислотных остатков области варианта CasX замещены заряженными аминокислотами.

[0652] Вариант осуществления 26. Вариант CasX по любому из вариантов

осуществления 18-23, в котором от 2 до 15 аминокислотных остатков области варианта CasX замещены полярными аминокислотами.

[0653] Вариант осуществления 27. Вариант CasX по любому из вариантов осуществления 18-23, в котором от 2 до 15 аминокислотных остатков области варианта CasX замещены аминокислотами, которые образуют стопку с основаниями ДНК или РНК.

[0654] Вариант осуществления 28. Вариант CasX по любому из вариантов осуществления 1-5, дополнительно содержащий замену NTSB и/или домена спиральный 1b из другого CasX.

[0655] Вариант осуществления 29. Вариант CasX по варианту осуществления 28, в котором замещенный NTSB и/или домен спиральный 1b получен из референсного CasX с SEQ ID NO: 1.

[0656] Вариант осуществления 30. Вариант CasX по любому из вариантов осуществления 1-29, имеющий последовательность, выбранную из группы, состоящей из SEQ ID NO: 89-101, 247-337, 411-592 и 760-982, или последовательность, имеющую идентичность последовательности с ней по меньшей мере примерно 50%, по меньшей мере примерно 60%, по меньшей мере примерно 70%, по меньшей мере примерно 80%, по меньшей мере примерно 90%, или по меньшей мере примерно 95%, или по меньшей мере примерно 96%, или по меньшей мере примерно 97%, или по меньшей мере примерно 98%, или по меньшей мере примерно 99%.

[0657] Вариант осуществления 31. Вариант CasX по любому из вариантов осуществления 1-30, дополнительно содержащий один или более сигналов ядерной локализации (NLS).

[0658] Вариант осуществления 32. Вариант CasX по варианту осуществления 31, в котором один или более NLS выбраны из группы последовательностей, состоящей из PKKKRKV (SEQ ID NO: 352), KRPAATKKAGQAKKKK (SEQ ID NO: 353), PAAKRVKLD (SEQ ID NO: 354), RQRRNELKRSP (SEQ ID NO: 355), NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY (SEQ ID NO: 356), RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV (SEQ ID NO: 357), VSRKRPRP (SEQ ID NO: 358), PPKKARED (SEQ ID NO: (359), PQPKKKPL (SEQ ID NO: 360), SALIKKKKKMAP (SEQ ID NO: 361), DRLRR (SEQ ID NO: 362), PKQKKRK (SEQ ID NO: 363), RKLKKKIKKL (SEQ ID NO: 364), REKKKFLKRR (SEQ ID NO: 365), KRKGDEVDGVDEVAKKKSKK (SEQ ID NO: 366), RKCLQAGMNLEARKTKK (SEQ ID NO: 367), PRPRKIPR (SEQ ID NO: 368), PPRKKRTVV (SEQ ID NO: 369), NLSKKKKRKREK (SEQ ID NO: 370), RRPSRPFRKP (SEQ ID NO: 371), KRPRSPSS (SEQ ID NO: 372), KRGINDRNFWRGENERKTR (SEQ ID NO: 373), PRPPKMARYDN (SEQ ID NO: 374), KRSFSKAF (SEQ ID NO: 375), KLKIKRPVK (SEQ ID NO: 376), PKKKRKVPPPPAAKRVKLD (SEQ ID NO: 377), PKTRRRPRRSQRKRPPT (SEQ ID NO: 378), SRRRKANPTKLSENAKKLAKEVEN (SEQ ID NO: 379), KTRRRPRRSQRKRPPT (SEQ ID NO: 380), RRKKRRPRRKKRR (SEQ ID NO: 381), PKKKSRKPKKKSRK (SEQ ID NO: 382), HKKKHPDASVNFSEFSK (SEQ ID NO: 383), QRPGPYDRPQRPGPYDRP (SEQ ID

NO: 384), LSPSLSPLLSPSLSPL (SEQ ID NO: 385), RGKGGKGLGKGGAKRHRK (SEQ ID NO: 386), PKRGRGRPKRGRGR (SEQ ID NO: 387), PKKKRKVPPPPKKKRKV (SEQ ID NO: 389), PAKRARRGYKC (SEQ ID NO: 63), KLGPRKATGRW (SEQ ID NO: 64), PRRKREE (SEQ ID NO: 65), PYRGRKE (SEQ ID NO: 66), PLRKRPRR (SEQ ID NO: 67), PLRKRPRRGSPLRKRPRR (SEQ ID NO:68), PAAKRVKLDGGKRTADGSEFESPKKKRKV (SEQ ID NO:69), PAAKRVKLDGGKRTADGSEFESPKKKRKVGIHGVPAA (SEQ ID NO:70), PAAKRVKLDGGKRTADGSEFESPKKKRKVAEAAAKEAAAKEAAAKA (SEQ ID NO:71), PAAKRVKLDGGKRTADGSEFESPKKKRKVPG (SEQ ID NO:72), KRKGSPERGERKRHW (SEQ ID NO:73), KRTADSQHSTPPKTKRKVEFEPKKKRKV (SEQ ID NO:74) и PKKKRKVGGSKRTADSQHSTPPKTKRKVEFEPKKKRKV (SEQ ID NO:75), и необязательно один или более NLS присоединены к варианту CasX или к смежному NLS с помощью линкерного пептида, причем линкерный пептид выбран из группы, состоящей из (G)n (SEQ ID NO: 1023), (GS)n (SEQ ID NO: 1024), (GSGGS)n (SEQ ID NO: 399), (GGSGGS)n (SEQ ID NO: 400), (GGGS)n (SEQ ID NO: 401), GGSG (SEQ ID NO: 402), GGSGG (SEQ ID NO: 403), GSGSG (SEQ ID NO: 404), GSGGG (SEQ ID NO: 405), GGGSG (SEQ ID NO: 406), GSSSG (SEQ ID NO: 407), GPGP (SEQ ID NO: 408), GGP, PPP, PPAPPA (SEQ ID NO: 409), PPPG (SEQ ID NO: 24), PPPGPPP (SEQ ID NO: 410), PPP(GGGS)n (SEQ ID NO: 25), (GGGS)nPPP (SEQ ID NO: 26), AEAAAKEAAAKEAAAKA (SEQ ID NO: 1025) и TPPKTKRKVEFE (SEQ ID NO: 27), где n составляет от 1 до 5.

[0659] Вариант осуществления 33. Вариант CasX по варианту осуществления 31 или варианту осуществления 32, в котором один или более NLS расположены на C-конце белка CasX или рядом с ним.

[0660] Вариант осуществления 34. Вариант CasX по варианту осуществления 31 или варианту осуществления 32, в котором один или более NLS расположены на N-конце белка CasX или рядом с ним.

[0661] Вариант осуществления 35. Вариант CasX по варианту осуществления 31 или варианту осуществления 32, содержащий по меньшей мере два NLS, причем по меньшей мере два NLS расположены на N-конце белка CasX или рядом с ним и на C-конце белка CasX или рядом с ним.

[0662] Вариант осуществления 36. Вариант CasX по любому из вариантов осуществления 2-35, причем кратность улучшения одной или более из улучшенных характеристик варианта CasX составляет от по меньшей мере примерно 1,1 до примерно 100 раз или более относительно референсного белка CasX с SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3 по сравнению с анализом *in vitro* в сопоставимых условиях.

[0663] Вариант осуществления 37. Вариант CasX по любому из вариантов осуществления 2-35, причем кратность улучшения одной или более из улучшенных характеристик варианта CasX составляет по меньшей мере примерно 1,1, по меньшей мере примерно 2, по меньшей мере примерно 10, по меньшей мере примерно 100 раз или более относительно белка CasX с SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 270 или SEQ ID NO: 336 по сравнению с анализом *in vitro* в сопоставимых условиях.

[0664] Вариант осуществления 38. Вариант CasX по любому из вариантов осуществления 2-37, причем улучшенная характеристика включает в себя эффективность редактирования, и вариант CasX имеет кратность улучшения эффективности редактирования целевой нуклеиновой кислоты от 1,1 до 100 раз по сравнению с белком CasX с SEQ ID NO: 270 или SEQ ID NO: 336 по сравнению с анализом *in vitro* в сопоставимых условиях.

[0665] Вариант осуществления 39. Вариант CasX по любому из вариантов осуществления 1-38, причем RNP, содержащий вариант CasX, демонстрирует более высокие эффективность редактирования и/или связывание последовательности в целевой нуклеиновой кислоте, когда любая из PAM-последовательностей TTC, ATC, GTC или CTC расположена на расстоянии 1 нуклеотида от 5'-конца нецелевой цепи протоспейсера, имеющей идентичность с нацеливающей последовательностью гНК, в системе клеточного анализа *in vitro* по сравнению с эффективностью редактирования и/или связыванием RNP, содержащего референсный белок CasX, в сопоставимой системе анализа.

[0666] Вариант осуществления 40. Вариант CasX по варианту осуществления 39, причем PAM-последовательность представляет собой TTC.

[0667] Вариант осуществления 41. Вариант CasX по варианту осуществления 39, причем PAM-последовательность представляет собой ATC.

[0668] Вариант осуществления 42. Вариант CasX по варианту осуществления 39, причем PAM-последовательность представляет собой CTC.

[0669] Вариант осуществления 43. Вариант CasX по варианту осуществления 39, причем PAM-последовательность представляет собой GTC.

[0670] Вариант осуществления 44. Вариант CasX по любому из вариантов осуществления 39-43, причем улучшенные эффективность редактирования и/или связывание с целевой нуклеиновой кислотой у RNP, содержащего вариант CasX, имеют кратность улучшения от по меньшей мере примерно 1,1 до примерно 100 раз относительно RNP, содержащего референсный CasX.

[0671] Вариант осуществления 45. Вариант CasX по любому из вариантов осуществления 1-44, содержащий от 400 до 2000 аминокислот.

[0672] Вариант осуществления 46. Вариант CasX по любому из вариантов осуществления 1-45, причем вариант белка CasX содержит нуклеазный домен, обладающий никазной активностью.

[0673] Вариант осуществления 47. Вариант CasX по любому из вариантов осуществления 1-45, причем вариант белка CasX содержит нуклеазный домен, обладающий активностью двухцепочечного расщепления.

[0674] Вариант осуществления 48. Вариант CasX по любому из вариантов осуществления 1-37, причем белок CasX представляет собой каталитически неактивный белок CasX (dCasX), и при этом dCasX и гНК сохраняют способность связываться с целевой нуклеиновой кислотой, и необязательно причем белок dCasX содержит последовательность SEQ ID NO: 44-62.

[0675] Вариант осуществления 49. Вариант CasX по варианту осуществления 48, причем dCasX содержит мутацию по остаткам:

a. D672, и/или E769, и/или D935, соответствующим белку CasX с SEQ ID NO: 1; или

b. D659, и/или E756, и/или D922, соответствующим белку CasX с SEQ ID NO: 2.

[0676] Вариант осуществления 50. Вариант CasX по варианту осуществления 49, в котором мутация представляет собой замену остатка на аланин.

[0677] Вариант осуществления 51. Вариант CasX по любому из вариантов осуществления 1-50, содержащий первый домен из первого белка CasX и второй домен из второго белка CasX, отличного от первого белка CasX.

[0678] Вариант осуществления 52. Вариант CasX по варианту осуществления 51, в котором первый домен выбран из группы, состоящей из доменов NTSB, TSL, спирального I, спирального II, OBD и RuvC.

[0679] Вариант осуществления 53. Вариант CasX по варианту осуществления 51, в котором второй домен выбран из группы, состоящей из доменов NTSB, TSL, спирального I, спирального II, OBD и RuvC.

[0680] Вариант осуществления 54. Вариант CasX по любому из вариантов осуществления 51-53, в котором первый и второй домены не являются одним и тем же доменом.

[0681] Вариант осуществления 55. Вариант CasX по любому из вариантов осуществления 1-50, содержащий по меньшей мере один химерный домен, который содержит первую часть из первого белка CasX и вторую часть из второго белка CasX, отличного от первого белка CasX.

[0682] Вариант осуществления 56. Вариант CasX по варианту осуществления 55, в котором по меньшей мере один химерный домен выбран из группы, состоящей из доменов NTSB, TSL, спирального I, спирального II, OBD и RuvC.

[0683] Вариант осуществления 57. Вариант CasX по варианту осуществления 56, в котором по меньшей мере один химерный домен содержит химерный домен RuvC.

[0684] Вариант осуществления 58. Вариант CasX по любому из вариантов осуществления 1-57, содержащий гетерологичный белок или его домен, слитый с CasX.

[0685] Вариант осуществления 59. Вариант CasX по варианту осуществления 58, в котором гетерологичный белок или его домен представляет собой редактор оснований.

[0686] Вариант осуществления 60. Вариант CasX по варианту осуществления 59, в котором редактор оснований представляет собой аденозиндезаминазу, цитозиндезаминазу или гуаниноксидазу.

[0687] Вариант осуществления 61. Вариант каркаса референсной гидовой нуклеиновой кислоты (вариант гНК), способный связывать референсный белок CasX или вариант CasX, причем:

a. вариант гНК содержит по меньшей мере одну модификацию по сравнению с каркасной последовательностью референсной гидовой нуклеиновой кислоты и

b. вариант гНК демонстрирует одну или более улучшенных характеристик по

сравнению с каркасом референсной гидовой нуклеиновой кислоты,

причем необязательно вариант содержит последовательность, выбранную из группы, состоящей из SEQ ID NO: 2101-2332.

[0688] Вариант осуществления 62. Вариант гНК по варианту осуществления 61, причем одна или более улучшенных характеристик выбраны из группы, состоящей из: улучшенной стабильности; улучшенной растворимости; улучшенной транскрипции гНК; улучшенной стойкости к нуклеазной активности; повышенной скорости свертывания гНК; сниженного образования побочных продуктов при свертывании; повышенного продуктивного свертывания; улучшенной аффинности связывания с белком CasX; улучшенной аффинности связывания с целевой нуклеиновой кислотой при нахождении в комплексе с белком CasX; улучшенного редактирования генов при нахождении в комплексе с белком CasX; улучшенной специфичности редактирования при нахождении в комплексе с белком CasX и улучшенной способности образовывать компетентный в отношении расщепления RNP при нахождении в комплексе с белком CasX.

[0689] Вариант осуществления 63. Вариант гНК по варианту осуществления 61 или 62, в котором каркас референсного гида содержит последовательность, выбранную из группы, состоящей из последовательностей SEQ ID NO: 4-16.

[0690] Вариант осуществления 64. Вариант гНК по любому из вариантов осуществления 61-63, в котором по меньшей мере одна модификация содержит:

a. по меньшей мере одну замену нуклеотида в области варианта гНК;

b. по меньшей мере одну делецию нуклеотида в области варианта гНК;

c. по меньшей мере одну вставку нуклеотида в области варианта гНК;

d. полную или частичную замену области варианта гНК;

e. полную или частичную делецию области варианта гНК; или

f. или любую комбинацию пп. (a)-(d).

[0691] Вариант осуществления 65. Вариант гНК по варианту осуществления 64, в котором область варианта гНК выбрана из группы, состоящей из удлиненной стеблевой петли, каркасной стеблевой петли, триплекса и псевдоузла.

[0692] Вариант осуществления 66. Вариант гНК по варианту осуществления 65, в котором каркасный стебель дополнительно содержит пузырек.

[0693] Вариант осуществления 67. Вариант гНК по варианту осуществления 65 или варианту осуществления 66, в котором каркас дополнительно содержит триплексную петлевую область.

[0694] Вариант осуществления 68. Вариант гНК по любому из вариантов осуществления 65-67, в котором каркас дополнительно содержит 5'-неструктурированную область.

[0695] Вариант осуществления 69. Вариант гНК по любому из вариантов осуществления 64-68, в котором по меньшей мере одна модификация содержит:

a. замену 1-15 последовательных или не последовательных нуклеотидов в варианте гНК в одной или более областях;

b. делецию 1-10 последовательных или не последовательных нуклеотидов в варианте гНК в одной или более областях;

c. вставку 1-10 последовательных или не последовательных нуклеотидов в варианте гНК в одной или более областях;

d. замену каркасной стеблевой петли или удлиненной стеблевой петли на последовательность стеблевой петли РНК из гетерологичного РНК-источника с проксимальными 5'- и 3'-концами или вставку в каркасную стеблевую петлю или удлиненную стеблевую петлю последовательности стеблевой петли РНК из гетерологичного РНК-источника с проксимальными 5'- и 3'-концами; или

e. любую комбинацию пп. (a)-(d).

[0696] Вариант осуществления 70. Вариант гНК по любому из вариантов осуществления 61-69, содержащий две или более модификаций в одной области.

[0697] Вариант осуществления 71. Вариант гНК по любому из вариантов осуществления 61-69, содержащий модификации в двух или более областях.

[0698] Вариант осуществления 72. Вариант гНК по любому из вариантов осуществления 61-71, содержащий область удлиненной стеблевой петли, содержащую по меньшей мере 10, по меньшей мере 100, по меньшей мере 500, по меньшей мере 1000 или по меньшей мере 10 000 нуклеотидов.

[0699] Вариант осуществления 73. Вариант гНК по варианту осуществления 72, в котором гетерологичная последовательность стеблевой петли РНК повышает стабильность гНК.

[0700] Вариант осуществления 74. Вариант гНК по варианту осуществления 72 или варианту осуществления 73, в котором гетерологичная стеблевая петля РНК способна связывать белок, структуру РНК, ДНК-последовательность или малую молекулу.

[0701] Вариант осуществления 75. Вариант гНК по варианту осуществления 74, в котором последовательность гетерологичной стеблевой петли РНК, вставленная в удлиненную стеблевую петлю, выбрана из шпильки MS2, шпильки Qβ, шпильки II U1, шпильки Uvsx или шпильки PP7, причем гетерологичная стеблевая петля способна связываться белок оболочки MS2, белок оболочки Qβ, частицу распознавания сигнала U1A, белок Uvsx фага T4 или белок оболочки PP7 соответственно.

[0702] Вариант осуществления 76. Вариант гНК по любому из вариантов осуществления 61-75, в котором модификация включает в себя вставку в удлиненную стеблевую петлю одного или более компонентов, выбранных из группы, состоящей из:

a. стебля IIB из Rev-отвечающего элемента (RRE);

b. стебля II-V из RRE;

c. стебля II из RRE;

d. Rev-связывающего элемента (RBE) стебля IIB и

e. полноразмерного RRE,

причем один или более компонентов способны связывать Rev.

[0703] Вариант осуществления 77. Вариант гНК по любому из вариантов

осуществления 61-76, дополнительно содержащий нацеливающую последовательность, причем нацеливающая последовательность комплементарна последовательности целевой нуклеиновой кислоты.

[0704] Вариант осуществления 78. Вариант гНК по варианту осуществления 77, в котором нацеливающая последовательность имеет 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34 или 35 нуклеотидов.

[0705] Вариант осуществления 79. Вариант гНК по варианту осуществления 78, в котором нацеливающая последовательность имеет 18, 19 или 20 нуклеотидов.

[0706] Вариант осуществления 80. Вариант гНК по варианту осуществления 79, в котором RNP, содержащий вариант гНК, имеющий нацеливающую последовательность из 18 нуклеотидов, демонстрирует эффективность редактирования, увеличенную по меньшей мере в 2 раза, по меньшей мере в 3 раза или по меньшей мере в 4 раза по сравнению с RNP, содержащим вариант гНК, имеющий нацеливающую последовательность из 20 нуклеотидов, при оценке клеточным анализом *in vitro* в сопоставимых условиях.

[0707] Вариант осуществления 81. Вариант гНК по варианту осуществления 79, в котором RNP, содержащий вариант гНК, имеющий нацеливающую последовательность из 19 нуклеотидов, демонстрирует эффективность редактирования, увеличенную по меньшей мере в 2 раза, по меньшей мере в 3 раза или по меньшей мере в 4 раза по сравнению с RNP, содержащим вариант гНК, имеющий нацеливающую последовательность из 20 нуклеотидов, при оценке клеточным анализом *in vitro* в сопоставимых условиях.

[0708] Вариант осуществления 82. Вариант гНК по любому из вариантов осуществления 77-79 или в котором нацеливающая последовательность имеет 20 нуклеотидов.

[0709] Вариант осуществления 83. Вариант гНК по любому из вариантов осуществления 77-80, в котором гНК представляет собой одномолекулярную гНК, содержащую последовательность каркаса, связанную с нацеливающей последовательностью.

[0710] Вариант осуществления 84. Вариант гНК по любому из вариантов осуществления 61-83, причем кратность улучшения одной или более из улучшенных характеристик варианта гНК составляет от по меньшей мере примерно 1,1 до примерно 100 раз или более относительно референсной гНК с SEQ ID NO: 4 или SEQ ID NO: 5.

[0711] Вариант осуществления 85. Вариант гНК по любому из вариантов осуществления 61-83, причем кратность улучшения одной или более из улучшенных характеристик варианта гНК составляет по меньшей мере примерно 1,1, по меньшей мере примерно 2, по меньшей мере примерно 10 или по меньшей мере примерно 100 раз или более относительно референсной гНК с SEQ ID NO: 4, референсной гНК с SEQ ID NO: 5, вариантным каркасом с SEQ ID NO: 2238, вариантным каркасом с SEQ ID NO: 2239, вариантного каркаса 174 (SEQ ID NO: 2238) или вариантного каркаса 175 (SEQ ID NO: 2239).

[0712] Вариант осуществления 86. Вариант гНК по любому из вариантов

осуществления 61-85, содержащий каркасную область, которая имеет идентичность последовательности по меньшей мере 60% с SEQ ID NO: 4 или SEQ ID NO: 5, за исключением области удлиненного стебля.

[0713] Вариант осуществления 87. Вариант гНК по любому из вариантов осуществления 61-85, содержащий каркасную стеблевую петлю, которая имеет идентичность последовательности по меньшей мере 60% с SEQ ID NO: 14.

[0714] Вариант осуществления 88. Вариант гНК по любому из вариантов осуществления 61-85, в котором каркас последовательности варианта гНК имеет идентичность последовательности по меньшей мере 20%, по меньшей мере 30%, по меньшей мере 40%, по меньшей мере 50%, по меньшей мере 60% или по меньшей мере 70%, по меньшей мере 80%, по меньшей мере 85%, по меньшей мере примерно 90%, по меньшей мере примерно 95%, по меньшей мере примерно 96%, по меньшей мере примерно 97%, по меньшей мере примерно 98% или по меньшей мере примерно 99% с SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 2238 или SEQ ID NO: 2239.

[0715] Вариант осуществления 89. Вариант гНК по любому из вариантов осуществления 61-85, в котором каркас последовательности варианта гНК содержит последовательность, выбранную из группы последовательностей SEQ ID NO: 2101-2332 или последовательности, которая имеет идентичность с ней по меньшей мере примерно 80%, по меньшей мере примерно 90%, по меньшей мере примерно 95%, по меньшей мере примерно 96%, по меньшей мере примерно 97%, по меньшей мере примерно 98%, по меньшей мере примерно 99%.

[0716] Вариант осуществления 90. Вариант гНК по варианту осуществления 89, в котором каркас последовательности варианта гНК состоит из последовательности, выбранной из группы, состоящей из SEQ ID NO: 2101-2332.

[0717] Вариант осуществления 91. Вариант гНК по любому из вариантов осуществления 61-90, дополнительно содержащий один или более рибозимов.

[0718] Вариант осуществления 92. Вариант гНК по варианту осуществления 91, в котором один или более рибозимов независимо слиты с концом варианта гНК.

[0719] Вариант осуществления 93. Вариант гНК по варианту осуществления 91 или варианту осуществления 92, в котором по меньшей мере один из одного или более рибозимов представляет собой рибозим вируса гепатита дельта (HDV), рибозим hammerhead, пистолетный рибозим, рибозим hatchet или рибозим вируса кольцевой пятнистости табака (TRSV).

[0720] Вариант осуществления 94. Вариант гНК по любому из вариантов осуществления 61-93, дополнительно содержащий термостабильную стеблевую петлю.

[0721] Вариант осуществления 95. Вариант гНК по любому из вариантов осуществления 61-94, в котором гНК химически модифицирована.

[0722] Вариант осуществления 96. Вариант гНК по любому из вариантов осуществления 61-95, в котором гНК содержит первую область из первой гНК и вторую область из второй гНК, отличной от первой гНК.

[0723] Вариант осуществления 97. Вариант гНК по варианту осуществления 96, в котором первая область выбрана из группы, состоящей из триплексной области, каркасной стеблевой петли и удлиненной стеблевой петли.

[0724] Вариант осуществления 98. Вариант гНК по варианту осуществления 96 или варианту осуществления 97, в котором вторая область выбрана из группы, состоящей из триплексной области, каркасной стеблевой петли и удлиненной стеблевой петли.

[0725] Вариант осуществления 99. Вариант гНК по любому из вариантов осуществления 96-98, в котором первая и вторая области не являются одной и той же областью.

[0726] Вариант осуществления 100. Вариант гНК по любому из вариантов осуществления 96-98, в котором первая гНК содержит последовательность SEQ ID NO: 4, а вторая гНК содержит последовательность SEQ ID NO: 5.

[0727] Вариант осуществления 101. Вариант гНК по любому из вариантов осуществления 61-100, содержащий по меньшей мере одну химерную область, которая содержит первую часть из первой гНК и вторую часть из второй гНК.

[0728] Вариант осуществления 102. Вариант гНК по варианту осуществления 101, в котором по меньшей мере одна химерная область выбрана из группы, состоящей из триплексной области, каркасной стеблевой петли и удлиненной стеблевой петли.

[0729] Вариант осуществления 103. Вариант гНК по варианту осуществления 61, содержащий последовательность с любой из любой из SEQ ID NO: 2101-2332.

[0730] Вариант осуществления 104. Пара для редактирования генов, содержащая вариантный белок CasX и первый вариант гНК.

[0731] Вариант осуществления 105. Пара для редактирования генов по варианту осуществления 104, в которой вариант белка CasX и вариант гНК способны связываться друг с другом в рибонуклеопротеиновом комплексе (RNP).

[0732] Вариант осуществления 106. Пара для редактирования генов по варианту осуществления 104, в которой вариант CasX и вариант гНК связаны друг с другом в рибонуклеопротеиновом комплексе (RNP).

[0733] Вариант осуществления 107. Пара для редактирования генов по любому из вариантов осуществления 104-107, в которой первая гНК содержит вариант гНК по любому из вариантов осуществления 77-103 и нацеливающую последовательность, причем нацеливающая последовательность комплементарна целевой нуклеиновой кислоте.

[0734] Вариант осуществления 108. Пара для редактирования генов по любому из вариантов осуществления 104-107, в которой вариант CasX содержит вариант CasX по любому из вариантов осуществления 1-60.

[0735] Вариант осуществления 109. Пара для редактирования генов по любому из вариантов осуществления 104-108, содержащая:

a. вариант гНК по любому из вариантов осуществления 77-103 и

b. вариант CasX по любому из вариантов осуществления 1-60.

[0736] Вариант осуществления 110. Пара для редактирования генов по варианту

осуществления 109, в которой пара для редактирования генов из варианта CasX и варианта гНК имеет одну или более улучшенных характеристик по сравнению с парой для редактирования генов, содержащей референсный белок CasX с SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3 и референсную гидовую нуклеиновую кислоту с SEQ ID NO: 4 или 5.

[0737] Вариант осуществления 111. Пара для редактирования генов по варианту осуществления 110, причем одна или более улучшенных характеристик включают в себя улучшенную стабильность комплекса CasX : гНК (RNP), улучшенную аффинность связывания между CasX и гНК, улучшенную кинетику образования комплекса RNP, более высокую процентную долю компетентных в отношении расщепления RNP, улучшенную аффинность связывания RNP с целевой нуклеиновой кислотой, способность использовать расширенный спектр PAM-последовательностей, улучшенное раскручивание целевой нуклеиновой кислоты, повышенную активность редактирования, улучшенную эффективность редактирования, улучшенную специфичность редактирования, повышенную нуклеазную активность, повышенную загрузку целевой цепи для двухцепочечного расщепления, сниженную загрузку целевой цепи для одноцепочечного никирования, сниженное нецелевое расщепление, улучшенное связывание нецелевой цепи ДНК или улучшенную стойкость к нуклеазной активности.

[0738] Вариант осуществления 112. Пара для редактирования генов по варианту осуществления 110 или варианту осуществления 111, причем улучшенная характеристика представляет собой повышенную эффективность редактирования целевой нуклеиновой кислоты с использованием неканонической PAM-последовательности - ATC, GTC или CTC - по сравнению с парой для редактирования генов, содержащей референсный белок CasX с SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3 или вариант CasX 119 или 491 и референсную гидовую нуклеиновую кислоту с SEQ ID NO: 4 или 5 или вариант гНК 174 (SEQ ID NO: 2238).

[0739] Вариант осуществления 113. Пара для редактирования генов по варианту осуществления 110, причем кратность улучшения по меньшей мере одной или более из улучшенных характеристик составляет от по меньшей мере примерно 1,1 раза до примерно 100 раз или более относительно пары для редактирования генов из референсного белка CasX и референсной гидовой нуклеиновой кислоты.

[0740] Вариант осуществления 114. Пара для редактирования генов по варианту осуществления 110, причем кратность улучшения одной или более из улучшенных характеристик варианта CasX составляет по меньшей мере примерно в 1,1 раза, по меньшей мере примерно в 2 раза, по меньшей мере примерно в 4 раза, по меньшей мере примерно в 6 раз, по меньшей мере примерно в 6 раз, по меньшей мере примерно в 10 раз или по меньшей мере примерно в 100 раз или более относительно пары для редактирования генов из референсного белка CasX и референсной гидовой нуклеиновой кислоты или варианта гНК 174 (SEQ ID NO: 2238).

[0741] Вариант осуществления 115. Пара для редактирования генов по варианту осуществления 110, причем улучшенная характеристика имеет кратность повышения

эффективности редактирования 4-9 раз по сравнению с референсной парой для редактирования с SEQ ID NO: 2 и SEQ ID NO: 5 или вариантом гНК 174 (SEQ ID NO: 2238) при анализе *in vitro* в сопоставимых условиях.

[0742] Вариант осуществления 116. Пара для редактирования генов по любому из вариантов осуществления 104-115, в которой RNP из варианта CasX и варианта гНК имеет более высокую процентную долю компетентных в отношении расщепления RNP по сравнению с RNP из референсного белка CasX и референсной гидовой нуклеиновой кислоты.

[0743] Вариант осуществления 117. Пара для редактирования генов по п. 116, в которой RNP из варианта CasX и варианта гНК имеет по меньшей мере в 2 раза, по меньшей мере в 3 раза, по меньшей мере в 4 раза или по меньшей мере в 5 раз более высокую процентную долю компетентных в отношении расщепления RNP по сравнению с RNP из референсного белка CasX и референсной гидовой нуклеиновой кислоты.

[0744] Вариант осуществления 118. Пара для редактирования генов по любому из вариантов осуществления 104-117, в которой RNP способен связывать целевую нуклеиновую кислоту и расщеплять ее.

[0745] Вариант осуществления 119. Пара для редактирования генов по любому из вариантов осуществления 104-118, в которой RNP из варианта CasX и варианта гНК демонстрирует соотношение специфичности в пределах по меньшей мере 10%, или по меньшей мере 15%, или по меньшей мере 20% по сравнению с референсной парой для редактирования с SEQ ID NO: 2 и SEQ ID NO: 5 или вариантом гНК 174 (SEQ ID NO: 2238) при анализе *in vitro* целевого/нецелевого редактирования в сопоставимых условиях.

[0746] Вариант осуществления 120. Пара для редактирования генов по любому из вариантов осуществления 104-111, в которой RNP способен связывать целевую нуклеиновую кислоту, но не способен расщеплять целевую нуклеиновую кислоту.

[0747] Вариант осуществления 121. Пара для редактирования генов по любому из вариантов осуществления 104-117, в которой RNP способен связывать целевую нуклеиновую кислоту и создавать один или более одноцепочечных ников в целевой нуклеиновой кислоте.

[0748] Вариант осуществления 122. Композиция, содержащая пару для редактирования генов по любому из вариантов осуществления 104-119, содержащая:

a. вторую пару для редактирования генов, содержащую вариант CasX по любому из вариантов осуществления 1-60; и

b. второй вариант гНК по любому из вариантов осуществления 61-103, причем второй вариант гНК имеет нацеливающую последовательность, комплементарную другой или перекрывающейся части целевой нуклеиновой кислоты по сравнению с нацеливающей последовательностью первой гНК.

[0749] Вариант осуществления 123. Вариант CasX, содержащий аминокислотную последовательность с любой из SEQ ID NO: 89-101, 247-337, 411-592 или 760-982.

[0750] Вариант осуществления 124. Вариант гНК, содержащий аминокислотную

последовательность с любой из SEQ ID NO: 2101-2332.

[0751] Вариант осуществления 125. Пара для редактирования генов по любому из вариантов осуществления 104-121 для применения в качестве лекарственного средства для лечения субъекта, имеющего заболевание.

Набор IV

[0752] Вариант осуществления 1. Каркас гидовой РНК (гРНК), содержащий последовательность, имеющую идентичность последовательности по меньшей мере 60%, по меньшей мере 70%, по меньшей мере 80%, по меньшей мере 90%, по меньшей мере 95%, по меньшей мере 98%, по меньшей мере 99% или 100% с любой из последовательностей, выбранных из группы, состоящей из SEQ ID NO: 2292, 2291, 2307, 2281-2290, 2293-2306, 2308-2332 и 23530-2398.

[0753] Вариант осуществления 2. Каркас гРНК по варианту осуществления 1, содержащий последовательность, выбранную из группы, состоящей из SEQ ID NO: 2292, 2291, 2307, 2281-2290, 2293-2306, 2308-2332 и 23530-2398.

[0754] Вариант осуществления 3. Каркас гРНК по варианту осуществления 1, содержащий последовательность, имеющую одну или более модификаций относительно SEQ ID NO: 2238, причем одна или более модификаций приводят к получению улучшенной характеристики.

[0755] Вариант осуществления 4. Каркас гРНК по варианту осуществления 3, в котором одна или более модификаций включают в себя одну или более нуклеотидных замен, вставок и/или делеций, представленных в таблице 19.

[0756] Вариант осуществления 5. Каркас гРНК по варианту осуществления 3 или варианту осуществления 4, причем улучшенная характеристика представляет собой одно или более функциональных свойств, выбранных из группы, состоящей из повышенной активности редактирования, повышенной стабильности стебля псевдоузла, повышенной стабильности триплексной области, повышенной стабильности каркасного стебля, стабильности удлиненного стебля, сниженного количества нецелевых интермедиатов свертывания и повышенной аффинности связывания с белком CRISPR класса 2, типа V, необязательно при анализе *in vitro*.

[0757] Вариант осуществления 6. Каркас гРНК по любому из вариантов осуществления 3-5, демонстрирующий улучшенную оценку обогащения ($\log_2$), которая на по меньшей мере примерно 2,0, по меньшей мере примерно 2,5, по меньшей мере примерно 3 или по меньшей мере примерно 3,5 выше, чем оценка каркаса гРНК с SEQ ID NO: 2238 в анализе *in vitro*.

[0758] Вариант осуществления 7. Каркас гРНК по варианту осуществления 1, содержащий последовательность, имеющую одну или более модификаций относительно SEQ ID NO: 2239, причем одна или более модификаций приводят к получению улучшенной характеристики.

[0759] Вариант осуществления 8. Каркас гРНК по варианту осуществления 7, в котором одна или более модификаций включают в себя одну или более нуклеотидных

замен, вставок и/или делеций, представленных в таблице 20.

[0760] Вариант осуществления 9. Каркас гРНК по варианту осуществления 7 или варианту осуществления 8, причем улучшенная характеристика представляет собой одно или более функциональных свойств, выбранных из группы, состоящей из повышенной активности редактирования, повышенной стабильности стебля псевдоузла, повышенной стабильности триплексной области, повышенной стабильности каркасного стебля, стабильности удлиненного стебля, сниженного количества нецелевых интермедиатов свертывания и повышенной аффинности связывания с белком CRISPR класса 2, типа V, необязательно при анализе *in vitro*.

[0761] Вариант осуществления 10. Каркас гРНК по любому из вариантов осуществления 7-9, демонстрирующий улучшенную оценку обогащения (log2), которая на по меньшей мере примерно 1,2, по меньшей мере примерно 1,5, по меньшей мере примерно 2,0, по меньшей мере примерно 2,5, по меньшей мере примерно 3 или по меньшей мере примерно 3,5 выше, чем оценка каркаса гРНК с SEQ ID NO: 2239 в анализе *in vitro*.

[0762] Вариант осуществления 11. Каркас гРНК по варианту осуществления 1, содержащий одну или более модификаций в положениях относительно последовательности SEQ ID NO: 2239, выбранных из группы, состоящей из C9, U11, C17, U24, A29, U54, G64, A88 и A95.

[0763] Вариант осуществления 12. Каркас гРНК по варианту осуществления 11, содержащий одну или более модификаций относительно последовательности SEQ ID NO: 2239, выбранных из группы, состоящей из C9U, U11C, C17G, U24C, A29C, вставки G в положении 54, вставки C в положении 64, A88G и A95G.

[0764] Вариант осуществления 13. Каркас гРНК по варианту осуществления 12, содержащий модификации относительно последовательности SEQ ID NO: 2239, состоящие из C9U, U11C, C17G, U24C, A29C, вставки G в положении 54, вставки C в положении 64, A88G и A95G.

[0765] Вариант осуществления 14. Каркас гРНК по любому из вариантов осуществления 7-13, причем улучшенная характеристика выбрана из группы, состоящей из стабильности стебля псевдоузла, стабильности триплексной области, стабильности пузырьков каркаса, стабильности удлиненного стебля и аффинности связывания с белком CRISPR класса 2, типа V.

[0766] Вариант осуществления 15. Каркас гРНК по варианту осуществления 14, в котором вставка C в положении 64 и замена A88G относительно последовательности SEQ ID NO: 2239 устраняют асимметричный элемент утолщения удлиненного стебля, повышая стабильность удлиненного стебля каркаса гРНК.

[0767] Вариант осуществления 16. Каркас гРНК по варианту осуществления 14, в котором замены U11C, U24C и A95G повышают стабильность триплексной области каркаса гРНК.

[0768] Вариант осуществления 17. Каркас гРНК по варианту осуществления 14, в котором замена A29C повышает стабильность стебля псевдоузла.

[0769] Вариант осуществления 18. Каркас гРНК по варианту осуществления 1 или варианту осуществления 2, в котором каркас гРНК содержит одну или более гетерологичных последовательностей РНК в удлиненном стебле.

[0770] Вариант осуществления 19. Каркас гРНК по варианту осуществления 18, в котором гетерологичная РНК выбрана из группы, состоящей из шпильки MS2, шпильки Qβ, шпильки II U1, шпильки Uvsx и стебля с петлей PP7 или вариантов их последовательностей.

[0771] Вариант осуществления 20. Каркас гРНК по варианту осуществления 18 или варианту осуществления 19, в котором гетерологичная последовательность РНК повышает стабильность гРНК.

[0772] Вариант осуществления 21. Каркас гРНК по варианту осуществления 18 или варианту осуществления 19, в котором гетерологичная РНК способна связывать белок, РНК, ДНК или малую молекулу.

[0773] Вариант осуществления 22. Каркас гРНК по любому из вариантов осуществления 18-21, в котором каркас гРНК содержит Rev-отвечающий элемент (RRE) или его часть.

[0774] Вариант осуществления 23. Каркас гРНК по варианту осуществления 22, в котором RRE или его часть выбраны из группы, состоящей из стебля IIB из RRE, имеющего последовательность UGGGCGCAGCGUCAAUGACGCUGACGGUACA (SEQ ID NO: 1280), стебля II-V из RRE, имеющего последовательность CAGGAAGCACUAUGGGCGCAGCGUCAAUGACGCUGACGGUACAGGCCAGACAAUU AUUGUCUGGUAUAGUGCAGCAGCAGAACAAUUUGCUGAGGGCUAUUGAGGCGCA ACAGCAUCUGUUGCAACUCACAGUCUGGGGCAUCAAGCAGCUCCAGGCAAGAAUC CUG (SEQ ID NO: 1282), стебля II из RRE, имеющего последовательность GCACUAUGGGCGCAGCGUCAAUGACGCUGACGGUACAGGCCAGACAAUUAUUGUC UGGUAUAGUGC (SEQ ID NO: 1281), Rev-связывающего элемента (RBE) стебля IIB, имеющего последовательность GCUGACGGUACAGGC (SEQ ID NO: 1284), и полноразмерного RRE, имеющего последовательность AGGAGCUUUGUUCCUUGGGUUCUUGGGAGCAGCAGGAAGCACUAUGGGCGCAGC GUCAAUGACGCUGACGGUACAGGCCAGACAAUUAUUGUCUGGUAUAGUGCAGCA GCAGAACAAUUUGCUGAGGGCUAUUGAGGCGCAACAGCAUCUGUUGCAACUCAC AGUCUGGGGCAUCAAGCAGCUCCAGGCAAGAAUCCUGGCUGUGGAAAGAUACCU AAAGGAUCAACAGCUCCU (SEQ ID NO: 1283).

[0775] Вариант осуществления 24. гРНК по любому из вариантов осуществления 1-23, в которой каркас гРНК содержит один или более тиминов (T).

[0776] Вариант осуществления 25. гРНК, содержащая каркас гРНК по любому из вариантов осуществления 1-24 и нацеливающую последовательность на 3'-конце каркаса гРНК, которая комплементарна последовательности целевой нуклеиновой кислоты.

[0777] Вариант осуществления 26. гРНК по варианту осуществления 25, в которой нацеливающая последовательность имеет 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 или 30 нуклеотидов.

[0778] Вариант осуществления 27. гРНК по варианту осуществления 26, в которой нацеливающая последовательность имеет 18, 19 или 20 нуклеотидов.

[0779] Вариант осуществления 28. гРНК по любому из вариантов осуществления 25-27, способная образовывать рибонуклеопротеиновый (RNP) комплекс с белком CRISPR класса 2, типа V.

[0780] Вариант осуществления 29. Сконструированный белок CRISPR класса 2, типа V, содержащий:

a. домен NTSB, содержащий последовательность QPASKKIDQNKLKPEMDEKGNLTTAGFACSQCGQPLFVYKLEQVSEKGKAYTNYFGRC NVAEHEKLILLAQLKPEKDSDEAVTYSLGKFGQ (SEQ ID NO: 2335) или последовательность, имеющую идентичность с ней по меньшей мере 80%, по меньшей мере 90%, по меньшей мере 95%, по меньшей мере 96%, по меньшей мере 97%, по меньшей мере 98% или по меньшей мере 99%;

b. домен спиральный I-II, содержащий последовательность RALDFYSIHVTKESTHPVKPLAQIAGNRYASGPVGKALSDACMGTIASFLSKYQDIIIEHQ KVVKGNQKRLESLRELAGKENLEYPSVTLPPQPHTKEGVDAYNEVIARVRMWVNLNL WQKLKLSRDDAKPLLRLKGFPSF (SEQ ID NO: 2336) или последовательность, имеющую идентичность с ней по меньшей мере 80%, по меньшей мере 90%, по меньшей мере 95%, по меньшей мере 96%, по меньшей мере 97%, по меньшей мере 98% или по меньшей мере 99%;

c. домен спиральный II, содержащий последовательность PLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAGYKRQEALRPYLSSEEDR KKGKKFARYQLGDLLLHLEKKHGEDWGKVYDEAWERIDKKVEGLSKHIKLEEERRSE DAQSKAALTDWLRAKASFVIEGLKEADKDEFCRCELKLQKWYGDLRGKPFAIEAE (SEQ ID NO: 2351) или последовательность, имеющую идентичность с ней по меньшей мере 80%, по меньшей мере 90%, по меньшей мере 95%, по меньшей мере 96%, по меньшей мере 97%, по меньшей мере 98% или по меньшей мере 99%; и

d. домен RuvC-I, содержащий последовательность SSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPTHILRIGESYKEKQRTIQAK KEVEQRRAGGYSRKYASKAKNLADDMVRNTARDLLYYAVTQDAMLIFENLSRGFGRQ GKRTFMAERQYTRMEDWLTAKLAYEGLPSKTYLSKTLAQYTSKTC (SEQ ID NO: 2352) или последовательность, имеющую идентичность с ней по меньшей мере 80%, по меньшей мере 90%, по меньшей мере 95%, по меньшей мере 96%, по меньшей мере 97%, по меньшей мере 98% или по меньшей мере 99%.

[0781] Вариант осуществления 30. Белок CRISPR класса 2, типа V по варианту осуществления 29, содержащий домен OBD-I, который содержит последовательность QEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPENIPQ (SEQ ID NO: 2342) или последовательность, имеющую идентичность с ней по меньшей мере 80%, по меньшей мере 90%, по меньшей мере 95%, по меньшей мере 96%, по меньшей мере 97%, по меньшей мере 98% или по меньшей мере 99%.

[0782] Вариант осуществления 31. Белок CRISPR класса 2, типа V по варианту осуществления 29 или варианту осуществления 30, содержащий домен OBD-II, который содержит последовательность NSILDISGFSKQYNCAFIWQKDGVKKLNLYLIINYFKGGKLRFKKIKPEAFEANRFYTVIN KKSGEIVPMEVNFNFDDPNLIILPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLY NRRTRQDEPALFVALTFERREVLD (SEQ ID NO: 2347) или последовательность, имеющую идентичность с ней по меньшей мере 80%, по меньшей мере 90%, по меньшей мере 95%, по меньшей мере 96%, по меньшей мере 97%, по меньшей мере 98% или по меньшей мере 99%.

[0783] Вариант осуществления 32. Белок CRISPR класса 2, типа V по любому из вариантов осуществления 29-31, содержащий домен спиральный I-I, который содержит последовательность PISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVA (SEQ ID NO: 2343) или последовательность, имеющую идентичность с ней по меньшей мере 80%, по меньшей мере 90%, по меньшей мере 95%, по меньшей мере 96%, по меньшей мере 97%, по меньшей мере 98% или по меньшей мере 99%.

[0784] Вариант осуществления 33. Белок CRISPR класса 2, типа V по любому из вариантов осуществления 29-32, содержащий домен TSL, который содержит последовательность SNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQITYYNRYKRQNVVKDLSV ELDRLSEESVNNDISSWTKGRSGEALSLLKKRFSHRPVQEKFVCLNCGFETH (SEQ ID NO: 2349) или последовательность, имеющую идентичность с ней по меньшей мере 80%, по меньшей мере 90%, по меньшей мере 95%, по меньшей мере 96%, по меньшей мере 97%, по меньшей мере 98% или по меньшей мере 99%.

[0785] Вариант осуществления 34. Белок CRISPR класса 2, типа V по любому из вариантов осуществления 29-33, содержащий домен RuvC-II, который содержит последовательность ADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDKRAFVETWQSFYRKKLKEVWKPA V (SEQ ID NO: 2350) или последовательность, имеющую идентичность с ней по меньшей мере 80%, по меньшей мере 90%, по меньшей мере 95%, по меньшей мере 96%, по меньшей мере 97%, по меньшей мере 98% или по меньшей мере 99%.

[0786] Вариант осуществления 35. Белок CRISPR класса 2, типа V по варианту осуществления 34, содержащий последовательность SEQ ID NO: 416 или последовательность, которая имеет идентичность с ней по меньшей мере 80%, по меньшей мере 90%, по меньшей мере 95%, по меньшей мере 96%, по меньшей мере 97%, по меньшей мере 98% или по меньшей мере 99%.

[0787] Вариант осуществления 36. Белок CRISPR класса 2, типа V по любому из вариантов осуществления 29-35, содержащий по меньшей мере одну модификацию в одном или более доменах.

[0788] Вариант осуществления 37. Белок CRISPR класса 2, типа V по варианту осуществления 36, причем по меньшей мере одна модификация включает в себя:

a. по меньшей мере одну замену аминокислоты в домене;

b. по меньшей мере одну делецию аминокислоты в домене;

c. по меньшей мере одну вставку аминокислоты в домене или

d. любую комбинацию пп. (a)-(c).

[0789] Вариант осуществления 38. Белок CRISPR класса 2, типа V по варианту осуществления 36 или варианту осуществления 37, содержащий модификацию в одном или более положениях аминокислот в домене NTSB относительно SEQ ID NO: 2335, выбранных из группы, состоящей из P2, S4, Q9, E15, G20, G33, L41, Y51, F55, L68, A70, E75, K88 и G90.

[0790] Вариант осуществления 39. Белок CRISPR класса 2, типа V по варианту осуществления 38, в котором одна или более модификаций в одном или более положениях аминокислот в домене NTSB выбраны из группы, состоящей из вставки G в положении 2, вставки I в положении 4, вставки L в положении 4, Q9P, E15S, G20D, делеции S в положении 30, G33T, L41A, Y51T, F55V, L68D, L68E, L68K, A70Y, A70S, E75A, E75D, E75P, K88Q и G90Q относительно SEQ ID NO: 2335.

[0791] Вариант осуществления 40. Белок CRISPR класса 2, типа V по любому из вариантов осуществления 36-39, содержащий модификацию в одном или более положениях аминокислот в домене спиральный I-II относительно SEQ ID NO: 2336, выбранных из группы, состоящей из I24, A25, Y29 G32, G44, S48, S51, Q54, I56, V63, S73, L74, K97, V100, M112, L116, G137, F138 и S140.

[0792] Вариант осуществления 41. Белок CRISPR класса 2, типа V по варианту осуществления 40, в котором одна или более модификаций в одном или более положениях аминокислот в домене спиральный I-II выбраны из группы, состоящей из вставки T в положении 24, вставки C в положении 25, Y29F, G32Y, G32N, G32H, G32S, G32T, G32A, G32V, делеции G в положении 32, G32S, G32T, G44L, G44H, S48H, S48T, S51T, Q54H, I56T, V63T, S73H, L74Y, K97G, K97S, K97D, K97E, V100L, M112T, M112W, M112R, M112K, L116K, G137R, G137K, G137N, вставки Q в положении 138 и S140Q относительно SEQ ID NO: 2336.

[0793] Вариант осуществления 42. Белок CRISPR класса 2, типа V по любому из вариантов осуществления 36-41, содержащий модификацию в одном или более положениях аминокислот в домене спиральный II относительно SEQ ID NO: 2351, выбранных из группы, состоящей из L2, V3, E4, R5, Q6, A7, E9, V10, D11, W12, W13, D14, M15, V16, C17, N18, V19, K20, L22, I23, E25, K26, K31, Q35, L37, A38, K41,R 42, Q43, E44, L46, K57, Y65, G68, L70, L71, L72, E75, G79, D81, W82, K84, V85, Y86, D87, I93, K95, K96, E98, L100, K102, I104, K105, E109, R110, D114, K118, A120, L121, W124, L125, R126, A127, A129, I133, E134, G135, L136, E138, D140, K141, D142, E143, F144, C145, C147, E148, L149, K150, L151, Q152, K153, L158, E166 и A167.

[0794] Вариант осуществления 43. Белок CRISPR класса 2, типа V по варианту осуществления 42, в котором одна или более модификаций в одном или более положениях аминокислот в домене спиральный II выбраны из группы, состоящей из вставки A в

положении 2, вставки H в положении 2, делеции L в положении 2 и делеции V в положении 3, V3E, V3Q, V3F, делеции V в положении 3, вставки D в положении 3, V3P, E4P, делеции E в положении 4, E4D, E4L, E4R, R5N, Q6V, вставки Q в положении 6, вставки G в положении 7, вставки H в положении 9, вставки A в положении 9, VD10, вставки T1 в положении 0, делеции V в положении 10, вставки F в положении 10, вставки D в положении 11, делеции D в положении 11, D11S, делеции W в положении 12, W12T, W12H, вставки P в положении 12, вставки Q в положении 13, вставки G в положении 12, вставки R в положении 13, W13P, W13D, вставки D в положении 13, W13L, вставки P в положении 14, вставки D в положении 14, делеции D в положении 14 и делеции M в положении 15, делеции M в положении 15, вставки T в положении 16, вставки P в положении 17, N18I, V19N, V19H, K20D, L22D, I23S, E25C, E25P, вставки G в положении 25, K26T, K27E, K31L, K31Y, Q35D, Q35P, вставки S в положении 37, делеции L в положении 37 и делеции A в положении 38, K41L, вставки R в положении 42, делеции Q в положении 43 и делеции E в положении 44, L46N, K57Q, Y65T, G68M, L70V, L71C, L72D, L72N, L72W, L72Y, E75F, E75L, E75Y, G79P, вставки E в положении 79, вставки T в положении 81, вставки R в положении 81, вставки W в положении 81, вставки Y в положении 81, вставки W в положении 82, вставки Y в положении 82, W82G, W82R, K84D, K84H, K84P, K84T, V85L, V85A, вставки L в положении 85, Y86C, D87G, D87M, D87P, I93C, K95T, K96R, E98G, L100A, K102H, I104T, I104S, I104Q, K105D, вставки K в положении 109, E109L, R110D, делеции R в положении 110, D114E, вставки D в положении 114, K118P, A120R, L121T, W124L, L125C, R126D, A127E, A127L, A129T, A129K, I133E, вставки C в положении 133, вставки S в положении 134, вставки G в положении 134, вставки R в положении 135, G135P, L136K, L136D, L136S, L136H, делеции E в положении 138, D140R, вставки D в положении 140, вставки P в положении 141, вставки D в положении 142, делеции E в положении 143+делеции F в положении 144, вставки Q в положении 143, F144K, делеции F в положении 144, делеции F в положении 144 и делеции C в положении 145, C145R, вставки G в положении 145, C145K, C147D, вставки V в положении 148, E148D, вставки H в положении 149, L149R, K150R, L151H, Q152C, K153P, L158S, E166L и вставки F в положении 167 относительно SEQ ID NO: 2351.

[0795] Вариант осуществления 44. Белок CRISPR класса 2, типа V по любому из вариантов осуществления 36-43, содержащий модификацию в одном или более положениях аминокислот в домене RuvC-I относительно SEQ ID NO: 2352, выбранных из группы, состоящей из I4, K5, P6, M7, N8, L9, V12, G49, K63, K80, N83, R90, M125 и L146.

[0796] Вариант осуществления 45. Белок CRISPR класса 2, типа V по варианту осуществления 44, в котором одна или более модификаций в одном или более положениях аминокислот в домене RuvC-I выбраны из группы, состоящей из вставки I в положении 4, вставки S в положении 5, вставки T в положении 6, вставки N в положении 6, вставки R в положении 7, вставки K в положении 7, вставки H в положении 8, вставки S в положении 8, V12L, G49W, G49R, S51R, S51K, K62S, K62T, K62E, V65A, K80E, N83G, R90H, R90G, M125S, M125A, L137Y, вставки P в положении 137, делеции L в положении 141, L141R,

L141D, вставки Q в положении 142, вставки R в положении 143, вставки N в положении 143, E144N, вставки P в положении 146, L146F, P147A, K149Q, T150V, вставки R в положении 152, вставки H153, T155Q, вставки H в положении 155, вставки R в положении 155, вставки L в положении 156, делеции L в положении 156, вставки W в положении 156, вставки A в положении 157, вставки F в положении 157, A157S, Q158K, делеции Y в положении 159, T160Y, T160F, вставки I в положении 161, S161P, T163P, вставки N в положении 163, C164K и C164M относительно SEQ ID NO: 2351.

[0797] Вариант осуществления 46. Белок CRISPR класса 2, типа V по любому из вариантов осуществления 36-45, содержащий модификацию в одном или более положениях аминокислот в домене OBD-I относительно SEQ ID NO: 2342, выбранных из группы, состоящей из I3, K4, R5, I6, N7, K8, K15, D16, N18, P27, M28, V33, R34, M36, R41, L47, R48, E52, P55 и Q56.

[0798] Вариант осуществления 47. Белок CRISPR класса 2, типа V по варианту осуществления 46, в котором одна или более модификаций в одном или более положениях аминокислот в домене OBD-I выбраны из группы, состоящей из вставки G в положении 3, I3G, I3E, вставки G в положении 4, K4G, K4P, K4S, K4W, K4W, R5P, вставки P в положении 5, вставки G в положении 5, R5S, вставки S в положении 5, R5A, R5P, R5G, R5L, I6A, I6L, вставки G в положении 6, N7Q, N7L, N7S, K8G, K15F, D16W, вставки F в положении 16, вставки F18, вставки P в положении 27, M28P, M28H, V33T, R34P, M36Y, R41P, L47P, вставки P в положении 48, E52P, вставки P в положении 55, делеции P в положении 55 и делеции Q в положении 56, Q56S, Q56P, вставки D в положении 56, вставки T в положении 56 и Q56P относительно SEQ ID NO: 2342.

[0799] Вариант осуществления 48. Белок CRISPR класса 2, типа V по любому из вариантов осуществления 36-47, содержащий модификацию в одном или более положениях аминокислот в домене OBD-II относительно SEQ ID NO: 2347, выбранных из группы, состоящей из S2, I3, L4, K11, V24, K37, R42, A53, T58, K63, M70, I82, Q92, G93, K110, L121, R124, R141, E143, V144 и L145.

[0800] Вариант осуществления 49. Белок CRISPR класса 2, типа V по варианту осуществления 48, в котором одна или более модификаций в одном или более положениях аминокислот в домене OBD-II выбраны из группы, состоящей из делеции S в положении 2, I3R, I3K, делеции I в положении 3 и делеции L4, делеции L в положении 4, K11T, вставки P в положении 24, K37G, R42E, вставки S в положении 53, вставки R в положении 58, делеции K в положении 63, M70T, I82T, Q92I, Q92F, Q92V, Q92A, вставки A в положении 93, K110Q, R115Q, L121T, вставки A в положении 124, вставки R в положении 141, вставки D в положении 143, вставки A в положении 143, вставки W в положении 144 и вставки A в положении 145 относительно SEQ ID NO: 2342.

[0801] Вариант осуществления 50. Белок CRISPR класса 2, типа V по любому из вариантов осуществления 36-49, содержащий модификацию в одном или более положениях аминокислот в домене TSL относительно SEQ ID NO: 2349, выбранных из группы, состоящей из S1, N2, C3, G4, F5, I7, K18, V58, S67, T76, G78, S80, G81, E82, S85, V96 и E98.

[0802] Вариант осуществления 51. Белок CRISPR класса 2, типа V по варианту осуществления 50, в котором одна или более модификаций в одном или более положениях аминокислот в домене OBD-II выбраны из группы, состоящей из вставки M в положении 1, делеции N в положении 2, вставки V в положении 2, C3S, вставки G в положении 4, вставки W в положении 4, F5P, вставки W в положении 7, K18G, V58D, вставки A в положении 67, T76E, T76D, T76N, G78D, делеции S в положении 80, делеции G в положении 81, вставки E в положении 82, вставки N в положении 82, S85I, V96C, V96T и E98D относительно SEQ ID NO: 2349.

[0803] Вариант осуществления 52. Белок CRISPR класса 2, типа V по любому из вариантов осуществления 29-51, демонстрирующий улучшенную характеристику относительно SEQ ID NO: 2, причем улучшенная характеристика включает в себя повышенную аффинность связывания с гРНК, повышенную аффинность связывания с целевой нуклеиновой кислотой, улучшенную способность использовать более широкий спектр PAM-последовательностей при редактировании целевой нуклеиновой кислоты, улучшенное раскручивание целевой нуклеиновой кислоты, повышенную активность редактирования, улучшенную эффективность редактирования, улучшенную специфичность редактирования для расщепления целевой нуклеиновой кислоты, сниженное нецелевое редактирование или расщепление целевой нуклеиновой кислоты, повышенную процентную долю эукариотического генома, которая может быть отредактирована, повышенную активность нуклеазы, повышенную загрузку целевой цепи для двухцепочечного расщепления, сниженную загрузку целевой цепи для одноцепочечного никирования, повышенное связывание нецелевой цепи ДНК, улучшенную стабильность белка, повышенную стабильность комплекса белок : гРНК (RNP) и улучшенные характеристики слияния.

[0804] Вариант осуществления 53. Белок CRISPR класса 2, типа V по варианту осуществления 52, причем улучшенная характеристика включает в себя повышенную активность расщепления последовательности целевой нуклеиновой кислоты, содержащей PAM-последовательность TTC, ATC, GTC или CTC.

[0805] Вариант осуществления 54. Белок CRISPR класса 2, типа V по варианту осуществления 53, причем улучшенная характеристика включает в себя повышенную активность расщепления последовательности целевой нуклеиновой кислоты, содержащей PAM-последовательность ATC или CTC, относительно активности расщепления последовательности SEQ ID NO: 416.

[0806] Вариант осуществления 55. Белок CRISPR класса 2, типа V по варианту осуществления 54, причем улучшенная активность расщепления представляет собой оценку обогащения (log2), которая на по меньшей мере примерно 1,5, по меньшей мере примерно 2,0, по меньшей мере примерно 2,5, по меньшей мере примерно 3, по меньшей мере примерно 3,5, по меньшей мере примерно 4, по меньшей мере примерно 4,5, по меньшей мере примерно 5, по меньшей мере примерно 6, по меньшей мере примерно 7, по меньшей мере примерно 8 или более выше, чем оценка последовательности SEQ ID NO:

416 в анализе *in vitro*.

[0807] Вариант осуществления 56. Белок CRISPR класса 2, типа V по варианту осуществления 54, причем улучшенная характеристика включает в себя повышенную активность расщепления последовательности целевой нуклеиновой кислоты, содержащей PAM-последовательность CTC, относительно последовательности SEQ ID NO: 416.

[0808] Вариант осуществления 57. Белок CRISPR класса 2, типа V по варианту осуществления 56, причем улучшенная активность расщепления представляет собой оценку обогащения (log2), которая на по меньшей мере примерно 2, по меньшей мере примерно 2,5, по меньшей мере примерно 3, по меньшей мере примерно 3,5, по меньшей мере примерно 4, по меньшей мере примерно 4,5, по меньшей мере примерно 5 или по меньшей мере примерно 6 или более выше, чем оценка последовательности SEQ ID NO: 416 в анализе *in vitro*.

[0809] Вариант осуществления 58. Белок CRISPR класса 2, типа V по варианту осуществления 53, причем улучшенная характеристика включает в себя повышенную активность расщепления последовательности целевой нуклеиновой кислоты, содержащей PAM-последовательность TTC, относительно последовательности SEQ ID NO: 416.

[0810] Вариант осуществления 59. Белок CRISPR класса 2, типа V по варианту осуществления 58, причем улучшенная активность расщепления представляет собой оценку обогащения, которая на по меньшей мере примерно 1,5, по меньшей мере примерно 2,0, по меньшей мере примерно 2,5, по меньшей мере примерно 3, по меньшей мере примерно 3,5, по меньшей мере примерно 4, по меньшей мере примерно 4,5, по меньшей мере примерно 5 или по меньшей мере примерно 6 log2 или более выше, чем у последовательности SEQ ID NO: 416 в анализе *in vitro*.

[0811] Вариант осуществления 60. Белок CRISPR класса 2, типа V по варианту осуществления 52, причем улучшенная характеристика включает в себя повышенную специфичность расщепления последовательности целевой нуклеиновой кислоты относительно последовательности SEQ ID NO: 416.

[0812] Вариант осуществления 61. Белок CRISPR класса 2, типа V по варианту осуществления 60, в котором повышенная специфичность представляет собой оценку обогащения, которая на по меньшей мере примерно 2,0, по меньшей мере примерно 2,5, по меньшей мере примерно 3, по меньшей мере примерно 3,5, по меньшей мере примерно 4, по меньшей мере примерно 4,5, по меньшей мере примерно 5 или по меньшей мере примерно 6 log2 или более выше, чем у последовательности SEQ ID NO: 416 в анализе *in vitro*.

[0813] Вариант осуществления 62. Белок CRISPR класса 2, типа V по варианту осуществления 52, причем улучшенная характеристика включает в себя сниженное нецелевое расщепление последовательности целевой нуклеиновой кислоты.

[0814] Вариант осуществления 63. Белок CRISPR класса 2, типа V по любому из вариантов осуществления 29-62, имеющий последовательность, выбранную из группы, состоящей из последовательностей SEQ ID NO: 415-592 и 1147-1231, как указано в таблице

3, или последовательность, имеющую идентичность последовательности с ней по меньшей мере примерно 50%, по меньшей мере примерно 60%, по меньшей мере примерно 70%, по меньшей мере примерно 80%, по меньшей мере примерно 90%, или по меньшей мере примерно 95%, или по меньшей мере примерно 96%, или по меньшей мере примерно 97%, или по меньшей мере примерно 98%, или по меньшей мере примерно 99%.

[0815] Вариант осуществления 64. Белок CRISPR класса 2, типа V по любому из вариантов осуществления 29-62, содержащий последовательность, выбранную из группы, состоящей из SEQ ID NO: 415-592 и 1147-1231, как указано в таблице 3.

[0816] Вариант осуществления 65. Белок CRISPR класса 2, типа V по любому из вариантов осуществления 29-64, содержащий один или более сигналов ядерной локализации (NLS).

[0817] Вариант осуществления 66. Белок CRISPR класса 2, типа V по варианту осуществления 65, в котором один или более NLS выбраны из группы последовательностей, состоящей из PKKKRKV (SEQ ID NO: 352), KRPAATKKAGQAKKKK (SEQ ID NO: 353), PAAKRVKLD (SEQ ID NO: 354), RQRRNELKRSP (SEQ ID NO: 355), NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY (SEQ ID NO: 356), RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV (SEQ ID NO: 357), VSRKRPRP (SEQ ID NO: 358), PPKKARED (SEQ ID NO: (359), PQPKKKPL (SEQ ID NO: 360), SALIKKKKKMAP (SEQ ID NO: 361), DRLRR (SEQ ID NO: 362), PKQKKRK (SEQ ID NO: 363), RKLKKKIKKL (SEQ ID NO: 364), REKKKFLKRR (SEQ ID NO: 365), KRKGDEVDGVDEVAKKKSKK (SEQ ID NO: 366), RKCLQAGMNLEARKTKK (SEQ ID NO: 367), PRPRKIPR (SEQ ID NO: 368), PPRKKRTVV (SEQ ID NO: 369), NLSKKKKRKREK (SEQ ID NO: 370), RRPSRPFRKP (SEQ ID NO: 371), KRPRSPSS (SEQ ID NO: 372), KRGINDRNFWRGENERKTR (SEQ ID NO: 373), PRPPKMARYDN (SEQ ID NO: 374), KRSFSKAF (SEQ ID NO: 375), KLKIKRPVK (SEQ ID NO: 376), PKKKRKVPPPPAAKRVKLD (SEQ ID NO: 377), PKTRRRPRRSQRKRPPT (SEQ ID NO: 378), SRRRKANPTKLSENAKKLAKEVEN (SEQ ID NO: 379), KTRRRPRRSQRKRPPT (SEQ ID NO: 380), RRKKRRPRRKKRR (SEQ ID NO: 381), PKKKSRKPKKKSRK (SEQ ID NO: 382), HKKKHPDASVNFSEFSK (SEQ ID NO: 383), QRPGPYDRPQRPGPYDRP (SEQ ID NO: 384), LSPSLSPLLSPSLSPL (SEQ ID NO: 385), RGKGGKGLGKGGAKRHRK (SEQ ID NO: 386), PKRGRGRPKRGRGR (SEQ ID NO: 387), PKKKRKVPPPPKKKRKV (SEQ ID NO: 389), PAKRARRGYKC (SEQ ID NO: 63), KLGPRKATGRW (SEQ ID NO: 64), PRRKREE (SEQ ID NO: 65), PYRGRKE (SEQ ID NO: 66), PLRKRPRR (SEQ ID NO: 67), PLRKRPRRGSPLRKRPRR (SEQ ID NO:68), PAAKRVKLDGGKRTADGSEFESPKKKRKV (SEQ ID NO:69), PAAKRVKLDGGKRTADGSEFESPKKKRKVGIHGVPAA (SEQ ID NO:70), PAAKRVKLDGGKRTADGSEFESPKKKRKVAEAAAKEAAAKEAAAKA (SEQ ID NO:71), PAAKRVKLDGGKRTADGSEFESPKKKRKVPG (SEQ ID NO:72), KRKGSPERGERKRHW (SEQ ID NO:73), KRTADSQHSTPPKTKRKVEFEPKKKRKV (SEQ ID NO:74) и PKKKRKVGGSKRTADSQHSTPPKTKRKVEFEPKKKRKV (SEQ ID NO:75), и необязательно при этом один или более NLS присоединены к белку CRISPR класса 2, типа

V или смежному NLS с помощью линкерного пептида, причем линкерный пептид выбран из группы, состоящей из SR, RS, (G)n (SEQ ID NO: 1023), (GS)n (SEQ ID NO: 1024), (GSGGS)n (SEQ ID NO: 399), (GGSGGS)n (SEQ ID NO: 400), (GGGS)n (SEQ ID NO: 401), GGSG (SEQ ID NO: 402), GGSGG (SEQ ID NO: 403), GSGSG (SEQ ID NO: 404), GSGGG (SEQ ID NO: 405), GGGSG (SEQ ID NO: 406), GSSSG (SEQ ID NO: 407), GPGP (SEQ ID NO: 408), GGP, PPP, PPAPPA (SEQ ID NO: 409), PPPG (SEQ ID NO: 24), PPPGPPP (SEQ ID NO: 410), PPP(GGGS)n (SEQ ID NO: 25), (GGGS)nPPP (SEQ ID NO: 26), AEAAAKEAAAKEAAAKA (SEQ ID NO: 1025) и TPPKTKRKVEFE (SEQ ID NO: 27), где n равно от 1 до 5.

[0818] Вариант осуществления 67. Белок CRISPR класса 2, типа V по варианту осуществления 65 или варианту осуществления 66, в котором один или более NLS расположены на C-конце белка или рядом с ним.

[0819] Вариант осуществления 68. Белок CRISPR класса 2, типа V по варианту осуществления 65 или варианту осуществления 66, в котором один или более NLS расположены на N-конце белка или рядом с ним.

[0820] Вариант осуществления 69. Белок CRISPR класса 2, типа V по варианту осуществления 65 или варианту осуществления 66, содержащий по меньшей мере два NLS, причем по меньшей мере два NLS расположены на N-конце белка или рядом с ним и на C-конце белка или рядом с ним.

[0821] Вариант осуществления 70. Белок CRISPR класса 2, типа V по любому из вариантов осуществления 29-69, способный образовывать рибонуклеопротеиновый комплекс (RNP) с гРНК.

[0822] Вариант осуществления 71. Белок CRISPR класса 2, типа V по варианту осуществления 70, причем RNP демонстрирует по меньшей мере одну или более улучшенных характеристик по сравнению с RNP из референсного белка с любой из SEQ ID NO: 1-3 и гРНК с SEQ ID NO: 4 или SEQ ID NO: 5.

[0823] Вариант осуществления 72. Белок CRISPR класса 2, типа V по варианту осуществления 71, причем улучшенная характеристика выбрана из группы, состоящей из повышенной аффинности связывания с гидовой нуклеиновой кислотой (гРНК); повышенной аффинности связывания с целевой нуклеиновой кислотой; улучшенной способности использовать более широкий спектр из одной или более PAM-последовательностей, включая ATC, CTC, GTC или TTC, при редактировании целевой нуклеиновой кислоты; повышенного раскручивания целевой нуклеиновой кислоты; повышенной активности редактирования; повышенной эффективности редактирования; повышенной специфичности редактирования целевой нуклеиновой кислоты; повышенной нуклеазной активности; повышенной загрузки целевой цепи для двухцепочечного расщепления; сниженной загрузки целевой цепи для одноцепочечного никирования; сниженного нецелевого расщепления целевой нуклеиновой кислоты; повышенного связывания нецелевой цепи нуклеиновой кислоты и повышенной стабильности комплекса белок : гРНК (RNP).

[0824] Вариант осуществления 73. Белок CRISPR класса 2, типа V по варианту осуществления 71 или варианту осуществления 72, причем улучшенная характеристика RNP имеет кратность повышения от по меньшей мере примерно 1,1 до примерно 100 000 раз относительно RNP из референсного белка с SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3 и гРНК с SEQ ID NO: 4 или SEQ ID NO: 5.

[0825] Вариант осуществления 74. Белок CRISPR класса 2, типа V по варианту осуществления 71 или варианту осуществления 72, причем улучшенная характеристика RNP имеет кратность повышения по меньшей мере примерно 10 раз, по меньшей мере примерно 100 раз, по меньшей мере примерно 1000 раз или по меньшей мере примерно 10 000 раз относительно RNP из референсного белка с SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3 и гРНК с SEQ ID NO: 4 или 5.

[0826] Вариант осуществления 75. Белок CRISPR класса 2, типа V по любому из вариантов осуществления 71-74, причем улучшенная характеристика RNP имеет кратность улучшения эффективности редактирования от 1,1 до 100 раз по сравнению с RNP из референсного белка с SEQ ID NO: 2 и гРНК, содержащей SEQ ID NO: 4 или 5.

[0827] Вариант осуществления 76. Пара для редактирования генов, содержащая гРНК и белок CRISPR класса 2, типа V, причем пара содержит:

a. гРНК по любому из вариантов осуществления 25-28 и

b. белок CRISPR класса 2, типа V по любому из вариантов осуществления 29-75.

[0828] Вариант осуществления 77. Пара для редактирования генов по варианту осуществления 76, в которой гРНК и белок CRISPR класса 2, типа V способны образовывать рибонуклеопротеиновый комплекс (RNP).

[0829] Вариант осуществления 78. Пара для редактирования генов по варианту осуществления 76 или варианту осуществления 77, в которой гРНК и белок CRISPR класса 2, типа V соединены вместе в рибонуклеопротеиновый комплекс (RNP).

[0830] Вариант осуществления 79. Пара для редактирования генов по варианту осуществления 77 или варианту осуществления 78, причем RNP из белка CRISPR класса 2, типа V и гРНК демонстрирует по меньшей мере одну или более улучшенных характеристик по сравнению с RNP из референсного белка с SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3 и гРНК, содержащей последовательность SEQ ID NO: 4 или SEQ ID NO: 5.

[0831] Вариант осуществления 80. Пара для редактирования генов по варианту осуществления 79, причем улучшенная характеристика выбрана из одного или более из группы, состоящей из повышенной аффинности связывания белка CRISPR класса 2, типа V с гРНК; повышенной аффинности связывания с целевой нуклеиновой кислотой; улучшенной способности использовать более широкий спектр из одной или более PAM-последовательностей, включая ATC, CTC, GTC или TTC, при редактировании целевой нуклеиновой кислоты; повышенного раскручивания целевой нуклеиновой кислоты; повышенной активности редактирования; повышенной эффективности редактирования; повышенной специфичности редактирования целевой нуклеиновой кислоты; повышенной нуклеазной активности; повышенной загрузки целевой цепи для двухцепочечного

расщепления; сниженной загрузки целевой цепи для одноцепочечного никирования; сниженного нецелевого расщепления целевой нуклеиновой кислоты; повышенного связывания нецелевой цепи нуклеиновой кислоты повышенной стабильности комплекса белок : гРНК (RNP) и улучшенных характеристик слияния.

[0832] Вариант осуществления 81. Пара для редактирования генов по варианту осуществления 79 или варианту осуществления 80, причем улучшенная характеристика RNP из белка CRISPR класса 2, типа V и гРНК имеет кратность повышения от по меньшей мере примерно 1,1 до примерно 100 раз или более относительно RNP из референсного белка с SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3 и гРНК, содержащей последовательность SEQ ID NO: 4 или SEQ ID NO: 5, в сопоставимой системе анализа *in vitro*.

[0833] Вариант осуществления 82. Пара для редактирования генов по варианту осуществления 79 или варианту осуществления 80, причем улучшенная характеристика белка CRISPR класса 2, типа V имеет кратность повышения по меньшей мере примерно 1,1 раза, по меньшей мере примерно 2 раза, по меньшей мере примерно 10 раз, по меньшей мере примерно 100 раз или более относительно референсного белка с SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3 и гРНК, содержащей последовательность SEQ ID NO: 4 или SEQ ID NO: 5, в сопоставимой системе анализа *in vitro*.

[0834] Вариант осуществления 83. Пара для редактирования генов по любому из вариантов осуществления 77-82, причем RNP, содержащий белок CRISPR класса 2, типа V и гРНК, демонстрирует более высокие эффективность редактирования и/или связывание последовательности целевой нуклеиновой кислоты в целевой нуклеиновой кислоте, когда любая из PAM-последовательностей TTC, ATC, GTC или CTC расположена на расстоянии 1 нуклеотида от 5'-конца нецелевой цепи протоспейсера, имеющей идентичность с нацеливающей последовательностью гРНК, в системе клеточного анализа по сравнению с эффективностью редактирования и/или связыванием у RNP, содержащего референсный белок с SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3 и референсную гРНК, в сопоставимой системе анализа.

[0835] Вариант осуществления 84. Пара для редактирования генов по варианту осуществления 83, причем PAM-последовательность представляет собой TTC.

[0836] Вариант осуществления 85. Пара для редактирования генов по варианту осуществления 83, причем PAM-последовательность представляет собой ATC.

[0837] Вариант осуществления 86. Пара для редактирования генов по варианту осуществления 83, причем PAM-последовательность представляет собой CTC.

[0838] Вариант осуществления 87. Пара для редактирования генов по варианту осуществления 83, причем PAM-последовательность представляет собой GTC.

[0839] Вариант осуществления 88. Пара для редактирования генов по любому из вариантов осуществления 83-87, в которой RNP, содержащий белок CRISPR класса 2, типа V и гРНК, демонстрирует повышенную аффинность связывания с одной или более PAM-последовательностями, которая по меньшей мере в 1,5 раза, по меньшей мере в 2 раза, по меньшей мере в 4 раза, по меньшей мере в 10 раз, по меньшей мере в 20 раз, по меньшей

мере в 30 раз или по меньшей мере в 40 раз выше, чем аффинность связывания RNP из любого из референсных белков с SEQ ID NO:1-3 и гРНК с SEQ ID NO: 4 или SEQ ID NO: 5, при оценке в сопоставимой системе анализа *in vitro*.

[0840] Вариант осуществления 89. Пара для редактирования генов по любому из вариантов осуществления 77-88, в которой RNP, содержащий белок CRISPR класса 2, типа V и гРНК, демонстрирует повышенную эффективность редактирования, которая по меньшей мере в 1,5 раза, по меньшей мере в 2 раза, по меньшей мере в 4 раза, по меньшей мере в 10 раз, по меньшей мере в 20 раз, по меньшей мере в 30 раз или по меньшей мере в 40 раз выше, чем эффективность редактирования у RNP из любого из референсных белков с SEQ ID NO:1-3 и гРНК с SEQ ID NO: 4 или SEQ ID NO: 5, при оценке в сопоставимой системе анализа *in vitro*.

[0841] Вариант осуществления 90. Пара для редактирования генов по любому из вариантов осуществления 77-89, в которой белок CRISPR класса 2, типа V и гРНК способны образовывать RNP, имеющий на по меньшей мере примерно 5%, по меньшей мере примерно 10%, по меньшей мере примерно 15% или по меньшей мере примерно 20% более высокую процентную долю компетентной в отношении расщепления конформации по сравнению с RNP из любого из референсных белков с SEQ ID NO: 1-3 и гРНК с SEQ ID NO: 4 или SEQ ID NO: 5, при оценке в сопоставимой системе анализа *in vitro*.

[0842] Вариант осуществления 91. Пара для редактирования генов по любому из вариантов осуществления 77-90, в которой RNP, содержащий белок CRISPR класса 2, типа V и гРНК, демонстрирует скорость расщепления целевой нуклеиновой кислоты в анализе зависимости от времени *in vitro*, которая по меньшей мере приблизительно в 5 раз, по меньшей мере приблизительно в 10 раз или по меньшей мере приблизительно в 20 раз выше по сравнению с RNP из любого из референсных белков с SEQ ID NO: 1-3 и гРНК с SEQ ID NO: 4 или SEQ ID NO: 5, при оценке в сопоставимой системе анализа *in vitro*.

[0843] Вариант осуществления 92. Пара для редактирования генов по любому из вариантов осуществления 77-91, в которой RNP, содержащий белок CRISPR класса 2, типа V и гРНК, демонстрирует более высокую процентную долю редактирования целевой нуклеиновой кислоты в анализе зависимости от времени *in vitro*, которая по меньшей мере приблизительно в 5 раз, по меньшей мере приблизительно в 10 раз, по меньшей мере приблизительно в 20 раз или по меньшей мере приблизительно в 100 раз выше по сравнению с RNP из любого из референсных белков с SEQ ID NO: 1-3 и гРНК с SEQ ID NO: 4 или SEQ ID NO: 5 при оценке в сопоставимой системе анализа *in vitro*.

[0844] Вариант осуществления 93. Каталитически неактивный белок CRISPR класса 2, типа V, содержащий последовательность, выбранную из группы, состоящей из SEQ NO: 44-62 и 1232-1235, как показано в таблице 7, или последовательность, имеющую по меньшей мере 60%, по меньшей мере 70%, по меньшей мере 80%, по меньшей мере 90%, по меньшей мере 95%, по меньшей мере 98% или по меньшей мере 99%.

[0845] Вариант осуществления 94. Каталитически неактивный белок CRISPR класса 2, типа V, содержащий последовательность, выбранную из группы, состоящей из SEQ NO:

44-62 и 1232-1235, как указано в таблице 7.

[0846] Вариант осуществления 95. Белок CRISPR класса 2, типа V по варианту осуществления 93 или варианту осуществления 94, причем и RNP из каталитически неактивного белка CRISPR класса 2, типа V и гРНК по любому из вариантов осуществления 25-28 сохраняет способность связывать целевую нуклеиновую кислоту.

[0847] Вариант осуществления 96. Нуклеиновая кислота, содержащая последовательность, которая кодирует каркас гРНК по любому из вариантов осуществления 1-24 или гРНК по любому из вариантов осуществления 25-28.

[0848] Вариант осуществления 97. Нуклеиновая кислота, содержащая последовательность, которая кодирует белок CRISPR класса 2, типа V по любому из вариантов осуществления 29-75.

[0849] Вариант осуществления 98. Нуклеиновая кислота по варианту осуществления 97, в которой последовательность, которая кодирует белок CRISPR класса 2, типа V, является кодон-оптимизированной для экспрессии в эукариотической клетке.

[0850] Вариант осуществления 99. Вектор, содержащий гРНК по любому из вариантов осуществления 25-28, белок CRISPR класса 2, типа V по любому из вариантов осуществления 29-75 или нуклеиновую кислоту по любому из вариантов осуществления 96-98.

[0851] Вариант осуществления 100. Вектор по варианту осуществления 99, содержащий промотор.

[0852] Вариант осуществления 101. Вектор по варианту осуществления 99 или варианту осуществления 100, выбранный из группы, состоящей из ретровирусного вектора, лентивирусного вектора, аденовирусного вектора, аденоассоциированного вирусного (AAV) вектора, вектора на основе вируса простого герпеса (HSV), частицы для доставки CasX (XDP), плазмиды, миникольца, наноплазмиды, ДНК-вектора и РНК-вектора.

[0853] Вариант осуществления 102. Вектор по варианту осуществления 101, который представляет собой AAV-вектор.

[0854] Вариант осуществления 103. Вектор по варианту осуществления 102, причем AAV-вектор выбран из AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAV-Rh74 или AAVRh10.

[0855] Вариант осуществления 104. Вектор по варианту осуществления 101, который представляет собой ретровирусный вектор.

[0856] Вариант осуществления 105. Вектор по варианту осуществления 101, который представляет собой XDP, содержащую один или более компонентов полибелка gag.

[0857] Вариант осуществления 106. Вектор по варианту осуществления 105, в котором один или более компонентов полибелка gag выбраны из группы, состоящей из матриксного белка (MA), нуклеокапсидного белка (NC), капсидного белка (CA), пептида p1, пептида p6, пептида P2A, пептида P2B, пептида P10, пептида p12, пептида PP21/24, пептида P12/P3/P8, пептида P20 и сайта расщепления протеазой.

[0858] Вариант осуществления 107. Вектор по варианту осуществления 105 или

варианту осуществления 106, причем белок CRISPR класса 2, типа V и гРНК связаны друг с другом в RNP.

[0859] Вариант осуществления 108. Вектор по любому из вариантов осуществления 105-107, содержащий гликопротеиновый фактор тропизма.

[0860] Вариант осуществления 109. Вектор по варианту осуществления 108, причем гликопротеиновый фактор тропизма обладает аффинностью связывания с маркером клеточной поверхности клетки-мишени и облегчает проникновение XDP в клетку-мишень.

[0861] Вариант осуществления 110. Вектор по любому из вариантов осуществления 99-109, содержащий донорную матрицу.

[0862] Вариант осуществления 111. Клетка-хозяин, содержащая вектор по любому из вариантов осуществления 99-110.

[0863] Вариант осуществления 112. Клетка-хозяин по варианту осуществления 111, выбранная из группы, состоящей из клеток-фибробластов почки новорожденного хомяка (BHK), клеток эмбриональной почки человека 293 (HEK293), клеток эмбриональной почки человека 293T (HEK293T), клеток NS0, клеток SP2/0, клеток миеломы YO, мышиных клеток миеломы P3X63, клеток PER, клеток PER.C6, клеток гибридомы, клеток NIH3T3, клеток, происходящих от CV-1 (обезьяньих), с генетическим материалом SV40 (COS), клеток HeLa, клеток яичника китайского хомячка (CHO) и дрожжевых клеток.

[0864] Вариант осуществления 113. Способ модификации целевой нуклеиновой кислоты в клетке, включающий приведение целевой нуклеиновой кислоты клетки в контакт с: i) парой для редактирования генов по любому из вариантов осуществления 76-92; ii) парой для редактирования генов по любому из вариантов осуществления 76-92 вместе с донорной матрицей; iii) одной или более нуклеиновыми кислотами, кодирующими пару для редактирования генов по п. (i) или (ii); iv) вектором, содержащим нуклеиновую кислоту по п. (iii); v) XDP, содержащей пару для редактирования генов по п. (i) или (ii); или vi) комбинацией двух или более из пп. (i)-(v), причем приведение целевой нуклеиновой кислоты в контакт модифицирует целевую нуклеиновую кислоту.

[0865] Вариант осуществления 114. Способ по варианту осуществления 113, включающий приведение целевой молекулы в контакт со множеством пар для редактирования генов, содержащих первую и вторую или множество гРНК, содержащих нацеливающие последовательности, комплементарные разным или перекрывающимся областям целевой нуклеиновой кислоты.

[0866] Вариант осуществления 115. Способ по варианту осуществления 113, включающий приведение целевой молекулы в контакт со множеством нуклеиновых кислот, кодирующих пары для редактирования генов, содержащие первую и вторую или множество гРНК, содержащих нацеливающие последовательности, комплементарные разным или перекрывающимся областям целевой нуклеиновой кислоты.

[0867] Вариант осуществления 116. Способ по варианту осуществления 113, включающий приведение целевой молекулы в контакт со множеством XDP, содержащих пары для редактирования генов, которые содержат первую и вторую или множество гРНК,

содержащих нацеливающие последовательности, комплементарные разным или перекрывающимся областям целевой нуклеиновой кислоты.

[0868] Вариант осуществления 117. Способ по любому из вариантов осуществления 113, в котором приведение в контакт включает в себя связывание целевой нуклеиновой кислоты с парой для редактирования генов и введение одного или более одноцепочечных разрывов в целевую нуклеиновую кислоту, причем модификация включает в себя введение мутации, вставки или делеции в целевую нуклеиновую кислоту.

[0869] Вариант осуществления 118. Способ по любому из вариантов осуществления 113-116, в котором приведение в контакт включает в себя связывание целевой нуклеиновой кислоты и введение одного или более двухцепочечных разрывов в целевую нуклеиновую кислоту, причем модификация включает в себя введение мутации, вставки или делеции в целевую нуклеиновую кислоту.

[0870] Вариант осуществления 119. Способ по любому из вариантов осуществления 113-118, включающий приведение целевой нуклеиновой кислоты в контакт с нуклеотидной последовательностью нуклеиновой кислоты донорной матрицы, причем донорная матрица содержит нуклеотидную последовательность, имеющую гомологию с целевой нуклеиновой кислотой.

[0871] Вариант осуществления 120. Способ по варианту осуществления 119, в котором донорная матрица содержит гомологичные плечи на 5’- и 3’-концах донорной матрицы.

[0872] Вариант осуществления 121. Способ по варианту осуществления 119 или варианту осуществления 120, в котором донорную матрицу вставляют в целевую нуклеиновую кислоту в сайте разрыва посредством направляемой гомологией репарации.

[0873] Вариант осуществления 122. Способ по варианту осуществления 121, в котором донорную матрицу вставляют в целевую нуклеиновую кислоту в сайте разрыва посредством негомологичного соединения концов (NHEJ) или соединения концов, обусловленного микрогомологией (MMEJ).

[0874] Вариант осуществления 123. Способ по любому из вариантов осуществления 113-122, в котором модификация клетки происходит *in vitro*.

[0875] Вариант осуществления 124. Способ по любому из вариантов осуществления 113-122, в котором модификация клетки происходит in vivo.

[0876] Вариант осуществления 125. Способ по любому из вариантов осуществления 113-124, в котором клетка представляет собой эукариотическую клетку.

[0877] Вариант осуществления 126. Способ по варианту осуществления 125, в котором эукариотическая клетка выбрана из группы, состоящей из клетки грызуна, клетки мыши, клетки крысы, клетки примата и клетки примата, не являющегося человеком.

[0878] Вариант осуществления 127. Способ по варианту осуществления 125, в котором эукариотическая клетка представляет собой клетку человека.

[0879] Вариант осуществления 128. Способ по любому из вариантов осуществления 113-127, в котором клетка выбрана из группы, состоящей из эмбриональной стволовой

клетки, индуцированной плюрипотентной стволовой клетки, зародышевой клетки, фибробласта, олигодендроцита, глиальной клетки, гематопоэтической стволовой клетки, нейрональной клетки-предшественника, нейрона, мышечной клетки, костной клетки, гепатоцита, клетки поджелудочной железы, клетки сетчатки, раковой клетки, Т-клетки, В-клетки, NK-клетки, фетального кардиомиоцита, миофибробласта, мезенхимальной стволовой клетки, аутотрансплантированного гипертрофированного кардиомиоцита, адипоцита, тотипотентной клетки, плюрипотентной клетки, стволовой клетки крови, миобласта, стволовой клетки взрослого, клетки костного мозга, мезенхимальной клетки, паренхимной клетки, эпителиальной клетки, эндотелиальной клетки, мезотелиальной клетки, клетки-фибробласта, клетки-остеобласта, клетки-хондроцита, экзогенной клетки, эндогенной клетки, стволовой клетки, гематопоэтической стволовой клетки, клетки-предшественницы, полученной из костного мозга, миокардиальной клетки, скелетной клетки, фетальной клетки, недифференцированной клетки, мультипотентной клетки-предшественницы, унипотентной клетки-предшественницы, моноцита, миобласта сердечной мышцы, миобласта скелетной мышцы, макрофага, эндотелиальной клетки капилляров, ксеногенной клетки, аллогенной клетки, аутологичной клетки и постнатальной стволовой клетки.

[0880] Вариант осуществления 129. Способ по любому из вариантов осуществления 124-128, причем клетка находится в субъекте.

[0881] Вариант осуществления 130. Способ по варианту осуществления 129, в котором модификация происходит в клетках субъекта, имеющего мутацию в аллели гена, причем мутация вызывает заболевание или расстройство у субъекта.

[0882] Вариант осуществления 131. Способ по варианту осуществления 130, в котором модификация изменяет мутацию с образованием аллели гена дикого типа или приводит к экспрессии продукта функционального гена.

[0883] Вариант осуществления 132. Способ по варианту осуществления 130, в котором модификация приводит к нокдауну или нокауту аллели гена, вызывающей заболевание или расстройство у субъекта.

[0884] Вариант осуществления 133. Способ по любому из вариантов осуществления 129-132, в котором клетка является аутологичной по отношению к субъекту.

[0885] Вариант осуществления 134. Способ по любому из вариантов осуществления 129-132, в котором клетка является аллогенной аутологичной по отношению к субъекту.

[0886] Вариант осуществления 135. Способ по любому из вариантов осуществления 113-134, в котором вектор представляет собой аденоассоциированный вирусный (AAV) вектор.

[0887] Вариант осуществления 136. Способ по варианту осуществления 135, в котором AAV-вектор представляет собой AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV-Rh74 или AAVRh10.

[0888] Вариант осуществления 137. Способ по варианту осуществления 113, в котором вектор представляет собой лентивирусный вектор.

[0889] Вариант осуществления 138. Способ по любому из вариантов осуществления 113-137, в котором вектор вводят нуждающемуся в этом субъекту с применением терапевтически эффективной дозы.

[0890] Вариант осуществления 139. Способ по варианту осуществления 138, причем субъект выбран из группы, состоящей из мыши, крысы, свиньи и примата, не являющегося человеком.

[0891] Вариант осуществления 140. Способ по варианту осуществления 138, причем субъект является человеком.

[0892] Вариант осуществления 141. Способ по любому из вариантов осуществления 138-140, в котором вектор вводят субъекту в дозе по меньшей мере примерно 1×10^5 векторных геномов/кг (вг/кг), по меньшей мере примерно 1×10^6 вг/кг, по меньшей мере примерно 1×10^7 вг/кг, по меньшей мере примерно 1×10^8 вг/кг, по меньшей мере примерно 1×10^9 вг/кг, по меньшей мере примерно 1×10^{10} вг/кг, по меньшей мере примерно 1×10^{11} вг/кг, по меньшей мере примерно 1×10^{12} вг/кг, по меньшей мере примерно 1×10^{13} вг/кг, по меньшей мере примерно 1×10^{14} вг/кг, по меньшей мере примерно 1×10^{15} вг/кг или по меньшей мере примерно 1×10^{16} вг/кг.

[0893] Вариант осуществления 142. Способ по любому из вариантов осуществления 138-140, в котором вектор вводят субъекту в дозе от по меньшей мере примерно 1×10^5 вг/кг до примерно 1×10^{16} вг/кг, от по меньшей мере примерно 1×10^6 вг/кг до примерно 1×10^{15} вг/кг или от по меньшей мере примерно 1×10^7 вг/кг до примерно 1×10^{14} вг/кг.

[0894] Вариант осуществления 143. Способ по варианту осуществления 113, в котором вектор представляет собой XDP.

[0895] Вариант осуществления 144. Способ по варианту осуществления 143, в котором XDP вводят нуждающемуся в этом субъекту с применением терапевтически эффективной дозы.

[0896] Вариант осуществления 145. Способ по варианту осуществления 144, в котором XDP вводят субъекту в дозе по меньшей мере примерно 1×10^5 частиц/кг, по меньшей мере примерно 1×10^6 частиц/кг, по меньшей мере примерно 1×10^7 частиц/кг, по меньшей мере примерно 1×10^8 частиц/кг, по меньшей мере примерно 1×10^9 частиц/кг, по меньшей мере примерно 1×10^{10} частиц/кг, по меньшей мере примерно 1×10^{11} частиц/кг, по меньшей мере примерно 1×10^{12} частиц/кг, по меньшей мере примерно 1×10^{13} частиц/кг, по меньшей мере примерно 1×10^{14} частиц/кг, по меньшей мере примерно 1×10^{15} частиц/кг, по меньшей мере примерно 1×10^{16} частиц/кг.

[0897] Вариант осуществления 146. Способ по варианту осуществления 143, в котором XDP вводят субъекту в дозе от по меньшей мере примерно 1×10^5 частиц/кг до примерно 1×10^{16} частиц/кг, или от по меньшей мере примерно 1×10^6 частиц/кг до примерно 1×10^{15} частиц/кг, или от по меньшей мере примерно 1×10^7 частиц/кг до примерно 1×10^{14} частиц/кг.

[0898] Вариант осуществления 147. Способ по любому из вариантов осуществления 138-146, в котором вектор вводят путем введения, выбранным из группы, состоящей из

интрапаренохиматозного, внутривенного, внутриартериального, интрацеребровентрикулярного, интрацистернального, интратекального, интракраниального и внутрибрюшинного путей, причем способ введения представляет собой инъекцию, трансфузию или имплантацию.

[0899] Вариант осуществления 148. Способ по любому из вариантов осуществления 141-147, в котором вектор вводят субъекту в соответствии со схемой лечения, включающей в себя одну или более последовательных доз с применением терапевтически эффективной дозы вектора.

[0900] Вариант осуществления 149. Способ по варианту осуществления 148, причем терапевтически эффективную дозу вводят субъекту в виде двух или более доз в течение по меньшей мере двух недель, или по меньшей мере одного месяца, или по меньшей мере двух месяцев, или по меньшей мере трех месяцев, или по меньшей мере четырех месяцев, или по меньшей мере пяти месяцев, или по меньшей мере шести месяцев, или раз в год, или каждые 2 или 3 года.

[0901] Вариант осуществления 150. Клетка, содержащая целевую нуклеиновую кислоту, модифицированную парой для редактирования генов по любому из вариантов осуществления 76-92.

[0902] Вариант осуществления 151. Клетка, отредактированная способом по любому из вариантов осуществления 113-149.

[0903] Вариант осуществления 152. Клетка по варианту осуществления 150 или 151, причем клетка представляет собой прокариотическую клетку.

[0904] Вариант осуществления 153. Клетка по варианту осуществления 150 или 151, причем клетка представляет собой эукариотическую клетку.

[0905] Вариант осуществления 154. Клетка по варианту осуществления 153, причем эукариотическая клетка выбрана из группы, состоящей из клетки грызуна, клетки мыши, клетки крысы, клетки примата и клетки примата, не являющегося человеком.

[0906] Вариант осуществления 155. Клетка по варианту осуществления 153, причем эукариотическая клетка представляет собой клетку человека.

[0907] Вариант осуществления 156. Композиция, содержащая белок CRISPR класса 2, типа V по любому из вариантов осуществления 29-75.

[0908] Вариант осуществления 157. Композиция по варианту осуществления 156, содержащая гРНК по любому из вариантов осуществления 25-28.

[0909] Вариант осуществления 158. Композиция по варианту осуществления 157, в которой белок и гРНК связаны друг с другом в рибонуклеопротеиновом комплексе (RNP).

[0910] Вариант осуществления 159. Композиция по любому из вариантов осуществления 156-158, содержащая нуклеиновую кислоту донорной матрицы, причем донорная матрица содержит нуклеотидную последовательность, имеющую гомологию с целевой нуклеиновой кислотой.

[0911] Вариант осуществления 160. Композиция по любому из вариантов осуществления 156-159, содержащая буфер, ингибитор нуклеазы, ингибитор протеазы,

липосому, терапевтический агент, метку, реагент для визуализации метки или любую комбинацию вышеизложенного.

[0912] Вариант осуществления 161. Композиция, содержащая каркас гРНК по любому из вариантов осуществления 1-24 или гРНК по любому из вариантов осуществления 25-28.

[0913] Вариант осуществления 162. Композиция по варианту осуществления 161, содержащая белок CRISPR класса 2, типа V по любому из вариантов осуществления 29-75.

[0914] Вариант осуществления 163. Композиция по варианту осуществления 162, в которой белок CRISPR класса 2, типа V и гРНК связаны друг с другом в рибонуклеопротеиновом комплексе (RNP).

[0915] Вариант осуществления 164. Композиция по любому из вариантов осуществления 161-163, содержащая нуклеиновую кислоту донорной матрицы, причем донорная матрица содержит нуклеотидную последовательность, имеющую гомологию с целевой нуклеиновой кислотой.

[0916] Вариант осуществления 165. Композиция по любому из вариантов осуществления 161-164, содержащая буфер, ингибитор нуклеазы, ингибитор протеазы, липосому, терапевтический агент, метку, реагент для визуализации метки или любую комбинацию вышеизложенного.

[0917] Вариант осуществления 166. Композиция, содержащая пару для редактирования генов по любому из вариантов осуществления 76-92.

[0918] Вариант осуществления 167. Композиция по варианту осуществления 166, содержащая нуклеиновую кислоту донорной матрицы, причем донорная матрица содержит нуклеотидную последовательность, имеющую гомологию с целевой нуклеиновой кислотой.

[0919] Вариант осуществления 168. Композиция по варианту осуществления 166 или варианту осуществления 167, содержащая буфер, ингибитор нуклеазы, ингибитор протеазы, липосому, терапевтический агент, метку, реагент для визуализации метки или любую комбинацию вышеизложенного.

[0920] Вариант осуществления 169. Набор, содержащий белок CRISPR класса 2, типа V по любому из вариантов осуществления 29-75 и контейнер.

[0921] Вариант осуществления 170. Набор по варианту осуществления 169, содержащий каркас гРНК по любому из вариантов осуществления 1-24 или гРНК по любому из вариантов осуществления 25-28.

[0922] Вариант осуществления 171. Набор по варианту осуществления 169 или варианту осуществления 170, содержащий нуклеиновую кислоту донорной матрицы, причем донорная матрица содержит нуклеотидную последовательность, имеющую гомологию с целевой нуклеотидной последовательностью целевой нуклеиновой кислоты.

[0923] Вариант осуществления 172. Набор по любому из вариантов осуществления 169-171, содержащий буфер, ингибитор нуклеазы, ингибитор протеазы, липосому, терапевтический агент, метку, реагент для визуализации метки или любую комбинацию

вышеизложенного.

[0924] Вариант осуществления 173. Набор, содержащий каркас гРНК по любому из вариантов осуществления 1-24 или гРНК по любому из вариантов осуществления 25-28.

[0925] Вариант осуществления 174. Набор по варианту осуществления 173, содержащий белок CRISPR класса 2, типа V по любому из вариантов осуществления 29-75.

[0926] Вариант осуществления 175. Набор по варианту осуществления 173 или варианту осуществления 174, содержащий нуклеиновую кислоту донорной матрицы, причем донорная матрица содержит нуклеотидную последовательность, имеющую гомологию с целевой нуклеотидной последовательностью целевой нуклеиновой кислоты.

[0927] Вариант осуществления 176. Набор по любому из вариантов осуществления 173-175, содержащий буфер, ингибитор нуклеазы, ингибитор протеазы, липосому, терапевтический агент, метку, реагент для визуализации метки или любую комбинацию вышеизложенного.

[0928] Вариант осуществления 177. Набор, содержащий пару для редактирования генов по любому из вариантов осуществления 76-92.

[0929] Вариант осуществления 178. Набор по варианту осуществления 177, содержащий нуклеиновую кислоту донорной матрицы, причем донорная матрица содержит нуклеотидную последовательность, имеющую гомологию с целевой нуклеиновой кислотой.

[0930] Вариант осуществления 179. Набор по варианту осуществления 177 или варианту осуществления 178, содержащий буфер, ингибитор нуклеазы, ингибитор протеазы, липосому, терапевтический агент, метку, реагент для визуализации метки или любую комбинацию вышеизложенного.

[0931] Вариант осуществления 180. Сконструированный белок CRISPR класса 2, типа V, содержащий любую из последовательностей, перечисленных в таблице 3.

[0932] Вариант осуществления 181. гРНК, содержащая любую из последовательностей варианта каркаса гРНК, перечисленных в таблице 2.

[0933] Вариант осуществления 182. гРНК по варианту осуществления 181, в которой один или более урацилов (U) вариантов каркаса гРНК с последовательностями из таблицы 2 заменены на тимины ы(T).

[0934] Вариант осуществления 183. гРНК по варианту осуществления 182, содержащая нацеливающую последовательность по меньшей мере от 10 до 30 нуклеотидов, комплементарную целевой нуклеиновой кислоте.

[0935] Вариант осуществления 184. гРНК по варианту осуществления 183, в которой нацеливающая последовательность имеет 20 нуклеотидов.

[0936] Вариант осуществления 185. гРНК по варианту осуществления 183, в которой нацеливающая последовательность имеет 19 нуклеотидов.

[0937] Вариант осуществления 186. гРНК по варианту осуществления 183, в которой нацеливающая последовательность имеет 18 нуклеотидов.

[0938] Вариант осуществления 187. гРНК по варианту осуществления 183, в которой

нацеливающая последовательность имеет 17 нуклеотидов.

[0939] Вариант осуществления 188. гРНК по варианту осуществления 183, в которой нацеливающая последовательность имеет 16 нуклеотидов.

[0940] Вариант осуществления 189. гРНК по варианту осуществления 183, в которой нацеливающая последовательность имеет 15 нуклеотидов.

[0941] Вариант осуществления 190. Способ лечения заболевания у нуждающегося в этом субъекта, включающий введение субъекту фармацевтически эффективного количества композиции, содержащей: (a) сконструированный белок CRISPR класса 2, типа V по любому из вариантов осуществления 29-75 и 180 и (b) гРНК по любому из вариантов осуществления 25-28 и 181-189.

[0942] Вариант осуществления 191. Композиция, содержащая: (a) сконструированный белок CRISPR класса 2, типа V по любому из вариантов осуществления 29-75 и 180 и (b) гРНК по любому из вариантов осуществления 25-28 и 181-189 для применения в качестве лекарственного средства для лечения субъекта, имеющего заболевание.

[0943] Следующие примеры носят лишь иллюстративный характер и не предназначены для ограничения каких-либо аспектов настоящего описания любым образом.

ПРИМЕРЫ

Пример 1. Создание вариантов конструктов CasX

[0944] Чтобы создать конструкт CasX 488 (последовательности в таблице 9), кодон-оптимизированный конструкт CasX 119 (на основе конструкта CasX Stx2, кодирующего CasX *Planctomycetes* с SEQ ID NO: 2, с аминокислотными заменами и делециями) клонировали в плазмиду назначения (pStX) с применением стандартных способов клонирования. Чтобы создать конструкт CasX 491 (последовательности в таблице 9), кодон-оптимизированный конструкт CasX 484 (на основе конструкта CasX Stx2, кодирующего CasX *Planctomycetes* с SEQ ID NO: 2, с заменами и делециями определенных аминокислот, со слитным NLS и связанными гидовыми и ненацеливающими последовательностями) клонировали в плазмиду назначения (pStX) с применением стандартных способов клонирования. Конструкт CasX 1 (CasX с SEQ ID NO: 1) клонировали в вектор назначения с применением стандартных способов клонирования. Для построения CasX 488 ДНК конструкта CasX 119 амплифицировали методом ПЦР в двух реакциях с применением ДНК-полимеразы Q5 в соответствии с протоколом производителя и подходящих универсальных праймеров. Для построения кодон-оптимизированного CasX 491 ДНК конструкта CasX 484 амплифицировали методом ПЦР в двух реакциях с применением ДНК-полимеразы Q5 в соответствии с протоколом производителя и подходящих праймеров. Конструкт CasX 1 амплифицировали методом ПЦР в двух реакциях с применением ДНК-полимеразы Q5 в соответствии с протоколом производителя и подходящих универсальных праймеров. Каждый из продуктов ПЦР очищали посредством гель-экстракции из 1% агарозного геля (Gold Bio, кат. № A-201-500) с помощью набора

Zymoclean Gel DNA Recovery Kit в соответствии с протоколом производителя. Затем соответствующие фрагменты соединяли вместе с применением сборки Гибсона (New England BioLabs, кат. № E2621S) в соответствии с протоколом производителя. Сборными продуктами в pStx1 трансформировали химически компетентные бактериальные клетки Turbo Competent *E. coli*, высевали клетки на чашки с LB-агаром, содержащие канамицин. Отдельные колонии собирали и проводили мини-выделение с помощью набора Qiagen spin Miniprep Kit в соответствии с протоколом производителя. Полученные плазмиды секвенировали по методу Сэнгера для обеспечения правильности сборки. Затем правильные клоны субклонировали в экспрессионный вектор млекопитающих pStx34 с применением клонирования с помощью ферментов рестрикции. Основная цепь pStx34 и клоны CasX 488 и 491 в pStx1 расщепляли с помощью XbaI и BamHI соответственно. Расщепленная основная цепь и соответствующие фрагменты вставки очищали посредством гель-экстракции из 1% агарозного геля (Gold Bio, кат. № A-201-500) с помощью набора Zymoclean Gel DNA Recovery Kit в соответствии с протоколом производителя. После этого очищенные основную цепь и вставку лигировали друг с другом с помощью T4 лигазы (New England Biolabs, кат. № M0202L) в соответствии с протоколом производителя. Лигированными продуктами трансформировали химически компетентные бактериальные клетки Turbo Competent *E. coli*, и клетки высевали на чашки с LB-агаром, содержащие карбенициллин. Отдельные колонии собирали и проводили мини-выделение с помощью набора Qiagen spin Miniprep Kit в соответствии с протоколом производителя. Полученные плазмиды секвенировали по методу Сэнгера для обеспечения правильности сборки.

[0945] Для построения CasX 515 (последовательности в таблице 9) ДНК конструкта CasX 491 амплифицировали методом ПЦР в двух реакциях с применением ДНК-полимеразы Q5 в соответствии с протоколом производителя, используя подходящие праймеры. Для построения CasX 527 (последовательности в таблице 9) ДНК конструкта CasX 491 амплифицировали методом ПЦР в двух реакциях с применением ДНК-полимеразы Q5 в соответствии с протоколом производителя и подходящих праймеров. Продукты ПЦР очищали посредством гель-экстракции из 1% агарозного геля с помощью набора Zymoclean Gel DNA Recovery Kit в соответствии с протоколом производителя. Основную цепь pStX расщепляли с помощью XbaI и SpeI для удаления фрагмента ДНК размером 2931 пара нуклеотидов между двумя сайтами в плазмиде pStx56. Фрагмент расщепленной основной цепи очищали посредством гель-экстракции из 1% агарозного геля с помощью набора Zymoclean Gel DNA Recovery Kit в соответствии с протоколом производителя. Затем фрагменты вставки и основной цепи соединяли вместе с использованием сборки Гибсона (New England BioLabs, кат. № E2621S) в соответствии с протоколом производителя. Сборными продуктами в pStx56 трансформировали химически компетентные бактериальные клетки Turbo Competent *E. coli*, и клетки высевали на чашки с LB-агаром, содержащие канамицин. Отдельные колонии собирали и проводили мини-выделение с помощью набора Qiagen spin Miniprep Kit в соответствии с протоколом производителя. Полученные плазмиды секвенировали по методу Сэнгера для обеспечения

правильности сборки. pStX34 содержит промотор EF-1α для белка, а также селективный маркер, обеспечивающий устойчивость как к пуромицину, так и к карбенициллину. pStX56 содержит промотор EF-1α для белка, а также селективный маркер, обеспечивающий устойчивость как к пуромицину, так и к канамицину. Последовательности, кодирующие нацеливающие последовательности, которые нацелены на интересующий ген, были разработаны на основе местоположений PAM CasX. Нацеливающую ДНК-последовательность заказывали в виде одноцепочечных ДНК-олигонуклеотидов (оцДНК) (Integrated DNA Technologies), состоящих из нацеливающей последовательности и обратного комплемента этой последовательности. Был проведен отжиг этих двух олигонуклеотидов, и их клонировали в pStX по отдельности или вместе методом сборки Golden Gate с применением ДНК-лигазы T4 и соответствующего фермента рестрикции для плазмиды. Продуктами сборки Golden Gate трансформировали химически компетентные или электрокомпетентные клетки, такие как NEB Turbo Competent *E. coli* (NEB, кат. № C2984I), и клетки высевали на чашки с LB-агаром, содержащие соответствующий антибиотик. Отдельные колонии собирали и проводили мини-выделение с помощью набора Qiaprep spin Miniprep Kit в соответствии с протоколом производителя. Полученные плазмиды секвенировали по методу Сэнгера для обеспечения правильности лигирования.

[0946] Для построения CasX 535-537 (последовательности в таблице 9) ДНК конструкта CasX 515 амплифицировали методом ПЦР в двух реакциях для каждого конструкта с применением ДНК-полимеразы Q5 в соответствии с протоколом производителя. Для CasX 535 для амплификации применяли соответствующие праймеры. Для CasX 536 применяли соответствующие праймеры. Для CasX 537 применяли соответствующие праймеры. Продукты ПЦР очищали посредством гель-экстракции из 1% агарозного геля с помощью набора Zymoclean Gel DNA Recovery Kit в соответствии с протоколом производителя. Основную цепь pStX расщепляли с помощью XbaI и SpeI для удаления фрагмента ДНК размером 2931 пара нуклеотидов между двумя сайтами в плазмиде pStx56. Фрагмент расщепленной основной цепи очищали посредством гель-экстракции из 1% агарозного геля с помощью набора Zymoclean Gel DNA Recovery Kit в соответствии с протоколом производителя. Затем фрагменты вставки и основной цепи соединяли вместе с использованием сборки Гибсона в соответствии с протоколом производителя. Сборными продуктами в pStx56 трансформировали химически компетентные бактериальные клетки Turbo Competent *E. coli*, и клетки высевали клетки на чашки с LB-агаром, содержащим канамицин. Отдельные колонии собирали и проводили мини-выделение с помощью набора Qiagen spin Miniprep Kit в соответствии с протоколом производителя. Полученные плазмиды секвенировали по методу Сэнгера для обеспечения правильности сборки. pStX34 содержит промотор EF-1α для белка, а также селективный маркер, обеспечивающий устойчивость как к пуромицину, так и к карбенициллину. pStX56 содержит промотор EF-1α для белка, а также селективный маркер, обеспечивающий устойчивость как к пуромицину, так и к канамицину. Последовательности, кодирующие нацеливающие последовательности, которые нацелены на интересующий ген, были

разработаны на основе местоположений PAM CasX. Нацеливающую ДНК-последовательность заказывали в виде одноцепочечных ДНК-олигонуклеотидов (оцДНК) (Integrated DNA Technologies), состоящих из нацеливающей последовательности и обратного комплемента этой последовательности. Был проведен отжиг этих двух олигонуклеотидов, и их клонировали в pStX по отдельности или вместе методом сборки Golden Gate с применением ДНК-лигазы T4 и соответствующего фермента рестрикции для плазмиды. Продуктами сборки Golden Gate трансформировали химически или электрокомпетентные клетки, такие как NEB Turbo Competent *E. coli*, и клетки высевали на чашки с LB-агаром, содержащим соответствующий антибиотик. Отдельные колонии собирали и проводили мини-выделение с помощью набора Qiaprep spin Miniprep Kit в соответствии с протоколом производителя. Полученные плазмиды секвенировали по методу Сэнгера для обеспечения правильности лигирования.

[0947] Все последующие варианты CasX, такие как CasX 544 и CasX 660-664, 668, 670, 672, 676 и 677, клонировали с применением той же методологии, которая описана выше, с использованием сборки Гибсона с внутренними праймерами, несущими специфические мутации, и универсальными прямыми и обратными праймерами (различия заключались в сконструированных праймерах, несущих специфические мутации, а также в том, какой базовый конструкт CasX применялся). Контрольные плазмиды SaCas9 и SpyCas9 получали аналогично описанным выше плазмидам pStX, причем белок-кодирующие и гидовые области pStX заменяли на соответствующие белок-кодирующую область и гид. Нацеливающие последовательности для SaCas9 и SpyCas9 были либо взяты из литературы, либо рационально разработаны в соответствии с известными способами.

[0948] Экспрессию и извлечение конструктов CasX проводили с применением стандартных методик, как в обобщенном виде описано ниже.

Очистка:

[0949] Замороженные пробы размораживали в течение ночи при 4 ˚C с перемешиванием на магнитной мешалке. Вязкость полученного лизата снижали посредством обработки и завершали лизис гомогенизацией, дважды обработав в NanoDeBEE (BEE International) под давлением 20 000 фунтов на квадратный дюйм. Лизат осветляли центрифугированием при 50 000 × g, 4 ˚C в течение 30 минут и собирали надосадочную жидкость. Очищенную надосадочную жидкость наносили на колонку Heparin 6 Fast Flow (Cytiva) с помощью AKTA Pure FPLC (Cytiva). Колонку промывали 5 объемами колонки буфера Heparin A (50 мМ HEPES-NaOH, 250 мМ NaCl, 5 мМ $MgCl_2$, 0,5 мМ TCEP, 10% глицерин, pH 8), затем 3 объемами колонки буфера Heparin B (буфер A с концентрацией NaCl, доведенной до 500 мМ). Белок элюировали 1,75 объема колонки буфера Heparin C (буфер A с концентрацией NaCl, доведенной до 1 М). Элюат наносили на колонку с StrepTactis HP (Cytiva) с помощью системы для ЖХБР. Колонку промывали 10 объемами колонки буфера Strep (50 мМ HEPES-NaOH, 500 мМ NaCl, 5 мМ $MgCl_2$, 0,5 мМ TCEP, 10% глицерин, pH 8). Белок элюировали из колонки с применением 1,65 объема колонки буфера Strep с добавлением 2,5 мМ дестиобиотина. Фракции, содержащие CasX,

объединяли, концентрировали при 4 ˚C с помощью центрифужного концентратора с пределом отсечения по молекулярной массе 50 кДа (Amicon) и очищали посредством эксклюзионной хроматографии на колонке Superdex 200 пг (Cytiva). Колонку уравновешивали буфером SEC (25 мМ фосфата натрия, 300 мМ NaCl, 1 мМ TCEP, 10% глицерин, pH 7,25) и проводили ЖХБР. Фракции, содержащие CasX, которые элюировались при соответствующей молекулярной массе, объединяли, концентрировали при 4 ˚C с помощью центрифужного концентратора с пределом отсечения по молекулярной массе 50 кД, делили на аликвоты и мгновенно замораживали в жидком азоте перед хранением при -80 ˚C.

[0950] Вариант CasX 488: средний выход составил 2,7 мг очищенного белка CasX на литр культуры при чистоте 98,8% по оценке с помощью окрашивания коллоидным Кумасси.

[0951] Вариант CasX 491: средний выход составил 12,4 мг очищенного белка CasX на литр культуры при чистоте 99,4% по оценке с помощью окрашивания коллоидным Кумасси.

[0952] Вариант CasX 515: средний выход составил 7,8 мг очищенного белка CasX на литр культуры при чистоте 90% по оценке с помощью окрашивания коллоидным Кумасси.

[0953] Вариант CasX 526: средний выход составил 13,79 мг на литр культуры при чистоте 93%. Чистоту оценивали с помощью окрашивания коллоидным Кумасси.

[0954] Вариант CasX 668: средний выход составил 3,32 мг на литр культуры при чистоте 93%. Чистоту оценивали с помощью окрашивания коллоидным Кумасси.

[0955] Вариант CasX 672: средний выход составил 6,50 мг на литр культуры при чистоте 88%. Чистоту оценивали с помощью окрашивания коллоидным Кумасси.

[0956] Вариант CasX 676: средний выход составил 5,05 мг на литр культуры при чистоте 92%. Чистоту оценивали с помощью окрашивания коллоидным Кумасси.

[0957] Вариант CasX 677: средний выход составил 2,93 мг на литр культуры при чистоте 81%. Чистоту оценивали с помощью окрашивания коллоидным Кумасси.

Таблица 9. Последовательности ДНК и аминокислотные последовательности вариантов CasX

Конструкт	SEQ ID NO ДНК	SEQ ID NO аминокислоты
CasX 488	1027	1045
CasX 491	1028	1046
CasX 515	1029	1047
CasX 527	1030	1048
CasX 535	1031	1049
CasX 536	1032	1050
CasX 537	1033	1051
CasX 583	1034	1052

Конструкт	SEQ ID NO ДНК	SEQ ID NO аминокислоты
CasX 660	1035	1053
CasX 661	1036	1054
CasX 662	1037	1055
CasX 663	1038	1056
CasX 664	1039	1057
CasX 668	1040	1058
CasX 670	1041	1059
CasX 672	1042	1060
CasX 676	1043	1061
CasX 677	1044	1062

Пример 2. Создание гидовых РНК

[0958] Для создания одномолекулярных гидовых РНК и нацеливающих последовательностей создавали матрицы для транскрипции *in vitro* посредством проведения ПЦР с полимеразой Q5, матричными праймерами для каждой основной цепи и праймерами для амплификации с промотором T7 и нацеливающей последовательностью. Последовательности ДНК-праймеров для промотора T7, гидовой и нацеливающей последовательности для гидов и нацеливающих последовательностей представлены в таблице 10 ниже. Гиды sg1, sg2, sg32, sg64, sg174 и sg235 соответствуют SEQ ID NO: 4, 5, 2104, 2106, 2238 и 2292 соответственно, за исключением того, что sg2, sg32 и sg64 были модифицированы дополнительным G на 5'-конце для повышения эффективности транскрипции (сравните последовательности в таблице 10 с таблицей 2). Нацеливающая последовательность 7.37 нацелена на бета 2-микроглобулин (B2M). После ПЦР-амплификации матрицы очищали и выделяли посредством экстракции фенолом - хлороформом - изоамиловым спиртом с последующим осаждением этанолом.

[0959] Транскрипцию *in vitro* проводили в буфере, содержащем 50 мМ Tris, pH 8,0, 30 мМ $MgCl_2$, 0,01% Triton X-100, 2 мМ спермидина, 20 мМ DTT, 5 мМ нуклеозидтрифосфатов (НТФ), 0,5 мкМ матрицы и 100 мкг/мл РНК-полимеразы T7. Реакционные смеси инкубировали при 37°C в течение ночи. Добавляли 20 единиц ДНКазы I (Promega, № M6101) на 1 мл объема транскрипционной смеси и инкубировали в течение одного часа. РНК-продукты очищали электрофорезом в ПААГ в денатурирующих условиях, осаждали этанолом и ресуспендировали в 1× фосфатно-солевом буферном растворе. Для осуществления свертывания огРНК пробы нагревали до 70 °C в течение 5 мин, а затем охлаждали до комнатной температуры. В реакционные смеси добавляли $MgCl_2$ до конечной концентрации 1 мМ, нагревали до 50°C в течение 5 мин, а затем охлаждали до комнатной температуры. Конечные продукты - гидовые РНК - хранили при -80 °C.

Таблица 10. Последовательности ДНК-праймеров для промотора T7, гидовой и нацеливающей последовательности для гидов

Праймер		SEQ ID NO
Праймер с промотором T7	GAAATTAATACGACTCACTATA (используется для всех)	169
	РНК-продукт	
Прямой праймер основной цепи sg2	GGUACUGGCGCUUUUAUCUCAUUACUUU GAGAGCCAUCACCAGCGACUAUGUCGUA UGGGUAAAGCGCUUAUUUAUCGGAGAGA AAUCCGAUAAAUAAGAAGCAUCAAAGGG CCGAGAUGUCUCGCUCCG	1063
Обратный праймер основной цепи sg2		
Праймер для спейсера sg2.7.37		
Прямой праймер основной цепи sg32	GGUACUGGCGCUUUUAUCUCAUUACUUU GAGAGCCAUCACCAGCGACUAUGUCGUA UGGGUAAAGCGCCCUCUUCGGAGGGAAG CAUCAAAGGGCCGAGAUGUCUCG	1064
Обратный праймер основной цепи sg32		
Праймер для спейсера sg32.7.37		
Прямой праймер основной цепи sg64	GGUACUGGCGCCUUUAUCUCAUUACUUU GAGAGCCAUCACCAGCGACUAUGUCGUA UGGGUAAAGCGCUUACGGACUUCGGUCC GUAAGAAGCAUCAAAGGGCCGAGAUGUC UCGCUCCG	1065
Обратный праймер основной цепи sg64		
Праймер для спейсера sg64.7.37		
Прямой праймер основной цепи sg174	ACUGGCGCUUUUAUCUgAUUACUUUGAG AGCCAUCACCAGCGACUAUGUCGUAgUGG GUAAAGCUCCCUCUUCGGAGGGAGCAUC AAAGGGCCGAGAUGUCUCGCUCCG	1066
Обратный праймер основной цепи sg174		
Праймер для спейсера sg174.7.37		
Прямой праймер основной цепи sg235	ACUGGCGCUUCUAUCUGAUUACUCUGAG CGCCAUCACCAGCGACUAUGUCGUAGUG GGUAAAGCCGCUUACGGACUUCGGUCCG UAAGAGGCAUCAGAG	1067
Обратный праймер основной цепи sg235		
Праймер для спейсера sg235.7.37		

Пример 3. Оценка аффинности связывания с гидовой РНК

[0960] Очищенный CasX дикого типа и улучшенный CasX будут инкубировать с синтетической одномолекулярной гидовой РНК, содержащей Cy7.5-фрагмент на 3’-конце, в низкосолевом буферном растворе, содержащем хлорид магния и гепарин для предотвращения неспецифического связывания и агрегации. Концентрация огРНК будет постоянной - 10 пМ, в то время как белок будут титровать от 1 пМ до 100 мкМ в отдельных реакциях связывания. После того как реакция придет в равновесное состояние, пробы пропустят через вакуумный коллектор для анализа связывания с фильтром, используя

нитроцеллюлозную мембрану и положительно заряженную нейлоновую мембрану, которые связывают белок и нуклеиновую кислоту соответственно. На мембранах проведут визуализацию для идентификации гидовой РНК, и будут определены фракции связанной и несвязанной РНК по количеству флуоресценции на нитроцеллюлозной и нейлоновой мембране при каждой концентрации белка для расчета константы диссоциации комплекса белок-огРНК. Также будет проведен эксперимент с улучшенными вариантами огРНК для определения того, влияют ли эти мутации на аффинность к гиду белков дикого типа и мутантных белков. Авторы изобретения также выполнят анализ методом сдвига электрофоретической подвижности для качественного сравнения с результатами анализа связывания с фильтром и подтверждения того, что основным фактором ассоциации белков и РНК является связывание в растворе, а не агрегация.

Пример 4. Оценка аффинности связывания с целевой ДНК

[0961] Очищенный CasX дикого типа и улучшенный CasX будут объединены в комплекс с одномолекулярной гидовой РНК, несущей нацеливающую последовательность, комплементарную целевой нуклеиновой кислоте. RNP-комплекс будут инкубировать с двухцепочечной целевой ДНК, содержащей PAM и соответствующую последовательность целевой нуклеиновой кислоты с Cy7.5-меткой на 5'-конце целевой цепи, в низкосолевом буферном растворе, содержащем хлорид магния и гепарин для предотвращения неспецифического связывания и агрегации. Концентрация целевой ДНК будет постоянной - 1 нМ, в то время как RNP будут титровать от 1 пМ до 100 мкМ в отдельных реакциях связывания. После того как реакция придет в равновесное состояние, пробы пропустят через нативный 5% полиакриламидный гель для отделения связанной и несвязанной целевой ДНК. Будет проведена визуализация геля для определения сдвигов подвижности целевой ДНК и будут рассчитаны фракции связанной и несвязанной ДНК при каждой концентрации белка для определения константы диссоциации тройного комплекса RNP-целевая ДНК.

Пример 5. Оценка различий в распознавании PAM *in vitro.* [Используйте SCRB-038 для ФИГ. №]

1. Сравнение референсных CasX и вариантов CasX

[0962] Анализ расщепления *in vitro* проводили с помощью CasX 2, CasX 119 и CasX 438 в комплексе с sg174.7.37, по существу как описано в примере 8. Применяли флуоресцентно-меченые целевые дцДНК со спейсером 7.37 и PAM-последовательность TTC, CTC, GTC или ATC (последовательности приведены в таблице 11). Пробы отбирали во временных точках 0,25, 0,5, 1, 2, 5, 10, 30 и 60 минут. Для визуализации гелей применяли Cytiva Typhoon и проводили количественную оценку с помощью программного обеспечения IQTL 8.2. Кажущиеся константы скорости реакции первого порядка для расщепления нецелевой цепи ($k_{расщепл.}$) определяли для каждого комплекса CasX : огРНК на каждой мишени. Константы скорости для мишеней с PAM без TTC сравнивали с мишенями с PAM TTC, чтобы определить, изменялось ли относительное предпочтение к каждой PAM в данном варианте белка.

[0963] Для всех вариантов мишень с TTC поддерживала самую высокую скорость расщепления, затем следовала мишень с ATC, затем - с CTC и, наконец, с GTC (ФИГ. 10A-D, таблица 12). Для каждой комбинации варианта CasX и PAM NTC приведена скорость расщепления $k_{расщепл.}$ Для всех PAM без NTC в скобках показано отношение скорости расщепления к скорости при наличии TTC для этого варианта. Все PAM без TTC демонстрировали существенно сниженные скорости расщепления (в > 10 раз во всех случаях). Соотношение скорости расщепления данной PAM без TTC и PAM TTC для конкретного варианта оставалось по существу постоянным у всех вариантов. В случае мишени с CTC обеспечивалось расщепление со скоростью 3,5-4,3% от скорости расщепления мишени с TTC; в случае мишени с GTC обеспечивалось расщепление со скоростью 1,0-1,4%; а в случае мишени с ATC обеспечивалось расщепление со скоростью 6,5-8,3%. Исключение составляет вариант 491, для которого кинетика расщепления при наличии PAM TTC слишком быстрая, чтобы ее можно было точно измерить, что искусственно уменьшает кажущуюся разницу между PAM TTC и PAM без TTC. Сравнение относительных скоростей для 491 при наличии PAM GTC, CTC и ATC, которые попадают в поддающийся измерению диапазон, дает соотношения, сопоставимые с полученными для других вариантов при сравнении по PAM без TTC, что согласуется с повышением скоростей в этом ряду. В целом, различия между вариантами не являются достаточно существенными, чтобы предположить, что у них будут изменены относительные предпочтения к различным PAM NTC. Однако более высокие базальные скорости расщепления у вариантов позволяют расщеплять мишени с PAM ATC или CTC практически полностью в течение 10 минут, а кажущиеся $k_{расщепл.}$ сопоставимы или выше, чем $k_{расщепл.}$ у CasX 2 на PAM TTC (таблица 12). Такая повышенная скорость расщепления может пересекать пороговое значение, необходимое для эффективного редактирования генома в человеческой клетке, что объясняет очевидное увеличение гибкости требований к PAM для этих вариантов.

Таблица 11. Последовательности ДНК-субстратов, применяемых в анализе *in vitro* расщепления PAM

Гид*	**ДНК-последовательность**	**SEQ ID NO**
7.37 TTC PAM TS	AGCGCGAGCACAGCTAAGGCCACGGAGCGAGACATCTCG GCCC**GAA**TGCTGTCAGCTTCA	1068
7.37 TTC PAM NTS	TGAAGCTGACAGCA**TTC**GGGCCGAGATGTCTCGCTCCGTG GCCTTAGCTGTGCTCGCGCT	1069
7.37 CTC PAM TS	AGCGCGAGCACAGCTAAGGCCACGGAGCGAGACATCTCG GCCC**GAG**TGCTGTCAGCTTCA	1070
7.37 CTC PAM NTS	TGAAGCTGACAGCA**CTC**GGGCCGAGATGTCTCGCTCCGT GGCCTTAGCTGTGCTCGCGCT	1071
7.37 GTC PAM TS	AGCGCGAGCACAGCTAAGGCCACGGAGCGAGACATCTCG GCCC**GAC**TGCTGTCAGCTTCA	1072
7.37 GTC PAM NTS	TGAAGCTGACAGCA**GTC**GGGCCGAGATGTCTCGCTCCGT GGCCTTAGCTGTGCTCGCGCT	1073

7.37 ATC PAM TS	AGCGCGAGCACAGCTAAGGCCACGGAGCGAGACATCTCG GCCC**GAT**TGCTGTCAGCTTCA	1074
7.37 ATC PAM NTS	TGAAGCTGACAGCA**ATC**GGGCCGAGATGTCTCGCTCCGT GGCCTTAGCTGTGCTCGCGCT	1075

* PAM-последовательности для каждого из них выделены жирным шрифтом. TS - целевая цепь. NTS - нецелевая цепь.

Таблица 12. Кажущиеся скорости расщепления PAM NTC для вариантов CasX

Вариант	**TTC**	**CTC**	**GTC**	**ATC**
2	0,267 мин$^{-1}$	9,29E-3 мин$^{-1}$ (0,035)	3,75E-3 мин$^{-1}$ (0,014)	1,87E-2 мин$^{-1}$ (0,070)
119	8,33 мин$^{-1}$	0,303 мин$^{-1}$ (0,036)	8,64E-2 мин$^{-1}$ (0,010)	0,540 мин$^{-1}$ (0,065)
438	4,94 мин$^{-1}$	0,212 мин$^{-1}$ (0,043)	1,31E-2 мин$^{-1}$ (0,013)	0,408 мин$^{-1}$ (0,083)
491	16,42 мин$^{-1}$	8,605 мин$^{-1}$ (0,524)	2,447 мин$^{-1}$ (0,149)	11,33 мин$^{-1}$ (0,690)

2. Сравнение распознавания PAM с применением одного варианта CasX

[0964] Материалы и методы. Применяли флуоресцентно-меченые целевые дцДНК со спейсером 7.37 и PAM-последовательности TTC, CTC, GTC, ATC, TTT, CTT, GTT или ATT (последовательности приведены в таблице 13). Были заказаны олигонуклеотиды с аминокислотной модификацией 5'-конца, которые метили с помощью N-гидроксисукцинимидного (NHS) сложного эфира Cy7.5 в случае олигонуклеотидов целевой цепи и NHS сложного эфира Cy5.5 в случае олигонуклеотидов нецелевой цепи. Целевые дцДНК получали посредством смешивания олигонуклеотидов в соотношении 1 : 1 в 1× буфере для расщепления (20 мМ Tris-HCl, pH 7,5, 150 мМ NaCl, 1 мМ TCEP, 5% глицерина, 10 мМ $MgCl_2$), нагревания до 95°C в течение 10 минут и охлаждения раствора до комнатной температуры.

[0965] Вариант 491 CasX объединяли в комплекс с sg174.7.37. Гид разбавляли в 1× буфере для расщепления до конечной концентрации 1,5 мкМ, а затем добавляли белок до конечной концентрации 1 мкМ. RNP инкубировали при 37°C в течение 10 минут и затем помещали на лед.

[0966] Анализ расщепления проводили посредством разбавления RNP в буфере для расщепления до конечной концентрации 200 нМ и добавления целевой дцДНК до конечной концентрации 10 нМ. Пробы отбирали во временных точках 0,25, 0,5, 1, 2, 5 и 10 минут и гасили реакцию посредством добавления к равному объему 95% формамида и 20 мМ ЭДТА. Продукты расщепления разделяли в 10% ПААГ с мочевиной. Для визуализации гелей использовали Amersham Typhoon и проводили количественную оценку с помощью программного обеспечения IQTL 8.2. Кажущиеся константы скорости реакции первого порядка для расщепления нецелевой цепи ($k_{расщепл.}$) определяли для каждой целевой ДНК с применением GraphPad Prism.

Результаты:

[0967] Исследовали относительную скорость расщепления 491.174 RNP на различных PAM. В дополнение к тому, что эти данные также помогут в прогнозировании

эффективности расщепления целевых и потенциальных нецелевых последовательностей в клетках, они также позволяют корректировать скорость расщепления синтетических целевых последовательностей. Авторы изобретения предположили, что в случае самоограничивающихся AAV-векторов, когда в вектор могут быть добавлены новые протоспейсеры для обеспечения самонацеливания, скорость расщепления эписом может быть увеличена или уменьшена путем изменения PAM.

[0968] Мы тестировали скорость расщепления RNP в отношении различных субстратов дцДНК, которые были идентичны по последовательности, за исключением PAM. Этот экспериментальный подход должен позволить изолированно рассмотреть влияние собственно PAM, не смешивая распознавание PAM с эффектами, обусловленными спейсерной последовательностью и геномным контекстом. Были протестированы все PAM NTC и NTT. Как и ожидалось, RNP расщепляли целевую ДНК с PAM TTC быстрее, превращая по существу ее всю в продукт к первой временной точке (ФИГ. 11A). CTC расщеплялась примерно в два раза медленнее, хотя быстрое расщепление TTC усложняет определение точного значения $k_{расщепл.}$ в таких условиях анализа, которые оптимизированы для регистрации более широкого диапазона скоростей расщепления (ФИГ. 11A, таблица 14). Целевая GTC расщеплялась медленнее всего среди PAM NTC, при этом скорость расщепления была примерно в шесть раз ниже, чем в случае целевой TTC. Все PAM NTT расщеплялись медленнее, чем все PAM NTC, при этом TTT расщеплялась наиболее эффективно, а после нее следовала GTT (ФИГ. 11B, таблица 14). Относительная эффективность расщепления GTT среди всех PAM NTT по сравнению с низкой скоростью расщепления GTC в сравнении со всеми PAM NTC демонстрирует, что распознавание отдельных нуклеотидов PAM зависит от контекста, причем конкретный нуклеотид в одном положении в PAM влияет на зависимость предпочтения последовательности от других положений.

[0969] Тестируемые в настоящем изобретении PAM-последовательности дают скорости расщепления, охватывающие три порядка величины, с сохранением активности расщепления в одной и той же спейсерной последовательности. Эти данные демонстрируют, что скорости расщепления в данной синтетической целевой последовательности могут быть легко модифицированы посредством изменения ассоциированной PAM, что позволяет корректировать активность саморасщепления для обеспечения эффективного нацеливания на геномную целевую последовательность перед расщеплением и элиминацией AAV эписомы.

Таблица 13. Последовательности ДНК-субстратов, применяемых в анализе *in vitro* расщепления PAM*

PAM и цепь	Спейсерная и PAM-последовательность	SEQ ID NO
7.37 TTC PAM TS	AGCGCGAGCACAGCTAAGGCCACGGAGCGAGACATCT CGGCCC**GAA**TGCTGTCAGCTTCA	1068

7.37 TTC PAM NTS	TGAAGCTGACAGCA**TTC**GGGCCGAGATGTCTCGCTCC GTGGCCTTAGCTGTGCTCGCGCT	1069
7.37 CTC PAM TS	AGCGCGAGCACAGCTAAGGCCACGGAGCGAGACATCT CGGCCC**GAG**TGCTGTCAGCTTCA	1070
7.37 CTC PAM NTS	TGAAGCTGACAGCA**CTC**GGGCCGAGATGTCTCGCTCC GTGGCCTTAGCTGTGCTCGCGCT	1071
7.37 GTC PAM TS	AGCGCGAGCACAGCTAAGGCCACGGAGCGAGACATCT CGGCCC**GAC**TGCTGTCAGCTTCA	1072
7.37 GTC PAM NTS	TGAAGCTGACAGCA**GTC**GGGCCGAGATGTCTCGCTCC GTGGCCTTAGCTGTGCTCGCGCT	1073
7.37 ATC PAM TS	AGCGCGAGCACAGCTAAGGCCACGGAGCGAGACATCT CGGCCC**GAT**TGCTGTCAGCTTCA	1074
7.37 ATC PAM NTS	TGAAGCTGACAGCA**ATC**GGGCCGAGATGTCTCGCTCC GTGGCCTTAGCTGTGCTCGCGCT	1075
7.37 TTT PAM TS	AGCGCGAGCACAGCTAAGGCCACGGAGCGAGACATCT CGGCCC**AAA**TGCTGTCAGCTTCA	1076
7.37 TTT PAM NTS	TGAAGCTGACAGCA**TTT**GGGCCGAGATGTCTCGCTCCG TGGCCTTAGCTGTGCTCGCGCT	1077
7.37 CTT PAM TS	AGCGCGAGCACAGCTAAGGCCACGGAGCGAGACATCT CGGCCC**TAG**TGCTGTCAGCTTCA	1078
7.37 CTT PAM NTS	TGAAGCTGACAGCA**CTT**GGGCCGAGATGTCTCGCTCC GTGGCCTTAGCTGTGCTCGCGCT	1079
7.37 GTT PAM TS	AGCGCGAGCACAGCTAAGGCCACGGAGCGAGACATCT CGGCCC**TAC**TGCTGTCAGCTTCA	1080
7.37 GTT PAM NTS	TGAAGCTGACAGCA**GTT**GGGCCGAGATGTCTCGCTCC GTGGCCTTAGCTGTGCTCGCGCT	1081
7.37 ATT PAM TS	AGCGCGAGCACAGCTAAGGCCACGGAGCGAGACATCT CGGCCC**TAT**TGCTGTCAGCTTCA	1082
7.37 ATT PAM NTS	TGAAGCTGACAGCA**ATT**GGGCCGAGATGTCTCGCTCC GTGGCCTTAGCTGTGCTCGCGCT	1083

* Показаны ДНК-последовательности, применяемые для создания каждого субстрата дцДНК. PAM-последовательности для каждого из них выделены жирным шрифтом. TS - целевая цепь. NTS - нецелевая цепь.

Таблица 14. Кажущиеся скорости расщепления PAM NTC и NTT для CasX 491.174

PAM	TTC	ATC	CTC	GTC	TTT	ATT	CTT	GTT
$k_{расщепл.}$ ($мин^{-1}$)	15,6*	6,66	9,45	2,52	1,33	0,0675	0,0204	0,330

* Скорость расщепления TTC превышает разрешение этого анализа, поэтому полученное значение $k_{расщепл.}$ следует рассматривать в качестве нижней границы.

Пример 6. Оценка нуклеазной активности для двухцепочечного расщепления

[0970] Очищенные CasX дикого типа и сконструированные варианты CasX будут объединены в комплекс с одномолекулярной гидовой РНК, несущей фиксированную нацеливающую последовательность для HRS. RNP-комплексы добавят к буферу, содержащему $MgCl_2$ в конечной концентрации 100 нМ, и будут инкубировать с двухцепочечной целевой ДНК, несущей Cy7.5-метку на 5'-конце целевой или нецелевой цепи, в концентрации 10 нМ. Аликвоты реакционных смесей будут отбирать в фиксированных временных точках и гасить реакцию посредством добавления равного объема 50 мМ ЭДТА и 95% формамида. Будет проведено разделение проб в полиакриламидном геле в денатурирующих условиях для разделения расщепленных и нерасщепленных ДНК-субстратов. Результаты визуализируют и определят скорости расщепления целевых и нецелевых цепей белком дикого типа и сконструированными вариантами. Для более четкого различения между изменениями в связывании с целевой последовательностью и скоростью катализа самой нуклеолитической реакции концентрацию белка будут титровать в диапазоне от 10 нМ до 1 мкМ, и при каждой концентрации будут определять скорости расщепления для получения псевдокривой Михаэлиса - Ментен и определения kкат.*и K_M*. Изменения K_M* указывают на изменение связывания, а изменения kкат.* указывают на изменение катализа.

Пример 7. Анализ PASS для идентификации вариантов белка CasX с разной специфичностью в отношении PAM-последовательности

[0971] Были проведены эксперименты для идентификации специфичности к PAM-последовательности белков CasX 2 (SEQ ID NO: 2), 491, 515, 533, 535, 668 и 672. Для обеспечения этого клеточную линию HEK293 PASS_V1.01 или PASS_V1.02 обрабатывали вышеуказанными белками CasX в по меньшей мере двух повторных экспериментах и проводили секвенирование нового поколения (NGS) для расчета процентной доли редактирования с применением разнообразных спейсеров в их заданных целевых сайтах.

[0972] Материалы и методы. Для анализа вариантов клонированного белка применяли мультиплексный объединенный подход с помощью системы PASS. Вкратце, были созданы две объединенные клеточные линии HEK293, названные PASS_V1.01 и PASS_V1.02. Каждая клетка в пределах пула содержала интегрированную в геном одномолекулярную гидовую РНК (огРНК), спаренную со специфическим целевым сайтом. После трансфекции конструктами, экспрессирующими белок, редактирование по специфической мишени специфическим спейсером можно количественно оценивать с помощью NGS. Каждая пара гида и мишени была разработана таким образом, чтобы можно было получать данные, относящиеся к активности, специфичности и нацеливаемости RNP-комплекса CasX - гид.

[0973] Спаренные последовательности спейсер - целевой сайт были синтезированы компанией Twist Biosciences и получены в виде эквимолярного пула олигонуклеотидов. Этот пул амплифицировали методом ПЦР и клонировали методом Golden Gate для создания конечной библиотеки плазмид, названной p77. Каждая плазмида содержала элемент,

экспрессирующий огРНК и целевой сайт, вместе с экспрессирующим GFP элементом. Экспрессирующий огРНК элемент состоял из промотора U6, приводящего в действие транскрипцию каркаса гРНК 174 (SEQ ID NO: 2238), далее следовала спейсерная последовательность, которая должна нацеливать RNP из гида и варианта CasX на необходимый целевой сайт. Были сконструированы и синтезированы 250 возможных уникальных спаренных синтетических последовательностей спейсер - целевой сайт. Затем из этой библиотеки плазмид получали пул лентивируса с помощью системы LentiX (Takara Bio USA, Inc) в соответствии с инструкциями производителя. После этого полученный вирусный препарат количественно оценивали с помощью количественной полимеразной цепной реакции (количественной ПЦР) и трансдуцировали в стандартную клеточную линию HEK293 при низкой множественности инфекции, чтобы создать однокопийные интегрирования. Затем полученную клеточную линию очищали с помощью цитометрии посредством сортировки клеток с активацией флуоресценции (FACS) для завершения получения PASS_V1.01 или PASSS_V1.02. После этого клеточную линию высевали в шестилуночный планшет и в двух повторностях либо обрабатывали водой, либо трансфицировали 2 мкг плазмиды p67, доставляемой с помощью реагента для трансфекции Lipofectamine (ThermoFisher) в соответствии с инструкциями производителя. Плазмида p67 содержит промотор EF-1альфа, направляющий экспрессию белка CasX, меченного последовательностью сигнала ядерной локализации SV40. Через два дня обработанные клетки собирали, лизировали и экстрагировали геномную ДНК с помощью набора для выделения геномной ДНК (Zymo Research). Затем геномную ДНК амплифицировали методом ПЦР с помощью специально разработанных праймеров для создания ампликонов, совместимых с NGS компании Illumina, и проводили секвенирование на приборе NextSeq. Прочтения проб демультиплексировали и фильтровали в зависимости от их качества. После этого количественно определяли показатели результатов редактирования (доля прочтений с инделами) для каждой синтетической последовательности спейсер - целевой сайт в обработанных пробах.

[0974] Для оценки специфичности PAM-последовательности для белка CasX были категоризированы показатели результатов редактирования для четырех разных PAM-последовательностей. Для целевых сайтов с PAM TTC количественно оценивали 48 разных пар спейсер - целевой сайт; для целевых сайтов с PAM ATC, CTC и GTC количественно оценивали 14, 22 и 11 отдельных целевых сайтов соответственно. Для некоторых белков CasX повторные эксперименты проводили несколько десятков раз в течение нескольких месяцев. Для каждого из этих экспериментов рассчитывали среднюю эффективность редактирования для каждого из вышеописанных спейсеров. Затем рассчитали среднюю эффективность редактирования по четырем категориям PAM-последовательности по данным всех таких экспериментов, а также стандартное отклонение этих измерений.

[0975] Результаты: В таблице 15 приведена средняя эффективность редактирования для разных категорий PAM и вариантов белка CasX, а также стандартное отклонение этих измерений. Также указано число измерений для каждой категории. Эти данные

свидетельствуют о том, что сконструированные варианты CasX 491 и 515 являются специфичными в отношении канонической PAM-последовательности TTC, тогда как другие сконструированные варианты CasX с большей или меньшей эффективностью действовали на протестированные PAM-последовательности. В частности, средний ранговый порядок предпочтений PAM для CasX 491 представляет собой TTC >> ATC > CTC > GTC или TTC >> ATC > GTC > CTC для CasX 515, в то время как CasX 2 дикого типа демонстрирует средний ранговый порядок TTC >> GTC > CTC > ATC. Следует отметить, что для PAM-последовательностей с более низкой эффективностью редактирования ошибка этих средних измерений является высокой. Напротив, варианты CasX 535, 668 и 672 обладают значительно более широким спектром распознавания PAM с ранговым порядком TTC > CTC > ATC > GTC. Наконец, CasX 533 демонстрирует полностью измененный порядок ранжирования относительно характерного для CasX дикого типа (WT): ATC > CTC >> GTC > TTC. Эти данные можно применять для конструирования максимально активных терапевтических молекул CasX для интересующей целевой ДНК-последовательности.

[0976] В условиях экспериментов был идентифицирован набор белков CasX, которые более эффективно расщепляют двухцепочечную ДНК в человеческих клетках в целевых ДНК-последовательностях, ассоциированных с PAM-последовательностью TTC, ATC, CTC или GTC, что свидетельствует в пользу вариантов CasX с измененным спектром специфичности к PAM относительно

Таблица 15. Средний показатель редактирования, осуществляемого избранными белками CasX на спейсерах, ассоциированных с PAM-последовательностями TTC, ATC, CTC или GTC

Название CasX	**PAM-последовательность**	**Средняя процентная доля редактирования**	**Стандартное отклонение**	**Число измерений**
2	ATC	0,40	1,35	336
2	CTC	0,46	2,29	528
2	GTC	0,69	6,27	264
2	TTC	5,28	7,34	1152
491	ATC	6,86	8,29	364
491	CTC	4,54	6,40	572
491	GTC	3,40	6,68	286
491	TTC	40,41	23,13	1248
515	ATC	4,47	5,49	252
515	CTC	3,36	4,80	396

515	GTC	3,65	10,75	198
515	TTC	36,75	24,89	864
533	ATC	47,50	15,86	96
533	CTC	25,90	14,74	28
533	GTC	6,34	8,36	44
533	TTC	0,87	3,05	22
535	ATC	9,70	10,20	56
535	CTC	11,77	13,59	88
535	GTC	7,62	15,04	44
535	TTC	29,29	18,78	192
668	ATC	44,69	24,40	56
668	CTC	46,14	26,57	88
668	GTC	30,48	24,06	44
668	TTC	55,34	28,59	192
672	ATC	25,51	20,85	56
672	CTC	30,05	22,95	88
672	GTC	14,21	13,38	44
672	TTC	52,36	27,64	192

Пример 8. Анализ расщепления *in vitro* с использованием CasX : гРНК

1. Сборка RNP

[0977] Очищенный RNP дикого типа и RNP из CasX и одномолекулярной гидовой РНК (огРНК) либо получали непосредственно перед экспериментами, либо получали, мгновенно замораживали в жидком азоте и хранили при -80 °C для последующего применения. Для получения RNP-комплексов белок CasX инкубировали с огРНК при молярном соотношении 1 : 1,2. Вкратце, огРНК добавляли в буфер № 1 (25 мМ NaPi, 150 мМ NaCl, 200 мМ трегалозы, 1 мМ $MgCl_2$), затем к раствору огРНК медленно при перемешивании на вихревой мешалке добавляли CasX и инкубировали при 37°C в течение 10 мин для образования RNP-комплексов. RNP-комплексы перед применением фильтровали через фильтры Costar 8160 с размером пор 0,22 мкм, предварительно смоченные 200 мкл буфера № 1. При необходимости пробу RNP концентрировали с помощью фильтра 0,5 мл Ultra с пределом отсечения 100 кДа (Millipore, кат. № UFC510096) до получения требуемого объема. Образование компетентных RNP оценивали так, как описано ниже.

2. Определение компетентных в отношении расщепления фракций вариантов белка по сравнению с референсным CasX дикого типа

[0978] Способность вариантов CasX образовывать активные RNP в сравнении с референсным CasX определяли с помощью анализа расщепления *in vitro*. Мишень ДНК гена бета-2 микроглобулина (B2M) 7.37 для анализа расщепления была создана следующим образом. Олигонуклеотиды ДНК с последовательностью TGAAGCTGACAGCATTCGGGCCGAGATGTCTCGCTCCGTGGCCTTAGCTGTGCTCGC GCT (нецелевая цепь, NTS (SEQ ID NO: 1069)) и AGCGCGAGCACAGCTAAGGCCACGGAGCGAGACATCTCGGCCCGAATGCTGTCAGC TTCA (целевая цепь, TS (SEQ ID NO: 1068)) приобретали с 5'-флуоресцентными метками (LI-COR IRDye 700 и 800 соответственно). Целевые дцДНК образовывали посредством смешивания олигонуклеотидов в соотношении 1 : 1 в 1× буфере для расщепления (20 мМ Tris-HCl, pH 7,5, 150 мМ NaCl, 1 мМ TCEP, 5% глицерина, 10 мМ $MgCl_2$), нагревания до 95°C в течение 10 минут и охлаждения раствора до комнатной температуры.

[0979] RNP CasX восстанавливали с помощью указанных CasX и гидов (см. графики) в конечной концентрации 1 мкМ при 1,5-кратном избытке указанного гида, если не указано иное, в 1× буфере для расщепления (20 мМ Tris-HCl, pH 7,5, 150 мМ NaCl, 1 мМ TCEP, 5% глицерина, 10 мМ $MgCl_2$) при 37°C в течение 10 мин перед перемещением в лед до готовности к применению. Применяли целевую последовательность 7.37 вместе с огРНК, имеющей спейсеры, комплементарные целевой последовательности 7.37.

[0980] Готовили реакционные смеси с конечной концентрацией RNP 100 нМ и конечной концентрацией целевой последовательности 100 нМ. Реакции проводили при 37 °C и инициировали добавлением целевой ДНК 7.37. Аликвоты отбирали через 5, 10, 30, 60 и 120 минут и гасили реакцию добавлением 95% формамида с 20 мМ ЭДТА. Пробы денатурировали нагреванием при 95°C в течение 10 минут и разделяли в 10% ПААГ с мочевиной. Гели либо фотографировали с помощью LI-COR Odyssey CLx и количественно оценивали с помощью программного обеспечения LI-COR Image Studio, либо фотографировали с помощью Cytiva Typhoon и количественно оценивали с помощью программного обеспечения Cytiva IQTL. Полученные данные откладывали на графике и анализировали с помощью Prism. Авторы изобретения предположили, что в условиях анализа CasX работает как фермент однократного действия, на что указывает такой отмеченный факт, что субстехиометрические количества фермента неспособны расщеплять более чем стехиометрическое количество мишени даже при продолжительных временных интервалах, а вместо этого достигают плато, которое по масштабу соответствует количеству присутствующего фермента. Таким образом, фракция целевой последовательности, расщепленная в течение длительного времени с помощью эквимолярного количества RNP, указывает на то, какая фракция RNP была образована надлежащим образом и обладает расщепляющей активностью. Результаты для дорожек с продуктами расщепления аппроксимировали двухфазной моделью скорости, так как реакция расщепления явно отклоняется от монофазной модели в данном диапазоне

концентраций, и для каждой из трех независимых повторов определяли плато. Рассчитали среднее значение и стандартное отклонение для определения активной фракции (таблица 16).

[0981] Кажущиеся активные (компетентные) фракции определяли для RNP, образованных из CasX 2+гид 174+спейсер 7.37, CasX 119+гид 174+спейсер 7.37, CasX 457+гид 174+спейсер 7.37, CasX 488+гид 174+спейсер 7.37 и CasX 491+гид 174+спейсер 7.37, как показано на ФИГ. 1. Определенные активные фракции показаны в таблице 16. Все варианты CasX имели фракции с более высокой активностью, чем CasX 2 дикого типа, что указывает на то, что в тестовых условиях сконструированные варианты CasX образуют значительно более активные и стабильные RNP с идентичным гидом по сравнению с CasX дикого типа. Это может быть связано с повышенной аффинностью к огРНК, повышенной стабильностью или растворимостью в присутствии огРНК или большей стабильностью компетентной в отношении расщепления конформации сконструированного комплекса CasX : огРНК. На увеличение растворимости RNP указывало заметное уменьшение наблюдаемого осадка, образованного при добавлении CasX 457, CasX 488 или CasX 491 к огРНК, по сравнению с CasX 2.

3. Анализы расщепления *in vitro* - определение фракций, компетентных в отношении расщепления, для вариантов одномолекулярных гидов относительно референсных одномолекулярных гидов

[0982] С помощью того же протокола для CasX 2.2.7.37, CasX 2.32.7.37, CasX 2.64.7.37 и CasX 2.174.7.37 также определяли, что фракции, компетентные в отношении расщепления, составляют 16±3%, 13±3%, 5±2% и 22±5%, как показано на ФИГ. 2 и в таблице 16.

[0983] Второй набор гидов тестировали в других условиях для лучшего отделения вклада гида в образование RNP. Гиды 174, 175, 185, 186, 196, 214 и 215 со спейсером 7.37 смешивали с CasX 491 в конечных концентрациях 1 мкМ для гида и 1,5 мкМ для белка, а не с избытком гида, как ранее. Результаты представлены на ФИГ. 3 и в таблице 16. В таких условиях ограничения гида многие из этих гидов демонстрировали дополнительное улучшение по сравнению с гидом 174, при этом у гидов 185 и 196 компетентные фракции достигли 91±4% и 91±1% соответственно по сравнению с 80±9% у гида 174.

[0984] Данные указывают на то, что как варианты CasX, так и варианты огРНК способны образовывать более высокую долю активных RNP с гидовой РНК по сравнению с CasX дикого типа и огРНК дикого типа.

[0985] Кажущиеся скорости расщепления вариантов CasX 119, 457, 488 и 491 по сравнению с референсным CasX дикого типа определяли с помощью флуоресцентного анализа *in vitro* для расщепления целевой последовательности 7.37.

4. Анализ расщепления *in vitro* - определение $k_{расщепл.}$ для вариантов CasX по сравнению с референсным CasX дикого типа

[0986] RNP CasX восстанавливали с применением указанных CasX (см. ФИГ. 4) в конечной концентрации 1 мкМ при 1,5-кратном избытке указанного гида в 1× буфере для

расщепления (20 мМ Tris-HCl, pH 7,5, 150 мМ NaCl, 1 мМ TCEP, 5% глицерина, 10 мМ $MgCl_2$), при 37°C в течение 10 мин перед перемещением в лед до готовности к применению. Готовили реакционные смеси с конечной концентрацией RNP 200 нМ и конечной концентрацией целевой последовательности 10 нМ. Реакции проводили при 37 °C, если не указано иное, и инициировали добавлением целевой ДНК. Аликвоты отбирали через 0,25, 0,5, 1, 2, 5 и 10 минут и гасили реакцию посредством добавления 95% формамида с 20 мМ ЭДТА. Пробы денатурировали нагреванием при 95°C в течение 10 минут и разделяли в 10% ПААГ с мочевиной. Гели фотографировали с помощью LI-COR Odyssey CLx и количественно оценивали с помощью программного обеспечения LI-COR Image Studio или фотографировали с помощью Cytiva Typhoon и количественно оценивали с помощью программного обеспечения Cytiva IQTL. Полученные данные откладывали на графике и анализировали с помощью Prism, и кажущуюся константу скорости реакции первого порядка для расщепления нецелевой цепи ($k_{расщепл.}$) определяли для каждой повторной комбинации CasX : огРНК по отдельности. Среднее значение и стандартное отклонение трех повторов с независимой аппроксимацией представлены в таблице 16, и результаты для дорожек с продуктами расщепления показаны на ФИГ. 5.

[0987] Кажущиеся константы скорости расщепления для CasX 2 дикого типа и вариантов CasX 119, 457, 488 и 491 определяли с помощью гида 174 и спейсера 7.37 в каждом анализе (см. таблицу 16 и ФИГ. 4). Все варианты CasX имели улучшенные скорости расщепления относительно CasX 2 дикого типа. CasX 457 расщеплял медленнее, чем 119, несмотря на наличие более высокой компетентной фракции, как определено выше. CasX 488 и CasX 491 имели самые высокие скорости расщепления с большим отрывом от остальных вариантов; так как целевая ДНК была почти полностью расщеплена к первой временной точке, истинная скорость расщепления превышает разрешение этого анализа, и представленное значение $k_{расщепл.}$ следует принимать в качестве нижней границы.

[0988] Данные указывают на то, что варианты CasX имеют более высокий уровень активности, при этом скорости $k_{расщепл.}$ достигают превышения в по меньшей мере 30 раз по сравнению с CasX 2 дикого типа.

5. Анализ расщепления *in vitro*: сравнение вариантов гидов с гидами дикого типа

[0989] Также был проведен анализ расщепления с референсным CasX 2 дикого типа и референсным гидом 2 в сравнении с вариантами гидов 32, 64 и 174 для определения того, улучшают ли варианты расщепление. Эксперименты проводили так, как описано выше. Поскольку многие из полученных RNP не достигали полного расщепления мишени за время испытания, авторы изобретения определяли начальные скорости реакции (V0), а не константы скорости реакции первого порядка. В первых двух временных точках (15 и 30 секунд) результаты вписывались в линию для каждой комбинации CasX : огРНК и повтора. Определяли среднее значение и стандартное отклонение угла наклона для трех повторов.

[0990] В условиях анализа V0 для CasX 2 с гидами 2, 32, 64 и 174 составляли 20,4±1,4 нМ/мин, 18,4±2,4 нМ/мин, 7,8±1,8 нМ/мин и 49,3±1,4 нМ/мин (см. таблицу 16, а также ФИГ. 5 и ФИГ. 6). Гид 174 показал значительное улучшение скорости расщепления у

полученного RNP (примерно в 2,5 раза относительно гида 2, см. ФИГ. 6), а активность гидов 32 и 64 была аналогична активности гида 2 или ниже. Следует отметить, что гид 64 обеспечивает скорость расщепления ниже, чем у гида 2, но его функциональность намного лучше in vivo (данные не представлены). Вероятно, некоторые изменения последовательности, внесенные для получения гида 64, улучшают транскрипцию in vivo за счет нуклеотида, участвующего в образовании триплекса. Улучшенная экспрессия гида 64, возможно, объясняет его улучшенную активность in vivo, хотя его сниженная стабильность может приводить к неправильному сверытванию *in vitro*.

[0991] Дополнительные эксперименты проводили с гидами 174, 175, 185, 186, 196, 214 и 215 со спейсером 7.37 и CasX 491 для определения относительных скоростей расщепления. Для снижения показателей кинетики расщепления до диапазона, поддающегося измерению с помощью нашего анализа, смеси реакций расщепления инкубировали при 10 °C. Результаты представлены на ФИГ. 7 и в таблице 16. В этих условиях гид 215 был единственным, который обеспечивал более высокую скорость расщепления, чем гид 174. Гид 196, который демонстрировал самую высокую активную фракцию RNP в условиях ограничения гида, имел кинетику, по существу аналогичную кинетике гида 174; это еще раз подчеркивает, что разные варианты приводят к улучшению разных характеристик.

[0992] Данные подтверждают, что в условиях анализа применение большинства вариантов гидов с CasX приводит к получению RNP с более высоким уровнем активности, чем в случае гида дикого типа, с улучшениями в начальной скорости расщепления в диапазоне от примерно 2 раз до > 6 раз. Числовые значения в таблице 16 указывают, слева направо, вариант CasX, каркас огРНК и спейсерную последовательность конструкта RNP. В названиях конструктов RNP в таблице ниже вариант белка CasX, каркас гида и спейсер указаны в порядке слева направо.

6. Анализ расщепления *in vitro*: сравнение скорости расщепления и компетентной фракции для 515.174 и 526.174 относительно референса 2.2.

[0993] Авторы изобретения хотели сравнить сконструированные варианты белка CasX 515 и 526 в комплексе со сконструированным вариантом одномолекулярного гида 174 относительно референсного белка 2 дикого типа (SEQ ID NO: 2) и минимально модифицированного генно-инженерными методами варианта гида 2 (SEQ ID NO: 5). RNP-комплексы собирали так, как описано выше, с 1,5-кратным избытком гида. Анализы расщепления для определения $k_{расщепл.}$ и компетентной фракции выполняли так, как описано выше, в обоих случаях при 37 °C, и применяли разные временные точки для определения компетентной фракции для RNP дикого типа в сравнении со сконструированным RNP из-за значительно различающихся значений времени, необходимых для практически полного завершения реакций.

[0994] Полученные данные четко демонстрируют значительные улучшения активности RNP, когда используют как сконструированный белок, так и сконструированный гид. RNP 515.174 и 526.174 имели компетентные фракции 76% и 91%

соответственно по сравнению с 16% для RNP 2.2 (ФИГ. 8, таблица 16). В кинетическом анализе как 515.174, так и 526.174 разрезали по существу всю целевую ДНК к первой временной точке, давая показатели, превышающие разрешение анализа, и приводя к расчетным скоростям расщепления 17,10 и 19,87 мин$^{-1}$ соответственно (ФИГ. 9, таблица 16). В отличие от этого, RNP 2.2 разрезал в среднем менее 60% целевой ДНК к конечной временной точке 10 минут и имеет расчетное значение $k_{расщепл.}$ почти на два порядка величины ниже, чем сконструированные RNP. Модификации, внесенные в белок и гид, приводят к получению RNP, которые более стабильны, с большей вероятностью образовывают активные частицы, а также разрезают ДНК более эффективно при расчете показателя на одну частицу.

Таблица 16. Результаты анализа расщепления и анализа образования RNP

Конструкт RNP	**$k_{расщепл.}$***	**Начальная скорость***	**Компетентная фракция**
2.2.7.37		20,4±1,4 нМ/мин	16±3%
2.32.7.37		18,4±2,4 нМ/мин	13±3%
2.64.7.37		7,8±1,8 нМ/мин	5±2%
2.174.7.37	0,51±0,01 мин$^{-1}$	49,3±1,4 нМ/мин	22±5%
119.174.7.37	6,29±2,11 мин$^{-1}$		35±6%
457.174.7.37	3,01±0,90 мин$^{-1}$		53±7%
488.174.7.37	15,19 мин$^{-1}$		67%
491.174.7.37	16,59 мин$^{-1}$/0,293 мин$^{-1}$ (10 °C)		83%/17% (с ограниченным количеством гида)
491.175.7.37	0,089 мин$^{-1}$ (10 °C)		5% (с ограниченным количеством гида)
491.185.7.37	0,227 мин$^{-1}$ (10 °C)		44% (с ограниченным количеством гида)
491.186.7.37	0,099 мин$^{-1}$ (10 °C)		11% (с ограниченным количеством гида)
491.196.7.37	0,292 мин$^{-1}$ (10 °C)		46% (с ограниченным количеством гида)
491.214.7.37	0,284 мин$^{-1}$ (10 °C)		30% (с ограниченным количеством гида)
491.215.7.37	0,398 мин$^{-1}$ (10 °C)		38% (с ограниченным количеством гида)
515.174.7.37	17,10 мин$^{-1}$**		76%

526.174.7.37	19,87 мин$^{-1}$**		91%

* Среднее ± стандартное отклонение

** Скорость превышает разрешение анализа

Пример 9. Исследование влияния длины спейсера на кинетику расщепления *in vitro*

[0995] Исследовали активность расщепления *in vitro* рибонуклеопротеиновыми комплексами (RNP) из двух вариантов CasX и гидовой РНК со спейсерами различной длины для определения того, какая длина спейсера обеспечивает наиболее эффективное расщепление целевой нуклеиновой кислоты, и изменяется ли предпочтительная длина спейсеров для белков.

Методы:

[0996] Исследовали активность расщепления *in vitro* рибонуклеопротеиновыми комплексами (RNP) из CasX и гидовой РНК со спейсерами различной длины для определения того, какая длина спейсера обеспечивает наиболее эффективное расщепление целевой нуклеиновой кислоты.

[0997] Вариант CasX 515 и 526 очищали так, как описано выше. Гиды с каркасом 174 (SEQ ID NO: 2238) получали посредством транскрипции *in vitro* (IVT). Матрицы для IVT образовывали с помощью ПЦР с использованием полимеразы Q5 (NEB M0491) в соответствии с рекомендованным протоколом, матричных олигонуклеотидов для каждой основной цепи каркаса и праймеров для амплификации с промотором T7 и спейсером 7.37 (GGCCGAGATGTCTCGCTCCG (SEQ ID NO: 1084); нацеливающий tdTomato) длиной 20 нуклеотидов или усеченным с 3’-конца до 18 или 19 нуклеотидов. Спейсерные последовательности, а также олигонуклеотиды, применяемые для создания каждой матрицы, показаны в таблице 17. Затем полученные матрицы применяли с РНК-полимеразой T7 для получения гидовых РНК в соответствии со стандартными протоколами. Гиды очищали с помощью электрофореза в полиакриламидном геле в денатурирующих условиях и заново проводили свертывание перед применением.

[0998] RNP CasX восстанавливали посредством разбавления CasX до 1 мкМ в 1× буфере для расщепления (20 мМ Tris-HCl, pH 7,5, 150 мМ NaCl, 1 мМ TCEP, 5% глицерина, 10 мМ $MgCl_2$), добавления огРНК до 1,2 мкМ и инкубации при 37°C в течение 10 мин перед перемещением в лед до готовности к применению. Флуоресцентно-меченую целевую ДНК 7.37 приобретали в виде отдельных олигонуклеотидов в компании Integrated DNA Technologies (последовательности в таблице 17) и получали целевую дцДНК посредством нагревания эквимолярной смеси двух комплементарных цепей в 1× буфере для расщепления и медленного охлаждения до комнатной температуры.

[0999] RNP разводили в буфере для расщепления до конечной концентрации 200 нМ и инкубировали при 10°C без встряхивания. Реакции расщепления инициировали добавлением целевой ДНК 7.37 до конечной концентрации 10 нМ. Пробы отбирали во временных точках 0,25, 0,5, 1, 2, 5, 10 и 30 минут. Отобранные во временных точках пробы гасили добавлением равного объема 95% формамида с 20 мМ ЭДТА. Пробы

денатурировали нагреванием при 95°C в течение 10 минут и разделяли в 10% ПААГ с мочевиной. Для визуализации гелей использовали Amersham Typhoon и проводили анализ с помощью программного обеспечения IQTL. Полученные данные откладывали на графике и анализировали с помощью Prism. Расщепление нецелевой цепи аппроксимировали простой экспоненциальной функцией для определения кажущейся константы скорости реакции первого порядка ($k_{расщепл.}$).

Результаты:

[1000] Проводили сравнение скоростей расщепления для вариантов CasX 515 и 526 в комплексе с огРНК с 18-, 19- или 20-нуклеотидными спейсерами, чтобы определить, какая длина спейсера приводит к наиболее эффективному расщеплению в случае каждого варианта белка. В соответствии с другими экспериментами, проведенными с огРНК, полученными транскрипцией *in vitro*, наилучшим для обоих вариантов белка с точки зрения эффективности был гид с 18-нуклеотидным спейсером (ФИГ. 12A и B, таблица 18). Спейсер длиной 18 нукл. обеспечивал в 1,4 раза более высокую скорость, чем спейсер длиной 20 нукл. в случае белка 515, и в 3 раза более высокую скорость, чем спейсер длиной 20 нукл. в случае белка 526. Активность спейсера длиной 19 нукл. была промежуточной в случае обоих белков, хотя, опять же, различие было более выраженным для варианта 526. В целом, наблюдалось, что спейсеры короче 20 нукл. обладают повышенной активностью при использовании целого диапазона белков, спейсеров и способов доставки, но степень улучшения и оптимальная длина спейсера варьировались. Эти данные показывают, что два сконструированных белка, которые являются достаточно похожими по последовательности (отличаются только двумя остатками), могут претерпевать изменения активности, вызванные длиной спейсера, которые сходны по направлению, но существенно различаются по степени.

Таблица 17. Значимые последовательности и олигонуклеотиды

Нецелевая цепь целевой последовательности 7.37	IR700-TGAAGCTGACAGCATTCGGGCCGAGATGTCTCGCTCCGTGGCCTTAGCTGTGCTCGCGCT (SEQ ID NO:1069)
Целевая цепь целевой последовательности 7.37	IR800-AGCGCGAGCACAGCTAAGGCCACGGAGCGAGACATCTCGGCCCGAATGCTGTCAGCTTCA (SEQ ID NO: 1068)
Спейсерная последовательность длиной 20 нукл.	GGCCGAGATGTCTCGCTCCG (SEQ ID NO:1085)
Спейсерная последовательность длиной 18 нукл.	GGCCGAGATGTCTCGCTC (SEQ ID NO: 1086)
Спейсерная последовательность длиной 19 нукл.	GGCCGAGATGTCTCGCTCC (SEQ ID NO: 1087)
Матричный прямой для каркаса 174	GAAATTAATACGACTCACTATAACTGGCGCTTTTATCTGATTACTTTGAGAGCCATCACCAGCGACTATG

	TCGTAGTGGGTAAAGCT (SEQ ID NO: 1088)
Матричный обратный для каркаса 174	CTTTGATGCTCCCTCCGAAGAGGGAGCTTTACCCA CTACGACATAGTCGC (SEQ ID NO: 1089)
Праймер для амплификации T7	GAAATTAATACGACTCACTATA (SEQ ID NO: 169)
Праймер для каркаса 174 со спейсером длиной 20 нукл.	CGGAGCGAGACATCTCGGCCCTTTGATGCTCCCTC C (SEQ ID NO: 1090)
Праймер для каркаса 174 со спейсером длиной 18 нукл.	GAGCGAGACATCTCGGCCCTTTGATGCTCCCTCC (SEQ ID NO: 1091)
Праймер для каркаса 174 со спейсером длиной 19 нукл.	GGAGCGAGACATCTCGGCCCTTTGATGCTCCCTCC (SEQ ID NO: 1092)

Таблица 18. Скорости расщепления RNP с укороченными спейсерами

Длина спейсера	515 $k_{-расщепл.}$ ($мин^{-1}$)	526 $k_{-расщепл.}$ ($мин^{-1}$)
18	0,215	0,427
19	0,182	0,282
20	0,150	0,143

Пример 10. Оценка аффинности связывания с гидовой РНК

[1001] Очищенный CasX дикого типа и улучшенный CasX будут инкубировать с синтетической одномолекулярной гидовой РНК, содержащей Cy7.5-фрагмент на 3’-конце, в низкосолевом буферном растворе, содержащем хлорид магния и гепарин для предотвращения неспецифического связывания и агрегации. Концентрация огРНК будет постоянной - 10 пМ, в то время как белок будут титровать от 1 пМ до 100 мкМ в отдельных реакциях связывания. После того как реакция придет в равновесное состояние, пробы пропустят через вакуумный коллектор для анализа связывания с фильтром, используя нитроцеллюлозную мембрану и положительно заряженную нейлоновую мембрану, которые связывают белок и нуклеиновую кислоту соответственно. На мембранах проведут визуализацию для идентификации гидовой РНК, и будут определены фракции связанной и несвязанной РНК по количеству флуоресценции на нитроцеллюлозной и нейлоновой мембране при каждой концентрации белка для расчета константы диссоциации комплекса белок-огРНК. Также будет проведен эксперимент с улучшенными вариантами огРНК для определения того, влияют ли эти мутации на аффинность к гиду белков дикого типа и мутантных белков. Авторы изобретения также выполнят анализ методом сдвига электрофоретической подвижности для качественного сравнения с результатами анализа связывания с фильтром и подтверждения того, что основным фактором ассоциации белков и РНК является связывание в растворе, а не агрегация.

Пример 11. Оценка аффинности связывания с целевой ДНК

[1002] Очищенный CasX дикого типа и улучшенный CasX будут объединены в комплекс с одномолекулярной гидовой РНК, несущей нацеливающую последовательность, комплементарную целевой нуклеиновой кислоте. RNP-комплекс будут инкубировать с двухцепочечной целевой ДНК, содержащей PAM и соответствующую последовательность целевой нуклеиновой кислоты с Cy7.5-меткой на 5’-конце целевой цепи, в низкосолевом буферном растворе, содержащем хлорид магния и гепарин для предотвращения

неспецифического связывания и агрегации. Концентрация целевой ДНК будет постоянной - 1 нМ, в то время как RNP будут титровать от 1 пМ до 100 мкМ в отдельных реакциях связывания. После того как реакция придет в равновесное состояние, пробы пропустят через нативный 5% полиакриламидный гель для отделения связанной и несвязанной целевой ДНК. Будет проведена визуализация геля для определения сдвигов подвижности целевой ДНК и будут рассчитаны фракции связанной и несвязанной ДНК при каждой концентрации белка для определения константы диссоциации тройного комплекса RNP-целевая ДНК. Ожидается, что эксперименты продемонстрируют улучшенную аффинность связывания RNP, содержащего вариант CasX и вариант гРНК, по сравнению с RNP, содержащим референсный CasX и референсную гРНК.

Пример 12. Оценка улучшенных характеристик экспрессии и растворимости вариантов CasX для получения RNP

[1003] CasX дикого типа и модифицированные варианты CasX будут экспрессированы в BL21 (DE3) *E. coli* в идентичных условиях. Все белки будут находиться под контролем IPTG-индуцибельного промотора T7. Клетки будут выращивать до OD 0,6 в среде ТВ при 37 °C, после чего температура выращивания будет снижена до 16°C и экспрессия будет индуцирована добавлением 0,5 мМ IPTG. Клетки будут собирать после 18 часов экспрессии. Растворимые белковые фракции будут экстрагированы и проанализированы в геле для ДСН-ПААГ. Относительные уровни экспрессии растворимого CasX будут определять посредством окрашивания Кумасси. Белки будут очищены параллельно в соответствии с протоколом, приведенным выше, и будет проведено сравнение конечного выхода очищенного белка. Для определения растворимости очищенного белка конструкты будут концентрировать в буфере для хранения вплоть до начала осаждения белка. Осажденный белок будет удален центрифугированием и будет измерена конечная концентрация растворимого белка для определения максимальной растворимости каждого варианта. Наконец, будут получены комплексы вариантов CasX с одномолекулярной гидовой РНК, которые будут сконцентрированы вплоть до начала образования осадка. Осажденные RNP будут удалены центрифугированием и будет измерена конечная концентрация растворимых RNP для определения максимальной растворимости каждого варианта, связанного с гидовой РНК.

Пример 13. Эволюция платформы гидового каркаса гидовых РНК

[1004] Были проведены эксперименты для идентификации вариантов гидовых каркасов гидовых РНК, которые демонстрируют улучшенную активность расщепления двухцепочечной ДНК (дцДНК). Для обеспечения этого была разработана масштабная библиотека вариантов каркаса, и объединенная в пул библиотека была протестирована на способность вызывать функциональный нокаут репортерного гена в человеческих клетках. Варианты каркаса, приводящие к улучшенному нокауту, определяли посредством секвенирования функциональных элементов в пуле и последующего компьютерного анализа.

Материалы и методы

Составление библиотеки

Оценка стабильности вторичной структуры РНК

[1005] RNAfold (v 2.4.14) (Lorzz R, et al. ViennaRNA Package 2.0. Algorithms Mol Biol.6:26 (2011)) использовали для прогнозирования стабильности вторичной структуры последовательностей РНК, аналогично тому, как это было выполнено в Jarmoskaite I., et al. A quantitative and predictive model for RNA binding by human pumilio proteins. Mol Cell. 74(5):966 (2019). Для оценки значения ΔΔG_BC рассчитывали свободную энергию ансамбля (ΔG) для неограниченного ансамбля, затем рассчитывали свободную энергию ансамбля (ΔG) для *ограниченного* ансамбля. ΔΔG_BC представляет собой разность между значениями ΔG ограниченного и неограниченного ансамблей. Использовали ограничивающую строку, которая отражает спаривание оснований стебля псевдоузла, каркасного стебля и удлиненного стебля и требует, чтобы основания триплекса были неспаренными.

Расчет стабильности вторичной структуры стебля псевдоузла

[1006] Стабильность структуры псевдоузла рассчитывали для положений 3-33, охватывающих весь стебель и петлю, используя последовательность петли триплекса из гидового каркаса 175. Кроме того, была создана ограничивающая строка, которая принудительно обеспечивала спаривание оснований псевдоузла и отсутствие спаривания оснований в петле триплекса. Таким образом, изменения стабильности могли быть обусловлены только различиями в последовательности стебля псевдоузла. Например, последовательность псевдоузла AAAACG_CGTTTT была превращена в последовательность стебля и петли посредством вставки последовательности петли триплекса CUUUAUCUCAUUACUUUGA (SEQ ID NO: 158) таким образом, чтобы конечная последовательность была AAAACG**CUUUAUCUCAUUACUUUGA**CGTTTT (SEQ ID NO: 159), а ограничивающая строка была: ‘((((((xxxxxxxxxxxxxxxxxxx))))))’ (SEQ ID NO: 160, где x=n).

Молекулярная биология

Молекулярная биология процесса создания библиотеки

[1007] Разработанная библиотека вариантов каркаса гидовой РНК была синтезирована и получена из Twist Biosciences, а затем ее амплифицировали с помощью ПЦР с праймерами, специфичными к библиотеке. Эти праймеры амплифицируют дополнительную последовательность на 5’- и 3’-концах библиотеки для введения сайтов распознавания последовательности для фермента рестрикции SapI. ПЦР проводили с помощью ДНК-полимеразы Q5 (New England Biolabs) в соответствии с инструкциями производителя. Типичными условиями ПЦР были следующие: 10 нг ДНК матричной библиотеки, 1× буфер для ДНК-полимеразы Q5, 300 нМ дНТФ, 300 нМ каждого праймера, 0,25 мкл ДНК-полимеразы Q5 в реакционной смеси объемом 50 мкл. Типичная программа термоциклера будет представлять собой: цикл при 95°C в течение 5 мин; затем 20 циклов при 98°C в течение 15 с, 65°C в течение 20 с, 72°C в течение 1 мин; с заключительным добавочным периодом 2 мин при 72 °C. Амплифицированный продукт ДНК очищали с

помощью набора для очистки и концентрирования ДНК DNA Clean and Concentrator kit (Zymo Research). Затем этот ПЦР-ампликон, а также плазмиду pKB4 расщепляли ферментом рестрикции SapI (New England Biolabs), и оба фрагмента независимо очищали электрофорезом в агарозном геле с последующей экстракцией из геля (Zymo) в соответствии с инструкциями производителя. После этого библиотеки лигировали с использованием ДНК-лигазы T4 (New England Biolabs), очищали с помощью набора DNA Clean and Concentrator kit (Zymo) и трансформировали ими электрокомпетентные клетки MegaX DH10B T1R Electrocomp Cells (ThermoFisher Scientific); все процедуры проводили в соответствии с инструкциями производителя. Трансформированные библиотеки восстанавливали в течение одного часа в среде SOC, затем выращивали в течение ночи при 37 ºC при встряхивании в 5 мл среды 2xyt. После этого проводили мини-выделение плазмидной ДНК из культур (QIAGEN). Затем плазмидную ДНК дополнительно клонировали посредством расщепления ферментом рестрикции Esp3I (New England Biolabs) с последующим лигированием с отожженными олигонуклеотидами, имеющими комплементарные липкие концы из одноцепочечной ДНК и несущими требуемую спейсерную последовательность для нацеливания на GFP. Олигонуклеотиды имели модификации в виде фосфорилирования 5'-концов, и их отжигали путем нагревания до 95°C в течение 1 мин с последующим снижением температуры со скоростью два градуса в минуту до достижения конечной температуры 25 °C. Лигирование проводили в виде реакции сборки Golden Gate, где типичные условия реакции были следующими: 1 мкг предварительно расщепленной библиотеки плазмид, 1 мкМ отожженных олигонуклеотидов, 2 мкл ДНК-лигазы T4, 2 мкл Esp3I и 1× буфер для ДНК-лигазы T4 в общем объеме водного раствора 40 мкл. Реакционную смесь обрабатывали в термоциклере: 25 циклов чередования 37°C в течение 3 минут и 16°C в течение 5 минут. Как описано выше, библиотеку очищали, трансформировали ей клетки, клетки выращивали в течение ночи и проводили мини-выделение. Затем полученную библиотеку плазмид использовали для получения лентивируса.

Скрининг библиотеки

Получение лентивируса (LV)

[1008] Лентивирусные частицы получали посредством трансфекции клеток LentiX HEK293T, посеянных предварительно за 24 ч, при конфлюэнтности 70-90%. Плазмиды, содержащие объединенную библиотеку, вводили в лентивирусную векторную систему второго поколения, содержащую упаковочную плазмиду и оболочечную плазмиду VSV-G, с полиэтиленимином в бессывороточной среде. Для получения частиц среду меняют через 12 часов после трансфекции, а вирусы собирают через 36-48 ч после трансфекции. Содержащий вирусы супернатант фильтровали с использованием полиэфирсульфоновых (PES) мембранных фильтров с диаметром пор 0,45 мкм и разбавляли в культуральной среде (при необходимости) перед добавлением к клеткам-мишеням.

[1009] Через 72 часа после фильтрации аликвоты содержащего лентивирусы супернатанта титровали с помощью количественной ПЦР TaqMan. Вирусную геномную

РНК выделяли с использованием экстракции смесью фенола и хлороформа (TRIzol) с последующим осаждением спиртом. Качественную и количественную оценку экстракции выполняли с помощью спектрофотометра NanoDrop. Затем всю остаточную плазмидную ДНК расщепляли с помощью ДНКазы I непосредственно перед получением кДНК с помощью обратной транскриптазы ThermoFischer SuperScript IV. Вирусную кДНК подвергали последовательным разведениям вплоть до 1 : 1000 и объединяли с праймерами на основе WPRE и мастер-миксом TaqMan перед проведением количественной ПЦР на приборе Bio-Rad CFX96. Все разведения проб добавляют в двух повторностях и усредняют показания перед расчетом титра по стандартной кривой, построенной на основе известных количеств плазмиды. Во всех случаях в качестве отрицательного контроля для измерений используют воду.

Скрининг LV (трансдукция, поддержание, гейтирование, сортировка, выделение гДНК)

[1010] Проводят пассаж целевых репортерных клеток за 24-48 часов перед трансдукцией, чтобы обеспечить наличие клеточного деления. В точке проведения трансдукции клетки трипсинизировали, подсчитывали и разбавляли до соответствующей плотности. Клетки ресуспендировали в не подвергавшемся обработке профильтрованном супернатанте, содержащем лентивирусы с библиотекой или контролем, при низком значении множественности заражения (MOI) (0,1-5, по вирусному геному) для сведения к минимуму двойных интеграций лентивируса. Смеси лентивирусов с клетками высевали при конфлюэнтности 40-60% перед инкубацией при 37 °C, 5% CO_2. Успешно трансдуцированные клетки отбирали через 48 ч после трансдукции с помощью пуромицина в концентрации 1-3 мкг/мл в течение 4-6 дней с последующим восстановлением в среде HEK или Fb.

[1011] После отбора клетки суспендировали в 4',6-диамидино-2-фенилиндоле (DAPI) и фосфатно-солевом буфере (PBS). Затем клетки фильтровали с помощью пробирки для FACS с крышкой с нейлоновой сеткой компании Corning (кат. № 352235) и сортировали на приборе Sony MA900. Клетки сортировали по наличию нокдауна флуоресцентного репортера, в дополнение к гейтированию на одиночные живые клетки посредством стандартных способов. Отсортированные клетки из эксперимента лизировали, и экстрагировали геном с использованием набора Zymo Quick-DNA Miniprep Plus в соответствии с протоколом производителя.

Обработка для секвенирования нового поколения (NGS)

[1012] Геномную ДНК амплифицировали методом ПЦР с праймерами, специфичными к ДНК, кодирующей гидовую РНК, с образованием целевого ампликона. Эти праймеры содержат дополнительную последовательность на 5'-концах для введения прочтения Illumina и 2 последовательностей. Типичными условиями ПЦР были следующие: 2 мкг гДНК, 1× буфер Kapa HiFi, 300 нМ дНТФ, 300 нМ каждого праймера, 0,75 мкл ДНК-полимеразы Kapa HiFi Hotstart в 50 мкл реакционной смеси. В термоциклере: цикл при 95°C в течение 5 мин; затем 15 циклов при 98°C в течение 15 с, 62°C в течение 20 с, 72°C в

течение 1 мин; с заключительным добавочным периодом 2 мин при 72 °С. Амплифицированный продукт ДНК очищают с помощью набора для очистки ДНК Ampure XP DNA cleanup kit. Вторую стадию ПЦР проводили с индексирующими адаптерами для обеспечения мультиплексирования на платформе Illumina. Объединяли 20 мкл очищенного продукта с предыдущей стадии с 1× буфером Kapa GC, 300 нМ дНТФ, 200 нМ каждого праймера, 0,75 мкл ДНК-полимеразы Kapa HiFi Hotstart в 50 мкл реакционной смеси. В термоциклере: цикл при 95°С в течение 5 мин; затем 5-16 циклов при 98°С в течение 15 с, 65°С в течение 15 с, 72°С в течение 30 с; с заключительным добавочным периодом 2 мин при 72 °С. Амплифицированный продукт ДНК очищают с помощью набора для очистки ДНК Ampure XP DNA cleanup kit. Качественную и количественную оценку ампликона проводили с помощью набора Fragment Analyzer DNA analyzer kit (Agilent, дцДНК 35-1500 п. о.). Ампликоны секвенировали на Illumina Miseq (v3, 150 циклов одноконцевого секвенирования) в соответствии с инструкциями производителя.

Анализ NGS (обработка проб и анализ данных)

[1013] В прочтениях были обрезаны адаптерные последовательности с помощью Cutadapt (версия 2.1), и из каждого прочтения была экстрагирована гидовая последовательность (содержащая последовательность каркаса и спейсерную последовательность) (также используя адаптеры, связанные с cutadapt v 2.1, для экстракции последовательности между вышележащей и нижележащей последовательностями ампликона). Были подсчитаны уникальные последовательности гидовых РНК, а затем каждую последовательность каркаса сравнивали со списком сконструированных последовательностей и последовательностью гидовых каркасов 174 (SEQ ID NO: 2238) и 175 (SEQ ID NO: 2239) для определения идентичности каждой.

[1014] Число прочтений для каждой уникальной последовательности гидовой РНК нормализовали по глубине секвенирования с использованием нормализации средним. Обогащение рассчитывали для каждой последовательности путем деления нормализованного числа прочтений в каждой GFP-отрицательной пробе на нормализованное число прочтений в ассоциированной наивной пробе. Для обоих выбранных вариантов (R2 и R4) GFP-отрицательную и наивную популяции обрабатывали для NGS в три отдельных дня, получая значение обогащения для каждого каркаса в трех повторностях. Суммарную оценку обогащения на каркас рассчитывали после суммирования числа прочтений для наивных и GFP-отрицательных проб в трех повторностях.

[1015] Две оценки обогащения из разных выбранных вариантов объединяли в средневзвешенное значение индивидуальных оценок обогащения $\log_2$, взвешенных по их относительной представленности в наивной популяции.

[1016] Для расчета ошибки для оценки обогащения $\log_2$ рассчитывали 95%-й доверительный интервал средней оценки обогащения трех повторных проб. Эти ошибки распространяются при комбинировании значений обогащения для двух отдельных выбранных вариантов.

Результаты и обсуждение

Составление библиотеки, заказ и клонирование

[1017] Библиотека вариантов гидовых РНК была разработана для тестирования изменения РНК-каркаса как с помощью несмещенного подхода, так и с помощью целевого подхода, который сфокусирован на ключевых модулях в каркасе РНК.

[1018] В несмещенной части библиотеки все однонуклеотидные замены, вставки и делеции были разработаны для каждого остатка гидовых каркасов 174 (SEQ ID NO: 2238) и 175 (SEQ ID NO: 2239) (~ 2800 отдельных последовательностей). Двойные мутанты были разработаны для того, чтобы специфически сконцентрироваться на областях, которые, возможно, могут взаимодействовать; таким образом, если в структуре CryoEM (PDBid: 6NY2) два остатка были вовлечены в каноническое или неканоническое взаимодействие спаривания оснований или было спрогнозировано, что два остатка образуют пару в структуре с наименьшей энергией, прогнозируемой с помощью RNAfold (вер. 2.4.14), то соответствующие остатки в гидовых каркасах 174 и 174 были подвергнуты мутации (включая все возможные замены, вставки и делеции обоих остатков). Также проводили мутации остатков, смежных с этими «взаимодействующими» остатками; однако в этих случаях были включены только замены каждого из двух остатков. В окончательной библиотеке были разработаны ~27 тыс. последовательностей с двумя мутациями относительно гидовых каркасов 174 или 175.

[1019] В части библиотеки, предназначенной для специфического мутагенеза ключевых областей каркаса РНК, модификации были разработаны для области псевдоузла, триплексной области, пузырька каркаса и удлиненного стебля (см. ФИГ. 18 для идентификации областей). В каждой из этих целевых секций библиотеки весь домен был подвергнут мутации с помощью подхода, управляемого гипотезами (ФИГ. 19). Например, для триплексной области каждый из триплетов оснований, составляющих триплекс, был подвергнут мутации с образованием другого формирующего триплекс мотива (см. ФИГ. 20). Этот тип мутагенеза отличается от того, который использовали в пузырьке каркасного стебля, в котором при мутировании были выполнены все возможные замены оснований, окружающих пузырек (т. е. до 5 мутаций относительно гидовых последовательностей 174 или 175). В противоположность этому 5 пар оснований, составляющих стебель псевдоузла, были полностью заменены на альтернативную последовательность со спариванием оснований Уотсона - Крика (мутагенизированы до 10 отдельных оснований).

[1020] Окончательная целевая секция библиотеки была предназначена для оптимизации последовательностей, которые с более высокой вероятностью образуют вторичные структуры, пригодные для связывания белка. Вкратце, стабильность вторичной структуры последовательности прогнозировалась при двух условиях: 1) в отсутствие каких-либо ограничений, 2) при ограничениях, которые гарантировали образование ключевых элементов вторичной структуры, таких как стебель псевдоузла, каркасный стебель и удлиненный стебель (см. «Материалы и методы»). Наша гипотеза заключалась в том, что разница в стабильности между этими двумя условиями (называемая в данном случае

ΔΔG_BC) будет минимальной для последовательностей, которые более пригодны для связывания белка, и, таким образом, следует искать последовательности, в которых эта разница является минимальной.

[1021] Разработанная библиотека была заказана в компании Twist (~ 40 тыс. отдельных последовательностей) и синтезирована с включением сайтов Golden Gate для клонирования в основную цепь лентивирусной плазмиды, которая также экспрессировала белок STX119 (см. «Материалы и методы»). Спейсерную последовательность, нацеливающую на ген GFP, клонировали в вектор библиотеки, эффективно создающий одномолекулярные гидовые РНК из каждого варианта каркаса РНК для нацеливания на ген GFP. Представительность сконструированных вариантов библиотеки оценивали с помощью секвенирования нового поколения (см. «Материалы и методы»).

Скрининг и оценка библиотеки

[1022] Библиотека плазмид, содержащая варианты гидовых РНК и один белок CasX (версия 119), была изготовлена в лентивирусных частицах (см. «Материалы и методы»); частицы титровали на основании числа копий вирусных геномов с использованием анализа количественной ПЦР (см. «Материалы и методы»). Клеточную линию, стабильно экспрессирующую GFP, трансдуцировали библиотекой лентивирусных частиц при низкой множественности заражения (MOI) для обеспечения того, чтобы в каждую клетку был интегрирован не более чем один член библиотеки. Был отобран пул клеток, содержащий только те клетки, в которых произошла интеграция в геном. Наконец, клеточная популяция была отсортирована по наличию экспрессии GFP, и была получена популяция GFP-отрицательных клеток. Такие GFP-отрицательные клетки содержали членов библиотеки, которые эффективно нацеливали CasX RNP на ген белка GFP, вызывая индел и последующую утрату функции.

[1023] Геномную ДНК из несортированной популяции клеток («наивные») и GFP-отрицательной популяции обрабатывали для выделения последовательности членов библиотеки гидовых РНК в каждой клетке. Для определения представительности гидовых РНК в наивной и GFP-отрицательной популяциях выполняли секвенирование нового поколения. Оценки обогащения рассчитывали для каждого члена библиотеки путем деления представленности члена библиотеки в GFP-отрицательной популяции на его представленность в наивной популяции: высокая оценка обогащения указывает на то, что член библиотеки гораздо чаще присутствует в активной GFP-отрицательной популяции, чем в исходном пуле, и, таким образом, является активным вариантом, способным эффективно создавать индел в гене GFP (значение обогащения > 1, $\log_2$ обогащения > 0). Низкая оценка обогащения указывает на то, что член библиотеки мало представлен в активной GFP-отрицательной популяции по сравнению с наивной популяцией и, таким образом, неэффективен для создания индела (значение обогащения < 1, $\log_2$ обогащения < 0). В качестве заключительной статистической оценки сравнения рассчитывали относительное значение обогащения как обогащение члена библиотеки (в GFP-отрицательной популяции по сравнению с наивной популяцией), деленное на обогащение

референсной последовательности каркаса (в GFP-отрицательной популяции по сравнению с наивной популяцией). (При представлении в логарифмическом виде эти значения просто вычитаются.) Значения обогащения для референсных последовательностей каркаса показаны на ФИГ. 21.)

[1024] Скрининг выполняли множество раз с независимым получением лентивирусных частиц, трансдукцией клеток, отбором и сортировкой для получения наивных и GFP-отрицательных популяций и секвенированием для выяснения значений обогащения для каждого члена библиотеки. Эти скрининги были названы R2 и R4, и их результаты в значительной степени воспроизводили значения обогащения, полученные для однонуклеотидных вариантов на гидовых каркасах 174 и 175 (ФИГ. 22). Скрининг позволял идентифицировать множество возможных комбинаций мутаций, которыми была обогащена функциональная GFP-отрицательная популяция, и, следовательно, эти мутации могут приводить к функциональным RNP. Напротив, не наблюдалось обогащение гидами, которые содержали ненацеливающие спейсеры, что подтверждает то, что обогащение является селективным фактором отсечения (данные не представлены). Полный набор мутаций на гидовых каркасах 174 и 175, для которых наблюдалось обогащение, приведен в таблицах 19 и 20 соответственно. Эти списки показывают разнообразие последовательностей, которые сохраняют способность приводить к нацеленным функциональным RNP.

Однонуклеотидные мутации указывают на мутабельные области каркаса.

[1025] Для определения мутаций каркаса, которые приводят к аналогичной или улучшенной активности относительно гидовых каркасов 174 и 175, значения обогащения для однонуклеотидных замен, вставок или делеций наносили на график (ФИГ. 23). Как правило, однонуклеотидные изменения в 174 были более переносимыми, чем в 175, что, возможно, отражает более высокую активность 174 в этом контексте и, следовательно, более высокую переносимость мутаций, которые снижают активность (ФИГ. 21 и ФИГ. 24). Однонуклеотидные мутации, которые были благоприятны в 175, в подавляющем большинстве случаев также были благоприятны и в контексте 174 (ФИГ. 24), поэтому было принято, что значения для мутаций в каркасе 175 являются более строгим показателем эффектов мутаций. По результатам этого анализа были определены ключевые мутабельные области, как описано в следующих параграфах.

[1026] Наиболее заметной особенностью был удлиненный стебель, который демонстрировал значения обогащения аналогичные таковым у референсных последовательностей 174 или 175, что позволяет предположить, что каркас может переносить изменения в этой области, аналогично тому, что наблюдали в прошлом и что было спрогнозировано в структурном анализе CasX RNP, который показал, что удлиненный стебель имеет мало контактов с белком.

[1027] Триплексная петля была еще одной областью, которая демонстрировала высокое обогащение относительно референсного каркаса, в особенности при внесении мутаций в гидовый каркас 175 (например, в частности, мутаций по C15 или C17).

Примечательно, что положение C17 в 175 уже было подвергнуто мутации на G в каркасе 174, что представляет собой одну из двух мутаций с высоким обогащением по этому положению для каркаса 175.

[1028] Изменения любого члена предполагаемой пары в стебле псевдоузла - между G7 и A29 - в обоих случаях приводили к высокому обогащению относительно референсной молекулы, в особенности в гидовом каркасе 175. Эта пара представляет собой неканоническую пару G:A в обоих гидовых каркасах 174 и 175. Приводящая к наибольшему обогащению мутация в этих положениях находилась в гидовом каркасе 175: замена A29 на C или T; первая из которых приводит к образованию канонической пары Уотсона - Крика (G7:C29), а вторая из которых приводит к образованию неоднозначной пары GU (G7:U29), и можно ожидать, что обе замены повысят стабильность спирали относительно пары G:A. Высокое обогащение также было характерно для замены G7 на T, которая приводит к образованию канонической пары (U7:A29) в этом положении. Очевидно, что по этим положениям благоприятным является более стабильное спаривание. В целом, 5'-конец был мутабельным, при этом лишь немногие изменения приводили к уменьшению обогащения.

[1029] Наконец, вставка C в положении 54 в гидовом каркасе 175 приводила к высокому обогащению, а делеция A или вставленного G в аналогичном положении в гидовом каркасе 174 в обоих случаях давала значения обогащения, аналогичные наблюдавшимся для референсных молекул. В совокупности для гидового каркаса может быть предпочтительно наличие двух нуклеотидов в этом пузырьке каркасного стебля, но это предпочтение может не быть сильным. Эти результаты дополнительно рассмотрены в разделах ниже.

Стабильность стебля псевдоузла связана с активностью каркаса

[1030] Для дальнейшего изучения влияния стебля псевдоузла на активность каркаса стебель псевдоузла был модифицирован следующим образом: (1) пары оснований в стебле были переставлены таким образом, чтобы каждый новый псевдоузел имел тот же самый состав пар оснований, но расположенных в стебле в другом порядке; (2) пары оснований были полностью заменены случайной последовательностью, образующей пары Уотсона - Крика. Протестировали 291 стебель псевдоузла. Анализ первого набора последовательностей демонстрирует сильное предпочтение пары G-A в первом положении стебля псевдоузла относительно других возможных положений (2-6; в последовательности дикого типа она находится в положении 5; ФИГ. 25), при этом результаты показывают, что наличие пары G-A в каждом из положений 2-6 в стебле псевдоузла обычно оказывает неблагоприятное действие, давая низкое среднее обогащение. Наличие оснований G-A в положении 1, вероятно, стабилизирует стебель псевдоузла, позволяя остальной части спирали образовываться только из пар Уотсона - Крика со стэкинг-взаимодействиями. Этот результат дополнительно подтверждает, что для каркаса предпочтителен полностью спаренный стебель псевдоузла.

[1031] Значительное число последовательностей псевдоузлов имело положительное значение $\log_2$ обогащения, что позволяет предположить, что замена этой

последовательности альтернативными парами оснований была по существу переносимой (структура псевдоузла на ФИГ. 26). Для дальнейшей проверки гипотезы о том, что более стабильная спираль в стебле псевдоузла приведет к получению более активного каркаса, рассчитывали стабильность вторичной структуры каждого стебля псевдоузла («Материалы и методы»). Наблюдалась сильная связь между стабильностью псевдоузла и обогащением, и, следовательно, активностью (ФИГ. 27: более активные каркасы имеют стабильные стебли псевдоузла), причем гидовые каркасы со стабильными стеблями псевдоузла (< -7 ккал/моль) имели высокое обогащение, а гидовые каркасы с дестабилизированными стеблями псевдоузла (≥ -3 ккал/моль) имели очень низкое обогащение.

Двойные мутации указывают на мутабельные области гидового каркаса.

[1032] Было проведено изучение двойных мутаций в каждом референсном гидовом каркасе, чтобы дополнительно идентифицировать мутабельные области внутри каркаса и потенциальные мутации для улучшения активности каркаса. Сфокусировавшись только на одной паре положений - положениях 7 и 29, которые, как предполагается, образуют неканоническую пару G:A в стебле псевдоузла и благоприятствуют мутагенезу (см. разделы выше) - мы можем нанести на диаграмму все 64 двойных мутаций для этой пары положений (ФИГ. 28). Для этих двух положений предпочтительными являются канонические пары (например, замена на C в положении 7 и G в положении 29 создает пару G:C, характеризующуюся обогащением; замена на C в положении 7 и вставка G в положении 29 аналогично создают пару G:C, замена на A в положении 7 и U в положении 29 создает пару A:U). Ни одна пара вставок не характеризовалась обогащением, возможно, из-за того, что вставка канонической пары в данном положении не является достаточной для стабилизации спирали, учитывая, что пара G:A смещается в спирали вверх на одну позицию, а не удаляется полностью. Неожиданно было обнаружено, что несколько двойных мутаций, характеризующихся обогащением, не образовывали канонические пары; например, замены на U в положении 7 и C в положении 29 (которые образуют неканоническую пару U:C), замены на U в положении 7 и U в положении 29 (образуют пару U:U), а также несколько других (ФИГ. 28). Возможно, пара пурин:пурин по существу в большей степени разрушает спираль, чем другие неканонические пары. Действительно, замена на A в положении 7 и G в положении 29 снова образует пару A:G, которая в этом положении не вызывает обогащения.

[1033] Значения обогащения для двойных замен в каждом из ключевых структурных элементов гидового каркаса 175 определяли по тепловым картам, на которых каждое положение могло иметь до трех замен. Было определено, что каркасный стебель меньше всего переносит мутации, что позволяет предположить, что в этой области находится сильно ограниченная последовательность.

[1034] Результаты демонстрируют, что в гидовый каркас могут быть внесены существенные изменения, которые все еще могут привести к нокауту функционального гена при использовании в анализе редактирования. В частности, результаты демонстрируют ключевые положения, которые можно использовать для улучшения

активности посредством модификаций в гидовом каркасе, включая повышение стабильности вторичной структуры стебля псевдоузла внутри каркаса.

Таблица 19. Мутации гида 174 мутации и полученное в результате относительное обогащение

Log_2 обогащения	Мутации в каркасе гРНК 174* (SEQ ID NO: 2238)
От 3,25 до 3,5	G79A, A80G; T34A, G78T; G7T, G75A; G78A, A80T; ^C2,A33T; ^A1,C68T; TG3CT, CGC6TAG, GAG28CTA, CA32AG; TG3CA, GC7AA, GA28TT, CA32TG; ^C4,C6G, T12_,G17C, GAG28CCC, C32G, A80C; T9C, T14A, T71A, C73A; C70A, G77T
От 3,0 до 3,25	A29T, G78T; T9C, G17C, A27T, G79_; C2G, A21G; ^A81,^C81; T71A, C73A; T14C, T16G; ^T64,^G81; T9C, G17C,^TG65; C2G, T16A; G7C, TC14AT, G17A, T34A; G75A, G77A; G7C, A21T; T-.3.CA, GC.7.-T, G28_,-A.33.TG,^T84; T65C, C82T; GCTCCC63_,^AATGAAAA70,^TTTTCATT76,GGGAGC77_; ^C2,G7A, A27T; T9C, G17C, C67G; ^A78,^T78; T3C, GCG5AGA, AGC29GCT, A33G; T9C, G17C, G78C; T3C, GC5TG, AGC29CAA, A33G; T9A,^T68,G77A; G7A, T9G; T65A,^G77; ^G70,^C75; C2T, G79C; ^C66,G78A; A29C, G75A; C15A, A60G; C67G,^A78; T14C, G17T, G40A, A76G; T34A, CT64TC; ^A69,T69A; T45G, G79T; T69C,^C76; C2A, G54C; A13C, C15A, G74C; C70G,^A75; A76G, G77C; C67T, G78C; TG3CC, A29C, CA32GG; ^T7,A29C; C2A, T34A; ^A66,^A66; C66T, A80C; ^G17; ^C76,^A76; A29C; C15G, C67G, T72G; ^T70,^A70; C15G, T16G; C64T, C66A; T69G, G74C; ^A3,G74C; ^T65,^T80
От 2,9 до 3,0	A29C, A33T; C64T, G78A; ^C64,A80T; ^A74; T65A,^A80; ^T69,G75T; ^C79,^A79; A29G, T59A; T69G, G75C, G78A; ^G70,^A70; G7A, TC14CG, G17A, C64T; ^T69,^A76; T9C, G17C, C68T, T72C; ^T69,A76G; A33T, C66G; C66T, C67G; TTC71ACA,^GGATGT75; A13G, T14A; T69A, G74C; G74T, A76G; G77C, G78A; A27C, T84G; C2_,C66G; T71C, G75C; TC14AG, G78A; T3G, A33T; T9C, G28A; ^A1,C2T; C68T, T72C; TGGC3CCAG, C8A, GA26_,--A.33.TGG; C64T, C66G; ^A67,C67G; C68T, G74A, G77C; G7T; C2T, G78T; C68_,G77T; T25C, A29C; ^A78,G78A; ^C78,G78C; G7C, A60G; T34A, T45A; T3_,G7A,^A9,^T28,A29G, A33_; ^CAG70,T72G, G--.74.AGT; A27G, A29C; T9C, G17C, T47C; ^T19; ^A65,^T65; C67G, C68T
От 2,8 до 2,9	T3C, G5T, C8G, GA28CC, CCA31AAG; T69C, A76G; C66T, A80T; ^G13; C2_,T65G; G7C, T9G; T9C, G17C, TT71AC; C6G, A29T; ^C66,^C79; C70A, A76T; T3A, CG6AC, AG29GT, A33G; ^T7,T12_; ^T69,^A76,A88C; C35G, G58C; ^A79,^T79; T16_,C67T, G79_; G7T, T9A; A29T, C37T, C66G,^G77; C2_,G81A; C15G, T34G; T3_,^T9,^C28,A29C, C32_; ^T76,^A76; G7C, A27T; C2_,G79C; TGGC3ACAG, GA26_,--A.33.TGT; ^G65,^G77; ^AC1,GC5_,C8T, GA26_,G30A,^GT34; T9C, G17C, C66T, A80T; T71G, T72G; G4C, CT8GC, G17C, GA28AC, C32G, T69C, G75C; C41A, G51T; ^T78; T9C, G17C, T65A, A80C; AG29CA, C82G; T9G, C82T; T45A, T47C; C2T, T3A; T65A, A80G; C2G, G4A, C32T; G7C, T59G; T9C, T14G; C2G, A29C, T52A; T9C, G17C,-A.53.CC; T9C, T69_,A76_; C68A, G75C; A1G, A33T; T3_,^T9,G28_,^G32; ^G70,G75C; ^C54,G54C; ^T79,G79A; G17C, C70T, A76G; G77A; ^T69,A76C; T65A,^C80; ^A66,G79_; T9G,^G85; ^TGGAAGAT63,C---.66.TCGG, C68A,

	GGAGGGAG74_,^A83; ^T2; G7A, A29C; ^A69,^C76; C6A, A29C; C2_,T9C, G17C, GA79TG
От 2,7 до 2,8	T34A,^T37; A36T, T65C; C2_,T69G; C73A, G74C; G17_; ^G65,^A65; ^T67,C67T; C2G, A29T; T9C, G17C,^C66,^G74; C70A, T71C; T14A, C15T; G4C, C32G, G78C; T9C, G17C, T34A; ^A66,^C79; AGT53GTG; G79_; T9G, T14A; ^C64,^C80; T65C,^G66; ^GT1,G7A, T9C; A60T, G78T; T9C, G17C, C67A, G79A; TC65CG, A80G; T14C, T16C; T3_; ^CGAAC70,T71A, G74C; G7T, C8G; T3A, GC7CG, GA28CG, A33T; C66T, G78T; A1G, T9C, G17C; T69C, C70A; C70T, T72G; T69C, T71G, A80T; T16G, A29C; T11G, A29C; G17A,^TA75,A88C; G7T, G40A, A61G; ^AC81,A88C; ^A71; G5C, C8G, GA28AC, C31G, C73T; G74T, A76C; ^T68,^A76; C2_,C70A; T9C, G28T; G28T, A29C; ^C29; A29C, GA75AC; ^T52,G54C; G7A, T9C, G17C; T9C, G17C, G79A; --A.29.CAC; ^A68,G77A; ^T69; G7C, T9C; A80C, C82T; ^C75,G75T; T14A, A29C; T72C, C73A; T9C, G17C, C66A, G79_; C2_,A33T; ^T2,C64T; ^AT79,A88C; C66G, A80C; ^A67,^T78; ^G67,G78A; ^A76; A21G,^C66,^T77; C2A, A36G, T69A; G63T, T71C; T9C, G17C,-G.77.CT; ^T2,T34A; C68T,^C77; T9C, G17C, T72C; T69A, C70T; CT15TA, A18T; TGG3ACA, C8G, GA28CC, CCA31TGT; T9C, A29C; C6G, G30C; -T.3.AA, C67G; C73_; ^G68,^A76; T69C,^A76; A80G; T69C, A76_; ^G68,^T77
От 2,6 до 2,7	T9A, A29C; A76G; T9C, G17C, AG76CC; T9C,^A13; ^A67,^T78,A88C; C70A, T72A; C66G,^T79; ^T64,C64T; ^A70,C70G; ^G65,A80C; T9C, G17C, C66T, G78C; C2_,T9G; T69_,A76T; T3A, G7A, A29T, A33G; T45G, C68A; ^T65,^T80,A88C; C66G; C64T, T71G; C2G, G54T; A1G, T3A; ^G70,G75T; T65A,^T80; -T.3.AC, GC.7.A-,GAG28TGC, CA32GG, T72A,^G74,A76G; A21C; T69G,^A76; C68G, C70A; C67T, A80T; ^A70,^C75; T9A, T14C; T3A, CGC6TCT, GAG28AGA, A33T; G54_; C68T; ^T65,G79C; C2_; C67G, G79T; CT2TG, G7A, G77C; T71G, G74A; C66T, G81A; A29T; --A.29.CAT, A88C; T69_,^C76; T9C, G17C,^G68; ^A69,T69C; A29C, G30T; T69_; G17C; ^A67,G78C; T65A; ^G79,^T79; A76G, G77T; ^GC1,A88C; A27T, A29C; ^CA79,A88C; T69_,G75T; C38G,^C56,G77A; C68T, G77C; A29C, AG39GT, T52C; G79T, A80T; G7T, A61T; T16C; ^A13; G7C, C15G; G5C, C8G, GA28AC, C31G; C2_,G77C; A29C, T52A; G75C, A76G; T9C, G17C,^C76; C8G, A29C; TGG3GTC, C8G, GA28AC, CCA31GAC; C64_,^GTG67,C68A, G77T,^CAC79,G81_; ^T68,G79_; ^A70,C70A; T65A, AG76GA; ^C70,^C70; C68G, G77T; C6T, A29T; ^T81; ^G67,^A67; TGG3GCA, C8A, A29C, CCA31TGC; G7A, A27C, A29G, A80G; G78A; T52G, G54C; T9C, G17C, T65A, C67T; A1C,^C64,^T81; ^T80,A80C; C67A, C73T; C73T, A80C; C67A, T69C; G7A, A76T, A80G; C2_,C15G; T69C, G77T; CT2_,G79T; G7C,^G28; ^C79; ^A80; ^G1,^C1; ^G65,A80T; G7T, A29_; -T.3.AC, GC.7.A-,GAG28TGC, CA32GG, T65A; T9C,^T14,G17C,^C29; A29T, T69C; T9C, A29G; C64T, T65C; ^TG70,T71A, C73_,G75T; T65G, C66T; T59C,^C66; T72A, G74A; C2T, T72C; T71C, A76G; T65G, A80T; TG3_,^TG7,GA26_,^AG33
От 2,5 до 2,6	T9C, G17C,^G81; --A.29.CAT; C68T, A76G; A29C, G79A; G17C, C67G, C70T; ^G66,C66G; A29T, G63A, C66A; G28C, A29C; T3G, C67A; T69C, T71C; T3A, GC7CA, GA28TG, A33G; C70G, G74A; ^C2,G4C, C8_,^CGC28,CCA31_; C2_,C68T; C66A, A80T; T3A, G5C, GC7AA, GA28TT, C31G, A33T; T9C, G17C, T72G; T9C, G17C, A29C; ^C70,G75T; C66T; C66T, G78A; A36T, G54C, C68T; ^G9,A29T; A76C; T69C, G77C; ^A77,^G77; T71G, G74C; C67T; C73G; T71G, AG76GA; ^C64,T65C; T3G,

	C68A; G74C; C67T, T69A; ^A69; ^A66,C66T; T71C; T14G, T16C; T9C, G79T; T65C; ^C15,G17C; ^T65,^C79; C70G, T71G; G74C, G75T; C2_,C68G; G7T, A27G; ^CA76,A88C; ^T65,^A65; T9C, G17C, T45A; A18C,^A66; A80C; G7C, TC14CT, G17A; TG3GC, G7A, A29T, CA32GC; T16G, A29C, G63T, T71C; C2A, G54T, T71C; ^T8,A29C; T9C, TG16GC; C70T, G77T; G75T, A76G; T69A; T16A, A18G; G77A, G78C; A1_,T59C; T14G, T16G; ^A60,G81A; A29G, A83G; T34A, GA79TC; T69C, G75A; G7T, T59A; G7T, C82G; A36T, G81T; C2_,G81T; T14C, T72_; --A.29.CAC, A88C; TGG3_,^AAG9,GA28CT, C31G, A33G; G17C, A18G; C66G, G77A; ^C5,C6T, C8_,G28C, GC30CG; C82T; G54A,^G56; C2_,C66T; G17C, A18C; G17C, G54_; G28A, T65C; C6T, A29C; G7A, T9C,^T79; T9C, GA17CT; G74A, G75T; C68A, C70G; G42C, C50G; ^C70,^C75; ^T66,^C66; T3C, CGC6GCT, G28_,^A32,A33G; C73A, G74A; TG3AC, C6A, AG29CT, CA32GG; C67A, G79A; A76_; C73G, G74T; TG3CA, GC7AG, GA28CG, CA32TG; T9C, T14C, T71A, C73A; G81C; A1G, T16A; T69A,^G74; C68_; C2A, A60C; T9C, G54T; T14C, C15G; ^G66,^G66; T16C, A18G; ^G68,G77C; A29T,-G.78.CC; G7T,^T61; CT2_,T72G; A1G; T65C, C66A; G7C, T34A; ^C35,T59G; ^AG77,A88C; ^TG67,A88C;
От 2,4 до 2,5	G54C, T59A; T69G, G75C; C68A, A76G; ^AT65,A88C; C68T, G77T; G7T, A29C; T65A, T71A, G74A; T16A; ^C65,^A65; ^T67,G79_; ^G71; ^C18; ^C29,A29T; G79A; T69G, T71A; T71C, T72C; C2_,T3_; ^T67,G78T; CTCCCTCT64_,C73G, AGG76TTC,^TCCCA82; T65A, A83G; C70A, G74A; G7C, TC14AT, G17A, T34C; G7T, A33C, A36C, A76G; T-.3.CA, GC.7.A-,AG29GC, CA32TG; C2_,A80G; -T.3.AC, C6T, C8_,G28C, G30C, CA32GT; G7C, A83G; C2_,C67A; T3G, A29C, T34G, G77A; C2G, A21G, T65C; G40A, T59A; ^A66,^G66; G81A; C2_,A29G; ^T64,G81A; ^CGC2,CGC6_,GAG28_,^AGG33; ^C77; T69A, A76G; ^T78,^T78; C66A,^C79; C2_,G7A, T34A; T3C, C6T, G30A, A33G,^C55; GC7CG, GA28AG; T3C, G5C, GC7TA, G28T, C31G, A33G; ^T68,^C77; ^T77,G77A; A27G,^GT77; ^G66,^T79,A88C; T9C,^G69,^A76; C68T, G75C; ^T81,^T81; ^C66; T9C, G28C; T14A, A29C, C66T; ^A65; T3A, G5C, C8A, A29C, C31G, A33T; CT2_,T71A; G7C, C15G, A33T; G77A,^T78; G63T, C82A; G7A, C15G, G54A, A60C, G79T; ^A13,^G13; T72G, C73T; A36C, G54T; T3G, G7T; ^G65,T65C; T65G, C66G; G77C; T45G; C15A; C41T, G51A; T14A; C2T, G54T; A76T; T71A, A76G; ^G66,^T79; ^A7,A29C; TGG3AAC, C8G, GA28CC, CCA31GTG; ^A1; ^T29; T71G, G74T; T45A; ^AT78,A88C; ^A3,GG4CC, C8_,GA28CG, CCA31GAT; C66T, C70G; C2_,^A66,^C79; ^TA76,A88C; TG3GA, CG6GA, AG29TC, CA32TC; ^C80,A80G; G79C; C67G,^G77; ^C66,G79A; G7A, T16C,^T68; G7C, T9C, G17C, G75C; C2_,^T58; ^A65,^C80; A1G,-C.68.GA; G17C, T65G; TG3CC, C6T, C8G, GAG28ACA, CA32GG; T72C, G75A; C64G, A80T; G7A, C66T; C66G,^C79; C15A, G17A; ^AG66,A88C; A36_; G79T; T9C, G17C,^T58; T10G, A29T; ^G69,^C76; ^A69,A76_; G7A, A29G; A53_; T65G,^A80; C70A, C73A; T59C, G74T; C67A; G54T,^G56; ^G66; C2_,A29C; C38_,G54_; T3_,C6T,^C8,^G28,G30A, A33_; -TG.3.ACC, C6A, C8_,^CTC28,A29G, CCA31_; T9C, T14A; C64_; T14G, G54A; T71C, C73G, A83C; T9C, G17C; A53G, G54T; C66A, A80G; ^G63,^G81; ^G1; ^C78,^C78
От 2,3 до 2,4	G7A, T9C; ^T67,^G67; C2_,C67T; A80_; ^G1,A13C; ^G66,G79C; T69A, A76T; T9C, T14C; A76G, G78C; T16G, G17T; T69C, G77A; T65_,A80G; G7C, T14C, T34G; C66T, C67T; A53G,-A.80.TC; C67T, G77C; C73A,

	G74T; A36G, C68G; T9C, G17C,^C78; TGG3GCT, C8G,^AC28,CA32_; ^T18; ^C29,^T29; ^GGGCG63,C68T, TTCGGA71_,^CCGCC82; T9C, G17C, C66T, A80G; ^A67,^G67; C2_,G79T; T3A, CGC6GAT, GAG28ATC, A33T; C2_,A21G, G79C; C2_,A21G; C64T, G77A; C8A, G79C; C67G,^A78,A80C; T69C,^A70; G74T, G75C; ^T76,A76G; A76T, A80G; ^C64; ^C29,C50T; ^AGCTTA65,^ATTG68,T69A, T72C, G77T, GA--.79.AGCT; A29T, G30A; T65C, A80C; ^C76,^T76; T9C; ^G67,G79_; C68T, G79T; ^CTCA3,GCGC5_,^CAT28,CCA31_; ^GA70,A88C; -T.3.AC, GC.7.A-,GAG28TGC, CA32GG; A21G; ^G69,T69C; G7A,^C66,^G74; ^T65,^A79; T65G; G74A, A76C; G74T, G75A; ^G68,G77T; T9G, G79T; ^AG67,A88C; ^C81; ^A67,G78T; C37A, G57T; G54C, G79T; G75T, G77A; G40A, TAG52CCT; ^G15; C67A, C68A; A36T,^C55; A36T, T59A, T65C; C67T,^G68; T71C, A76C; G7C, A29G, T65A; ^A78; T69C, G75T; ^TC66,A88C; CT2_,T59A; T9C, G17C, T65G; C70G, G75T; C2_,C73T, G75A; TG3CC, C6G, C8T, G28A, G30C, CA32GG; C64G, C66T; T11C, A29C; T9C,^G15,G17C, T65C; T69G, G74T; ^GA65,A88C; G7C, A61G; ^T65,^A80; C68_,^C79; G7A,^T29,G79T; A27T; A1_,T9G, T59C; T14G,-G.79.TT; T14C, T16A; C70A, G74T; T65A, G78A; ^T65,^G77; T9C, G17C,^G68,G77C; C66A,^A79; G7T, T9C, G17C; ^G69,A76T; C2_,A21C; ^T29,A29T; ^G69,^T69; C6T, T10C, T84G; T65C, C67T; C15T; G78C; G7T, A27G, C44T; ^C68,^A68; A1G, T9C, G17C, A76G; A36T, T59A; T14A, T16A; ^C66,G79_; -T.3.AA, G7_,AG29GC, CA32TG; C8G,^A70,^T75; C66A; ^C64,A80_; T69C; T71G, A76T; CT68TC, G74A; G54C, C68T; T9C, G17C, G81T; C2_,A13G; T65A,^C81; ^C66,^A78; ^C70,^A75; ^T68,G77T; A29T, C50T, A53G, G79T; C68T, A76T; T16C, A18T, A80C; ^TGGAAGAT63,C---.66.TCGG, C68A, GGAGGGAG74_,^A83,A86C; -T.3.AA, G7_,AG29GC, CA32TT; T9G, A29G; C68A; A27C, A29T; A36T, G54C; ^A4; ^A73
От 2,2 до 2,3	^C66,A76_; ^G65; ^T1,T59C; A36T; T3C, GCG5CGA, AGC29TCG, A33G; T9C, TG16GC, C68T, G79A; G7A, T14A, G17A, T34A; T65G, G79T; G7C, TC14CT, G17A, T34A; T3C, C67A; G77C, G78T; C2T,^G56; C6T, A83G; G7T, C8A; C66G, G79_; TG3_,C--.8.TCG,^C28,^C30,CA32_; C67T, T69G; CT2_,T9C; G78T, G79A; C2_,T9C, C15A, G17C; T9C, G17C,^TG67; G75T, A76C; ^C76; G79A, A80T; TT71GG, G74C; C70_,G75C; ^G66,G79T; T34A, A60G; A29T, C64T, C66A; ^CT29,A88C; ^G69; A53C, G79T; ^T80,A80G; ^G67,C67A; C67A, G78C; T9C, G17C, C70T, T72A; -T.3.AC, GC.7.A-,A29_,A-.33.GT; C2G,^T58; A27G,^A70; A39G, G78C; -G.78.AA, A80C; C66G, C67A; ^G68,^A68; T69C, T71A; G7T, G40C, AG53GA; T9C, G17C, G79T; C8A, C66T; G74T, A80C; G7C, T14G; ^C77,G77T; G58T, G79C; T14C; ^T65,^A80,A88C; C68A,^C77; GC--.63.ATTA, CCC66ATT, G--GG.77.AATAT, GC81AT; T11G, A29T; T14A, T16G; T71C, G75A; ^T67,^C78; T65C, G81A; G79C, A80T; ^C66,^G74; A53C, G54A; ^C66,^C79,A88C; G79C, A80G; T9C, G17C,^C66,^G74,A88C; C2_,T16C; T69G; ^G68,^A76,A88C; T71A, G74C; G74T; G7T, C37A; ^CA68,A88C; ^T12,^G12; A29T, C64T, C70T; G7C, A29G; G7A, T14A; T69C, C70G; G79T, A80C; C2T, G54C; ^T58; G7T, G30A, G81T; A29C, A83G; C2_,T69C; T3C, G5C, G7C, A29G, C31G, A33G; T72G; C64A; T34G, T59C; A1G, A60C; T65A, G79A; A27T,^C29; ^G67,^G77; ^G68,C68A; C64G; C66T, G77A; ^C64,^A80; C2_,C73T; A29G; ^T7; A1_,A46C, T59C; T9C, G17C, A76T; G78C, A80G; ^C66,A76C; ^T29,^T29; A27T, CT68TC, G74A; G75C, A76C; ^TT81,A88C; ^G77,A80G; ^C5,G7T; ^C66,T69C; C15A, T16A; C73T; ^A65,^A80; ^T65,G79_; G40A, T52C; G7T, A60T;

	TG3GA, GC7CA, A29G, CA32TC; ^TA70,A88C; ^C66,^A66; ^G67; A36C,^T55,C68T; T65_; G63_,C82_; C2A, A29G
От 2,1 до 2,2	A83G; G75_; C68_,G79_; C2_,A46C; ^C4; ^A69,^A69; G42A, C50T; A53G,^T55; A36G,^C58; TG3AC, C8A, GA28TC, CA32GG, T59C, C66A; C2_,A46C, C66T; C64T, G81T; ^A68,G77T; ^T80,A80T; T25G, A29T; G4A, C32T, G54_; ^T68; A76C, G78A; T9C, T14C, G17C; CT2_,A33C; ^CA65,A88C; A60C; ^A69,^T69; T9C, G17C,-T.65.GC; A18C, A61G, A80C; CT15TG, A21C; T72G, A76T; G7C, A29C; ^G79,^C79; T69G,^T76; C70A, G74C; T9G, A29C; C2_,G54A; C15G, T72A, G74A; ^A75; T3_,C6T,^C8,A29_,C32A,^C34; ^C29,A80C; G74A, A76T; C68T, T69C; T3_,C64T; A80T; CT2_,T9A; ^C29,A36C; ^GA67,A88C; T9C, G17C, T59A; A60T, C64T; T65A, G79T; A29C, T65C; ^T7,A13C; C8A, C82T; A76G,^C77; T3G, GC7CT, GA28AG, CA32AC; ---TT.71.AAGAA, G75_; G7T, C15G; ^C79,^C79; TG3GA, CG6AC, A29G, CA32TC, C68T, T72C; T72C; G63C, C82T; ^TG56,G57T; T14C, A29T, A36T; ^T68,^T68; T69G, T71G; ^G66,C66T; ^G68,G77A; G54C, G79A; G7T, C67G; C66G, G78A; A60C, A76G, A80G; G40A,-A.76.CC; C2T, C67A,^T78; T9C, G17C, G77A, G79T; G77T, G78A; ^T78,^C78; ^T68,G77C; ^A67,^G77; C73T, G75A; A29T, C66A, G74T; C2G, A36G; T3G, G5A, GC7CA, A29G, C31T, A33C; T69A, T71C; ^CG2,G5_,C8_,--G.28.CGC, CA32_; ^GT79,A88C; C68A, G77T; C64T; G40A, G77C; C68G, C70G; C2T, G78A; T9C, G17C,^C66,A76C; G7T, A29G, C82T; C2_,T65G, A80G; TGG3GCT, C8G,^CC28,CC31_; A29G, T69C, A80G; T34A, A36_; T9C, G17C, A27G; C15T, T16C; G7T, T9C, G17C, G40A, TA52AT; A36G, T71A; C6T; ^G69,A76_; C66A, G79A; ^C68,^T68; A21T, C67A; A21C, T72G, G77T; T71G, A76G; C2T, G54A; T71G, G77A; T9C, G17C, A29G, G81A; G7A, A36T, G54C, C68T; T3A, T59A; ^G70; ^T77; ^T68,^C77,A88C; TC14GT, T72C; T9C, G17C, T72_; ^C73; G7C, T14C; A36T,^T58; G54T; T59C; A29C, C50T, A60T; G54A, C70G,^T75; ^C66,G77C; C15G, G17C; C64G,^C81; T3A, G5C, GC7AG, GA28CT, C31G, A33G; A29C, C32A; ^G28; A21G, A53G; G75A, A76T; G7C, TC14CT, G17C, T34A; G28A

* мутированные последовательности разделены знаком «;», а множественные мутации в одной последовательности разделены знаком «,»

Таблица 20. Мутации гида 175 мутации и полученное в результате относительное обогащение

Log_2 обогащения	**Мутации в каркасе 175* (SEQ ID NO: 2239)**
От 3,2 до 3,5	C73A,^T78; C6T, A29C, G71C,^G80
От 3,1 до 3,2	C17G, A87C; T3G, CGC6ACT, GAG28AGT, A33C; G7T, C9T, C17G, CG81GA; T16G, A29C; C9T, C17G, C65A, A87G
От 3,0 до 3,1	A68T, T83G; A27G, T92C; TGG3ATC, GC7AG, GA28CT, CCA31GAT; ^C65,A87G; G7T, A29T; T3G, GC7AA, GA28TT, A33C; C9T, C17G, C65_; G7T, T14G; ^G54,G78T; C9T, C17G,^A80; TC16AT, G64C
От 2,9 до 3,0	C15T, T34A; C9T, C17G, A88T; G7A, C15G; ^C76,^G76; CT2_,C15_,T58A; C2_,C15G; C9T, A29C; C9T, C17G, A85T, A88T; C9T, C17G,^CA63; G7T, C9G; A87T, A88C; C73G, G78A; A29T, A91G; TG3GA, G7A, A29G, CA32TC; ^G14,A29T, A87G; C9T, C17G, T74C; C2_,^A53

От 2,8 до 2,9	C9T, A33T; G7T, T67G, G82C; ^T5,C9_,GAGC28CGCA; G7T,^A68,^A82; G7T,^C60; T14G, A29C; A29T, T66A; T3A, CG6TC, AG29GA, A33T; C2T, TC75AT; ^CG76,A88C; G7T, T14A, T83_; -T.3.GA, C6T, C9_,G28C, G30C, CA32TC; CT2_,C15T; TG3_,^GT8,G30C, C32G; T14_,A29C; C9G, C17G, A29C, T79G; TG3AC, G7C, A29G, CA32GT, G86C, A88C; T3A, GC7CA, A29G, A33T; G7C, C80A
От 2,7 до 2,8	G7T, A91C; ^C2,G4C, G7_,A29_,C32G,^G34; CT2_,A88C; C65G, A88C; G7T,-T.79.AA; A29C; T3A, GC7CA, A29G; C8G, A29C, A88_; A29T; C2_,A29C; A29C, C31T, A33G; T14G, C15T; C9T, C15A; ^GA1,G7A, C15A, C17G; C15A, T16A; CT2_,A29C; C9T, C17G, G78_; C9T, C17G, G-.78.AT; C73T, C76G
От 2,6 до 2,7	C9T, C17G, C65_,^A84; C9T, C17G, G70T, C81A; T74A, T79A; T3C, C6T, AG29CA, A33G; G7A,^T29; C76G, G77C; GG77CA, A87G; T16G, A29T; T3A, G5A, A29C, C31T, A33G; C9T, C17G,^AA53; TG3CA, GC7AA, GA28TT, CA32TG; G7A, A29C; T3G, G7T; CT2_,A68G; T14_,A29T; C2_,C9T, C17G; ^G3,GC.7.-T, G28_,^C34; G7T,^T92; G7T,^G69,G82T; ^GGCAGATCTGA64,T66C, A68C, GA71AG,^C75,G77T, T79C, CGTAAGAA81_; T3A, C6G, AG29CC, A33T; C80T,^A81; C81T; CT2_,C17A; C15A, T16G; C2_,T16G; G71_,C80T; TG3AC, GC7AG, GA28CT, CA32GG; T3A, G5C, G7T, C31G, A33T; T3G, G7T, C9T, C17G; G64T, A85T; G7C, T14_; C9T, A29T; G7T,^G14; A88G,^C89; CT2_,A33T; C81T,^A82; C9T, C17G, A29C, C32A; C9T, C17G,^GA77
От 2,5 до 2,6	G7C, C15G; C9T, C17G, TC75GT; TG3CA, CG6GA, AG29TC, CA32GG; G7T; T14A, T16G; G7T, C9T, G71_,^T79; C15A; CT2_,A33T, C73_; C2A, C9T, C17G; CGC6TCA, GAG28TGA; C15G, A29C; C2_,T16G, A91C; ^T81,C81T; TG3AA, A29C, CA32TG; G4A, G7T, C32T; T3C, CGC6GCT, GAG28AGC, A33G; T3A, G7A, A29T, A33G; -G.4.CC, G7_,AGCC29GCGG; C65T, G86_; C9T,^A16; A36G,^C57; A1_,T16G; C6T, G7T; ^G14,A29T; ^AT16,A88C; C8G, A29C; ^G64,A87C; ^G70,^T79; T16A,^C29; TG3GA, C6G, C8T, GAG28ACC, CA32TC, G71T; G7T, A29C; T3G, GCG5AGT, GC30CT, A33C; ^C2,^T14,A29T; C9T, C17G, A88_; C9T, T16A
От 2,4 до 2,5	TGG3ACA, A29C, CCA31TGT; T3_,G5A, G7C,^G9,^C28,A29G, C31T, A33_; C15A, A29T; G64A,^T65; CT2_,A27G; ^A16,^T16; G7T, C15A; G7T, C9T, C17G; C2G, A29T, T66A; TG3GA, CGC6TTA, G28T, G30A, CA32TC; A1C, G82C; A27C, A29C; C9T, C17G,^GA71; T3C,^T6,CC.8.T-,C17G, GAG28AGA, A33G,^G54; ^T16,A27T, A29C; G64C,^A87; ^C14,A29C; ^A65,^T65; C2T, C9T, C17G; C9T, C17A; G70A, C81A; C2G, A36T; G5C, C8G, GA28CC, CC31GA; C6T, A29C; C80T,^G81; T-.3.CA, G7_,AG29GC, CA32TG; ^C78,G78A; G7A, T14_,CT65TC; -T.3.AA, G7_,AG29GC, CA32TG; ^C29,A29T; G7A, A29T; TG3GA, GC7CA, A29G, CA32TC; ^T64,G64A; C15A, A29C; T75A, G77T; ^A3,^T3; A27T, A29C; T14A, A29C; T74C, G77A; G7C, A29G; C9T, C17_; G5A, G7A, A29T, C31T; ^C63,^A63; G7T, A91G
От 2,3 до 2,4	CT2_,G64T, T66G; G28T, A29C; T3G, G5T, GC7CG, GA28CG, C31A; TG3AC, G7C, A29G, CA32GT; C9T, C15A, C17G, A29C,^TG55,G57A; ^C14,A29T; C9T, C17G, GC64TG; G7A,^T29,A36C; ^T16,^G54; TG3CA, C8A, GA28TC, CA32GG; G7T, C9T, C69G; C9T, C17G,^A70; A72_,T79G; T3A, G5T, C8T, GA28AC, C31A, A33T; C9T, C17G, A29C; ^G54; G7A, TC14CT, C17A; C9T, C17G; ^G70,^T79,A88C; ^A64,^G64; T14G, A29T;

	C9T, T16_; ^A14,^T14; ^AC1,GCG.5.--T, GC30_,^GT34; A29C, A91G; C2_,T14A; C9T,^A17; C9T, C17G, G78A; T3G, G5A, A29C, C31T, A33C; C9T,^G17; G7T, A29G; TG3GA, C6G, C8T, GAG28ACC, CA32TC; ^T1,CG6TC, C9T, C17G; C17A; ^T17,^A17; T3A, G5C, GC7AG, GA28CT, C31G, A33G; ^GC72,A88C; T3G, G7T, A33C; TG3CA, CG6GA, AG29TC, CA32TG; T3G, G5C, C8G, GA28CC, C31G, A33C; ^T3,C80G; C9T, C17G, T45G,^G54; C9T, C17G, A72C, T74G; G5C, C8G, GA28AC, C31G; A29T, G56T; G7T, C63A
От 2,2 до 2,3	A36T, A85C, A87T; T14A, C17G; C9T, C17G,^G54; G4C, C8G, GA28AC, C32G, A87G; ^T72; A85C, A87C; G7T, T92C; C9T, C17G,^C63; TG3AA, C6T, AG29CA, CA32TT; C9T, C17G, A85G, A88G; G64C,^G88; G7A,^T29,A68C; ^A13,T14C; C9T, C17G,^G54,A85C, A88C; -GG.4.CAT, C9_,GAGCC28CGATG; TG3AC, C6A, AG29CT, CA32GG; C9T,^C63; C9T, A88C; A27T, A29T; C9T, C17G,^G54,A91C; G86A, A88T; TG3CA, GC7AA, GA28TT, CA32TG, C69T; T74G, G77T; TGG3ACA, C8G, GA28CC, CCA31TGG; G7A, C17A,^G81; G7T, A59G; ^A65,^G86; C73T, G78T; ^C72,^T79; A1G, C9T, C17G; ^G1,C9T, C17G; ^G72,^C72; C2_,A29T; ^T14,A29T; ^G64,^T87; ^A65; ^C18,^T18; ^G64,A88C; C9A, A29C, G57T; G7C,^G28; G77A; G7A, TC14CT, C17G; C2_; G7C, T14A,^T86; C9T, C17G, A53G; T3G, GC7CT, GA28AG, G86T; C9T, C17G, A29C, A91G; C9T, T16_,A91C; CT2_,^G64,C65A; C15_; T16G, C17T; G7T, G28A
От 2,1 до 2,2	C9T, C17G, A29T; A87C; ^CT18,A88C; C9T, C17G,^G64; C17G; C15T; ^T16,T79C; ^A64,G64A; A1C, T3G, C9T, C17G; GA28CC,^T65; C15A, C17A; G78C, T79G; A29C, T58G; C2_,G7A,-C.65.AA; CT2_,A29T; T3A, A33T; G4A, CGC6GTA, G28T, G30C, C32T, T67_; C9T, C17G, C65_,A91C; ^T65,A87G; A88_; G7T, C9A; C9T, C17G, C65A; TG3GC, C6T, AG29CA, C32G; G7T, T16A; G7T, G70C, C80A; G7T, T14A; TG3AA, GC7CG, GA28CG, CA32TG; ^G54,A91C; C73_,G78_; T3C, GC5TG, C8T, GA26_,G30A,^CG34; ^CT3,A29C; C2T, T14G; G7C, A29T; C9T, TC16GG; T3G, C8T, GA28AC, A33C; ^G16,^T16; C9T, C17G, A36C; TGG3AAC, C8G, GAG28_,A---.33.GGGT; C9T, C17G, A87G; ^T72,T79G; ^G17,C17T; CT2_,A39C, A88C; T3G, A33C; T3_,A33G; C-.2.TG, TC75CA; G7C, C9T, C17G,^G92; C9T, C17G, G82C; C9A, A29C; C2_,C9T, C17G, A91C; C2_,A29C, A91C; CT2_,C9T, C17G; G7T, A60G; ^C71,^T71; C2_,G77T, A91C; C2_,A29G; ^T71,C80G; T3A, G7A, A29G, A33T; C9T, A29G
От 2,0 до 2,1	C65T,^A66; CT2_,C15_,T58A, A72C; C9T, C17G, C73A, C76A; C2_,A91C; C80T; T3A, G7C, C9T, C17G; ^C63,^G88; G7T, A61T; GC62_,C65G, T67G, A72T, T79A, AAGA.84.---C, G89C; T3G, C9T; T16A, C17A; C6T, A29T; T3C, GC5CG, C8T, GAGC28ACCG, A33G; G7A, C15T; ^T2; C15G; C9G, A29T; C15T, A29T; G7T,^C14; ^A64,A88T; A29C, G30A; C2_,A29C, A46C; C9T, C17G, A72T, G78A; ^A87,^T87; C9T, A59C; TG3AC, C8A, GA28TC, CA32GG; C9T, C17G,^G64,^G88; A29C, G71A, C80T; T3C, A29T, AC68TA; ^A17; C9T, C17G, G64T, T66C; G7A, T16G; C17T, C65G, G86C; C69T, G82C; A1T, C2A; T14A,^C29; ^A15,C15T; G7T, T16G; T3A, GC7CA, GA28TG, A33G; ^T81; T16C, A29C; A29C, A91C; G71A, A88T; ^C65,A87G, A91C; C9T, C17G, A29T,^A53; G71T; ^A80,^A80; C9T, C17G, A36G; C9T, C17G, T--.54.CTG; T16A, A29T; ^G77,T79C; C9T, C17G, G64C; TG3AC, CG6GA, AG29TC, CA32GG; A36T, C37T; A29C,^C65,A85_; C15G, A29T; ^A70,C81T; A29T, A33G; C73A, C80T; C9T, C17G, G82_; C9T; C69T,

	A84G; C2_,C9T, C17G, A46C
От 1,9 до 2,0	C2_,A29G, A91C; A68G, T83C; C9T, T14A, C17A,^AG85; ^T66,^G85; G62T, CT65_,C69A, G71A, C80T, G82T, A85C, AGC88_; T3_,G5T,^A8,-A.29.TC, C31A, A33_; G7A, T14C, C17A; T3G, CG6TC, AG29GA; ^T54; ^C8,^T8; G7T, AA87TG; A72C, C73A; C2_,C6T; ^C29; G71C, C81_; C9T, C17G, G64_,A88_; C2_,A88T; T3G, G5C, GC7TG, G28C, C31G; C9T, C15T, C17G, A36C; G7T, T34G; T14A; ^T73,^C78; ^G64; ^G15,C15T; A36C,^A57; A-.72.GC,^T79; T16A, A29C,^A58; C9T, C17G,^T52; C2_,A85T; ^C29,A29G; G7T, T14C; C2A,^T57; G7T, C15G, T34G; T14G, C17T; T14C, C15T; T3G, G5A, GC7TA, G28T, C31T, A33C; ^C71,^T79; ^T14,A29C; ^A1,A36C; ^C63,^G89; G7C, A91G; T14C, A29C; C9T, C17G, G78T, C80T; ^G69,G82C; TGG3GCA, G7T, CCA31TGC; C6T, A29C, G71C,^G80,A91C; A13C, A29C; ^C63,A88T; G7T, T14_; C2_,GG77AA; C9T, C17G, T58A; C2_,G77T; C2_,T3_; C9T, C17G,^AA53,A88C; G7T, C9T; G7A; CG6GC, AG29GC, C32A; C63T, TTA66GCC, GA71_,TC79_,TAA83GGC, A87C, G89_; G7C, C17G; C2_,A46C; C9G, A29T, C37T,^A56
От 1,8 до 1,9	^G69,A72C, G82C; ^G70,T79G; G7A, C15A; ^T36,^A57; ^G70,^C79; TGGCG3CACAT, GCCA30TGTG; G71A; TG3AC, C8A, A29C, CA32GT; T10G, A29C; ^A65,G77A,^G86; C9T, C17G, A88_,A91C; ^C78,^A78; G7T, C90T; T3G, G5A, GC7TG, G28C, C31T, A33C; G7T, C9G, G86T; A29C, C31T, A33C; A29C, G70A; A-.88.GC, A91C; ^A17,A36C; T3C, GCG5TGA, AGC29TCA, A33G; T3C, CGC6GCT, GAG28AGA, A33G, A88C; C35G,^C58; T74A, G78C; C9T, CA17GT; G7A, C17G; C9T, C17G,^GT70; CTG2_,A29C; C2_,A68G; ^T64,^T88; T3G, A33T; C2_,T16G, A29C; ^A1; A36T,^G55; C9T, C17G, C63A; C9T, A18G; C2T, A36T; ^A81,^A81; C9T, T14G, C17G; -A.72.CC, A91C; A29T, T79G; G7A, A29T, A59G; G7C,^C78; ^AG64,A88C; CT2_,C9T, C17G, C69T; C2_,A46C, A91C; ^C89,A91C; ^C29,A68C; C2_,G64T; -C.15.GT, A27C; CT2_,T10G, A88C; T14C, A29T; C9T, C17G, C76T; A84G, A87C; G7C, C9T, T14A, C17G, T34A; G70T, C81A; T14G; ^T3,A29T; G7T,^T29; A29T, C65A, T67G; G64C, A87G; C9T, T14A, C17G; ^T57,A87G; TGG3ATC, A29C, CCA31GAT
От 1,7 до 1,8	C2_,G70A; C9T, C17G,^GA77,A88C; C9G, C17G, A29C; ^T70,^T81; G7C, C9T, C17G; T3G, CGC6TTG, G28C, G30A, A33C; ^A16,A68T; C9T, C17G, T67C; G7T,^C14,A33C; G7A, T14_; ^C14,^T14; C9T, C17G, GG77TT; C2T, C80T; ^T64,A88_; ^G54,A68C; G7T, CT9AG; C9T, C17G, T79G; T79G, C80T; ^AT3,A88C; ^AG54,A88C; C2G, A33C; C2_,A88T, A91C; C9T, C17G, T58C; C2_,C73T; TGG3CCC, C8G, GA28CC, CCA31GGG; G7T, T10G; C9T, C17G,^A80,A91C; ^T64; T14_,A29C, A91C; G7A, G28T, AAAGCGCTTA59_; G7T, G71_; ^A17,^A17; T14_,A29T, A91C; C17G, A72G, T74C; ^T88; CT2_,A94C; A27G, A29C; A85T, A87G; C9T, C17G,^AA79; C9T, T14A, C17G, T34A,^G64,G86T; C9T, C17G, T45G; C2_,C9T, C17G, C65T; ^G3,G5C, C9_,GA28CG, C32A; T74G, G78T; TG3_,--C.8.GCT, G28_,^G33; A39T, T54A; C2_,A72G; C9T, C15T, C17G; TG3CA, CG6GA, AG29GC, CA32TG; G64C, A88G; C15A, C17G; C2_,C65A; ^G64,G86A; ^C29,A36C; G64T, T66A; TG3GT, A29C, C32A; ^A64; C81G; C9T, A72T, T79C; C9T, C17G, G77T
От 1,6 до 1,7	A72G; ^C14,A29C, A36C; T3C, C9T, C17G; G4C, C8G, GA28AC, C32G; C2_,G71C,^G80; C76T; C9T, T14A; C2G, C9T, C17G; G70T, C81G; C17G,^T54; A72C; C2_,C9G, C17G; TG3GC, C8T, GA28AC, C32G;

	TGG3GCT, C8G,^CC28,CC31_; C9T, C17G, A39T, A-.53.GC; ^T16; T67C, A87C; ^G81,C81T; C76G, G78C; A1C, G56A; TG3CA, GC7AG, GA28CT, CA32GG; C9T, C17G, C65G,^A87; G86A, A88C; G7T, C9T, C17G,^A72,G78A; ^G70,C80A; ^A17,A68C; C2_,C80G; ^C71,^T79,A88C; C9T, C17G,^T57; ^T2,C9T, C17G; T45G; G64C; T14_; C65T, G86A; C69T; ^C65; G64T, C65A; T3G, GC7CT, GA28AG; ^A1,^A53; T3A, G5C, GC7AT, GA28AT, C31G, A33T; C9T, C17G,^CA72; C9T, C17G, C73A, T79A; C2_,A53G; TGG3GTC, C8G, GA28CC, CC31GA; ^C5,G7T, C9T, C17G; G71T, C80T; C15T, T16G; G7C, C9T, C17G, C76A, G78T; G64T, T66C; ^C65,A91C; C73T; A72C, G78T; ^C63; A68G, C81T; ^GT87,A88C; C9T, C17G,^A78; T3A, GC5AG, C8T, GAGC28ACCT, A33T; ^A1,^T54; A29C, G56A; C2_,C80T; ^TA17,A88C; A72G, C73T; A29C, C31T, T83C; G7T, A27T; T3C, G7T, G40A,^T54; A88C; G64T, A87C; T3_,^T9,G28_,^G32; ^GT16,A88C; -T.3.AC, G7A, C9_,GAG28TGC, CA32GG, A84C, G86T; ^T65; C76A, G77T; ^G14,A29C; G64C, A88C; A72_,T79G, A91C; ^C29,A68C, A72C; TG3AT, GC7TT, G28A, CA32AT; C9T, C17G, T--.54.CTG, A88C; G7T, A59C; CC8GT, C17G; G7C, T14C,^T86; ^CA3,GC5_,C8G, GA26_,G30C,^TG34
От 1,5 до 1,6	T3A,^A5,G7_,AGC29GCT, A33G; C9T, C17G,^C73,G78C; G71A, A72G; AG27TA, A88T; G7T, A91T; ^T57,A91C; ^T2,A68C; ^T2,A36C; G7T, T10C; ^A64,A88G; TG3CA, C6T, C8T, GAG28ACA, CA32TG; ^T54,A68C, A72C; G7T, A61G; GCGC5CAAG, GAGC28CTTG; C6T, CT9TC, C17G, A29C; ^CA63,A88C; C2_,C9T, C17G, A36C; ^G64,^G86; ^CGGCAGAT65,T67G,^GC69,G70T, A72G,^GCTC75,G77T, T79C, CGTAA81_; C73T,^G74; T14G, T16A; ^AT14,A88C; G64C, A88T; C2_,A39T,^A55; C2_,C15T; ^G70,C81T; ^A81,C81T; ^T72,A72T; C2_,C69T; T75G, T79G; A88_,A91C; ^T7,G7T; G7A, A29T,^A77; CC8AT, C17G; C2_,T52C; G7A, C9T, TC16CG; G70A; C9T, C17G, AA87TC; ^A53,A91C; T3A, G5C, GC7CT, GA28AG, C31G, A33T; ^G70,^C79,A88C; ^T72,^G77; C9T, C17G, C69T; T-.3.CA, G7A, C9_,AG29GC, CA32TG; TGG.3.-AA,^G9,^CGC28,A29T, CCA31_; GCGC.62.--AA, T67C, C69A, GA71AC, TC79GT, G82T, AAGA.84.---G, GC89TT; A85G, A87G; TG3_,C--.8.TCG, GAG28CGA, C32G; T66C, A85G; ^A16,G86T, A88T; TT74GG, G--.77.AAC; C2_,T79C; C9T,^A13,C17G,^G54; ^C63,G64T; C2_,T83C; ^C73,^C73; -T.3.AA, G7_,A29_,A-.33.GT, G70A; ^T16,A91C; ^T64,^G64; T79C; C9T, C17G, G77A; ^T64,^T64; C2_,G71A; T14C, C17G; G7C, TC14CA, C17G; A85C, A88C; ^A3,GG4TC, C9_,GA28CG, CCA31AAT; --C.63.TTT, C65_,CGGA.69.T---,TCCG.79.---A, G86C, G89A; C9T, C17G,^C57; C15G, T16A; C9T, C17G,^CA64; AG39TA, T52C, T54A; C2A, A87G
От 1,4 до 1,5	-C.15.GT, A36C; A29C, T83C; G7T, A27G; ^C29,^C29; ^T80,^C80; TGGC3ACAG, GA26_,--A.33.TGG; A72G,^T73; C9T, C17G, T66A, A85G; ^C15,^G15; TG3_,--C.8.GCT, GAG28CGC, C32G; ^T19; G28A, A29C; ^G70,^G80; CT2_,A36C, A39C; C9T, C17G,^CC79; ^G54,A68C, A72C; ^CT78,A88C; T74G, G78C; TTC74AGG,^AT78; C9T, C17G, C76G; ^GGCAGCTCTGA64,T66C, A68C, GA71AG,^C75,G77T, T79C, CGTAAGAA81_; ^A1,A68C; ^A4; A72G, G78C; T3G, C8T, GA28CC, A33C; G7C,-C.80.AT; C9T, C17G, A59T; G26C, C93G; G7C, T14A,^T86,A91C; ^G64,^T87,A88C; A1G, A29C; C9T, C17G,^AT78; G28T, GCCA30TTTG; C2_,T75A, G78A; TG3GA, CG6AC, AG29GT, CA32TC; A36G,^C57,A91C; ^C72,A72C; C9T, C17G,^G82; A27T;

	TG3CC, CGC6TTG, G28C, G30A, CA32GG, C80G; ^A1,^A53,A88C; A72C, C80A; G7T, C73G; ^A15,A87G; T14_,^C29; G7A, T14_,A91C; C15T, T16A; C15T, C17G; C65_,A88_,A94C; ^A16; C9T, C17G,^G54,A68C; -T.3.AC, G5A, C9_,GAG28CGT, CA32GG; ^T15,^C15; C9T, T14A, C17G, T34C,^G64,G86T; ^T71,C80G, A91C; -C.15.GT, A68C; ^G87,^T87; C73_,G78_,A94C; C2G; G77C, T79A; G70C; A68G; ^T81,A91C; C9T, C17G, T79A; ^T72,^T72
От 1,3 до 1,4	T66A, A88C; C76G, G77T; A53G, A59C; CTG2_,G7T; A72_,^T79; ^AA80,A88C; TGG3CAA, C8G, GA28CC, CCA31TGG; ^C78,^T78; --G.28.TGA, T79C; ^T72,^G77,A88C; A72G,^C79; T3G, G5A, G7A, A29G, C31T, A33C; T14G, A21G; ^T2,A72C; G7T, T14G,^CG64; T3G, G71A; G64A, A87G; T3C, C6T, AG29CC, A33G; T45A; G7A, C9T, T14A, C17G; TG3CT, CGC6TAT, GAG28ATA, CA32AG; C9T, C17G,^T83; G7T, C9T, A53T; C9T, C17G, T75G; G7C, T14C, A72_; ^A65,A87G,^C89; C9T, C17G, G70C, C81G; G7T, A59T; AG29CA, A72T,^G77; T74C, G78A; C2A; C9T, C17G, C73T, T75G; ^G54,A72C; ^AA81,A88C; ^T54,A68C; C65A, G86A; ^A1,A72C; T3G, C9T, C17G; C2_,A33T; A87T; ^A65,^T86; A53G; A85G, A87C; T3G, G5C, GC7TG, G28C, C31G, TC75_; -T.3.AC, G7A, C9_,GAG28TGC, CA32GG, G71T; G7C, C15A; G64A, A85G, A88_; ^A74; ^TG64,A88C; A29C, A60T; C9T, C17G, C80G; ^T64,^A87; G7T,^A59; G77C, G78C; A72C,^T79; ^T73,^C78,A88C; ^C29,A91C; ^A64,A88C; ^G54,T58A; TGGCG3CACTT, GCCA30AGTG; C9T, C17G, A21T; G4C, C8G, GA28AC, C32G,^G82; A36C, A53G; C9T, C17G, G71T; C9T, CA17GT, T45A, G70C; ^A81; G7A, A72T; CT2_,T10G; G64T, A87G; ^G70,T79A; C2_,C9T, C17G, T52C; C2_,T45C; C9T, C17G,^C35,A36G; G7T, T58A; ^A73,^C73
От 1,2 до 1,3	C2G, C73G; G7T,^T14; T75C, C76T; ^A80,^C80; A1_,A46C; C9T, C17_,A91C; C35G,^C58,A68C; C2T, T3A; ^C29,A72C; T79G, C80A; G71A, C81_; G7T, G28T; CT2_,T45G; A29C,^G92; C9T, C17G, T67C, A84G; T3C,^T6,C9_,GAG28AGA, A33G; A36T; A85C, A88T; TG3GC, C6A, C8T, GAG28ACT, CA32GC; T10C, A29C; A1_,C2_; ^C65,A87T; A72T, C81T; C15A, T79A; ^GA1,G7A, C15A, C17G, A88C; ^A16,T16A; A29T, A60C; C76A, G78A; A29T, C31T; A29C, G86C; ^G70,T79G, A91C; ^T54,A72C; ^GAAC73,T74A, GG.77.C-; T14_,A29C, A46C; C9T, C17G,^A72,^A78; T14C, C15A; ^A17,^G17; C9T, C17G, CG76AC; T74C, T79C; G7A, TC14AA, C17A; ^T64,^A64; ^T81,^A81; C2A, A36T; C9T, C17G, G82T; T74A, G77A; ^A1,A33C, A36C; G7C, TC14CT, T34A; A36T, A53G; ^A65,^A84; A1_; G7T,^T60; T3A, G5C, G7T, C31G, A33T, T52G,^C54; T75G, G77T; G5C, G7A, A29T, C31T; TGGC3CCAG, C8T, GA26_,G30A,--A.33.TGG; C9T,^C17; C2_,T14A, A91C; G77A, G78T; ^G64,G86A, A91C; T16A, C17G; C9T, C17G, T34A; A87G; A39G,-T.54.GC; A39G,-T.54.GC, A91C; ^A5,C6T, C9_,G28C, GC30CT; A72C, G77A; C2_,A91C, A94C; C2_,G7C; A84G; C73A, G78T; ^T78,^A78; TGG3GTC, C8G, GA28AC, CCA31GAC; G7A,^G14; C76T, G77A; C2_,G7T; G7A, T14A; ^A17,A68C, A72C; TGG3CCA, GC7CG, GA28CG, CCA31TGG; T79G; ^A72,^C78; C15G, A29T, G57C, A59T; T14A,^G74; G7T, C65T, A87C; C9T, C17G, G70T

* мутированные последовательности разделены знаком «;», а множественные мутации в одной последовательности разделены знаком «,»

Пример 14. Анализ отбора CcdB позволяет идентифицировать варианты белка

CasX с улучшенным расщеплением дцДНК или улучшенной специфичностью спейсера при наличии PAM-последовательностей TTC, ATC и CTC

[1035] Были проведены эксперименты для идентификации набора вариантов, полученных из CasX 515 (SEQ ID NO: 416), которые являются биохимически компетентными и проявляют улучшенную активность или улучшенную специфичность спейсера по сравнению с CasX 515 при расщеплении двухцепочечной ДНК (дцДНК) по целевым ДНК-последовательностям, ассоциированным с PAM-последовательностью TTC, или ATC, или CTC. Для обеспечения, во-первых, определяли набор спейсеров с выживаемостью выше фоновых уровней в эксперименте по отбору CcdB с использованием CasX 515 и гидового каркаса 174. Во-вторых, проводили отбор CcdB с этими спейсерами для определения набора вариантов, полученных из CasX 515, которые являются биохимически компетентными в отношении расщепления дцДНК при наличии канонической PAM-последовательности TTC «дикого типа». В-третьих, проводили эксперименты по отбору CcdB для определения набора вариантов CasX 515, которые обеспечивают улучшенное расщепление дцДНК при наличии PAM-последовательностей либо типа ATC, либо типа CTC. В-четвертых, проводили эксперименты с отрицательным отбором плазмид для определения набора вариантов, полученных из CasX 515, которые приводили к улучшенной специфичности спейсера.

Материалы и методы

[1036] Для экспериментов по отбору CcdB 300 нг плазмидной ДНК (p73), экспрессирующей указанный белок CasX (или библиотеку) и огРНК, трансформировали методом электропорации клетки *E. coli* штамма BW25113, несущие плазмиду, экспрессирующую токсичный белок CcdB. После трансформации культуру оставляли для восстановления в обогащенной глюкозой среде в течение 20 минут при 37 °C со встряхиванием, после чего добавляли IPTG до конечной концентрации 1 мМ, и культуру дополнительно инкубировали еще в течение 40 минут. Затем восстановившуюся культуру титровали на чашках с агаром на среде LB (Teknova, кат. № L9315), содержащим антибиотик, селективный для плазмиды. Клетки титровали на чашках, содержащих либо глюкозу (токсин CcdB не экспрессируется), либо арабинозу (токсин CcdB экспрессируется), рассчитывали относительную выживаемость и наносили на график, как показано на ФИГ. 32. Затем культуру подвергали электропорации и восстановлению, как описано выше, и часть восстановившейся культуры была сохранена для титрования. Оставшуюся часть восстановившейся культуры разделяли после периода восстановления и выращивали в среде, содержащей либо глюкозу, либо арабинозу, чтобы собрать пробы объединенной библиотеки либо без отбора, либо с сильным отбором соответственно. Эти культуры собирали и экстрагировали пул плазмид из выживших клеток с использованием набора для мини-выделения плазмид Plasmid Miniprep Kit (QIAGEN) в соответствии с инструкциями производителя. Весь процесс повторяли для всех трех раундов отбора.

[1037] Выделяли окончательный пул плазмид и проводили ПЦР-амплификацию плазмиды p73 с использованием праймеров, специфичных для уникального молекулярного

идентификатора (UMI). Эти UMI-последовательности были разработаны таким образом, что каждый конкретный UMI связан с одной и только одной мутацией белка CasX 515. Для амплификации использовали типичные условия ПЦР. Пул вариантов CasX 515, содержал многие возможные аминокислотные замены, а также возможные вставки и делеции одного аминокислотного остатка, внесенные с применением подхода, называемого глубокой мутационной эволюцией (DME). . Амплифицированный продукт ДНК очищали с помощью набора для очистки ДНК Ampure XP DNA cleanup kit с элюированием в 30 мкл воды. Затем подготавливали ампликоны для секвенирования с помощью второй ПЦР, чтобы добавить адаптерные последовательности, совместимые с секвенированием нового поколения (NGS) либо на приборе MiSeq, либо на приборе NextSeq (Illumina), в соответствии с инструкциями производителя. Проводили NGS подготовленных проб. Возвращенные файлы с необработанными данными обрабатывали следующим образом: (1) последовательности обрезали для обеспечения качества и для адаптерных последовательностей; (2) последовательности из прочтения 1 и прочтения 2 объединяли в одну последовательность вставки; и (3) в каждой последовательности количественно определяли содержание UMI, связанного с мутацией относительно референсной последовательности CasX 515. Подсчитывали частоты встречаемости отдельных мутаций относительно CasX 515. Число мутаций после отбора делили на число мутаций до отбора, и для получения «оценки обогащения» использовали псевдоотсчет, равный десяти. Рассчитывали логарифм по основанию два ($\log_2$) этой оценки и наносили значения на диаграмму в виде тепловых карт, представленных на ФИГ. 33-36. Показаны биологические повторности для одного спейсера, а среднее значение для этих повторностей рассчитывали и наносили на диаграмму как общую оценку обогащения. Для эксперимента, представленного на ФИГ. 33, библиотеку пропускали через отбор CcdB с двумя спейсерами и PAM-последовательностью TTC, выполненный в трех повторностях (спейсер 23.2 AGAGCGTGATATTACCCTGT, SEQ ID NO: 161 и спейсер 23.13 CCCTTTGACGTTGGAGTCCA, SEQ ID NO: 162), и с одним спейсером и PAM-последовательностью TTC, выполненный в двух повторностях (спейсер 23.11 TCCCCGATATGCACCACCGG, SEQ ID NO: 154), и среднее значение трех повторных измерений наносили на диаграмму по шкале $\log_2$ обогащения на тепловой карте для измеряемых вариантов CasX 515. Варианты CasX 515, которые сохраняли полную способность к расщеплению по сравнению с CasX 515, демонстрировали значения $\log_2$ обогащения около нуля; варианты с потерей функции расщепления демонстрировали значения $\log_2$ ниже нуля, а варианты с улучшенным расщеплением при применении этого способа отбора давали значения $\log_2$ выше нуля по сравнению со значениями у CasX 515. Эксперименты для получения тепловых карт, представленных на ФИГ. 34-36, проводили с использованием следующих одиночных спейсеров (11.2 AAGTGGCTGCGTACCACACC, SEQ ID NO: 163; 23.27 GTACATCCACAAACAGACGA, SEQ ID NO: 164 и 23.19 CCGATATGCACCACCGGGTA, SEQ ID NO: 157 соответственно) для селективности.

[1038] В рамках экспериментов по отрицательному отбору плазмид проводили дополнительные раунды отбора бактерий из окончательного пула плазмид, полученного в

результате отбора CcdB с применением спейсеров для PAM-последовательности TTC. Общая схема отрицательного отбора заключается в обеспечении репликации только тех клеток *E. coli*, которые содержат две популяции плазмид одновременно. Первая плазмида (p73) экспрессирует белок CasX (экспрессия индуцируется гидрохлоридом ангидротетрациклина (ATc)) и orPHK (экспрессируется конститутивно), а также ген устойчивости к антибиотику (хлорамфениколу). Следует отметить, что эта плазмида также может применяться для анализов со стандартным прямым отбором, таким как CcdB, и что экспериментатор может при необходимости свободно изменять спейсерную последовательность. Вторая плазмида (p74) служит только для экспрессии гена устойчивости к антибиотику (канамицину), но была модифицирована таким образом, чтобы она содержала (или не содержала) целевые сайты, соответствующие спейсеру, закодированному в плазмиде p73. Дополнительно эти целевые сайты могут быть сконструированы таким образом, чтобы в них были встроены «неспаренные основания» относительно спейсерной последовательности, не образующие канонические пары оснований Уотсона - Крика между РНК спейсера и ДНК целевого сайта. Если RNP, экспрессируемый с плазмиды p73, способен расщеплять целевой сайт в плазмиде p74, клетка будет оставаться устойчивой только к хлорамфениколу. Напротив, если RNP не может расщеплять целевой сайт, клетка будет оставаться устойчивой как к хлорамфениколу, так и к канамицину. Наконец, описанная выше система репликации двух плазмид может быть достигнута двумя способами. В последовательных способах сначала в клетку может быть доставлена любая из плазмид, после чего штамм делают электрокомпетентным и доставляют вторую плазмиду (в обоих случаях - методом электропорации). Проведенные ранее исследования показали, что любой порядок доставки плазмид является достаточным для успешного отрицательного отбора, и выполнялись обе схемы: в эксперименте под названием Screen 5 проводили электропорацию плазмиды p73 в компетентные клетки, несущие плазмиду p74, а в эксперименте Screen 6 порядок был обратным. Культуры подвергали электропорации, давали восстановиться, титровали и выращивали в селективных условиях, как указано выше для одного цикла, а затем проводили извлечение плазмид с последующими амплификацией, NGS и расчетом обогащения, как описано выше.

[1039] Наконец, проводили дополнительные отборы CcdB аналогичным образом, но с гидовым каркасом 235 и альтернативными промоторами WGAN45, Ran2 и Ran4; во всех случаях нацеливание осуществлялось на плазмиду с геном токсического белка CcdB и использовался спейсер 23.2. Ожидается, что эти промоторы будут вызывать более слабую экспрессию гидовой РНК по сравнению с вышеописанными отборами CcdB, и, следовательно, как ожидается, они уменьшат общую концентрацию CasX RNP в бактериальной клетке. Этот физиологический эффект должен снижать общую выживаемость бактериальных клеток в условиях селективного анализа, увеличивая таким образом динамический диапазон оценок обогащения и более точно коррелируя с нуклеазной активностью RNP со спейсером 23.2 для PAM-последовательности TTC. Для

каждого промотора проводили три цикла отбора в трех повторностях, как указано выше, и каждый цикл экспериментов приводил к получению данных по обогащению, приведенных выше. Эти эксперименты далее называются Screen 7.

Результаты:

[1040] Результаты, представленные на ФИГ. 32, демонстрируют, что CasX 515 в комплексе с гидовым каркасом 174 способен расщеплять экспрессионную плазмиду с CcdB при нацеливании с применением спейсеров (перечисленных ниже), которые нацелены на ДНК-последовательности, ассоциированные с PAM-последовательностями TTC. Напротив, спейсеры, использующие альтернативные PAM-последовательности, демонстрировали гораздо более вариабельную выживаемость. Спейсеры для PAM-последовательности ATC (перечисленные ниже) обеспечивали диапазон выживаемости от нескольких процентов до намного менее 0,1%, а спейсеры для PAM-последовательности CTC (перечисленные ниже) обеспечивали выживаемость в диапазоне от > 50% до менее 1%. Наконец, спейсеры для PAM-последовательности GTC (перечисленные ниже) обеспечивали выживаемость на уровне всего 0,1% или ниже. Эти данные сравнительного анализа свидетельствуют в пользу плана эксперимента с таким процессом отбора и демонстрируют устойчивую мощность отбора бактериального анализа с CcdB. В частности, обогащение белками CasX, неспособными расщеплять двухцепочечную ДНК, снижено по меньшей мере на четыре порядка, тогда как белки CasX, которые биохимически компетентны в отношении расщепления, успешно пройдут анализ.

[1041] Тепловые карты, показанные на ФИГ. 33, применяли для идентификации набора вариантов CasX 515, которые являются биохимически компетентным в отношении расщепления дцДНК по целевым ДНК-последовательностям, ассоциированным с PAM-последовательностью TTC; тепловые карты на ФИГ. 34 и ФИГ. 35 применяли для идентификации тех вариантов, которые демонстрируют улучшенное расщепление дцДНК по целевым ДНК-последовательностям, ассоциированным с PAM-последовательностями CTC (спейсеры 11.2 и 23.27), тогда как тепловые карты на ФИГ. 36 получали с помощью того же набора вариантов для спейсера, ассоциированного с PAM-последовательностями ATC (спейсер 23.19).

[1042] Результаты, представленные на ФИГ. 33A-E, идентифицируют и категоризируют варианты CasX 515, компетентные в отношении редактирования трех целевых сайтов, ассоциированных с PAM-последовательностями TTC. Эти три набора данных, как по отдельности, так и в комбинации, представляют базовые биохимические различия между вариантами и идентифицируют области, представляющие интерес с точки зрения будущего конструирования улучшенных терапевтических средств с CasX для редактирования генома человека. В качестве доказательства этого в состав интактной библиотеки были равномерно включены внутренние контроли, например, со стоп-кодоном в каждом положении по всей длине белка. Неизменно наблюдалось, что эти стоп-кодоны утрачивались во время раундов отбора, что согласуется с ожиданием того, что частично усеченный CasX 515 не должен обеспечивать расщепление дцДНК. Аналогичным образом,

было обнаружено, что варианты с потерей активности, отраженные в данных тепловых карт, истощались в процессе отбора, и, таким образом, для них характерна выраженная потеря способности к расщеплению двухцепочечной ДНК в этом анализе. Однако варианты со значением обогащения один или выше (и соответствующим значением $\log_2$ обогащения нуль или выше) как минимум были нейтральны с точки зрения способности к биохимическому расщеплению. Важно отметить, что, если одна или более мутаций, идентифицированных в данном конкретном подмножестве вариантов, приводят к желаемым для терапевтической молекулы свойствам, эти мутации устанавливают зависимость структура-функция, которая, как было показано, согласуется с биохимической функцией. Более конкретно, эти мутации могут влиять на такие свойства, как транскрипция, трансляция, свертывание, стабильность белка CasX, образование рибонуклеопротеина (RNP), распознавание PAM, раскручивание двухцепочечной ДНК, расщепление нецелевой цепи и расщепление целевой цепи.

[1043] ФИГ. 34A-E, ФИГ. 35A-E и ФИГ. 36A-E применялись для идентификации и категоризации вариантов, компетентных по отношению к расщеплению по последовательностям, ассоциированным с PAM-последовательностями CTC и ATC. Варианты с обогащением в этих наборах данных (обогащение > 1, что эквивалентно значениям $\log_2$ обогащения приблизительно 0) представляют мутации, которые специфически улучшают расщепление целевых сайтов, ассоциированных с PAM-последовательностями CTC или ATC. Мутации, удовлетворяющие этим критериям, могут быть дополнительно подразделены на две общие группы: либо мутация улучшает скорость расщепления посредством улучшения распознавания PAM (тип 1), либо мутация улучшает общую скорость расщепления молекулы независимо от PAM-последовательности (тип 2).

[1044] В качестве примеров первого типа было обнаружено, что для мутаций по типу замены в положении 223 было характерно обогащение в несколько сотен раз во всех протестированных пробах. Это местоположение кодирует глицин в обоих референсных белках CasX дикого типа - CasX 1 и 2, который, согласно измерениям, располагается на расстоянии 6,34 ангстрема от положения -4 нуклеотида в ДНК нецелевой цепи в опубликованной структуре CryoEM для CasX 1 (PDB ID: 6NY2). Таким образом, эти мутации по типу замены в положении 223 физически приближены к измененному нуклеотиду новой PAM-последовательности и, вероятно, непосредственно взаимодействуют с ДНК. Дополнительно в пользу этого вывода свидетельствует то, что многие замены с обогащением кодировали аминокислоты, которые способны образовывать дополнительные водородные связи относительно замещенной аминокислоты (глицина). Эти результаты демонстрируют, что улучшенное распознавание новых PAM-последовательностей может быть достигнуто в белке CasX посредством введения мутаций, которые влияют на взаимодействие с одной или обеими цепями ДНК, в особенности если они физически приближены к PAM-последовательности ДНК (в пределах десяти ангстрем). Дополнительные элементы тепловых карт, показанных на ФИГ. 34-36, могут представлять мутации, обеспечивающие повышенное распознавание неканонических PAM-

последовательностей, но механизм их действия еще не исследован.

[1045] В качестве примеров мутации второго типа, результаты анализа тепловых карт на ФИГ. 34A-E, 35A-E и 36A-E применяли для идентификации мутаций, которые улучшают общую скорость расщепления по сравнению с CasX 515, но не обязательно влияют на специфическое распознавание PAM-последовательности ДНК. Например, вариант CasX 515, заключающийся во вставке аргинина в положении 27, согласно измерениям, имеет значение обогащения, превышающее один, при отборе с помощью спейсера 11.2 (для PAM-последовательности CTC) и спейсера 23.19 (для PAM-последовательности ATC). Ранее этот вариант был идентифицирован с помощью сравнительного отбора со спейсером для PAM-последовательности CTC, где эта мутация характеризовалась обогащением на несколько порядков величины (данные не представлены). В вышеуказанной структурной модели положение этой аминокислотной мутации физически приближено (9,29 ангстрема) к целевой цепи ДНК в положении -1. Эти оценки предполагают механизм, в котором зрелая R-петля, образованная CasX RNP с двухцепочечной ДНК, стабилизируется боковой цепью аргинина, возможно, с помощью ионных взаимодействий положительно заряженной боковой цепи с отрицательно заряженной основной цепью целевой цепи ДНК. Такое взаимодействие благоприятно для общей кинетики расщепления, и при этом не изменяет специфичности в отношении PAM. Эти данные подтверждают вывод о том, что некоторые мутации с обогащением, показанные на ФИГ. 34A-E, 35A-E и 36A-E, представляют варианты, которые улучшают общую активность расщепления CasX 515 посредством физического взаимодействия либо с одной, либо с обеими цепями ДНК, когда они физически приближены к ним (в пределах десяти ангстрем).

[1046] Данные подтверждают вывод о том, что многие мутации, которые, согласно измерениям, улучшают расщепление по последовательностям, ассоциированным с PAM-последовательностями CTC или ATC, и были идентифицированы по тепловым картам на ФИГ. 34A-E, 35A-E и 36A-E, могут быть классифицированы как принадлежащие к одному из двух типов мутаций, указанных выше. Что касается мутаций первого типа, то варианты с мутациями в положении 223 с высокой оценкой обогащения, полученной с использованием по меньшей мере одного из протестированных спейсеров для PAM-последовательности CTC, перечислены в таблице 21 с соответствующей максимальной оценкой обогащения. Для мутаций второго типа из тысяч вариантов с обогащением был систематически выбран меньший список мутаций. Для идентификации мутаций, которые с высокой вероятностью улучшат общую активность расщепления по сравнению с CasX 515, применяли следующий подход. Во-первых, были отфильтрованы мутации, которые имели наиболее стабильное обогащение при использовании спейсеров для PAM-последовательностей CTC или ATM. Для оценки обогащения каждой мутации для каждого спейсера была определена нижняя граница (НГ). НГ определяли как объединенное значение $\log_2$ оценки обогащения в трех биологических повторностях минус стандартное отклонение $\log_2$ оценок обогащения для отдельных повторностей. Во-вторых, было отобрано

подмножество этих мутаций, у которых НГ > 1 для по меньшей мере двух из трех независимых экспериментальных наборов данных (один отбор с PAM-последовательностью ATC и два отбора с PAM-последовательностью CTC). В-третьих, это подмножество мутаций дополнительно уменьшали, исключив те из них, для которых было получено отрицательное значение $\log_2$ обогащения в любом из трех отборов с PAM-последовательностью TTC. Наконец, отдельные мутации вручную выбирали на основании комбинации структурных признаков и сильной оценки обогащения в по меньшей мере одном эксперименте. Полученные 274 мутации, удовлетворяющие этим критериям, перечислены в таблице 22 вместе с максимальным наблюдаемым значением $\log_2$ оценки обогащения по результатам двух экспериментов с PAM-последовательностью CTC или одного эксперимента с PAM-последовательностью ATC, представленных на тепловых картах ФИГ. 34A-E, 35A-E и 36A-E; также указан домен, в котором расположена мутация.

[1047] В отличие от мутаций класса I, существует другая категория мутаций, которые улучшают способность CasX RNP различать целевые и нецелевые сайты в геномной ДНК, которые определяются спейсерной последовательностью; это мутации класса II, которые улучшают нуклеазную активность белка CasX посредством улучшения специфичности спейсера. Были проведены два дополнительных эксперимента для специфического определения мутаций класса II, причем эти эксперименты состояли из отрицательного отбора плазмид и давали в результате оценки обогащения, представляющие чувствительность созданного варианта по сравнению с CasX 515 к одной паре неспаренных оснований между спейсерной последовательностью гидовой РНК и предполагаемой целевой ДНК. Полученные оценки обогащения ранжировали для всех наблюдаемых мутаций в экспериментальных данных и проводили последующий анализ для идентификации подмножества мутаций, вероятно, способных улучшать специфичность спейсера, связанного с белком CasX, без значимого снижения нуклеазной активности в отношении требуемого целевого сайта. Во-первых, мутации из эксперимента Screen 5 были ранжированы по их средней оценке обогащения по результатам трех технических повторностей с помощью спейсера 23.2. Эти мутации, которые были физически приближены к неспаренной паре нуклеотидов, как было определено из опубликованных моделей CasX RNP, связанного с целевым сайтом (PDB ID: 6NY2), удаляли с целью отсечения тех мутаций класса II, которые могут обеспечивать улучшения только специфично для спейсера 23.2, а не универсально для всех спейсеров. Наконец, эти мутации класса II отбрасывали, если мутация отрицательно влияла на активность расщепления по целевым сайтам с PAM-последовательностью TTC, если их среднее значение $\log_2$ обогащения по результатам трех отборов CcdB с PAM-последовательностью TTC было ниже нуля. Полученные мутации, соответствующие этим критериям, перечислены в таблице 23 вместе с максимальным наблюдаемым значением $\log_2$ оценки обогащения по результатам эксперимента Screen 5 и доменом, в котором расположена мутация. Кроме того, мутации класса II были идентифицированы в эксперименте по отрицательному отбору Screen 6. Эти мутации были аналогичным образом ранжированы по их средним оценкам

обогащения, но были использованы другие стадии фильтрации. В частности, мутации были идентифицированы в каждой из следующих категорий: мутации с наивысшими средними оценками обогащения при использовании спейсера 23.2, спейсера 23.11 или спейсера 23.13; мутации с наивысшими комбинированными средними оценками обогащения при использовании спейсера 23.2 и спейсера 23.11; мутации с наивысшими комбинированными средними оценками обогащения при использовании спейсера 23.11 и спейсера 23.13 или мутации с наивысшими комбинированными средними оценками обогащения при использовании спейсера 23.2 в эксперименте Screen 5 и спейсера 23.2 в эксперименте Screen 6. Эти полученные мутации перечислены в таблице 23 вместе с максимальным наблюдаемым значением $\log_2$ оценки обогащения по результатам эксперимента Screen 6 и доменом, в котором расположена мутация.

[1048] В дополнение к мутациям класса I или класса II существует еще одна категория мутаций, в отношении которых непосредственно наблюдалась способность улучшать активность редактирования дцДНК, ассоциированной с PAM-последовательностями TTC. Эти мутации, называемые мутациями класса III, демонстрировали улучшенную нуклеазную активность за счет обеспечения оценок обогащения выше, чем у CasX 515 при нацеливании на плазмиду CcdB с помощью спейсера 23.2 в эксперименте Screen 7. Для идентификации подмножества этих мутаций с высоким обогащением, представляющих особый интерес, применяли стадию компьютерной фильтрации. В частности, были идентифицированы мутации, которые имели среднее значение обогащения по трем повторностям выше нуля для каждого из трех протестированных промоторов. Наконец, особенности оценок обогащения по всей длине аминокислотной последовательности применяли для идентификации дополнительных мутаций в положениях с обогащением. Примеры интересующих особенностей включали в себя следующие: вставки или делеции в областях соединения доменов белка с целью облегчения топологических изменений; замены аминокислоты на пролин для введения изгиба в полипептидную основную цепь; замены аминокислоты на положительно заряженную аминокислоту для добавления ионной связи между белком и отрицательно заряженной основной цепью нуклеиновой кислоты - либо гидовой РНК, либо любой из цепей целевой ДНК; делеции аминокислоты, при которых обе последовательные делеции характеризуются высоким обогащением; замены в положении, которое содержит множество замен, характеризующихся высоким обогащением; замены аминокислоты на аминокислоту, характеризующуюся высоким обогащением, на самом крае N-конца белка. Эти полученные мутации перечислены в таблице 24 вместе с максимальным наблюдаемым значением $\log_2$ оценки обогащения по результатам эксперимента Screen 6 и доменом, в котором расположена мутация.

Таблица 21. Мутации в CasX 515 (SEQ ID NO: 416), которые улучшают активность расщепления последовательностей, ассоциированных с PAM-последовательностями CTC, посредством физического взаимодействия с нуклеотидами PAM-последовательности ДНК

Положение	Референсный	Альтернативный	Максимальное наблюдаемое значение log_2 обогащения в отборах Ccdb	Домен
223	G	Y	4,6	спиральный I-II
223	G	N	5,7	спиральный I-II
223	G	H	4,2	спиральный I-II
223	G	S	4,6	спиральный I-II
223	G	T	3,8	спиральный I-II
223	G	A	6,3	спиральный I-II
223	G	V	3,6	спиральный I-II

Таблица 22. Мутации в CasX 515 (SEQ ID NO: 416), целенаправленно идентифицированные во всех наборах данных по улучшению активность расщепления последовательностей, ассоциированных с PAM-последовательностями ATC и CTC

Положение	Референсный	Альтернативный	Максимальное наблюдаемое значение log_2 обогащения в отборах CcdB	Домен
3	-	G	3,0	OBD-I
3	I	G	3,5	OBD-I
3	I	E	4,5	OBD-I
4	-	G	2,5	OBD-I
4	K	G	2,5	OBD-I
4	K	P	3,1	OBD-I
4	K	S	3,3	OBD-I
4	K	W	2,8	OBD-I
5	-	P	3,5	OBD-I
5	-	G	3,1	OBD-I
5	R	S	3,7	OBD-I
5	-	S	2,2	OBD-I
5	R	A	3,2	OBD-I
5	R	P	3,6	OBD-I
5	R	G	3,2	OBD-I
5	R	L	2,7	OBD-I
6	I	A	3,3	OBD-I
6	-	G	3,7	OBD-I
7	N	Q	3,1	OBD-I

7	N	L	2,7	OBD-I
7	N	S	3,7	OBD-I
8	K	G	3,3	OBD-I
15	K	F	3,0	OBD-I
16	D	W	2,8	OBD-I
16	-	F	4,2	OBD-I
18	-	F	3,5	OBD-I
28	M	H	2,5	OBD-I
33	V	T	2,0	OBD-I
34	R	P	3,6	OBD-I
36	M	Y	2,4	OBD-I
41	R	P	2,2	OBD-I
47	L	P	2,2	OBD-I
52	E	P	3,2	OBD-I
55	-	P	2,7	OBD-I
55	PQ	--	3,0	OBD-I
56	Q	S	1,9	OBD-I
56	-	D	2,5	OBD-I
56	-	T	2,8	OBD-I
56	Q	P	3,9	OBD-I
58	-	A	2,2	спиральный I-I
63	R	S	3,0	спиральный I-I
63	R	Q	2,7	спиральный I-I
72	D	E	2,7	спиральный I-I
81	L	V	2,8	спиральный I-I
81	L	T	2,7	спиральный I-I
85	W	G	3,2	спиральный I-I
85	W	F	2,7	спиральный I-I
85	W	E	2,9	спиральный I-I
85	W	D	3,1	спиральный I-I
85	W	A	2,8	спиральный I-I
85	W	Q	3,0	спиральный I-I
85	W	R	3,7	спиральный I-I
88	F	M	2,4	спиральный I-I
89	Q	D	2,5	спиральный I-I
93	V	L	1,9	спиральный I-I
109	Q	P	1,8	NTSB

115	E	S	1,8	NTSB
120	G	D	2,4	NTSB
133	G	T	2,2	NTSB
141	L	A	2,2	NTSB
168	L	K	3,1	NTSB
170	A	Y	2,7	NTSB
170	A	S	1,7	NTSB
175	E	A	2,0	NTSB
175	E	D	2,8	NTSB
175	E	P	3,8	NTSB
223	G	-	1,4	спиральный I-II
223	G	S	8,8	спиральный I-II
223	G	T	3,7	спиральный I-II
242	S	T	1,9	спиральный I-II
247	I	T	1,8	спиральный I-II
254	V	T	2,5	спиральный I-II
265	L	Y	1,9	спиральный I-II
288	K	G	4,2	спиральный I-II
288	K	S	4,0	спиральный I-II
291	V	L	2,6	спиральный I-II
303	M	T	2,3	спиральный I-II
303	M	W	2,7	спиральный I-II
328	G	N	3,3	спиральный I-II
331	S	Q	2,7	спиральный I-II
334	-	A	2,3	спиральный II
334	LV	--	3,0	спиральный II
335	V	E	2,8	спиральный II
335	V	Q	2,7	спиральный II
335	V	F	2,5	спиральный II
335	V	-	3,2	спиральный II
336	E	P	2,9	спиральный II
336	E	-	3,1	спиральный II
336	E	D	2,7	спиральный II
336	E	L	2,4	спиральный II
336	E	R	2,7	спиральный II
337	R	N	2,5	спиральный II
338	Q	V	2,5	спиральный II

338	-	Q	3,0	спиральный II
339	-	G	2,6	спиральный II
341	-	H	2,9	спиральный II
341	-	A	2,0	спиральный II
342	V	D	2,7	спиральный II
342	-	T	2,3	спиральный II
342	V	-	3,0	спиральный II
342	-	F	2,5	спиральный II
343	-	D	3,3	спиральный II
343	D	-	2,0	спиральный II
344	W	-	3,1	спиральный II
344	W	T	2,8	спиральный II
344	W	H	2,8	спиральный II
344	-	P	3,0	спиральный II
344	-	G	2,6	спиральный II
345	-	R	3,2	спиральный II
345	W	P	3,1	спиральный II
345	W	D	2,3	спиральный II
345	-	D	2,9	спиральный II
345	W	L	2,3	спиральный II
346	-	P	2,4	спиральный II
346	-	D	2,9	спиральный II
347	M	-	2,6	спиральный II
348	-	T	3,3	спиральный II
350	N	I	2,3	спиральный II
351	V	N	2,8	спиральный II
351	V	H	3,1	спиральный II
352	K	D	2,2	спиральный II
354	L	D	3,1	спиральный II
355	I	S	2,6	спиральный II
357	E	C	2,1	спиральный II
357	E	P	2,8	спиральный II
358	K	T	2,8	спиральный II
359	K	E	2,7	спиральный II
363	K	L	3,3	спиральный II
363	K	Y	2,2	спиральный II
367	Q	D	2,8	спиральный II

367	Q	P	3,0	спиральный II
369	-	S	2,6	спиральный II
369	LA	--	2,4	спиральный II
373	K	L	2,2	спиральный II
374	-	R	2,0	спиральный II
397	Y	T	2,5	спиральный II
400	G	M	2,0	спиральный II
402	L	V	2,4	спиральный II
403	L	C	2,3	спиральный II
404	L	D	2,5	спиральный II
404	L	N	2,5	спиральный II
404	L	W	2,3	спиральный II
404	L	Y	2,1	спиральный II
407	E	F	2,6	спиральный II
407	E	L	2,2	спиральный II
407	E	Y	2,6	спиральный II
411	G	P	2,6	спиральный II
411	-	E	3,2	спиральный II
413	-	T	2,7	спиральный II
413	-	R	2,4	спиральный II
413	-	W	3,0	спиральный II
413	-	Y	3,7	спиральный II
414	-	W	2,6	спиральный II
414	-	Y	3,1	спиральный II
414	W	G	3,0	спиральный II
414	W	R	2,6	спиральный II
416	K	D	2,7	спиральный II
416	K	H	2,0	спиральный II
416	K	P	2,6	спиральный II
416	K	T	2,3	спиральный II
417	V	L	2,6	спиральный II
417	V	A	2,5	спиральный II
418	Y	C	2,7	спиральный II
419	D	G	3,2	спиральный II
419	D	M	2,4	спиральный II
419	D	P	2,4	спиральный II
425	I	C	2,2	спиральный II

427	K	T	2,4	спиральный II
428	K	R	2,5	спиральный II
430	E	G	1,9	спиральный II
432	L	A	1,9	спиральный II
434	K	H	2,2	спиральный II
436	I	T	2,4	спиральный II
436	I	S	3,0	спиральный II
436	I	Q	2,7	спиральный II
437	K	D	3,1	спиральный II
442	R	D	2,5	спиральный II
442	R	-	2,7	спиральный II
446	D	E	2,3	спиральный II
446	-	D	2,3	спиральный II
450	K	P	2,3	спиральный II
452	A	R	2,0	спиральный II
453	L	T	3,2	спиральный II
456	W	L	2,2	спиральный II
457	L	C	2,2	спиральный II
459	A	L	2,0	спиральный II
461	A	T	2,7	спиральный II
461	A	K	2,1	спиральный II
465	I	E	3,1	спиральный II
465	-	C	2,9	спиральный II
466	-	S	3,5	спиральный II
466	-	G	2,5	спиральный II
467	-	R	2,4	спиральный II
467	G	P	2,0	спиральный II
468	L	K	3,6	спиральный II
468	L	D	3,2	спиральный II
468	L	S	3,0	спиральный II
468	L	H	3,3	спиральный II
470	E	-	2,4	спиральный II
472	D	R	2,2	спиральный II
472	-	D	2,4	спиральный II
473	-	P	2,6	спиральный II
474	-	D	2,7	спиральный II
475	EF	--	2,8	спиральный II

475	-	Q	2,7	спиральный II
476	F	K	2,8	спиральный II
476	F	-	2,2	спиральный II
477	-	G	2,8	спиральный II
479	C	D	3,1	спиральный II
480	-	V	2,2	спиральный II
480	E	D	2,3	спиральный II
481	-	H	2,2	спиральный II
481	L	R	2,9	спиральный II
482	K	R	2,1	спиральный II
483	L	H	2,7	спиральный II
484	Q	C	2,1	спиральный II
485	K	P	3,0	спиральный II
490	L	S	2,8	спиральный II
498	E	L	2,1	спиральный II
499	-	F	1,6	спиральный II
511	K	T	6,8	OBD-II
524	-	P	2,4	OBD-II
553	-	S	2,4	OBD-II
558	-	R	1,9	OBD-II
570	M	T	2,7	OBD-II
582	I	T	1,9	OBD-II
592	Q	I	2,1	OBD-II
592	Q	F	2,8	OBD-II
592	Q	V	2,0	OBD-II
592	Q	A	2,9	OBD-II
641	-	R	2,3	OBD-II
643	-	D	2,7	OBD-II
644	-	W	2,5	OBD-II
645	-	A	2,4	OBD-II
650	-	I	2,5	RuvC-I
651	-	S	2,4	RuvC-I
652	-	T	2,4	RuvC-I
652	-	N	2,3	RuvC-I
653	-	R	2,3	RuvC-I
653	-	K	2,2	RuvC-I
654	-	H	2,2	RuvC-I

654	-	S	2,3	RuvC-I
658	V	L	1,9	RuvC-I
695	G	W	1,4	RuvC-I
695	G	R	3,5	RuvC-I
708	K	S	3,0	RuvC-I
708	K	T	2,9	RuvC-I
708	K	E	3,1	RuvC-I
711	V	A	1,6	RuvC-I
726	K	E	2,0	RuvC-I
729	N	G	2,8	RuvC-I
736	R	H	2,7	RuvC-I
736	R	G	2,4	RuvC-I
771	M	S	3,7	RuvC-I
771	M	A	3,3	RuvC-I
792	L	F	2,5	RuvC-I
868	V	D	1,9	TSL
877	-	A	2,0	TSL
886	T	E	1,8	TSL
886	T	D	2,5	TSL
886	T	N	1,6	TSL
888	G	D	2,5	TSL
890	S	-	3,0	TSL
891	G	-	2,7	TSL
892	-	E	2,0	TSL
892	-	N	2,9	TSL
895	S	I	1,7	TSL
908	E	D	1,7	TSL
932	S	M	2,5	RuvC-II
932	S	V	2,6	RuvC-II
944	-	L	1,4	RuvC-II
947	-	G	1,9	RuvC-II
949	T	-	1,9	RuvC-II
951	G	I	3,7	RuvC-II

Таблица 23. Мутации в CasX 515 (SEQ ID NO: 416), целенаправленно идентифицированные во всех наборах данных по улучшению специфичности спейсера

Положение	**Референсный**	**Альтернативный**	**Максимальное наблюдаемое значение log_2**	**Домен**

			обогащения в отрицательных отборах	
6	I	L	2,25	OBD-I
48	-	P	2	OBD-I
87	-	G	3,96	спиральный I-I
90	K	V	4,84	спиральный I-I
155	F	V	2,13	NTSB
215	-	T	2,03	спиральный I-II
216	-	C	3,03	спиральный I-II
220	Y	F	2,1	спиральный I-II
264	S	H	3,16	спиральный I-II
329	-	Q	2,71	спиральный I-II
343	D	S	2,69	спиральный II
346	DM	--	2,96	спиральный II
349	-	P	2,06	спиральный II
357	-	G	2,11	спиральный II
375	QE	--	2,34	спиральный II
378	L	N	2,38	спиральный II
389	K	Q	2,29	спиральный II
417	-	L	2,75	спиральный II
441	E	L	2,36	спиральный II
458	R	D	2,2	спиральный II
459	A	E	2,65	спиральный II
476	FC	--	2,34	спиральный II
503	IL	--	2,15	OBD-II
537	K	G	2,85	OBD-II
621	L	T	2,45	OBD-II
624	-	A	3	OBD-II
783	L	Y	2,08	RuvC-I
783	-	P	2,6	RuvC-I
787	L	-	2,49	RuvC-I
787	L	R	3,58	RuvC-I
787	L	D	5,58	RuvC-I
788	-	Q	2,65	RuvC-I
789	-	R	2,5	RuvC-I
789	-	N	2,71	RuvC-I
790	E	N	2,45	RuvC-I

792	-	P	2,85	RuvC-I
793	P	A	2,93	RuvC-I
795	K	Q	2,45	RuvC-I
796	T	V	2,75	RuvC-I
798	-	R	4,07	RuvC-I
799	-	H	2,79	RuvC-I
801	T	Q	3,16	RuvC-I
801	-	H	3,34	RuvC-I
801	-	R	2,86	RuvC-I
802	-	L	2,88	RuvC-I
802	L	-	2,87	RuvC-I
802	-	W	3,08	RuvC-I
803	-	A	3,19	RuvC-I
803	-	F	3,14	RuvC-I
803	A	S	5,79	RuvC-I
804	Q	K	3,05	RuvC-I
805	Y	-	3,29	RuvC-I
806	T	Y	3,07	RuvC-I
806	T	F	2,49	RuvC-I
807	-	I	3,21	RuvC-I
807	S	P	2,61	RuvC-I
809	T	P	3,2	RuvC-I
809	-	N	3,1	RuvC-I
810	C	K	3,19	RuvC-I
810	C	M	3,08	RuvC-I
811	-	M	2,51	TSL
812	N	-	3,07	TSL
812	-	V	2,68	TSL
813	C	S	2,3	TSL
814	-	G	3,15	TSL
814	-	W	3,04	TSL
815	F	P	3,09	TSL
817	-	W	2,87	TSL
828	K	G	1,99	TSL
906	V	C	2,01	TSL

Таблица 24. Мутации в CasX 515 (SEQ ID NO: 416), целенаправленно идентифицированные во всех наборах данных по улучшению активности расщепления в PAM-последовательностях TTC

Положение	**Референс-ный**	**Альтер-нативный**	**Максимальное наблюдаемое значение log_2 обогащения в отборах Ccdb**	**Домен**
4	K	W	3,51	OBD-I
5	R	P	4,01	OBD-I
27	-	P	4,69	OBD-I
28	M	P	3,69	OBD-I
56	Q	P	3,78	OBD-I
85	W	A	3,96	спиральный I-I
102	-	G	4,75	NTSB
104	-	I	4,43	NTSB
104	-	L	4,52	NTSB
130	S	-	4,02	NTSB
151	Y	T	3,46	NTSB
168	L	D	3,32	NTSB
168	L	E	4,08	NTSB
188	K	Q	4,96	NTSB
190	G	Q	4,1	NTSB
223	G	-	1,63	спиральный I-II
235	G	L	4,64	спиральный I-II
235	G	H	4,97	спиральный I-II
239	S	H	3,93	спиральный I-II
239	S	T	4,97	спиральный I-II
245	Q	H	5	спиральный I-II
288	K	D	5,08	спиральный I-II
288	K	E	4,79	спиральный I-II
303	M	R	3,71	спиральный I-II
303	M	K	3,29	спиральный I-II
307	L	K	3,55	спиральный I-II
328	G	R	3,91	спиральный I-II

328	G	K	4,58	спиральный I-II
334	-	H	5,65	спиральный II
335	-	D	5,5	спиральный II
335	V	P	5,1	спиральный II
345	-	Q	5,22	спиральный II
441	-	K	5,07	спиральный II
477	C	R	2,94	спиральный II
477	C	K	3,49	спиральный II
502	S	-	4,04	OBD-II
503	I	R	3,72	OBD-II
503	I	K	Не обнаружено	OBD-II
504	L	-	4,24	OBD-II
542	R	E	4,54	OBD-II
563	K	-	3,25	OBD-II
593	-	A	1,83	OBD-II
610	K	Q	3,46	OBD-II
615	R	Q	3,67	OBD-II
643	-	A	2,42	OBD-II
697	S	R	2,67	RuvC-I
697	S	K	2,55	RuvC-I
906	V	T	4,65	TSL

Пример 15. Оценка влияния длины спейсера на редактирование в клетках при доставке в виде RNP

[1049] Цель эксперимента заключалась в определении влияния длины спейсера (нацеливающей последовательности) на редактирование целевой нуклеиновой кислоты с помощью RNP из CasX и гидов, доставляемых внутриклеточно.

[1050] Вариант CasX 491 очищали так, как описано выше. Гидовые РНК с каркасом 174 получали посредством транскрипции *in vitro* (IVT). Матрицы для IVT создавали с помощью ПЦР с использованием полимеразы Q5 (NEB M0491) в соответствии с рекомендованным протоколом, матричных олигонуклеотидов для каждой основной цепи каркаса и праймеров для амплификации с промотором T7 и спейсером либо 15.3 (CAAACAAATGTGTCACAAAG, SEQ ID NO: 165), либо 15.5

(GGAATAATGCTGTTGTTGAA, SEQ ID NO: 166) с полным размером (20 нуклеотидов) или усеченным на один или два нуклеотида с 3'-конца соответствующего спейсера (последовательности в таблице 26). Последовательности праймеров, используемых для создания матриц для IVT, показаны в таблице 25. Затем полученные матрицы применяли с РНК-полимеразой T7 для получения гидовых РНК в соответствии со стандартными протоколами. Гиды очищали с помощью электрофореза в полиакриламидном геле в денатурирующих условиях и заново проводили свертывание перед применением. Отдельные RNP собирали посредством смешивания белка с 1,2-кратным молярным избытком гида в 25 мМ натрий-фосфатном буфере (pH 7,25), содержащем 300 мМ NaCl, 1 мМ $MgCl_2$ и 200 мМ трегалозу. RNP инкубировали при 37 °C в течение 10 минут, а затем очищали с помощью эксклюзионной хроматографии размеров и осуществляли замену на буфер, содержащий 25 мМ натрий-фосфатный буфер (pH 7,25), 150 мМ NaCl, 1 мМ $MgCl_2$ и 200 мМ трегалозу (буфер 1). Концентрацию RNP определяли после очистки с использованием реактива для анализа белков Pierce 660 нм.

[1051] Очищенные RNP тестировали на способность редактировать локус T-клеточного рецептора α (TCRα) в клетках Jurkat. Доставку RNP осуществляли посредством электропорации с использованием системы Lonza 4-D nucleofector. Ресуспендировали 700 000 клеток в 20 мкл буфера Lonza SE и добавляли к RNP, разбавленному в буфере 1 до соответствующей концентрации и конечного объема 2 мкл. Клетки подвергали электропорации с помощью 96-луночной системы Lonza 96-well Shuttle System, используя протокол CL-120. Клетки восстанавливались при 37 °C в предварительно уравновешенной среде RPMI, а затем пробу, обработанную при каждом условии электропорации, разделяли на три лунки 96-луночного планшета. Среду клеток заменяли на свежую среду RPMI через один день после нуклеофекции. На третий день после нуклеофекции клетки окрашивали с помощью меченного Alexa Fluor 647 антитела к TCR α/β (BioLegend) и оценивали потерю поверхностного TCR α/β с использованием проточного цитометра Attune Nxt. Часть клеток Jurkat не дает положительного окрашивания на TCR α/β в отсутствие редактирования. Для учета этого и оценки фактической процентной доли клеток, в которых произошел нокаут TCRα вследствие редактирования, применяли формулу $TCR_{KO=}(TCR\text{-}_{набл.} - TCR\text{-}_{отриц.}) / (1 - TCR\text{-}_{отриц.})$, где TCR_{OK} представляет собой расчетную частоту нокаута TCRα, $TCR\text{-}_{набл.}$ представляет собой наблюдаемую фракцию клеток с отрицательным результатом окрашивания на TCR в экспериментальной пробе, а $TCR\text{-}_{отриц.}$ представляет собой фракцию клеток с отрицательным результатом окрашивания на TCR в контрольной пробе, не содержавшей RNP. Эта формула предполагает, что клетки, которые экспрессируют и не экспрессируют TCRα/β, подвергаются редактированию с одинаковой частотой. Скорректированную долю клеток с нокаутом TCRα наносили на график зависимости от концентрации RNP с использованием Prism. Для каждого спейсера при трех длинах строили кривые зависимости ответа от дозы, используя общие параметры, за исключением EC_{50}. Указанные p-значения представляют собой вероятность того, что кривая зависимости от дозы для спейсера длиной 20 нукл. и такая же кривая для сравниваемого укороченного

спейсера могут быть смоделированы с одним и тем же параметром EC_{50}.

Таблица 25. Олигонуклеотиды, используемые для создания матриц для IVT

Конструкт	ДНК-последовательность
Матричный прямой для каркаса 174	GAAATTAATACGACTCACTATAACTGGCGCTTTTATCTG ATTACTTTGAGAGCCATCACCAGCGACTATGTCGTAGTG GGTAAAGCT (SEQ ID NO: 1088)
Матричный обратный для каркаса 174	CTTTGATGCTCCCTCCGAAGAGGGAGCTTTACCCACTAC GACATAGTCGC (SEQ ID NO: 1089)
Праймер для амплификации T7	GAAATTAATACGACTCACTATA (SEQ ID NO: 169)
Праймер для спейсера 15.3 длиной 20 нукл.	CTTTGTGACACATTTGTTTGCTTTGATGCTCCCTCC (SEQ ID NO: 1093)
Праймер для спейсера 15.3 длиной 19 нукл.	TTTGTGACACATTTGTTTGCTTTGATGCTCCCTCC (SEQ ID NO: 1094)
Праймер для спейсера 15.3 длиной 18 нукл.	TTGTGACACATTTGTTTGCTTTGATGCTCCCTCC (SEQ ID NO: 1095)
Праймер для спейсера 15.5 длиной 20 нукл.	TTCAACAACAGCATTATTCCCTTTGATGCTCCCTCC (SEQ ID NO: 1096)
Праймер для спейсера 15.5 длиной 19 нукл.	TCAACAACAGCATTATTCCCTTTGATGCTCCCTCC (SEQ ID NO: 1097)
Праймер для спейсера 15.5 длиной 18 нукл.	CAACAACAGCATTATTCCCTTTGATGCTCCCTCC (SEQ ID NO: 1098)

Таблица 26. Спейсерные последовательности

Конструкт	Последовательность РНК
Спейсер 15.3 длиной 20 нукл.	CAAACAAAUGUGUCACAAAG (SEQ ID NO: 1099)
Спейсер 15.3 длиной 19 нукл.	CAAACAAAUGUGUCACAAA (SEQ ID NO: 1100)
Спейсер 15.3 длиной 18 нукл.	CAAACAAAUGUGUCACAA (SEQ ID NO: 1101)
Спейсер 15.5 длиной 20 нукл.	GGAAUAAUGCUGUUGUUGAA (SEQ ID NO: 1102)
Спейсер 15.5 длиной 19 нукл.	GGAAUAAUGCUGUUGUUGA (SEQ ID NO: 1103)
Спейсер 15.5 длиной 18 нукл.	GGAAUAAUGCUGUUGUUG (SEQ ID NO: 1104)

[1052] Результаты: RNP CasX собирали с использованием варианта CasX 491 и гидов, состоящих из каркаса 174 со спейсером либо 15.3, либо 15.5, оба из которых нацеливают на константную область гена TCRα. Гиды с полноразмерными спейсерами длиной 20 нукл., а также укороченными спейсерами длиной 19 и 18 нукл. тестировали для определения того, поддерживают ли более короткие спейсеры повышенное редактирование при доставке предварительно собранных RNP методом нуклеофекции для редактирования *ex vivo*. RNP тестировали в 2-кратных разведениях в диапазоне от 0,3125 мкМ до 2,5 мкМ в реакционной смеси для нуклеофекции объемом 22 мкл. Редактирование оценивали с помощью проточной цитометрии через три дня после нуклеофекции. В случае обеих спейсерных последовательностей RNP с укороченными спейсерами в основном редактировали более эффективно, чем RNP со спейсерами длиной 20 нукл., в диапазоне доз

(ФИГ. 37A-B, кривые доза - ответ). В случае спейсера 15.3 спейсеры длиной 18 нукл. и 19 нукл. имели значения EC_{50} 0,225 мкМ и 0,299 мкМ соответственно по сравнению с 1,414 мкМ для спейсера длиной 20 нукл. ($p < 0,0001$ для обоих укороченных вариантов; F-критерий для экстра-суммы квадратов). В случае спейсера 15.5 спейсер длиной 18 нукл. имел EC_{50} 0,519 мкМ по сравнению с 0,938 мкМ у спейсера длиной 20 нукл. ($p=0,0001$), а спейсер длиной 19 нукл. был более похож на спейсер длиной 20 нукл., имея EC_{50} 0,808 мкМ ($p=0,0762$). Несмотря на тот факт, что в случае спейсера 15.3 длиной 19 нукл. редактирование было похоже на редактирование с использованием спейсера длиной 18 нукл., а спейсер 15.5 длиной 19 нукл. был больше похож на соответствующий спейсер длиной 20 нукл., для обоих изученных спейсеров направление тенденции оставалось постоянным и указывает на то, что использование гидов со спейсерами длиной 18 нукл. может представлять собой общую стратегию для усиления редактирования, когда молекула для редактирования CasX доставляется в виде предварительно собранного RNP, и демонстрирует, что более короткие спейсеры длиной 18 или 19 нукл. поддерживают повышенную активность по сравнению со спейсером длиной 20 нукл. при редактировании *ex vivo* с участием RNP.

Пример 16. Анализ позволяет идентифицировать варианты белка CasX со специфичностью в отношении новой PAM-последовательности

[1053] Цель эксперимента заключалась в определении специфичностей в отношении PAM-последовательностей для вариантов последовательностей белка CasX 515 (SEQ ID NO: 416). Для обеспечения этого клеточную линию HEK293 PASS_V1.01 обрабатывали белком CasX 491 (SEQ ID NO:336) или 515 или вариантом белка 515 вместе с гидом 174 (SEQ ID NO: 2238) и проводили секвенирование нового поколения (NGS) для расчета процента редактирования при использовании различных спейсеров и ассоциированных с ними целевых сайтов.

[1054] Материалы и методы. Для анализа вариантов клонированного белка применяли мультиплексный объединенный подход с помощью системы PASS. Вкратце, была создана объединенная клеточная линия HEK, названная PASS_V1.01. Каждая клетка в пределах пула содержала интегрированную в геном одномолекулярную гидовую РНК (огРНК), спаренную со специфическим целевым сайтом. После трансфекции конструктами, экспрессирующими белок, редактирование по специфической мишени специфическим спейсером можно количественно оценивать с помощью NGS. Каждая пара гида и мишени была разработана таким образом, чтобы можно было получать данные, относящиеся к активности, специфичности и нацеливаемости RNP-комплекса CasX - гид.

[1055] Спаренные последовательности спейсер - целевой сайт были синтезированы компанией Twist Biosciences и получены в виде эквимолярного пула олигонуклеотидов. Этот пул амплифицировали методом ПЦР и клонировали методом Golden Gate для создания конечной библиотеки плазмид, названной p77. Каждая плазмида содержала элемент, экспрессирующий огРНК и целевой сайт, вместе с экспрессирующим GFP элементом. Экспрессирующий огРНК элемент состоял из промотора U6, приводящего в действие

транскрипцию каркаса гРНК 174 (SEQ ID NO: 2238), далее следовала спейсерная последовательность, которая должна нацеливать RNP из гида и варианта CasX на парный целевой сайт. Были сконструированы и синтезированы 250 возможных уникальных спаренных синтетических последовательностей спейсер - целевой сайт. Затем из этой библиотеки плазмид получали пул лентивируса с помощью системы LentiX (Takara Bio USA, Inc) в соответствии с инструкциями производителя. После этого полученный вирусный препарат количественно оценивали с помощью количественной полимеразной цепной реакции (количественной ПЦР) и трансдуцировали в стандартную клеточную линию HEK293 при низкой множественности инфекции, чтобы создать однокопийные интегрирования. Затем полученную клеточную линию очищали с помощью цитометрии посредством сортировки клеток с активацией флуоресценции (FACS) для завершения получения PASS_V1.01. Затем эту клеточную линию высевали в шестилуночный планшет и в двух повторностях либо обрабатывали водой, либо трансфицировали 2 мкг плазмиды p67, доставляемой с помощью реагента для трансфекции Lipofectamine (ThermoFisher) в соответствии с инструкциями производителя. Плазмида p67 содержит промотор EF-1α, направляющий экспрессию белка CasX, меченного последовательностью сигнала ядерной локализации SV40. Через два дня обработанные клетки PASS_V1.01 собирали, лизировали и экстрагировали геномную ДНК с помощью набора для выделения геномной ДНК (Zymo Research). Затем геномную ДНК амплифицировали методом ПЦР с помощью специально разработанных праймеров для создания ампликонов, совместимых с NGS компании Illumina, и проводили секвенирование на приборе NextSeq. Прочтения проб демультиплексировали и фильтровали в зависимости от их качества. После этого количественно определяли показатели результатов редактирования (доля прочтений с инделами) для каждой синтетической последовательности спейсер - целевой сайт в обработанных пробах.

[1056] Для оценки специфичности PAM-последовательности к молекуле были категоризированы показатели результатов редактирования для четырех разных PAM-последовательностей. Для целевых сайтов с PAM TTC количественно оценивали 48 разных пар спейсер - целевой сайт; для целевых сайтов с PAM ATC, CTC и GTC количественно оценивали 14, 22 и 11 отдельных целевых сайтов соответственно. Для каждого из этих спейсеров рассчитывали среднюю эффективность редактирования и стандартную ошибку среднего для двух биологических повторов. Затем рассчитывали среднюю эффективность редактирования по четырем категориям PAM-последовательности, а также стандартную ошибку среднего.

[1057] Результаты: В таблице 27 приведены средние значения эффективности редактирования для описанных выше спейсеров при нацеливании с вариантом белка CasX 515, рассчитанные как среднее значение для двух экспериментов. Указаны название спейсера и ассоциированная PAM-последовательность. В таблице 28 перечислены те же данные для варианта белка CasX 534. Кроме того, средняя эффективность редактирования была рассчитана для каждой из четырех категорий PAM-последовательности.

Таблица 27. Средний показатель редактирования, осуществляемого CasX 515 со спейсерами, ассоциированными с PAM-последовательностями TTC, ATC, CTC или GTC

Название спейсера	**Средняя процентная доля редактирования, PAM TTC**	**Средняя процентная доля редактирования, PAM ATC**	**Средняя процентная доля редактирования, PAM CTC**	**Средняя процентная доля редактирования, PAM GTC**
Rho_11.2	32,60	3,03	0,00	4,01
Rho_11.24	44,61	0,49	2,55	0,13
SOD1_8.89	18,29	0,00	0,00	0,00
SOD1_8.57	25,65	6,34	9,01	2,52
DMD_16.30	36,53	3,24	0,68	1,07
DMD_16.32	33,76	1,20	0,01	2,21
BCL11A_21.1	38,17	5,19	11,24	22,85
PCSK9_6.8	28,94	5,48	9,84	2,35
B2M_7.37	48,25	4,31	3,27	4,24
CD70_26.3	23,89	1,71	4,96	0,33
DMD_16.29	1,29	-	-	-
DMD_16.31	21,69	-	-	-
HTT_5.1	32,77	-	-	-
DMPK_20.7	19,89	-	-	-
B2M_7.9	42,44	-	-	-
CD70_26.5	27,52	-	-	-
FSHD_25.1	17,74	-	-	-
PMP22_18.22	19,65	-	-	-
PTBP1_28.1	4,00	-	-	-
VCV000162497_alt	11,40	-	-	-
VCV000037426_alt	7,00	-	-	-
VCV000052759_alt	13,14	-	-	-
VCV000031157_alt	0,00	-	-	-
VCV000217157_alt	17,94	-	-	-

VCV000048146_alt	18,05	-	-	-
VCV000039614_alt	33,14	-	-	-
VCV000217646_alt	37,75	-	-	-
VCV000408199_alt	27,21	-	-	-
VCV000431973_alt	16,05	-	-	-
VCV000004193_alt	2,85	-	-	-
VCV000003920_alt	34,92	-	-	-
VCV000224616_alt	36,90	-	-	-
VCV000004787_alt	34,83	-	-	-
VCV000013115_alt	26,04	-	-	-
VCV000195355_alt	32,05	-	-	-
VCV000190123_alt	36,17	-	-	-
VCV000048535_alt	21,74	-	-	-
VCV000468648_alt	49,12	-	-	-
VCV000000908_alt	32,78	-	-	-
VCV000066016_alt	20,45	-	-	-
VCV000184702_alt	14,97	-	-	-
VCV000003825_alt	27,40	-	-	-
VCV000014243_alt	32,03	-	-	-
VCV000046102_alt	41,95	-	-	-
VCV000050962_alt	47,92	-	-	-
VCV000011908_alt	38,86	-	-	-
VCV000252575_alt	24,77	-	-	-
SOD1_G93A_alt	22,54	-	-	-
C9orf72_29.1	-	20,52	-	-
VCV000092361_alt	-	6,13	-	-
VCV000096602_alt	-	0,56	-	-
DUX4_25.3_alt	-	4,02	-	-
VCV000012379_alt	-	-	0,52	-

VCV000067838_alt	-	-	0,64	-
VCV000246282_alt	-	-	0,00	-
VCV000000157_alt	-	-	2,01	-
VCV000550954_alt	-	-	1,58	-
VCV000189075_alt	-	-	4,63	-
VCV000006756_alt	-	-	2,61	-
VCV000100629_alt	-	-	3,39	-
VCV000372498_alt	-	-	2,97	-
VCV000002019_alt	-	-	1,56	-
HTT_slot_1_alt	-	-	3,13	-
HTT_slot_2_alt	-	-	0,00	-
HTT_slot_3_alt	-	-	-	2,00

Таблица 28. Средний показатель редактирования, осуществляемого CasX 534 со спейсерами, ассоциированными с PAM-последовательностями TTC, ATC, CTC или GTC

Название спейсера	**Средняя процентная доля редактирования, PAM TTC**	**Средняя процентная доля редактирования, PAM ATC**	**Средняя процентная доля редактирования, PAM CTC**	**Средняя процентная доля редактирования, PAM GTC**
Rho_11.2	0,70	0,24	0,25	8,00
Rho_11.24	2,49	3,16	5,29	11,50
SOD1_8.89	0,06	0,04	0,00	0,00
SOD1_8.57	12,46	8,04	6,30	17,45
DMD_16.30	0,14	4,80	0,93	6,86
DMD_16.32	0,60	0,58	0,02	1,64
BCL11A_21.1	3,28	10,46	17,16	38,28
PCSK9_6.8	1,73	7,44	11,18	29,00
B2M_7.37	1,48	19,34	6,92	5,26
CD70_26.3	0,08	3,13	8,24	12,89
DMD_16.29	0,02	-	-	-
DMD_16.31	0,11	-	-	-

HTT_5.1	0,80	-	-	-
DMPK_20.7	0,37	-	-	-
B2M_7.9	8,14	-	-	-
CD70_26.5	5,92	-	-	-
FSHD_25.1	0,27	-	-	-
PMP22_18.22	0,07	-	-	-
PTBP1_28.1	0,03	-	-	-
VCV000162497_alt	0,03	-	-	-
VCV000037426_alt	3,75	-	-	-
VCV000052759_alt	0,00	-	-	-
VCV000031157_alt	48,50	-	-	-
VCV000217157_alt	2,26	-	-	-
VCV000048146_alt	0,14	-	-	-
VCV000039614_alt	0,27	-	-	-
VCV000217646_alt	0,72	-	-	-
VCV000408199_alt	0,59	-	-	-
VCV000431973_alt	0,13	-	-	-
VCV000004193_alt	0,01	-	-	-
VCV000003920_alt	1,40	-	-	-
VCV000224616_alt	0,20	-	-	-
VCV000004787_alt	0,35	-	-	-
VCV000013115_alt	0,00	-	-	-
VCV000195355_alt	0,05	-	-	-
VCV000190123_alt	0,68	-	-	-
VCV000048535_alt	1,10	-	-	-
VCV000468648_alt	2,69	-	-	-
VCV000000908_alt	0,86	-	-	-
VCV000066016_alt	0,00	-	-	-
VCV000184702_alt	0,02	-	-	-

VCV000003825_alt	1,40	-	-	-
VCV000014243_alt	0,94	-	-	-
VCV000046102_alt	0,66	-	-	-
VCV000050962_alt	0,08	-	-	-
VCV000011908_alt	1,05	-	-	-
VCV000252575_alt	1,41	-	-	-
SOD1_G93A_alt	5,13	-	-	-
C9orf72_29.1	-	17,85	-	-
VCV000092361_alt	-	1,82	-	-
VCV000096602_alt	-	1,22	-	-
DUX4_25.3_alt	-	5,50	-	-
VCV000012379_alt	-	-	1,32	-
VCV000067838_alt	-	-	0,88	-
VCV000246282_alt	-	-	0,00	-
VCV000000157_alt	-	-	3,17	-
VCV000550954_alt	-	-	3,12	-
VCV000189075_alt	-	-	4,87	-
VCV000006756_alt	-	-	2,53	-
VCV000100629_alt	-	-	2,60	-
VCV000372498_alt	-	-	3,67	-
VCV000002019_alt	-	-	3,86	-
HTT_slot_1_alt	-	-	4,53	-
HTT_slot_2_alt	-	-	0,03	-
HTT_slot_3_alt	-	-	-	4,21

[1058] На ФИГ. 38 представлена гистограмма, демонстрирующая среднюю эффективность редактирования по категориям PAM и по вариантам белка CasX, при этом стандартная ошибка среднего по двум экспериментам нанесена в виде плашек погрешностей. Эти данные свидетельствуют о том, что оба из CasX 491 и 515 являются специфичными к канонической PAM-последовательности TTC, тогда как другие варианты CasX с большей или меньшей эффективностью действовали в присутствии протестированных PAM-последовательностей. В частности, наблюдалось, что разные

аминокислотные замены в положении 223 CasX поддерживали нахождение разных оснований в положении непосредственно на 5'-конце нуклеотидной последовательности PAM. Вариант CasX с самой высокой активностью редактирования приведен для каждой из этих четырех PAM-последовательностей в таблице 29. Указаны название варианта и аминокислотная замена, а также кратность улучшения редактирования относительно CasX 515 для указанной PAM-последовательности. Эти данные свидетельствуют о том, что характеристики боковой цепи аминокислоты в данном положении имеет критическое значение для распознавания PAM. Вероятно, каждый из четырех нуклеотидов является наиболее комплементарным своей сети связей или химическому окружению в этой области белка CasX. Эти данные можно применять для конструирования максимально активных терапевтических молекул CasX для интересующей целевой ДНК-последовательности.

[1059] В условиях экспериментов был идентифицирован набор вариантов белка CasX 515, которые более эффективно расщепляют двухцепочечную ДНК в человеческих клетках в целевых ДНК-последовательностях, ассоциированных с PAM-последовательностью TTC, ATC, CTC или GTC, что свидетельствует в пользу возможности создания вариантов CasX с расширенным спектром специфичности к PAM относительно CasX дикого типа.

Таблица 29. Мутации в CasX 515, которые улучшают активность расщепления в присутствии PAM-последовательностей NTC

Белок CasX	**Положение**	**Референсный**	**Альтернативный**	**PAM-последовательность**	**Кратность улучшения по сравнению с 515**
515	223	G	н/д	TTC	1,0
528	223	G	Y	ATC	6,7
535	223	G	S	CTC	5,2
534	223	G	H	GTC	3,2

Пример 17. Анализ PASS для сравнения вариантов белка CasX с улучшенной эффективностью редактирования в присутствии PAM-последовательностей TTC, ATC, CTC или GTC

[1060] Цель эксперимента заключалась в сравнении эффективности редактирования генома в человеческих клетках с использованием сконструированных и новых вариантов белка CasX 2 (SEQ ID NO: 2). Для обеспечения этого клеточную линию HEK293 PASS_V1.00 обрабатывали референсным белком CasX 2 дикого типа или сконструированными вариантами 119 (SEQ ID NO: 270), 491 (SEQ ID NO: 336) и 515 (SEQ ID NO: 416) или вариантами последовательности 515 и проводили секвенирование нового поколения (NGS) для расчета процента редактирования при использовании различных спейсеров и ассоциированных с ними целевых сайтов.

[1061] Материалы и методы. Для анализа вариантов клонированного белка

применяли мультиплексный объединенный подход с помощью системы PASS. Вкратце, была создана объединенная клеточная линия HEK, названная PASS_V1.00. Каждая клетка в пределах пула содержала интегрированную в геном последовательность, кодирующую одномолекулярную гидовую РНК (огРНК; SEQ ID NO: 2238), спаривающуюся со специфическим целевым сайтом. После трансфекции конструктами, экспрессирующими белок, редактирование по специфической мишени специфическим спейсером можно количественно оценивать с помощью NGS. Каждая пара гида и мишени была разработана таким образом, чтобы можно было получать данные, относящиеся к активности, специфичности и нацеливаемости RNP-комплекса CasX - гид.

[1062] Спаренные последовательности спейсер - целевой сайт были синтезированы компанией Twist Biosciences и получены в виде эквимолярного пула олигонуклеотидов. Этот пул амплифицировали методом ПЦР и клонировали методом Golden Gate для создания конечной библиотеки плазмид, названной p66. Каждая плазмида содержала элемент, экспрессирующий огРНК, и целевой сайт вместе с экспрессирующим GFP элементом и элементом, экспрессирующим фактор устойчивости к гигромицину. Экспрессирующий огРНК элемент состоял из промотора U6, приводящего в действие транскрипцию каркаса гРНК 174 (SEQ ID NO: 2238), далее следовала спейсерная последовательность, которая должна нацеливать RNP из гида и варианта CasX на парный целевой сайт. Были сконструированы и синтезированы 250 возможных уникальных спаренных синтетических последовательностей спейсер - целевой сайт. Затем из этой библиотеки плазмид получали пул лентивируса с помощью системы LentiX (Takara Bio USA, Inc) в соответствии с инструкциями производителя. После этого полученный вирусный препарат количественно оценивали с помощью количественной полимеразной цепной реакции (количественной ПЦР) и трансдуцировали в стандартную клеточную линию HEK293 при низкой множественности инфекции, чтобы создать однокопийные интегрирования. Затем полученную клеточную линию очищали с помощью цитометрии посредством сортировки клеток с активацией флуоресценции (FACS) для завершения получения PASS_V1.00. После этого эту клеточную линию высевали в шестилуночный планшет и либо обрабатывали водой, либо трансфицировали 2 мкг плазмиды p67, доставляемой с помощью реагента для трансфекции Lipofectamine (ThermoFisher) в соответствии с инструкциями производителя. Плазмида p67 содержит промотор EF-1α, направляющий экспрессию белка CasX, меченного последовательностью сигнала ядерной локализации SV40. Через пять дней обработанные клетки PASS_V1.00 собирали, лизировали, и геномную ДНК экстрагировали с помощью набора для выделения геномной ДНК (Zymo Research). Затем геномную ДНК амплифицировали методом ПЦР с помощью специально разработанных праймеров для создания ампликонов, совместимых с NGS компании Illumina, и проводили секвенирование на приборе NextSeq. Прочтения проб демультиплексировали и фильтровали в зависимости от их качества. После этого количественно определяли показатели результатов редактирования (доля прочтений с инделами) для каждой синтетической последовательности спейсер - целевой сайт в обработанных пробах.

[1063] Для оценки эффективности редактирования для молекулы были категоризированы показатели результатов редактирования для четырех разных PAM-последовательностей. Для целевых сайтов с PAM TTC количественно оценивали 48 разных пар спейсер - целевой сайт; для целевых сайтов с PAM ATC, CTC и GTC количественно оценивали 14, 22 и 11 отдельных целевых сайтов соответственно. Для каждого из этих спейсеров рассчитывали эффективность редактирования и нормализовали относительно фонового сигнала посредством вычитания показателя редактирования, наблюдаемого в пробе, обработанной водой.

[1064] Результаты: На ФИГ. 30 представлен скрипичный график, демонстрирующий эффективность редактирования по разным категориям PAM и вариантам белка CasX. Эффективность редактирования отдельного спейсера показана в виде точки, а средняя эффективность редактирования указана в виде горизонтального отрезка. Клетки, подвергнутые липофекции любой нуклеазой, демонстрировали более высокое среднее редактирование в целевых сайтах с PAM TTC (горизонтальный отрезок), чем в случае нуклеазы дикого типа CasX 2, за исключением CasX 528 (SEQ ID NO: 428). Относительное предпочтение любой данной нуклеазы к четырем разным PAM-последовательностям также представлено на скрипичных графиках. Нуклеазы CasX 527 (SEQ ID NO: 427), 528 (SEQ ID NO: 428) и 529 (SEQ ID NO: 429) демонстрируют предпочтения PAM, по существу отличающиеся от таковых у нуклеазы CasX 2 дикого типа. Эти данные свидетельствуют о том, что определенные области аминокислотной последовательности CasX имеют критическое значение для распознавания PAM, и эти данные можно применять для конструирования максимально активных терапевтических молекул CasX для интересующей целевой ДНК-последовательности.

[1065] В условиях эксперимента были идентифицированы сконструированные варианты CasX 119, 491 и 515 или варианты последовательности 515, которые обеспечивают улучшенное расщепление двухцепочечной ДНК в человеческих клетках по целевым ДНК-последовательностям, ассоциированным с PAM-последовательностями TTC, ATC, CTC или GTC, по сравнению с белком CasX 2 дикого типа. Эти данные подтверждают возможность создания вариантов CasX с расширенным спектром специфичности к PAM относительно референсного CasX дикого типа.

Пример 18. Анализ PASS для сравнения вариантов белка CasX с улучшенной максимальной эффективностью редактирования в присутствии PAM-последовательностей TTC, ATC, CTC или GTC

[1066] Цель эксперимента заключалась в сравнении вариантов белка CasX для определения максимальной эффективности редактирования в присутствии PAM-последовательностей TTC, ATC, CTC или GTC по избранным последовательностям целевых нуклеиновых кислот. Для обеспечения этого клеточную линию HEK293 PASS_V1.00 обрабатывали сконструированным белком CasX 491 (SEQ ID NO: 336) или дополнительно сконструированными вариантами CasX 532 (SEQ ID NO: 432) или 533 (SEQ ID NO: 433) и проводили секвенирование нового поколения (NGS) для расчета процента

редактирования при использовании различных спейсеров и ассоциированных с ними сайтов целевых нуклеиновых кислот.

[1067] Материалы и методы. Для анализа вариантов клонированного белка применяли мультиплексный объединенный подход с помощью системы PASS. Вкратце, была создана объединенная клеточная линия HEK, названная PASS_V1.00. Каждая клетка в пределах пула содержала интегрированную в геном одномолекулярную гидовую РНК (огРНК), спаренную со специфическим целевым сайтом. После трансфекции конструктами, экспрессирующими белок, редактирование по специфической мишени специфическим спейсером можно количественно оценивать с помощью NGS. Каждая пара гида и мишени была разработана таким образом, чтобы можно было получать данные, относящиеся к активности, специфичности и нацеливаемости RNP-комплекса CasX - гид.

[1068] Спаренные последовательности спейсер - целевой сайт были синтезированы компанией Twist Biosciences и получены в виде эквимолярного пула олигонуклеотидов. Этот пул амплифицировали методом ПЦР и клонировали методом Golden Gate для создания конечной библиотеки плазмид, названной p66. Каждая плазмида содержала элемент, экспрессирующий огРНК, и целевой сайт вместе с экспрессирующим GFP элементом и элементом, экспрессирующим фактор устойчивости к гигромицину. Экспрессирующий огРНК элемент состоял из промотора U6, приводящего в действие транскрипцию каркаса гРНК 174 (SEQ ID NO: 2238), далее следовала спейсерная последовательность, которая должна нацеливать RNP из гида и варианта CasX на парный целевой сайт. Были сконструированы и синтезированы 250 возможных уникальных спаренных синтетических последовательностей спейсер - целевой сайт. Затем из этой библиотеки плазмид получали пул лентивируса с помощью системы LentiX (Takara Bio USA, Inc) в соответствии с инструкциями производителя. После этого полученный вирусный препарат количественно оценивали с помощью количественной полимеразной цепной реакции (количественной ПЦР) и трансдуцировали в стандартную клеточную линию HEK293 при низкой множественности инфекции, чтобы создать однокопийные интегрирования. Затем полученную клеточную линию очищали с помощью цитометрии посредством сортировки клеток с активацией флуоресценции (FACS) для завершения получения PASS_V1.00. Затем эту клеточную линию высевали в шестилуночный планшет и в двух повторностях либо обрабатывали водой, либо трансфицировали 2 мкг плазмиды p67, доставляемой с помощью реагента для трансфекции Lipofectamine (ThermoFisher) в соответствии с инструкциями производителя. Плазмида p67 содержит промотор EF-1α, направляющий экспрессию белка CasX, меченного последовательностью сигнала ядерной локализации SV40. Через пять дней обработанные клетки PASS_V1.00 собирали, лизировали, и геномную ДНК экстрагировали с помощью набора для выделения геномной ДНК (Zymo Research). Затем геномную ДНК амплифицировали методом ПЦР с помощью специально разработанных праймеров для создания ампликонов, совместимых с NGS компании Illumina, и проводили секвенирование на приборе NextSeq. Прочтения проб демультиплексировали и фильтровали в зависимости от их качества. После этого количественно определяли

показатели результатов редактирования (доля прочтений с инделами) для каждой синтетической последовательности спейсер - целевой сайт в обработанных пробах.

[1069] Для оценки эффективности редактирования для молекулы были категоризированы показатели результатов редактирования для четырех разных PAM-последовательностей. Для целевых сайтов с PAM TTC количественно оценивали 48 разных пар спейсер - целевой сайт; для целевых сайтов с PAM ATC, CTC и GTC количественно оценивали 14, 22 и 11 отдельных целевых сайтов соответственно. Для каждого из этих спейсеров рассчитывали эффективность редактирования и нормализовали относительно фонового сигнала посредством вычитания показателя редактирования, наблюдаемого в пробе, обработанной водой. Затем рассчитывали среднее значение для двух биологических повторностей, а также стандартную ошибку среднего (станд. ош. среднего). Наконец, для каждой категории PAM (TTC, ATC, CTC или GTC) рассчитывали среднюю эффективность редактирования для всех спейсеров, а также распространяющуюся станд. ош. среднего.

[1070] Результаты: На ФИГ. 31 представлена гистограмма, демонстрирующая эффективность редактирования по категориям PAM и по вариантам белка CasX. Клетки, подвергнутые липофекции Cas 532 или 533, демонстрировали более высокое среднее редактирование, чем CasX 491, в случае каждой из PAM-последовательностей, за исключением Cas 533 и целевых сайтов с PAM-последовательностью TTC. Эти данные свидетельствуют о том, что варианты CasX 532 или 533 могут улучшать активность редактирования в интересующих терапевтических мишенях.

[1071] Было выявлено, что сконструированные варианты CasX 532 и 533 обладают улучшенной эффективностью редактирования в человеческих клетках по сравнению с CasX 491, когда целевые ДНК-последовательности ассоциированы с PAM-последовательностью TTC, ATC, CTC или GTC. Эти данные подтверждают, что варианты CasX 532 и 533 могут улучшать эффективность редактирования в интересующих терапевтических мишенях по сравнению с CasX 491, в особенности в случае тех мишеней, которые ассоциированы с неканонической PAM-последовательностью.

Пример 19. Анализ PASS для идентификации вариантов белка CasX с повышенной активностью редактирования относительно CasX 491 или CasX 119

[1072] Цель эксперимента заключалась в идентификации вариантов CasX с улучшенным редактированием в человеческих клетках относительно CasX 491 или 119. Для обеспечения этого клеточную линию HEK293 PASS_V1.01 обрабатывали белком CasX 2 дикого типа или сконструированными вариантами белка CasX 119 или 491 (SEQ ID NO: 336) или другим вариантом белка CasX и проводили секвенирование нового поколения (NGS) для расчета процента редактирования при использовании различных спейсеров и ассоциированных с ними целевых сайтов.

[1073] Материалы и методы. Для анализа вариантов клонированного белка применяли мультиплексный объединенный подход с помощью системы PASS. Вкратце, была создана объединенная клеточная линия HEK, названная PASS_V1.01. Каждая клетка в пределах пула содержала интегрированную в геном последовательность, кодирующую

одномолекулярную гидовую РНК (огРНК: SEQ ID NO: 2238), спаренную со специфическим целевым сайтом (перечислены в таблице 30). После трансфекции конструктами, экспрессирующими белок, редактирование по специфической мишени специфическим спейсером можно количественно оценивать с помощью NGS. Каждая пара гида и мишени была разработана таким образом, чтобы можно было получать данные, относящиеся к активности, специфичности и нацеливаемости RNP-комплекса CasX - гид.

[1074] Спаренные последовательности спейсер - целевой сайт были синтезированы компанией Twist Biosciences и получены в виде эквимолярного пула олигонуклеотидов. Этот пул амплифицировали методом ПЦР и клонировали методом Golden Gate для создания конечной библиотеки плазмид, названной p77. Каждая плазмида содержала элемент, экспрессирующий огРНК и целевой сайт, вместе с экспрессирующим GFP элементом. Экспрессирующий огРНК элемент состоял из промотора U6, приводящего в действие транскрипцию каркаса гРНК 174 (SEQ ID NO: 2238), далее следовала спейсерная последовательность, которая должна нацеливать RNP из гида и варианта CasX на парный целевой сайт. Были сконструированы и синтезированы 250 возможных уникальных спаренных синтетических последовательностей спейсер - целевой сайт. Затем из этой библиотеки плазмид получали пул лентивируса с помощью системы LentiX (Takara Bio USA, Inc) в соответствии с инструкциями производителя. После этого полученный вирусный препарат количественно оценивали с помощью количественной полимеразной цепной реакции (количественной ПЦР) и трансдуцировали в стандартную клеточную линию HEK293 при низкой множественности инфекции, чтобы создать однокопийные интегрирования. Затем полученную клеточную линию очищали с помощью цитометрии посредством сортировки клеток с активацией флуоресценции (FACS) для завершения получения PASS_V1.01. Затем эту клеточную линию высевали в шестилуночный планшет и либо в двух повторностях, либо в одной пробе обрабатывали водой или трансфицировали 2 мкг плазмиды p67, доставляемой с помощью реагента для трансфекции Lipofectamine (ThermoFisher) в соответствии с инструкциями производителя. Плазмида p67 содержит промотор EF-1α, направляющий экспрессию белка CasX, меченного последовательностью сигнала ядерной локализации SV40, а также ген устойчивости к пуромицину. Через один день клетки переносили в среду, селективную в отношении устойчивости к пуромицину (Sigma). Еще через четыре дня обработанные клетки PASS_V1.01 собирали, лизировали, и геномную ДНК экстрагировали с помощью набора для выделения геномной ДНК (Zymo Research). Затем геномную ДНК амплифицировали методом ПЦР с помощью специально разработанных праймеров для создания ампликонов, совместимых с NGS компании Illumina, и проводили секвенирование на приборе NextSeq. Прочтения проб демультиплексировали и фильтровали в зависимости от их качества. После этого количественно определяли показатели результатов редактирования (доля прочтений с инделами) для каждой синтетической последовательности спейсер - целевой сайт в обработанных пробах.

[1075] Для оценки активности редактирования нуклеазой CasX целевых сайтов в

человеческих клетках проводили количественную оценку 48 целевых сайтов, ассоциированных с PAM-последовательностью TTC. Для каждого из этих спейсеров рассчитывали среднюю эффективность редактирования и стандартную ошибку среднего для двух биологических повторностей там, где указано. Также рассчитывали среднюю эффективность редактирования для 48 спейсеров и распространяющуюся стандартную ошибку среднего там, где указано.

[1076] Результаты: На ФИГ. 39 представлена гистограмма, демонстрирующая среднюю эффективность редактирования относительно CasX 491 для избранных нуклеаз CasX по 48 разным целевым сайтам с PAM-последовательностью TTC в человеческих клетках. Планками погрешностей показана распространяемая стандартная ошибка среднего по двум экспериментам. Эти данные показывают, что как CasX 119, так и CasX 491 являются существенно более эффективными, чем CasX 2 дикого типа. Кроме того, CasX 515 значительно не отличается по эффективности редактирования от CasX 491. Неожиданно было обнаружено, что CasX 527 демонстрировал улучшенную эффективность по сравнению с 491 в присутствии PAM-последовательностей TTC ($p=0,0000635$ при анализе с помощью двухстороннего t-критерия Уэлча). Нуклеаза CasX 527 была сконструирована для демонстрации улучшенной эффективности редактирования в присутствии PAM-последовательностей ATC, CTC или GTC, потенциально посредством стабилизации структуры R-петли содержащего CasX рибонуклеопротеина (RNP), в отношении целевых сайтов из двухцепочечной ДНК с неканоническими PAM-последовательностями. CasX 527 содержит аминокислотную вставку аргинина в положении 26 CasX 491. Это положение физически приближено к взаимодействию петли распознавания PAM в CasX (положение аминокислоты 223) с нуклеотидами PAM-последовательности в нецелевой цепи (NTS) ДНК.

[1077] На ФИГ. 40 на схеме показаны положения петли распознавания PAM и мутация в CasX 527 (положение 26) на опубликованной структуре CryoEM гомологичного референсного белка CasX 1 (SEQ ID NO:1; идентификационный номер Protein Data Bank: 6NY2). Вероятно, дополнительные ионные взаимодействия между вставленным аргинином и NTS ДНК способствуют улучшенной стабильности в присутствии любой из PAM-последовательностей TTC, ATC, CTC или GTC, что приводит к улучшенной общей эффективности редактирования с участием этих спейсеров.

[1078] На ФИГ. 41 представлен скрипичный график сравнения эффективности редактирования CasX 2 и CasX 527 с избранными вариантами CasX 515 в присутствии 48 спейсеров для ассоциированных PAM-последовательностей TTC; при этом медианная эффективность редактирования представлена в виде горизонтального отрезка. Как описано выше, ранее было показано, что CasX 527 обладает эффективностью редактирования, которая равна или лучше, чем у CasX 491, и в данном случае наблюдали, что несколько новых вариантов CasX имели дополнительно улучшенную эффективность редактирования относительно CasX 527. Неожиданно было обнаружено, что частоты редактирования для CasX 583 были особенно однородными и высокими. Это может быть результатом

значительного улучшения стабильности структуры R-петли таким образом, что вариабельность показателей редактирования, обычно наблюдаемая для разных спейсеров, была преимущественно устранена. В поддержку этой гипотезы свидетельствует то, что CasX 583 отличается от CasX 515 наличием замены гидрофобного лейцина в положении 168 на положительно заряженный лизин в домене связывания нецелевой цепи (NTSB) в белке, что может облегчать образование дополнительных ионных связей целевой ДНК с NTS. Эта область не структурирована в структуре CryoEM, описанной выше, и домен NTSB был отмечен для ясности. В таблице 30 перечислены эффективности редактирования для 48 спейсеров для ассоциированных PAM-последовательностей TTC при нацеливании варианта белка CasX 527 или 583; можно видеть повышенную эффективность редактирования CasX 583 по большинству целевых сайтов в условиях анализа. Указаны название спейсера и ассоциированная PAM-последовательность.

[1079] На ФИГ. 42 представлена гистограмма, демонстрирующая среднюю эффективность редактирования относительно CasX 491 для избранных нуклеаз CasX по 48 разным целевым сайтам с PAM-последовательностью TTC в человеческих клетках. Планками погрешностей показана распространяемая стандартная ошибка среднего по двум экспериментам. Серой пунктирной линией показана активность редактирования для CasX 119. Эти данные указывают на то, что варианты CasX от 429 до 458 демонстрируют вариабельную эффективность редактирования относительно CasX 119, а в некоторых случаях демонстрируют улучшенную эффективность редактирования относительно CasX 119, который обеспечивает редактирование на уровне 70,8% относительно CasX 491. В частности, CasX 450 был существенно более эффективным, чем CasX 119, обеспечивая редактирование на уровне 95,9% относительно CasX 491, и он содержит четыре мутации по типу замены относительно последовательности CasX 119. Эти четыре мутации по типу замены следующие: D732N, E385P, Y857R, I658V. Важно отметить, что CasX 449 состоит из такой же последовательности, но у него отсутствует замена I658V, и он значительно менее эффективен, обеспечивая редактирование на уровне всего 58,1% относительно CasX 491. Это сравнение указывает на то, что эта мутация имеет критическое значение для повышения активности. Эти данные демонстрируют, что улучшенная активность редактирования достижима при внесении в CasX 119 комбинаций мутаций по типу замены. Примечательно, что такие повышения активности происходят отдельно от улучшений, наблюдаемых в результате обмена доменами между гомологичными последовательностями белка CasX. В частности, CasX 484 отличается от CasX 491 только тем, что домены NTSB и спиральный Ib в CasX 2 заменены на домены CasX 1, что сопровождается повышением активности с 62,0% до 100,0%. Эти данные показывают, что нуклеазная активность CasX 119 может быть усилена благодаря комбинациям отдельных мутаций по типу замены или обмену доменов с гомологичными белками CasX.

[1080] В условиях экспериментов был идентифицирован набор вариантов белка CasX 491 или 515, которые обеспечивают улучшенное расщепление двухцепочечной ДНК в человеческих клетках по целевым ДНК-последовательностям, ассоциированным с PAM-

последовательностью TTC, и приведены доказательства для конкретных положений или комбинаций положений для мутаций, которые могут применяться с целью дальнейшего конструирования вариантов молекул CasX с повышенной активностью в отношении интересующих целевых ДНК-последовательностей.

Таблица 30. Процент редактирования CasX 527 или CasX 583 в присутствии 48 спейсеров, ассоциированных с PAM-последовательностями TTC

Название спейсера	**Процент редактирования CasX 527**	**Процент редактирования CasX 583**
Rho_11.24	26,0	38,9
SOD1_8.89	25,9	38,9
SOD1_8.57	42,4	22,2
DMD_16.29	5,0	0,0
DMD_16.30	33,0	62,5
DMD_16.31	44,1	70,3
DMD_16.32	19,1	66,3
BCL11A_21.1	53,1	74,0
PCSK9_6.8	25,9	57,9
HTT_5.1	45,3	74,6
DMPK_20.7	19,9	56,3
B2M_7.9	53,9	72,7
B2M_7.37	32,5	59,9
CD70_26.5	37,7	66,0
CD70_26.3	27,2	78,0
FSHD_25.1	19,4	0,0
PMP22_18.22	31,9	66,4
PTBP1_28.1	17,4	0,0
Rho_11.2	32,4	59,5
VCV000162497_alt	22,6	27,4
VCV000037426_alt	34,7	34,7
VCV000052759_alt	43,8	84,8

VCV000031157_alt	27,5	0,0
VCV000217157_alt	17,7	21,4
VCV000048146_alt	36,8	53,3
VCV000039614_alt	42,9	73,1
VCV000217646_alt	39,7	81,8
VCV000408199_alt	28,5	64,2
VCV000431973_alt	40,6	62,2
VCV000004193_alt	18,7	15,8
VCV000003920_alt	31,1	68,1
VCV000224616_alt	20,6	58,9
VCV000004787_alt	37,9	61,7
VCV000013115_alt	0,0	100,0
VCV000195355_alt	45,3	83,3
VCV000190123_alt	49,1	78,1
VCV000048535_alt	32,9	61,2
VCV000468648_alt	38,4	77,9
VCV000000908_alt	30,7	64,3
VCV000066016_alt	0,0	0,0
VCV000184702_alt	41,8	8,1
VCV000003825_alt	30,2	55,7
VCV000014243_alt	36,2	77,9
VCV000046102_alt	47,3	74,6
VCV000050962_alt	24,9	67,4
VCV000011908_alt	30,0	64,8
VCV000252575_alt	17,9	38,3
SOD1_G93A_alt	37,5	80,2

Пример 20. Анализ PASS для идентификации вариантов белка CasX с повышенной специфичностью относительно CasX 491

[1081] Цель эксперимента заключалась в том, чтобы идентифицировать варианты CasX с улучшенной специфичностью к расщеплению целевых сайтов относительно

нецелевых сайтов в человеческих клетках. Для обеспечения этого клеточную линию HEK293 PASS_V1.01 обрабатывали белком CasX 2 дикого типа или сконструированными вариантами белка CasX 119 или 491 (SEQ ID NO: 336) или другим вариантом белка CasX вместе с одномолекулярной гидовой РНК (огРНК; SEQ ID NO: 2238) и проводили секвенирование нового поколения (NGS) для расчета процента редактирования при использовании различных спейсеров и ассоциированных целевых сайтов.

[1082] Материалы и методы. Для анализа вариантов клонированного белка применяли мультиплексный объединенный подход с помощью системы PASS, как описано в примере 19.

[1083] Для оценки активности редактирования, а также специфичности нуклеазы CasX к целевым сайтам в человеческих клетках проводили количественную оценку двух наборов целевых сайтов. Сначала была проведена количественная оценка 48 целевых сайтов, ассоциированных с PAM-последовательностью TTC, и для каждого из этих спейсеров рассчитывали среднюю эффективность редактирования и стандартную ошибку среднего для двух биологических повторностей. Также рассчитывали среднюю эффективность редактирования для 48 спейсеров и распространяющуюся стандартную ошибку среднего. Во-вторых, была проведена количественная оценка 28 наборов из двух пар спейсеров, ассоциированных с PAM-последовательностью TTC, и целевых сайтов. Каждый набор пар спейсер - целевой сайт состоял из фиксированной спейсерной последовательности и двух разных целевых сайтов. Два целевых сайта различались одним несовпадающим нуклеотидом в одном из двадцати положений целевого сайта. Один целевой сайт (целевая пара спейсер - целевой сайт) был полностью комплементарен спейсерной последовательности, а другой (нецелевая пара спейсер - целевой сайт) содержал неспаренную пару оснований между РНК спейсера и целевой цепью ДНК. Для каждого из этих спейсеров рассчитывали среднюю эффективность редактирования и стандартную ошибку среднего для двух биологических повторностей. Рассчитывали отношение эффективности нецелевого и целевого редактирования сайтов, а также распространяющуюся стандартную ошибку среднего для каждого из этих 28 наборов целевых сайтов. Этот показатель определяется как соотношение специфичности. Наконец, рассчитывали среднее соотношение специфичности для 28 наборов целевых сайтов, а также распространяющуюся стандартную ошибку среднего.

[1084] Результаты: На ФИГ. 43 представлена гистограмма, демонстрирующая среднюю эффективность редактирования относительно CasX 491, а также среднее соотношение специфичности для избранных нуклеаз CasX. Планками погрешностей показана распространяемая стандартная ошибка среднего по двум экспериментам. Эти данные показывают, что как CasX 119, так и CasX 491 являются существенно более эффективными, чем CasX 2 дикого типа, и что CasX 119 сохраняет специфичность при внесении однонуклеотидных неспаренных участков в спейсер (низкое соотношение специфичности) по сравнению с CasX 2 (среднее соотношение специфичности 0,171 для CasX 2 по сравнению с 0,182 для CasX 119). Напротив, CasX 491 потерял значительную

долю специфичности при внесении однонуклеотидных неспаренных участков в этих условиях (среднее соотношение специфичности 0,446). Дополнительная мутация в CasX 491 (вставка пролина в положении 793) с созданием CasX 515 несколько улучшает специфичность молекулы, хотя этот результат и не является статистически значимым. Более того, вставка аргинина в положении 26 CasX 491 с созданием CasX 527 значимо и существенно ухудшает специфичность молекулы (среднее соотношение специфичности 0,839). Однако важно отметить, что дополнительные одиночные мутации в CasX 515 дополнительно улучшают специфичность молекулы. Все из CasX 535, 537, 542, 543 и 544 значительно улучшают специфичность. CasX 544 улучшает специфичность в наибольшей степени, до среднего соотношения специфичности 0,183, что значительно не отличается от показателя CasX 2 дикого типа. Одновременно CasX 544 сохраняет среднюю эффективность редактирования 97,7% по сравнению с таковой у CasX 491, и эти показатели значительно не различаются. Эти данные демонстрируют, что одиночные мутации, внесенные в CasX 491 или 515, могут изменять либо активность, либо специфичность, либо обе эти характеристики нуклеазы при редактировании целевых ДНК-последовательностей в человеческих клетках. На ФИГ. 68 представлена схема, иллюстрирующая качественную связь между протестированными комбинациями мутаций и их воздействием как на активность, так и на специфичность. В таблице 31 представлены количественные данные по средней процентной доле редактирования для CasX 491 в отношении 28 наборов целевых и нецелевых пар спейсер - целевой сайт, иллюстрирующие среднее соотношение эффективности целевого и нецелевого редактирования. В таблице 32 представлены количественные данные по средней активности редактирования для избранных вариантов CasX относительно CasX 2 или варианта CasX 491, а также по среднему соотношению специфичности для каждого варианта в виде абсолютного значения от нуля до единицы. Во всех случаях сконструированные варианты показали улучшенную эффективность редактирования по сравнению с референсным CasX 2 (SEQ ID NO: 2) с кратностью повышения редактирования от 4 раз до почти 7 раз в условиях анализа, в то время как CasX 527 (SEQ ID NO: 427) и 532 (SEQ ID NO: 432) показали улучшения относительно CasX 491 (SEQ ID NO: 336). CasX 542 (SEQ ID NO: 442), 543 (SEQ ID NO: 443) и 544 (SEQ ID NO: 444) по существу имели такую же активность, как и CasX 491, но сохраняли специфичность, сравнимую с таковой у референсного CasX 2.

[1085] В условиях экспериментов был идентифицирован набор вариантов белка CasX 491 или 515, которые обеспечивают улучшенное расщепление двухцепочечной ДНК в человеческих клетках по целевым ДНК-последовательностям, ассоциированным с PAM-последовательностью TTC. Кроме того, посредством измерения целевого и нецелевого редактирования сайтов количественно определили специфичность этих вариантов белка. Эти данные можно применять для конструирования максимально активных и максимально специфичных терапевтических вариантов молекул CasX для интересующей целевой ДНК-последовательности.

Таблица 31. Средняя процентная доля редактирования для CasX 491 в отношении 28

наборов целевых и нецелевых пар спейсер - целевой сайт

Название спейсера	**Средний процент целевого редактирования для CasX 491**	**Средний процент нецелевого редактирования для CasX 491**
VCV000162497_alt	15,4	2,0
VCV000037426_alt	21,9	0,1
VCV000052759_alt	35,4	25,8
VCV000031157_alt	12,8	5,4
VCV000217157_alt	11,7	6,9
VCV000048146_alt	21,8	0,3
VCV000039614_alt	38,0	13,3
VCV000217646_alt	38,2	9,2
VCV000408199_alt	31,2	47,5
VCV000431973_alt	35,7	0,0
VCV000004193_alt	7,4	0,2
VCV000003920_alt	32,0	24,2
VCV000224616_alt	33,1	30,7
VCV000004787_alt	41,8	32,7
VCV000013115_alt	16,2	43,7
VCV000195355_alt	42,6	10,6
VCV000190123_alt	39,1	6,0
VCV000048535_alt	31,9	6,6
VCV000468648_alt	35,7	35,1
VCV000000908_alt	45,2	10,6
VCV000066016_alt	40,6	1,8
VCV000184702_alt	16,8	0,0
VCV000003825_alt	29,9	27,1
VCV000014243_alt	37,0	5,9
VCV000046102_alt	38,1	17,0

VCV000050962_alt	31,6	8,5
VCV000011908_alt	43,5	7,3
VCV000252575_alt	21,3	44,1

Таблица 31. Активность и специфичность белков CasX

Белок CasX	**Средняя эффективность редактирования относительно CasX 2 (%)**	**Средняя эффективность редактирования относительно CasX 491 (%)**	**Среднее соотношение специфичности**
2	100	17	0,17
119	429	71	0,18
491	606	100	0,45
515	602	99	0,37
527	694	115	0,84
532	677	112	0,63
535	541	89	0,27
536	513	85	0,29
537	470	78	0,25
542	575	95	0,19
543	539	89	0,17
544	592	98	0,18

Пример 21. Варианты улучшенных гидовых РНК демонстрируют повышенную целевую активность в отношении мышиного и человеческого локуса экзона 1 гена RHO *in vitro*

[1086] Были проведены эксперименты для идентификации новых вариантов сконструированных гидовых РНК с повышенной активностью в отношении разных геномных мишеней, включая терапевтически значимый мышиный и человеческий экзон 1 гена Rho. В предыдущих анализах было идентифицировано множество разных «гипервариабельных» областей (например, стеблевая петля) в пределах последовательностей каркаса, которые обладают потенциалом в отношении значительного повышения эффективности редактирования, а также специфичности (последовательности в таблице 33). Кроме того, проводили скрининги для идентификации вариантов каркаса, которые увеличивали бы общую активность нашей системы CRISPR в AAV-векторе при применении множества разных комбинаций PAM-последовательностей и спейсеров без запуска нецелевого или неспецифического редактирования. Достижение повышенной

эффективности редактирования по сравнению с текущими эталонными векторами позволило бы снизить дозы вирусных векторов, применяемых в исследованиях *in vivo*, повысив безопасность AAV-опосредованных систем CasX - гид.

Материалы и методы

[1087] Новые последовательности вариантов CasX и вариантов каркаса гРНК были вставлены в трансгенный конструкт на основе AAV для валидации плазмиды и вирусного вектора. Авторы изобретения по существу разделили трансген AAV между ITR на разные части, которые состояли из нашего терапевтического карго (варианты CasX и гРНК+спейсер) и вспомогательных элементов (например, промоторов, NLS, поли(A)), имеющих отношение к экспрессии в клетках млекопитающих. Каждая часть в геноме AAV была отделена сайтами ферментов рестрикции для обеспечения модульного клонирования. Части были заказаны в виде фрагментов генов в компании Twist, их амплифицировали с помощью ПЦР и расщепляли соответствующими ферментами рестрикции, очищали, а затем лигировали в вектор, расщепленный теми же ферментами. После этого новыми конструктами AAV трансформировали химически компетентные клетки *E. coli* (Turbos или Stbl3s), которые после восстановления при 37°C в течение 1 часа высевали на чашки с агаром на среде LB, содержащие канамицин. Одиночные колонии собирали, проводили мини-выделение ДНК и секвенировали по методу Сэнгера. Затем конструкты с подтвержденной последовательностью клонировали с помощью метода сборки Golden Gate и фермента рестрикции BbsI со спейсером 12.7 (нацеливающий tdTomato: CTGCATTCTAGTTGTGGTTT, SEQ ID NO: 194). Спейсеры получали посредством отжига двух олигонуклеотидов и разбавления в воде. Затем повторяли процессы трансформации и мини-выделения ДНК и снова проверяли последовательности клонированных векторов со спейсерами. Проводили макси-выделение валидированных конструктов. Для оценки качества макси-выделения конструкты обрабатывали в двух отдельных реакциях расщепления: с XmaI (который вносит разрезы по нескольким сайтам в каждом из ITR) и XhoI, который вносит один разрез в геном AAV. Эти гидролизаты и неразрезанный конструкт разделяли в 1% агарозном геле и визуализировали на ChemiDoc. Если плазмида находилась на > 90% в суперспиральной форме, имела правильный размер, а ITR были интактными, далее конструкт испытывали с помощью нуклеофекции и впоследствии использовали для получения AAV-вектора.

Таблица 33. Гидовые последовательности, клонированные в плазмиды p59.491.U6.X.Y. (X - гид; Y - спейсер)

Конструкт гида	Спейсерная последова-тельность	Гидовые последовательности	Последовательность гид+спейсер

174.12.7	CTGCATTCTAGTTGTGGTTT (SEQ ID NO: 194)	ACTGGCGCTTTTATCTGATTACTTTGAGAGCCATCACCAGCGACTATGTCGTAGTGGGTAAAGCTCCTCTTCGGAGGGAGCATCAAAG (SEQ ID NO: 702)	ACTGGCGCTTTTATCTGATTACTTTGAGAGCCATCACCAGCGACTATGTCGTAGTGGGTAAAGCTCCCTCTTCGGAGGGAGCATCAAAGCTGCATTCTAGTTGTGGTTT (SEQ ID NO: 727)
229.12.7	CTGCATTCTAGTTGTGGTT (SEQ ID NO: 195)	ACTGGCACTTTTATCTGATTACTTTGAGAGCCATCACCAGCGACTATGTCGTATGGGTAAAGCGCTTACGGACTTCGGTCCGTAAGAAGCATCAAAG (SEQ ID NO: 703)	ACTGGCACTTTTATCTGATTACTTTGAGAGCCATCACCAGCGACTATGTCGTATGGGTAAAGCGCTTACGGACTTCGGTCCGTAAGAAGCATCAAAGCTGCATTCTAGTTGTGGTTT (SEQ ID NO: 728)
230.12.7	CTGCATTCTAGTTGTGGTT (SEQ ID NO: 195)	ACTGGCACTTCTATCTGATTACTCTGAGAGCCATCACCAGCGACTATGTCGTATGGGTAAAGCGCTTACGGACTTCGGTCCGTAAGAAGCATCAGA (SEQ ID NO: 704)	ACTGGCACTTCTATCTGATTACTCTGAGAGCCATCACCAGCGACTATGTCGTATGGGTAAAGCGCTTACGGACTTCGGTCCGTAAGAAGCATCAGAGCTGCATTCTAGTTGTGGTTT (SEQ ID NO: 729)
231.12.7	CTGCATTCTAGTTGTGGTT (SEQ ID NO: 195)	ACTGGCGCTTCTATCTGATTACTCTGAGAGCCATCACCAGCGACTATGTCGTATGGGTAAAGCCGCTTACGGACTTCGGTCCGTAAGAGGCATCAGAG (SEQ ID NO: 705)	ACTGGCGCTTCTATCTGATTACTCTGAGAGCCATCACCAGCGACTATGTCGTATGGGTAAAGCCGCTTACGGACTTCGGTCCGTAAGAGGCATCAGAGCTGCATTCTAGTTGTGGTTT (SEQ ID NO: 730)
232.12.7	CTGCATTCTAGTTGTGGTT (SEQ ID NO: 195)	ACTGGCACTTCTATCTGATTACTCTGAGCGCCATCACCAGCGACTATGTCGTATGGGTAAAGCCGCTTACGGACTTCGGTCCGTAAGAGGCATCAGAG (SEQ ID NO: 706)	ACTGGCACTTCTATCTGATTACTCTGAGCGCCATCACCAGCGACTATGTCGTATGGGTAAAGCCGCTTACGGACTTCGGTCCGTAAGAGGCATCAGAGCTGCATTCTAGTTGTGGTTT (SEQ ID NO: 731)

233.12.7	CTGCATTCTAGTTGTGGTT (SEQ ID NO: 195)	ACTGGCGCTTCTATCTGATTACTCTGAGCGCCATCACCAGCGACTATGTCGTATGGGTAAAGCCGCTTACGGACTTCGGTCCGTAAGAGGCATCAGAG (SEQ ID NO: 707)	ACTGGCGCTTCTATCTGATTACTCTGAGCGCCATCACCAGCGACTATGTCGTATGGGTAAAGCCGCTTACGGACTTCGGTCCGTAAGAGGCATCAGAGCTGCATTCTAGTTGTGGTTT (SEQ ID NO: 732)
234.12.7	CTGCATTCTAGTTGTGGTT (SEQ ID NO: 195)	ACTGGCGCTTCTATCTGATTACTCTGAGCGCCATCACCAGCGACTATGTCGTATGGGTAAAGCGCCTTACGGACTTCGGTCCGTAAGGAGCATCAGAG (SEQ ID NO: 708)	ACTGGCGCTTCTATCTGATTACTCTGAGCGCCATCACCAGCGACTATGTCGTATGGGTAAAGCGCCTTACGGACTTCGGTCCGTAAGGAGCATCAGAGCTGCATTCTAGTTGTGGTTT (SEQ ID NO: 733)
235.12.7	CTGCATTCTAGTTGTGGTT (SEQ ID NO: 195)	ACTGGCGCTTCTATCTGATTACTCTGAGCGCCATCACCAGCGACTATGTCGTAGTGGGTAAAGCCGCTTACGGACTTCGGTCCGTAAGAGGCATCAGAG (SEQ ID NO: 709)	ACTGGCGCTTCTATCTGATTACTCTGAGCGCCATCACCAGCGACTATGTCGTAGTGGGTAAAGCCGCTTACGGACTTCGGTCCGTAAGAGGCATCAGAGCTGCATTCTAGTTGTGGTT (SEQ ID NO: 734)
236.12.7	CTGCATTCTAGTTGTGGTT (SEQ ID NO: 195)	ACGGGACTTTCTATCTGATTACTCTGAAGTCCCTCACCAGCGACTATGTCGTATGGGTAAAGCCGCTTACGGACTTCGGTCCGTAAGAGGCATCAGAG (SEQ ID NO: 710)	ACGGGACTTTCTATCTGATTACTCTGAAGTCCCTCACCAGCGACTATGTCGTATGGGTAAAGCCGCTTACGGACTTCGGTCCGTAAGAGGCATCAGAGCTGCATTCTAGTTGTGGTT (SEQ ID NO: 735)
237.12.7	CTGCATTCTAGTTGTGGTT (SEQ ID NO: 195)	ACCTGTAGTTCTATCTGATTACTCTGACTACAGTCACCAGCGACTATGTCGTATGGGTAAAGCCGCTTACGGACTTCGGTCCGTAAGAGGCATCAGAG (SEQ ID NO: 711)	ACCTGTAGTTCTATCTGATTACTCTGACTACAGTCACCAGCGACTATGTCGTATGGGTAAAGCCGCTTACGGACTTCGGTCCGTAAGAGGCATCAGAGCTGCATTCTAGTTGTGGTT (SEQ ID NO: 736)

174.11.30	AAGGGGC TCCGCAC CACGCC (SEQ ID NO: 204)	ACTGGCGCTTTTATCTG ATTACTTTGAGAGCCAT CACCAGCGACTATGTC GTAGTGGGTAAAGCTC CCTCTTCGGAGGGAGC ATCAAAG (SEQ ID NO: 702)	ACTGGCGCTTTTATCTGATTAC TTTGAGAGCCATCACCAGCGA CTATGTCGTAGTGGGTAAAGC TCCCTCTTCGGAGGGAGCATC AAAGAAGGGGCTCCGCACCA CGCC (SEQ ID NO: 737)
229.11.30	AAGGGGC TCCGCAC CACGCC (SEQ ID NO: 204)	ACTGGCACTTTTATCTG ATTACTTTGAGAGCCAT CACCAGCGACTATGTC GTATGGGTAAAGCGCT TACGGACTTCGGTCCGT AAGAAGCATCAAAG (SEQ ID NO: 703)	ACTGGCACTTTTATCTGATTAC TTTGAGAGCCATCACCAGCGA CTATGTCGTATGGGTAAAGCG CTTACGGACTTCGGTCCGTAA GAAGCATCAAAGAAGGGGCT CCGCACCACGCC (SEQ ID NO: 738)
230.11.30	AAGGGGC TCCGCAC CACGCC (SEQ ID NO: 204)	ACTGGCACTTCTATCTG ATTACTCTGAGAGCCAT CACCAGCGACTATGTC GTATGGGTAAAGCGCT TACGGACTTCGGTCCGT AAGAAGCATCAGA (SEQ ID NO: 704)	ACTGGCACTTCTATCTGATTA CTCTGAGAGCCATCACCAGCG ACTATGTCGTATGGGTAAAGC GCTTACGGACTTCGGTCCGTA AGAAGCATCAGAAAGGGGCT CCGCACCACGCC (SEQ ID NO: 739)
231.11.30	AAGGGGC TCCGCAC CACGCC (SEQ ID NO: 204)	ACTGGCGCTTCTATCTG ATTACTCTGAGAGCCAT CACCAGCGACTATGTC GTATGGGTAAAGCCGC TTACGGACTTCGGTCCG TAAGAGGCATCAGAG (SEQ ID NO: 705)	ACTGGCGCTTCTATCTGATTA CTCTGAGAGCCATCACCAGCG ACTATGTCGTATGGGTAAAGC CGCTTACGGACTTCGGTCCGT AAGAGGCATCAGAGAAGGGG CTCCGCACCACGCC (SEQ ID NO: 740)
232.11.30	AAGGGGC TCCGCAC CACGCC (SEQ ID NO: 204)	ACTGGCACTTCTATCTG ATTACTCTGAGCGCCAT CACCAGCGACTATGTC GTATGGGTAAAGCCGC TTACGGACTTCGGTCCG TAAGAGGCATCAGAG (SEQ ID NO: 706)	ACTGGCACTTCTATCTGATTA CTCTGAGCGCCATCACCAGCG ACTATGTCGTATGGGTAAAGC CGCTTACGGACTTCGGTCCGT AAGAGGCATCAGAGAAGGGG CTCCGCACCACGCC (SEQ ID NO: 741)

233.11.30	AAGGGGC TCCGCAC CACGCC (SEQ ID NO: 204)	ACTGGCGCTTCTATCTG ATTACTCTGAGCGCCAT CACCAGCGACTATGTC GTATGGGTAAAGCCGC TTACGGACTTCGGTCCG TAAGAGGCATCAGAG (SEQ ID NO: 707)	ACTGGCGCTTCTATCTGATTA CTCTGAGCGCCATCACCAGCG ACTATGTCGTATGGGTAAAGC CGCTTACGGACTTCGGTCCGT AAGAGGCATCAGAGAAGGGG CTCCGCACCACGCC (SEQ ID NO: 742)
234.11.30	AAGGGGC TCCGCAC CACGCC (SEQ ID NO: 204)	ACTGGCGCTTCTATCTG ATTACTCTGAGCGCCAT CACCAGCGACTATGTC GTATGGGTAAAGCGCC TTACGGACTTCGGTCCG TAAGGAGCATCAGAG (SEQ ID NO: 708)	ACTGGCGCTTCTATCTGATTA CTCTGAGCGCCATCACCAGCG ACTATGTCGTATGGGTAAAGC GCCTTACGGACTTCGGTCCGT AAGGAGCATCAGAGAAGGGG CTCCGCACCACGCC (SEQ ID NO: 743)
235.11.30	AAGGGGC TCCGCAC CACGCC (SEQ ID NO: 204)	ACTGGCGCTTCTATCTG ATTACTCTGAGCGCCAT CACCAGCGACTATGTC GTAGTGGGTAAAGCCG CTTACGGACTTCGGTCC GTAAGAGGCATCAGAG (SEQ ID NO: 709)	ACTGGCGCTTCTATCTGATTA CTCTGAGCGCCATCACCAGCG ACTATGTCGTAGTGGGTAAAG CCGCTTACGGACTTCGGTCCG TAAGAGGCATCAGAGAAGGG GCTCCGCACCACGCC (SEQ ID NO: 744)
236.11.30	AAGGGGC TCCGCAC CACGCC (SEQ ID NO: 204)	ACGGGACTTTCTATCTG ATTACTCTGAAGTCCCT CACCAGCGACTATGTC GTATGGGTAAAGCCGC TTACGGACTTCGGTCCG TAAGAGGCATCAGAG (SEQ ID NO: 710)	ACGGGACTTTCTATCTGATTA CTCTGAAGTCCCTCACCAGCG ACTATGTCGTATGGGTAAAGC CGCTTACGGACTTCGGTCCGT AAGAGGCATCAGAGAAGGGG CTCCGCACCACGCC (SEQ ID NO: 745)
237.11.30	AAGGGGC TCCGCAC CACGCC (SEQ ID NO: 204)	ACCTGTAGTTCTATCTG ATTACTCTGACTACAGT CACCAGCGACTATGTC GTATGGGTAAAGCCGC TTACGGACTTCGGTCCG TAAGAGGCATCAGAG (SEQ ID NO: 711)	ACCTGTAGTTCTATCTGATTAC TCTGACTACAGTCACCAGCGA CTATGTCGTATGGGTAAAGCC GCTTACGGACTTCGGTCCGTA AGAGGCATCAGAGAAGGGGC TCCGCACCACGCC (SEQ ID NO: 746)

174.11.31	AAGTGGC TCCGCAC CACGCC (SEQ ID NO: 215)	ACTGGCGCTTTTATCTG ATTACTTTGAGAGCCAT CACCAGCGACTATGTC GTAgTGGGTAAAGCTCC CTCTTCGGAGGGAGCA TCAAAG (SEQ ID NO: 702)	ACTGGCGCTTTTATCTGATTAC TTTGAGAGCCATCACCAGCGA CTATGTCGTAgTGGGTAAAGC TCCCTCTTCGGAGGGAGCATC AAAGAAGTGGCTCCGCACCAC GCC (SEQ ID NO: 747)
235.11.31	AAGTGGC TCCGCAC CACGCC (SEQ ID NO: 215)	ACTGGCGCTTCTATCTG ATTACTCTGAGCGCCAT CACCAGCGACTATGTC GTAGTGGGTAAAGCCG CTTACGGACTTCGGTCC GTAAGAGGCATCAGAG (SEQ ID NO: 709)	ACTGGCGCTTCTATCTGATTA CTCTGAGCGCCATCACCAGCG ACTATGTCGTAGTGGGTAAAG CCGCTTACGGACTTCGGTCCG TAAGAGGCATCAGAGAAGTG GCTCCGCACCACGCC (SEQ ID NO: 748)
174.11.1	AAGGGGC TGCGTAC CACACC (SEQ ID NO: 216)	ACTGGCGCTTTTATCTG ATTACTTTGAGAGCCAT CACCAGCGACTATGTC GTAGTGGGTAAAGCTC CCTCTTCGGAGGGAGC ATCAAAG (SEQ ID NO: 702)	ACTGGCGCTTTTATCTGATTAC TTTGAGAGCCATCACCAGCGA CTATGTCGTAGTGGGTAAAGC TCCCTCTTCGGAGGGAGCATC AAAGAAGGGGCTGCGTACCA CACC (SEQ ID NO: 749)
235.11.1	AAGGGGC TGCGTAC CACACC (SEQ ID NO: 216)	ACTGGCGCTTCTATCTG ATTACTCTGAGCGCCAT CACCAGCGACTATGTC GTAGTGGGTAAAGCCG CTTACGGACTTCGGTCC GTAAGAGGCATCAGAG (SEQ ID NO: 709)	ACTGGCGCTTCTATCTGATTA CTCTGAGCGCCATCACCAGCG ACTATGTCGTAGTGGGTAAAG CCGCTTACGGACTTCGGTCCG TAAGAGGCATCAGAGAAGGG GCTGCGTACCACACC (SEQ ID NO: 750)

235.NT	GGGTCTTCGAGAAGACCC (SEQ ID NO: 218)	ACTGGCGCTTCTATCTGATTACTCTGAGCGCCATCACCAGCGACTATGTCGTAGTGGGTAAAGCCGCTTACGGACTTCGGTCCGTAAGAGGCATCAGAG (SEQ ID NO: 709)	ACTGGCGCTTCTATCTGATTACTCTGAGCGCCATCACCAGCGACTATGTCGTAGTGGGTAAAGCCGCTTACGGACTTCGGTCCGTAAGAGGCATCAGAGGGGTCTTCGAGAAGACCC (SEQ ID NO: 751)

Репортерные клеточные линии

[1088] Иммортализованную линию нейрональных клеток-предшественников, выделенную из Ai9-tdTomato, культивировали в суспензии в предварительно уравновешенной среде для мышиных NPC (mNPC) (DMEM/F12 с GlutaMax, 10 мМ HEPES, 1× MEM с заменимыми аминокислотами, 1× пенициллин/стрептомицин, 1 : 1000 2-меркаптоэтанол, 1× добавка B-27, без витамина A, 1× N2 с добавлением факторов роста bFGF и EGF). Перед тестированием клетки снимали с помощью аккутазы при осторожном ресуспендировании, отслеживая полное разделение нейросфер. Затем реакцию в суспензии клеток гасили средой, осаждали клетки, центрифугировали и ресуспендировали в свежей среде. Клетки подсчитывали и непосредственно использовали для нуклеофекции или 10 000 клеток инкубировали в 96-луночном планшете, покрытом PLF (1× гидробромид поли-DL-орнитина, 10 мг/мл в стерильной деионизованной H_2O, 1× ламинин и 1× фибронектин), 2 дня перед трансдукцией AAV.

[1089] Двойную репортерную клеточную линию HEK293T получали посредством внесения в клетки HEK293T двух трансгенных кассет, которые конститутивно экспрессировали экзон 1 человеческого гена RHO, присоединенный к GFP, и экзон 1 человеческого гена P23H.RHO, присоединенный к mScarlet. Модифицированные клетки размножали последовательными пассажами каждые 3-5 дней и выдерживали в среде для фибробластов (FB), состоящей из среды Игла в модификации Дульбекко (DMEM; Corning Cellgro, № 10-013-CV) с добавлением 10% эмбриональной бычьей сыворотки (FBS; Seradigm, № 1500-500) и 100 ЕД/мл пенициллина и 100 мг/мл стрептомицина (100× Pen-Strep; GIBCO, № 15140-122), которая может дополнительно включать в себя пируват натрия (100×, Thermofisher, № 11360070), заменимые аминокислоты (100×, Thermofisher, № 11140050), буфер HEPES (100×, Thermofisher, № 15630080) и 2-меркаптоэтанол (1000×, Thermofisher, № 21985023). Клетки инкубировали при 37°C и 5% CO_2. Через 1-2 недели клетки GFP+/mScarlet+ были отсортированы и помещены в среду FB. Репортерные линии размножали последовательными пассажами каждые 3-5 дней и сохраняли в среде FB в инкубаторе при 37°C и 5% CO_2. Репортерные клоны получали методом ограничивающего разведения. Для характеризации клональных линий применяли проточную цитометрию, геномное секвенирование и функциональную модификацию локуса RHO с использованием ранее валидированной молекулы CasX, нацеливающей на RHO. Оптимальные репортерные линии были идентифицированы как линии, которые i) имели по одной правильно

интегрированной копии WTRHO.GFP и mutRHO.mScarlet на клетку, ii) сохраняли время удвоения, эквивалентное таковому у немодифицированных клеток, и iii) приводили к уменьшению флуоресценции GFP и mScarlet после разрушения гена RHO при анализе с применением способов, описанных ниже.

Нуклеофекция

[1090] Цис-плазмиды AAV, направляющие экспрессию системы CasX - каркас - гид, вводили нуклеофекцией в mNPC с помощью набора для нуклеофекции в 96-луночном планшете Lonza P3 Primary Cell 96-well Nucleofector Kit. Для линии ARPE-19 использовали раствор Lonza SF и добавку. Плазмиды разбавляли до концентрации 200 нг/мкл, 100 нг/мкл. К раствору P3 или SF, содержащему 200 000 клеток tdTomato mNPC или клеток ARPE-19 соответственно, добавляли по 5 мкл ДНК на конструкт. Объединенный раствор применяли для нуклеофекции с помощью системы Lonza 4D Nucleofectral System в соответствии с инструкциями производителя. После нуклеофекции процесс в растворе останавливали соответствующей культуральной средой. Затем раствор делили на аликвоты в трех повторностях (приблизительно 67 000 клеток на лунку) в 96-луночном планшете. Через 48 часов после трансфекции к обработанным клеткам добавляли свежую среду для mNPC, содержащую факторы роста. Через 5 дней после трансфекции клетки tdTomato mNPC снимали и оценивали активность с помощью FACS.

Получение AAV

[1091] Суспензионные клетки HEK293T адаптировали из исходных HEK293T и выращивали в среде FreeStyle 293. Для целей скрининга культуры в малом масштабе (20-30 мл, культивируемые в колбах Эрленмейера объемом 125 мл с перемешиванием при 110 об/мин) разбавляли до плотности 1,5e+6 клеток/мл в день трансфекции. Не содержащие эндотоксины плазмиды pAAV с трансгеном, фланкированным повторами ITR, применяли для трансфекции совместно с плазмидами, поставляющими аденовирусные вспомогательные гены для репликации и геном AAV rep/cap, с использованием PEIMax (Polysciences) в бессывороточной среде OPTIMEM. Через 3 часа после трансфекции в культуры добавляли 10% среду CDM4HEK293 (HyClone). Через три дня культуры центрифугировали при 1000 об/мин в течение 10 минут для отделения супернатанта от клеточного осадка. Супернатант смешивали с 40% PEG/2,5 M NaCl (конечная концентрация 8%) и инкубировали на льду в течение по меньшей мере 2 часов для осаждения вирусных частиц AAV. Клеточный осадок, содержащий большинство AAV-векторов, ресуспендировали в лизирующей среде (0,15 M NaCl, 50 мМ Tris-HCl, 0,05% Tween, pH 8,5), обрабатывали ультразвуком на льду (15 секунд, амплитуда 30%) и обрабатывали Benzonase (250 ЕД/мкл, Novagen) в течение 30 минут при 37 °C. Затем неочищенный лизат и обработанный PEG супернатант центрифугировали при 4000 об/мин в течение 20 минут при 4°C для ресуспендирования осажденных PEG AAV (осадка) в неочищенным лизате без клеточного дебриса (супернатанте). Дополнительно осветляли с помощью фильтра с диаметром пор 0,45 мкм.

[1092] Для определения титра вирусного генома 1 мкл вирусов в неочищенном

лизате подвергали расщеплению с помощью ДНКазы и протеазы K с последующей количественной ПЦР. Пять мкл вирусной суспензии после расщепления использовали в реакции количественной ПЦР объемом 25 мкл, состоящей из мастер-микса IDT PrimeTime и набора праймеров и зонда 6'-FAM/Zen/IBFQ (IDT), разработанных для амплификации области с промотором CMV (прямой 5'-CATCTACGTATTAGTCATCGCTATTACCA-3' (SEQ ID NO: 752); обратный 5'-GAAATCCCCGTGAGTCAAACC-3'(SEQ ID NO: 753), зонд 5'-TCAATGGGCGTGGATAG-3'(SEQ ID NO: 754)) или фрагмента размером 62 п. н., расположенного в AAV2-ITR (прямой 5'-GGAACCCCTAGTGATGGAGTT -3' (SEQ ID NO: 755); обратный 5'-CGGCCTCAGTGAGCGA-3' (SEQ ID NO: 756), зонд 5'-CACTCCCTCTCTGCGCGCTCG-3' (SEQ ID NO: 757)). Для расчета титра (вирусный геном (вг)/мл) вирусных проб в качестве референсных стандартов использовали десятикратные серийные разведения (по 5 мкл каждого: от 2e+9 до 2e+4 копий ДНК/мл). Программа для количественной ПЦР была установлена следующим образом: начальная стадия денатурации при 95°C в течение 5 минут с последующими 40 циклами денатурации при 95°C в течение 1 мин и отжигом/удлинением при 60°C в течение 1 мин.

Трансдукция AAV

[1093] Клетки mNPC по 10 000 клеток/лунка высевали в 96-луночные планшеты в лунки с покрытием PLF за 48 часов до трансдукции AAV. Все эксперименты в условиях вирусной инфекции проводили в трех повторностях с нормализованным количеством вг среди экспериментальных векторов, в сериях 3-кратных разведений при значениях множественности заражения (MOI) в диапазоне от примерно 1,0e+6 до 1,0e+4 вг/клетка. Расчеты были основаны на расчетном количестве 20 000 клеток на лунку во время трансфекции. В каждую лунку вносили конечный объем 50 мкл AAV-векторов, разведенных в предварительно уравновешенной среде для mNPC с добавлением факторов роста bFGF/EGF (конечная концентрация 20 нг/мл). Через 48 часов после трансфекции проводили полную замену среды на свежую среду с добавлением факторов роста. Активность редактирования (количественное определение tdt+ клеток) оценивали с помощью FACS через 5 дней после трансфекции.

Оценка активности редактирования с помощью FACS

[1094] Через 5 дней после трансфекции обработанные клетки tdTomato mNPC или клетки ARPE-19 в 96-луночных планшетах промывали dPBS и обрабатывали 50 мкл TrypLE и трипсина (0,25%) в течение 15 и 5 минут соответственно. После диссоциации клеток реакции в обработанных лунках останавливали средой, содержащей DMEM, 10% FBS и 1× пенициллин/стрептомицин. Ресуспендированные клетки переносили в 96-луночные круглодонные планшеты и центрифугировали в течение 5 мин на 1000g. Затем клеточные осадки ресуспендировали в dPBS, содержащем 1× DAPI, и планшеты загружали в автоматический проточный цитометр Attune NxT. Проточный цитометр Attune NxT запускали с использованием следующих параметров гейтирования: FSC-A × SSC-A для отбора клеток, FSC-H × FSC-A для отбора одиночных клеток, FSC-A × VL1-A для отбора DAPI-отрицательных живых клеток и FSC-A × YL1-A для отбора tdTomato-положительных

клеток.

NGS-анализ инделов в локусе экзона 1 гена mRHO

[1095] Через 5 дней после трансфекции обработанные клетки tdTomato mNPC в 96-луночных планшетах промывали dPBS и обрабатывали 50 мкл TrypLE и трипсина (0,25%) в течение 15 и 5 минут соответственно. После диссоциации клеток реакции в обработанных лунках останавливали средой, содержащей DMEM, 10% FBS и 1× пенициллин/стрептомицин. Затем клетки центрифугировали, и полученные клеточные осадки промывали PBS перед обработкой для экстракции гДНК с помощью набора Zymo для мини-выделения ДНК в соответствии с инструкциями производителя. Для оценки уровней редактирования, происходящего в локусе экзона 1 мышиного гена RHO, амплификацией получали ампликоны из 200 нг гДНК с набором праймеров (прямой 5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNGCAGCCTTGGTCTCTGT CTACG-3' (SEQ ID NO: 758); обратный 5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCCCCAGTCTCTCTGCTCATACC-3' (SEQ ID NO: 759)), очищали с помощью гранул (Beckman coulter, Agencourt Ampure XP) и затем повторно амплифицировали для встраивания последовательности адаптера Illumina и уникального молекулярного идентификатора (UMI) размером 16 нукл. Качественную и количественную оценку ампликона проводили с помощью набора Fragment Analyzer DNA analyzer kit (Agilent, дцДНК 35-1500 п. о.). Ампликоны секвенировали на Illumina Miseq в соответствии с инструкциями производителя. Необработанные файлы fastq, полученные в результате секвенирования, обрабатывали следующим образом: (1) последовательности обрезали для обеспечения качества и для адаптерных последовательностей с помощью программы cutadapt (вер. 2.1); (2) последовательности из прочтения 1 и прочтения 2 объединяли в одну последовательность вставки с помощью программы flash2 (вер. 2.2.00) и (3) консенсусные последовательности вставки проводили через программу CRISPResso2 (вер. 2.0.29) вместе с ожидаемой последовательностью ампликона и спейсерной последовательностью. В этой программе количественно определяется процентная доля прочтений, которые были модифицированы в окне вокруг 3'-конца спейсера (окно размером 30 п. о., центрированное по -3 п. о. от 3'-конца спейсера). Активность CasX количественно определяли как общую процентную долю прочтений, которые содержат вставки, замены и/или делеции в любом месте в пределах этого окна.

Результаты:

[1096] Были проведены разные эксперименты по редактированию для количественной оценки целевого расщепления, опосредованного CasX 491 в паре с новыми вариантами каркаса гРНК (гидами 174 и 229-237) с разными спейсерами, нацеливающими на множество интересующих геномных локусов. Конструкты клонировали в основную цепь AAV p59, фланкированную последовательностями ITR2, что позволяло осуществлять экспрессию белка Cas 491 под контролем промотора CMV, а также каркаса-спейсера под контролем человеческого промотора U6. Репортерные клеточные линии mNPC-tdT использовали для оценки эффективности внесения двойных разрезов, опосредованных

одним спейсером, в локусе tdTomato (спейсер 12.7, PAM-последовательность TTC), а также эффективности внесения одиночного разреза в локусе экзона 1 эндогенного мышиного гена RHO (спейсер 11.30, PAM-последовательность CTCN). Двойную репортерную систему, интегрированную в клеточную линию, полученную из ARPE-19, также применяли для оценки целевого редактирования экзогенно экспрессируемого локуса человеческого гена RHO дикого типа (спейсеры 11.1, PAM-последовательность CTC).

[1097] Варианты каркаса гРНК со спейсерами 12.7 и 11.30 тестировали посредством нуклеофекции в линии мышиных клеток NPC при двух разных дозах, указанных на ФИГ. 14 и ФИГ. 15 соответственно. Активность конструктов сравнивали с активностью имеющегося эталонного каркаса гРНК 174. В обоих целевых локусах конструкты вариантов гидовых каркасов 231, 233, 234 и 235 демонстрировали более высокие уровни активности, чем конструкт, содержащий каркас 174. Каркас 235 демонстрировал 2-кратное увеличение активности в локусе экзона 1 mRHO по сравнению с каркасом 174. Авторы настоящего изобретения дополнительно подтвердили, что каркас 235 неизменно улучшает активность без повышения нецелевого расщепления, проведя нуклеофекцию двойной репортерной клеточной линии ARPE-19 конструктами p59.491.174.11.1 и p59.491.235.11.1, а также контролем с нецелевым спейсером. Спейсер 11.1 нацеливал на экзогенно экспрессируемый ген mRHO-GFP. Каркас 235 демонстрировал 3-кратное повышение активности по сравнению с каркасом 174 (9% по сравнению с 3% Rho-GFP-клеток соответственно). Аллель-специфичность оценивали с учетом процента популяции клеток P23H-RHO-Scarlet, у которых последовательность отличается от дикого типа на 1 п. н.

[1098] Наконец, авторы изобретения стремились продемонстрировать, что эти варианты каркаса эффективно упакованы в AAV и сохраняли активность при доставке с помощью вирусных частиц. Клетки mNPC, трансдуцированные AAV-векторами, экспрессирующими варианты гидовых каркасов 174 и 235 со спейсером 11.30 (целевой, мышиный ген RHO дикого типа) и 11.31 (нецелевой, мышиный P23 RHO), продемонстрировали повышенную активность конструктов, содержащих вариант каркаса 235, по сравнению с каркасом 174 на целевом локусе (> 5-кратное увеличение, ФИГ. 45A и B) при MOI 3,0e+5, при этом никаких нецелевых инделов обнаружено не было.

[1099] Результаты подтверждают, что могут быть сконструированы варианты каркаса с новыми структурными мутациями, имеющие повышенную активность в двойных репортерных системах с терапевтически релевантными геномными мишенями, такими как локусы экзона 1 мышиного и человеческого RHO. Кроме того, в то время как новый охарактеризованный каркас показал в общем > 2-кратное увеличение активности, не было обнаружено нецелевого расщепления области, имеющей 1 неспаренную пару нуклеотидов со спейсером. Это важно для аллель-специфической терапевтической стратегии, например, в случае adRP P23H Rho, где мутированная аллель отличается от последовательности дикого типа 1 нуклеотидом, и на нее нацеливает спейсер 11.31. Это исследование дополнительно подтверждает возможность применения гидового каркаса 235 в AAV-векторах, разработанных для спасения гена P23H RHO и генотоксических исследований, а

также для других терапевтических целей.

Пример 22. Демонстрация того, что каталитически неактивный CasX не редактирует эндогенный локус B2M *in vitro*

[1100] Будут проведены эксперименты для демонстрации того, что каталитически неактивный CasX не способен редактировать эндогенный ген бета-2-микроглобулина (B2M) *in vitro*.

Материалы и методы

Создание каталитически неактивных конструктов CasX (dCasX) и их клонирование

[1101] В этих экспериментах будут использованы варианты CasX 491, 527, 668 и 676 с вариантом каркаса 174. Также будет включен положительный контроль с Cas9 и соответствующим гидом. Для получения каталитически неактивного CasX 491 (dCasX491; CAS096; SEQ ID NO: 1107) и каталитически неактивного CasX 527 (dCasX527; CAS142; SEQ ID NO: 1109) участвующие в катализе остатки D659, E756, D921 домена RuvC варианта CasX491 и участвующие в катализе остатки D660, E757 и D922 домена RuvC варианта CasX 527 будут подвержены мутации с заменой на аланин для устранения эндонуклеазной активности. Аналогичным образом, будут выполнены мутации участвующих в катализе остатков D660, E757, D923 с заменой на аланин в домене RuvC вариантов CasX 668 и 676 для получения каталитически неактивного CasX 668 (dCasX668; CAS401; SEQ ID NO: XX) и каталитически неактивного CasX 676 (dCasX676; CAS402; SEQ ID NO: XX). Полученные плазмиды (аминокислотные последовательности вариантов dCasX перечислены в таблице 7) будут содержать конструкты со следующей конфигурацией: Ef1α-SV40NLS-вариант dCasX-SV40NLS. Плазмиды также будут содержать последовательности, кодирующие вариант каркаса гРНК 174, имеющий либо спейсер (спейсер 7.37; GGCCGAGAUGUCUCGCUCCG, SEQ ID NO: 1105), нацеливающий на эндогенный локус B2M, либо нецелевой контроль (спейсер 0.0; CGAGACGUAAUUACGUCUCG; SEQ ID NO: 1106).

[1102] Конструкции, содержащие последовательности, кодирующие dCasX491, dCasX527, dCasX668 и dCasX676, будут заказаны в виде олигонуклеотидов и собраны методом ПЦР с перекрывающимися праймерами с последующей изотермической сборкой для получения плазмид, кодирующих каталитически неактивные варианты CasX. После изотермической сборки полученными плазмидами будут трансформированы химически компетентные клетки *E. coli*, которые после восстановления при 37°C в течение 1 часа будут высеяны на чашки с агаром на среде LB, содержащие канамицин. Отдельные колонии будут собраны для проведения ПЦР для отбора колоний и секвенирования по методу Сэнгера. Будет проведено миди-выделение конструктов с валидированными последовательностями для последующей трансфекции в клетки HEK293T.

Трансфекция плазмид в клетки HEK293T

[1103] Клетки HEK293T будут высеяны при плотности 30 000 клеток на каждую лунку 96-луночного планшета. На следующий день в каждой лунке будет проведена кратковременная трансфекция с использованием липофектамина со 100 нг плазмиды

каталитически неактивного варианта, содержащей конструкт CasX : гРНК, кодирующий вариант CasX 491, dCasX491, dCasX527, dCasX668 или dCasX676 (последовательности в таблице 7), причем гРНК будет нести либо нецелевой спейсер 0.0, либо нацеливающий спейсер 7.37 для локуса B2M. Каждый конструкт будет протестирован в трех повторностях. Через 24 часа после трансфекции клетки будут отобраны с помощью 2 мкг/мл пуромицина. Через шесть дней после трансфекции клетки будут собраны для анализа редактирования с помощью секвенирования нового поколения (NGS) и анализа экспрессии белка B2M методом иммуноокрашивания B2M с последующей проточной цитометрией. Экспрессию B2M будут определять с использованием антитела (BioLegend), которое позволит обнаружить B2M-зависимый белок HLA, экспрессируемый на клеточной поверхности. Количественную оценку HLA+ клеток будут проводить с помощью проточного цитометра Attune NxT.

Обработка и анализ NGS

[1104] Геномная ДНК (гДНК) из собранных клеток будет извлечена с помощью набора Zymo Quick-DNA Miniprep Plus в соответствии с инструкциями производителя. Целевые ампликоны будут образованы посредством амплификации интересующих областей из 200 нг извлеченной гДНК с набором праймеров, специфичных к человеческому локусу B2M. Эти ген-специфичные праймеры содержат дополнительную последовательность на 5'-конце для введения адаптера Illumina и уникального молекулярного идентификатора длиной 16 нуклеотидов. Амплифицированные продукты ДНК будут очищены с помощью набора для очистки ДНК Ampure XP DNA cleanup kit. Качественную и количественную оценку ампликона будут проводить с помощью набора для анализа ДНК DNA analyzer kit (Agilent, дцДНК 35-1500 п. н.). Ампликоны будут секвенированы на Illumina Miseq в соответствии с инструкциями производителя. Необработанные файлы sfq, полученные в результате секвенирования, будут подвергнуты контролю качества данных и обработаны с использованием программ cutadapt вер. 2.1, flash2 вер. 2.2.00 и CRISPResso2 вер. 2.0.29. Каждая последовательность будет количественно оценена на содержание вставки или делеции (индела) по сравнению с референсной последовательностью в окне вокруг 3'-конца спейсера (окно 30 п. н., центрированное относительно положения -3 п. н. от 3'-конца спейсера). Активность CasX будет количественно определена как общий процент прочтений, которые содержат вставки, замены и/или делеции в любом месте в пределах этого окна для каждой пробы.

Результаты:

[1105] Ожидается, что эксперименты продемонстрируют редактирование в локусе B2M, опосредованное каталитически активным CasX 491, которое приведет к сниженной экспрессии белка B2M. С другой стороны, как ожидается, ни один из dCasX491, dCasX527, dCasX668 и dCasX676 не будет демонстрировать редактирование в локусе B2M. Учитывая предполагаемый стерический эффект каталитически неактивной молекулы CasX на начальный сайт транскрипции локуса B2M, ожидается небольшое подавление экспрессии белка B2M в экспериментах со всеми ферментативно неактивными молекулами CasX.

Пример 23. Улучшение аффинности связывания шпильки MS2 усиливает активность редактирования XDP

[1106] Были проведены эксперименты для определения того, можно ли улучшить активность редактирования XDP с помощью стратегии рекрутирования, в соответствии с которой гРНК в RNP-комплексе CasX : гРНК содержит функционализированный удлиненный стебель РНК со шпилькой MS2, имеющей высокую аффинность к Gag-MS2 РНК-связывающему белку (RBP). Связывание шпильки РНК с MS2 RBP позволяет рекрутировать карго RNP CasX в частицу XDP. После доставки XDP в клетку-мишень для редактирования, как ожидается, комплекс шпилька РНК-MS2 RBP диссоциирует, позволяя CasX транслоцироваться в ядро. Таким образом, повышение стабильности комплекса белка MS2 - РНК способствует образованию XDP, что может быть достигнуто посредством изменения последовательностей MS2 РНК-связывающего белка или шпильки РНК для повышения аффинности связывания между этими компонентами.

[1107] Чтобы дополнительно исследовать этот принцип, гРНК, содержащие варианты шпильки РНК с различными аффинностями к MS2 RBP, исследовали с использованием высокоэффективного биохимического анализа *in vitro* для оценки равновесного связывания и кинетики диссоциации (Buenrostro et al., Quantitative analysis of RNA-protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. Nat Biotechnol. 32(6):562 (2014)). Варианты шпильки гРНК и соответствующие им значения K_D (константы диссоциации) приведены в таблице 34; последовательности гидовых плазмид, кодирующих разные варианты шпильки РНК MS2, представлены в таблице 35, а последовательности шпилек MS2 представлены в таблице 36. Были проведены эксперименты для исследования того, будут ли гРНК, содержащие варианты шпильки MS2 с улучшенной аффинностью связывания, усиливать образование XDP или активность редактирования. В частности, множество вариантов шпильки MS2 с различными равновесными аффинностями связывания оценивали на предмет их влияния на активность и титр XDP. В эти эксперименты также были включены несколько не связывающихся вариантов.

Таблица 34. Каркасы гРНК, содержащие варианты шпильки MS2 с различными аффинностями, и их значения константы диссоциации (K_D). Конкретные положения указанных нуклеотидных мутаций относятся к положениям базовой шпильки MS2 (каркас 188), представленной на ФИГ. 48

№ каркаса	**Положения указанных замен нуклеотидов в пределах шпильки MS2 (каркас 188)**	**K_D (нМ)**	**K_{OFF} (1/с)**	**$-\Delta\log(K_{OFF})/-\Delta\Delta G$**
Варианты шпильки MS2 с высокой аффинностью				
188		2,558	0,001	
251		2,558		
296	-13C, 1G	1,881	0,001	0,510
297	-15G, 3C	2,112	0,001	0,062

298	-15G, -13C, 1G, 3C			
299	-15G, -13C, -8C, -3G, 1G, 3C			
300	-8C, -3G	2,686	0,002	> 1
Варианты шпильки MS2 со средневысокой аффинностью				
304	-13U	9,346	0,002	0,286
307	-11U, -1A	9,226	0,003	0,579
313	-6C	9,274	0,002	0,400
Варианты шпильки MS2 со средней аффинностью				
301	-8U, -3A	34,084	0,007	0,654
303	-15C, -13U		0,002	0,089
305	-13C	17,634	0,002	0,163
306	-1U		0,015	0,854
310	-5U	36,912	0,016	0,910
314	-14G, 3G		0,002	0,120
Варианты шпильки MS2 со средненизкой аффинностью				
308	-11U, -1G	77,562	0,018	0,754
Варианты шпильки MS2 с низкой аффинностью				
309	-11U, -1U	453,563	*н/д*	*н/д*
311	-5A	415,477	*н/д*	*н/д*
302	-8A, -3G	1489,244	*н/д*	*н/д*
Варианты шпильки MS2 без аффинности				
312	-5G	12 506,440	*н/д*	*н/д*
315	-10G	18 018,92728	*н/д*	*н/д*

Таблица 35. Последовательности плазмид XDP

Идентификатор плазмиды	**Описание**	**ДНК-последовательность или SEQ ID NO**
pXDP161	Gag-(-1)-PR	1111
pXDP164	Gag-MS2	1112
pXDP353	Gag-MS2(N55K)	1113
pXDP344	CasX 491 NLS 240	1114
pSG50	Каркас 188	ACTGGCGCTTTTATCTGATTACTTTGAGAGCCATCACCAGCGACTATGTCGTAGTGGGTAAAGCTCACATGAGGATCACCCATGTGAGCATCAAAGCTGCATTCTAGTTGTGGTTT (SEQ ID NO: 1115)
pSG73	Каркас 251	ACTGGCGCTTTTATCTGATTACTTTGAGAGCCATCACCAGCGACTATGTCGTAGTGGGTAAAGCTGCACTATGGGCGCAGCTCATGAGGATCACCCATGAGCTGACGGTACAGGCCACATGAGGATCACCCATGTGGTATAGTGCAGCATCAAAGCTGCATTCTAGTTGTGGTTT (SEQ ID NO: 1116)
pSG170	Каркас 296	ACTGGCGCTTTTATCTGATTACTTTGAGAGCCATCACCAGCGACTATGTCGTAGTGGGTAAAGCTCACCTGAGGATCACCCAGGTGAGCATCAAAGCTGCATTCTAGTTGTGGTTTTTTTTT (SEQ ID NO: 1117)
pSG171	Каркас 297	ACTGGCGCTTTTATCTGATTACTTTGAGAGCCATCA

		CCAGCGACTATGTCGTAGTGGGTAAAGCTCGCATG AGGATCACCCATGCGAGCATCAAAGCTGCATTCTA GTTGTGGTTTTTTTTT (SEQ ID NO: 1118)
pSG172	Каркас 298	ACTGGCGCTTTTATCTGATTACTTTGAGAGCCATCA CCAGCGACTATGTCGTAGTGGGTAAAGCTCGCCTG AGGATCACCCAGGCGAGCATCAAAGCTGCATTCTA GTTGTGGTTTTTTTTT (SEQ ID NO: 1119)
pSG173	Каркас 299	ACTGGCGCTTTTATCTGATTACTTTGAGAGCCATCA CCAGCGACTATGTCGTAGTGGGTAAAGCTCGCCTG AGCATCAGCCAGGCGAGCATCAAAGCTGCATTCTA GTTGTGGTTTTTTTTT (SEQ ID NO: 1120)
pSG174	Каркас 300	ACTGGCGCTTTTATCTGATTACTTTGAGAGCCATCA CCAGCGACTATGTCGTAGTGGGTAAAGCTCACATG AGCATCAGCCATGTGAGCATCAAAGCTGCATTCTA GTTGTGGTTTTTTTTT (SEQ ID NO: 1121)
pSG175	Каркас 301	ACTGGCGCTTTTATCTGATTACTTTGAGAGCCATCA CCAGCGACTATGTCGTAGTGGGTAAAGCTCACATG AGTATCAACCATGTGAGCATCAAAGCTGCATTCTA GTTGTGGTTTTTTTTT (SEQ ID NO: 1122)
pSG176	Каркас 302	ACTGGCGCTTTTATCTGATTACTTTGAGAGCCATCA CCAGCGACTATGTCGTAGTGGGTAAAGCTCACATG AGAATCAGCCATGTGAGCATCAAAGCTGCATTCTA GTTGTGGTTTTTTTTT (SEQ ID NO: 1123)
pSG177	Каркас 303	ACTGGCGCTTTTATCTGATTACTTTGAGAGCCATCA CCAGCGACTATGTCGTAGTGGGTAAAGCTCCCTTGA GGATCACCCATGTGAGCATCAAAGCTGCATTCTAGT TGTGGTTTTTTTT (SEQ ID NO: 1124)
pSG178	Каркас 304	ACTGGCGCTTTTATCTGATTACTTTGAGAGCCATCA CCAGCGACTATGTCGTAGTGGGTAAAGCTCACTTG AGGATCACCCATGTGAGCATCAAAGCTGCATTCTA GTTGTGGTTTTTTTTT (SEQ ID NO: 1125)
pSG179	Каркас 305	ACTGGCGCTTTTATCTGATTACTTTGAGAGCCATCA CCAGCGACTATGTCGTAGTGGGTAAAGCTCACCTG AGGATCACCCATGTGAGCATCAAAGCTGCATTCTA GTTGTGGTTTTTTTTT (SEQ ID NO: 1126)
pSG180	Каркас 306	ACTGGCGCTTTTATCTGATTACTTTGAGAGCCATCA CCAGCGACTATGTCGTAGTGGGTAAAGCTCACATG AGGATCACCTATGTGAGCATCAAAGCTGCATTCTA GTTGTGGTTTTTTTTT (SEQ ID NO: 1127)
pSG181	Каркас 307	ACTGGCGCTTTTATCTGATTACTTTGAGAGCCATCA CCAGCGACTATGTCGTAGTGGGTAAAGCTCACATT AGGATCACCAATGTGAGCATCAAAGCTGCATTCTA GTTGTGGTTTTTTTTT (SEQ ID NO: 1128)
pSG182	Каркас 308	ACTGGCGCTTTTATCTGATTACTTTGAGAGCCATCA CCAGCGACTATGTCGTAGTGGGTAAAGCTCACATT AGGATCACCGATGTGAGCATCAAAGCTGCATTCTA GTTGTGGTTTTTTTTT (SEQ ID NO: 1129)
pSG183	Каркас 309	ACTGGCGCTTTTATCTGATTACTTTGAGAGCCATCA CCAGCGACTATGTCGTAGTGGGTAAAGCTCACATT AGGATCACCTATGTGAGCATCAAAGCTGCATTCTA GTTGTGGTTTTTTTTT (SEQ ID NO: 1130)
pSG184	Каркас 310	ACTGGCGCTTTTATCTGATTACTTTGAGAGCCATCA

		CCAGCGACTATGTCGTAGTGGGTAAAGCTCACATG AGGATTACCCATGTGAGCATCAAAGCTGCATTCTA GTTGTGGTTTTTTTTT (SEQ ID NO: 1131)
pSG185	Каркас 311	ACTGGCGCTTTTATCTGATTACTTTGAGAGCCATCA CCAGCGACTATGTCGTAGTGGGTAAAGCTCACATG AGGATAACCCATGTGAGCATCAAAGCTGCATTCTA GTTGTGGTTTTTTTTT (SEQ ID NO: 1132)
pSG186	Каркас 312	ACTGGCGCTTTTATCTGATTACTTTGAGAGCCATCA CCAGCGACTATGTCGTAGTGGGTAAAGCTCACATG AGGATGACCCATGTGAGCATCAAAGCTGCATTCTA GTTGTGGTTTTTTTTT (SEQ ID NO: 1133)
pSG187	Каркас 313	ACTGGCGCTTTTATCTGATTACTTTGAGAGCCATCA CCAGCGACTATGTCGTAGTGGGTAAAGCTCACATG AGGACCACCCATGTGAGCATCAAAGCTGCATTCTA GTTGTGGTTTTTTTTT (SEQ ID NO: 1134)
pSG188	Каркас 314	ACTGGCGCTTTTATCTGATTACTTTGAGAGCCATCA CCAGCGACTATGTCGTAGTGGGTAAAGCTCAGATG AGGATCACCCATGGGAGCATCAAAGCTGCATTCTA GTTGTGGTTTTTTTTT (SEQ ID NO: 1135)
pSG189	Каркас 315	ACTGGCGCTTTTATCTGATTACTTTGAGAGCCATCA CCAGCGACTATGTCGTAGTGGGTAAAGCTCACATG GGGATCACCCATGTGAGCATCAAAGCTGCATTCTA GTTGTGGTTTTTTTTT (SEQ ID NO: 1136)

Таблица 36. Варианты последовательностей шпильки MS2

№ каркаса	**Последовательности MS2**	**SEQ ID NO**
188	ACAUGAGGAUCACCCAUGU	1137
251		
296	ACCUGAGGAUCACCCAGGU	1138
297	GCAUGAGGAUCACCCAUGC	1139
298	GCCUGAGGAUCACCCAGGC	1140
299	GCCUGAGCAUCAGCCAGGC	1141
300	ACAUGAGCAUCAGCCAUGU	1142
304	ACUUGAGGAUCACCCAUGU	1143
307	ACAUUAGGAUCACCAAUGU	1144
313	ACAUGAGGACCACCCAUGU	1145

Материалы и методы

[1108] Все плазмиды, кодирующие белки CasX, относятся к варианту CasX 491. Все XDP были псевдотипированы с использованием 10% VSV-G (процентная доля плазмиды VSV-G относительно других структурных плазмид XDP). Структуры свернутой молекулы РНК получали с помощью веб-сервера RNAfold и программного обеспечения VARNA. Способы получения XDP описаны в настоящем документе, а также в WO2021113772A1, включенной в полном объеме путем ссылки.

Клонирование структурных плазмид

[1109] Вкратце, для получения структурных плазмид XDP удаляли последовательность Gag-pol из pXDP1 (UC Berkeley), и амплифицированные и очищенные фрагменты, кодирующие CasX 491, ВИЧ-1 или компоненты CP генома MS2, клонировали в основные цепи плазмиды с использованием набора для клонирования In-Fusion HD

Cloning Kit (Takara) в соответствии с протоколами производителя. Сборными продуктами трансформировали химически компетентные клетки Turbo Competent *E. coli*, которые после восстановления при 37 °C высевали на чашки с агаром на основе среды LB, содержащие ампициллин. Отдельные колонии собирали, проводили мини-выделение ДНК и секвенировали по методу Сэнгера, чтобы проверить правильность сборки. Последовательности плазмид перечислены в таблице 35.

Клонирование гидовых плазмид

[1110] Все гидовые плазмиды, содержащие варианты шпильки РНК MS2, включали нацеливающий на tdTomato спейсер 12,7 (CUGCAUUCUAGUUGUGGUUU; SEQ ID NO: 1146). Нацеливающий на tdTomato спейсер клонировали так, как описано ранее. Вкратце, спейсер получили посредством отжига двух олигонуклеотидов и клонировали методом сборки Golden Gate с соответствующими ферментами рестрикции в плазмиду pSG с альтернативным каркасом. Клонированные спейсеры подвергали трансформации, проводили мини-выделение ДНК и секвенировали по методу Сэнгера для проверки.

Клонирование плазмиды гликопротеина pGP2

[1111] Вкратце, последовательности, кодирующие гликопротеин VSV-G и промотор CMV, и основную цепь, полученную из плазмиды с геном устойчивости к канамицину, амплифицировали и очищали так, как описано ранее. Эти конструкты клонировали в основные цепи плазмиды с помощью набора In-Fusion® HD Cloning Kit (Takara) в соответствии с протоколами производителя. Сборными продуктами трансформировали химически компетентные клетки Turbo Competent *E. coli*, которые высевали на чашки с агаром на основе среды LB, содержащие канамицин, и инкубировали при 37 °C. Отдельные колонии собирали, проводили мини-выделение ДНК и секвенировали по методу Сэнгера, чтобы проверить правильность сборки.

Получение XDP

[1112] Вкратце, клетки HEK293T Lenti-X высевали в чашки диаметром 15 см в количестве 20×10^6 клеток на чашку за 24 часа до трансфекции и выращивали до достижения 70-90% конфлюэнтности. На следующий день клетки Lenti-X трансфицировали следующими плазмидами с использованием PEI Max (Polypus): структурные плазмиды XDP (также кодирующие варианты CasX), варианты гидовых плазмид и pGP2 для псевдотипирования XDP. Через 24 часа после трансфекции среду заменяли на Opti-MEM (Thermo Fisher). Среду, содержащую XDP, собирали через 72 часа после трансфекции и фильтровали через фильтр из ПЭС 0,45 мкм. Супернатант концентрировали и очищали посредством центрифугирования. XDP ресуспендировали в 500 мкл среды DMEM/F12 с добавлением GlutaMAX, HEPES, заменимых аминокислот (NEAA), пенициллина/стрептомицина, 2-меркаптоэтанола, B-27 без витамина A и N2.

Трансдукция XDP нейрональных клеток-предшественников (NPC), экспрессирующих белок tdTomato

[1113] Клетки tdTomato NPC выращивали в среде DMEM/F12 с добавлением GlutaMAX, HEPES, NEAA, пенициллина/стрептомицина, 2-меркаптоэтанола, B-27 без

витамина A и N2. Клетки собирали с использованием реактива для диссоциации клеток с аккутазой StemPro Accutase и высевали на 96-луночные планшеты с PLF-покрытием. Через 48 часов клетки трансдуцировали XDP, содержащими спейсер для нацеливания на tdTomato: сначала использовали профильтрованный супернатант с ресуспендированным вирусом и далее 5 полулогарифмических разведений. Затем клетки центрифугировали в течение 15 минут при 1000g. Трансдуцированные NPC выращивали в течение 96 часов перед анализом с помощью проточной цитометрии флуоресценции tdTomato в качестве маркера редактирования в локусе tdTomato, при этом EC_{50} определяли как количество частиц XDP, необходимых для достижения редактирования в 50% клеток, по данным проточной цитометрии. Анализ проводили 2-3 раза для каждой пробы с аналогичными результатами.

Результаты:

[1114] XDP, состоящие из Gag-MS2, Gag-pro, CasX, вариантов каркаса гРНК и VSV-G, получали либо с исходной MS2 (MS2 дикого типа), либо с вариантом MS2 с высокой аффинностью (MS2 353). Затем у полученных XDP оценивали их эффективность редактирования в локусе tdTomato в клетках NPC. На ФИГ. 49 представлена процентная доля редактирования в локусе tdTomato, измеренная по флуоресценции белка tdTomato с использованием проточной цитометрии, когда для трансдукции NPC использовали 0,007 мкл сконцентрированных препаратов XDP. В дополнение к базовым контрольным каркасам гРНК 188 и 251 высокоаффинные варианты каркасов 296 и 298 продемонстрировали повышенную активность в присутствии как MS2 дикого типа, так и MS2 353, при этом значения K_D находились в диапазоне от 1,8 до 2,1 нМ. Кроме того, варианты каркасов 303, 304, 305, 307, 310 и 313 со средней аффинностью, со значениями K_D в диапазоне от 9,2 до 36,9 нМ продемонстрировали многообещающие значения эффективности редактирования. На ФИГ. 50 представлены результаты определения EC_{50} для разных каркасов гРНК, включающих конфигурации MS2 дикого типа и MS 353. Варианты каркасов 296, 297 и 305 демонстрировали немного более высокую активность по сравнению с каркасом 188, - преимущество, которое было более явным при конфигурации MS2 353. На ФИГ. 51 представлена отчетливая корреляция между аффинностью (K_D) шпильки MS2 в гРНК и получаемой в результате активностью XDP (EC_{50}), со значением R^2=0,81 ($p < 0,001$). XDP, содержащие MS2, имеющую аффинность < 35 нМ, обеспечивали эффективное рекрутирование и упаковку CasX RNP в XDP. Однако в условиях этих экспериментов не наблюдалось корреляции между аффинностью (K_D) шпильки MS2 в гРНК и получаемым в результате титром XDP (ФИГ. 52).

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Каркас гидовой РНК (гРНК), содержащий последовательность, имеющую идентичность последовательности по меньшей мере 60%, по меньшей мере 70%, по меньшей мере 80%, по меньшей мере 90%, по меньшей мере 95%, по меньшей мере 98%, по меньшей мере 99% или 100% с любой из последовательностей, выбранных из группы, состоящей из SEQ ID NO: 2292, 2291, 2307, 2281-2290, 2293-2306, 2308-2332 и 23530-2398.

2. Каркас гРНК по п. 1, содержащий последовательность, выбранную из группы, состоящей из SEQ ID NO: 2292, 2291, 2307, 2281-2290, 2293-2306, 2308-2332 и 23530-2398.

3. Каркас гРНК по п. 1, содержащий последовательность, имеющую одну или более модификаций относительно SEQ ID NO: 2238, причем одна или более модификаций приводят к получению улучшенной характеристики.

4. Каркас гРНК по п. 3, в котором одна или более модификаций включают в себя одну или более нуклеотидных замен, вставок и/или делеций, представленных в таблице 19.

5. Каркас гРНК по п. 3 или п. 4, причем улучшенная характеристика представляет собой одно или более функциональных свойств, выбранных из группы, состоящей из повышенной активности редактирования, повышенной стабильности стебля псевдоузла, повышенной стабильности триплексной области, повышенной стабильности каркасного стебля, стабильности удлиненного стебля, сниженного числа нецелевых интермедиатов свертывания и повышенной аффинности связывания с белком CRISPR класса 2, типа V, необязательно при анализе in vitro.

6. Каркас гРНК по любому из пп. 3-5, демонстрирующий улучшенную оценку обогащения (log_2), которая на по меньшей мере примерно 2,0, по меньшей мере примерно 2,5, по меньшей мере примерно 3 или по меньшей мере примерно 3,5 выше, чем оценка каркаса гРНК с SEQ ID NO: 2238 в анализе in vitro.

7. Каркас гРНК по п. 1, содержащий последовательность, имеющую одну или более модификаций относительно SEQ ID NO: 2239, причем одна или более модификаций приводят к получению улучшенной характеристики.

8. Каркас гРНК по п. 7, в котором одна или более модификаций содержат одну или более нуклеотидных замен, вставок и/или делеций, представленных в таблице 20.

9. Каркас гРНК по п. 7 или п. 8, причем улучшенная характеристика представляет собой одно или более функциональных свойств, выбранных из группы, состоящей из повышенной активности редактирования, повышенной стабильности стебля псевдоузла, повышенной стабильности триплексной области, повышенной стабильности каркасного стебля, стабильности удлиненного стебля, сниженного числа нецелевых интермедиатов свертывания и повышенной аффинности связывания с белком CRISPR класса 2, типа V, необязательно при анализе in vitro.

10. Каркас гРНК по любому из пп. 7-9, демонстрирующий улучшенную оценку обогащения (log_2), которая на по меньшей мере примерно 1,2, по меньшей мере примерно 1,5, по меньшей мере примерно 2,0, по меньшей мере примерно 2,5, по меньшей мере примерно 3 или по меньшей мере примерно 3,5 выше, чем оценка каркаса гРНК с SEQ ID

NO: 2239 в анализе in vitro.

11. Каркас гРНК по п. 1, содержащий одну или более модификаций в положениях относительно последовательности SEQ ID NO: 2239, выбранных из группы, состоящей из C9, U11, C17, U24, A29, U54, G64, A88 и A95.

12. Каркас гРНК по п. 11, содержащий одну или более модификаций относительно последовательности SEQ ID NO: 2239, выбранных из группы, состоящей из C9U, U11C, C17G, U24C, A29C, вставки G в положении 54, вставки C в положении 64, A88G и A95G.

13. Каркас гРНК по п. 12, содержащий модификации относительно последовательности SEQ ID NO: 2239, состоящие из C9U, U11C, C17G, U24C, A29C, вставки G в положении 54, вставки C в положении 64, A88G и A95G.

14. Каркас гРНК по любому из пп. 7-13, причем улучшенная характеристика выбрана из группы, состоящей из стабильности стебля псевдоузла, стабильности триплексной области, стабильности пузырьков каркаса, стабильности удлиненного стебля и аффинности связывания с белком CRISPR класса 2, типа V.

15. Каркас гРНК по п. 14, в котором вставка C в положении 64 и замена A88G относительно последовательности SEQ ID NO: 2239 устраняют асимметричный элемент утолщения удлиненного стебля, повышая стабильность удлиненного стебля каркаса гРНК.

16. Каркас гРНК по п. 14, в котором замены U11C, U24C и A95G повышают стабильность триплексной области каркаса гРНК.

17. Каркас гРНК по п. 14, в котором замена A29C повышает стабильность стебля псевдоузла.

18. Каркас гРНК по п. 1 или п. 2, содержащий одну или более гетерологичных РНК-последовательностей в удлиненном стебле.

19. Каркас гРНК по п. 18, в котором гетерологичная РНК выбрана из группы, состоящей из шпильки MS2, шпильки Qβ, шпильки II U1, шпильки Uvsx и стебля с петлей PP7 или вариантов их последовательностей.

20. Каркас гРНК по п. 18 или п. 19, в котором гетерологичная последовательность РНК повышает стабильность гРНК.

21. Каркас гРНК по п. 18 или п. 19, в котором гетерологичная РНК способна связывать белок, РНК, ДНК или малую молекулу.

22. Каркас гРНК по любому из пп. 18-21, содержащий Rev-отвечающий элемент (RRE) или его часть.

23. Каркас гРНК по п. 22, в котором RRE или его часть выбраны из группы, состоящей из стебля IIB RRE, имеющего последовательность UGGGCGCAGCGUCAAUGACGCUGACGGUACA (SEQ ID NO: 1280), стебля II-V из RRE, имеющего последовательность CAGGAAGCACUAUGGGCGCAGCGUCAAUGACGCUGACGGUACAGGCCAGACAAUU AUUGUCUGGUAUAGUGCAGCAGCAGAACAAUUUGCUGAGGGCUAUUGAGGCGCA ACAGCAUCUGUUGCAACUCACAGUCUGGGGCAUCAAGCAGCUCCAGGCAAGAAUC CUG (SEQ ID NO: 1282), стебля II из RRE, имеющего последовательность

GCACUAUGGGCGCAGCGUCAAUGACGCUGACGGUACAGGCCAGACAAUUAUUGUC
UGGUAUAGUGC (SEQ ID NO: 1281), Rev-связывающего элемента (RBE) стебля IIB, имеющего последовательность GCUGACGGUACAGGC (SEQ ID NO: 1284), и полноразмерного RRE, имеющего последовательность
AGGAGCUUUGUUCCUUGGGUUCUUGGGAGCAGCAGGAAGCACUAUGGGCGCAGC
GUCAAUGACGCUGACGGUACAGGCCAGACAAUUAUUGUCUGGUAUAGUGCAGCA
GCAGAACAAUUUGCUGAGGGCUAUUGAGGCGCAACAGCAUCUGUUGCAACUCAC
AGUCUGGGGCAUCAAGCAGCUCCAGGCAAGAAUCCUGGCUGUGGAAAGAUACCU
AAAGGAUCAACAGCUCCU (SEQ ID NO: 1283).

24. гРНК по любому из пп. 1-23, в которой каркас гРНК содержит один или более тиминов (T).

25. гРНК, содержащая каркас гРНК по любому из пп. 1-24 и нацеливающую последовательность на 3'-конце каркаса гРНК, которая комплементарна последовательности целевой нуклеиновой кислоты.

26. гРНК по п. 25, в которой нацеливающая последовательность имеет 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 или 30 нуклеотидов.

27. гРНК по п. 26, в которой нацеливающая последовательность имеет 18, 19 или 20 нуклеотидов.

28. гРНК по любому из пп. 25-27, способная образовывать рибонуклеопротеиновый (RNP) комплекс с белком CRISPR класса 2, типа V.

29. Сконструированный белок CRISPR класса 2, типа V, содержащий:

a. домен NTSB, содержащий последовательность QPASKKIDQNKLKPEMDEKGNLTTAGFACSQCGQPLFVYKLEQVSEKGKAYTNYFGRC
NVAEHEKLILLAQLKPEKDSDEAVTYSLGKFGQ (SEQ ID NO: 2335) или последовательность, имеющую идентичность с ней по меньшей мере 80%, по меньшей мере 90%, по меньшей мере 95%, по меньшей мере 96%, по меньшей мере 97%, по меньшей мере 98% или по меньшей мере 99%;

b. домен спиральный I-II, содержащий последовательность RALDFYSIHVTKESTHPVKPLAQIAGNRYASGPVGKALSDACMGTIASFLSKYQDIIIEHQ
KVVKGNQKRLESLRELAGKENLEYPSVTLPPQPHTKEGVDAYNEVIARVRMWVNLNL
WQKLKLSRDDAKPLLRLKGFPSF (SEQ ID NO: 2336) или последовательность, имеющую идентичность с ней по меньшей мере 80%, по меньшей мере 90%, по меньшей мере 95%, по меньшей мере 96%, по меньшей мере 97%, по меньшей мере 98% или по меньшей мере 99%;

c. домен спиральный II, содержащий последовательность PLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAGYKRQEALRPYLSSEEDR
KKGKKFARYQLGDLLLHLEKKHGEDWGKVYDEAWERIDKKVEGLSKHIKLEEERRSE
DAQSKAALTDWLRAKASFVIEGLKEADKDEFCRCELKLQKWYGDLRGKPFAIEAE
(SEQ ID NO: 2351) или последовательность, имеющую идентичность с ней по меньшей мере 80%, по меньшей мере 90%, по меньшей мере 95%, по меньшей мере 96%, по меньшей

мере 97%, по меньшей мере 98% или по меньшей мере 99%; и

d. домен RuvC-I, содержащий последовательность SSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPTHILRIGESYKEKQRTIQAK KEVEQRRAGGYSRKYASKAKNLADDMVRNTARDLLYYAVTQDAMLIFENLSRGFGRQ GKRTFMAERQYTRMEDWLTAKLAYEGLPSKTYLSKTLAQYTSKTC (SEQ ID NO: 2352) или последовательность, имеющую идентичность с ней по меньшей мере 80%, по меньшей мере 90%, по меньшей мере 95%, по меньшей мере 96%, по меньшей мере 97%, по меньшей мере 98% или по меньшей мере 99%..

30. Белок CRISPR класса 2, типа V по п. 29, содержащий домен OBD-I, который содержит последовательность QEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPENIPQ (SEQ ID NO: 2342) или последовательность, имеющую идентичность с ней по меньшей мере 80%, по меньшей мере 90%, по меньшей мере 95%, по меньшей мере 96%, по меньшей мере 97%, по меньшей мере 98% или по меньшей мере 99%.

31. Белок CRISPR класса 2, типа V по п. 29 или п. 30, содержащий домен OBD-II, который содержит последовательность

NSILDISGFSKQYNCAFIWQKDGVKKLNLYLIINYFKGGKLRFKKIKPEAFEANRF YTVINKKSGEIVPMEVNFNFDDPNLIILPLAFGKRQGREFIWNDLLSLETGSLKLANGRVI EKTLYNRRTRQDEPALFVALTFERREVLD (SEQ ID NO: 2347) или последовательность, имеющую идентичность с ней по меньшей мере 80%, по меньшей мере 90%, по меньшей мере 95%, по меньшей мере 96%, по меньшей мере 97%, по меньшей мере 98% или по меньшей мере 99%.

32. Белок CRISPR класса 2, типа V по любому из пп. 29-31, содержащий домен спиральный I-I, который содержит последовательность PISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVA (SEQ ID NO: 2343) или последовательность, имеющую идентичность с ней по меньшей мере 80%, по меньшей мере 90%, по меньшей мере 95%, по меньшей мере 96%, по меньшей мере 97%, по меньшей мере 98% или по меньшей мере 99%.

33. Белок CRISPR класса 2, типа V по любому из пп. 29-32, содержащий домен TSL, который содержит последовательность SNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQITYYNRYKRQNVVKDLSV ELDRLSEESVNNDISSWTKGRSGEALSLLKKRFSHRPVQEKFVCLNCGFETH (SEQ ID NO: 2349) или последовательность, имеющую идентичность с ней по меньшей мере 80%, по меньшей мере 90%, по меньшей мере 95%, по меньшей мере 96%, по меньшей мере 97%, по меньшей мере 98% или по меньшей мере 99%.

34. Белок CRISPR класса 2, типа V по любому из пп. 29-33, содержащий домен RuvC-II, который содержит последовательность

ADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDKRAFVETWQSFYRKKLKEV WKPAV (SEQ ID NO: 2350) или последовательность, имеющую идентичность с ней по меньшей мере 80%, по меньшей мере 90%, по меньшей мере 95%, по меньшей мере 96%,

по меньшей мере 97%, по меньшей мере 98% или по меньшей мере 99%.

35. Белок CRISPR класса 2, типа V по п. 34, содержащий последовательность SEQ ID NO: 416 или последовательность, которая имеет идентичность с ней по меньшей мере 80%, по меньшей мере 90%, по меньшей мере 95%, по меньшей мере 96%, по меньшей мере 97%, по меньшей мере 98% или по меньшей мере 99%.

36. Белок CRISPR класса 2, типа V по любому из пп. 29-35, содержащий по меньшей мере одну модификацию в одном или более доменах.

37. Белок CRISPR класса 2, типа V по п. 36, причем по меньшей мере одна модификация включает в себя:

a. по меньшей мере одну замену аминокислоты в домене;

b. по меньшей мере одну делецию аминокислоты в домене;

c. по меньшей мере одну вставку аминокислоты в домене; или

d. любую комбинацию пп. (a)-(c).

38. Белок CRISPR класса 2, типа V по п. 36 или п. 37, содержащий модификацию в одном или более положениях аминокислот в домене NTSB относительно SEQ ID NO: 2335, выбранных из группы, состоящей из P2, S4, Q9, E15, G20, G33, L41, Y51, F55, L68, A70, E75, K88 и G90.

39. Белок CRISPR класса 2, типа V по п. 38, в котором одна или более модификаций в одном или более положениях аминокислот в домене NTSB выбраны из группы, состоящей из вставки G в положении 2, вставки I в положении 4, вставки L в положении 4, Q9P, E15S, G20D, делеции S в положении 30, G33T, L41A, Y51T, F55V, L68D, L68E, L68K, A70Y, A70S, E75A, E75D, E75P, K88Q и G90Q относительно SEQ ID NO: 2335.

40. Белок CRISPR класса 2, типа V по любому из пп. 36-39, содержащий модификацию в одном или более положениях аминокислот в домене спиральный I-II относительно SEQ ID NO: 2336, выбранных из группы, состоящей из I24, A25, Y29 G32, G44, S48, S51, Q54, I56, V63, S73, L74, K97, V100, M112, L116, G137, F138 и S140.

41. Белок CRISPR класса 2, типа V по п. 40, в котором одна или более модификаций в одном или более положениях аминокислот в домене спиральный I-II выбраны из группы, состоящей из вставки T в положении 24, вставки C в положении 25, Y29F, G32Y, G32N, G32H, G32S, G32T, G32A, G32V, делеции G в положении 32, G32S, G32T, G44L, G44H, S48H, S48T, S51T, Q54H, I56T, V63T, S73H, L74Y, K97G, K97S, K97D, K97E, V100L, M112T, M112W, M112R, M112K, L116K, G137R, G137K, G137N, вставки Q в положении 138 и S140Q относительно SEQ ID NO: 2336.

42. Белок CRISPR класса 2, типа V по любому из пп. 36-41, содержащий модификацию в одном или более положениях аминокислот в домене спиральный II относительно SEQ ID NO: 2351, выбранных из группы, состоящей из L2, V3, E4, R5, Q6, A7, E9, V10, D11, W12, W13, D14, M15, V16, C17, N18, V19, K20, L22, I23, E25, K26, K31, Q35, L37, A38, K41,R 42, Q43, E44, L46, K57, Y65, G68, L70, L71, L72, E75, G79, D81, W82, K84, V85, Y86, D87, I93, K95, K96, E98, L100, K102, I104, K105, E109, R110, D114, K118, A120, L121, W124, L125, R126, A127, A129, I133, E134, G135, L136, E138, D140, K141, D142,

E143, F144, C145, C147, E148, L149, K150, L151, Q152, K153, L158, E166 и A167.

43. Белок CRISPR класса 2, типа V по п. 42, в котором одна или более модификаций в одном или более положениях аминокислот в домене спиральный II выбраны из группы, состоящей из вставки A в положении 2, вставки H в положении 2, делеции L в положении 2 и делеции V в положении 3, V3E, V3Q, V3F, делеции V в положении 3, вставки D в положении 3, V3P, E4P, делеции E в положении 4, E4D, E4L, E4R, R5N, Q6V, вставки Q в положении 6, вставки G в положении 7, вставки H в положении 9, вставки A в положении 9, VD10, вставки T1 в положении 0, делеции V в положении 10, вставки F в положении 10, вставки D в положении 11, делеции D в положении 11, D11S, делеции W в положении 12, W12T, W12H, вставки P в положении 12, вставки Q в положении 13, вставки G в положении 12, вставки R в положении 13, W13P, W13D, вставки D в положении 13, W13L, вставки P в положении 14, вставки D в положении 14, делеции D в положении 14 и делеции M в положении 15, делеции M в положении 15, вставки T в положении 16, вставки P в положении 17, N18I, V19N, V19H, K20D, L22D, I23S, E25C, E25P, вставки G в положении 25, K26T, K27E, K31L, K31Y, Q35D, Q35P, вставки S в положении 37, делеции L в положении 37 и делеции A в положении 38, K41L, вставки R в положении 42, делеции Q в положении 43 и делеции E в положении 44, L46N, K57Q, Y65T, G68M, L70V, L71C, L72D, L72N, L72W, L72Y, E75F, E75L, E75Y, G79P, вставки E в положении 79, вставки T в положении 81, вставки R в положении 81, вставки W в положении 81, вставки Y в положении 81, вставки W в положении 82, вставки Y в положении 82, W82G, W82R, K84D, K84H, K84P, K84T, V85L, V85A, вставки L в положении 85, Y86C, D87G, D87M, D87P, I93C, K95T, K96R, E98G, L100A, K102H, I104T, I104S, I104Q, K105D, вставки K в положении 109, E109L, R110D, делеции R в положении 110, D114E, вставки D в положении 114, K118P, A120R, L121T, W124L, L125C, R126D, A127E, A127L, A129T, A129K, I133E, вставки C в положении 133, вставки S в положении 134, вставки G в положении 134, вставки R в положении 135, G135P, L136K, L136D, L136S, L136H, делеции E в положении 138, D140R, вставки D в положении 140, вставки P в положении 141, вставки D в положении 142, делеции E в положении 143+делеции F в положении 144, вставки Q в положении 143, F144K, делеции F в положении 144, делеции F в положении 144 и делеции C в положении 145, C145R, вставки G в положении 145, C145K, C147D, вставки V в положении 148, E148D, вставки H в положении 149, L149R, K150R, L151H, Q152C, K153P, L158S, E166L и вставки F в положении 167 относительно SEQ ID NO: 2351.

44. Белок CRISPR класса 2, типа V по любому из пп. 36-43, содержащий модификацию в одном или более положениях аминокислот в домене RuvC-I относительно SEQ ID NO: 2352, выбранных из группы, состоящей из I4, K5, P6, M7, N8, L9, V12, G49, K63, K80, N83, R90, M125 и L146.

45. Белок CRISPR класса 2, типа V по п. 44, в котором одна или более модификаций в одном или более положениях аминокислот в домене RuvC-I выбраны из группы, состоящей из вставки I в положении 4, вставки S в положении 5, вставки T в положении 6, вставки N в положении 6, вставки R в положении 7, вставки K в положении 7, вставки H в

положении 8, вставки S в положении 8, V12L, G49W, G49R, S51R, S51K, K62S, K62T, K62E, V65A, K80E, N83G, R90H, R90G, M125S, M125A, L137Y, вставки P в положении 137, делеции L в положении 141, L141R, L141D, вставки Q в положении 142, вставки R в положении 143, вставки N в положении 143, E144N, вставки P в положении 146, L146F, P147A, K149Q, T150V, вставки R в положении 152, вставки H153, T155Q, вставки H в положении 155, вставки R в положении 155, вставки L в положении 156, делеции L в положении 156, вставки W в положении 156, вставки A в положении 157, вставки F в положении 157, A157S, Q158K, делеции Y в положении 159, T160Y, T160F, вставки I в положении 161, S161P, T163P, вставки N в положении 163, C164K и C164M относительно SEQ ID NO: 2351.

46. Белок CRISPR класса 2, типа V по любому из пп. 36-45, содержащий модификацию в одном или более положениях аминокислот в домене OBD-I относительно SEQ ID NO: 2342, выбранных из группы, состоящей из I3, K4, R5, I6, N7, K8, K15, D16, N18, P27, M28, V33, R34, M36, R41, L47, R48, E52, P55 и Q56.

47. Белок CRISPR класса 2, типа V по п. 46, в котором одна или более модификаций в одном или более положениях аминокислот в домене OBD-I выбраны из группы, состоящей из вставки G в положении 3, I3G, I3E, вставки G в положении 4, K4G, K4P, K4S, K4W, K4W, R5P, вставки P в положении 5, вставки G в положении 5, R5S, вставки S в положении 5, R5A, R5P, R5G, R5L, I6A, I6L, вставки G в положении 6, N7Q, N7L, N7S, K8G, K15F, D16W, вставки F в положении 16, вставки F18, вставки P в положении 27, M28P, M28H, V33T, R34P, M36Y, R41P, L47P, вставки P в положении 48, E52P, вставки P в положении 55, делеции P в положении 55 и делеции Q в положении 56, Q56S, Q56P, вставки D в положении 56, вставки T в положении 56 и Q56P относительно SEQ ID NO: 2342.

48. Белок CRISPR класса 2, типа V по любому из пп. 36-47, содержащий модификацию в одном или более положениях аминокислот в домене OBD-II относительно SEQ ID NO: 2347, выбранных из группы, состоящей из S2, I3, L4, K11, V24, K37, R42, A53, T58, K63, M70, I82, Q92, G93, K110, L121, R124, R141, E143, V144 и L145.

49. Белок CRISPR класса 2, типа V по п. 48, в котором одна или более модификаций в одном или более положениях аминокислот в домене OBD-II выбраны из группы, состоящей из делеции S в положении 2, I3R, I3K, делеции I в положении 3 и делеции L4, делеции L в положении 4, K11T, вставки P в положении 24, K37G, R42E, вставки S в положении 53, вставки R в положении 58, делеции K в положении 63, M70T, I82T, Q92I, Q92F, Q92V, Q92A, вставки A в положении 93, K110Q, R115Q, L121T, вставки A в положении 124, вставки R в положении 141, вставки D в положении 143, вставки A в положении 143, вставки W в положении 144 и вставки A в положении 145 относительно SEQ ID NO: 2342.

50. Белок CRISPR класса 2, типа V по любому из пп. 36-49, содержащий модификацию в одном или более положениях аминокислот в домене TSL относительно SEQ ID NO: 2349, выбранных из группы, состоящей из S1, N2, C3, G4, F5, I7, K18, V58, S67, T76, G78, S80, G81, E82, S85, V96 и E98.

51. Белок CRISPR класса 2, типа V по п. 50, в котором одна или более модификаций в одном или более положениях аминокислот в домене OBD-II выбраны из группы, состоящей из вставки M в положении 1, делеции N в положении 2, вставки V в положении 2, C3S, вставки G в положении 4, вставки W в положении 4, F5P, вставки W в положении 7, K18G, V58D, вставки A в положении 67, T76E, T76D, T76N, G78D, делеции S в положении 80, делеции G в положении 81, вставки E в положении 82, вставки N в положении 82, S85I, V96C, V96T и E98D относительно SEQ ID NO: 2349.

52. Белок CRISPR класса 2, типа V по любому из пп. 29-51, демонстрирующий улучшенную характеристику относительно SEQ ID NO: 2, причем улучшенная характеристика включает в себя повышенную аффинность связывания с гРНК, повышенную аффинность связывания с целевой нуклеиновой кислотой, улучшенную способность использовать более широкий спектр последовательностей смежного с протоспейсером мотива (PAM) при редактировании целевой нуклеиновой кислоты, улучшенное раскручивание целевой нуклеиновой кислоты, повышенную активность редактирования, улучшенную эффективность редактирования, улучшенную специфичность редактирования для расщепления целевой нуклеиновой кислоты, сниженное нецелевое редактирование или расщепление целевой нуклеиновой кислоты, повышенную процентную долю эукариотического генома, которая может быть отредактирована, повышенную активность нуклеазы, повышенную загрузку целевой цепи для двухцепочечного расщепления, сниженную загрузку целевой цепи для одноцепочечного никирования, повышенное связывание нецелевой цепи ДНК, улучшенную стабильность белка, повышенную стабильность комплекса белок : гРНК (RNP) и улучшенные характеристики слияния.

53. Белок CRISPR класса 2, типа V по п. 52, причем улучшенная характеристика включает в себя повышенную активность расщепления последовательности целевой нуклеиновой кислоты, содержащей PAM-последовательность TTC, ATC, GTC или CTC.

54. Белок CRISPR класса 2, типа V по п. 53, причем улучшенная характеристика включает в себя повышенную активность расщепления последовательности целевой нуклеиновой кислоты, содержащей PAM-последовательность ATC или CTC, относительно активности расщепления у последовательности SEQ ID NO: 416.

55. Белок CRISPR класса 2, типа V по п. 54, причем улучшенная активность расщепления представляет собой оценку обогащения (log_2), которая на по меньшей мере примерно 1,5, по меньшей мере примерно 2,0, по меньшей мере примерно 2,5, по меньшей мере примерно 3, по меньшей мере примерно 3,5, по меньшей мере примерно 4, по меньшей мере примерно 4,5, по меньшей мере примерно 5, по меньшей мере примерно 6, по меньшей мере примерно 7, по меньшей мере примерно 8 или более выше, чем оценка у последовательности SEQ ID NO: 416 в анализе in vitro.

56. Белок CRISPR класса 2, типа V по п. 54, причем улучшенная характеристика включает в себя повышенную активность расщепления последовательности целевой нуклеиновой кислоты, содержащей PAM-последовательность CTC, относительно

последовательности SEQ ID NO: 416.

57. Белок CRISPR класса 2, типа V по п. 56, причем улучшенная активность расщепления представляет собой оценку обогащения (log_2), которая на по меньшей мере примерно 2, по меньшей мере примерно 2,5, по меньшей мере примерно 3, по меньшей мере примерно 3,5, по меньшей мере примерно 4, по меньшей мере примерно 4,5, по меньшей мере примерно 5 или по меньшей мере примерно 6 или более выше, чем оценка у последовательности SEQ ID NO: 416 в анализе in vitro.

58. Белок CRISPR класса 2, типа V по п. 53, причем улучшенная характеристика включает в себя повышенную активность расщепления последовательности целевой нуклеиновой кислоты, содержащей PAM-последовательность TTC, относительно последовательности SEQ ID NO: 416.

59. Белок CRISPR класса 2, типа V по п. 58, причем улучшенная активность расщепления представляет собой оценку обогащения, которая на по меньшей мере примерно 1,5, по меньшей мере примерно 2,0, по меньшей мере примерно 2,5, по меньшей мере примерно 3, по меньшей мере примерно 3,5, по меньшей мере примерно 4, по меньшей мере примерно 4,5, по меньшей мере примерно 5 или по меньшей мере примерно 6 log_2 или более выше, чем у последовательности SEQ ID NO: 416 в анализе in vitro.

60. Белок CRISPR класса 2, типа V по п. 52, причем улучшенная характеристика включает в себя повышенную специфичность расщепления последовательности целевой нуклеиновой кислоты относительно последовательности SEQ ID NO: 416.

61. Белок CRISPR класса 2, типа V по п. 60, в котором повышенная специфичность представляет собой оценку обогащения, которая на по меньшей мере примерно 2,0, по меньшей мере примерно 2,5, по меньшей мере примерно 3, по меньшей мере примерно 3,5, по меньшей мере примерно 4, по меньшей мере примерно 4,5, по меньшей мере примерно 5 или по меньшей мере примерно 6 log_2 или более выше, чем у последовательности SEQ ID NO: 416 в анализе in vitro.

62. Белок CRISPR класса 2, типа V по п. 52, причем улучшенная характеристика включает в себя сниженное нецелевое расщепление последовательности целевой нуклеиновой кислоты.

63. Белок CRISPR класса 2, типа V по любому из пп. 29-62, имеющий последовательность, выбранную из группы, состоящей из последовательностей SEQ ID NO: 415-592 и 1147-1231, как указано в таблице 3, или последовательность, имеющую идентичность последовательности с ней по меньшей мере примерно 50%, по меньшей мере примерно 60%, по меньшей мере примерно 70%, по меньшей мере примерно 80%, по меньшей мере примерно 90%, или по меньшей мере примерно 95%, или по меньшей мере примерно 96%, или по меньшей мере примерно 97%, или по меньшей мере примерно 98%, или по меньшей мере примерно 99%.

64. Белок CRISPR класса 2, типа V по любому из пп. 29-62, содержащий последовательность, выбранную из группы, состоящей из SEQ ID NO: 415-592 и 1147-1231, как указано в таблице 3.

65. Белок CRISPR класса 2, типа V по любому из пп. 29-64, содержащий один или более сигналов ядерной локализации (NLS).

66. Белок CRISPR класса 2, типа V по п. 65, в котором один или более NLS выбраны из группы последовательностей, состоящей из PKKKRKV (SEQ ID NO: 352), KRPAATKKAGQAKKKK (SEQ ID NO: 353), PAAKRVKLD (SEQ ID NO: 354), RQRRNELKRSP (SEQ ID NO: 355), NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY (SEQ ID NO: 356), RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV (SEQ ID NO: 357), VSRKRPRP (SEQ ID NO: 358), PPKKARED (SEQ ID NO: (359), PQPKKKPL (SEQ ID NO: 360), SALIKKKKKMAP (SEQ ID NO: 361), DRLRR (SEQ ID NO: 362), PKQKKRK (SEQ ID NO: 363), RKLKKKIKKL (SEQ ID NO: 364), REKKKFLKRR (SEQ ID NO: 365), KRKGDEVDGVDEVAKKKSKK (SEQ ID NO: 366), RKCLQAGMNLEARKTKK (SEQ ID NO: 367), PRPRKIPR (SEQ ID NO: 368), PPRKKRTVV (SEQ ID NO: 369), NLSKKKKRKREK (SEQ ID NO: 370), RRPSRPFRKP (SEQ ID NO: 371), KRPRSPSS (SEQ ID NO: 372), KRGINDRNFWRGENERKTR (SEQ ID NO: 373), PRPPKMARYDN (SEQ ID NO: 374), KRSFSKAF (SEQ ID NO: 375), KLKIKRPVK (SEQ ID NO: 376), PKKKRKVPPPPAAKRVKLD (SEQ ID NO: 377), PKTRRRPRRSQRKRPPT (SEQ ID NO: 378), SRRRKANPTKLSENAKKLAKEVEN (SEQ ID NO: 379), KTRRRPRRSQRKRPPT (SEQ ID NO: 380), RRKKRRPRRKKRR (SEQ ID NO: 381), PKKKSRKPKKKSRK (SEQ ID NO: 382), HKKKHPDASVNFSEFSK (SEQ ID NO: 383), QRPGPYDRPQRPGPYDRP (SEQ ID NO: 384), LSPSLSPLLSPSLSPL (SEQ ID NO: 385), RGKGGKGLGKGGAKRHRK (SEQ ID NO: 386), PKRGRGRPKRGRGR (SEQ ID NO: 387), PKKKRKVPPPPKKKRKV (SEQ ID NO: 389), PAKRARRGYKC (SEQ ID NO: 63), KLGPRKATGRW (SEQ ID NO: 64), PRRKREE (SEQ ID NO: 65), PYRGRKE (SEQ ID NO: 66), PLRKRPRR (SEQ ID NO: 67), PLRKRPRRGSPLRKRPRR (SEQ ID NO:68), PAAKRVKLDGGKRTADGSEFESPKKKRKV (SEQ ID NO:69), PAAKRVKLDGGKRTADGSEFESPKKKRKVGIHGVPAA (SEQ ID NO:70), PAAKRVKLDGGKRTADGSEFESPKKKRKVAEAAAKEAAAKEAAAKA (SEQ ID NO:71), PAAKRVKLDGGKRTADGSEFESPKKKRKVPG (SEQ ID NO:72), KRKGSPERGERKRHW (SEQ ID NO:73), KRTADSQHSTPPKTKRKVEFEPKKKRKV (SEQ ID NO:74) и PKKKRKVGGSKRTADSQHSTPPKTKRKVEFEPKKKRKV (SEQ ID NO:75), и необязательно при этом один или более NLS присоединены к белку CRISPR класса 2, типа V или смежному NLS с помощью линкерного пептида, причем линкерный пептид выбран из группы, состоящей из SR, RS, (G)n (SEQ ID NO: 1023), (GS)n (SEQ ID NO: 1024), (GSGGS)n (SEQ ID NO: 399), (GGSGGS)n (SEQ ID NO: 400), (GGGS)n (SEQ I DNO: 401), GGSG (SEQ ID NO: 402), GGSGG (SEQ ID NO: 403), GSGSG (SEQ ID NO: 404), GSGGG (SEQ ID NO: 405), GGGSG (SEQ ID NO: 406), GSSSG (SEQ ID NO: 407), GPGP (SEQ ID NO: 408), GGP, PPP, PPAPPA (SEQ ID NO: 409), PPPG (SEQ ID NO: 24), PPPGPPP (SEQ ID NO: 410), PPP(GGGS)n (SEQ ID NO: 25), (GGGS)nPPP (SEQ ID NO: 26), AEAAAKEAAAKEAAAKA (SEQ ID NO: 1025) и TPPKTKRKVEFE (SEQ ID NO: 27), где n равно от 1 до 5.

67. Белок CRISPR класса 2, типа V по п. 65 или п. 66, в котором один или более NLS расположены на С-конце белка или рядом с ним.

68. Белок CRISPR класса 2, типа V по п. 65 или п. 66, в котором один или более NLS расположены на N-конце белка или рядом с ним.

69. Белок CRISPR класса 2, типа V по п. 65 или п. 66, содержащий по меньшей мере два NLS, причем по меньшей мере два NLS расположены на N-конце белка или рядом с ним и на С-конце белка или рядом с ним.

70. Белок CRISPR класса 2, типа V по любому из пп. 29-69, способный образовывать рибонуклеопротеиновый комплекс (RNP) с гРНК.

71. Белок CRISPR класса 2, типа V по п. 70, причем RNP демонстрирует по меньшей мере одну или более улучшенных характеристик по сравнению с RNP из референсного белка с любой из SEQ ID NO: 1-3 и гРНК с SEQ ID NO: 4 или SEQ ID NO: 5.

72. Белок CRISPR класса 2, типа V по 71, причем улучшенная характеристика выбрана из группы, состоящей из повышенной аффинности связывания с гидовой нуклеиновой кислотой (гРНК); повышенной аффинности связывания с целевой нуклеиновой кислотой; улучшенной способности использовать более широкий спектр из одной или более PAM-последовательностей, включая ATC, CTC, GTC или TTC, при редактировании целевой нуклеиновой кислоты; повышенного раскручивания целевой нуклеиновой кислоты; повышенной активности редактирования; повышенной эффективности редактирования; повышенной специфичности редактирования целевой нуклеиновой кислоты; повышенной нуклеазной активности; повышенной загрузки целевой цепи для двухцепочечного расщепления; сниженной загрузки целевой цепи для одноцепочечного никирования; сниженного нецелевого расщепления целевой нуклеиновой кислоты; повышенного связывания нецелевой цепи нуклеиновой кислоты и повышенной стабильности комплекса белок : гРНК (RNP).

73. Белок CRISPR класса 2, типа V по п. 71 или п. 72, причем улучшенная характеристика RNP имеет кратность повышения от по меньшей мере примерно 1,1 до примерно 100 000 раз относительно RNP из референсного белка с SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3 и гРНК с SEQ ID NO: 4 или SEQ ID NO: 5.

74. Белок CRISPR класса 2, типа V по п. 71 или п. 72, причем улучшенная характеристика RNP имеет кратность повышения по меньшей мере примерно 10 раз, по меньшей мере примерно 100 раз, по меньшей мере примерно 1000 раз или по меньшей мере примерно 10 000 раз относительно RNP из референсного белка с SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3 и гРНК с SEQ ID NO: 4 или 5.

75. Белок CRISPR класса 2, типа V по любому из пп. 71-74, причем улучшенная характеристика RNP имеет кратность улучшения эффективности редактирования от 1,1 до 100 раз по сравнению с RNP из референсного белка с SEQ ID NO: 2 и гРНК, содержащей SEQ ID NO: 4 или 5.

76. Пара для редактирования генов, содержащая гРНК и белок CRISPR класса 2, типа V, причем пара содержит:

a. гРНК по любому из пп. 25-28; и

b. белок CRISPR класса 2, типа V по любому из пп. 29-75.

77. Пара для редактирования генов по п. 76, в которой гРНК и белок CRISPR класса 2, типа V способны образовывать рибонуклеопротеиновый комплекс (RNP).

78. Пара для редактирования генов по п. 76 или п. 77, в которой гРНК и белок CRISPR класса 2, типа V связаны вместе в рибонуклеопротеиновый комплекс (RNP).

79. Пара для редактирования генов по п. 77 или п. 78, причем RNP из белка CRISPR класса 2, типа V и гРНК демонстрирует по меньшей мере одну или более улучшенных характеристик по сравнению с RNP из референсного белка с SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3 и гРНК, содержащей последовательность SEQ ID NO: 4 или SEQ ID NO: 5.

80. Пара для редактирования генов по п. 79, причем улучшенная характеристика выбрана из одного или более из группы, состоящей из повышенной аффинности связывания белка CRISPR класса 2, типа V с гРНК; повышенной аффинности связывания с целевой нуклеиновой кислотой; улучшенной способности использовать более широкий спектр из одной или более PAM-последовательностей, включая ATC, CTC, GTC или TTC, при редактировании целевой нуклеиновой кислоты; повышенного раскручивания целевой нуклеиновой кислоты; повышенной активности редактирования; повышенной эффективности редактирования; повышенной специфичности редактирования целевой нуклеиновой кислоты; повышенной нуклеазной активности; повышенной загрузки целевой цепи для двухцепочечного расщепления; сниженной загрузки целевой цепи для одноцепочечного никирования; сниженного нецелевого расщепления целевой нуклеиновой кислоты; повышенного связывания нецелевой цепи нуклеиновой кислоты повышенной стабильности комплекса белок : гРНК (RNP) и улучшенных характеристик слияния.

81. Пара для редактирования генов по п. 79 или п. 80, причем улучшенная характеристика RNP из белка CRISPR класса 2, типа V и гРНК имеет кратность повышения от по меньшей мере примерно 1,1 до примерно 100 раз или более относительно RNP из референсного белка с SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3 и гРНК, содержащей последовательность SEQ ID NO: 4 или SEQ ID NO: 5, в сопоставимой системе анализа in vitro.

82. Пара для редактирования генов по п. 79 или п. 80, причем улучшенная характеристика белка CRISPR класса 2, типа V имеет кратность повышения по меньшей мере примерно 1,1 раза, по меньшей мере примерно 2 раза, по меньшей мере примерно 10 раз, по меньшей мере примерно 100 раз или более относительно референсного белка с SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3 и гРНК, содержащей последовательность SEQ ID NO: 4 или SEQ ID NO: 5, в сопоставимой системе анализа in vitro.

83. Пара для редактирования генов по любому из пп. 77-82, причем RNP, содержащий белок CRISPR класса 2, типа V и гРНК, демонстрирует более высокие эффективность редактирования и/или связывание последовательности целевой нуклеиновой кислоты в целевой нуклеиновой кислоте, когда любая из PAM-

последовательностей TTC, ATC, GTC или CTC расположена на расстоянии 1 нуклеотида от 5'-конца нецелевой цепи протоспейсера, имеющей идентичность с нацеливающей последовательностью гРНК, в системе клеточного анализа по сравнению с эффективностью редактирования и/или связыванием у RNP, содержащего референсный белок с SEQ ID NO: 1, SEQ ID NO: 2 или SEQ ID NO: 3 и референсную гРНК, в сопоставимой системе анализа.

84. Пара для редактирования генов по п. 83, причем PAM-последовательность представляет собой TTC.

85. Пара для редактирования генов по п. 83, причем PAM-последовательность представляет собой ATC.

86. Пара для редактирования генов по п. 83, причем PAM-последовательность представляет собой CTC.

87. Пара для редактирования генов по п. 83, причем PAM-последовательность представляет собой GTC.

88. Пара для редактирования генов по любому из пп. 83-87, в которой RNP, содержащий белок CRISPR класса 2, типа V и гРНК, демонстрирует повышенную аффинность связывания с одной или более PAM-последовательностями, которая в по меньшей мере 1,5 раза, по меньшей мере 2 раза, по меньшей мере 4 раза, по меньшей мере 10 раз, по меньшей мере 20 раз, по меньшей мере 30 раз или по меньшей мере 40 раз выше, чем аффинность связывания RNP из любого из референсных белков с SEQ ID NO:1-3 и гРНК с SEQ ID NO: 4 или SEQ ID NO: 5, при оценке в сопоставимой системе анализа in vitro.

89. Пара для редактирования генов по любому из пп. 77-88, в которой RNP из белка CRISPR класса 2, типа V и гРНК демонстрирует повышенную эффективность редактирования, которая в по меньшей мере 1,5 раза, по меньшей мере 2 раза, по меньшей мере 4 раза, по меньшей мере 10 раз, по меньшей мере 20 раз, по меньшей мере 30 раз или по меньшей мере 40 раз выше, чем эффективность редактирования у RNP из любого из референсных белков с SEQ ID NO:1-3 и гРНК с SEQ ID NO: 4 или SEQ ID NO: 5, при оценке в сопоставимой системе анализа in vitro.

90. Пара для редактирования генов по любому из пп. 77-89, в которой белок CRISPR класса 2, типа V и гРНК способны образовывать RNP, имеющий на по меньшей мере примерно 5%, по меньшей мере примерно 10%, по меньшей мере примерно 15% или по меньшей мере примерно 20% более высокую процентную долю компетентной в отношении расщепления конформации по сравнению с RNP из любого из референсных белков с SEQ ID NO: 1-3 и гРНК с SEQ ID NO: 4 или SEQ ID NO: 5, при оценке в сопоставимой системе анализа in vitro.

91. Пара для редактирования генов по любому из пп. 77-90, в которой RNP, содержащий белок CRISPR класса 2, типа V и гРНК, демонстрирует скорость расщепления целевой нуклеиновой кислоты в анализе зависимости от времени in vitro, которая в по меньшей мере приблизительно 5 раз, по меньшей мере приблизительно 10 раз или по меньшей мере приблизительно 20 раз выше по сравнению с RNP из любого из референсных

белков с SEQ ID NO: 1-3 и гРНК с SEQ ID NO: 4 или SEQ ID NO: 5, при оценке в сопоставимой системе анализа in vitro.

92. Пара для редактирования генов по любому из пп. 77-91, в которой RNP, содержащий белок CRISPR класса 2, типа V и гРНК, демонстрирует более высокую процентную долю редактирования целевой нуклеиновой кислоты в анализе зависимости от времени in vitro, которая в по меньшей мере приблизительно 5 раз, по меньшей мере приблизительно 10 раз, по меньшей мере приблизительно 20 раз или по меньшей мере приблизительно 100 раз выше по сравнению с RNP из любого из референсных белков с SEQ ID NO: 1-3 и гРНК с SEQ ID NO: 4 или SEQ ID NO: 5 при оценке в сопоставимой системе анализа in vitro.

93. Каталитически неактивный белок CRISPR класса 2, типа V, содержащий последовательность, выбранную из группы, состоящей из SEQ NO: 44-62 и 1232-1235, как показано в таблице 7, или последовательность, имеющую по меньшей мере 60%, по меньшей мере 70%, по меньшей мере 80%, по меньшей мере 90%, по меньшей мере 95%, по меньшей мере 98% или по меньшей мере 99%.

94. Каталитически неактивный белок CRISPR класса 2, типа V, содержащий последовательность, выбранную из группы, состоящей из SEQ NO: 44-62 и 1232-1235, как указано в таблице 7.

95. Белок CRISPR класса 2, типа V по п. 93 или п. 94, причем и RNP из каталитически неактивного белка CRISPR класса 2, типа V, и гРНК по любому из пп. 25-28 сохраняет способность связывать целевую нуклеиновую кислоту.

96. Нуклеиновая кислота, содержащая последовательность, которая кодирует каркас гРНК по любому из пп. 1-24 или гРНК по любому из пп. 25-28.

97. Нуклеиновая кислота, содержащая последовательность, которая кодирует белок CRISPR класса 2, типа V по любому из пп. 29-75.

98. Нуклеиновая кислота по п. 97, в которой последовательность, которая кодирует белок CRISPR класса 2, типа V, является кодон-оптимизированной для экспрессии в эукариотической клетке.

99. Вектор, содержащий гРНК по любому из пп. 25-28, белок CRISPR класса 2, типа V по любому из пп. 29-75 или нуклеиновую кислоту по любому из пп. 96-98.

100. Вектор по п. 99, содержащий промотор.

101. Вектор по п. 99 или п. 100, выбранный из группы, состоящей из ретровирусного вектора, лентивирусного вектора, аденовирусного вектора, аденоассоциированного вирусного (AAV) вектора, вектора на основе вируса простого герпеса (HSV), частицы для доставки CasX (XDP), плазмиды, миникольца, наноплазмиды, ДНК-вектора и РНК-вектора.

102. Вектор по п. 101, представляющий собой AAV-вектор.

103. Вектор по п. 102, причем AAV-вектор выбран из AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAV-Rh74 или AAVRh10.

104. Вектор по п. 101, представляющий собой ретровирусный вектор.

105. Вектор по п. 101, представляющий собой XDP, содержащую один или более

компонентов полибелка gag.

106. Вектор по п. 105, в котором один или более компонентов полибелка gag выбраны из группы, состоящей из матриксного белка (MA), нуклеокапсидного белка (NC), капсидного белка (CA), пептида p1, пептида p6, пептида P2A, пептида P2B, пептида P10, пептида p12, пептида PP21/24, пептида P12/P3/P8, пептида P20 и сайта расщепления протеазой.

107. Вектор по п. 105 или п. 106, причем белок CRISPR класса 2, типа V и гРНК связаны друг с другом в RNP.

108. Вектор по любому из пп. 105-107, содержащий гликопротеиновый фактор тропизма.

109. Вектор по п. 108, причем гликопротеиновый фактор тропизма обладает аффинностью связывания с маркером клеточной поверхности клетки-мишени и облегчает проникновение XDP в клетку-мишень.

110. Вектор по любому из пп. 99-109, содержащий донорную матрицу.

111. Клетка-хозяин, содержащая вектор по любому из пп. 99-110.

112. Клетка-хозяин по п. 111, выбранная из группы, состоящей из клеток-фибробластов почки новорожденного хомяка (BHK), клеток эмбриональной почки человека 293 (HEK293), клеток эмбриональной почки человека 293T (HEK293T), клеток NS0, клеток SP2/0, клеток миеломы YO, мышиных клеток миеломы P3X63, клеток PER, клеток PER.C6, клеток гибридомы, клеток NIH3T3, клеток, происходящих от CV-1 (обезьяньих), с генетическим материалом SV40 (COS), клеток HeLa, клеток яичника китайского хомячка (CHO) или дрожжевых клеток.

113. Способ модификации целевой нуклеиновой кислоты в клетке, включающий приведение целевой нуклеиновой кислоты клетки в контакт с: i) парой для редактирования генов по любому из пп. 76-92; ii) парой для редактирования генов по любому из пп. 76-92 вместе с донорной матрицей; iii) одной или более нуклеиновыми кислотами, кодирующими пару для редактирования генов по п. (i) или (ii); iv) вектором, содержащим нуклеиновую кислоту по п. (iii); v) XDP, содержащей пару для редактирования генов по п. (i) или (ii); или vi) комбинацией двух или более из пп. (i)-(v), причем приведение целевой нуклеиновой кислоты в контакт модифицирует целевую нуклеиновую кислоту.

114. Способ по п. 113, включающий приведение целевой молекулы в контакт со множеством пар для редактирования генов, содержащих первую и вторую или множество гРНК, содержащих нацеливающие последовательности, комплементарные разным или перекрывающимся областям целевой нуклеиновой кислоты.

115. Способ по п. 113, включающий приведение целевой молекулы в контакт со множеством нуклеиновых кислот, кодирующих пары для редактирования генов, содержащие первую и вторую или множество гРНК, содержащих нацеливающие последовательности, комплементарные разным или перекрывающимся областям целевой нуклеиновой кислоты.

116. Способ по п. 113, включающий приведение целевой молекулы в контакт со

множеством XDP, содержащих пары для редактирования генов, которые содержат первую и вторую или множество гРНК, содержащих нацеливающие последовательности, комплементарные разным или перекрывающимся областям целевой нуклеиновой кислоты.

117. Способ по любому из п. 113, в котором приведение в контакт включает в себя связывание целевой нуклеиновой кислоты с парой для редактирования генов и введение одного или более одноцепочечных разрывов в целевую нуклеиновую кислоту, причем модификация включает в себя внесение мутации, вставки или делеции в целевую нуклеиновую кислоту.

118. Способ по любому из пп. 113-116, в котором приведение в контакт включает в себя связывание целевой нуклеиновой кислоты и введение одного или более двухцепочечных разрывов в целевую нуклеиновую кислоту, причем модификация включает в себя введение мутации, вставки или делеции в целевую нуклеиновую кислоту.

119. Способ по любому из пп. 113-118, включающий приведение целевой нуклеиновой кислоты в контакт с нуклеотидной последовательностью нуклеиновой кислоты донорной матрицы, причем донорная матрица содержит нуклеотидную последовательность, имеющую гомологию с целевой нуклеиновой кислотой.

120. Способ по п. 119, в котором донорная матрица содержит гомологичные плечи на 5’- и 3’-концах донорной матрицы.

121. Способ по п. 119 или п. 120, в котором донорную матрицу вставляют в целевую нуклеиновую кислоту в сайте разрыва посредством направляемой гомологией репарации.

122. Способ по п. 121, в котором донорную матрицу вставляют в целевую нуклеиновую кислоту в сайте разрыва посредством негомологичного соединения концов (NHEJ) или соединения концов, обусловленного микрогомологией (MMEJ).

123. Способ по любому из пп. 113-122, в котором модификация клетки происходит *in vitro*.

124. Способ по любому из пп. 113-122, в котором модификация клетки происходит *in vivo*.

125. Способ по любому из пп. 113-124, в котором клетка представляет собой эукариотическую клетку.

126. Способ по п. 125, в котором эукариотическая клетка выбрана из группы, состоящей из клетки грызуна, клетки мыши, клетки крысы, клетки примата и клетки примата, не являющегося человеком.

127. Способ по п. 125, в котором эукариотическая клетка представляет собой клетку человека.

128. Способ по любому из пп. 113-127, в котором клетка выбрана из группы, состоящей из эмбриональной стволовой клетки, индуцированной плюрипотентной стволовой клетки, зародышевой клетки, фибробласта, олигодендроцита, глиальной клетки, гематопоэтической стволовой клетки, нейрональной клетки-предшественника, нейрона, мышечной клетки, костной клетки, гепатоцита, клетки поджелудочной железы, клетки сетчатки, раковой клетки, Т-клетки, В-клетки, NK-клетки, фетального кардиомиоцита,

миофибробласта, мезенхимальной стволовой клетки, аутотрансплантированного гипертрофированного кардиомиоцита, адипоцита, тотипотентной клетки, плюрипотентной клетки, стволовой клетки крови, миобласта, стволовой клетки взрослого, клетки костного мозга, мезенхимальной клетки, паренхимной клетки, эпителиальной клетки, эндотелиальной клетки, мезотелиальной клетки, клетки-фибробласта, клетки-остеобласта, клетки-хондроцита, экзогенной клетки, эндогенной клетки, стволовой клетки, гематопоэтической стволовой клетки, клетки-предшественника из костного мозга, миокардиальной клетки, скелетной клетки, фетальной клетки, недифференцированной клетки, мультипотентной клетки-предшественника, унипотентной клетки-предшественника, моноцита, миобласта сердечной мышцы, миобласта скелетной мышцы, макрофага, эндотелиальной клетки капилляров, ксеногенной клетки, аллогенной клетки, аутологичной клетки и постнатальной стволовой клетки.

129. Способ по любому из пп. 124-128, причем клетка находится в субъекте.

130. Способ по п. 129, в котором модификация происходит в клетках субъекта, имеющего мутацию в аллели гена, причем мутация вызывает заболевание или расстройство у субъекта.

131. Способ по п. 130, в котором модификация изменяет мутацию с образованием аллели гена дикого типа или приводит к экспрессии продукта функционального гена.

132. Способ по п. 130, в котором модификация приводит к нокдауну или нокауту аллели гена, вызывающей заболевание или расстройство у субъекта.

133. Способ по любому из пп. 129-132, в котором клетка является аутологичной по отношению к субъекту.

134. Способ по любому из пп. 129-132, в котором клетка является аллогенной аутологичной по отношению к субъекту.

135. Способ по любому из пп. 113-134, в котором вектор представляет собой аденоассоциированный вирусный (AAV) вектор.

136. Способ по п. 135, в котором AAV-вектор представляет собой AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV-Rh74 или AAVRh10.

137. Способ по п. 113, в котором вектор представляет собой лентивирусный вектор.

138. Способ по любому из пп. 113-137, в котором вектор вводят нуждающемуся в этом субъекту с применением терапевтически эффективной дозы.

139. Способ по п. 138, причем субъект выбран из группы, состоящей из мыши, крысы, свиньи и примата, не являющегося человеком.

140. Способ по п. 138, причем субъект является человеком.

141. Способ по любому из пп. 138-140, в котором вектор вводят субъекту в дозе по меньшей мере примерно 1×10^5 векторных геномов/кг (вг/кг), по меньшей мере примерно 1×10^6 вг/кг, по меньшей мере примерно 1×10^7 вг/кг, по меньшей мере примерно 1×10^8 вг/кг, по меньшей мере примерно 1×10^9 вг/кг, по меньшей мере примерно 1×10^{10} вг/кг, по меньшей мере примерно 1×10^{11} вг/кг, по меньшей мере примерно 1×10^{12} вг/кг, по меньшей мере примерно 1×10^{13} вг/кг, по меньшей мере примерно 1×10^{14} вг/кг, по

меньшей мере примерно 1×10^{15} вг/кг или по меньшей мере примерно 1×10^{16} вг/кг.

142. Способ по любому из пп. 138-140, в котором вектор вводят субъекту в дозе от по меньшей мере примерно 1×10^{5} вг/кг до примерно 1×10^{16} вг/кг, от по меньшей мере примерно 1×10^{6} вг/кг до примерно 1×10^{15} вг/кг или от по меньшей мере примерно 1×10^{7} вг/кг до примерно 1×10^{14} вг/кг.

143. Способ по п. 113, в котором вектор представляет собой XDP.

144. Способ по п. 143, в котором XDP вводят нуждающемуся в этом субъекту с применением терапевтически эффективной дозы.

145. Способ по п. 144, в котором XDP вводят субъекту в дозе по меньшей мере примерно 1×10^{5} частиц/кг, по меньшей мере примерно 1×10^{6} частиц/кг, по меньшей мере примерно 1×10^{7} частиц/кг, по меньшей мере примерно 1×10^{8} частиц/кг, по меньшей мере примерно 1×10^{9} частиц/кг, по меньшей мере примерно 1×10^{10} частиц/кг, по меньшей мере примерно 1×10^{11} частиц/кг, по меньшей мере примерно 1×10^{12} частиц/кг, по меньшей мере примерно 1×10^{13} частиц/кг, по меньшей мере примерно 1×10^{14} частиц/кг, по меньшей мере примерно 1×10^{15} частиц/кг, по меньшей мере примерно 1×10^{16} частиц/кг.

146. Способ по п. 143, в котором XDP вводят субъекту в дозе от по меньшей мере примерно 1×10^{5} частиц/кг до примерно 1×10^{16} частиц/кг, или от по меньшей мере примерно 1×10^{6} частиц/кг до примерно 1×10^{15} частиц/кг, или от по меньшей мере примерно 1×10^{7} частиц/кг до примерно 1×10^{14} частиц/кг.

147. Способ по любому из пп. 138-146, в котором вектор вводят путем введения, выбранным из группы, состоящей из интрапаренохиматозного, внутривенного, внутриартериального, интрацеребровентрикулярного, интрацистернального, интратекального, интракраниального и внутрибрюшинного путей, причем способ введения представляет собой инъекцию, трансфузию или имплантацию.

148. Способ по любому из пп. 141-147, в котором вектор вводят субъекту в соответствии со схемой лечения, содержащей одну или более последовательных доз с применением терапевтически эффективной дозы вектора.

149. Способ по п. 148, причем терапевтически эффективную дозу вводят субъекту в виде двух или более доз в течение периода по меньшей мере двух недель, или по меньшей мере одного месяца, или по меньшей мере двух месяцев, или по меньшей мере трех месяцев, или по меньшей мере четырех месяцев, или по меньшей мере пяти месяцев, или по меньшей мере шести месяцев, или раз в год, или каждые 2 или 3 года.

150. Клетка, содержащая целевую нуклеиновую кислоту, модифицированную парой для редактирования генов по любому из пп. 76-92.

151. Клетка, отредактированная способом по любому из пп. 113-149.

152. Клетка по п. 150 или 151, представляющая собой прокариотическую клетку.

153. Клетка по п. 150 или 151, представляющая собой эукариотическую клетку.

154. Клетка по п. 153, причем эукариотическая клетка выбрана из группы, состоящей из клетки грызуна, клетки мыши, клетки крысы, клетки примата и клетки примата, не являющегося человеком.

155. Клетка по п. 153, причем эукариотическая клетка представляют собой клетку человека.

156. Композиция, содержащая белок CRISPR класса 2, типа V по любому из пп. 29-75.

157. Композиция по п. 156, содержащая гРНК по любому из пп. 25-28.

158. Композиция по п. 157, в которой белок и гРНК связаны друг с другом в рибонуклеопротеиновом комплексе (RNP).

159. Композиция по любому из пп. 156-158, содержащая нуклеиновую кислоту донорной матрицы, причем донорная матрица содержит нуклеотидную последовательность, имеющую гомологию с целевой нуклеиновой кислотой.

160. Композиция по любому из пп. 156-159, содержащая буфер, ингибитор нуклеазы, ингибитор протеазы, липосому, терапевтический агент, метку, реагент для визуализации метки или любую комбинацию вышеизложенного.

161. Композиция, содержащая каркас гРНК по любому из пп. 1-24 или гРНК по любому из пп. 25-28.

162. Композиция по п. 161, содержащая белок CRISPR класса 2, типа V по любому из пп. 29-75.

163. Композиция по п. 162, в которой белок CRISPR класса 2, типа V и гРНК связаны друг с другом в рибонуклеопротеиновом комплексе (RNP).

164. Композиция по любому из пп. 161-163, содержащая нуклеиновую кислоту донорной матрицы, причем донорная матрица содержит нуклеотидную последовательность, имеющую гомологию с целевой нуклеиновой кислотой.

165. Композиция по любому из пп. 161-164, содержащая буфер, ингибитор нуклеазы, ингибитор протеазы, липосому, терапевтический агент, метку, реагент для визуализации метки или любую комбинацию вышеизложенного.

166. Композиция, содержащая пару для редактирования генов по любому из пп. 76-92.

167. Композиция по п. 166, содержащая нуклеиновую кислоту донорной матрицы, причем донорная матрица содержит нуклеотидную последовательность, имеющую гомологию с целевой нуклеиновой кислотой.

168. Композиция по п. 166 или п. 167, содержащая буфер, ингибитор нуклеазы, ингибитор протеазы, липосому, терапевтический агент, метку, реагент для визуализации метки или любую комбинацию вышеизложенного.

169. Набор, содержащий белок CRISPR класса 2, типа V по любому из пп. 29-75 и контейнер.

170. Набор по п. 169, содержащий каркас гРНК по любому из пп. 1-24 или гРНК по любому из пп. 25-28.

171. Набор по п. 169 или п. 170, содержащий нуклеиновую кислоту донорной матрицы, причем донорная матрица содержит нуклеотидную последовательность, имеющую гомологию с целевой нуклеотидной последовательностью целевой нуклеиновой

кислоты.

172. Набор по любому из пп. 169-171, содержащий буфер, ингибитор нуклеазы, ингибитор протеазы, липосому, терапевтический агент, метку, реагент для визуализации метки или любую комбинацию вышеизложенного.

173. Набор, содержащий каркас гРНК по любому из пп. 1-24 или гРНК по любому из пп. 25-28.

174. Набор по п. 173, содержащий белок CRISPR класса 2, типа V по любому из пп. 29-75.

175. Набор по п. 173 или п. 174, содержащий нуклеиновую кислоту донорной матрицы, причем донорная матрица содержит нуклеотидную последовательность, имеющую гомологию с целевой нуклеотидной последовательностью целевой нуклеиновой кислоты.

176. Набор по любому из пп. 173-175, содержащий буфер, ингибитор нуклеазы, ингибитор протеазы, липосому, терапевтический агент, метку, реагент для визуализации метки или любую комбинацию вышеизложенного.

177. Набор, содержащий пару для редактирования генов по любому из пп. 76-92.

178. Набор по п. 177, содержащий нуклеиновую кислоту донорной матрицы, причем донорная матрица содержит нуклеотидную последовательность, имеющую гомологию с целевой нуклеиновой кислотой.

179. Набор по п. 177 или п. 178, содержащий буфер, ингибитор нуклеазы, ингибитор протеазы, липосому, терапевтический агент, метку, реагент для визуализации метки или любую комбинацию вышеизложенного.

180. Сконструированный белок CRISPR класса 2, типа V, содержащий любую из последовательностей, перечисленных в таблице 3.

181. гРНК, содержащая любую из последовательностей варианта каркаса гРНК, перечисленных в таблице 2.

182. гРНК по п. 181, в которой один или более урацилов (U) вариантов каркаса гРНК с последовательностями из таблицы 2 заменены тиминами (T).

183. гРНК по п. 182, содержащая нацеливающую последовательность по меньшей мере от 10 до 30 нуклеотидов, комплементарную целевой нуклеиновой кислоте.

184. гРНК по п. 183, в которой нацеливающая последовательность имеет 20 нуклеотидов.

185. гРНК по п. 183, в которой нацеливающая последовательность имеет 19 нуклеотидов.

186. гРНК по п. 183, в которой нацеливающая последовательность имеет 18 нуклеотидов.

187. гРНК по п. 183, в которой нацеливающая последовательность имеет 17 нуклеотидов.

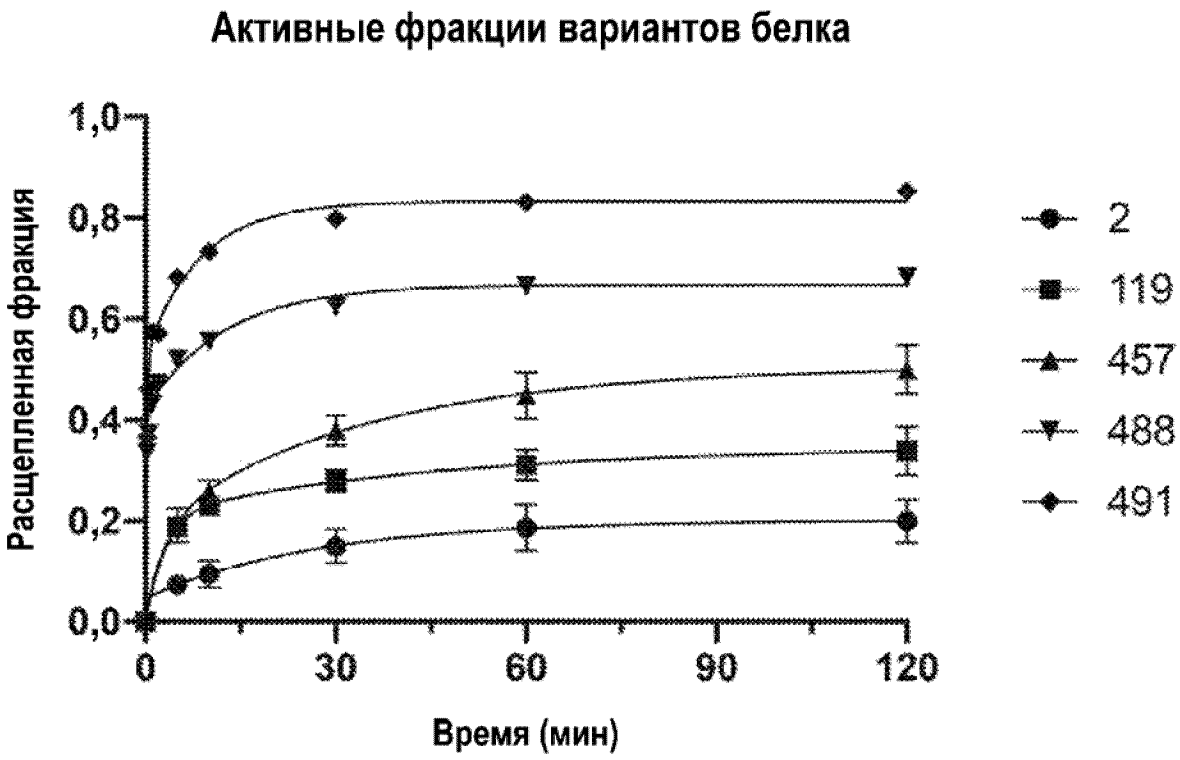
188. гРНК по п. 183, в которой нацеливающая последовательность имеет 16 нуклеотидов.

189. гРНК по п. 183, в которой нацеливающая последовательность имеет 15 нуклеотидов.

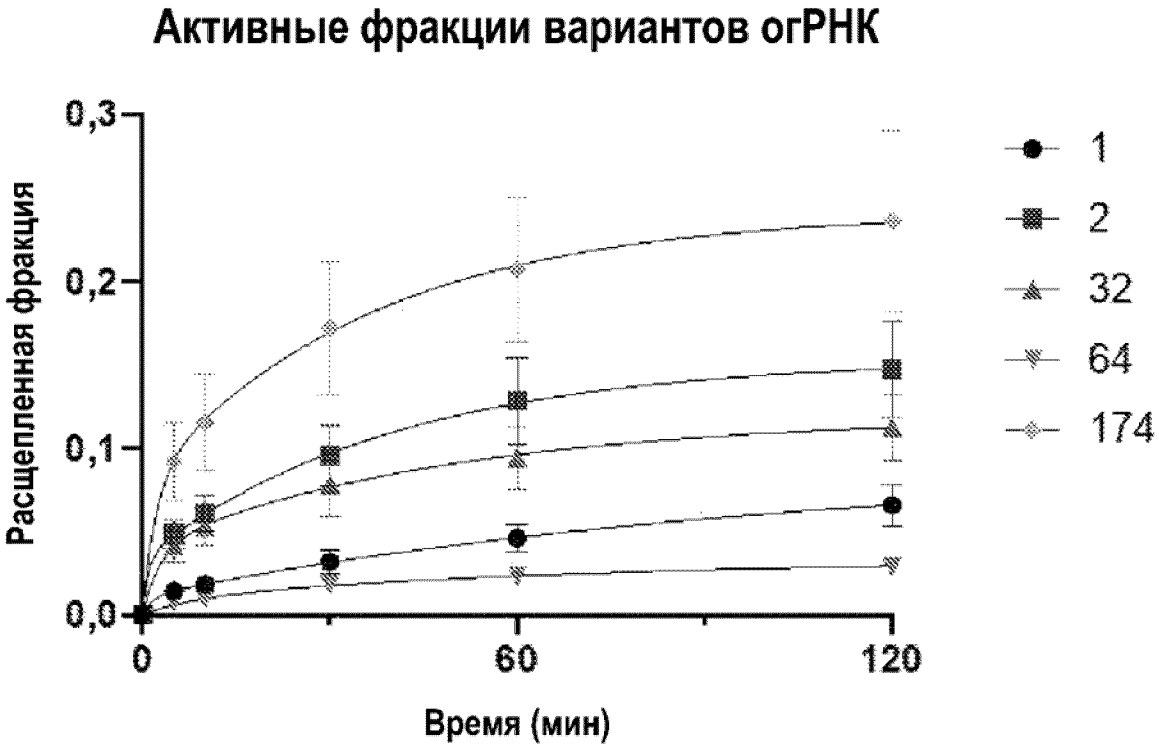
190. Способ лечения заболевания у нуждающегося в этом субъекта, включающий введение субъекту фармацевтически эффективного количества композиции, содержащей: (a) сконструированный белок CRISPR класса 2, типа V по любому из пп. 29-75 и 180 и (b) гРНК по любому из пп. 25-28 и 181-189.

191. Композиция, содержащая: (a) сконструированный белок CRISPR класса 2, типа V по любому из пп. 29-75 и 180 и (b) гРНК по любому из пп. 25-28 и 181-189 для применения в качестве лекарственного средства для лечения субъекта, имеющего заболевание.

По доверенности


Активные фракции вариантов белка
Расщепленная фракция
1,0
0,8
0,6
0,4
0,2
0,0
0
30
60
90
120
Время (мин)
2
119
457
488
491

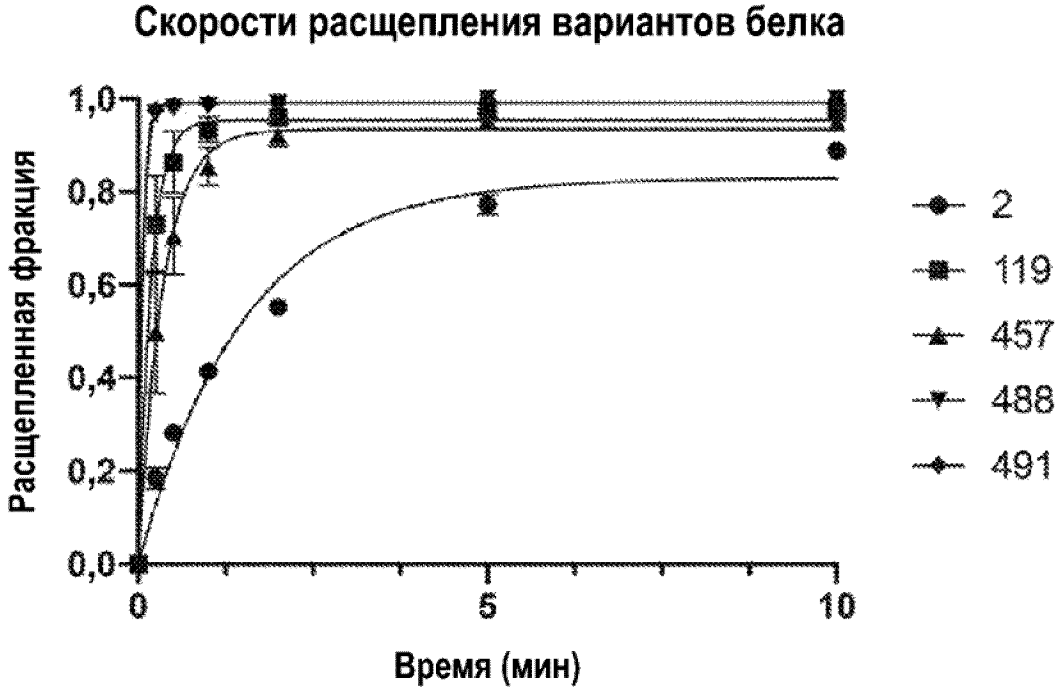
ФИГ. 1


Активные фракции вариантов огРНК
Расщепленная фракция
0,3
0,2
0,1
0,0
0
60
120
Время (мин)
1
2
32
64
174

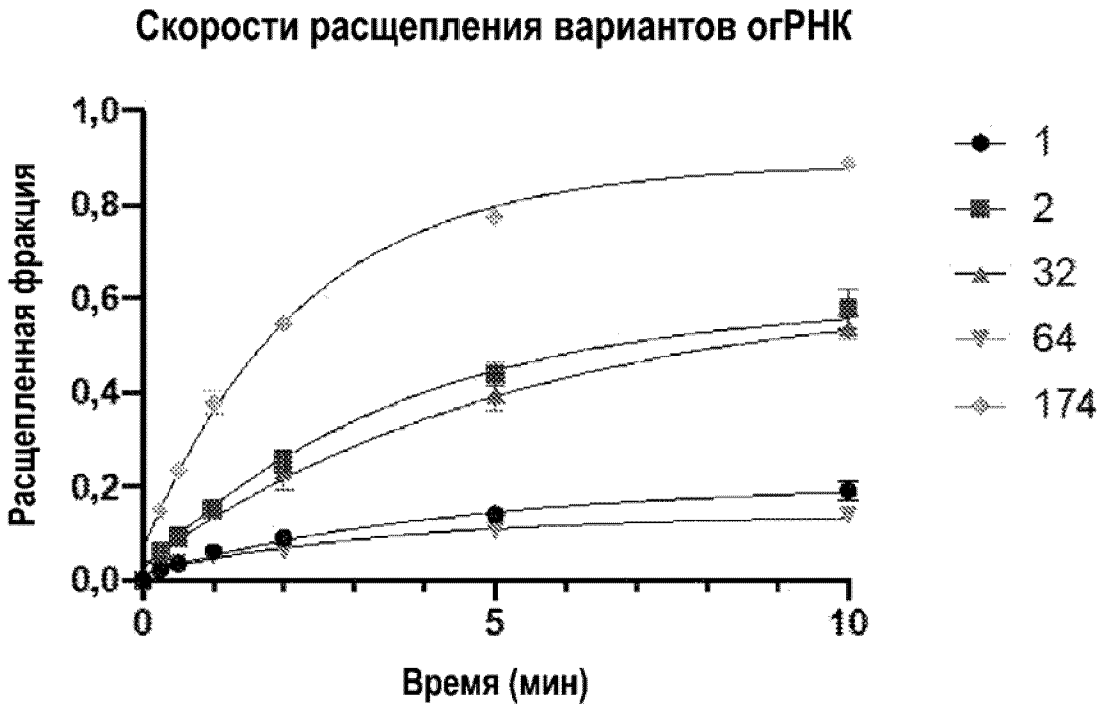
ФИГ. 2


Компетентные фракции в условиях ограничения гида
Расщепленная фракция
0,5
0,4
0,3
0,2
0,1
0,0
0
30
60
90
120
Время (мин)
174
175
185
186
196
214
215

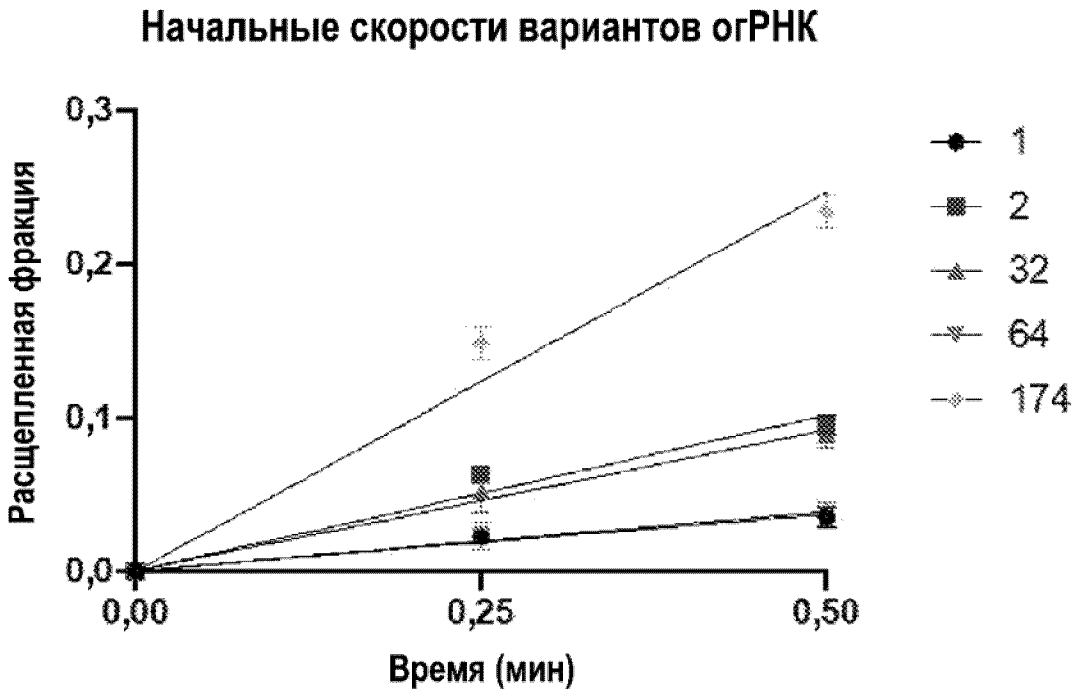
ФИГ. 3


Скорости расщепления вариантов белка
Расщепленная фракция
1,0
0,8
0,6
0,4
0,2
0,0
0
5
10
Время (мин)
2
119
457
488
491

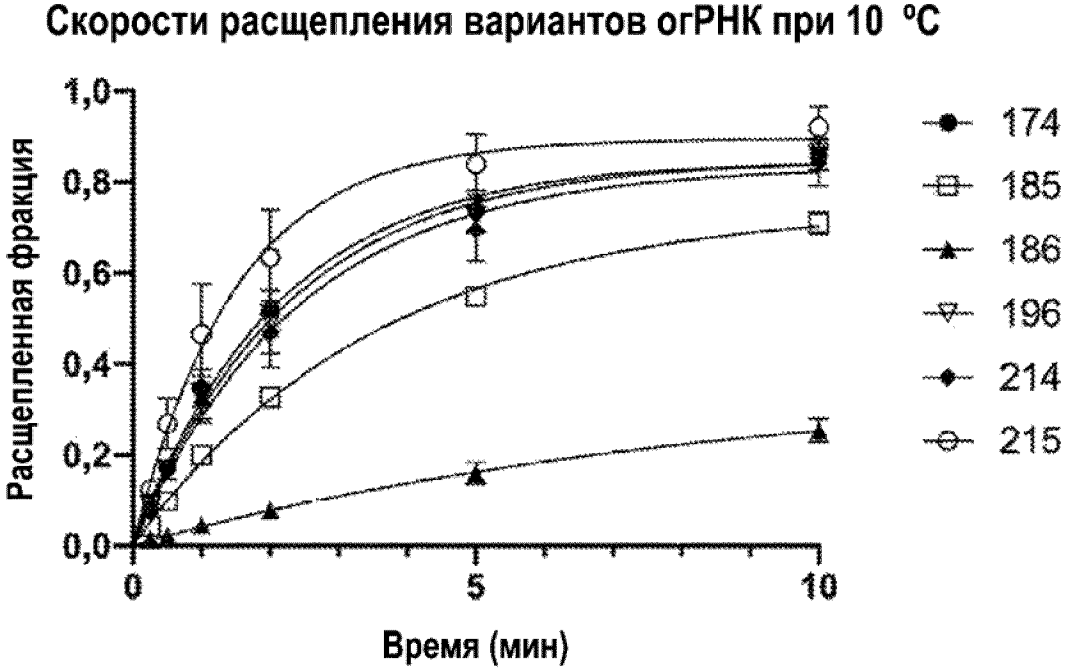
ФИГ. 4


Скорости расщепления вариантов огРНК
Расщепленная фракция
1,0
0,8
0,6
0,4
0,2
0,0
0
5
10
Время (мин)
1
2
32
64
174

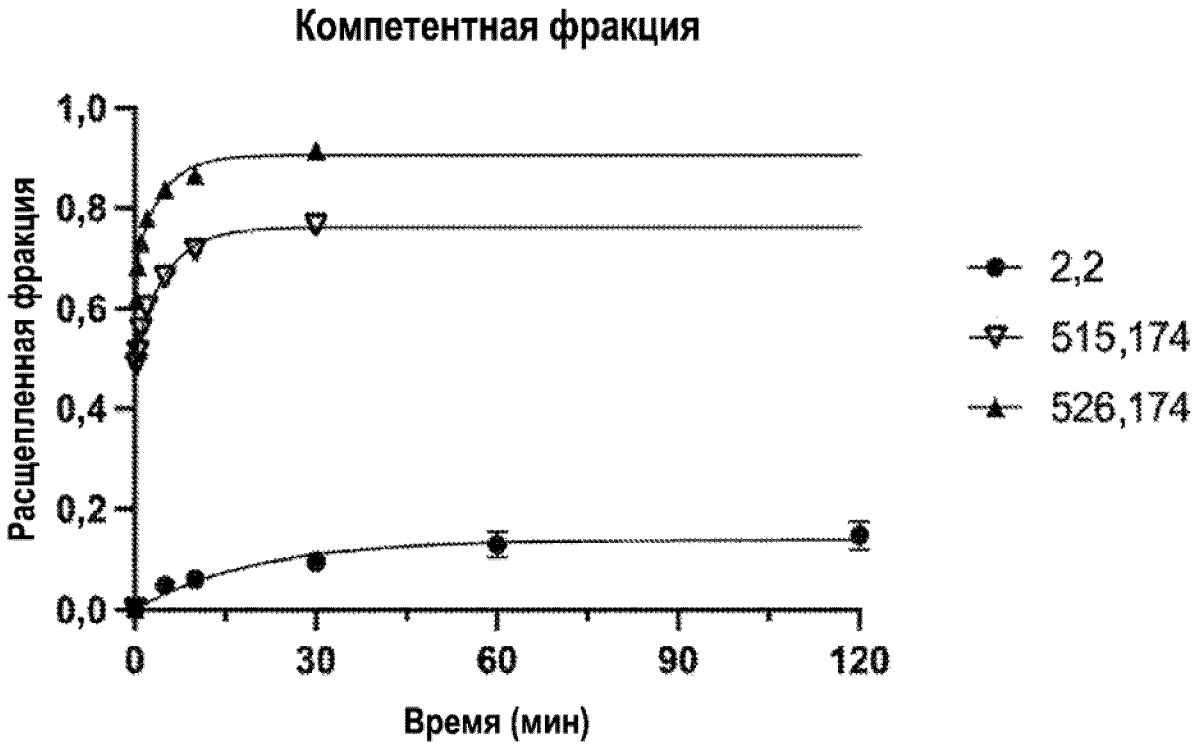
ФИГ. 5


Начальные скорости вариантов огРНК
Расщепленная фракция
0,3
0,2
0,1
0,0
0,00
0,25
0,50
Время (мин)
1
2
32
64
174

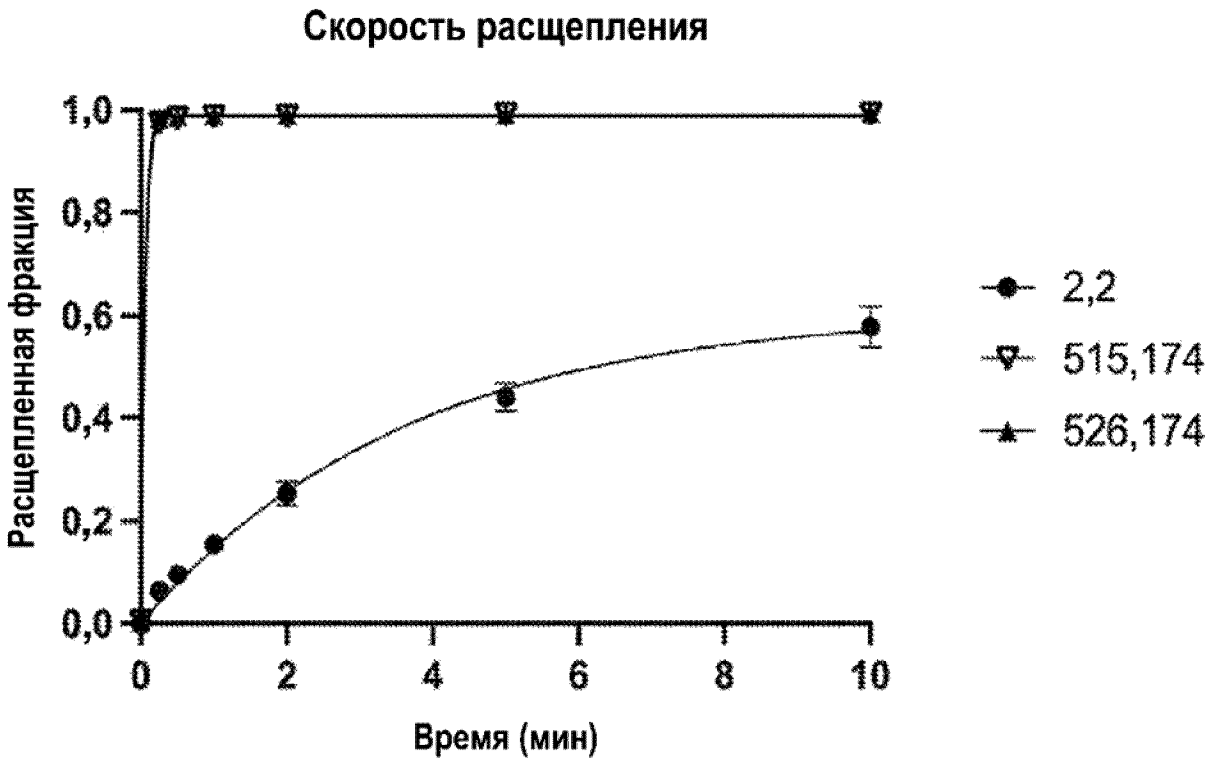
ФИГ. 6


Скорости расщепления вариантов огРНК при 10 °C
Расщепленная фракция
1,0
0,8
0,6
0,4
0,2
0,0
0
5
10
Время (мин)
174
185
186
196
214
215

ФИГ. 7


Компетентная фракция
Расщепленная фракция
1,0
0,8
0,6
0,4
0,2
0,0
0
30
60
90
120
Время (мин)
2,2
515,174
526,174

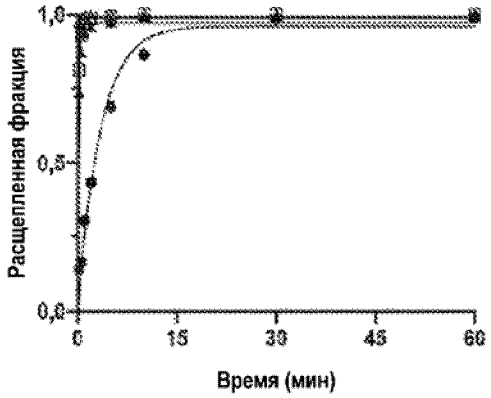
ФИГ. 8


Скорость расщепления
Расщепленная фракция
1,0
0,8
0,6
0,4
0,2
0,0
0
2
4
6
8
10
Время (мин)
2,2
515,174
526,174

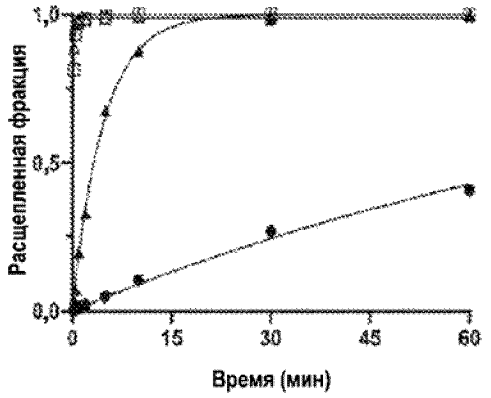
ФИГ. 9

Скорость расщепления на PAM NTC

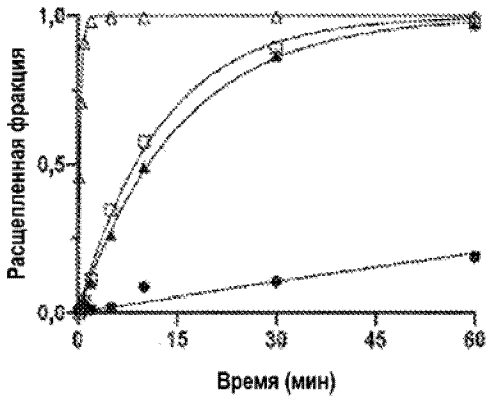
ФИГ. 10A



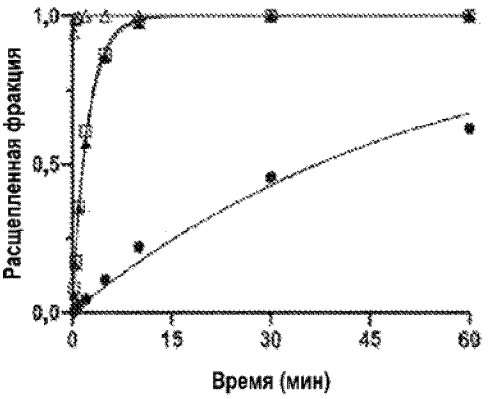
ФИГ. 10B



ФИГ. 10C

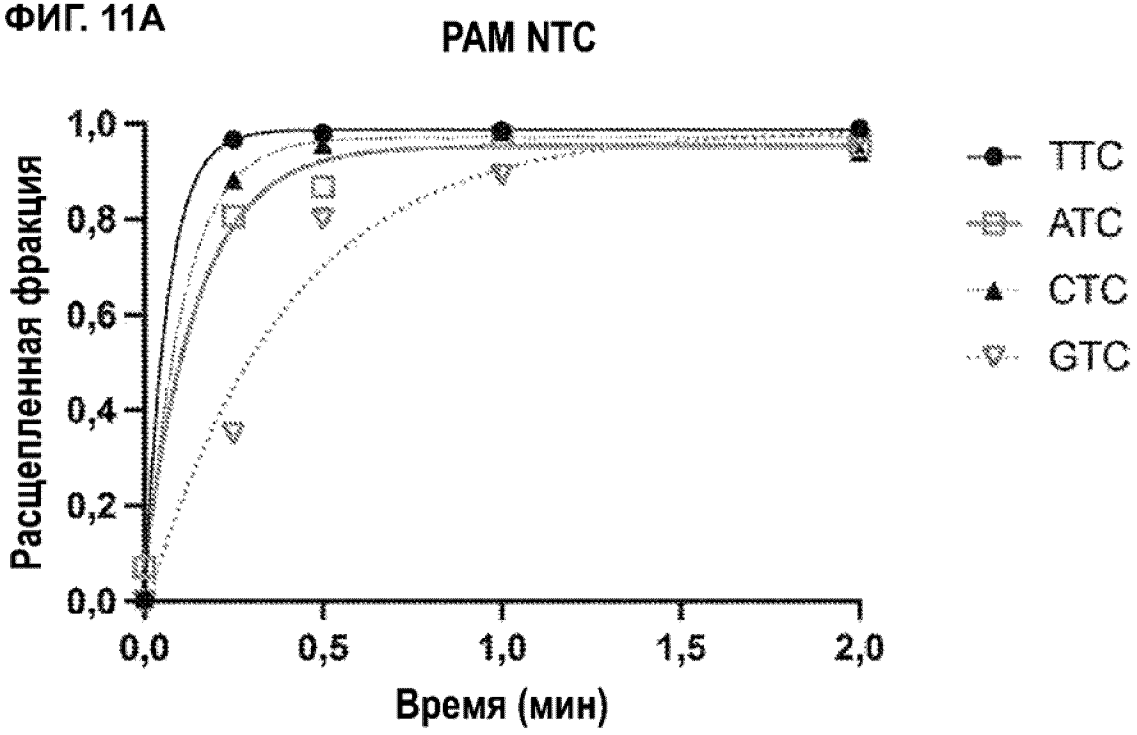


ФИГ. 10D

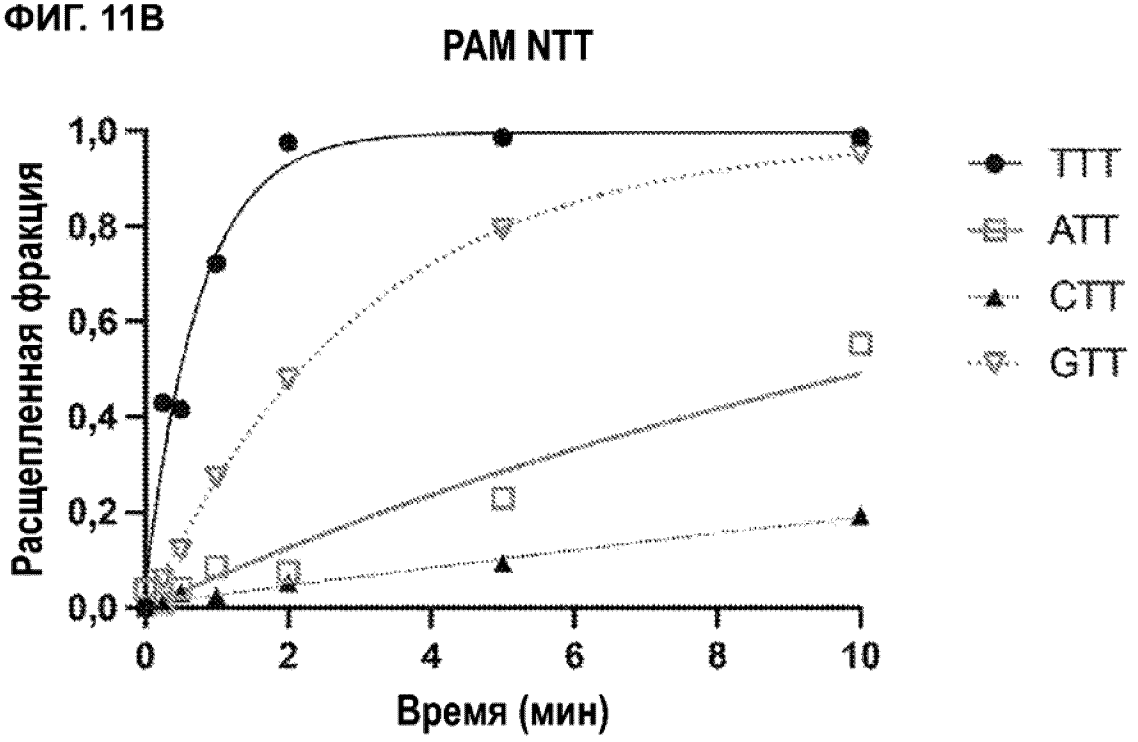




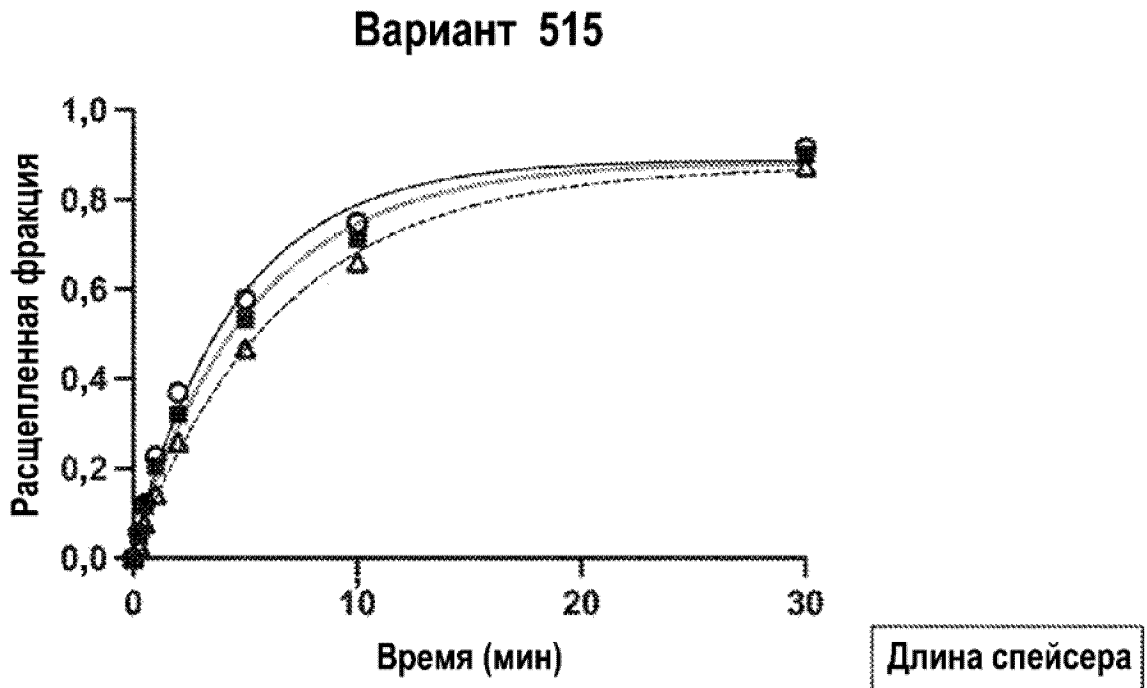
ФИГ. 11A



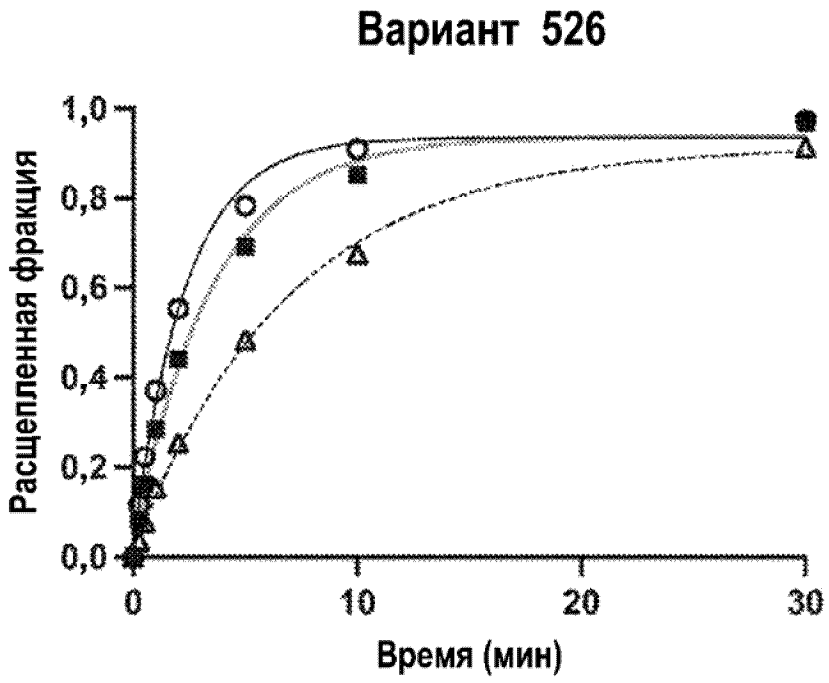
ФИГ. 11B

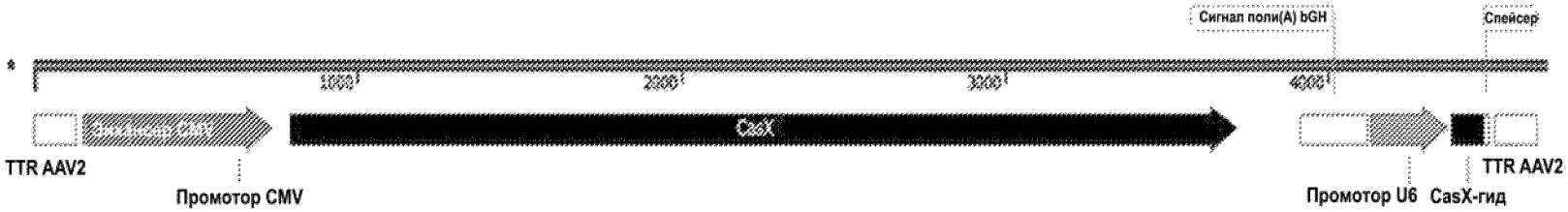


ФИГ. 12A

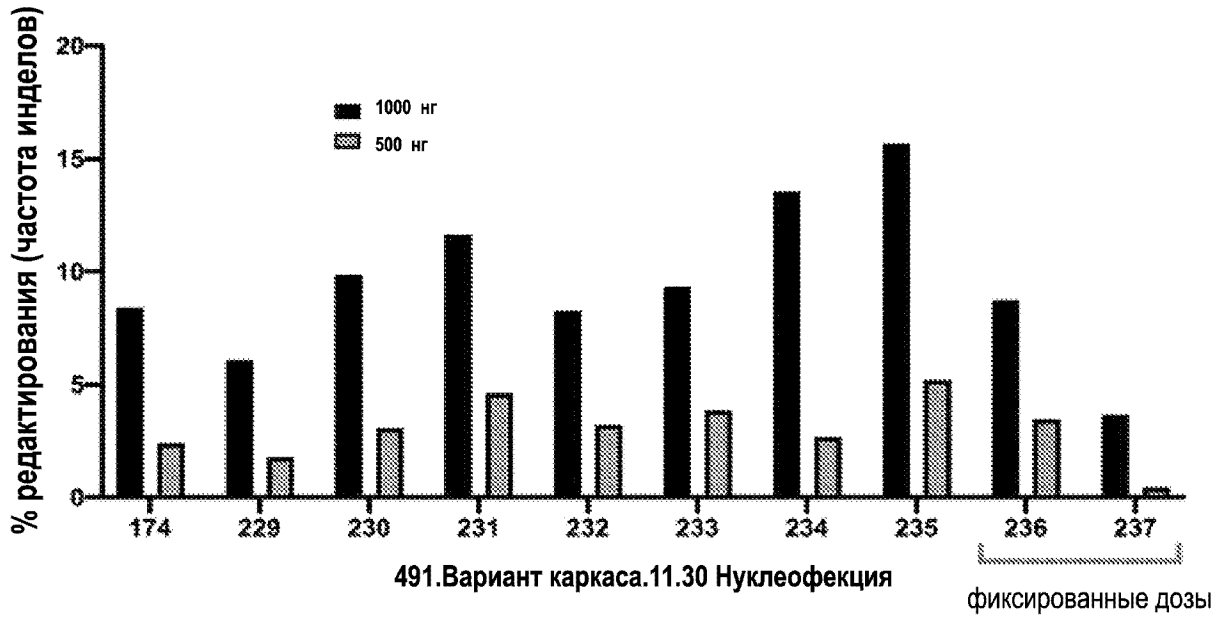


ФИГ. 12B

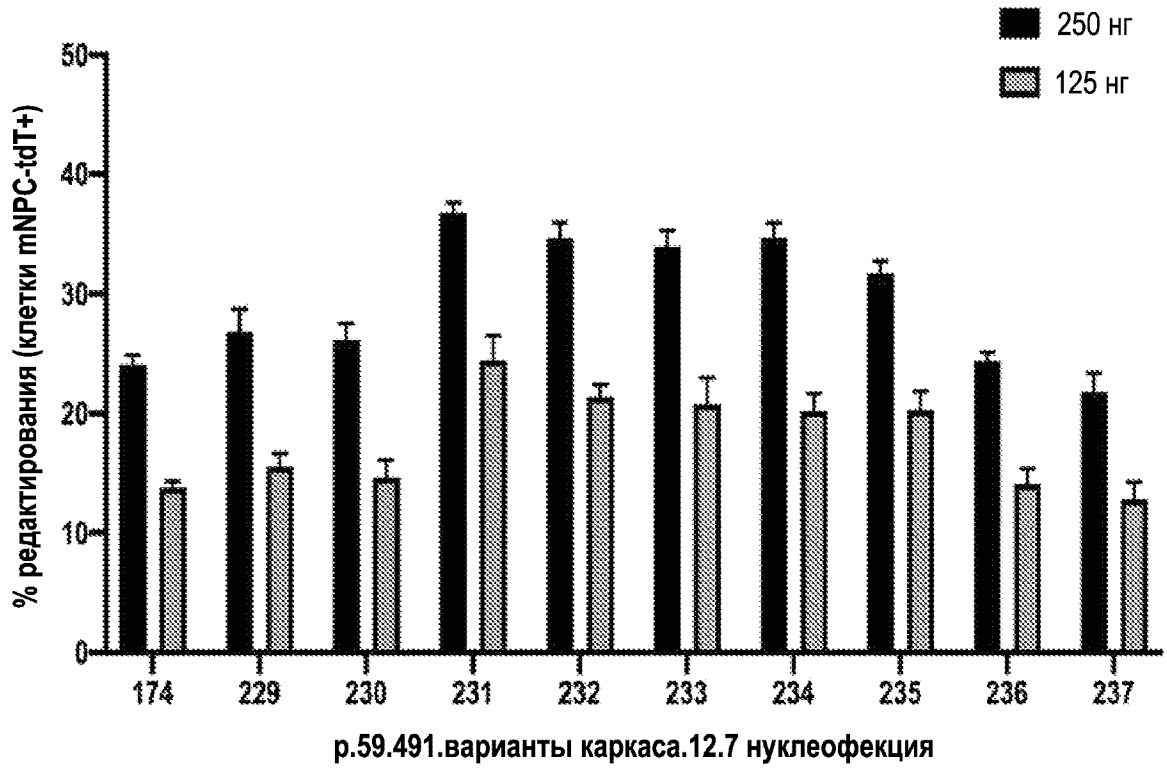



Сигнал поли(A) bGH
Спейсер
1000
2000
3000
4000
Энхансер CMV
CasX
TTR AAV2
Промотор CMV
Промотор U6
CasX-гид
TTR AAV2

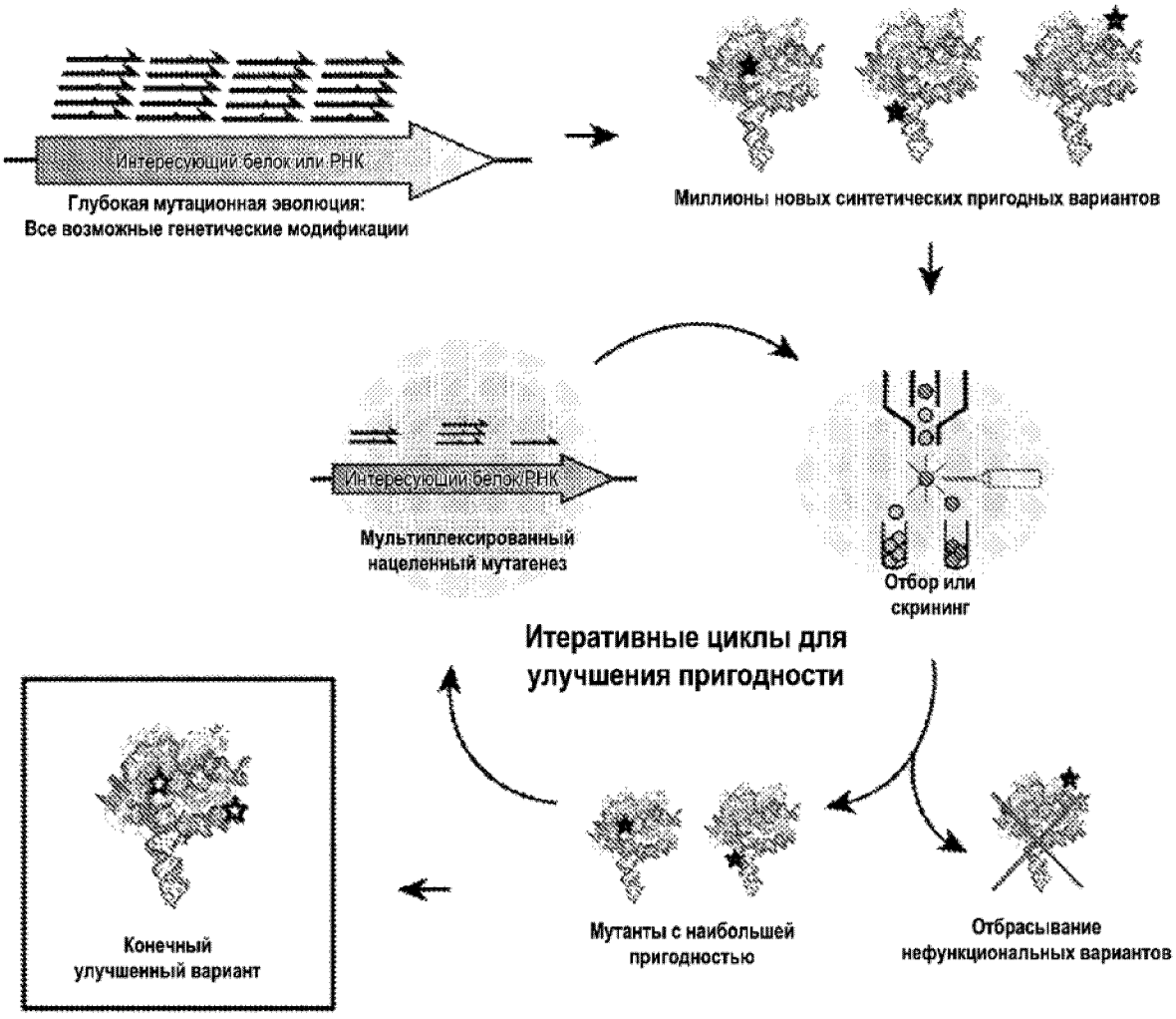
ФИГ. 13


% редактирования (частота инделов)
20
15
10
5
0
1000 нг
500 нг
174
229
230
231
232
233
234
235
236
237
491.Вариант каркаса.11.30 Нуклеофекция
фиксированные дозы

ФИГ. 14


250 нг
125 нг
% редактирования (клетки mNPC-tdT+)
0
10
20
30
40
50
174
229
230
231
232
233
234
235
236
237
p.59.491.варианты каркаса.12.7 нуклеофекция

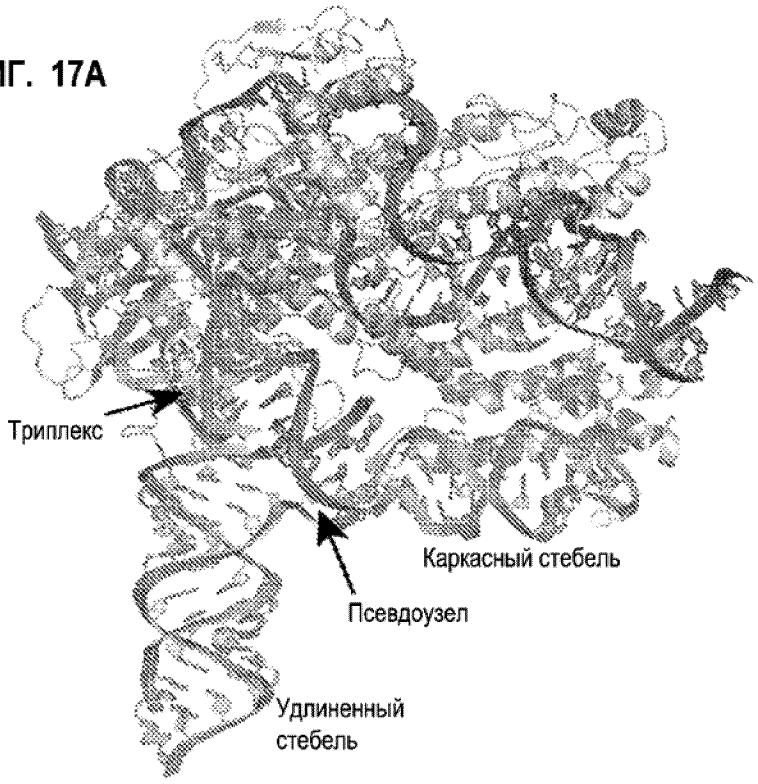
ФИГ. 15



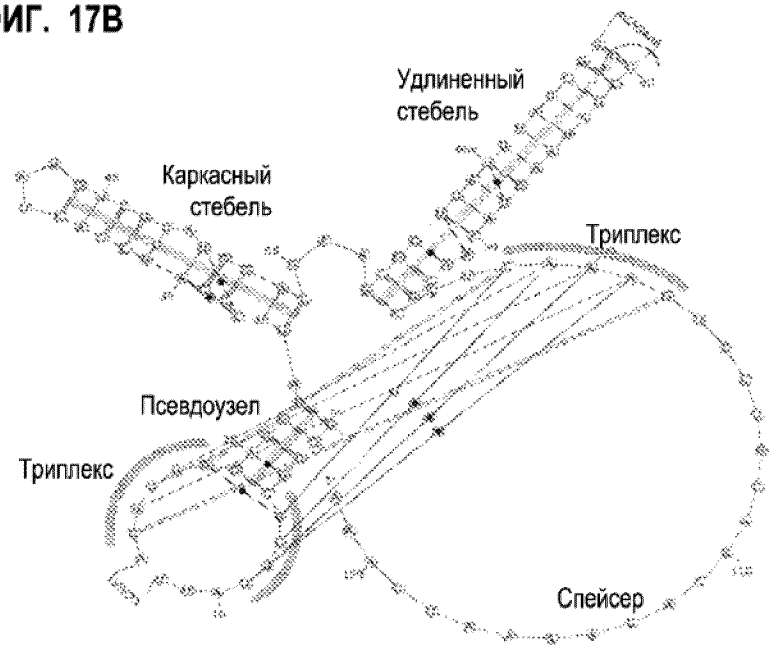
Способы получения вариантов белка CasX и вариантов гидовой РНК

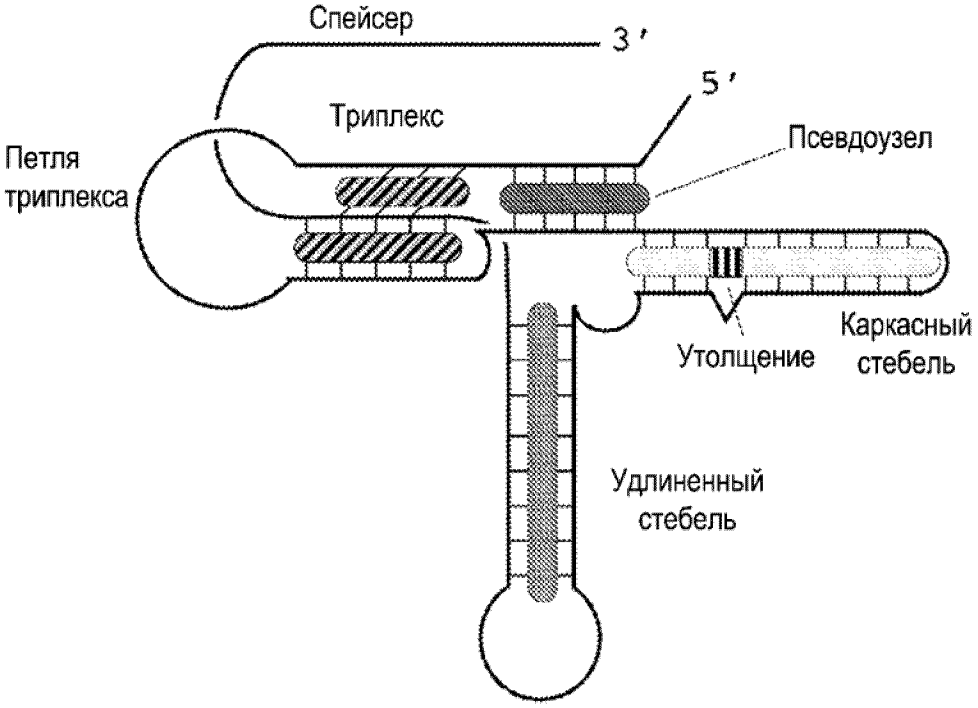
ФИГ. 16

ФИГ. 17A

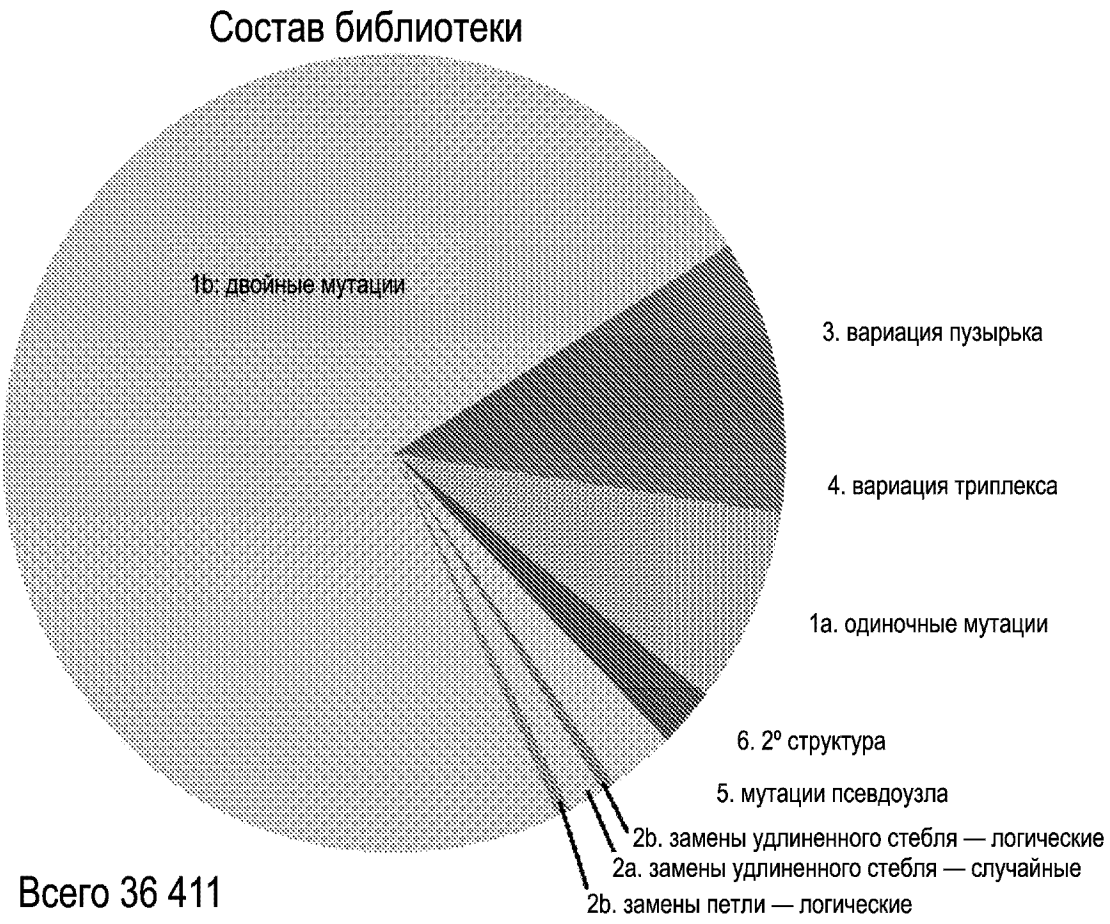


ФИГ. 17B

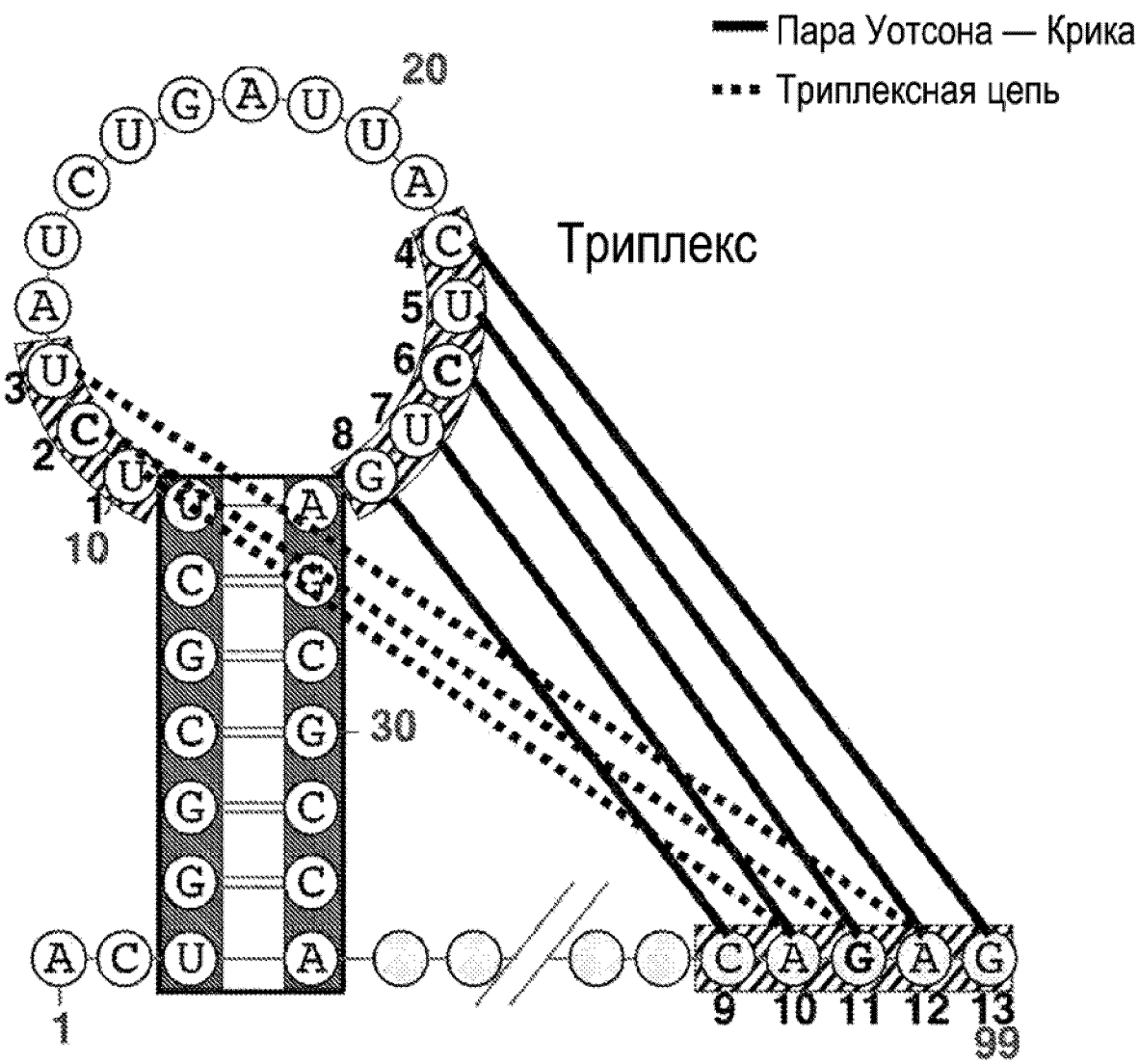



Спейсер
3'
5'
Триплекс
Псевдоузел
Петля
триплекса
Утолщение
Каркасный
стебель
Удлиненный
стебель

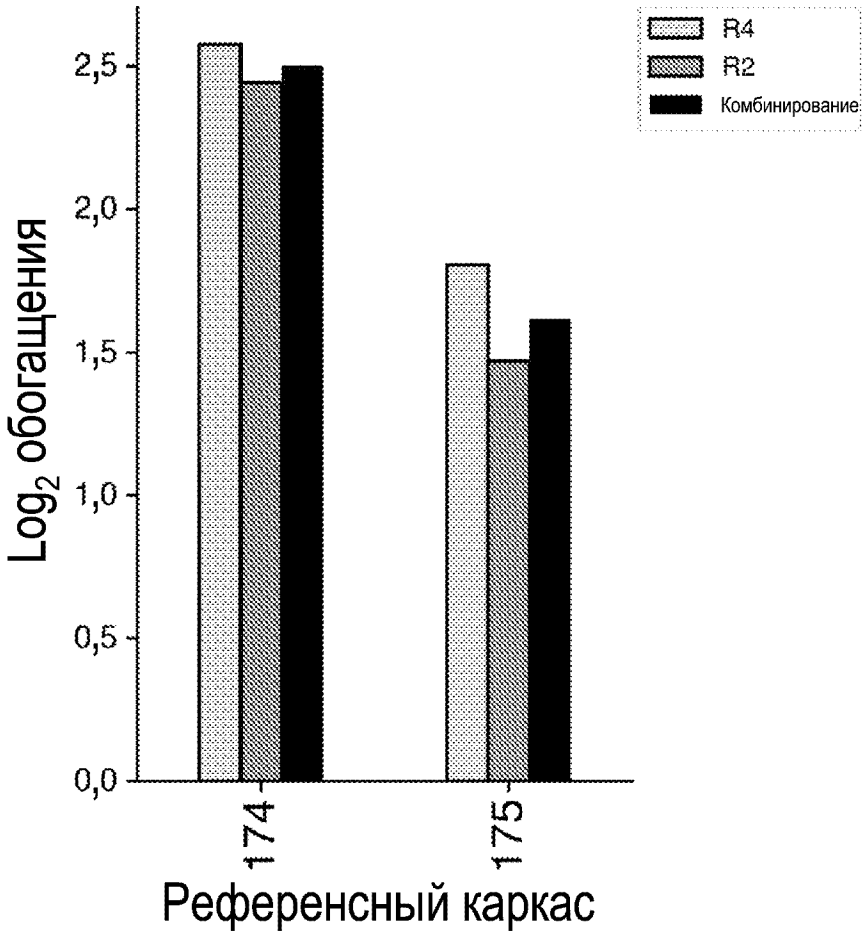
ФИГ. 18


Состав библиотеки
1b. двойные мутации
3. вариация пузырька
4. вариация триплекса
1a. одиночные мутации
6. 2º структура
5. мутации псевдоузла
2b. замены удлиненного стебля — логические
2a. замены удлиненного стебля — случайные
2b. замены петли — логические
Всего 36 411

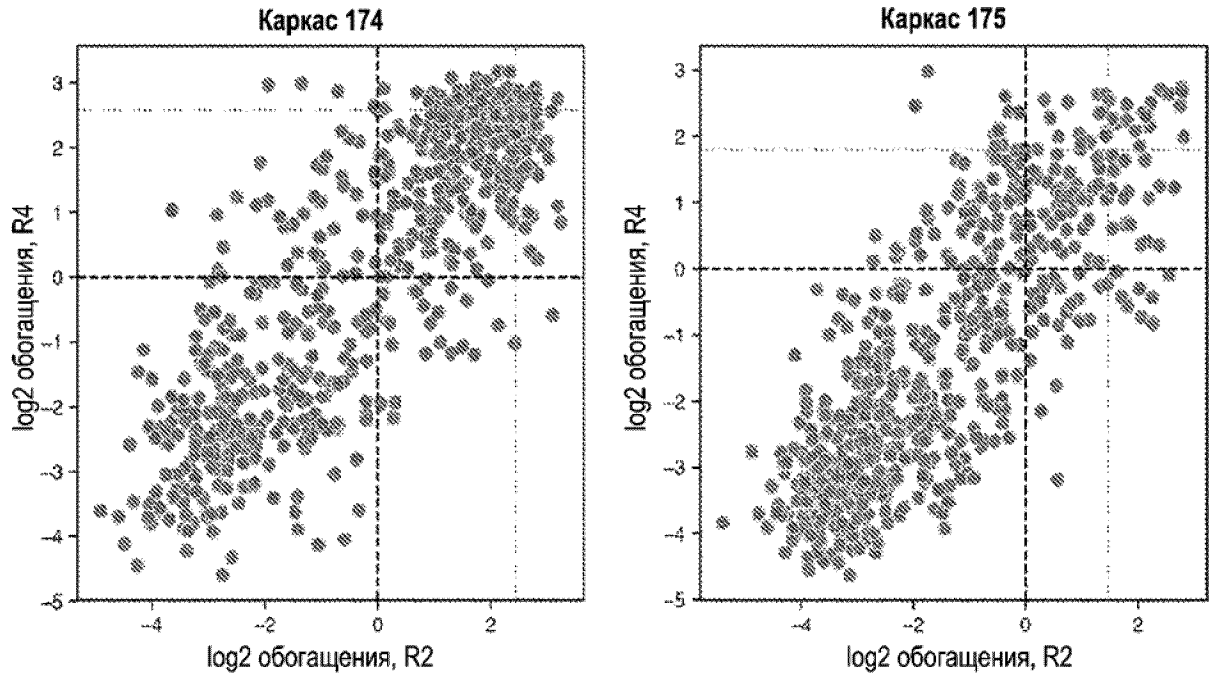
ФИГ. 19


Пара Уотсона — Крика
Триплексная цепь
Триплекс
20
4
5
6
7
8
3
2
1
10
30
A C U A
1
C A G A G
9 10 11 12 13
99

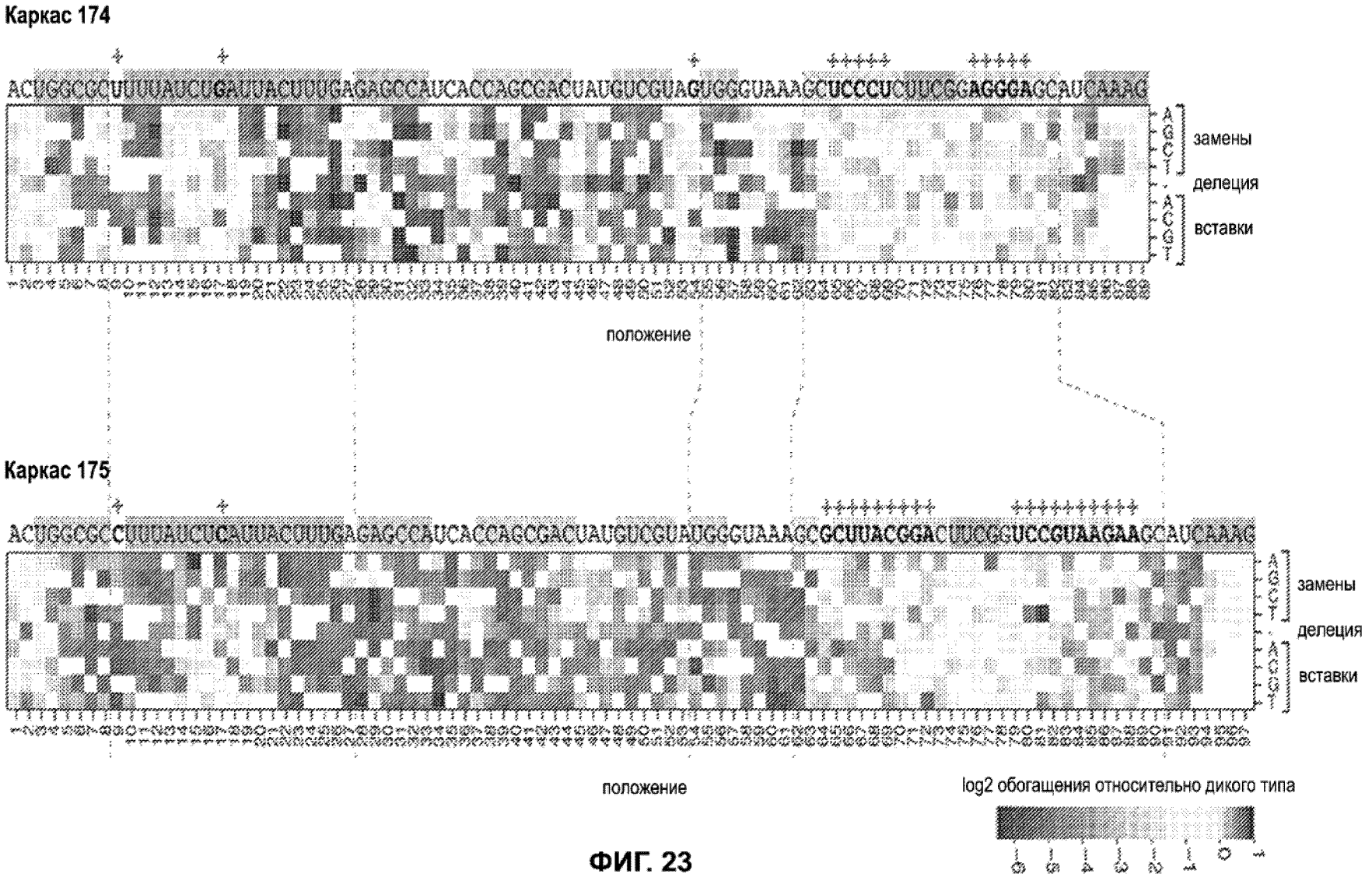
ФИГ. 20


R4
R2
Комбинирование
2,5
2,0
1,5
1,0
0,5
0,0
Log2 обогащения
174
175
Референсный каркас

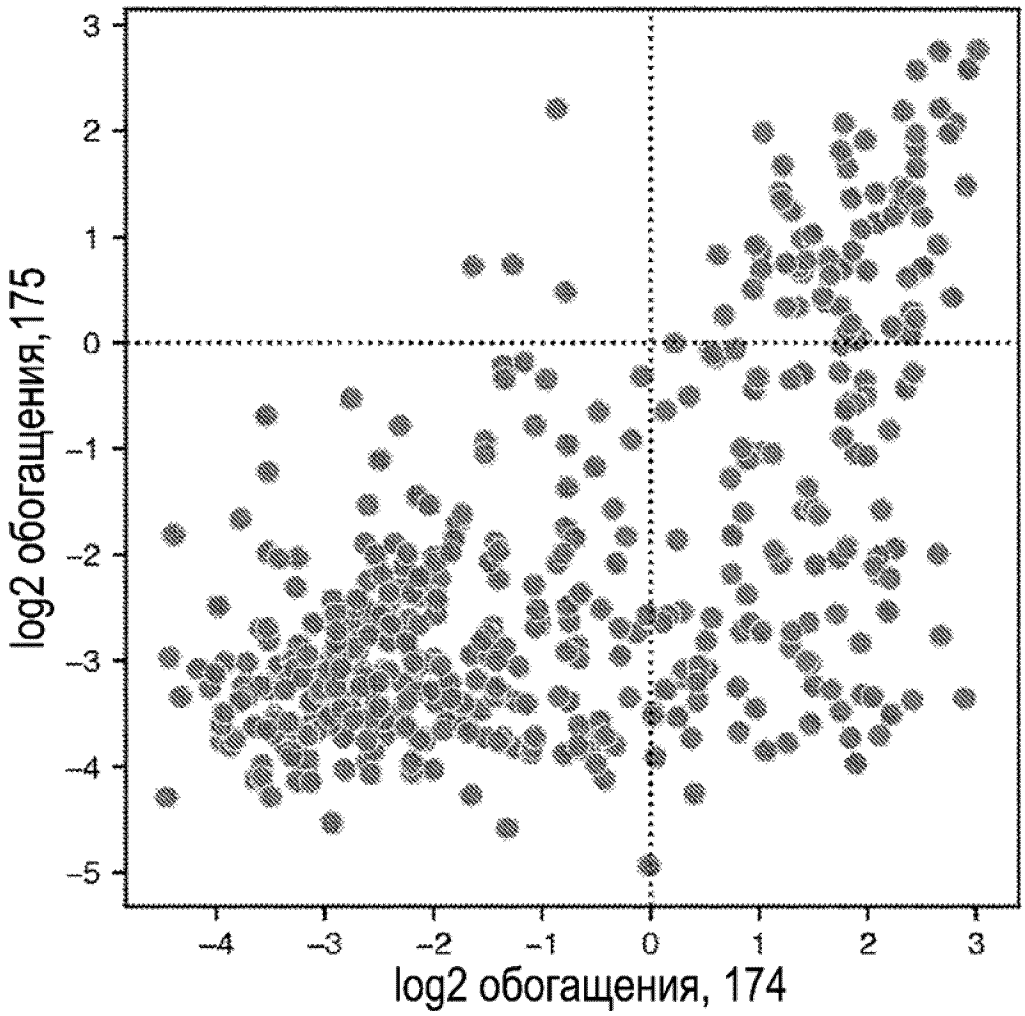
ФИГ. 21


Каркас 174
log2 обогащения, R4
log2 обогащения, R2
3
2
1
0
-1
-2
-3
-4
-5
-4
-2
0
2
Каркас 175
log2 обогащения, R4
log2 обогащения, R2
3
2
1
0
-1
-2
-3
-4
-5
-4
-2
0
2

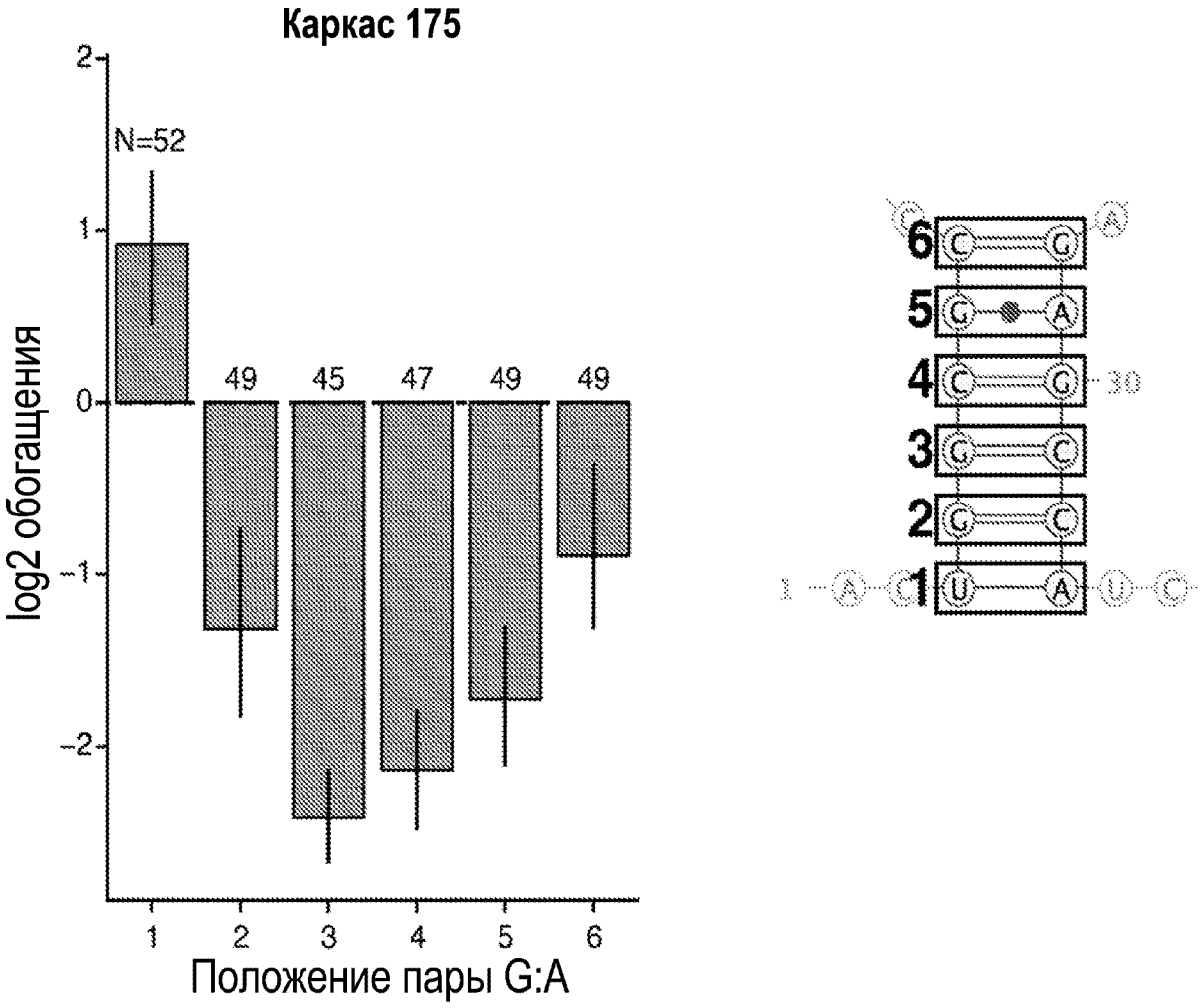
ФИГ. 22


Каркас 174
ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCAUCAAAG
замены
делеция
вставки
положение
Каркас 175
ACUGGCGCCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCCGUAAGAAGCAUCAAAG
замены
делеция
вставки
положение
log2 обогащения относительно дикого типа
6
5
4
3
2
1
0
-1

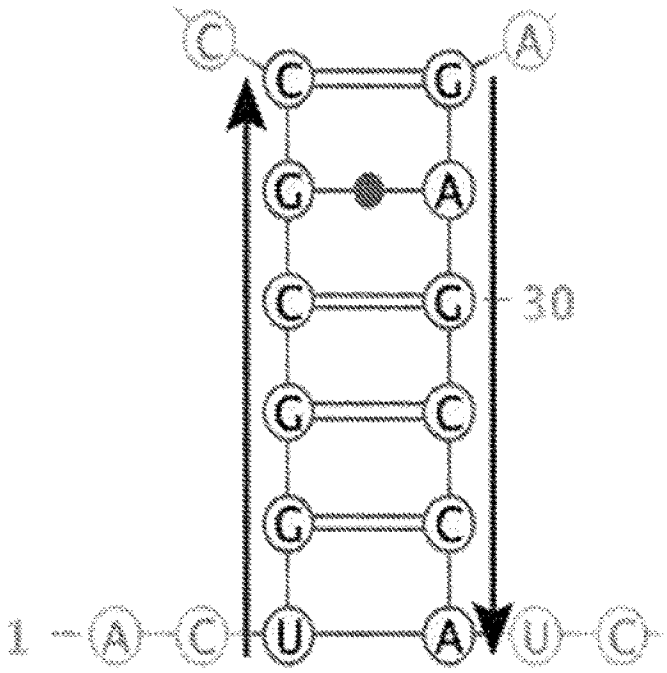
ФИГ. 23


log2 обогащения,175
log2 обогащения, 174
3
2
1
0
−1
−2
−3
−4
−5
−4
−3
−2
−1
0
1
2
3

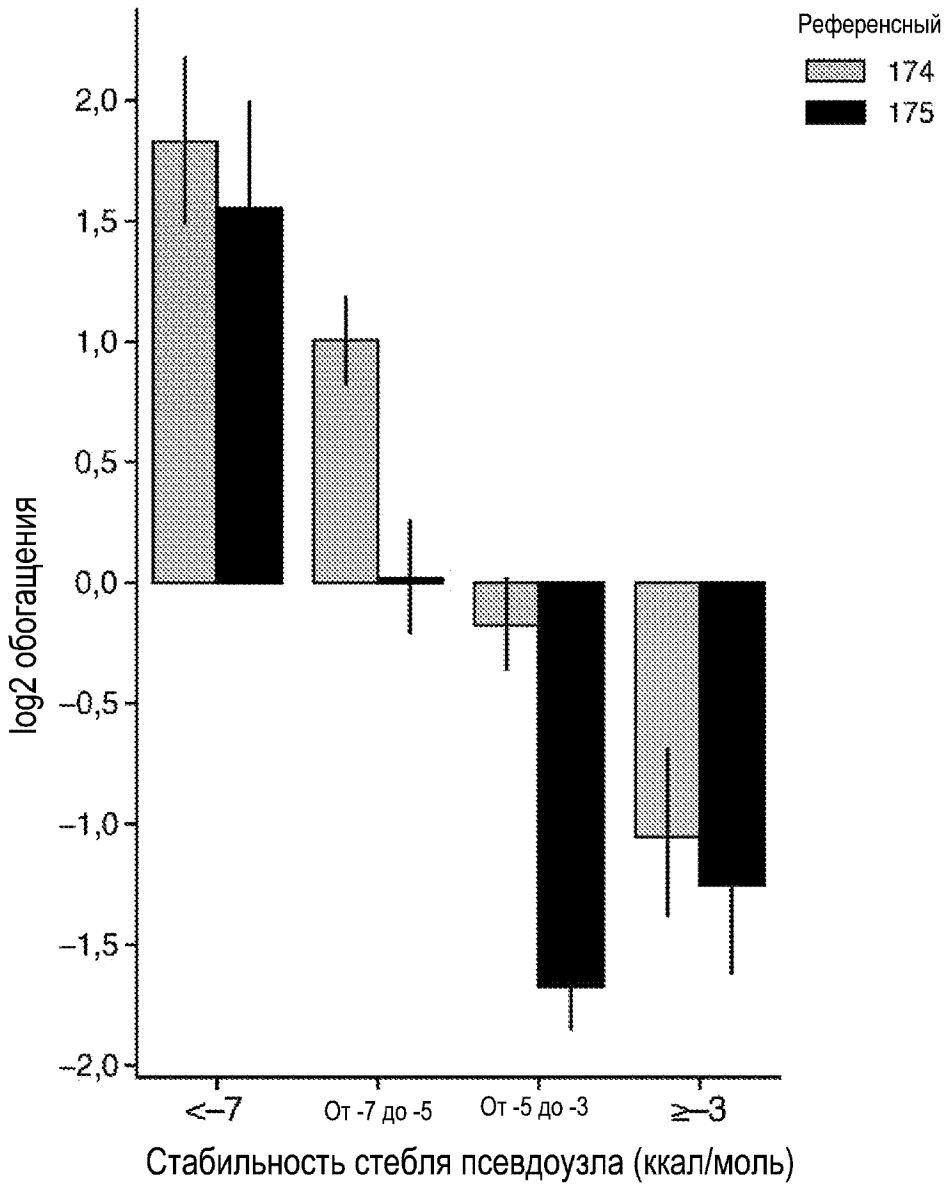
ФИГ. 24


Каркас 175
log2 обогащения
N=52
49
45
47
49
49
Положение пары G:A
6
5
4
3
2
1
30

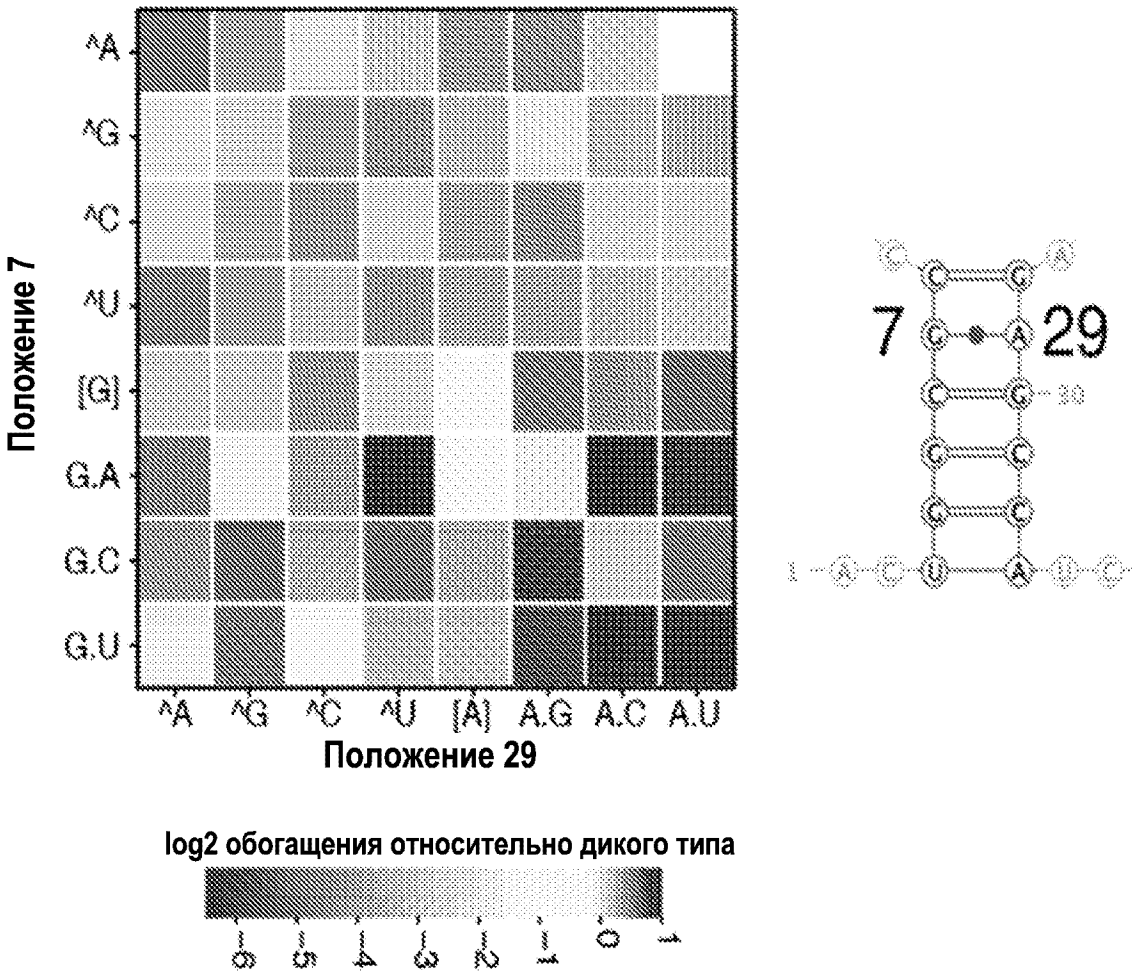
ФИГ. 25



ФИГ. 26

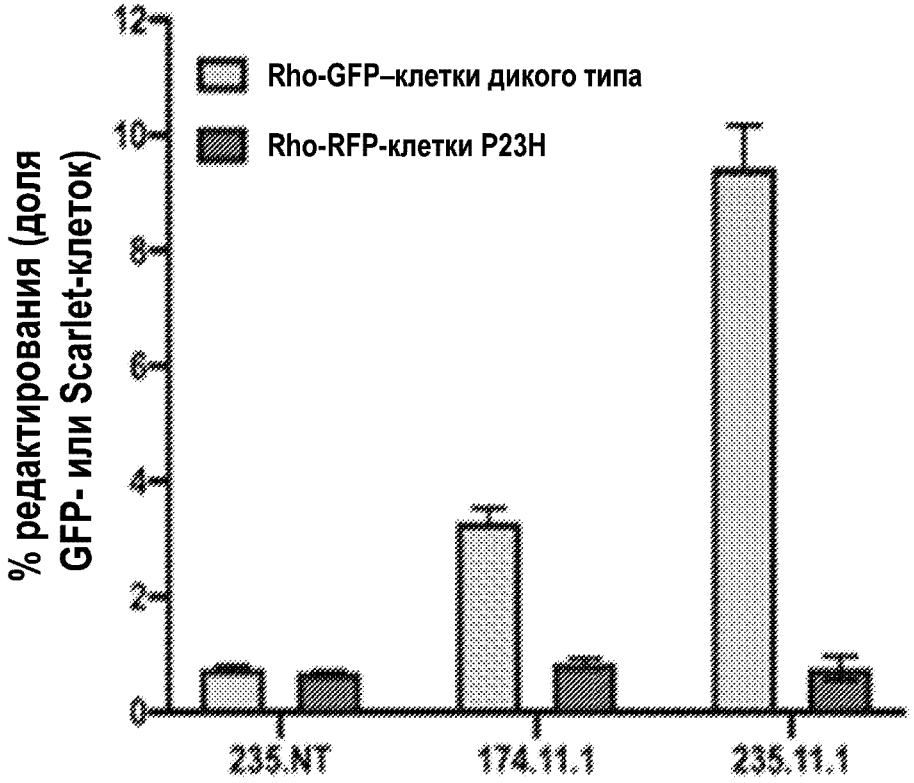

Референсный
174
175
2,0
1,5
1,0
0,5
0,0
−0,5
−1,0
−1,5
−2,0
log2 обогащения
<−7
От -7 до -5
От -5 до -3
≥−3
Стабильность стебля псевдоузла (ккал/моль)

ФИГ. 27

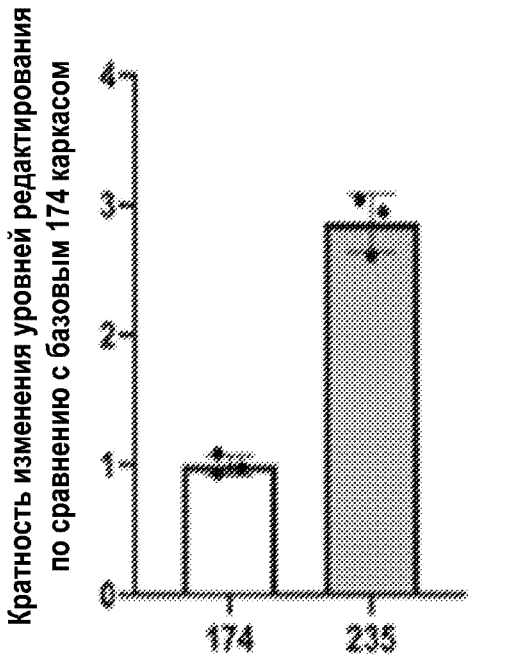

Положение 7
^A
^G
^C
^U
[G]
G.A
G.C
G.U
^A ^G ^C ^U [A] A.G A.C A.U
Положение 29
log2 обогащения относительно дикого типа
7
29

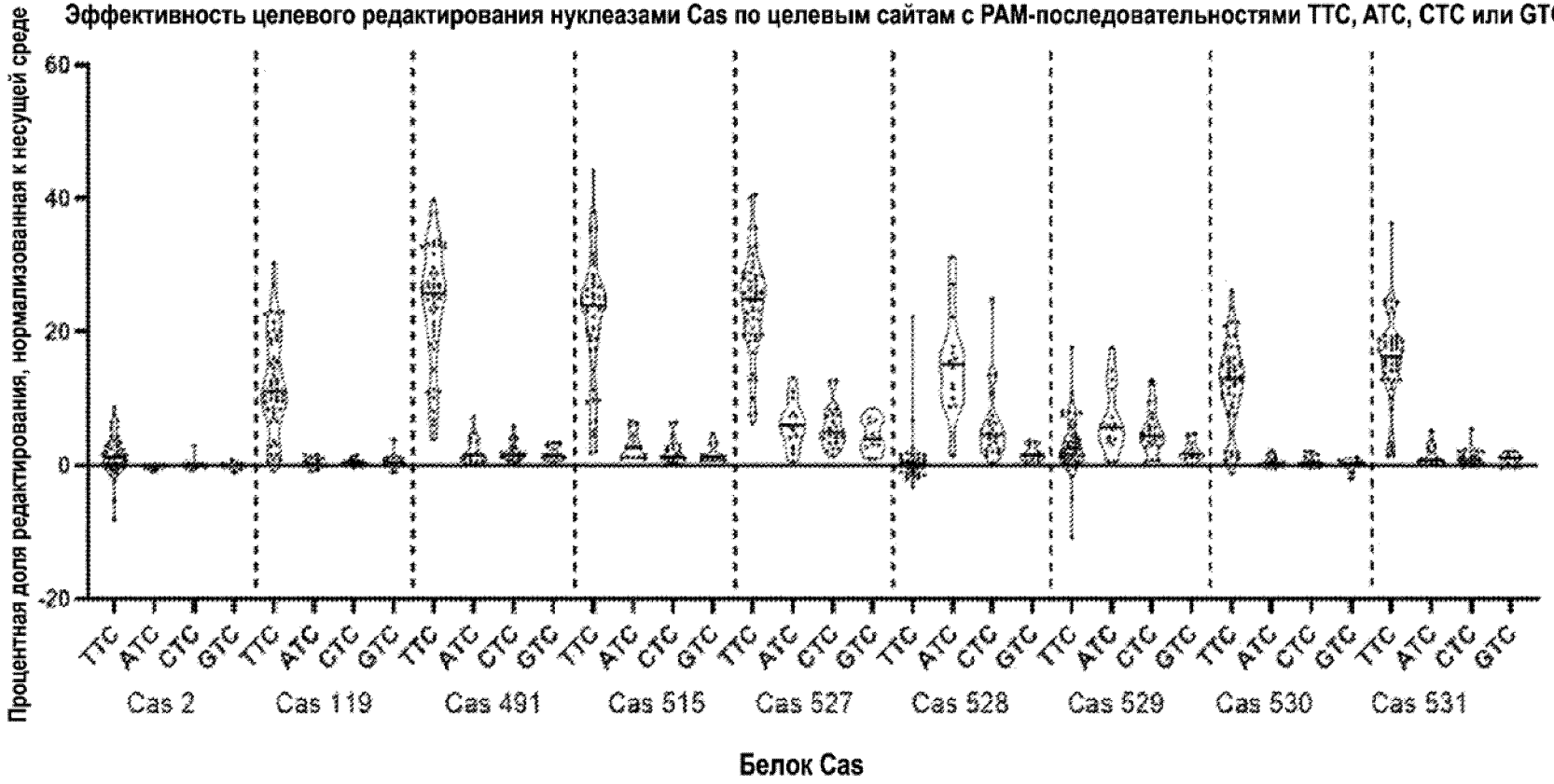
ФИГ. 28

ФИГ. 29A

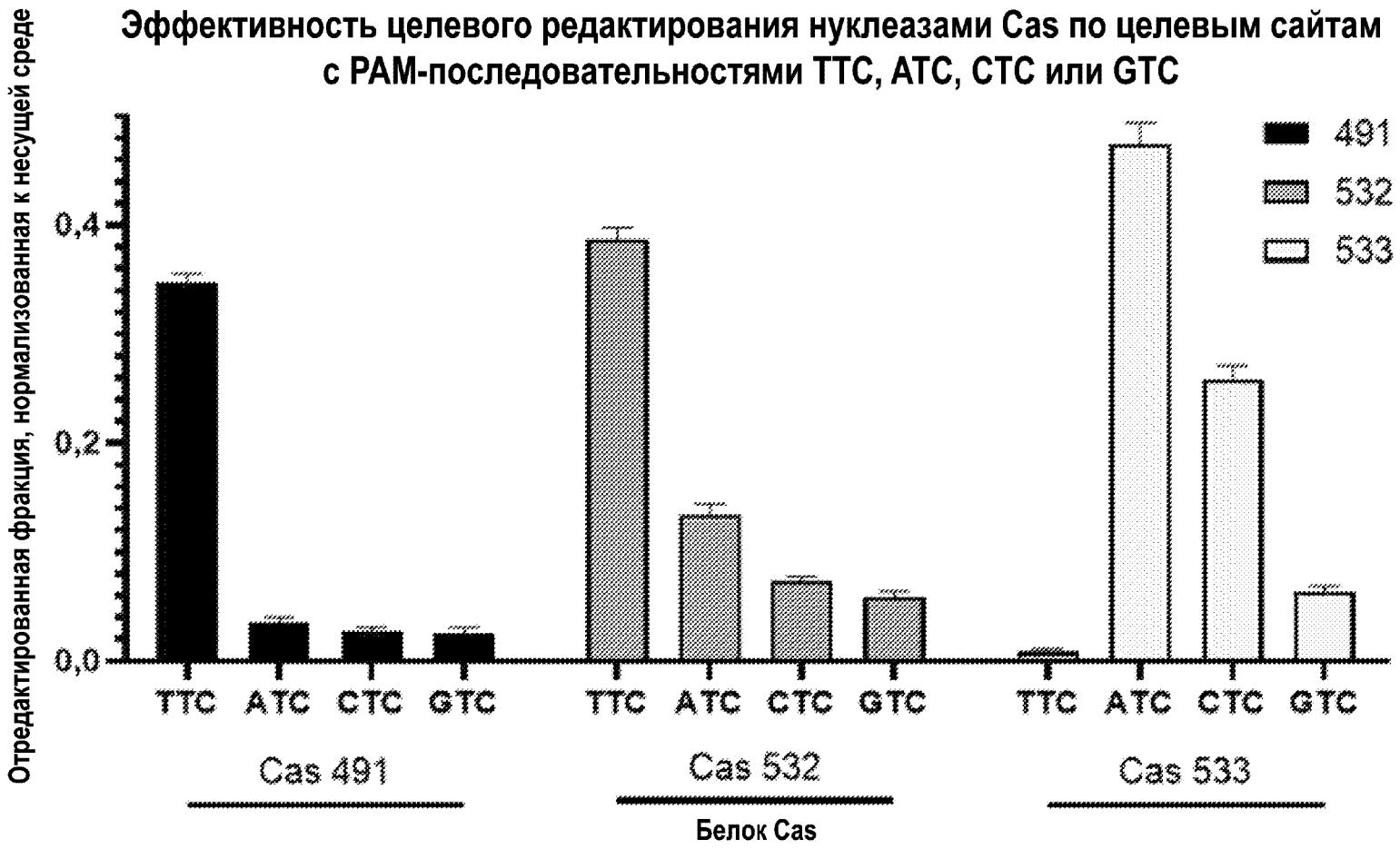


ФИГ. 29B

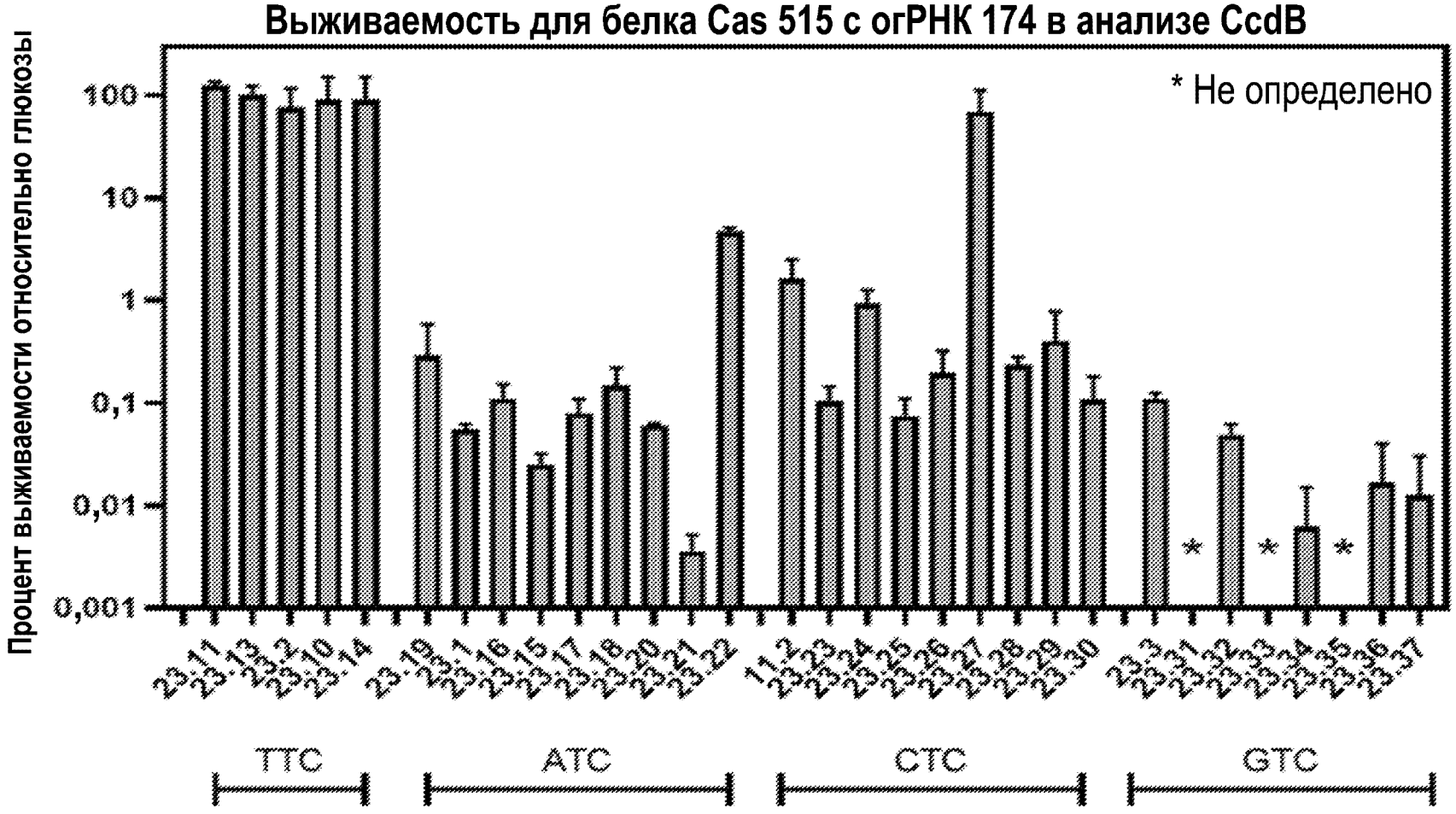



Эффективность целевого редактирования нуклеазами Cas по целевым сайтам с PAM-последовательностями TTC, ATC, CTC или GTC
Процентная доля редактирования, нормализованная к несущей среде
60
40
20
0
-20
TTC
ATC
CTC
GTC
Cas 2
Cas 119
Cas 491
Cas 515
Cas 527
Cas 528
Cas 529
Cas 530
Cas 531
Белок Cas

ФИГ. 30


Эффективность целевого редактирования нуклеазами Cas по целевым сайтам с PAM-последовательностями TTC, ATC, CTC или GTC
Отредактированная фракция, нормализованная к несущей среде
0,0
0,2
0,4
491
532
533
TTC
ATC
CTC
GTC
Cas 491
Cas 532
Cas 533
Белок Cas

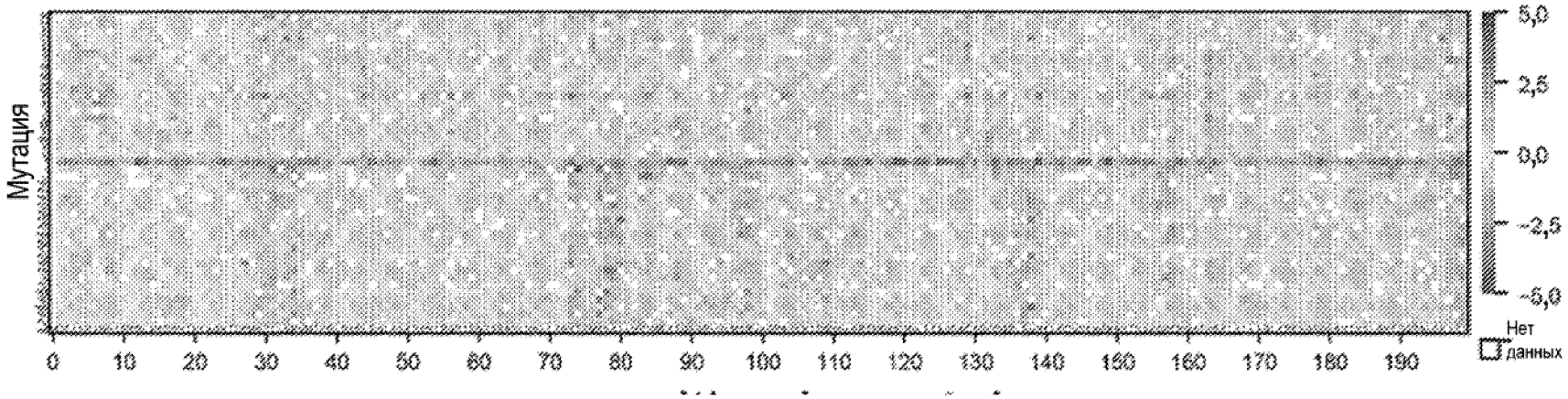
ФИГ. 31



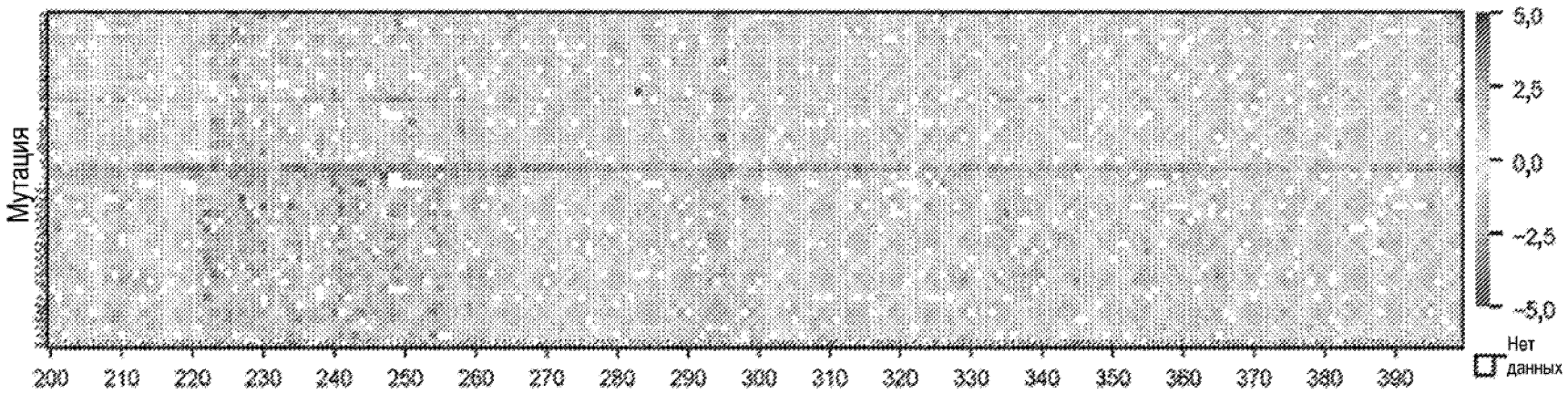
Плазмида p.73.515.174, нацеленная на CcdB, плазмида p58 с указанным спейсером

ФИГ. 32

ФИГ. 33A

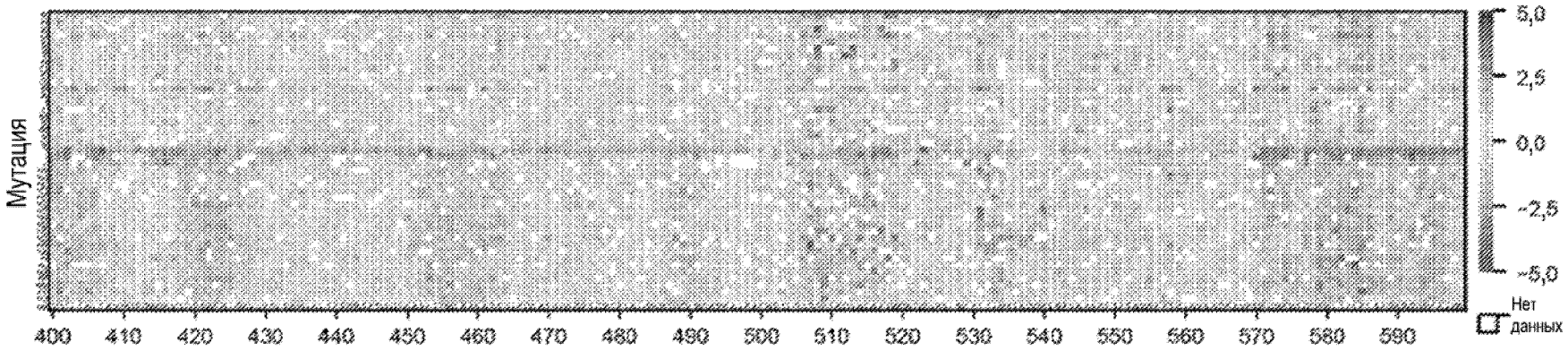


ФИГ. 33B

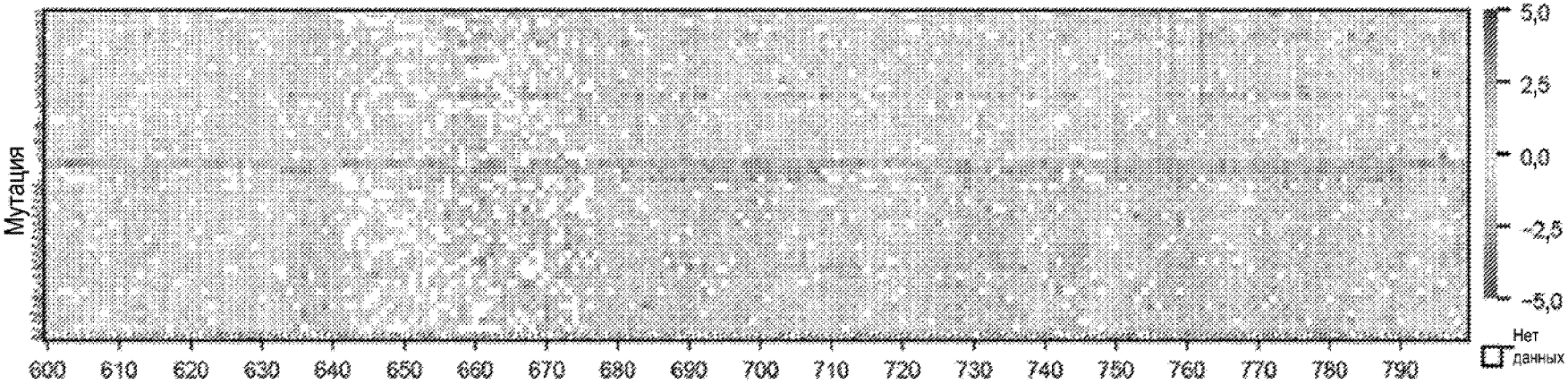


Положение в белке

ФИГ. 33C

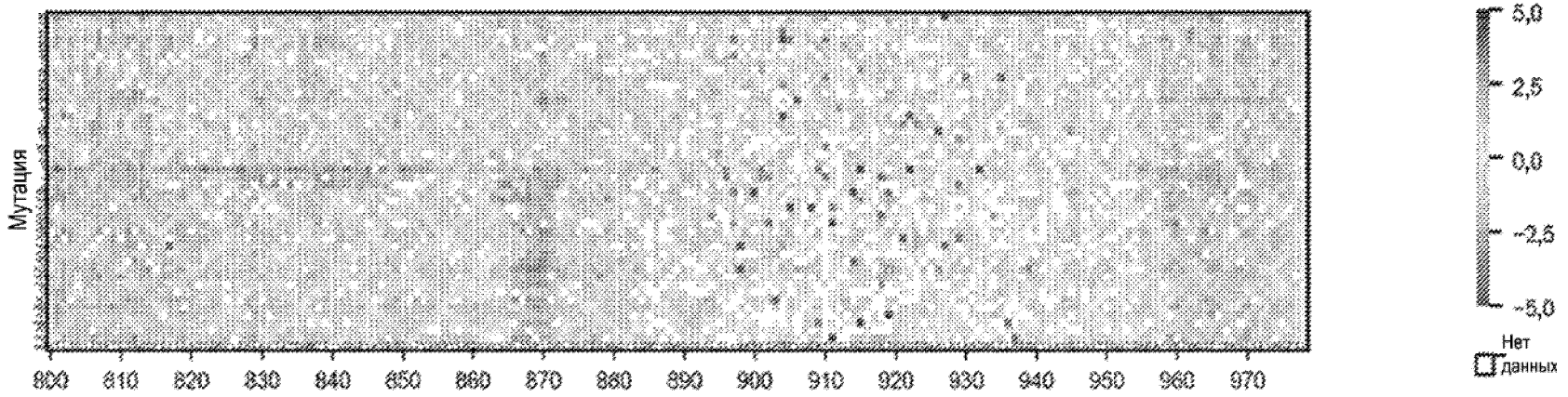


ФИГ. 33D

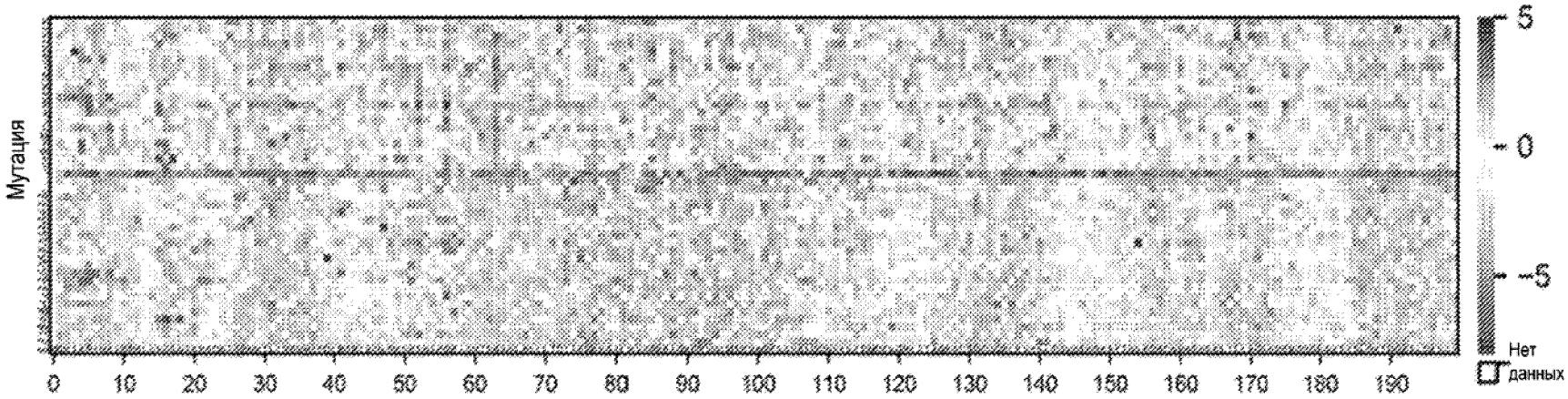


Положение в белке

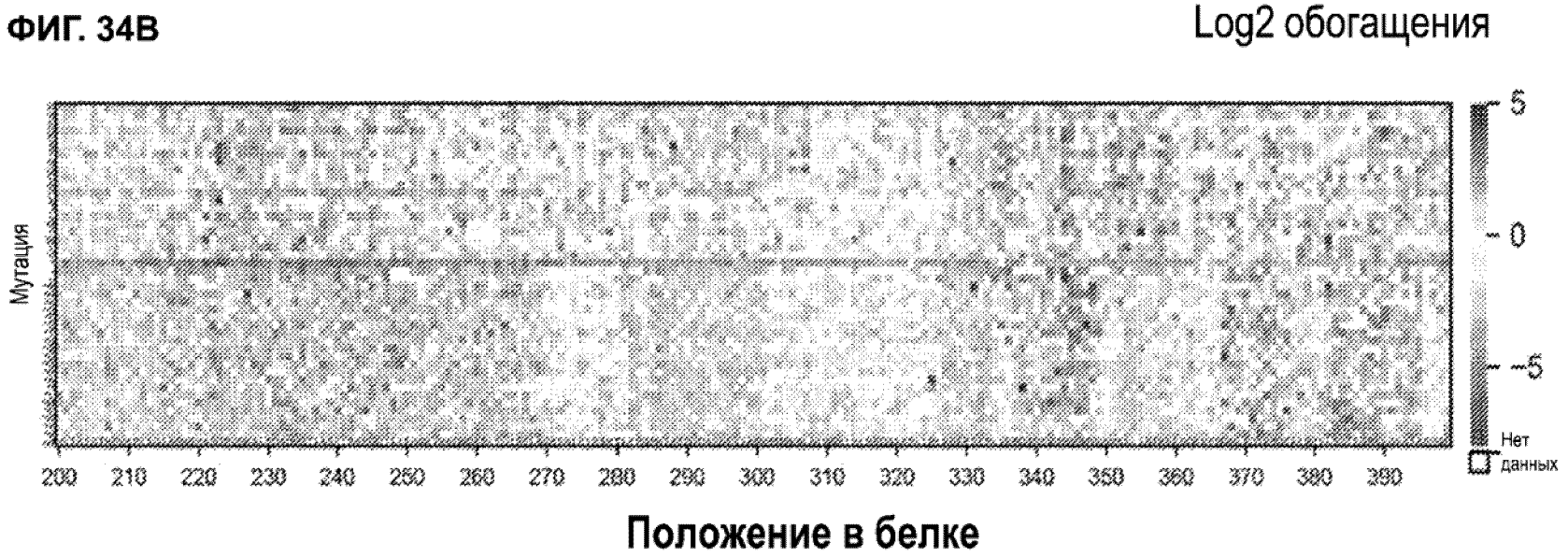
ФИГ. 33E


Log2 обогащения
5,0
2,5
0,0
-2,5
-5,0
Нет данных
Мутация
800
810
820
830
840
850
860
870
880
890
900
910
920
930
940
950
960
970
Положение в белке

ФИГ. 34A

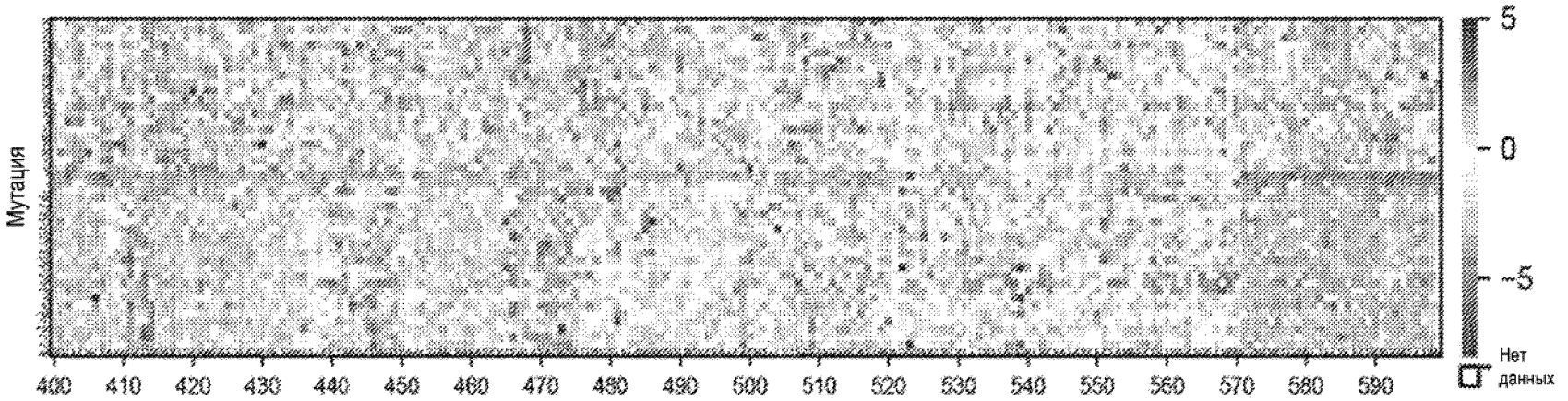


ФИГ. 34B



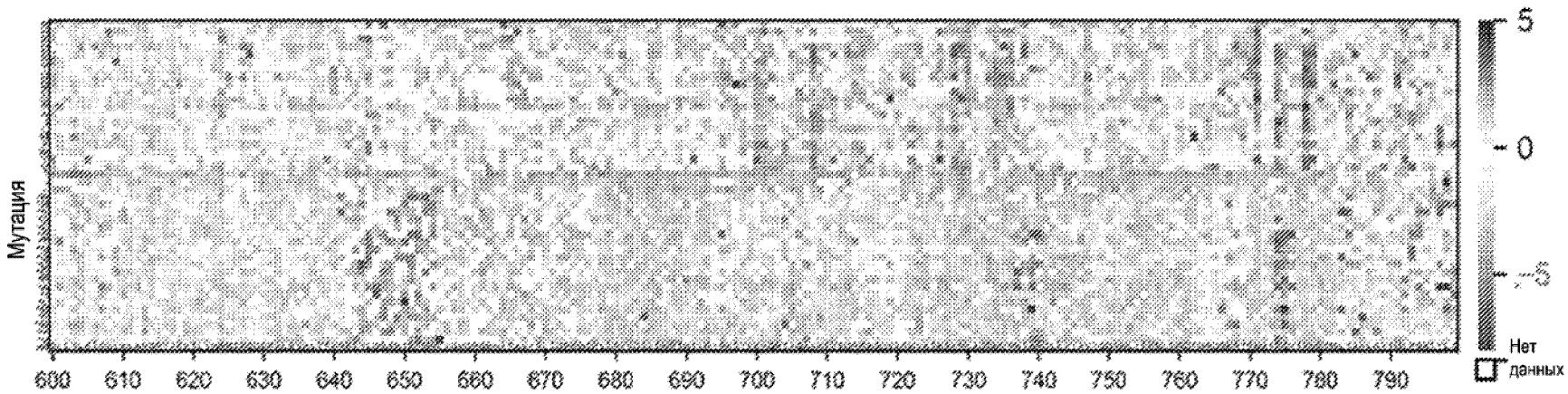
Log2 обогащения

ФИГ. 34C

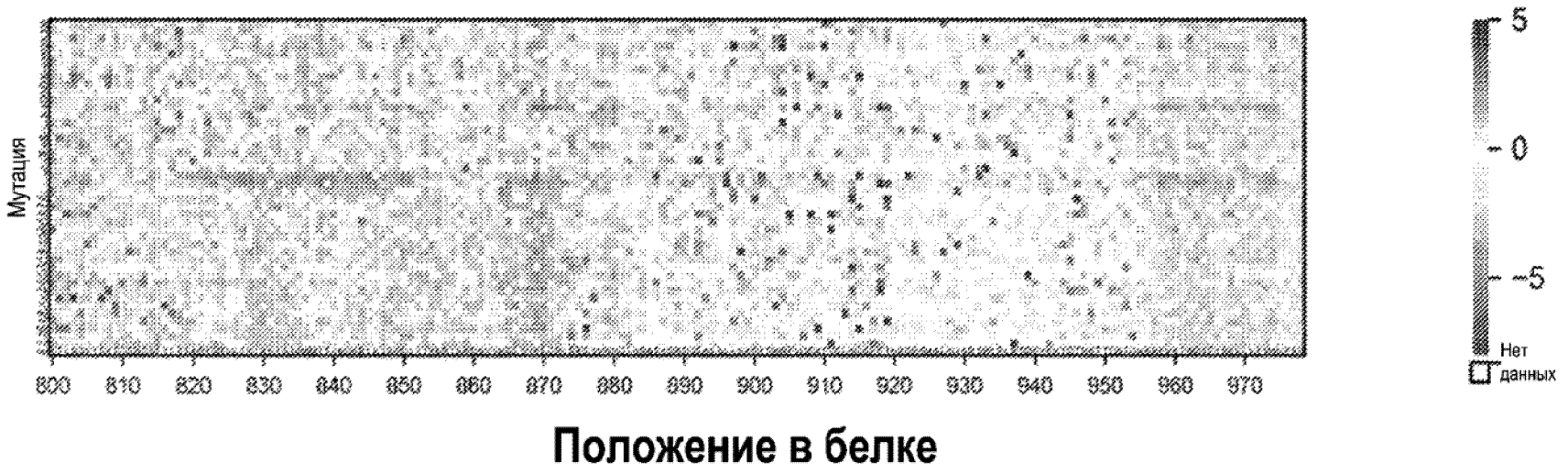


Log2 обогащения

ФИГ. 34D

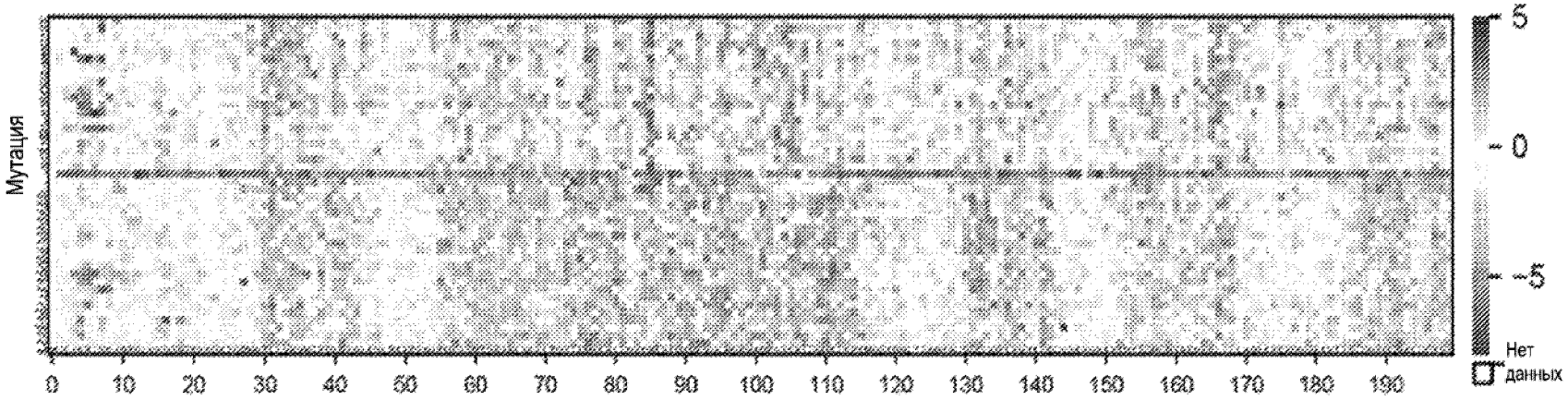


Положение в белке

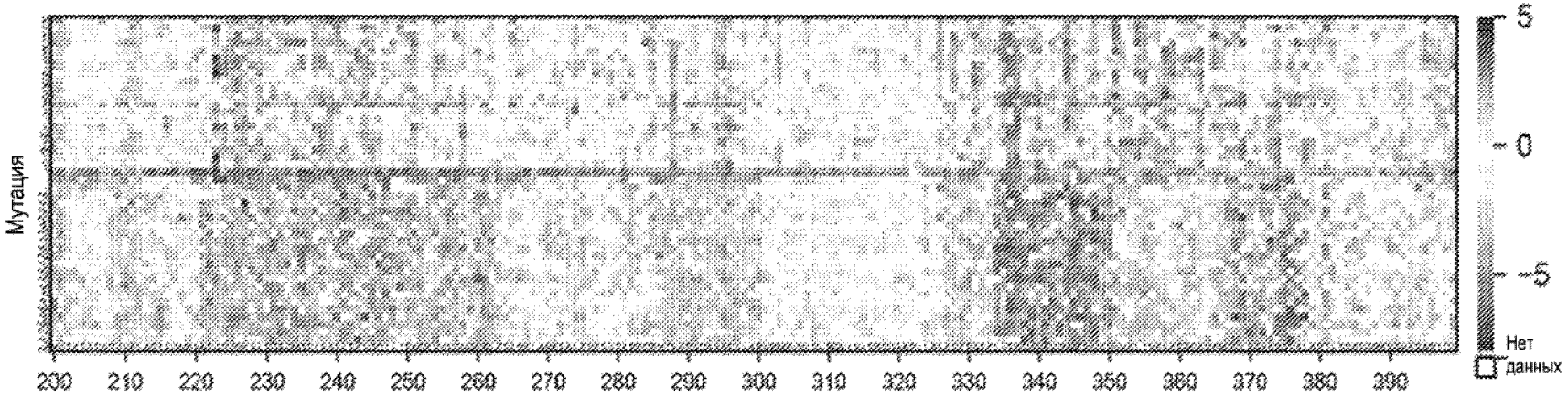

Log2 обогащения
5
0
-5
Нет данных
Мутация
800
810
820
830
840
850
860
870
880
890
900
910
920
930
940
950
960
970
Положение в белке

ФИГ. 34E

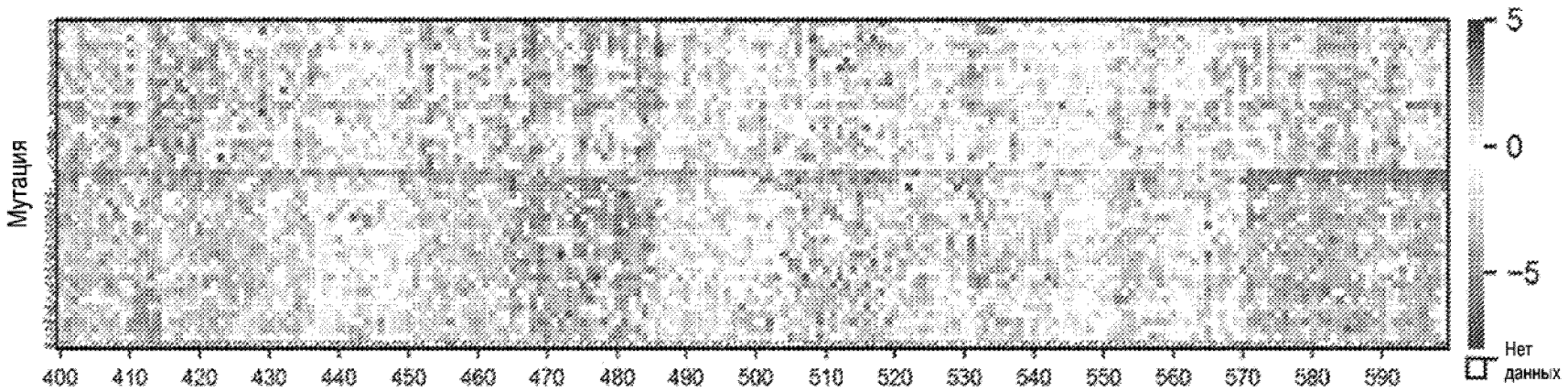
ФИГ. 35A



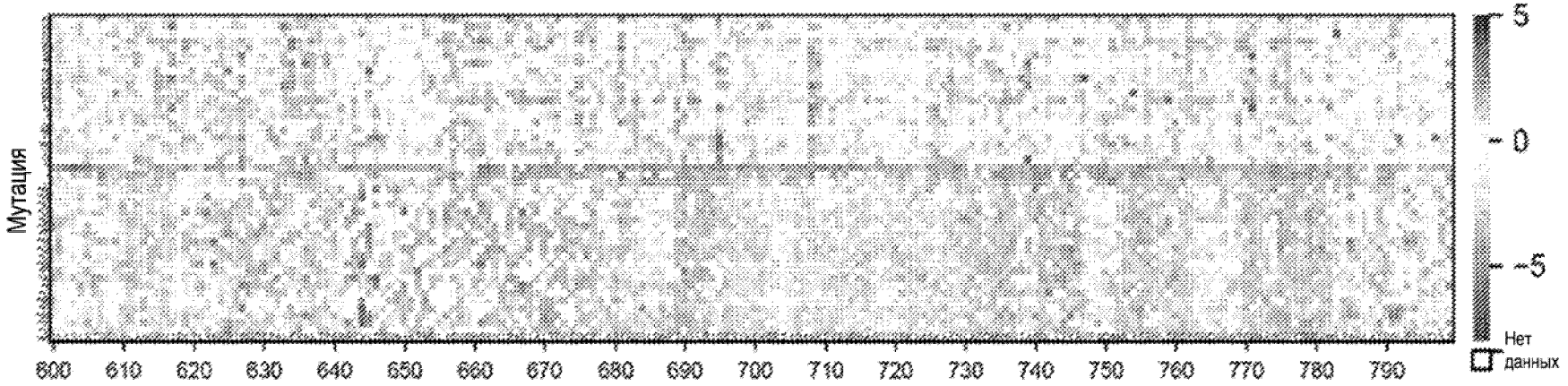
ФИГ. 35B



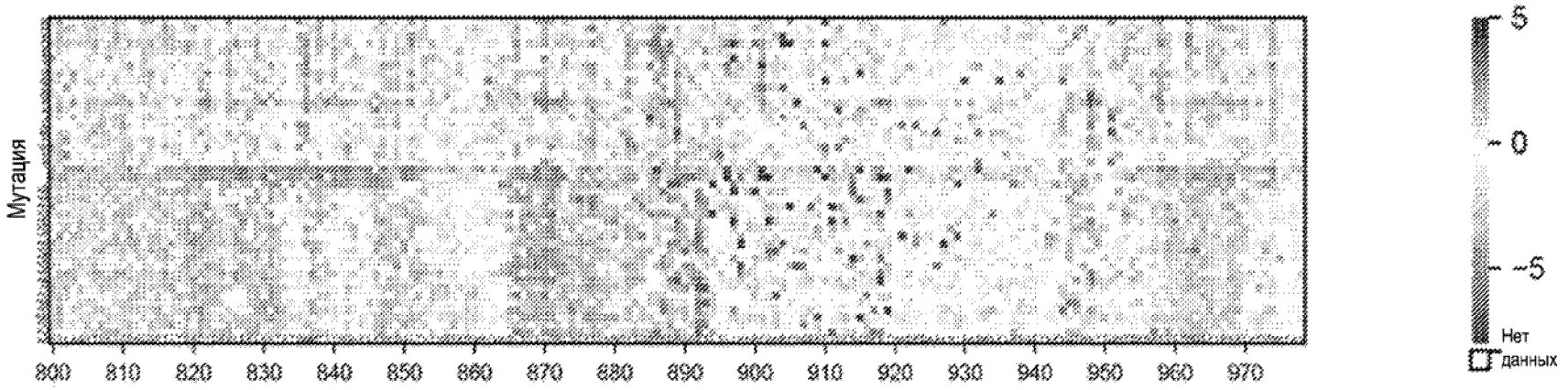
ФИГ. 35C



ФИГ. 35D

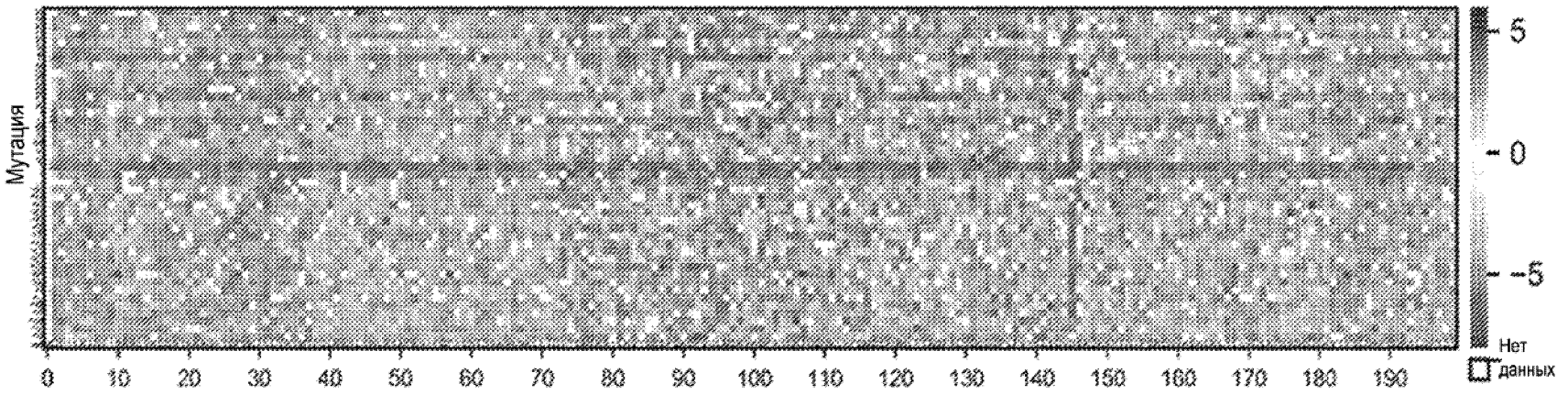


Положение в белке

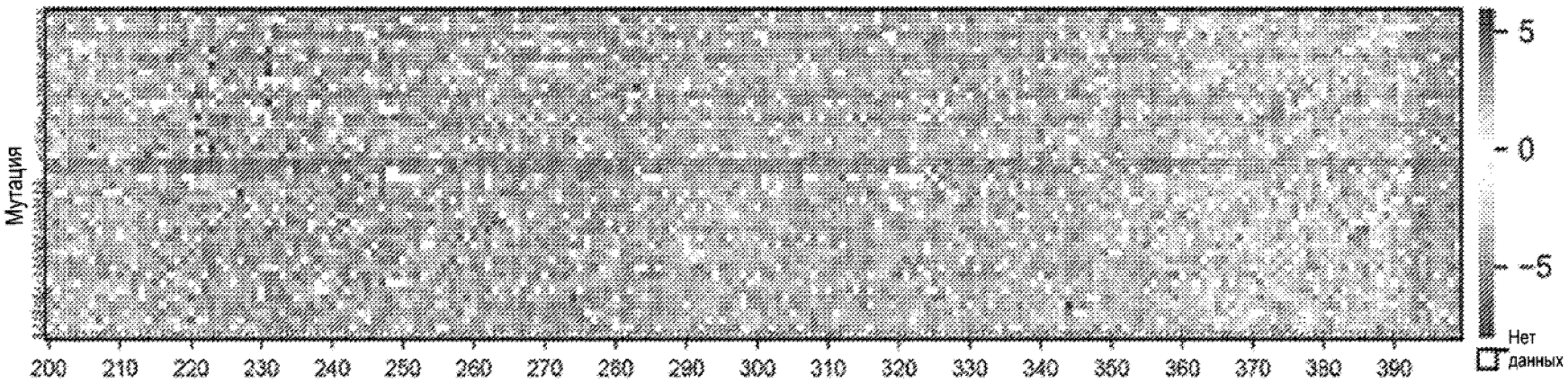

Log2 обогащения
5
0
−5
Нет данных
Мутация
800
810
820
830
840
850
860
870
880
890
900
910
920
930
940
950
960
970
Положение в белке

ФИГ. 35E

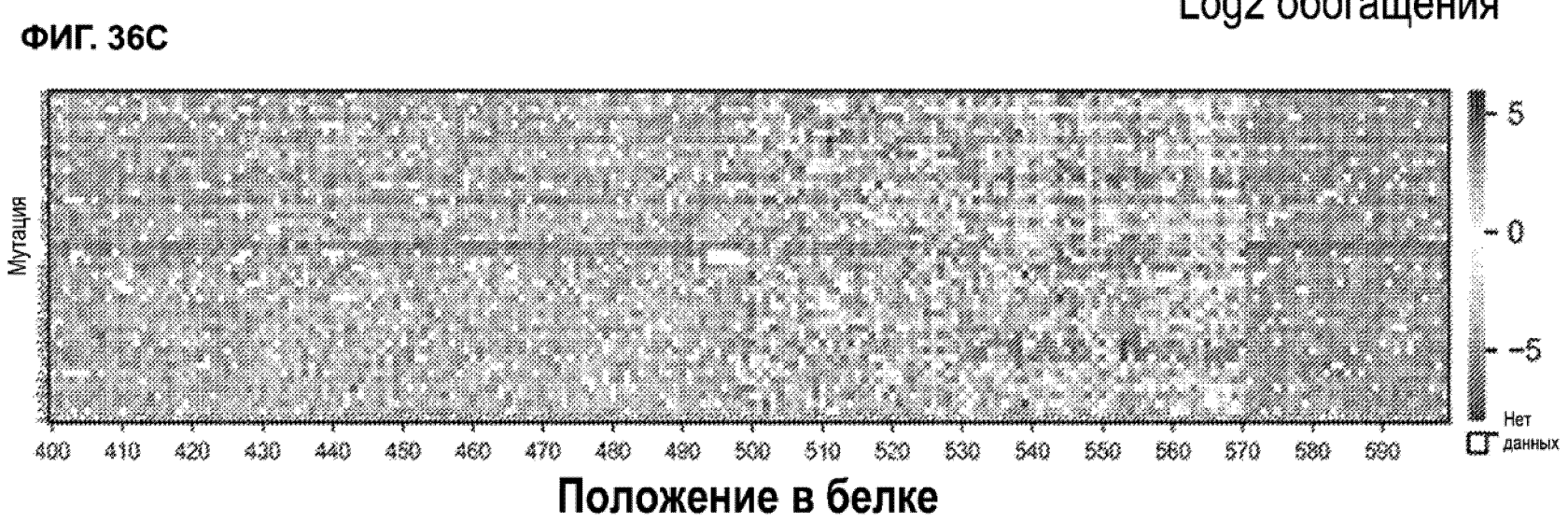
ФИГ. 36A



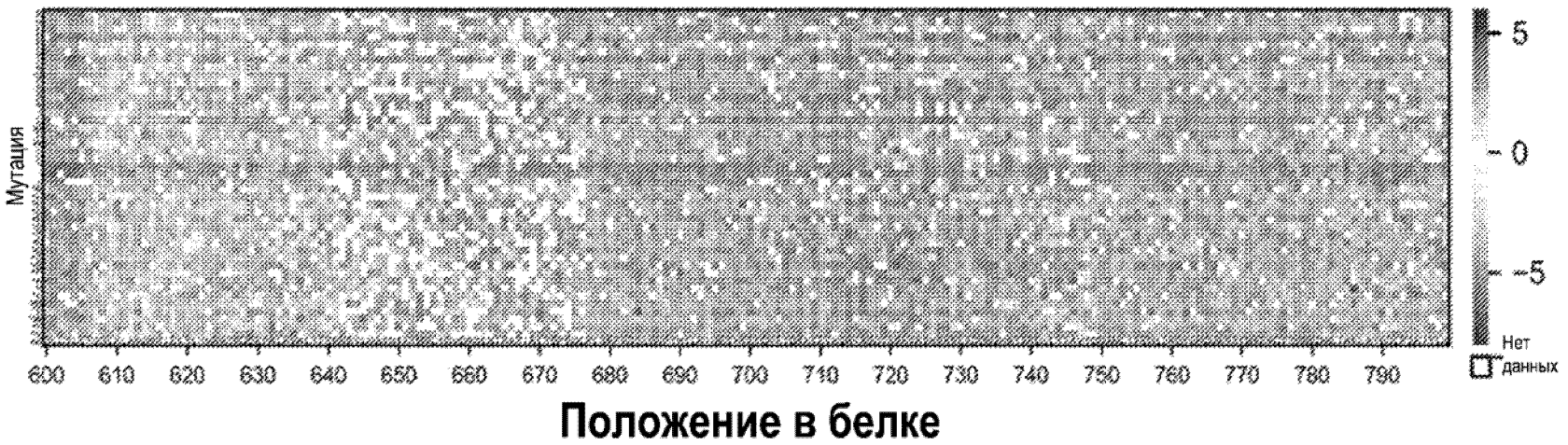
ФИГ. 36B

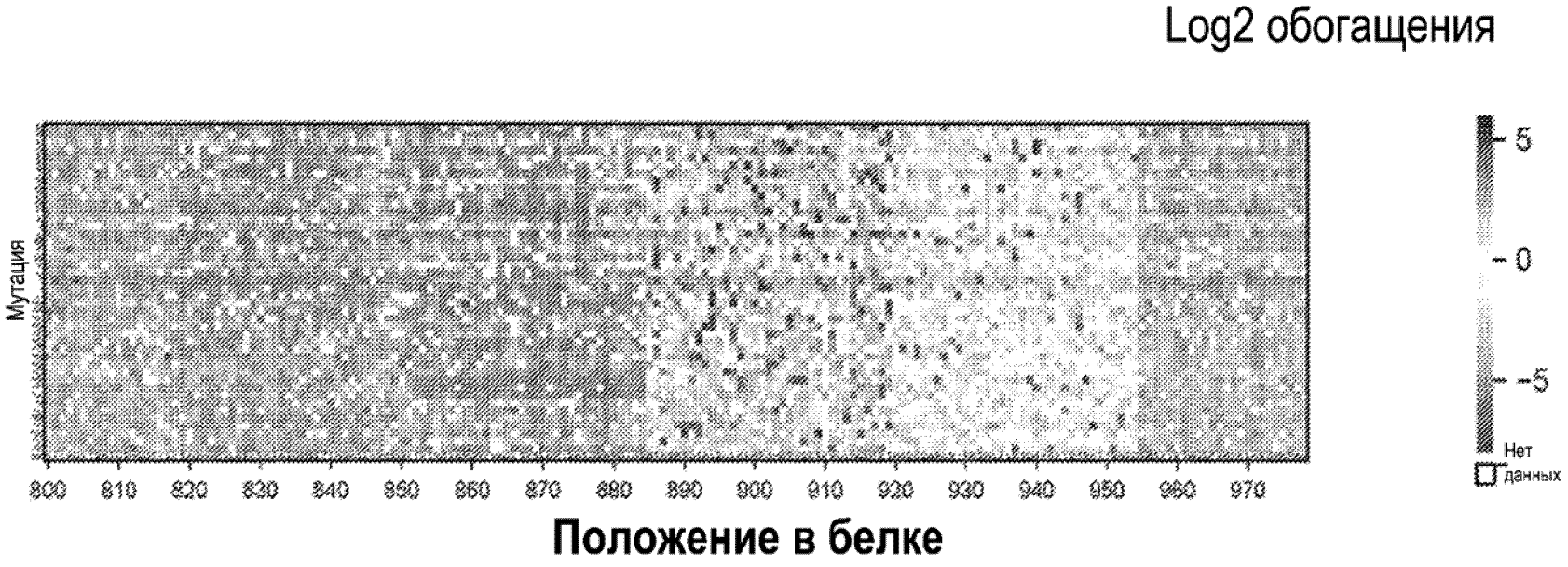


ФИГ. 36C



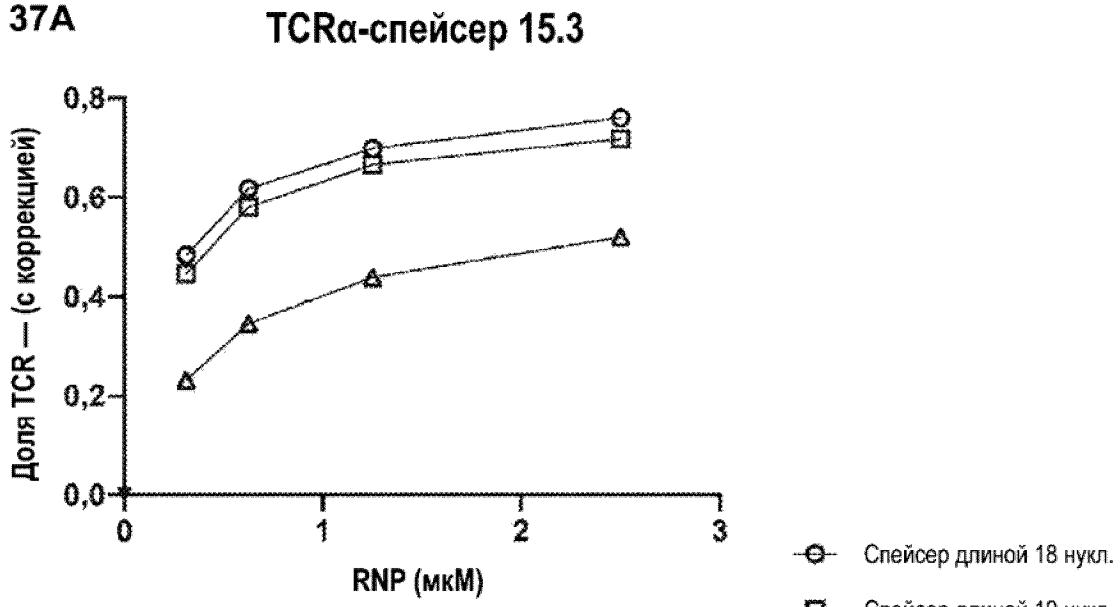
ФИГ. 36D



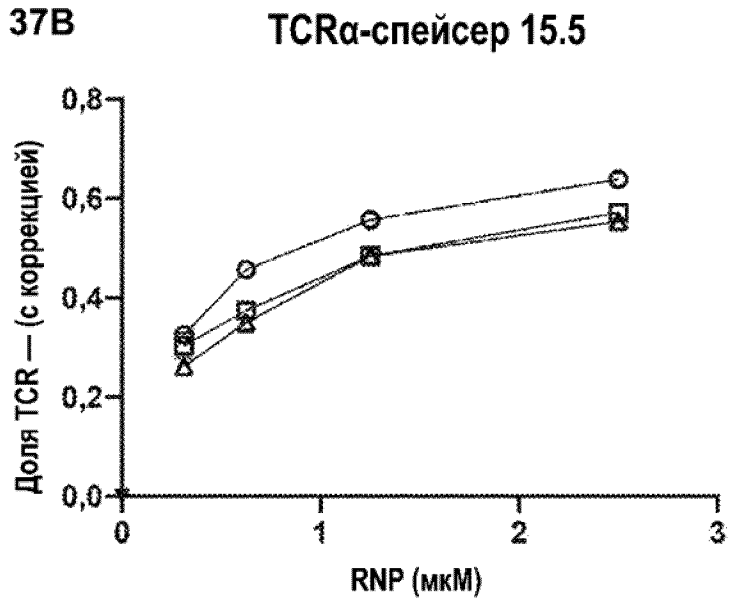

Log2 обогащения
5
0
-5
Нет данных
Мутация
800 810 820 830 840 850 860 870 880 890 900 910 920 930 940 950 960 970
Положение в белке

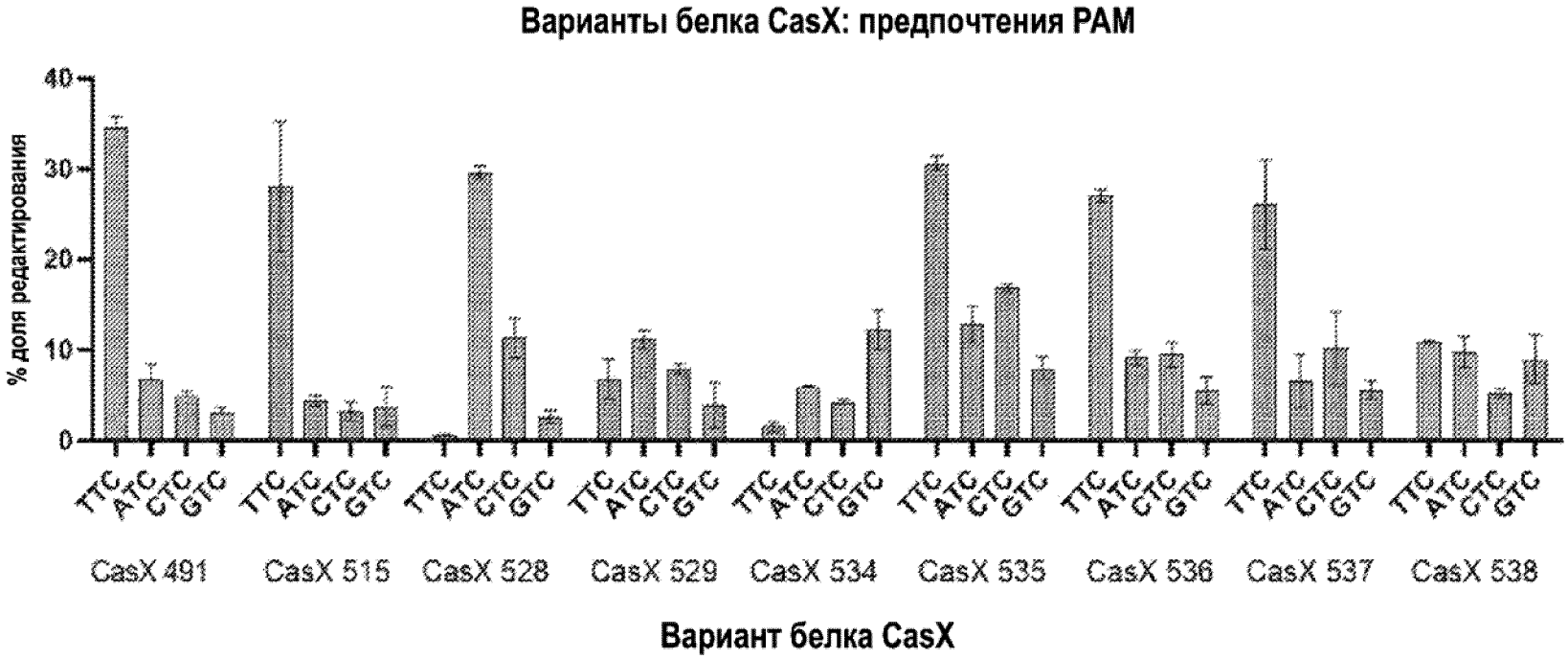
ФИГ. 36E

ФИГ. 37A



ФИГ. 37B

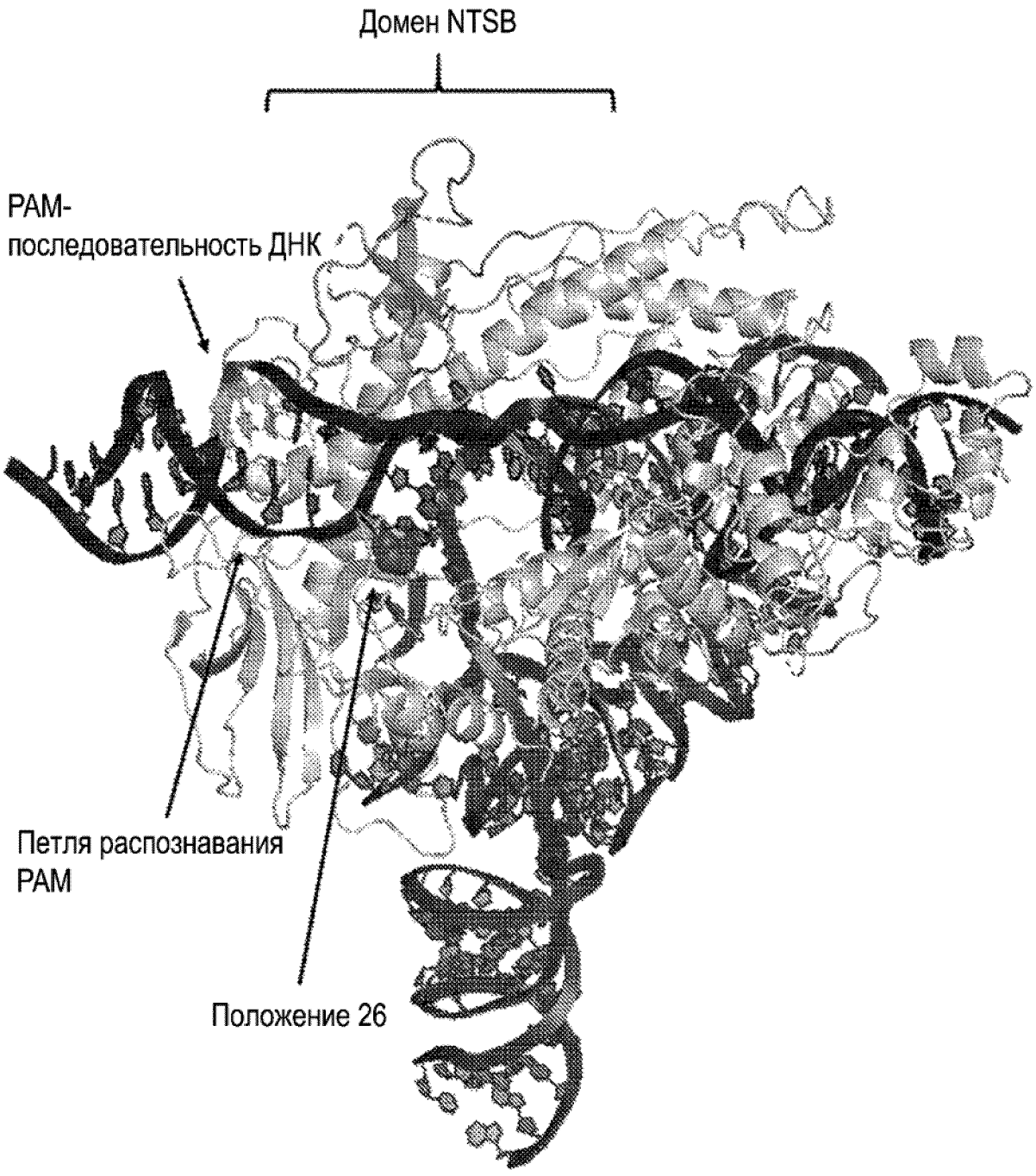



Варианты белка CasX: предпочтения PAM
% доля редактирования
40
30
20
10
0
TTC
ATC
CTC
GTC
CasX 491
CasX 515
CasX 528
CasX 529
CasX 534
CasX 535
CasX 536
CasX 537
CasX 538
Вариант белка CasX

ФИГ. 38


Средняя активность редактирования относительно CasX 491
****
Средняя активность редактирования относительно CasX 491
1,0
0,5
0,0
2
119
491
515
527
532
535
536
537
542
543
544
Белок CasX

ФИГ. 39


Домен NTSB
PAM-
последовательность ДНК
Петля распознавания
PAM
Положение 26

ФИГ. 40


Эффективность специфического редактирования нуклеазами Cas для 48 целевых сайтов TTC PAM
Процентная доля редактирования
100
80
60
40
20
0
2
527
545
551
552
553
554
556
557
558
541
563
565
576
579
581
583
589
Белок CasX

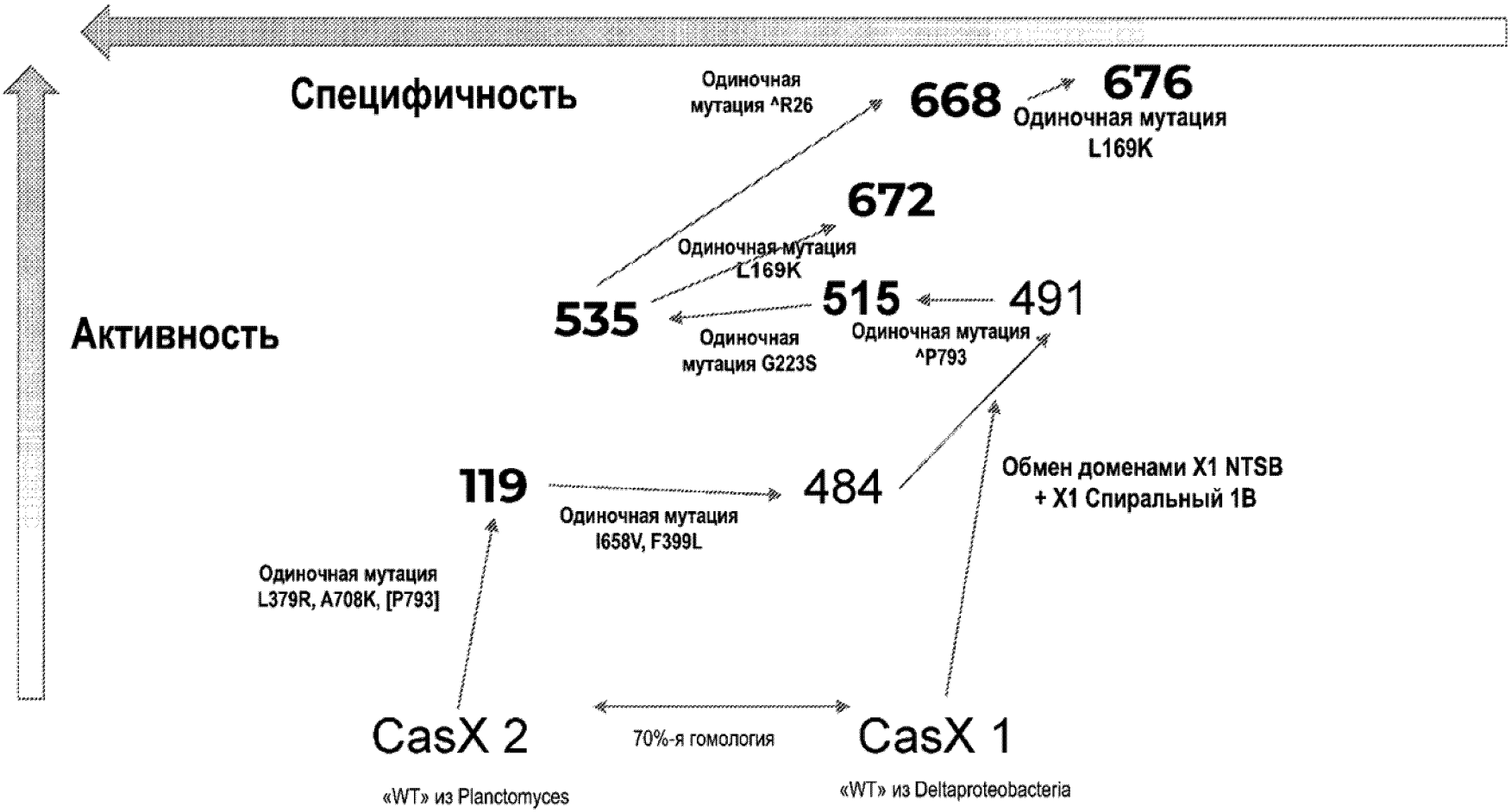
ФИГ. 41


Средняя активность редактирования относительно CasX 491
Средняя активность редактирования относительно CasX 491
1,0
0,5
0,0
2
119
429
432
433
434
435
437
438
440
441
442
443
444
446
447
448
449
450
451
452
453
454
455
456
457
458
484
491
Белок CasX

ФИГ. 42

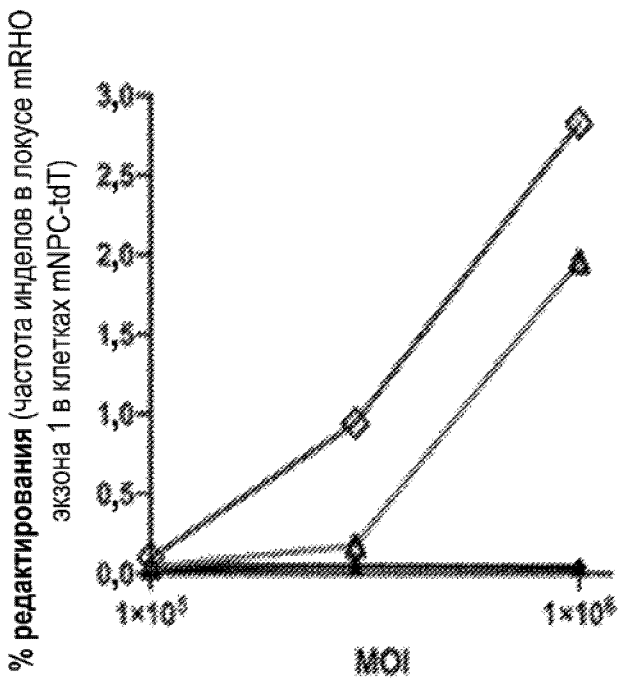

Активность и специфичность белков CasX
Активность редактирования относительно 491
Соотношение специфичности
Активность редактирования относительно CasX 491 и соотношение специфичности
1,0
0,5
0,0
2
119
491
515
527
532
535
536
537
542
543
544
Белок Cas

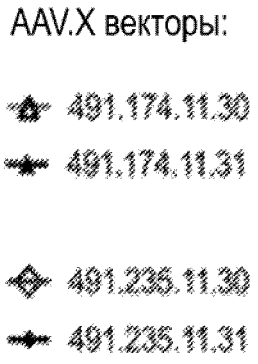
ФИГ. 43


Специфичность
Активность
Одиночная мутация ^R26
668
676
Одиночная мутация L169K
672
Одиночная мутация L169K
535
515
491
Одиночная мутация G223S
Одиночная мутация ^P793
119
484
Обмен доменами X1 NTSB + X1 Спиральный 1B
Одиночная мутация I658V, F399L
Одиночная мутация L379R, A708K, [P793]
CasX 2
70%-я гомология
CasX 1
«WT» из Planctomyces
«WT» из Deltaproteobacteria

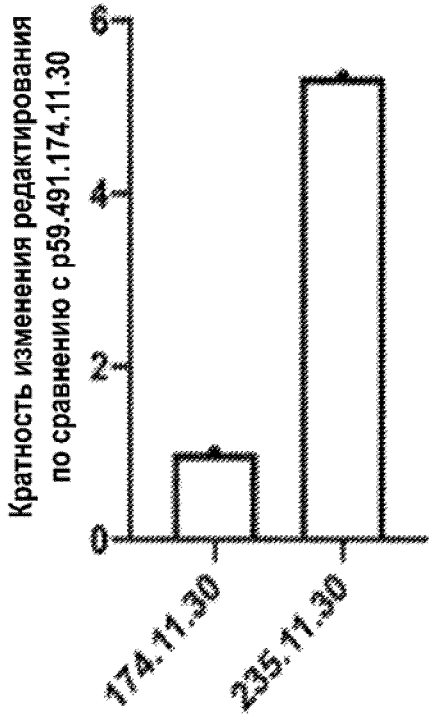
ФИГ. 44

ФИГ. 45А

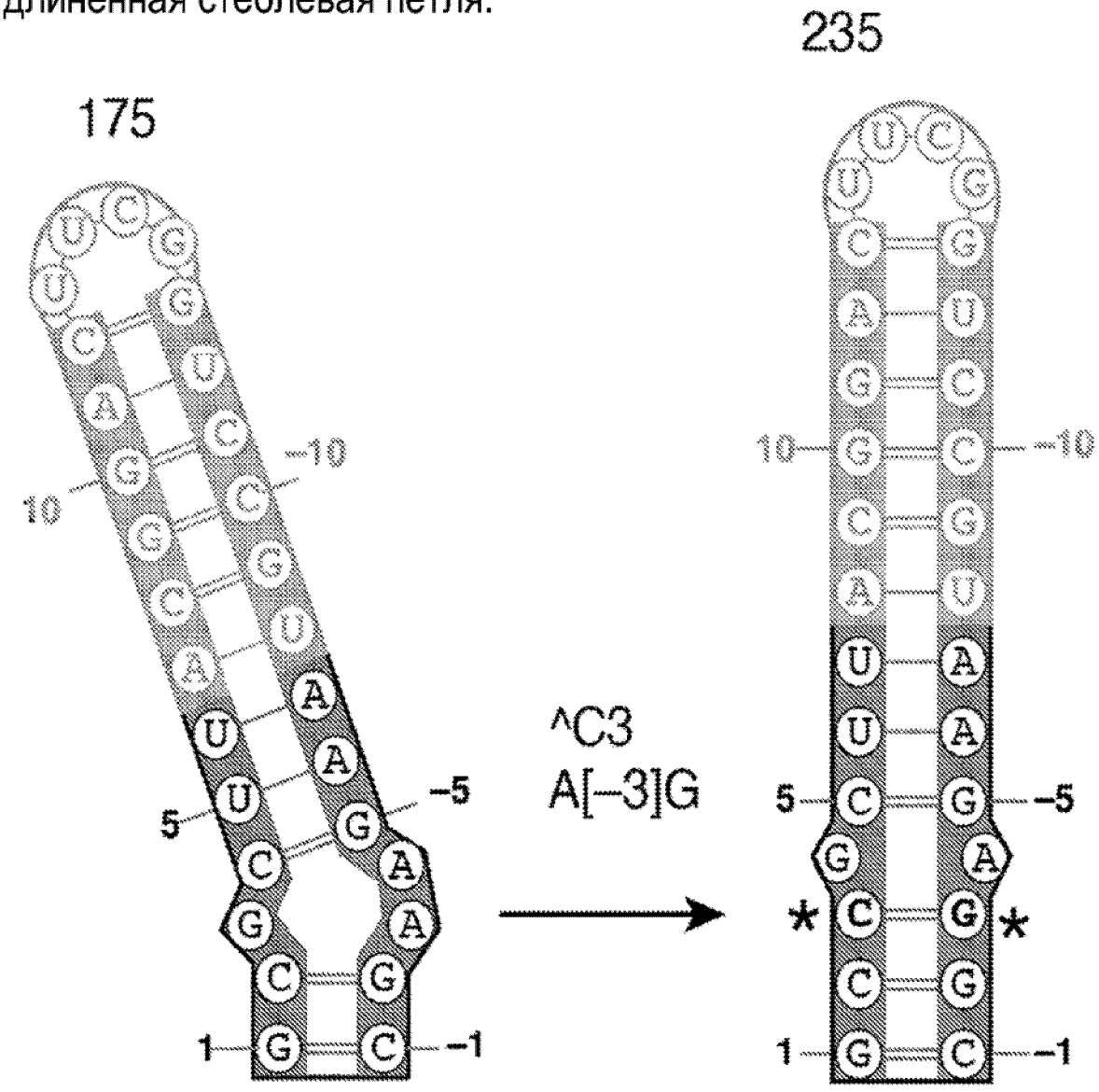




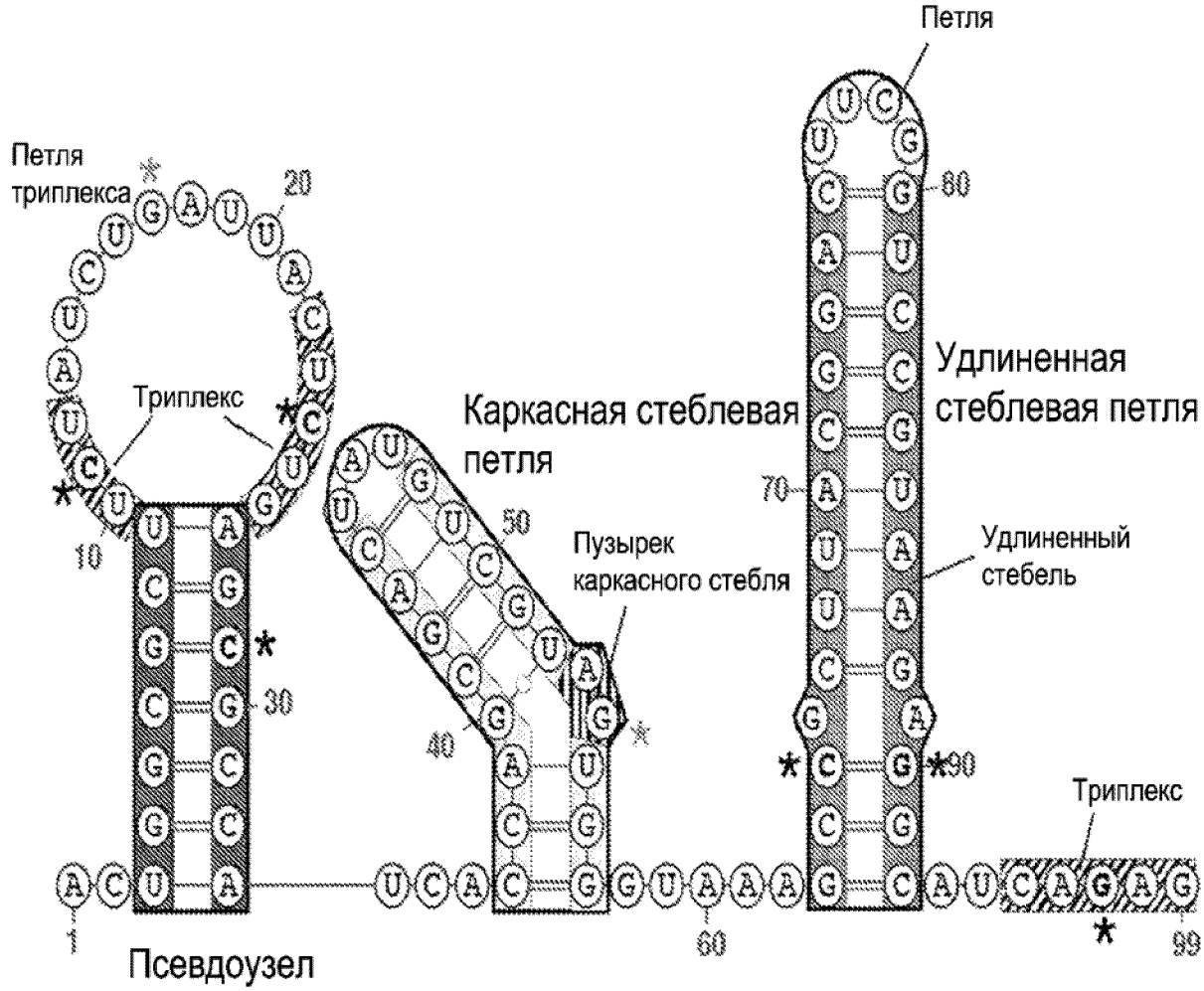
ФИГ. 45В



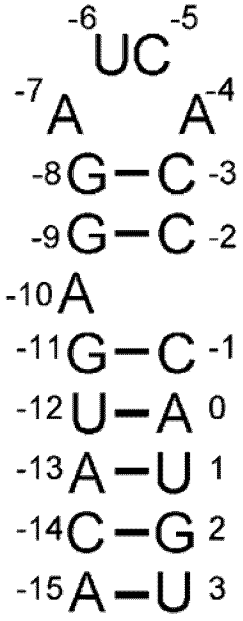
Удлиненная стеблевая петля:



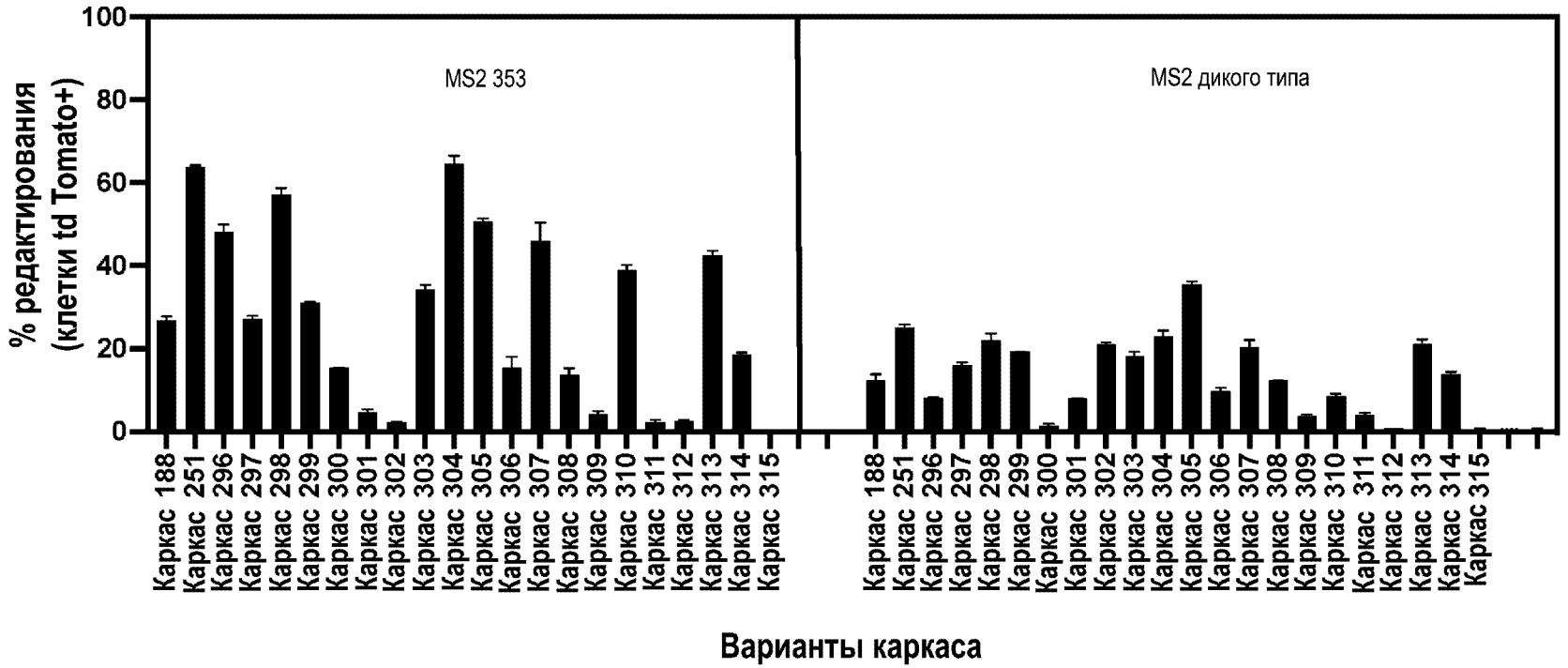
ФИГ. 46


Петля
Петля триплекса
Триплекс
Каркасная стеблевая петля
Пузырек каркасного стебля
Удлиненная стеблевая петля
Удлиненный стебель
Триплекс
Псевдоузел
1
10
20
30
40
50
60
70
80
90
99

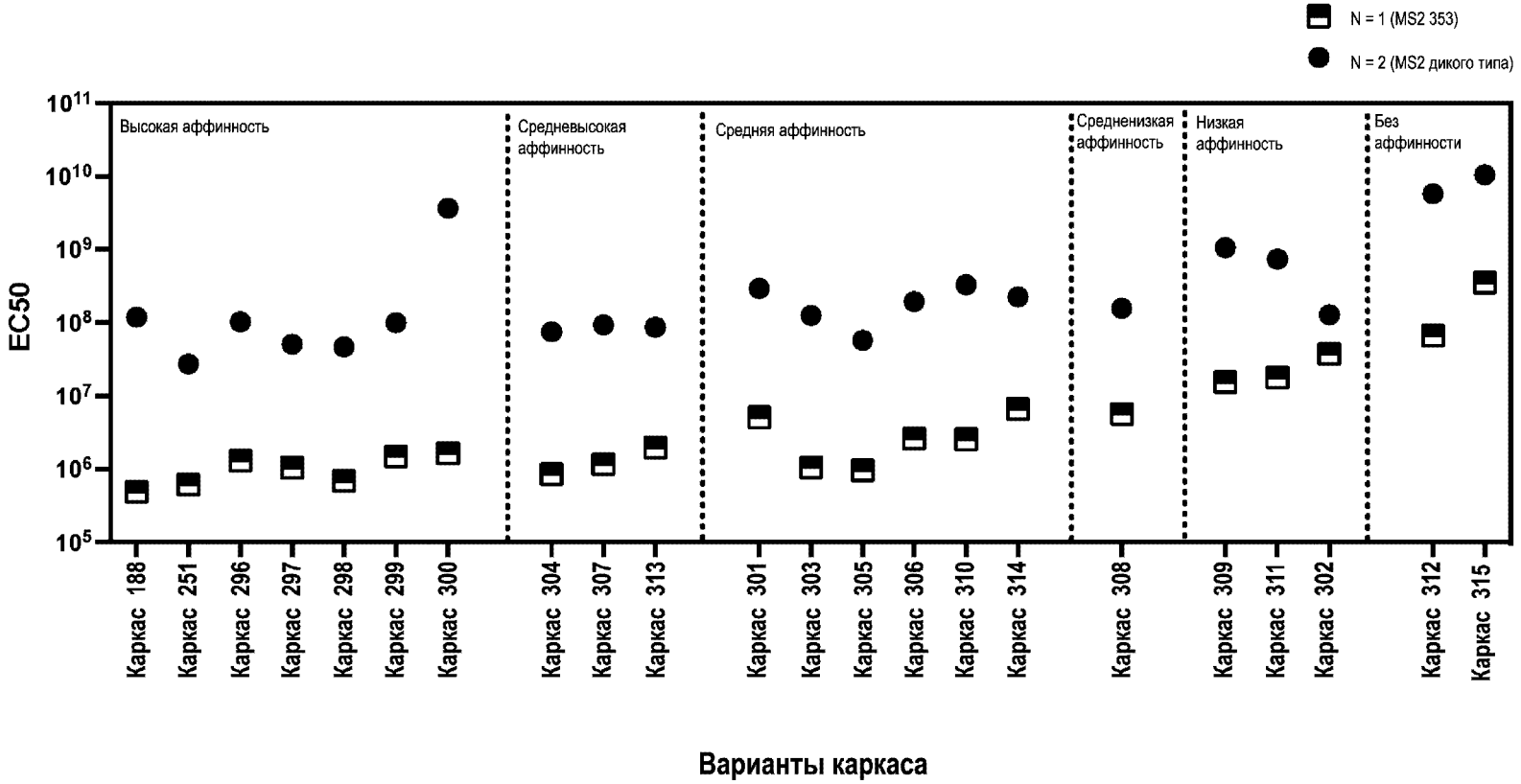
ФИГ. 47



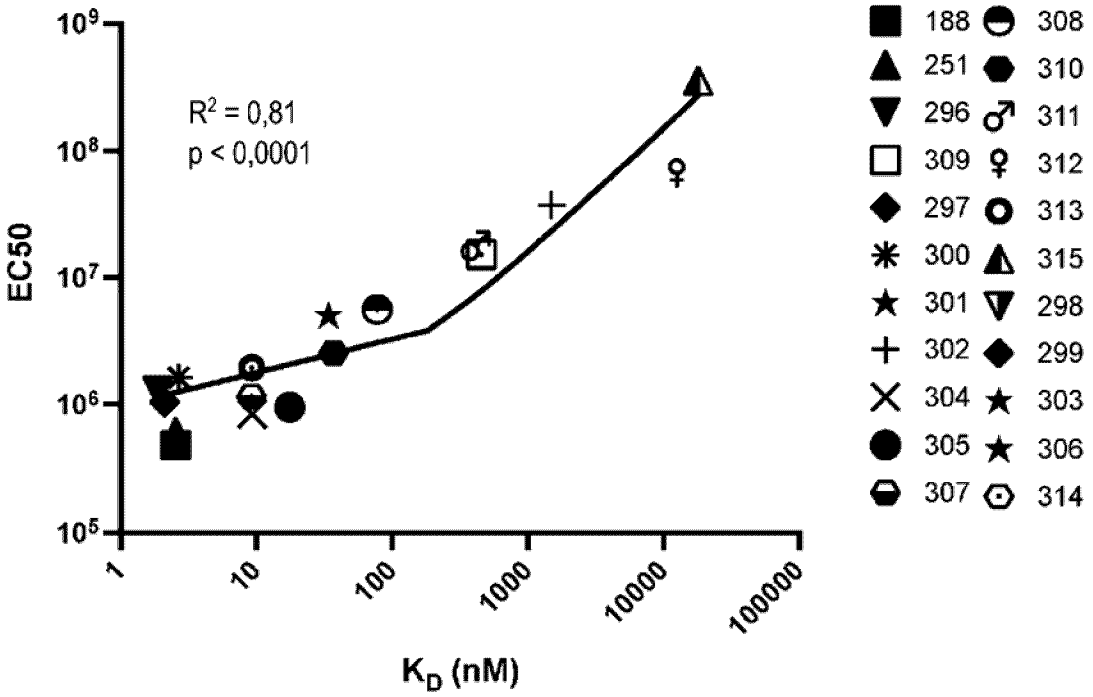
ФИГ. 48


% редактирования (клетки td Tomato+)
100
80
60
40
20
0
MS2 353
MS2 дикого типа
Каркас 188
Каркас 251
Каркас 296
Каркас 297
Каркас 298
Каркас 299
Каркас 300
Каркас 301
Каркас 302
Каркас 303
Каркас 304
Каркас 305
Каркас 306
Каркас 307
Каркас 308
Каркас 309
Каркас 310
Каркас 311
Каркас 312
Каркас 313
Каркас 314
Каркас 315
Варианты каркаса

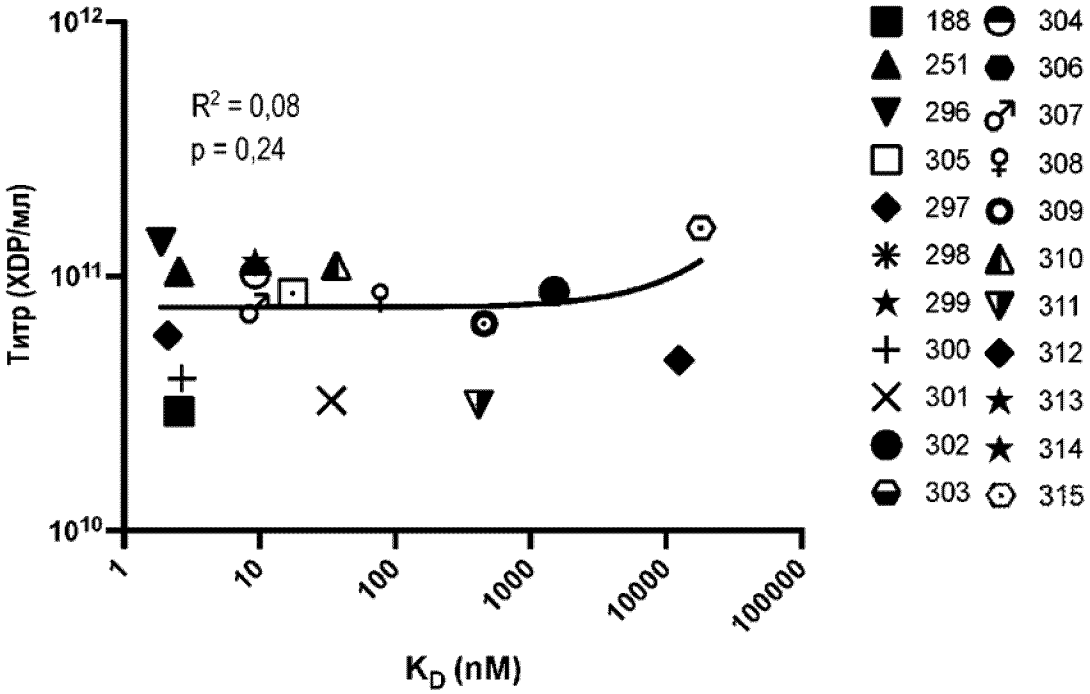
ФИГ. 49


N = 1 (MS2 353)
N = 2 (MS2 дикого типа)
EC50
10^{11}
10^{10}
10^{9}
10^{8}
10^{7}
10^{6}
10^{5}
Высокая аффинность
Средневысокая аффинность
Средняя аффинность
Средненизкая аффинность
Низкая аффинность
Без аффинности
Каркас 188
Каркас 251
Каркас 296
Каркас 297
Каркас 298
Каркас 299
Каркас 300
Каркас 304
Каркас 307
Каркас 313
Каркас 301
Каркас 303
Каркас 305
Каркас 306
Каркас 310
Каркас 314
Каркас 308
Каркас 309
Каркас 311
Каркас 302
Каркас 312
Каркас 315
Варианты каркаса

ФИГ. 50


10^9
10^8
10^7
10^6
10^5
EC50
R^2 = 0,81
p < 0,0001
1
10
100
1000
10000
100000
K_D (nM)
188
308
251
310
296
311
309
312
297
313
300
315
301
298
302
299
304
303
305
306
307
314

ФИГ. 51


10^{12}
10^{11}
10^{10}
Титр (XDP/мл)
R^2 = 0,08
p = 0,24
1
10
100
1000
10000
100000
K_D (nM)
188
251
296
305
297
298
299
300
301
302
303
304
306
307
308
309
310
311
312
313
314
315

ФИГ. 52