

(19)



Евразийское
патентное
ведомство

(21) 202491422 (13) A2

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ

(43) Дата публикации заявки
2024.08.30

(51) Int. Cl. C12N 15/10 (2006.01)
C12N 9/22 (2006.01)
C12Q 1/6816 (2018.01)

(22) Дата подачи заявки
2020.08.11

(54) РНК-НАПРАВЛЯЕМЫЕ НУКЛЕАЗЫ, ИХ АКТИВНЫЕ ФРАГМЕНТЫ И ВАРИАНТЫ, А ТАКЖЕ СПОСОБЫ ПРИМЕНЕНИЯ

(31) 62/885,483; 62/901,875; 63/030,088

(32) 2019.08.12; 2019.09.18; 2020.05.26

(33) US

(62) 202290537; 2020.08.11

(71) Заявитель:
ЛАЙФЭДИТ ТЕРАПЬЮТИКС, ИНК.
(US)

(72) Изобретатель:

Боуэн Тайсон Д., Кроли Александра
Брайнер, Элич Теда Д., Койл Майкл
(US)

(74) Представитель:

Веселицкий М.Б., Кузенкова Н.В.,
Каксис Р.А., Белоусов Ю.В., Куликов
А.В., Кузнецова Е.В., Соколов Р.А.,
Кузнецова Т.В. (RU)

(57) В изобретении предусматривают композиции и способы связывания с представляющей интерес последовательностью-мишенью. Композиции применяют для расщепления или модификации представляющей интерес последовательности-мишени, визуализации представляющей интерес последовательности-мишени и модификации экспрессии представляющей интерес последовательности. Композиции содержат полипептиды РНК-направляемых нуклеаз, РНК CRISPR, трансактивирующие РНК CRISPR, направляющие РНК и кодирующие их молекулы нуклеиновой кислоты. Также описывают векторы и клетки-хозяева, содержащие молекулы нуклеиновой кислоты. Кроме того, предложены системы CRISPR для связывания представляющей интерес последовательности-мишени, причем система CRISPR содержит полипептид РНК-направляемой нуклеазы и одну или несколько направляющих РНК. Также предусмотрены способы и наборы для обнаружения последовательности ДНК-мишени.

A2

202491422

202491422

A2

РНК-НАПРАВЛЯЕМЫЕ НУКЛЕАЗЫ, ИХ АКТИВНЫЕ ФРАГМЕНТЫ И ВАРИАНТЫ, А ТАКЖЕ СПОСОБЫ ПРИМЕНЕНИЯ

5

Область техники

Настоящее изобретение относится к области молекулярной биологии и генному редактированию.

10 Ссылка на перечень последовательностей, представленный в виде
текстового файла через EFS-Web

Настоящая заявка содержит Перечень последовательностей, который был представлен в формате ASCII через систему EFS-Web электронной подачи заявок Управления по патентам и товарным знакам США (USPTO) и настоящим полностью включен посредством ссылки. Указанная копия ASCII, созданная 11
15 августа 2020 г., называется L103438_1170WO_0049_2_Seq_List.txt и имеет размер 489326 байтов.

Предшествующий уровень техники

Целевое редактирование или модификация генома быстро становится важным инструментом фундаментальных и прикладных исследований.
20 Первоначальные методы включали нуклеазы для инженерии, такие как мегануклеазы, слитые цинк-пальцевые белки или TALEN, требующие получения химерных нуклеаз со сконструированными, программируемыми, специфичными в отношении последовательности ДНК-связывающими доменами, специфичными для каждой конкретной последовательности-мишени. РНК-
25 направляемые нуклеазы, такие как белки Cas, ассоциированные с кластеризованными регулярно расположенными короткими палиндромными повторами (CRISPR – Clustered Regularly Interspaced Short Palindromic Repeats) бактериальной системы CRISPR-Cas ((CRISPR)-associated – Cas), позволяют нацеливаться на определенные последовательности путем образования
30 комплексов нуклеаз с направляющей РНК, которая специфически гибридизируется с конкретной целевой последовательностью. Получение направляющих РНК, специфичных в отношении мишени, менее затратно и более эффективно, чем создание химерных нуклеаз для каждой последовательности-мишени. Такие РНК-направляемые нуклеазы можно использовать для

редактирования геномов необязательно путем введения специфического для последовательности двухцепочечного разрыва, который восстанавливается с помощью подверженного ошибкам негомологичного соединения концов (NHEJ – non-homologous end-joining) для введения мутации в определенное место генома.

5 В другом варианте гетерологичная ДНК может быть введена в геномный сайт через направленную по гомологии репарацию.

Краткое описание изобретения

Предложены композиции и способы связывания представляющей интерес последовательности-мишени. Композиции могут быть применены для

10 расщепления или модификации представляющей интерес последовательности-мишени, обнаружения представляющей интерес последовательности-мишени и модификации экспрессии представляющей интерес последовательности.

Композиции содержат полипептиды РНК-направляемой нуклеазы (RGN – RNA-guided nuclease), CRISPR РНК (crРНК – CRISPR РНК), транс-активирующие

15 CRISPR РНК (tracrРНК – trans-activating CRISPR РНК), направляющие РНК (gРНК – guide РНК), молекулы нуклеиновой кислоты, кодирующие их, а также векторы и клетки-хозяева, содержащие молекулы нуклеиновой кислоты. Также предложены системы CRISPR для связывания представляющей интерес последовательности-мишени, причем система CRISPR содержит полипептид

20 РНК-направляемой нуклеазы и одну или несколько направляющих РНК. Таким образом, способы, предусмотренные в настоящем изобретении, направлены на связывание представляющей интерес последовательности-мишени и, в некоторых вариантах осуществления настоящего изобретения, на расщепление или модификацию представляющей интерес последовательности-мишени.

25 Представляющая интерес последовательность-мишень может быть модифицирована, например, в результате негомологичного соединения концов или направленной по гомологии репарации с введенной донорной последовательностью. Кроме того, предусмотрены способы и наборы для обнаружения последовательности-мишени ДНК молекулы ДНК с

30 использованием детекторной одноцепочечной ДНК.

Подробное описание изобретение

Многие модификации и другие варианты осуществления настоящего изобретения, приведенные ниже, понятны специалистам в данной области

техники, к которой относится это изобретение, с учетом принципов, представленных в предшествующих описаниях и связанных с ними фигурах. Таким образом, следует учитывать, что изобретения не должны ограничиваться конкретными вариантами осуществления, и что модификации и другие варианты осуществления предназначены для включения в рамки охвата прилагаемых вариантов осуществления изобретения. Хотя в описании используются специфические термины, они используются только для описания в общепринятом смысле, а не в целях ограничения.

I. Обзор

РНК-направляемые нуклеазы (RGN) позволяют целенаправленно манипулировать одним сайтом в геноме, и полезны в плане нацеливания на гены для терапевтических и исследовательских целей. У различных организмов, включая млекопитающих, РНК-направляемые нуклеазы используют для инженерии генома, например, путем стимуляции негомологичного соединения концов и гомологичной рекомбинации. Композиции и способы, описанные в настоящем изобретении, применимы для создания одноцепочечных или двухцепочечных разрывов в полинуклеотидах, модификации полинуклеотидов, обнаружения конкретного сайта в полинуклеотиде или модификации экспрессии конкретного гена.

Описанные в настоящем изобретении РНК-направляемые нуклеазы могут изменять экспрессию генов путем модификации последовательности-мишени. В конкретных вариантах осуществления настоящего изобретения РНК-направляемые нуклеазы направляются к последовательности-мишени с помощью направляющей РНК (gРНК – guide RNA) как части РНК-направляемой нуклеазной системы кластерированных регулярно расположенных коротких палиндромных повторов (CRISPR – Clustered Regularly Interspaced Short Palindromic Repeats). RGN рассматривают как «РНК-направляемые», поскольку направляющие РНК образуют комплекс с РНК-направляемыми нуклеазами для направления РНК-направляемой нуклеазы на связывание с последовательностью-мишенью, а в некоторых вариантах осуществления настоящего изобретения для внедрения одноцепочечного или двухцепочечного разрыва в целевую последовательность (последовательность-мишень). После расщепления последовательности-мишени разрыв может быть репарирован

таким образом, что последовательность ДНК целевой последовательности модифицируется в процессе репарации. Таким образом, в настоящем изобретении предусмотрены способы применения РНК-направляемых нуклеаз для модификации последовательности-мишени в ДНК клеток-хозяев. Например, РНК-направляемые нуклеазы можно использовать для модификации последовательности-мишени в геномном локусе эукариотических клеток или прокариотических клеток.

II. РНК-направляемые нуклеазы

В настоящем изобретении предусматривают РНК-направляемые нуклеазы.

10 Термин РНК-направляемая нуклеаза (RGN) относится к полипептиду, который связывается с определенной последовательностью нуклеотидов-мишеней специфичным для последовательности образом и направляется к целевой нуклеотидной последовательности с помощью молекулы направляющей РНК, которая образует комплекс с полипептидом и гибридизируется с
15 последовательностью-мишенью. Хотя РНК-направляемая нуклеаза способна расщеплять последовательность-мишень при связывании, термин «РНК-направляемая нуклеаза» также охватывает нуклеазы, направляемые РНК, которые способны связываться с целевой последовательностью, но не расщепляют ее. Расщепление последовательности-мишени РНК-направляемой нуклеазой может привести к одноцепочечному или двухцепочечному разрыву.
20 РНК-направляемые нуклеазы, способные расщеплять только одну цепь двухцепочечной молекулы нуклеиновой кислоты, называются никазами.

К РНК-направляющим нуклеазам, описанным в настоящем изобретении, относятся APG05733.1, APG06207.1, APG01647.1, APG08032.1, APG05712.1,
25 APG01658.1, APG06498.1, APG09106.1, APG09882.1, APG02675.1, APG01405.1, APG06250.1, APG06877.1, APG09053.1, APG04293.1, APG01308.1, APG06646.1, APG09748, APG07433.1, аминокислотные последовательности которых представлены соответственно как SEQ ID NO: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 117, 137 или 235, и их активные фрагменты или
30 варианты, которые сохраняют способность связываться с последовательностью нуклеотидов-мишенью специфическим для РНК-направляемой последовательности образом. В некоторых из этих вариантов осуществления настоящего изобретения активный фрагмент или вариант RGN из APG05733.1, APG06207.1, APG01647.1, APG08032.1, APG05712.1, APG01658.1, APG06498.1,

APG09106.1, APG09882.1, APG02675.1, APG01405.1, APG06250.1, APG06877.1, APG09053.1, APG04293.1, APG01308.1, APG06646.1, APG09748 или APG07433.1 способен расщеплять одно- или двухцепочечную последовательность-мишень. В некоторых вариантах осуществления настоящего изобретения активный вариант RGN из APG05733.1, APG06207.1, APG01647.1, APG08032.1, APG05712.1, APG01658.1, APG06498.1, APG09106.1, APG09882.1, APG02675.1, APG01405.1, APG06250.1, APG06877.1, APG09053.1, APG04293.1, APG01308.1, APG06646.1, APG09748 или APG07433.1 содержит аминокислотную последовательность, которая по меньшей мере на 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или более идентична аминокислотной последовательности, представленной в виде SEQ ID NO: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 117, 137 или 235. В некоторых вариантах осуществления настоящего изобретения активный фрагмент RGN из APG05733.1, APG06207.1, APG01647.1, APG08032.1, APG05712.1, APG01658.1, APG06498.1, APG09106.1, APG09882.1, APG02675.1, APG01405.1, APG06250.1, APG06877.1, APG09053.1, APG04293.1, APG01308.1, APG06646.1, APG09748 или APG07433.1 содержит по меньшей мере 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050 или более непрерывно расположенных аминокислотных остатков аминокислотной последовательности, представленной в виде SEQ ID NO: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 117, 137 или 235. РНК-направляемые нуклеазы, предусмотренные в настоящем изобретении, могут содержать по меньшей мере один нуклеазный домен (например, домен ДНКазы, РНКазы) и по меньшей мере один РНК-распознающий и/или РНК-связывающий домен для взаимодействия с направляющими РНК. Дополнительные домены, которые могут быть обнаружены в РНК-направляемых нуклеазах, предусмотренные в настоящем изобретении, включают, но не ограничиваются ими: ДНК-связывающие домены, домены геликазы, домены взаимодействия белок-белок и домены димеризации. В конкретных вариантах осуществления настоящего изобретения РНК-направляемые нуклеазы, предусмотренные в настоящем изобретении, могут содержать по меньшей мере 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% из ДНК-связывающих доменов, доменов геликазы, доменов взаимодействия белок-белок и доменов димеризации.

Нуклеотидная последовательность-мишень связывается с РНК-направляемой нуклеазой, предусмотренной в настоящем изобретении, и гибридизируется с направляющей РНК, связанной с РНК-направляемой нуклеазой. Затем последовательность-мишень может быть расщеплена РНК-направляемой нуклеазой, если полипептид обладает нуклеазной активностью. Термины «расщепить» или «расщепление» относятся к гидролизу по меньшей мере одной фосфодиэфирной связи в каркасе целевой последовательности нуклеотидов, что может привести к одноцепочечным или двухцепочечным разрывам в целевой последовательности. Раскрытые в настоящее время RGN могут расщеплять нуклеотиды внутри полинуклеотида, функционируя как эндонуклеаза, или могут быть экзонуклеазой, удаляя последовательные нуклеотиды с конца (5'- и/или 3'-конца) полинуклеотида. В других вариантах осуществления настоящего изобретения описанные RGN могут расщеплять нуклеотиды в целевой последовательности в любом положении полинуклеотида, функционируя одновременно и как эндонуклеаза, и как экзонуклеаза. Расщепление полинуклеотида-мишени описанными в настоящем документе RGN может привести к смещенным разрывам или тупым концам.

Описанные в настоящее время РНК-направляемые нуклеазы могут представлять собой последовательности дикого типа, полученные из видов бактерий или архей. В другом варианте РНК-направляемые нуклеазы могут быть вариантами или фрагментами полипептидов дикого типа. RGN дикого типа можно модифицировать, например, для изменения активности нуклеазы или изменения специфичности РАМ. В некоторых вариантах осуществления настоящего изобретения применяют РНК-направляемую нуклеазу, не встречающуюся в природе.

В некоторых вариантах осуществления настоящего изобретения РНК-направляемая нуклеаза действует как никаза, расщепляя только одну цепь целевой нуклеотидной последовательности. Такие РНК-направляемые нуклеазы имеют один функционирующий нуклеазный домен. В некоторых из этих вариантов осуществления настоящего изобретения были мутированы дополнительные нуклеазные домены таким образом, что нуклеазная активность снижена или устранена.

В других вариантах осуществления настоящего изобретения РНК-направляемая нуклеаза полностью лишена нуклеазной активности или проявляет

пониженную нуклеазную активность, и в настоящем изобретении ее называют мертвой нуклеазой или неактивной нуклеазой. Любой способ, известный в данной области техники для введения мутаций в аминокислотную последовательность, такой как ПЦР-опосредованный мутагенез и сайт-направленный мутагенез, может быть использован для получения RGN без нуклеазной активности. См., например, публикацию US № 2014/0068797 и US 9790490, сущность которых включена в настоящее изобретение посредством ссылки.

РНК-направляемые нуклеазы, не обладающие нуклеазной активностью, можно использовать для доставки слитых полипептидов, полинуклеотидов или малых молекул в определенное место генома. В некоторых из этих вариантов осуществления настоящего изобретения полипептид RGN или направляющая РНК могут быть слиты с детектируемой меткой, чтобы обеспечить обнаружение конкретной последовательности. В качестве примера, не ограничивающего области охвата настоящего изобретения, RGN без нуклеазной активности можно слить с обнаруживаемой меткой (например, флуоресцентным белком) и нацелить на конкретную последовательность, связанную с заболеванием, чтобы обнаружить последовательность, связанную с заболеванием.

В другом варианте RGN без нуклеазной активности можно нацелить на определенные места генома, чтобы изменить экспрессию нужной последовательности. В некоторых вариантах осуществления настоящего изобретения связывание РНК-направляемой мертвой нуклеазы с последовательностью-мишенью приводит к подавлению экспрессии последовательности-мишени или гена, находящегося под контролем транскрипции последовательностью-мишенью путем вмешательства в связывание РНК-полимеразы или факторов транскрипции в целевой области генома. В других вариантах осуществления настоящего изобретения RGN (например, RGN без нуклеазной активности) или его комплексная направляющая РНК дополнительно содержит модулятор экспрессии, который при связывании с последовательностью-мишенью служит либо для подавления, либо для активации экспрессии последовательности-мишени или гена, находящегося под транскрипционным контролем последовательностью-мишенью. В некоторых из этих вариантов осуществления настоящего изобретения модулятор экспрессии

модулирует экспрессию целевой последовательности или регулируемого гена посредством эпигенетических механизмов.

В некоторых вариантах осуществления настоящего изобретения мертвые RGN или RGN, обладающий только никазой активностью, могут быть нацелены на определенные участки генома для модификации последовательности полинуклеотида-мишени путем слияния с полипептидом, редактирующим основание, например, полипептидом дезаминазы или его активным вариантом или фрагментом, который дезаминирует нуклеотидное основание, что приводит к превращению одного нуклеотидного основания в другое. Полипептид, редактирующий основание, может быть слит с RGN на её N-конце или C-конце. Кроме того, полипептид, редактирующий основания, может быть слит с RGN через пептидный линкер. Пример, не ограничивающий рамок охвата настоящего изобретения, полипептида дезаминазы, который можно использовать для таких композиций и способов, включает цитидиндезаминазу или аденозиндезаминазу (такую как дезаминаза, редактирующая основания аденозина, описанная Gaudelli с соавт., *Nature*, 2017, 551:464-471, US 2017/0121693, US 2018/0073012, WO/2018/027078, или любая из дезаминаз, описанных в PCT/US2019/068079, сущность каждой из которых включена в настоящее изобретение в виде ссылки). Кроме того, в данной области техники известно, что некоторые слитые белки RGN и фермента, редактирующего основания, могут также содержать по меньшей мере один полипептид, стабилизирующий урацил, который увеличивает скорость мутации цитидина, дезоксицитидина или цитозина в тимидин, дезокситимидин или тимин в молекуле нуклеиновой кислоты с помощью дезаминазы. Не ограничивающие рамок охвата настоящего изобретения примеры полипептидов, стабилизирующих урацил, включают полипептиды, описанные в патентной заявке US 63/052175, поданной 15 июля 2020 г., и домен ингибитора урацилгликозилазы (UGI – uracil glycosylase inhibitor) (SEQ ID NO: 261), который может повышать эффективность редактирования оснований (US 10167547, включенный в настоящее изобретение в виде ссылки). Таким образом, слитый белок может содержать RGN, описанную в настоящем изобретении, или её вариант, дезаминазу и необязательно по меньшей мере один полипептид, стабилизирующий урацил, такой как UGI.

РНК-направляемые нуклеазы, слитые с полипептидом или доменом, могут быть разделены или соединены линкером. Термин «линкер», используемый в

настоящем изобретении, относится к химической группе или молекуле, связывающей две молекулы или два фрагменты молекулы, например, связывающий домен и домен расщепления нуклеазы. В некоторых вариантах осуществления настоящего изобретения линкер соединяет связывающий домен gРНК РНК-направляемой нуклеазы и полипептид, редактирующий основания, такой как дезаминаза. В некоторых вариантах осуществления настоящего изобретения линкер соединяет мертвую RGN и дезаминазу. Как правило, линкер расположен между двумя группами, молекулами или фрагментами молекул или фланкирован ими и соединен с каждым из них ковалентной связью, таким образом, объединяя их. В некоторых вариантах осуществления настоящего изобретения линкер представляет собой аминокислоту или множество аминокислот (например, пептид или белок). В некоторых вариантах осуществления настоящего изобретения линкер представляет собой органическую молекулу, группу, полимер или химический фрагмент. В некоторых вариантах осуществления настоящего изобретения длина линкера составляет 5-100 аминокислот, например, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150 или 150-200 аминокислот в длину. Также рассматриваются более длинные или короткие линкеры.

Описанные в настоящее время РНК-направляемые нуклеазы могут содержать по меньшей мере один сигнал ядерной локализации (NLS – nuclear localization signal) для усиления транспорта RGN в ядро клетки. Сигналы ядерной локализации известны в данной области и, как правило, включают последовательность основных аминокислот (см., например, Lange с соавт., *J. Biol. Chem.*, 2007, 282:5101-5105). В некоторых вариантах осуществления настоящего изобретения RGN содержит 2, 3, 4, 5, 6 или более сигналов ядерной локализации. Сигнал (сигналы) ядерной локализации могут быть гетерологичными NLS. Примеры сигналов ядерной локализации, не ограничивающие рамок охвата настоящего изобретения и применимые для описанных в настоящее время RGN являются сигналами ядерной локализации большого Т-антигена SV40, нуклеоплазмينا и с-Мус (см., например, Ray с соавт. *Bioconjug Chem*, 2015, 26(6):1004-7). В определенных вариантах осуществления настоящего изобретения RGN включают последовательности NLS, представленные как SEQ ID NO: 125 или 127. RGN может содержать одну

или несколько последовательностей NLS на N-конце, С-конце или как на N-конце, так и на С-конце. Например, RGN может содержать две последовательности NLS в N-концевой области и четыре последовательности NLS в С-концевой области.

5 Другие последовательности сигналов локализации, известные в данной области, которые локализуют полипептиды в определенных субклеточных местоположениях, также могут быть использованы для нацеливания RGN, включая, но ими не ограничиваясь, последовательности пластидной локализации, последовательности митохондриальной локализации и сигнальные последовательности двойного нацеливания, которые нацелены на RGN. как в 10 пластидах, так и в митохондриях (см., например, Nassoury, Morse, *Biochim Biophys Acta*, 2005, 1743:5-19; Kunze, Berger, *Front Physiol*, 2015, dx.doi.org/10.3389/fphys.2015.00259; Herrmann, Neupert, *IUBMB Life*, 2003, 55:219-225; Soll, *Curr Opin Plant Biol*, 2002, 5:529-535; Carrie, Small, *Biochim Biophys Acta*, 2013, 1833:253-259; Carrie с соавт., *FEBS J*, 2012, 276:1187-1195; 15 Silva-Filho, *Curr Opin Plant Biol*, 2003, 6:589-595; Peeters, Small, *Biochim Biophys Acta*, 2001, 1541:54-63; Murcha с соавт., *J Exp Bot*, 2014, 65:6301-6335; Mackenzie, *Trends Cell Biol*, 2005, 15:548-554; Glaser с соавт., *Plant Mol Biol*, 1998, 38:311-338).

20 В некоторых вариантах осуществления настоящего изобретения известные в настоящее время РНК-направляемые нуклеазы содержат по меньшей мере один проникающий в клетку домен, который способствует поглощению RGN клетками. Проникающие в клетку домены известны в данной области и обычно состоят из участков положительно заряженных аминокислотных остатков (т.е., 25 поликатионные проникающие в клетку домены), чередующихся полярными аминокислотными остатками и неполярными аминокислотными остатками (т.е., амфипатические проникающие в клетку домены), или гидрофобных аминокислотных остатков (т.е., гидрофобные проникающие в клетку домены) (см., например, Milletti F., *Drug Discov Today* 2012, 17:850-860). Примером, не 30 ограничивающим рамки охвата настоящего изобретения, проникающего в клетку домена является трансактивирующий активатор транскрипции (ТАТ) из вируса иммунодефицита человека 1.

Сигнал ядерной локализации, сигнал пластидной локализации, сигнал митохондриальной локализации, сигнал локализации с двойной

направленностью и/или проникающий в клетку домен могут быть расположены на амино-конце (N-конце), карбоксильном конце (C-конце), или во внутренней локализации РНК-направляемой нуклеазы.

Известные в настоящее время RGN могут быть слиты с эффекторным доменом, таким как домен расщепления, домен дезаминазы или домен модулятора экспрессии, прямо или косвенно через линкерный пептид. Такой домен может быть расположен на N-конце, C-конце или во внутренней локализации в РНК-направляемой нуклеазы. В некоторых из этих вариантов осуществления настоящего изобретения компонент RGN слитого белка является мертвой нуклеазой RGN.

В некоторых вариантах осуществления настоящего изобретения слитый белок RGN содержит домен расщепления, который представляет собой любой домен, способный расщеплять полинуклеотид (т.е., РНК, ДНК или гибрид РНК/ДНК), и включает, но ими перечень не ограничивается, эндонуклеазы рестрикции и хоуминг-эндонуклеазы, такие как эндонуклеазы типа II (например, FokI) (см., например, Belfort с соавт., *Nucleic Acids Res*, 1997, 25:3379-3388; кн.: «Nucleases», под ред. Linn с соавт., изд. Cold Spring Harbor Laboratory Press, 1993).

В других вариантах осуществления настоящего изобретения слитый белок RGN содержит домен дезаминазы, который дезаминирует нуклеотидное основание, что приводит к превращению одного нуклеотидного основания в другое, и включает, но не ограничивается ими, редакторы оснований цитидиндезаминазу или аденозиндезаминазу (см., например, Gaudelli с соавт., *Nature* 2017, 551:464-471, US 2017/0121693, US 2018/0073012, US 9840699, WO/2018/027078).

В некоторых вариантах осуществления настоящего изобретения эффекторный домен слитого белка RGN может быть доменом модулятора экспрессии, который является доменом либо для усиления, либо для подавления транскрипции. Домен модулятора экспрессии может представлять собой домен эпигенетической модификации, домен репрессора транскрипции или домен активации транскрипции.

В некоторых подобных вариантах осуществления настоящего изобретения модулятор экспрессии слитого белка RGN содержит домен эпигенетической модификации, который ковалентно модифицирует ДНК или гистонные белки,

изменяя структуру гистонов и/или хромосомную структуру без изменения последовательности ДНК, что приводит к изменениям в экспрессии генов (т.е., повышающая или понижающая регуляция). Неограничительные примеры эпигенетических модификаций включают ацетилирование или метилирование остатков лизина, метилирование аргинина, фосфорилирование серина и треонина, убиквитинирование лизина и сумоилирование гистонных белков, а также метилирование и гидроксиметилирование остатков цитозина в ДНК. Неограничительные примеры доменов эпигенетической модификации включают домены гистон-ацетилтрансферазы, домены гистон-деацетилазы, домены гистон-метилтрансферазы, домены гистон-деметилазы, домены ДНК-метилтрансферазы и домены ДНК-деметилазы.

В других вариантах осуществления настоящего изобретения модулятор экспрессии слитого белка содержит домен репрессора транскрипции, который взаимодействует с элементами контроля транскрипции и/или белками, регулирующими транскрипцию, такими как РНК-полимеразы и факторы транскрипции, для снижения или терминации транскрипции по меньшей мере одного гена. Домены репрессоров транскрипции известны в данной области и включают, но не ограничиваются ими, Sp1-подобные репрессоры, домены IκB и связанную коробку Krüppel (KRAB – Krüppel associated box).

В еще одних вариантах осуществления настоящего изобретения модулятор экспрессии слитого белка содержит домен активации транскрипции, который взаимодействует с элементами контроля транскрипции и/или белками, регулирующими транскрипцию, такими как РНК-полимеразы и факторы транскрипции, для увеличения или активации транскрипции по меньшей мере одного гена. Домены активации транскрипции известны в данной области и включают, но не ограничиваются ими, домен активации вируса простого герпеса VP16 и домен активации NFAT.

Описываемые в настоящем изобретении полипептиды RGN могут содержать обнаруживаемую метку или метку для очистки. Детектируемая метка или метка для очистки может быть расположена на N-конце, C-конце или внутри РНК-направляемой нуклеазы и соединена либо прямо, либо косвенно через линкерный пептид. В некоторых из этих вариантов осуществления настоящего изобретения компонент RGN слитого белка представляет собой мертвую

нуклеазу RGN. В других вариантах осуществления компонент RGN слитого белка представляет собой RGN с никазной активностью.

5 Детектируемой меткой является молекула, с помощью которой можно визуализировать или наблюдать иным образом. Детектируемая метка может быть слита с RGN в виде слитого белка (например, флуоресцентного белка) или может представлять собой небольшую молекулу, конъюгированную с полипептидом RGN, которую можно обнаружить визуальными или другими способами. Детектируемые метки, которые могут быть слиты с известными в настоящее время RGN в виде слитого белка, включают любой детектируемый белковый домен, включая, но не ограничиваясь им, флуоресцентный белок или белковый домен, который можно обнаружить с помощью специфического антитела. Неограничительные примеры флуоресцентных белков включают зеленые флуоресцентные белки (например, GFP, EGFP, ZsGreen1) и желтые флуоресцентные белки (например, YFP, EYFP, ZsYellow1). Неограничительные
10
15
15 примеры меток, представляющих малые молекулы, включают радиоактивные метки, такие как ^3H и ^{35}S .

Полипептиды RGN также могут содержать метку очистки, которая представляет собой любую молекулу, которую можно использовать для выделения белка или слитого белка из смеси (например, биологического образца, культуральной среды). Неограничительные примеры меток очистки включают биотин, мус, белок связывания мальтозы (MBP – maltose binding protein) и глутатион-S-трансферазу (GST – glutathione-S-transferase).

II. Направляющая РНК

Настоящее изобретение предусматривает направляющие РНК и кодирующие их полинуклеотиды. Термин «направляющая РНК» относится к нуклеотидной последовательности, имеющей достаточную комплементарность с целевой нуклеотидной последовательностью для гибридизации с целевой последовательностью и прямого связывания ассоциированной РНК-направляемой нуклеазы с целевой нуклеотидной последовательностью. Таким образом, соответствующая направляющая РНК RGN представляет собой одну или несколько молекул РНК (как правило, одну или две), которые могут связываться с RGN и направлять RGN для связывания с определенной последовательностью нуклеотидов-мишеней, и в тех случаях, когда RGN обладает никазной или нуклеазной активностью, также расщепляют целевую
25
30

нуклеотидную последовательность. Обычно направляющая РНК содержит РНК CRISPR (сгРНК), а в некоторых вариантах осуществления настоящего изобретения – транс-активирующую РНК CRISPR (tracrРНК). Нативные направляющие РНК, которые содержат как сгРНК, так и tracrРНК, обычно содержат две отдельные молекулы РНК, которые гибридизируются друг с другом посредством повторяющейся последовательности сгРНК и антиповторяющейся последовательности tracrРНК.

Нативные прямые повторяющиеся последовательности в массиве CRISPR обычно имеют длину от 28 до 37 пар оснований, хотя длина может варьировать от примерно 23 до примерно 55 п.н. Спейсерные последовательности в массиве CRISPR обычно имеют длину от примерно 32 до примерно 38 п.н., хотя длина может составлять от примерно 21 до примерно 72 п.н. Каждый массив CRISPR обычно содержит менее 50 единиц последовательности повторов-спейсеров CRISPR. CRISPR транскрибируются как часть длинного транскрипта, называемого первичным транскриптом CRISPR, который включает большую часть массива CRISPR. Первичный транскрипт CRISPR расщепляется белками Cas с образованием сгРНК или, в некоторых случаях, с образованием пре-сгРНК, которые далее процессируются дополнительными белками Cas в зрелые сгРНК. Зрелые сгРНК содержат спейсерную последовательность и повторяющуюся последовательность CRISPR. В некоторых вариантах осуществления настоящего изобретения, в которых пре-сгРНК процессируются в зрелые (или процессированные) сгРНК, созревание включает удаление примерно от одного до примерно шести или более 5', 3' или 5' и 3' нуклеотидов. В целях редактирования генома или нацеливания на конкретную представляющую интерес последовательность нуклеотидов-мишеней эти нуклеотиды, которые удаляются во время созревания молекулы пре-сгРНК, не являются необходимыми для создания или конструирования направляющей РНК.

РНК CRISPR (сгРНК) содержит спейсерную последовательность и повторяющуюся последовательность CRISPR. «Спейсерная последовательность» представляет собой нуклеотидную последовательность, которая непосредственно гибридизируется с представляющей интерес последовательностью-мишенью. Спейсерную последовательность конструируют таким образом, чтобы она была полностью или частично комплементарной представляющей интерес последовательности-мишени. В различных вариантах

осуществления настоящего изобретения спейсерная последовательность может содержать от примерно 8 нуклеотидов до примерно 30 нуклеотидов или более. Например, спейсерная последовательность может составлять примерно 8, примерно 9, примерно 10, примерно 11, примерно 12, примерно 13, примерно 14, 5 примерно 15, примерно 16, примерно 17, примерно 18, примерно 19, примерно 20, примерно 21, примерно 22, примерно 23, примерно 24, примерно 25, примерно 26, примерно 27, примерно 28, примерно 29, примерно 30 или более нуклеотидов в длину. В некоторых вариантах осуществления настоящего изобретения спейсерная последовательность может содержать от примерно 10 до 10 примерно 26 нуклеотидов в длину, или примерно от 12 до примерно 30 нуклеотидов в длину. В особенно предпочтительных вариантах осуществления настоящего изобретения длина спейсерной последовательности составляет примерно 30 нуклеотидов. В некоторых вариантах осуществления настоящего изобретения степень комплементарности между спейсерной 15 последовательностью и соответствующей последовательностью-мишенью при оптимальном выравнивании с использованием подходящего алгоритма выравнивания составляет примерно 50%, примерно 60%, примерно 70%, примерно 75%, примерно 80%, примерно 81 %, примерно 82%, примерно 83%, примерно 84%, примерно 85%, примерно 86%, примерно 87%, примерно 88%, 20 примерно 89%, примерно 90%, примерно 91%, примерно 92%, примерно 93%, примерно 94%, примерно 95%, примерно 96%, примерно 97%, примерно 98%, примерно 99% или более. В особенно предпочтительных вариантах осуществления настоящего изобретения спейсерная последовательность не имеет вторичной структуры, которую можно предсказать с использованием 25 любого подходящего алгоритма сворачивания полинуклеотидов, известного в данной области, включая, например, помимо прочего, mFold (см., например, Zuker, Stiegler, *Nucleic Acids Res.*, 1981, 9:133-148) и PНКfold (см., например, Gruber с соавт., *Cell*, 2008, 106(1):23-24).

Повторяющаяся последовательность ПНК CRISPR содержит нуклеотидную 30 последовательность, которая включает область с достаточной комплементарностью для гибридизации с tracrПНК. В различных вариантах осуществления настоящего изобретения последовательность повторов CRISPR ПНК может содержать от примерно 8 нуклеотидов до примерно 30 нуклеотидов или более. Например, повторяющаяся последовательность ПНК CRISPR может

содержать примерно 9, примерно 10, примерно 11, примерно 12, примерно 13, примерно 14, примерно 15, примерно 16, примерно 17, примерно 18, примерно 19, примерно 20, примерно 21, примерно 22, примерно 23, примерно 24, примерно 25, примерно 26, примерно 27, примерно 28, примерно 29, примерно 30 или более нуклеотидов в длину. В некоторых вариантах осуществления настоящего изобретения повторяющаяся последовательность CRISPR может содержать примерно 21 нуклеотид в длину. В некоторых вариантах осуществления настоящего изобретения степень комплементарности между повторяющейся последовательностью CRISPR и соответствующей последовательностью tracrPНК при оптимальном выравнивании с использованием подходящего алгоритма выравнивания составляет примерно 50%, примерно 60%, примерно 70%, примерно 75%, примерно 80%, примерно 81%, примерно 82%, примерно 83%, примерно 84%, примерно 85%, примерно 86%, примерно 87%, примерно 88%, примерно 89%, примерно 90%, примерно 91%, примерно 92%, примерно 93%, примерно 94%, примерно 95%, примерно 96%, примерно 97%, примерно 98%, примерно 99% или более. В некоторых вариантах осуществления настоящего изобретения повторяющаяся последовательность CRISPR содержит нуклеотидную последовательность SEQ ID NO: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 118, 240, 273 или 287, или ее активный вариант или фрагмент, который при включении в направляющую РНК способен направлять специфичное к последовательности связывание ассоциированной РНК-направляемой нуклеазы, предусмотренной в настоящем изобретении, с представляющей интерес последовательностью-мишенью. В некоторых вариантах осуществления настоящего изобретения активный вариант повторяющейся последовательности CRISPR последовательности дикого типа содержит нуклеотидную последовательность, которая по меньшей мере на 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или более идентична последовательности нуклеотидной последовательности, представленной как SEQ ID NO: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 118, 240, 273 или 287. В некоторых вариантах осуществления настоящего изобретения активный фрагмент повторяющейся последовательности CRISPR последовательности дикого типа содержит по меньшей мере 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 или 20

последовательных нуклеотидов нуклеотидной последовательности, представленной как SEQ ID NO: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 118, 240, 273 или 287.

В некоторых вариантах осуществления настоящего изобретения crPНК не является природной. В некоторых из этих вариантов осуществления настоящего изобретения специфическая повторяющаяся последовательность CRISPR не связана со сконструированной последовательностью спейсера в природе, и повторяющаяся последовательность CRISPR рассматривают как гетерологичную последовательность в отношении последовательности спейсера. В некоторых вариантах осуществления настоящего изобретения спейсерная последовательность является сконструированной последовательностью, которая не существует в природе.

Транс-активирующая молекула PНК CRISPR или tracrPНК содержит нуклеотидную последовательность, включающую область, которая имеет достаточную комплементарность для гибридизации с повторяющейся последовательностью CRISPR в crPНК, которая упоминается в настоящем описании как область антиповтора. В некоторых вариантах осуществления настоящего изобретения молекула tracrPНК дополнительно содержит область, имеющую вторичную структуру (например, стебель-петля) или формы вторичной структуры при гибридизации с соответствующей ей crPНК. В определенных вариантах осуществления настоящего изобретения область tracrPНК, полностью или частично комплементарная повторяющейся последовательности CRISPR, находится на 5'-конце молекулы, а 3'-конец tracrPНК содержит вторичную структуру. Эта область вторичной структуры обычно включает несколько шпилек, в том числе некусную шпильку, которая находится рядом с анти-повторяющейся последовательностью. Некусная шпилька часто имеет консервативную последовательность нуклеотидов в основании стержня шпильки с мотивом UNANNA, CNANNG, CNANNU, UNANNG, UNANNC или CNANNU (SEQ ID NO: 8, 37, 45, 53, 68 и 102, соответственно), обнаруженную во многих некусных шпильках в tracrPНК. Часто на 3'-конце tracrPНК имеются концевые шпильки, которые могут различаться по структуре и количеству, но часто содержат Rho-независимую шпильку терминатора транскрипции с большим содержанием GC, за которой следует отрезок U на 3'-конце. См., например, Briner с соавт., *Molecular Cell*

2014, 56:333-339, Briner, Barrangou *Cold Spring Harb Protoc*, 2016, doi: 10.1101/pdb.top090902, и US 2017/0275648, сущность каждой из которых включена в настоящее описание в виде ссылки.

В различных вариантах осуществления настоящего изобретения область анти-повторов tracrPНК, которая полностью или частично комплементарна последовательности повторов CRISPR, содержит от примерно 8 нуклеотидов до примерно 30 нуклеотидов или более. Например, область спаривания оснований между последовательностью анти-повторов tracrPНК и последовательностью повторов CRISPR может составлять примерно 8, примерно 9, примерно 10, примерно 11, примерно 12, примерно 13, примерно 14, примерно 15, примерно 16, примерно 17. , примерно 18, примерно 19, примерно 20, примерно 21, примерно 22, примерно 23, примерно 24, примерно 25, примерно 26, примерно 27, примерно 28, примерно 29, примерно 30 или более нуклеотидов в длину. В определенных вариантах осуществления настоящего изобретения область анти-повторов tracrPНК, которая полностью или частично комплементарна последовательности повторов CRISPR, имеет в длину около 20 нуклеотидов. В некоторых вариантах осуществления настоящего изобретения степень комплементарности между повторяющейся последовательностью CRISPR и соответствующей последовательностью анти-повторов tracrPНК при оптимальном выравнивании с использованием подходящего алгоритма выравнивания составляет примерно 50%, примерно 60%, примерно 70%, примерно 75%, примерно 80 %, примерно 81%, примерно 82%, примерно 83%, примерно 84%, примерно 85%, примерно 86%, примерно 87%, примерно 88%, примерно 89%, примерно 90%, примерно 91%, примерно 92%, примерно 93%, примерно 94%, примерно 95%, примерно 96%, примерно 97%, примерно 98%, примерно 99% или более.

В различных вариантах осуществления настоящего изобретения вся tracrPНК может содержать примерно от 60 нуклеотидов до примерно более чем 140 нуклеотидов в длину. Например, tracrPНК может включать примерно 60, примерно 65, примерно 70, примерно 75, примерно 80, примерно 85, примерно 90, примерно 95, примерно 100, примерно 105, примерно 110, примерно 115, примерно 120, примерно 125, примерно 130, примерно 135, примерно 140 или более нуклеотидов в длину. В некоторых вариантах осуществления настоящего изобретения tracrPНК содержит примерно от 80 до примерно 90 нуклеотидов,

включая примерно 80, примерно 81, примерно 82, примерно 83, примерно 84, примерно 85, примерно 86, примерно 87, примерно 88, примерно 89 и примерно 90 нуклеотидов в длину. В определенных вариантах осуществления настоящего изобретения tracrРНК содержит примерно 85 нуклеотидов в длину.

5 В определенных вариантах осуществления настоящего изобретения tracrРНК включает нуклеотидную последовательность SEQ ID NO: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112 или 119, или ее активный вариант или фрагмент, который при включении в направляющую РНК способен направлять специфичное к последовательности связывание ассоциированной РНК-направляемой нуклеазы, предусмотренной в настоящем изобретении, с представляющей интерес последовательностью-мишенью. В некоторых вариантах осуществления настоящего изобретения активный вариант последовательности tracrРНК дикого типа содержит нуклеотидную последовательность, которая по меньшей мере на 40%, 45%, 50%, 55%, 60%, 15 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или более процентов идентична нуклеотидной последовательности, представленной как SEQ ID NO: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112, 119, 241, 274 или 286. В некоторых вариантах осуществления настоящего изобретения активный фрагмент последовательности tracrРНК последовательности дикого типа содержит по меньшей мере 5, 10, 15, 20, 25, 30, 20 35, 40, 45, 50, 55, 60, 65, 70, 75, 80 или более непрерывно расположенных нуклеотидов в нуклеотидной последовательности, представленной как SEQ ID NO: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112, 119, 241, 274 или 286.

25 Две полинуклеотидные последовательности можно считать по существу комплементарными, когда две последовательности гибридизируются друг с другом в жестких условиях. Аналогичным образом полагают, что RGN связывается с конкретной последовательностью-мишенью специфичным для последовательности образом, если направляющая РНК, связанная с RGN, 30 связывается с последовательностью-мишенью в жестких условиях. Под «жесткими условиями» или «жесткими условиями гибридизации» подразумевают условия, при которых две полинуклеотидные последовательности будут гибридизоваться друг с другом в заметно большей степени, чем с другими последовательностями (например, по меньшей мере в 2

раза по сравнению с фоном). Жесткие условия зависят от последовательности и будут разными в разных обстоятельствах. Обычно жесткими условиями будут те, в которых концентрация соли составляет менее примерно 1,5 М иона Na, обычно примерно от 0,01 до 1,0 М концентрации иона Na (или других солей) при рН от 7,0 до 8,3, а температура составляет по меньшей мере примерно 30°C для коротких последовательностей (например, от 10 до 50 нуклеотидов) и по меньшей мере около 60°C для длинных последовательностей (например, более 50 нуклеотидов). Жесткие условия также могут быть достигнуты при добавлении дестабилизирующих агентов, таких как формамид. Примеры условий низкой жесткости включают гибридизацию с буферным раствором 30-35% формамида, 1 М NaCl, 1% SDS (додецилсульфата натрия) при 37°C и промывку в 1X-2X SSC (20X SSC = 3,0 М NaCl/0,3 М тринатрия цитрата) при температуре от 50 до 55°C. Типичные условия умеренной строгости включают гибридизацию в 40-45% формамиде, 1,0 М NaCl, 1% SDS при 37°C и промывку в 0,5X-1X SSC при 55-60°C. Типичные условия высокой жесткости включают гибридизацию в 50% формамиде, 1 М NaCl, 1% SDS при 37°C и промывку в 0,1X SSC при 60-65°C. Необязательно, промывочные буферы могут содержать от примерно 0,1% до примерно 1% SDS. Продолжительность гибридизации обычно составляет менее примерно 24 ч, обычно от примерно 4 до примерно 12 ч. Продолжительность времени промывки будет по меньшей мере достаточной для достижения равновесия.

T_m представляет собой температуру (при определенной ионной силе и рН), при которой 50% комплементарной последовательности-мишени гибридизируется с идеально подобранной последовательностью. Для гибридов ДНК-ДНК T_m можно приблизительно понять из уравнения Meinkoth и Wahl, *Anal. Biochem.* 1984, 138:267-284: $T_m = 81,5^\circ\text{C} + 16,6 (\log M) + 0,41 (\%GC) - 0,61 (\% \text{ формамида}) - 500/L$; где M – молярность одновалентных катионов, %GC – процентное содержание гуанозиновых и цитозиновых нуклеотидов в ДНК, % формамида — процентное содержание формамида в растворе для гибридизации, а L – длина гибрида в парах оснований. Обычно жесткие условия выбирают таким образом, чтобы они были примерно на 5°C ниже температуры плавления (T_m) для конкретной последовательности и ее дополнения при определенной ионной силе и рН. Однако в очень жестких условиях можно использовать гибридизацию и/или промывку при температуре на 1, 2, 3 или 4°C ниже

температуры плавления (T_m); в умеренно жестких условиях можно использовать гибридизацию и/или промывку при температуре на 6, 7, 8, 9 или 10°C ниже температуры плавления (T_m); в условиях низкой жесткости можно использовать гибридизацию и/или промывку при температуре на 11, 12, 13, 14, 15 или 20°C ниже температуры плавления (T_m). Используя уравнение, композиции для гибридизации и промывки, а также желаемую T_m , специалисты с обычной квалификацией поймут, что вариации жесткости растворов для гибридизации и/или промывки по существу описаны. Подробное руководство по гибридизации нуклеиновых кислот можно найти в кн.: Tijssen «Laboratory Techniques in Biochemistry and Molecular Biology – Hybridization with Nucleic Acid Probes», 1993, часть I, глава 2 (Elsevier, Нью-Йорк); кн.: «Current Protocols in Molecular Biology», под ред. Ausubel с соавт., 1995, глава 2 (изд. Greene Publishing and Wiley-Interscience, Нью-Йорк). См. кн.: Sambrook с соавт. «Molecular Cloning: A Laboratory Manual» (1989, 2е изд., изд-во Cold Spring Harbor Laboratory Press, Plainview, Нью-Йорк).

Направляющая РНК может представлять систему однонаправляющей РНК или из двунаправляющей РНК. Однонаправляющая РНК содержит $sgRNA$ и $tracrRNA$ в одной молекуле РНК, тогда как система двунаправляющей РНК включает $sgRNA$ и $tracrRNA$, присутствующие в двух разных молекулах РНК, гибридизированных друг с другом по крайней мере через часть последовательности повторов CRISPR в $sgRNA$ и по меньшей мере часть $tracrRNA$, которые могут быть полностью или частично комплементарны повторяющейся последовательности CRISPR в $sgRNA$. В некоторых из этих вариантов осуществления настоящего изобретения, где направляющая РНК представляет собой однонаправляющую РНК, $sgRNA$ и $tracrRNA$ разделены линкерной нуклеотидной последовательностью. Как правило, линкерная нуклеотидная последовательность представляет собой последовательность, которая не включает комплементарные основания, чтобы избежать образования вторичной структуры внутри или включения нуклеотидов линкерной нуклеотидной последовательности. В некоторых вариантах осуществления настоящего изобретения линкерная нуклеотидная последовательность между $sgRNA$ и $tracrRNA$ имеет по меньшей мере 3, по меньшей мере 4, по меньшей мере 5, по меньшей мере 6, по меньшей мере 7, по меньшей мере 8, по меньшей мере 9, по меньшей мере 10, по меньшей мере 11, по меньшей мере 12 или более

нуклеотидов в длину. В определенных вариантах осуществления настоящего изобретения линкерная нуклеотидная последовательность однонаправляющей РНК включает по меньшей мере 4 нуклеотида в длину. В некоторых вариантах осуществления настоящего изобретения линкерная нуклеотидная последовательность представлена как SEQ ID NO: 123.

Однонаправляющая РНК или двунаправляющая РНК может быть синтезирована химически или путем транскрипции *in vitro*. Анализы для определения специфичного по последовательности связывания между RGN и направляющей РНК известны в данной области и включают, но не ограничиваются ими, анализы связывания *in vitro* между экспрессированной RGN и направляющей РНК, которые могут быть помечены обнаруживаемой меткой (например, биотином), используемой в анализе с осаждением, в котором улавливают комплекс направляющей РНК:RGN по обнаруживаемой метке (например, гранулами со стрептавидином). Контрольная направляющая РНК с последовательностью или структурой, не имеющей отношения к направляющей РНК, может использоваться в качестве отрицательного контроля для неспецифического связывания RGN с РНК. В некоторых вариантах осуществления настоящего изобретения направляющая РНК представлена как SEQ ID NO: 4, 12, 19, 26, 33, 41, 49, 57, 64, 72, 78, 85, 92, 98, 106, 113 или 120, где спейсерная последовательность может представлять собой какую-либо последовательность и обозначается как поли-N-последовательность.

В других вариантах осуществления настоящего изобретения направляющая РНК может быть введена в клетку-мишень, органеллу или эмбрион в виде молекулы РНК. Направляющая РНК может быть транскрибирована *in vitro* или синтезирована химически. В других вариантах осуществления настоящего изобретения нуклеотидная последовательность, кодирующая направляющую РНК, вводится в клетку, органеллу или эмбрион. В некоторых из этих вариантов осуществления настоящего изобретения нуклеотидная последовательность, кодирующая направляющую РНК, функционально связана с промотором (например, промотором РНК-полимеразы III). Промотор может быть нативным промотором или гетерологичным по отношению к нуклеотидной последовательности, кодирующей направляющую РНК.

В различных вариантах осуществления настоящего изобретения направляющая РНК может быть введена в клетку-мишень, органеллу или

эмбрион в виде рибонуклеопротеинового комплекса, как описано в настоящем изобретении, где направляющая РНК связана с полипептидом РНК-направляемой нуклеазы.

Направляющая РНК направляет ассоциированную РНК-направляемую нуклеазу к конкретной представляющей интерес нуклеотидной последовательности-мишени посредством гибридизации направляющей РНК с целевой нуклеотидной последовательностью. Нуклеотидная последовательность-мишень может содержать ДНК, РНК или их комбинацию и может быть одноцепочечной или двухцепочечной. Нуклеотидной последовательностью-мишенью может быть геномная ДНК (т.е., хромосомная ДНК), плазмидная ДНК или молекула РНК (например, матричная РНК, рибосомальная РНК, транспортная РНК, микроРНК, малая интерферирующая РНК). Нуклеотидная последовательность-мишень может быть связана (и в некоторых вариантах осуществления настоящего изобретения расщеплена) с помощью РНК-направляемой нуклеазы *in vitro* или в клетке. Хромосомальная последовательность, на которую нацелен RGN, может быть ядерной, плазмидной или митохондриальной хромосомальной последовательностью. В некоторых вариантах осуществления настоящего изобретения последовательность нуклеотидов-мишеней уникальна в целевом геноме.

Нуклеотидная последовательность-мишень примыкает к мотиву, примыкающему к протоспейсеру (РАМ – protospacer adjacent motif). Смежный мотив протоспейсера обычно находится в пределах от примерно 1 до примерно 10 нуклеотидов от последовательности нуклеотидов-мишеней, включая примерно 1, примерно 2, примерно 3, примерно 4, примерно 5, примерно 6, примерно 7, примерно 8, примерно 9 или примерно 10 нуклеотидов от целевой нуклеотидной последовательности. РАМ может быть с 5'- или 3'-конца целевой последовательности. В некоторых вариантах осуществления настоящего изобретения РАМ представляет собой 3'-конец целевой последовательности для известных RGN. Обычно РАМ является консенсусной последовательностью примерно из 3-4 нуклеотидов, но в определенных вариантах осуществления настоящего изобретения может иметь в длину 2, 3, 4, 5, 6, 7, 8, 9 или более нуклеотидов. В различных вариантах осуществления настоящего изобретения последовательность РАМ, распознаваемая описываемыми RGN, содержит

консенсусную последовательность, представленную как SEQ ID NO: 7, 15, 22, 29, 36, 44, 52, 60, 67, 81, 88, 101, 109 или 116.

В определенных вариантах осуществления настоящего изобретения РНК-направляемая нуклеаза, имеющая последовательность SEQ ID NO: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110 или 117, или ее активный вариант или фрагмент связывает соответственно целевую нуклеотидную последовательность, смежную с последовательностью PAM, представленную как SEQ ID NO: 7, 15, 22, 29, 36, 44, 52, 60, 67, 81, 88, 101, 109 или 116. В некоторых вариантах осуществления настоящего изобретения РНК-направляемая нуклеаза, имеющая SEQ ID NO: 54 или 137, или ее активный вариант или фрагмент, связывает целевую нуклеотидную последовательность, смежную с последовательностью PAM, представленную как SEQ ID NO: 147. В некоторых подобных вариантах осуществления настоящего изобретения RGN связывается с направляющей последовательностью, включающей повторяющуюся последовательность CRISPR, представленную как SEQ ID NO: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111 или 118, соответственно, или ее активным вариантом или фрагментом, и последовательность *tracrP*НК, представленную как SEQ ID NO: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112 или 119, соответственно, или ее активный вариант или фрагмент.

Системы RGN, описанные ниже в примере 1 и табл. 1 и 2 настоящего описания.

В данной области техники известно, что специфичность последовательности PAM для данного фермента нуклеазы зависит от концентрации фермента (см., например, Karvelis с соавт., *Genome Biol*, 2015, 16:253), и ее можно модифицировать путем изменения промотора, используемого для экспрессии RGN, или количества рибонуклеопротеинового комплекса, доставляемого в клетку, органеллу или эмбрион.

При распознавании соответствующей последовательности PAM, RGN может расщеплять целевую нуклеотидную последовательность в определенном сайте расщепления. В настоящем изобретении сайт расщепления состоит из двух конкретных нуклеотидов в целевой нуклеотидной последовательности, между которыми нуклеотидная последовательность расщепляется с помощью RGN. Сайт расщепления может содержать 1^й и 2^й, 2^й и 3^й, 3^й и 4^й, 4^й и 5^й, 5^й и 6^й, 7^й и 8^й или 8^й и 9^й нуклеотид из PAM в любом направлении, 5' или 3'. В некоторых вариантах осуществления настоящего изобретения сайт расщепления может

находиться на расстоянии более 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 или 20 нуклеотидов от PAM либо в 5'-, либо в 3'-направлении. В некоторых вариантах осуществления настоящего изобретения сайт расщепления может находиться на расстоянии 4 нуклеотидов от PAM. В других вариантах осуществления настоящего изобретения сайт расщепления находится на расстоянии 15 нуклеотидов от PAM. Поскольку RGN могут расщеплять целевую нуклеотидную последовательность, что приводит к образованию ступенчатых концов, в некоторых вариантах осуществления настоящего изобретения сайт расщепления определяют на основании расстояния двух нуклеотидов от PAM на положительной (+) цепи полинуклеотида и расстояния двух нуклеотидов от PAM на отрицательной (-) цепи полинуклеотида.

PAM рассматривают как отличительную черту РНК-направляемых нуклеаз систем CRISPR типа II (Szczelkun с соавт., *PNAS*, 2014, 111: 9798-9803; Sternberg с соавт., 2014, *Nature* 507: 62-67). Следует отметить, что хотя APG06646.1 и APG04293.1 функционируют как РНК-направляемые нуклеазы и обладают многими из тех же доменов, что и нуклеазы CRISPR Cas9 типа II, каждая из них не имеет типичного PAM-взаимодействующего домена ((PID – PAM-Interacting domain); база данных InterPro: IPR032237; база данных Pfam: PF16595). Соответственно, APG06646.1 и APG04293.1 также не обладают характерной потребностью в мотиве PAM, которая представляет собой последовательность из 2-5 нуклеотидов согласно описанному выше. Вместо этого эти белки имеют уникальные домены распознавания ДНК на своих С-концах (остатки 821-1092 в APG06646.1 (последовательность полной длины представлена как SEQ ID NO: 117); остатки 1064-1401 в APG04293.1 (последовательность полной длины представлена как SEQ ID NO: 103)). Такие уникальные домены распознавания ДНК позволяют нуклеазам расщеплять по геномному сайту-мишени на основе однонуклеотидного мотива вблизи геномной последовательности-мишени (SEQ ID NO: 109; см. табл. 2).

APG04293.1 также обладает уникальным сигнатурным доменом из 133 аминокислотных остатков, расположенным ближе к N-концу (остатки 144-276). Функция этого домена неизвестна не только у нуклеаз CRISPR Cas9 типа II, но и вообще в данной области техники.

III. Нуклеотиды, кодирующие РНК-направляемые нуклеазы, CRISPR РНК и/или tracrРНК

В настоящем изобретении предложены полинуклеотиды, содержащие известные в настоящее время РНК CRISPR, tracrРНК и/или sgРНК, и
5 полинуклеотиды, содержащие нуклеотидную последовательность, кодирующую установленные в настоящее время РНК-направляемые нуклеазы, CRISPR РНК, tracrРНКs и/или sgРНК. Известные полинуклеотиды включают полинуклеотиды, содержащие или кодирующие повторяющуюся последовательность CRISPR, содержащую нуклеотидную последовательность SEQ ID NO: 2, 10, 17, 24, 31, 39,
10 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 118, 240, 273 или 287 или их активный вариант или фрагмент, которые будучи включенными в направляющую РНК способны направлять специфическое в отношении последовательности связывание ассоциированной РНК-направляемой нуклеазы с представляющей интерес последовательностью-мишенью. Все описанные полинуклеотиды
15 являются включающими или кодирующими tracrРНК, содержащими нуклеотидные последовательности SEQ ID NO: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112, 119, 241, 274 или 286 или их активный вариант или фрагмент, которые при включении в направляющую РНК способны направлять специфическое по последовательности связывание РНК-направляющей нуклеазы
20 с целевой последовательностью, представляющей интерес. Также предусмотрены полинуклеотиды, которые кодируют РНК-направляемую нуклеазу, содержащую аминокислотную последовательность, представленную как SEQ ID NO: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 117, 137 или 235, а также их активные фрагменты или варианты, которые сохраняют
25 способность связываться с целевой нуклеотидной последовательностью РНК-направляемым специфическим в отношении последовательности образом.

Использование термина «полинуклеотид» не предназначено для ограничения настоящего изобретения полинуклеотидами, представляющими ДНК. Специалистам в данной области понятно, что полинуклеотиды могут
30 включать рибонуклеотиды (РНК) и комбинации рибонуклеотидов и дезоксирибонуклеотидов. Такие дезоксирибонуклеотиды и рибонуклеотиды включают как встречающиеся в природе молекулы, так и синтетические аналоги. К ним относятся пептидные нуклеиновые кислоты (ПНК или PNA – peptide nucleic acid), химеры ПНК-ДНК, заблокированные нуклеиновые кислоты (ЗНК

или LNA – locked nucleic acid) и последовательности, связанные с фосфотиоратом. Полинуклеотиды, описанные в настоящем изобретении, также охватывают все формы последовательностей, включая, помимо прочего, одноцепочечные формы, двухцепочечные формы, гибриды ДНК-РНК, триплексные структуры, структуры типа «стебель-петля» и т.п.

Молекулы нуклеиновых кислот, кодирующие RGN, могут быть оптимизированы по кодонам для экспрессии в представляющем интерес организме. «Оптимизированная по кодонам» кодирующая последовательность представляет собой полинуклеотидную кодирующую последовательность, для которой сконструирована частота изменения кодонов, имитирующая частоту использования кодонов, предназначенную для имитации частоты использования предпочтительных кодонов или условий транскрипции в конкретной клетке-хозяине. Повышается экспрессия в конкретной клетке-хозяине или организме в результате изменения одного или нескольких кодонов на уровне нуклеиновой кислоты таким образом, что транслируемая аминокислотная последовательность не изменяется. Молекулы нуклеиновых кислот могут быть оптимизированы полностью или частично. В данной области есть таблицы кодонов и другие ссылки, содержащие информацию о предпочтениях для широкого круга организмов (см., например, Campbell, Gowri, *Plant Physiol.*, 1990, 92:1-11, где приводят обсуждение применения кодонов, предпочтительных для растений). В данной области техники известны методы синтеза генов, предпочтительных для растений. См., например, US 5380831, US 5436391, Murray с соавт., *Nucleic Acids Res.*, 1989, 17:477-498, включенные в настоящее изобретение в виде ссылки.

Полинуклеотиды, кодирующие RGN, crРНК, tracrРНКs и/или sgРНК, предусмотренные в настоящем документе, могут быть в экспрессионных кассетах для экспрессии *in vitro* или экспрессии в представляющих интерес клетке, органелле, эмбрионе или организме. Кассета может включать 5'- и 3'-регуляторные последовательности, функционально связанные с полинуклеотидом, кодирующим RGN, crРНК, tracrРНКs и/или sgРНК, предусмотренные в настоящем изобретении, что обеспечивает экспрессию полинуклеотида. Кассета может дополнительно содержать по меньшей мере один дополнительный ген или генетический элемент, одновременно трансформируемый в организме. Если включены дополнительные гены или элементы, компоненты функционально связаны. Термин «функционально

связаны» предназначен для обозначения функциональной связи между двумя или более элементами. Например, действующая связь между промотором и представляющей интерес кодирующей областью (например, областью, кодирующей RGN, crPНК, tracrPНК и/или sgPНК) является функциональной связью, которая позволяет экспрессировать представляющую интерес кодирующую область. Функционально связанные элементы могут быть смежными или находиться дистанционно. При использовании для обозначения соединения двух областей, кодирующих белок, термин «функционально связанные» подразумевает, что кодирующие области находятся в одной и той же рамке считывания. В другом варианте, дополнительный ген (гены) или элемент (элементы) могут быть представлены на нескольких экспрессионных кассетах. Например, последовательность нуклеотидов, кодирующая известные в настоящее время RGN, может находиться на одной кассете экспрессии, тогда как последовательность нуклеотидов, кодирующая crPНК, tracrPНК или всю направляющую PНК, может находиться на другой кассете экспрессии. Такая экспрессионная кассета снабжена множеством сайтов рестрикции и/или сайтов рекомбинации для инсерции полинуклеотидов, подлежащих транскрипционной регуляции регуляторных областей. Кассета экспрессии может дополнительно содержать селективируемый маркерный ген.

Кассета экспрессии может включать в направлении транскрипции 5'-3' область инициации транскрипции (и, в некоторых вариантах осуществления настоящего изобретения, трансляции), т.е. промотор, RGN-, crPНК-, tracrPНК- и/или sgPНК-кодирующие полипептид по настоящему изобретению, а также область терминации транскрипции (и, в некоторых вариантах осуществления настоящего изобретения, трансляции), т.е. область терминации, функционирующие в представляющем интересе организме. Промоторы по настоящему изобретению способны направлять или стимулировать экспрессию кодирующей последовательности в клетке-хозяине. Регуляторные области (например, промоторы, области регуляции транскрипции и области терминации трансляции) могут быть эндогенными или гетерологичными по отношению к клетке-хозяину или друг к другу. Используемый в настоящем изобретении термин «гетерологичный» в отношении последовательности представляет собой последовательность, которая происходит от другого вида или, если она принадлежит к тому же виду, существенно изменена по сравнению с ее нативной

формой по составу и/или геномному локусу в результате преднамеренного вмешательства человека. В настоящем изобретении химерный ген содержит кодирующую последовательность, функционально связанную с областью инициации транскрипции, которая гетерологична в отношении кодирующей последовательности.

Подходящие области терминации можно получить из T_i-плазмиды из *A. tumefaciens*, такие как области терминации октопинсинтазы и нопалинсинтазы. См. также Guerineau с соавт., *Mol. Gen. Genet*, 1991, 262:141-144; Proudfoot, *Cell*, 1991, 64:671-674; Sanfacon с соавт., *Genes Dev*, 1991, 5:141-149; Mogen с соавт., *Plant Cell*, 1990, 2:1261-1272; Munroe с соавт., *Gene*, 1990, 91:151-158; Ballas с соавт., *Nucleic Acids Res*, 1989, 17:7891-7903; Joshi с соавт., *Nucleic Acids Res*, 1987, 15:9627-9639.

К дополнительным регуляторным сигналам относятся, но ими не ограничиваются, стартовые сайты инициации транскрипции, операторы, активаторы, энхансеры, другие регуляторные элементы, рибосомальные сайты связывания, кодоны инициации, сигналы терминации и другие. См., например, US 5039523 и US 4853331; EPO 0480762A2; Sambrook с соавт., кн.: «Molecular Cloning: A Laboratory Manual», 1992, под ред. Maniatis с соавт. (изд. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, Нью-Йорк), в дальнейшем ссылка дается как «Sambrook 11»; кн.: «Advanced Bacterial Genetics» под ред. Davis с соавт., 1980 (изд. Cold Spring Harbor Laboratory Press), Cold Spring Harbor, Нью-Йорк, и приведенные в них ссылки.

При приготовлении кассеты экспрессии можно манипулировать различными фрагментами ДНК, чтобы обеспечить последовательности ДНК в надлежащей ориентации и, при необходимости, в надлежащей рамке считывания. С этой целью для соединения фрагментов ДНК могут быть использованы адаптеры или линкеры, или могут быть задействованы другие манипуляции для обеспечения удобных сайтов рестрикции, удаления лишней ДНК, удаления сайтов рестрикции и т.п. Для этой цели могут быть применены *in vitro* мутагенез, репарация праймеров, рестрикция, отжиг, повторные замены, например, транзиции и трансверсии.

В практике настоящего изобретения можно использовать ряд промоторов. Промоторы могут быть выбраны исходя из требуемого результата. Нуклеиновые кислоты можно комбинировать с конститутивными, индуцируемыми,

специфичными для стадии роста, специфичными для типа клеток, предпочтительными для тканей, специфичными для тканей или другими промоторами для экспрессии в представляющем интересе организме. См., например, промоторы, описанные в WO 99/43838, US 8575425; US 7790846; US 8147856; US 8586832; US 7772369; US 7534939; US 6072050; US 5659026; US 5608149; US 5608144; US 5604121; US 5569597; US 5466785; US 5399680; US 5268463; US 5608142; US 6177611; эти патенты включены в настоящее изобретение в виде ссылок.

Для экспрессии в растениях к конститутивным промоторам также относятся промотор CaMV 35S (Odell с соавт., *Nature*, 1985, 313:810-812); актин риза (McElroy с соавт., *Plant Cell*, 1990, 2:163-171); убиквитин (Christensen с соавт., *Plant Mol. Bio*, 1989, 12:619-632; Christensen с соавт., *Plant Mol. Biol*, 1992, 18:675-689); pEMU (Last с соавт., *Theor. Appl. Genet*, 1991, 81:581-588); MAS (Velten с соавт., *EMBO J*, 1984, 3:2723-2730).

Примерами индуцируемых промоторов являются промотор Adh1, который индуцируется гипоксией или холодовым стрессом, промотор Hsp70, который индуцируется тепловым стрессом, промотор PPDK и промотор пепкарбоксилазы, которые оба индуцируются светом. Также применимы химически индуцируемые промоторы, такие как промотор In2-2, который индуцируется антидотом гербицидов (US 5364780), промотор Axig1, который индуцируется ауксином и специфичен для тапетума, но также активен в каллюсе (PCT US01/22169), стероид-чувствительные промоторы (см., например, промотор ERE, который индуцируется эстрогеном, и промотор, индуцируемый глюкокортикоидами, описанными в публикациях Schena с соавт., *Proc. Natl. Acad. Sci. USA*, 1991, 88:10421-10425; McNellis с соавт., *Plant J*. 1998, 14(2):247-257), и промоторы, индуцируемые тетрациклином и репрессируемые тетрациклином (см., например, Gatz с соавт., *Mol. Gen. Genet*. 1991, 227:229-237; US 5814618; US 5789156); сведения приведены в настоящем изобретении в виде ссылок.

Тканеспецифичные или предпочтительные для тканей промоторы можно использовать для целевой экспрессии экспрессионной конструкции в конкретной ткани. В некоторых вариантах осуществления настоящего изобретения промоторы, специфичные или предпочтительные для определенных тканей, активны в растительной ткани. Примеры промоторов, находящихся под контролем развития у растений, включают промоторы, которые предпочтительно

инициируют транскрипцию в определенных тканях, таких как листья, корни, плоды, семена или цветы. «Тканеспецифичный» промотор представляет собой промотор, который инициирует транскрипцию только в определенных тканях. В отличие от конститутивной экспрессии генов, тканеспецифичная экспрессия является результатом нескольких взаимодействующих уровней генной регуляции. Как таковые, промоторы из гомологичных или близкородственных видов растений могут быть предпочтительными для использования для достижения эффективной и надежной экспрессии трансгенов в конкретных тканях. В некоторых вариантах осуществления настоящего изобретения кассета экспрессии включает промотор, специфичный в отношении определенной ткани. Понятие «промотор, предпочтительный в отношении определенной ткани» означает, что он инициирует транскрипцию предпочтительно в данной ткани, но не обязательно полностью или исключительно в определенной ткани.

В некоторых вариантах осуществления настоящего изобретения молекулы нуклеиновой кислоты, кодирующие RGN, crPНК и/или tracrPНК, содержат промотор, специфичный для типа клеток. Промотор, «специфичный в отношении типа клеток», является промотором, который в первую очередь управляет экспрессией в определенных типах клеток в одном или нескольких органах. Некоторые примеры растительных клеток, в которых в первую очередь могут быть активны специфические для типа клеток промоторы, функционирующие в растениях, включают, например, клетки ВЕТЛ, сосудистые клетки в корнях, листьях, клетки стебля и ствольные клетки. Молекулы нуклеиновой кислоты также могут включать промоторы, предпочтительные для определенных типов клеток. Промотор «предпочтительного типа клеток» представляет собой промотор, который главным образом управляет экспрессией в основном, но не обязательно полностью или исключительно в определенных типах клеток в одном или нескольких органах. Некоторые примеры клеток растений, в которых могут быть предпочтительно активны промоторы предпочтительного типа клеток, функционирующие в растениях, включают, например, клетки ВЕТЛ, сосудистые клетки в корнях, листьях, клетки стебля и ствольные клетки.

Последовательности нуклеиновых кислот, кодирующие RGN, crPНК, tracrPНК и/или sgPНК, могут быть функционально связаны с последовательностью промотора, которая распознается фаговой РНК -

полимеразой, например, для синтеза мРНК *in vitro*. В таких вариантах осуществления настоящего изобретения транскрибированная *in vitro* РНК может быть очищена для использования в методах по настоящему изобретению.

Например, промоторная последовательность может представлять собой промоторную последовательность фагов T7, T3 или SP6 или вариант промоторной последовательности T7, T3 или SP6. В таких вариантах осуществления настоящего изобретения экспрессированный белок и/или РНК могут быть очищены для использования в методах модификации генома, описанных в настоящем изобретении.

В определенных вариантах осуществления настоящего изобретения полинуклеотид, кодирующий RGN, crРНК, tracrРНК и/или sgРНК, также может быть связан с сигналом полиаденилирования (например, сигналом поли(А) SV40 и другими сигналами, функционирующими в растениях) и/или по меньшей мере с одной последовательностью терминации транскрипции. Кроме того, последовательность, кодирующая RGN, также может быть связана с последовательностью (или последовательностями), кодирующей по меньшей мере один сигнал ядерной локализации, по меньшей мере один домен, проникающий в клетку, и/или по меньшей мере один сигнальный пептид, способный доставлять белки в определенные субклеточные местоположения, как описано в настоящем изобретении.

Полинуклеотид, кодирующий RGN, crРНК, tracrРНК, и/или sgРНК, может присутствовать в векторе или множестве векторов. Термин «вектор» относится к полинуклеотидной композиции для переноса, доставки или введения нуклеиновой кислоты в клетку-хозяин. Подходящие векторы включают плазмидные векторы, фагмиды, космиды, искусственные/мини-хромосомы, транспозоны и вирусные векторы (например, лентивирусные векторы, аденоассоциированные вирусные векторы, бакуловирусный вектор). Вектор может содержать дополнительные последовательности контроля экспрессии (например, энхансерные последовательности, последовательности Козака, последовательности полиаденилирования, последовательности терминации транскрипции), селективируемые маркерные последовательности (например, гены устойчивости к антибиотикам), точки начала репликации и другие.

Дополнительную информацию можно найти в кн.: Ausubel с соавт., «Current Protocols in Molecular Biology», изд-во John Wiley & Sons, Нью-Йорк, 2003; в

кн.: Sambrook и Russell, «Molecular Cloning: A Laboratory Manual», изд-во Cold Spring Harbor Press, Cold Spring Harbor, Нью-Йорк, 3е издание, 2001.

Вектор также может содержать селективный маркерный ген для селекции трансформированных клеток. Селектируемые маркерные гены используют для селекции трансформированных клеток или тканей. Маркерные гены включают гены, кодирующие устойчивость к антибиотикам, например, гены, кодирующие неомицинофосфотрансферазу II (NEO) и гигромицинофосфотрансферазу (HPT), а также гены, придающие устойчивость к гербицидным соединениям, таким как глюфосинат аммония, бромоксинил, имидазолиноны и 2,4-дихлорфеноксиацетат (2,4-Д).

В некоторых вариантах осуществления настоящего изобретения касета экспрессии или вектор экспрессии, содержащий последовательность, кодирующую полипептид RGN, может дополнительно содержать последовательность, кодирующую сгРНК и/или trасгРНК, или сгРНК и trасгРНК, объединенные для создания направляющей РНК. Последовательность (последовательности), кодирующая сгРНК и/или trасгРНК, может быть функционально связана по меньшей мере с одной последовательностью контроля транскрипции для экспрессии сгРНК и/или trасгРНК в представляющем интересе организме или клетке-хозяине. Например, полинуклеотид, кодирующий сгРНК и/или trасгРНК, может быть функционально связан с последовательностью промотора, которая распознается РНК-полимеразой III (Pol III). Примеры подходящих промоторов Pol III включают, но не ограничиваются ими, U6, U3, H1 и 7SL РНК промоторы млекопитающих и U6 и U3 промоторы риса.

Отмечено, что экспрессирующие конструкции, содержащие нуклеотидные последовательности, кодирующие RGN, сгРНК, trасгРНК и/или sgРНК, можно использовать для трансформации представляющих интерес организмов. Методы трансформации включают введение нуклеотидной конструкции в представляющий интерес организм. Под «введением» подразумевается введение нуклеотидной конструкции в клетку-хозяина таким образом, что конструкция получает доступ внутрь клетки-хозяина. Способы по настоящему изобретению не требуют конкретного метода введения нуклеотидной конструкции в организм-хозяин, требуется только для того, чтобы нуклеотидная конструкция получила доступ внутрь по меньшей мере одной клетки организма-хозяина.

Клетка-хозяин может быть эукариотической или прокариотической клеткой. В определенных вариантах осуществления настоящего изобретения эукариотическая клетка-хозяин является клеткой растения, клеткой млекопитающего или клеткой насекомого. Методы введения нуклеотидных конструкций в растения и другие клетки-хозяева известны в данной области техники, включая, но, не ограничиваясь ими, способы стабильной трансформации, способы транзиторной трансформации и способы, опосредованные вирусами.

Применение методов позволяет получить трансформированный организм, такой как растение, включая целые растения, а также органы растения (например, листья, стебли, корни и т.д.), семена, растительные клетки, пропагулы, зародыши и их потомство. Клетки растений могут быть дифференцированными или недифференцированными (например, каллус, клетки суспензионной культуры клеток, протопласты, клетки листьев, клетки корня, клетки флоэмы, пыльца).

«Трансгенные организмы», или «трансформированные организмы», или «стабильно трансформированные» организмы, или клетки, или ткани относятся к организмам, в которые внедрен или интегрирован полинуклеотид, кодирующий RGN, сгРНК и/или trасгРНК по настоящему изобретению. Установлено, что другие экзогенные или эндогенные последовательности нуклеиновых кислот или фрагментов ДНК также могут быть включены в клетку-хозяина. Трансформация с помощью *Agrobacterium*, опосредованная агробактериями или методом молекулярной биолистики, остается двумя преимущественно используемыми подходами для трансформации растительных клеток. Тем не менее, трансформация клетки-хозяина может быть осуществлена путем инфекции, трансфекции, микроинъекции, электропорации, микропроекции, биолистики или бомбардировки частицами, электропорации, применения волокон силикагеля/углерода, с помощью ультразвука, с применением ПЭГ, совместного осаждения фосфатом кальция, методом поликатионного ДМСО, DEAE-декстранового метода, опосредованного вирусами метода, опосредованного липосомами метода и других. Для введения с помощью вирусов полинуклеотида, кодирующего RGN, сгРНК и/или trасгРНК, используют лентивирусное, аденовирусное и аденоассоциированное опосредованное вирусами введение и экспрессию, а также используют каулимовирусы, геминивирусы и РНК-вирусы

растений.

Протоколы проведения трансформации, а также протоколы введения полипептидов или полинуклеотидных последовательностей в растения, могут варьировать в зависимости от типа клетки-хозяина (например, клетки
 5 однодольного или двудольного растения), предназначенной для трансформации. Методы трансформации известны в данной области и включают те, которые описаны в US 8575425; US 7692068; US 8802934; US 7541517; сущность каждого из этих патентов включена в настоящее изобретение в виде ссылки. См., также, Rakoczy-Trojanowska M., *Cell Mol Biol Lett*, 2002, 7:849-858; Jones с соавт., *Plant*
 10 *Methods* 2005, 1:5; Rivera с соавт., *Physics of Life Reviews*, 2012, 9:308-345; Bartlett с соавт., *Plant Methods*, 2008, 4:1-12; Bates G.W., *Methods in Molecular Biology*, 1999, 111:359-366; Binns, Thomashow, *Annual Reviews in Microbiology*, 1988, 42:575-606; Christou P., *The Plant Journal*, 1992, 2:275-281; Christou P., *Euphytica*, 1995, 85:13-27; Tzfira с соавт., *TRENDS in Genetics*, 2004, 20:375-383;
 15 Yao с соавт., *Journal of Experimental Botany*, 2006, 57:3737-3746; Zupan, Zambryski, *Plant Physiology*, 1995, 107:1041-1047; Jones с соавт., *Plant Methods* 2005, 1:5.

Трансформация может привести к стабильному или временному включению нуклеиновой кислоты в клетку. Предполагают, что «стабильная трансформация»
 20 означает, что введенная в клетку-хозяина нуклеотидная конструкция интегрирует в геном клетки-хозяина и способна передаваться по наследству. Подразумевают, что «временная трансформация» означает, что полинуклеотид вводится в клетку-хозяина и не интегрирует в ее геном.

Методы трансформации хлоропластов известны в данной области техники.
 25 См., например, Svab с соавт., *Proc. Natl. Acad. Sci. USA*, 1990, 87:8526-8530; Svab, Maliga *Proc. Natl. Acad. Sci. USA*, 1993, 90:913-917; Svab, Maliga, *EMBO J.*, 1993, 12:601-606. Этот метод основан на доставке ДНК, содержащей селективный маркер, с помощью пушки частиц и нацеливании ДНК на пластидный геном посредством гомологичной рекомбинации. Кроме того,
 30 пластидная трансформация может быть осуществлена путем трансактивации молчащего переносимого пластидами трансгена посредством предпочтительной в тканях экспрессии кодируемой в ядре и управляемой пластидой РНК-полимеразы. Такая система описана в публикации McBride с соавт., *Proc. Natl. Acad. Sci. USA*, 1994, 91:7301-7305.

Трансформированные клетки могут быть выращены в трансгенный организм, такой как растение, руководствуясь обычными методами. См., например, McCormick с соавт. (1986) *Plant Cell Reports* 5:81-84. Затем эти растения могут быть выращены и либо опылены одним и тем же трансформированным штаммом, либо разными штаммами, и идентифицирован полученный гибрид, имеющий конститутивную экспрессию требуемой фенотипической характеристики. Можно вырастить два или более поколений, чтобы гарантировать, что экспрессия желаемого фенотипического признака стабильно поддерживается и передается по наследству, а затем можно собрать семена, чтобы гарантировать достижение желаемого фенотипического признака. Таким образом, настоящее изобретение обеспечивает трансформированные семена (также называемые «трансгенными семенами»), имеющие нуклеотидную конструкцию по настоящему изобретению, например кассету экспрессии по настоящему изобретению, стабильно включенную в их геном.

В другом варианте, трансформированные клетки могут быть введены в организм. Эти клетки могли происходить из организма, в котором клетки трансформировали *ex vivo*.

Предусмотренные в настоящем изобретении последовательности можно использовать для трансформации любых видов растений, включая, но не ограничиваясь ими, однодольные и двудольные растения. Примеры представляющих интерес растений включают, но не ограничиваются ими, кукурузу, сорго, пшеницу, подсолнечник, помидоры, крестоцветные, перец, картофель, хлопок, рис, сою, сахарную свеклу, сахарный тростник, табак, ячмень и масличный рапс, *Brassica* sp., люцерну, рожь, просо, сафлор, арахис, сладкий картофель, маниоку, кофе, кокос, ананас, цитрусовые деревья, какао, чай, банан, авокадо, инжир, гуава, манго, оливки, папайю, кешью, макадамию, миндаль, овес, овощи, декоративные растения и хвойные деревья.

Овощи включают, но ими перечень не ограничивается, помидоры, салат, зеленую фасоль, лимскую фасоль, горох и представителей рода *Curcumins*, такие как огурец, мускусную дыню и мускусную дыню. Декоративные растения включают, но ими перечень не ограничивается, азалии, гортензии, гибискусы, розы, тюльпаны, нарциссы, петунии, гвоздики, пуансеттии и хризантемы. Предпочтительно, растениями по настоящему изобретению являются сельскохозяйственные растения (например, кукуруза, сорго, пшеница,

подсолнечник, помидоры, крестоцветные, перец, картофель, хлопок, рис, соя, сахарная свекла, сахарный тростник, табак, ячмень, масличный рапс и т.д.).

В настоящем изобретении термин «растение» включает клетки растений, протопласты растений, культуры клеток тканей растений, из которых растения
5 могут быть регенерированы, каллусы растений, заросли растений и клетки растений, которые являются интактными в растениях или частях растений, таких как зародыши, пыльца, семяпочки, семена, листья, цветы, ветки, плоды, зерна, колосья, початки, шелуха, стебли, корни, кончики корней, пыльники и т.п. Термин «зерно» означает зрелые семена, произведенные коммерческими
10 производителями для целей, отличных от выращивания или воспроизводства вида. Потомство, варианты и мутанты созданных растений также входят в рамки охвата настоящего изобретения при условии, что эти части содержат введенные полинуклеотиды. Кроме того, предусматривают переработанный растительный продукт или побочный продукт, который сохраняет последовательности,
15 раскрытые в настоящем документе, включая, например, соевый шрот.

Полинуклеотиды, кодирующие RGN, crPHK и/или tracrPHK, также можно использовать для трансформации любых прокариотических видов, включая, но, не ограничиваясь ими, археи и бактерии (например, *Bacillus* sp., *Klebsiella* sp.,
20 *Streptomyces* sp., *Rhizobium* sp., *Escherichia* sp., *Pseudomonas* sp., *Salmonella* sp., *Shigella* sp., *Vibrio* sp., *Yersinia* sp., *Mycoplasma* sp., *Agrobacterium* sp., *Lactobacillus* sp.).

Полинуклеотиды, кодирующие RGN, crPHK и/или tracrPHK, можно использовать для трансформации каких-либо эукариотических видов, включая, но ими не ограничиваясь, животных (например, млекопитающих, насекомых, рыб, птиц и рептилий), грибы, амёб, водорослей и дрожжей.
25

Для введения нуклеиновых кислот в клетки млекопитающих или ткани-мишени можно использовать обычные вирусные и невирусные методы переноса генов. Такие способы можно использовать для введения нуклеиновых кислот, кодирующих компоненты системы CRISPR, в клетки в культуре или в
30 организме-хозяине. Невирусные векторные системы доставки включают ДНК-плазмиды, РНК (например, транскрипт вектора, описанного в настоящем документе), голую нуклеиновую кислоту и нуклеиновую кислоту в комплексе с носителем доставки, таким как липосома. Вирусные векторные системы доставки включают ДНК- и РНК-вирусы, которые после поступления в клетку

имеют либо эписомальные, либо интегрированные геномы. Обзоры методов генной терапии представлены в публикациях: Anderson, *Science*, 1992, 256: 808-813; Nabel, Feigner, *TIBTECH*, 1993, 11:211-217; Mitani, Caskey, *TIBTECH*, 1993, 11:162-166; Dillon, *TIBTECH*, 1993, 11:167-175; Miller, *Nature*, 1992, 357:455-460; Van Brunt, *Biotechnology*, 1988, 6(10): 1149-1154; Vigne, *Restorative Neurology and Neuroscience*, 1995, 8:35-36; Kremer, Perricaudet, *British Medical Bulletin*, 1995, 51(1):31-44; кн.: Haddada с соавт., «Current Topics in Microbiology and Immunology», под ред. Doerfler, Bohm, 1995; Yu с соавт., *Gene Therapy*, 1994, 1:13-26 (1994).

10 Методы невирусной доставки нуклеиновых кислот включают липофекцию, нуклеофекцию, микроинъекцию, биолистику, виросомы, липосомы, иммунолипосомы, поликатион или липид: конъюгаты нуклеиновых кислот, голую ДНК, искусственные вирионы и усиленное агентом поглощение ДНК. Липофекция описана, например, в US 5049386, US 4946787, US 4897355, а
15 реагенты для липофекции можно приобрести на коммерческой основе (например, продукты Transfectam™ и Lipofectin™). Катионные и нейтральные липиды, которые подходят для эффективной липофекции полинуклеотидов с распознаванием рецепторов, включают описанные в WO 91/17424; WO 91/16024. Доставка может осуществляться в клетки (например, введение *in vitro* или *ex vivo*) или ткани-мишени (например, введение *in vivo*). Приготовление
20 комплексов липид:нуклеиновая кислота, включая целевые липосомы, такие как иммунолипидные комплексы, известны специалистам в данной области (см., например, Crystal, *Science*, 1995, 270:404-410; Blaese с соавт., *Cancer Gene Ther*, 1995, 2:291-297; Behr с соавт., *Bioconjugate Chem*, 1994, 5:382-389; Remy с
25 соавт., *Bioconjugate Chem*, 1994, 5:647-654; Gao с соавт., *Gene Therapy*, 1995, 2:710-722; Ahmad с соавт., *Cancer Res*, 1992, 52:4817-4820; US 4186183, US 4217344, US 4235871, US 4261975, US 4485054, US 4501728, US 4774085, US 4837028, US 4946787).

Использование систем на основе вирусов РНК или ДНК для доставки
30 нуклеиновых кислот использует преимущества высокоразвитых процессов для нацеливания вируса на определенные клетки в организме и доставки вирусной полезной нагрузки в ядро. Вирусные векторы можно вводить непосредственно пациентам (*in vivo*) или их можно использовать для лечения клеток *in vitro*, а модифицированные клетки можно необязательно вводить пациентам (*ex vivo*).

Обычные вирусные системы могут включать векторы ретровирусные, лентивирусные, аденовирусные, аденоассоциированные векторы и вируса простого герпеса для переноса генов. Интеграция в геном хозяина возможна с помощью методов переноса генов ретровируса, лентивируса и

5 аденоассоциированного вируса, что часто приводит к длительной экспрессии встроенного трансгена. Кроме того, высокая эффективность трансдукции наблюдалась во многих разных типах клеток и тканях-мишенях.

Тропизм ретровируса можно изменить за счет включения чужеродных белков оболочки, расширяя потенциальную популяцию клеток-мишеней.

10 Лентивирусные векторы представляют собой ретровирусные векторы, которые способны трансдуцировать или инфицировать неделящиеся клетки и обычно продуцируют высокие вирусные титры. Таким образом, выбор системы переноса ретровирусного гена будет зависеть от ткани-мишени. Ретровирусные векторы состоят из цис-действующих длинных концевых повторов с упаковывающей

15 способностью до 6-10 т.п.о. чужеродной последовательности. Минимальные цис-действующие LTR достаточны для репликации и упаковки векторов, которые затем используются для интеграции терапевтического гена в клетку-мишень для обеспечения постоянной экспрессии трансгена. Широко

используемые ретровирусные векторы включают векторы, основанные на вирусе лейкоза мышей (MuLV – murine leukemia virus), вирусе лейкоза гиббоновых обезьян (GaLV – gibbon ape leukemia virus), вирусе иммунодефицита обезьян (SIV – Simian Immuno deficiency virus), вирусе иммунодефицита человека (ВИЧ) и их комбинациях (см., например, Buchsner с соавт., *J. Viral.*, 1992, 66:2731-2739; Johann с соавт., *J. Viral.*, 1992, 66:1635-1640; Sommerfelt с соавт., *Viral.*, 20 1990, 176:58-59; Wilson с соавт., *J. Viral.*, 1989, 63:2374-2378; Miller с соавт., *J. Viral.*, 1991, 65:2220-2224; PCT/US94/05700).

В вариантах осуществления настоящего изобретения, где предпочтительна временная экспрессия, могут использоваться системы на основе аденовирусов. Векторы на основе аденовируса способны к очень высокой эффективности

30 трансдукции во многих типах клеток и не требуют клеточного деления. С такими векторами были получены высокие титр и уровни экспрессии. Этот вектор можно получать в больших количествах в относительно простой системе. Векторы аденоассоциированного вируса (AAV – Adeno-associated virus) также могут применяться для трансдукции клеток целевыми нуклеиновыми кислотами,

например, получение *in vitro* нуклеиновых кислот и пептидов, и для процедур генной терапии *in vivo* и *ex vivo* (см., например, West с соавт., *Virology* 1987, 160:38-47; US 4797368; WO 93/24641; Katin, *Human Gene Therapy*, 1994, 5:793-801; Muzyczka, *J. Clin. Invest.*, 1994, 94:1351. Конструкции рекомбинантных векторов AAV описаны в ряде публикаций, в том числе в US 5173414; Tratschin с соавт., *Mol. Cell. Biol.*, 1985, 5:3251-3260; Tratschin с соавт., *Mol. Cell. Biol.*, 1984, 4:2072-2081; Hermonat, Muzyczka, *PNAS*, 1984, 81:6466-6470; Samulski с соавт., *J. Viral.*, 1989, 63:03822-3828). Упаковывающие клетки обычно используют для формирования вирусных частиц, способных инфицировать клетку-хозяина. К таким клеткам относятся клетки 293, которые упаковывают аденовирус, и клетки ψ J2 или клетки PA317, которые упаковывают ретровирус.

Вирусные векторы, используемые в генной терапии, обычно создают путем получения клеточной линии, которая упаковывает вектор нуклеиновой кислоты в вирусную частицу. Векторы обычно содержат минимальные вирусные последовательности, необходимые для упаковки и последующей интеграции в хозяина, при этом другие вирусные последовательности заменяют экспрессионной кассетой для полинуклеотида (полинуклеотидов), подлежащих экспрессии. Отсутствующие вирусные функции обычно поставляются в виде транс упаковочной клеточной линией. Например, векторы AAV, используемые в генной терапии, обычно содержат только последовательности ITR из генома AAV, которые необходимы для упаковки и интеграции в геном хозяина. Вирусная ДНК упакована в клеточную линию, которая содержит хелперную плазмиду, кодирующую другие гены AAV, а именно *rep* и *cap*, но без последовательностей ITR.

Линия клеток также может быть инфицирована аденовирусом в качестве хелпера. Хелпер-вирус индуцирует репликацию вектора AAV и экспрессию генов AAV с плазмиды-хелпера. Вспомогательная плазида не упакована в значительных количествах из-за отсутствия последовательностей ITR. Контаминация аденовирусом может быть понижена, например, с помощью термической обработки, к которой аденовирус более чувствителен, чем AAV. Дополнительные методы доставки нуклеиновых кислот в клетки известны специалистам в данной области. См., например, US 20030087817, включенный в настоящее изобретение в виде ссылки.

В некоторых вариантах осуществления настоящего изобретения клетка-хозяин временно или постоянно трансфицирована одним или несколькими векторами, описанными в настоящем изобретении. В некоторых вариантах осуществления настоящего изобретения клетку трансфицируют так, как это происходит в природе у субъекта. В некоторых вариантах осуществления настоящего изобретения у субъекта берут трансфицированную клетку. В некоторых вариантах осуществления настоящего изобретения клетку получают из клеток, взятых у субъекта, таких как клеточная линия. В данной области техники известно большое разнообразие клеточных линий для тканевых культур. Примеры клеточных линий включают, но не ограничиваются ими, C8161, CCRF-CEM, MOLT, mIMCD-3, NHDF, HeLaS3, Huh1, Huh4, Huh7, HUVEC, HASMC, HEK_n, HEK_a, MiaPaCell, Panel, PC-3, TFl, CTLL-2, CIR, Rat6, CVI, RPTE, A10, T24, 182, A375, ARH-77, Calul, SW480, SW620, SKOV3, SK-UT, CaCo2, P388D1, SEM-K2, WEHI-231, HB56, TIB55, lurkat, *145.01*, LRMB, Bcl-1, BC-3, IC21, DLD2, Raw264.7, NRK, NRK-52E, MRC5, MEF, Hep G2, HeLa B, HeLa T4, COS, COS-1, COS-6, COS-M6A, клетки эпителия почки обезьяны BS-C-1, фибробласты эмбриона мыши BALB/3T3, 3T3 Swiss, 3T3-L1, фибробласты эмбриона человека 132-d5; фибробласты мыши 10.1, 293-T, 3T3, 721, 9L, A2780, A2780ADR, A2780cis, A172, A20, A253, A431, A-549, ALC, B16, B35, BCP-I клетки, BEAS-2B, bEnd.3, BHK-21, BR 293, BxPC3, C3H-10T1/2, C6/36, Cal-27, CHO, CHO-7, CHO-IR, CHO-K1, CHO-K2, CHO-T, CHO Dhfr^{-/-}, COR-L23, COR-L23/CPR, COR-L235010, CORL23/ R23, COS-7, COV-434, CML T1, CMT, CT26, D17, DH82, DU145, DuCaP, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, HEK-293, HeLa, Hepalclc7, HL-60, HMEC, HT-29, линия клеток Jurkat, клетки *IY*, клетки K562, Ku812, KCL22, KGI, KYO1, LNCap, Ma-Mel 1-48, MC-38, MCF-7, MCF-10A, MDA-MB-231, MDA-MB-468, MDA-MB-435, MDCKII, MDCKII, MOR/ 0.2R, MONO-MAC 6, MTD-1A, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NALM-1, NW-145, OPCN/OPCT линия клеток, Peeg, PNT-1A/ PNT 2, RenCa, RIN-5F, RMA/RMAS, клетки Saos-2, Sf-9, SkBr3, T2, T-47D, T84, линия клеток THP1, U373, U87, U937, VCaP, клетки Vero, WM39, WT-49, X63, YAC-1, YAR и их трансгенные вариации. Линии клеток можно получить из разных источников, известных специалистам в этой области (см., например, Американскую

коллекцию типовых культур (American Type Culture Collection – ATCC) (Манассас, штат Вирджиния)).

В некоторых вариантах осуществления настоящего изобретения клетку, трансфицированную одним или несколькими векторами, описанными в настоящем изобретении, используют для создания новой клеточной линии, содержащей одну или несколько последовательностей, полученных из вектора. В некоторых вариантах осуществления настоящего изобретения клетка, временно трансфицированная компонентами системы CRISPR, как описано в настоящем изобретении (например, путем временной трансфекции одного или нескольких векторов или трансфекции РНК), и модифицированная за счет активности комплекса CRISPR, используется для создания новой клеточной линии, содержащей клетки с модификацией, но лишенные какой-либо другой экзогенной последовательности. В некоторых вариантах осуществления настоящего изобретения клетки, временно или постоянно трансфицированные одним или несколькими векторами, описанными в настоящем изобретении, или клеточные линии, полученные из таких клеток, используют для оценки одного или нескольких тестируемых соединений.

В некоторых вариантах осуществления настоящего изобретения один или несколько векторов, описанных в настоящем изобретении, используют для получения трансгенного отличного от человека организма, трансгенного животного или растения. В некоторых вариантах осуществления настоящего изобретения трансгенным животным является млекопитающее, такое как мышь, крыса или кролик.

25 IV. Варианты и фрагменты полипептидов и полинуклеотидов

В настоящем описании представлены активные варианты и фрагменты встречающейся в природе (т.е. дикого типа) РНК-направляемой нуклеазы, аминокислотная последовательность которой представлена в виде SEQ ID NO: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 117, 137 или 235, а также активные варианты и фрагменты встречающихся в природе повторов CRISPR, таких как последовательность, представленная в виде SEQ ID NO: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 118, 240, 273 или 287, а также активный вариант и фрагменты встречающихся в природе tracrРНК, таких как последовательность, представленная как SEQ ID NO: 3, 11, 18, 25, 32, 40, 48, 56,

63, 71, 77, 84, 91, 97, 105, 112, 119, 241, 274 или 286 и кодирующие их полинуклеотиды.

Хотя активность варианта или фрагмента может быть изменена по сравнению с интересующим полинуклеотидом или полипептидом, вариант и фрагмент должны сохранять функциональность представляющего интерес полинуклеотида или полипептида. Например, вариант или фрагмент может иметь повышенную активность, пониженную активность, другой спектр активности или любое другое изменение активности по сравнению с представляющим интерес полинуклеотидом или полипептидом.

Фрагменты и варианты встречающихся в природе полипептидов RGN, таких как описанные в настоящем изобретении, будут сохранять специфичную для последовательности РНК-направляемую ДНК-связывающую активность. В определенных вариантах осуществления настоящего изобретения фрагменты и варианты природных полипептидов RGN, такие как описанные в настоящем изобретении, сохраняют нуклеазную активность (одноцепочечную или двухцепочечную).

Фрагменты и варианты природных повторов CRISPR, таких как описанные в настоящем изобретении, сохраняют способность, будучи частью направляющей РНК (содержащей *tracrRNA*), связывать и направлять РНК-направляемую нуклеазу (в комплексе с направляющей РНК) к целевой нуклеотидной последовательности специфичным для последовательности способом.

Термин «фрагмент» относится к части полинуклеотидной или полипептидной последовательности по настоящему изобретению. «Фрагменты» или «биологически активные части» включают полинуклеотиды, содержащие достаточное количество смежных нуклеотидов для сохранения биологической активности (т.е. связывания и направления RGN специфичным для последовательности образом к целевой нуклеотидной последовательности, когда она включена в состав нуклеотидной направляющей РНК). «Фрагменты» или «биологически активные части» включают полипептиды, содержащие достаточное количество смежных аминокислотных остатков для сохранения биологической активности (т.е. связывания с последовательностью нуклеотидов-мишеней специфичным для последовательности образом при образовании направляющей РНК). Фрагменты белков RGN включают те, которые короче, чем полноразмерные последовательности, из-за использования альтернативного

сайта старта ниже по цепи. Биологически активная часть белка RGN может быть полипептидом, который включает, например, 10, 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050 или более расположенных смежно аминокислотных остатков последовательностей SEQ ID NO: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 117, 137 или 235. Такие биологически активные части могут быть получены с помощью рекомбинантных методов и оценены по специфичной для последовательности РНК-направляющей ДНК-связывающей активности. Биологически активный фрагмент повторяющейся последовательности CRISPR может содержать по меньшей мере 8 последовательных аминокислот последовательностей SEQ ID NO: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 118, 240, 273 или 287. Биологически активная часть повторяющейся последовательности CRISPR может быть полинуклеотидом, включающим, например, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 или 20 непрерывно расположенных нуклеотидов последовательности SEQ ID NO: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 118, 240, 273 или 287. Биологически активная часть tracrPНК может быть полинуклеотидом, который содержит, например, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80 или более смежных нуклеотидов последовательностей SEQ ID NO: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112, 119, 241, 274 или 286.

Как правило, «варианты» означают по существу сходные последовательности. Для полинуклеотидов вариант включает делецию и/или добавление одного или нескольких нуклеотидов в одном или нескольких внутренних сайтах в нативном полинуклеотиде и/или замену одного или нескольких нуклеотидов в одном или нескольких сайтах в нативном полинуклеотиде. Используемый в настоящем изобретении термин «нативный» или «дикий тип» применительно к полинуклеотиду или полипептиду означает встречающуюся в природе нуклеотидную последовательность или аминокислотную последовательность, соответственно. Для полинуклеотидов консервативные варианты включают те последовательности, которые из-за вырожденности генетического кода кодируют нативную аминокислотную последовательность представляющего интерес гена. Встречающиеся в природе аллельные варианты, такие как эти, могут быть идентифицированы с использованием хорошо известных методов молекулярной биологии, таких как,

например, полимеразная цепная реакция (ПЦР) и методы гибридизации, описанные ниже. К вариантам полинуклеотидов также относятся полинуклеотиды, полученные синтетическим путем, такие как те, которые получены, например, с использованием сайт-направленного мутагенеза, но которые по-прежнему кодируют представляющий интерес полипептид или полинуклеотид. Как правило, варианты конкретного полинуклеотида, описанные в настоящем изобретении, по меньшей мере приблизительно на 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или более процентов идентичны последовательности этого конкретного полинуклеотида, что определено программами выравнивания последовательностей и параметрами, также описанными в настоящем описании.

Варианты конкретного полинуклеотида, описанного в настоящем изобретении (т.е., эталонный полинуклеотид), также можно оценивать путем сравнения процентной идентичности последовательности между полипептидом, кодируемым вариантным полинуклеотидом, и полипептидом, кодируемым эталонным полинуклеотидом. Процентная идентичность последовательностей между любыми двумя полипептидами может быть рассчитана с использованием программ выравнивания последовательностей и параметров, также описанных в настоящем изобретении. Если любую данную пару полинуклеотидов, описанных в настоящем изобретении, оценивают путем сравнения процента идентичности последовательностей, общих для двух полипептидов, которые они кодируют, процент идентичности последовательностей между двумя кодируемыми полипептидами составляет по меньшей мере примерно 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или более.

В определенных вариантах осуществления настоящего изобретения описанные в нем полинуклеотиды кодируют полипептид РНК-направляемой нуклеазы, содержащий аминокислотную последовательность, которая по меньшей мере на 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или более процентов идентична аминокислотной последовательности SEQ ID NO: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 117, 137 или 235.

Биологически активный вариант полипептида RGN по настоящему изобретению может отличаться примерно на 1-15 аминокислотных остатков, примерно на 1-10, например, примерно на 6-10, всего на 5, всего на 4, всего на 3, всего на 2 или всего на 1 аминокислотный остаток. В конкретных вариантах осуществления настоящего изобретения полипептиды могут содержать N-концевое или C-концевое усечение, которое может быть следствием делеции 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050 аминокислот или более с N- или C-конца полипептида.

В определенных вариантах осуществления настоящего изобретения предусмотренные в настоящем изобретении полинуклеотиды содержат или кодируют повтор CRISPR, содержащий нуклеотидную последовательность, которая по меньшей мере на 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или больше процентов идентична нуклеотидной последовательности, представленной как SEQ ID NO: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 118, 240, 273 или 287.

Раскрытые в настоящем документе полинуклеотиды могут содержать или кодировать tracrPНК, содержащую нуклеотидную последовательность, которая по меньшей мере на 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или больше процентов идентична нуклеотидной последовательности, представленной как SEQ ID NO: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112, 119, 241, 274 или 286.

Биологически активные варианты повтора CRISPR или tracrPНК по настоящему изобретению могут отличаться примерно на 1-15 нуклеотидов, примерно на 1-10, например, примерно на 6-10, всего на 5, всего на 4, всего на 3, всего на 2 или всего на 1 нуклеотид. В некоторых вариантах осуществления настоящего изобретения полинуклеотиды могут содержать 5'- или 3'-усечение, которое может быть следствием делеции 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80 нуклеотидов или более с 5'- или 3'-конца полинуклеотида.

Установлено, что модификации полипептидов RGN, повторов CRISPR и tracrPНК, предусмотренных в настоящем изобретении, могут быть осуществлены путем создания вариантов белков и полинуклеотидов. Изменения,

разработанные человеком, могут быть введены путем применения методов сайт-направленного мутагенеза. В другом варианте природные, еще неизвестные или еще не идентифицированные полинуклеотиды и/или полипептиды, структурно и/или функционально родственные описанным в настоящем изобретении последовательностям, также могут быть идентифицированы и они войдут в область охвата настоящего изобретения. Консервативные аминокислотные замены могут быть сделаны в неконсервативных областях, которые не изменяют функцию белков RGN. В качестве альтернативы могут быть осуществлены модификации, улучшающие активность RGN.

К вариантам полинуклеотидов и белков также относятся последовательности и белки, полученные в результате мутагенных и рекомбинантных процедур, таких как перетасовка ДНК. С помощью такой процедуры один или несколько различных белков RGN, описанных в настоящем изобретении (например, SEQ ID NO: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 117, 137 или 235), изменяют для создания нового белка RGN, обладающего требуемыми свойствами. Таким образом, библиотеки рекомбинантных полинуклеотидов создают из популяции полинуклеотидов с родственными последовательностями, включающих области последовательностей, которые имеют существенную идентичность последовательностей и могут быть гомологично рекомбинированы *in vitro* или *in vivo*. Например, используя такой подход, мотивы последовательности, кодирующие представляющие интерес домен, могут быть перетасованы между последовательностями RGN, представленными в настоящем изобретении, и другими известными генами RGN для получения нового гена, кодирующего белок с улучшенным представляющий интерес свойством, таким как увеличенная K_m в случае фермента. Стратегии такой перетасовки ДНК известны в данной области. См., например, Stemmer, *Proc. Natl. Acad. Sci. USA*, 1994, 91:10747-10751; Stemmer, *Nature*, 1994, 370:389-391; Cramer с соавт., *Nature Biotech.* 1997, 15:436-438; Moore с соавт., *J. Mol. Biol.* 1997, 272:336-347; Zhang с соавт., *Proc. Natl. Acad. Sci. USA*, 1997, 94:4504-4509; Cramer с соавт., *Nature*, 1998, 391:288-291; US 5605793; US 5837458. «Перетасованная» нуклеиновая кислота представляет собой нуклеиновую кислоту, полученную с помощью процедуры перетасовки, например, любой описанной в настоящем изобретении. Перетасованные нуклеиновые кислоты получают путем рекомбинации

(физической или виртуальной) двух или более нуклеиновых кислот (или специфических отрезков), например, искусственным и, необязательно, обратным образом. Обычно в процессах перетасовки используют один или несколько этапов скрининга для идентификации представляющих интерес нуклеиновых кислот; этот этап скрининга может быть выполнен до или после любого этапа рекомбинации. В некоторых (но не во всех) перетасованных вариантах осуществления настоящего изобретения желательно выполнить несколько раундов рекомбинации перед отбором, чтобы увеличить разнообразие пула, подлежащего скринингу. Общий процесс рекомбинации и отбора необязательно повторяется обратным образом (рекурсивно). В зависимости от контекста перетасовка может относиться к общему процессу рекомбинации и отбора или, наоборот, может просто относиться к рекомбинационным частям общего процесса.

В настоящем изобретении понятия «идентичность последовательности» или «идентичность» применительно к двум последовательностям полинуклеотидов или полипептидов относятся к остаткам в двух последовательностях, которые являются одинаковыми при выравнивании для максимального соответствия в указанном окне сравнения. Когда процент идентичности последовательности используют в отношении белков, признается, что положения остатков, которые не идентичны, часто отличаются консервативными заменами аминокислот, при этом аминокислотные остатки заменены другими аминокислотными остатками с аналогичными химическими свойствами (например, зарядом или гидрофобностью) и поэтому не изменяют функциональных свойств молекулы. Если последовательности отличаются консервативными заменами, процент идентичности последовательности может быть скорректирован в сторону увеличения, чтобы скорректировать консервативный характер замены. Последовательности, которые отличаются такими консервативными заменами, называют имеющими «подобие последовательностей» или «сходство». Возможности выполнения такой регулировки известны специалистам в данной области техники. Как правило, она включает оценку консервативной замены как частичного, а не полного несоответствия, тем самым увеличивая процент идентичности последовательности. Таким образом, например, если идентичной аминокислоте присваивается оценка 1, а неконсервативной замене присваивается нулевая оценка, консервативной замене присваивается оценка от

нуля до 1. Оценку консервативных замен рассчитывают, например, как в программе PC/GENE PC/GENE (фирма IntelliGenetics, Маунтин-Вью, Калифорния).

В настоящем изобретении «процент идентичности последовательности»
5 означает значение, определяемое путем сравнения двух оптимально
выровненных последовательностей в окне сравнения, при этом часть
полинуклеотидной последовательности в окне сравнения может содержать
инсерции или делеции (т.е., пробелы) по сравнению с эталонной
10 последовательность (которая не содержит инсерций или делеций) для
оптимального выравнивания двух последовательностей. Процент рассчитывается
путем определения количества позиций, в которых идентичные основания
нуклеиновой кислоты или аминокислотные остатки встречаются в обеих
последовательностях, чтобы получить количество совпадающих позиций, путем
15 деления количества совпадающих позиций на общее количество позиций в окне
сравнения и умножения результата на 100, чтобы получить процент
идентичности последовательности.

Если не указано иное, значения идентичности/сходства
последовательности, представленные в настоящем изобретении, относятся к
значению, полученному с помощью GAP Version 10 с использованием
20 следующих параметров: % идентичности и % сходства для нуклеотидной
последовательности с использованием GAP Weight = 50 и Length Weight = 3, а
также матрицы оценок nwsgapdna.cmp; % идентичности и % сходства для
аминокислотной последовательности с использованием GAP Weight = 8 и Length
Weight = 2, а также матрицы оценок BLOSUM62; или какой-либо другой
25 эквивалентной им программы. Под «эквивалентной программой»
подразумевается любая программа сравнения последовательностей, которая для
любых двух рассматриваемых последовательностей создает выравнивание,
имеющее идентичные совпадения нуклеотидов или аминокислотных остатков и
идентичный процент идентичности последовательностей по сравнению с
30 соответствующим выравниванием, созданным GAP Version 10.

Две последовательности являются «оптимально выровненными», когда они
выровнены для оценки сходства с использованием определенной матрицы
аминокислотных замен (например, BLOSUM62), штрафа за наличие гэпа и
штрафа за удлинение гэпа, чтобы получить максимально возможный балл для

этой пары последовательностей. Матрицы аминокислотных замен и их использование для количественной оценки сходства между двумя последовательностями хорошо известны в данной области и описаны например, Dayhoff с соавт. в кн: «*Atlas of Protein Sequence and Structure*», под ред. М.О. Dayhoff, 1978, т. 5, приложение 3, с. 345-352, Natl. Biomed. Res. Found.,
5 Вашингтон, округ Колумбия; Henikoff с соавт., *Proc. Natl. Acad. Sci. USA*, 1992, 89:10915-10919. Матрицу BLOSUM62 часто используют в качестве матрицы замещения по умолчанию в протоколах выравнивания последовательностей. Штраф за существование гэпа налагается за введение одного аминокислотного гэпа в одну из выровненных последовательностей, а штраф за удлинение гэпа налагается за каждую дополнительную пустую аминокислотную позицию, вставленную в уже открытый гэп. Выравнивание определяется положениями аминокислот каждой последовательности, с которых начинается и заканчивается выравнивание, и, необязательно, вставкой гэпа или нескольких гэпов в одну или
10 обе последовательности, чтобы получить максимально возможную оценку. Хотя оптимальное выравнивание и оценку можно выполнить вручную, этот процесс упрощается за счет использования реализованного компьютером алгоритма выравнивания, например, BLAST 2.0 с пробелами, описанного в Altschul с соавт., *Nucleic Acids Res.*, 1997, 25:3389-3402, который есть в свободном доступе
20 на веб-сайте Национального центра биотехнологической информации (www.ncbi.nlm.nih.gov). Оптимальные выравнивания, в том числе множественные выравнивания, могут быть подготовлены с использованием, например, PSI-BLAST, доступной на сайте www.ncbi.nlm.nih.gov и описанной в публикации Altschul с соавт., *Nucleic Acids Res.*, 1997, 25:3389-3402.

25 Что касается аминокислотной последовательности, которая оптимально выровнена с эталонной последовательностью, аминокислотный остаток «соответствует» положению в эталонной последовательности, с которым остаток спаривается при выравнивании. «Положение» обозначается числом, которое последовательно идентифицирует каждую аминокислоту в эталонной
30 последовательности на основе ее положения относительно N-конца. Из-за делеций, инсерций, усечений, слияний и т. д., которые необходимо учитывать при определении оптимального выравнивания, в общем случае количество аминокислотных остатков в тестируемой последовательности, определяемое простым подсчетом от N-конца, не обязательно совпадает с номером

соответствующей позиции в эталонной последовательности. Например, в случае делеции в выровненной тестовой последовательности не будет аминокислоты, соответствующей положению в эталонной последовательности в месте делеции. При наличии вставки в выровненной эталонной последовательности эта вставка не будет соответствовать ни одному положению аминокислоты в эталонной последовательности. В случае усечения или слияния могут быть участки аминокислот либо в эталонной, либо в выровненной последовательности, которые не соответствуют ни одной аминокислоте в соответствующей последовательности.

10

V. Антитела

Антитела к полипептидам RGN или рибонуклеопротеинам, включающим полипептиды RGN по настоящему изобретению, включая те, которые имеют аминокислотную последовательность, представленную как SEQ ID NO: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110 или 117, или их активные варианты или фрагменты также входят в область охвата настоящего изобретения. Способы получения антител хорошо известны в данной области (см., например, кн.: Harlow, Lane «Antibodies: A Laboratory Manual», 1988, изд-во Cold Spring Harbour Laboratory, Cold Spring Harbor, Нью-Йорк; патент US 4196265). Эти антитела можно использовать в наборах для обнаружения и выделения полипептидов RGN или рибонуклеопротеинов. Таким образом, данное изобретение обеспечивает наборы, содержащие антитела, которые специфически связываются с полипептидами или рибонуклеопротеинами, описанными в настоящем изобретении, включая, например, полипептиды, имеющие последовательность SEQ ID NO: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110 или 117.

15

20

25

VI. Системы и рибонуклеопротеиновые комплексы для связывания представляющей интерес последовательности-мишени и способы их получения

В настоящем изобретении предусматривают систему связывания представляющей интерес последовательности-мишени, которая содержит по меньшей мере одну направляющую РНК или кодирующую ее нуклеотидную последовательность, и по меньшей мере одну РНК-направляемую нуклеазу или кодирующую ее нуклеотидную последовательность. Направляющая РНК гибридизируется с представляющей интерес последовательностью-мишенью, а

30

также образует комплекс с полипептидом RGN, тем самым направляя полипептид RGN на связывание с последовательностью-мишенью. В некоторых из этих вариантов осуществления настоящего изобретения RGN включает аминокислотную последовательность SEQ ID NO: 1, 9, 16, 23, 30, 38, 46, 54, 61, 5 69, 75, 82, 89, 95, 103, 110, 117, 137 или 235 или ее активный вариант или фрагмент. В различных вариантах осуществления настоящего изобретения направляющая РНК содержит повторяющуюся последовательность CRISPR, включающую нуклеотидную последовательность SEQ ID NO: 2, 10, 17, 24, 31, 10 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 118, 240, 273 или 287 или ее активный вариант или фрагмент. В определенных вариантах осуществления настоящего изобретения направляющая РНК содержит tracrPНК, включающую нуклеотидную последовательность SEQ ID NO: 3, 11, 18, 25, 32, 40, 48, 56, 63, 15 71, 77, 84, 91, 97, 105, 112, 119, 241, 274 или 286, или ее активный вариант или фрагмент. Направляющая РНК системы может быть однонаправляющей РНК или двухнаправляющей РНК. В определенных вариантах осуществления настоящего изобретения система включает РНК-направляемую нуклеазу, гетерологичную направляющей РНК, при этом RGN и направляющая РНК в природе не образуют комплексов.

Система для связывания представляющей интерес последовательности-мишени, предусмотренная в настоящем изобретении, может быть 20 рибонуклеопротеиновым комплексом, который представляет собой по меньшей мере одну молекулу РНК, связанную по меньшей мере с одним белком. Предусмотренные в настоящем изобретении рибонуклеопротеиновые комплексы содержат по меньшей мере одну направляющую РНК в качестве компонента 25 РНК и РНК-направляемую нуклеазу в качестве белкового компонента. Такие рибонуклеопротеиновые комплексы могут быть выделены и очищены из клетки или организма, которые естественным образом экспрессируют полипептид RGN и сконструированы для экспрессии конкретной направляющей РНК, специфичной в отношении представляющей интерес последовательности- 30 мишени. В другом варианте рибонуклеопротеиновый комплекс может быть выделен и очищен из клетки или организма, которые были трансформированы полинуклеотидами, кодирующими полипептид RGN и направляющую РНК, и культивирован в условиях, обеспечивающих экспрессию полипептида RGN и направляющей РНК. Таким образом, предложены способы получения

полипептида RGN или рибонуклеопротеинового комплекса RGN. Такие способы включают культивирование клетки, содержащей нуклеотидную последовательность, кодирующую полипептид RGN, а в некоторых вариантах осуществления настоящего изобретения нуклеотидную последовательность, кодирующую направляющую РНК, в условиях, в которых экспрессируется полипептид RGN (и в некоторых вариантах осуществления настоящего изобретения направляющая РНК). Затем полипептид RGN или рибонуклеопротеин RGN можно выделить из лизата культивируемых клеток.

Методы очистки полипептида RGN или рибонуклеопротеинового комплекса RGN из лизата биологического образца известны в данной области (например, эксклюзионная и/или аффинная хроматография, 2D-PAGE, ВЭЖХ, обращенно-фазовая хроматография, иммунопреципитация). В конкретных методах полипептид RGN получают рекомбинантно, и он содержит метку для очистки, помогающий очистке, включая, но, не ограничиваясь ими, глутатион-S-трансферазу (GST – glutathione-S-transferase), хитин-связывающий белок (CBP – chitin binding protein), мальтозо-связывающий белок, тиоредоксин (TRX), поли(NANP), метку тандемной аффинной очистки (TAP – tandem affinity purification), myc, AcV5, AU1, AU5, E, ECS, E2, FLAG, HA, nus, Softag 1, Softag 3, Strep, SBP, Glu-Glu, HSV, KT3, S, S1, T7, V5, VSV-G, 6xHis, 10xHis, белок-носитель биотинкарбоксила (BCCP – biotin carboxyl carrier protein) и кальмодулин. Как правило, меченый полипептид RGN или рибонуклеопротеиновый комплекс RGN очищают с использованием иммобилизованной металлической аффинной хроматографии. Следует понимать, что могут быть использованы и другие подобные методы, известные в данной области, включая другие формы хроматографии или, например, иммунопреципитацию, отдельно или в комбинации.

«Выделенный» или «очищенный» полипептид или его биологически активная часть фактически свободны от компонентов, которые обычно сопровождают или взаимодействуют с полипептидом, присутствующим в его естественном окружении. Таким образом, выделенный или очищенный полипептид практически не содержит другого клеточного материала или материала культуральной среды, если он получен рекомбинантными методами, или практически не содержит химических предшественников или других химических веществ, если он синтезирован химическим путем. Белок,

практически не содержащий клеточного материала, включает препараты белка, содержащие менее примерно 30%, 20%, 10%, 5% или 1% (по сухому весу) контаминантного белка. Когда белок или его биологически активную часть по настоящему изобретению получают рекомбинантным путем, оптимально культуральная среда представляет собой менее примерно 30%, 20%, 10%, 5% или 1% (по сухому весу) химических предшественников или представляющих интерес химических веществ, а именно небелковых соединений.

Конкретные методы, предусмотренные в настоящем изобретении для связывания и/или расщепления представляющей интерес последовательности-мишени, включают использование собранного *in vitro* рибонуклеопротеинового комплекса RGN. Сборка рибонуклеопротеинового комплекса RGN *in vitro* может быть осуществлена с использованием любого метода, известного в данной области, в котором полипептид RGN контактирует с направляющей РНК в условиях, обеспечивающих связывание полипептида RGN с направляющей РНК. В настоящем изобретении применяемые термины «контакт», «контактирование», «контактирующий» относятся к помещению компонентов нужной реакции вместе в условиях, пригодных для осуществления этой реакции. Полипептид RGN, полученный в результате трансляции *in vitro*, может быть выделен и очищен из биологического образца, клеточного лизата или культуральной среды, или получен химическим синтезом. Полипептид RGN и направляющая РНК могут быть приведены в контакт в растворе (например, буферном солевом растворе) допуская сборку *in vitro* рибонуклеопротеинового комплекса RGN.

VII. Способы связывания, расщепления и модификации последовательности-мишени

В настоящем описании предусматривают способы связывания, расщепления и модификации целевой последовательности нуклеотидов, представляющей интерес. Способы включают доставку системы, содержащей по меньшей мере одну направляющую РНК или кодирующий ее полинуклеотид и по меньшей мере один полипептид RGN или кодирующий ее полинуклеотид, в целевую последовательность или клетку, органеллу или эмбрион, содержащие целевую последовательность. В некоторых из подобных вариантов осуществления настоящего изобретения RGN содержит аминокислотную последовательность, представленную как SEQ ID NO: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 117, 137 или 235 или ее активный вариант или фрагмент. В различных

вариантах осуществления настоящего изобретения направляющая РНК содержит повторяющуюся повторность CRISPR, содержащую нуклеотидную последовательность, представленную как SEQ ID NO: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 118, 240, 273 или 287 или ее активный вариант или фрагмент. В определенных вариантах осуществления настоящего изобретения направляющая РНК содержит tracrРНК, содержащую нуклеотидную последовательность, представленную как SEQ ID NO: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112, 119, 241, 274 или 286 или ее активный вариант или фрагмент. Направляющая РНК системы может быть однонаправляющей РНК или двунаправляющей РНК. RGN системы может быть мертвой RGN, обладающей никазой активностью, или может быть слитым полипептидом. В некоторых вариантах осуществления настоящего изобретения слитый полипептид содержит отредактированный по основаниям полипептид, например, цитидиндезаминазу или аденозиндезаминазу. В других вариантах осуществления настоящего изобретения RGN-слитый белок содержит обратную транскриптазу. В других вариантах осуществления настоящего изобретения RGN-слитый белок содержит полипептид, который рекрутирует компоненты функционального комплекса репарации нуклеиновых кислот, например, компонент метаболического пути эксцизионной репарации нуклеотидов (NER – nucleotide excision repair) или эксцизионной репарации нуклеотидов, связанной с транскрипцией (TC-NER – transcription coupled-nucleotide excision repair) (Wei с соавт., *PNAS USA* 2015, 112(27):E3495-504; Troelstra с соавт., *Cell* 1992, 71:939-953; Marnef с соавт., *J Mol Biol* 2017, 429(9):1277-1288) согласно предварительной заявке US No. 62/966,203, поданной 27 января 2020, сущность которой включена в настоящее изобретение в виде ссылки. В некоторых вариантах осуществления настоящего изобретения RGN-слитый белок содержит CSB (van den Boom с соавт., *J Cell Biol* 2004, 166(1):27-36; van Gool с соавт., *EMBO J* 1997, 16(19):5955-65; пример представлен ниже в виде последовательности SEQ ID NO: 268), которая является участником пути TC-NER (эксцизионная репарация нуклеотидов, связанная с транскрипцией) и участвует в рекрутировании других членов. В других вариантах осуществления настоящего изобретения этот RGNABP-слитый белок содержит активный домен CSB, такой как кислотный домен CSB, включающий аминокислотные остатки

356-394 последовательности SEQ ID NO: 268 (Teng с соавт., *Nat Commun* 2018, 9(1):4115).

В определенных вариантах осуществления настоящего изобретения RGN и/или направляющая РНК является гетерологичной в отношении клетки, органеллы или эмбриона, в которые вводят RGN и/или направляющую РНК (или полинуклеотид (нуклеотиды), кодирующие по меньшей мере одну RGN и направляющую РНК).

В тех вариантах осуществления настоящего изобретения, в которых способ включает доставку полинуклеотида, кодирующего направляющую РНК и/или полипептид RGN, на следующем этапе клетку или эмбрион можно культивировать в условиях, в которых экспрессируется направляющая РНК и/или полипептид RGN. В различных вариантах осуществления настоящего изобретения способ включает приведение целевой последовательности в контакт с рибонуклеопротеиновым комплексом RGN. Рибонуклеопротеиновый комплекс RGN может содержать RGN, которая является мертвой нуклеазой или обладает никазой активностью. В некоторых вариантах осуществления настоящего изобретения RGN рибонуклеопротеинового комплекса представляет собой слитый полипептид, содержащий полипептид, редактирующий основания. В некоторых вариантах осуществления настоящего изобретения способ включает введение в клетку, органеллу или эмбрион, содержащие целевую последовательность, рибонуклеопротеинового комплекса RGN. Рибонуклеопротеиновый комплекс RGN может быть очищен из биологического образца, получен рекомбинантно и затем очищен или собран *in vitro*, как описано в настоящем изобретении. В тех вариантах осуществления настоящего изобретения, где рибонуклеопротеиновый комплекс RGN, который контактирует с последовательностью-мишенью, или органеллой клетки, или эмбрионом, собирают *in vitro*, метод может дополнительно включать сборку комплекса *in vitro* до контакта с последовательностью-мишенью, клеткой, органеллой или эмбрионом.

Очищенный или собранный *in vitro* рибонуклеопротеиновый комплекс RGN можно ввести в клетку, органеллу или эмбрион с использованием любого метода, известного в данной области, включая электропорацию, но не ограничиваясь ею. В другом варианте полипептид RGN и/или полинуклеотид, кодирующий или включающий направляющую РНК, можно ввести в клетку,

органеллу или эмбрион с использованием любого метода, известного в данной области (например, метода электропорации).

При доставке или контакте с последовательностью-мишенью или клеткой, органеллой или эмбрионом, содержащим последовательность-мишень, направляющая РНК направляет RGN на связывание с последовательностью-мишенью специфичным для последовательности образом. В тех вариантах осуществления настоящего изобретения, где RGN обладает нуклеазной активностью, полипептид RGN расщепляет интересующую последовательность-мишень при связывании. Последовательность-мишень впоследствии может быть модифицирована с помощью механизмов эндогенной репарации, таких как негомологичное соединение концов или репарация, направленная на гомологию, с предусмотренным донорным полинуклеотидом.

Методы измерения связывания полипептида RGN с последовательностью-мишенью известны в данной области и включают анализы иммунопреципитации хроматина, анализы изменения подвижности в геле, анализы осаждения ДНК, репортерные анализы, анализы захвата и обнаружения на микропланшетах. Аналогичным образом, методы измерения расщепления или модификации последовательности-мишени известны в данной области и включают анализы расщепления *in vitro* или *in vivo*, в которых расщепление подтверждается с помощью ПЦР, секвенирования или гель-электрофореза с добавлением или без добавления соответствующей метки (например, радиоизотопа, флуоресцентного вещества) к целевой последовательности для облегчения обнаружения продуктов разрушения. В другом варианте можно использовать анализ реакции экспоненциальной амплификации, запускаемой разрывом (NTEXPAR – nicking triggered exponential amplification reaction). (см., например, Zhang с соавт., *Chem. Sci.*, 2016, 7:4951-4957). Расщепление *in vivo* можно оценить с помощью анализа Surveyor (Guschin с соавт., *Methods Mol Biol*, 2010, 649:247-256).

В некоторых вариантах осуществления настоящего изобретения методы включают использование одного типа RGN в комплексе с более чем одной направляющей РНК. Более чем одна направляющая РНК может нацеливаться на разные области одного гена или может нацеливаться на несколько генов.

В тех вариантах осуществления настоящего изобретения, где донорный полинуклеотид не предусмотрен, двухцепочечный разрыв, введенный полипептидом RGN, может быть репарирован с помощью процесса репарации

негомологичного соединения концов (NHEJ – non-homologous end-joining). Из-за подверженной ошибкам природы NHEJ восстановление двухцепочечного разрыва может привести к модификации целевой последовательности.

Используемый в настоящем изобретении термин «модификация» в отношении молекулы нуклеиновой кислоты относится к изменению нуклеотидной последовательности молекулы нуклеиновой кислоты, которое может быть делецией, инсерцией или заменой одного или нескольких нуклеотидов или их комбинацией. Модификация целевой последовательности может привести к экспрессии измененного белкового продукта или инактивации кодирующей последовательности.

В тех вариантах осуществления настоящего изобретения, где имеется донорный полинуклеотид, донорная последовательность в донорном полинуклеотиде может быть интегрирована или заменена целевой нуклеотидной последовательностью в ходе репарации введенного двухцепочечного разрыва, что приводит к введению экзогенной донорной последовательности. Таким образом, донорный полинуклеотид содержит донорную последовательность, которую желательно ввести в представляющую интерес последовательность-мишень. В некоторых вариантах осуществления настоящего изобретения донорная последовательность изменяет исходную последовательность нуклеотидов-мишеней таким образом, что вновь интегрированная донорная последовательность не будет распознаваться и расщепляться RGN. Интеграция донорной последовательности может быть усилена включением в донорный полинуклеотид фланкирующих последовательностей, которые имеют существенную идентичность последовательности с последовательностями, фланкирующими нуклеотидную последовательность-мишень, что делает возможным процесс репарации, направленный по гомологии. В тех вариантах осуществления настоящего изобретения, в которых полипептид RGN вводит двухцепочечные ступенчатые разрывы, донорный полинуклеотид может содержать донорную последовательность, фланкированную совместимыми выступами, что позволяет осуществлять прямое лигирование донорной последовательности с расщепленной целевой нуклеотидной последовательностью, содержащей выступы, посредством негомологичного процесса репарации во время репарации двухцепочечного разрыва.

В тех вариантах осуществления настоящего изобретения, в которых метод включает использование RGN, являющейся никазой (т.е., способной расщеплять только одну цепь двухцепочечного полинуклеотида), метод может включать введение двух никаз RGN, нацеленных на идентичные или перекрывающиеся целевые последовательности, и расщеплять разные нити полинуклеотида. Например, никаза RGN, которая расщепляет только положительную (+) цепь двухцепочечного полинуклеотида, может быть введена вместе со второй никазой RGN, которая расщепляет только отрицательную (-) цепь двухцепочечного полинуклеотида.

В различных вариантах осуществления настоящего изобретения предусматривают метод связывания целевой нуклеотидной последовательности и обнаружения целевой последовательности, включающий введение в клетку, органеллу или эмбрион по меньшей мере одной направляющей РНК или полинуклеотида, кодирующего ее, и по меньшей мере одного полипептида RGN или кодирующего ее полинуклеотида, экспрессирующего направляющую РНК и/или полипептид RGN (если введены кодирующие последовательности), где полипептид RGN представляет собой мертвую нуклеазу RGN и дополнительно содержит обнаруживаемую метку, кроме того метод дополнительно предусматривает выявление обнаруживаемой метки. Обнаруживаемая метка может быть слита с RGN в виде слитого белка (например, флуоресцентного белка) или может представлять собой небольшую молекулу, конъюгированную с полипептидом RGN или включенную в него, которую можно обнаружить визуально или другими средствами.

Также в настоящем изобретении предусматривают методы модулирования экспрессии целевой последовательности или представляющего интерес гена при регулировании целевой последовательности. Способы включают введение в клетку, органеллу или эмбрион по меньшей мере одной направляющей РНК или полинуклеотида, кодирующего ее, и по меньшей мере одного полипептида RGN или кодирующего его полинуклеотида, экспрессирующего направляющую РНК и/или полипептид RGN (если кодирующие последовательности введены), где полипептид RGN представляет собой мертвую нуклеазу RGN. В некоторых из этих вариантов осуществления настоящего изобретения мертвая нуклеаза RGN представляет собой слитый белок, содержащий домен модулятора экспрессии

(т.е., домен эпигенетической модификации, домен активации транскрипции или домен репрессора транскрипции), как описано в настоящем изобретении.

Настоящее описание также предусматривает методы связывания и/или модификации представляющей интерес нуклеотидной последовательности-мишени. Методы включают систему доставки, содержащую по меньшей мере одну направляющую РНК или полинуклеотид, кодирующий ее, и по меньшей мере один слитый полипептид, содержащий RGN по настоящему изобретению и полипептид, редактирующий основания, например, цитидиндезаминазу или аденозиндезаминазу, или полинуклеотид кодирующий слитый полипептид, к последовательности-мишени или клетке, органелле или эмбриону, содержащим последовательность-мишень.

Специалисту в данной области известно, что любой из описанных в настоящее время методов может быть использован для нацеливания на одну целевую последовательность или несколько целевых последовательностей. Таким образом, методы включают применение одного полипептида RGN в комбинации с несколькими отдельными направляющими РНК, которые могут нацеливаться на несколько отдельных последовательностей в пределах одного гена и/или нескольких генов. Также в настоящем изобретении описывают способы, в которых несколько различных направляющих РНК вводят в комбинацию с несколькими различными полипептидами RGN. Эти направляющие РНК и направляющие РНК/полипептидные системы РНК/RGN могут нацеливаться на несколько различных последовательностей в пределах одного гена и/или нескольких генов.

В одном объекте настоящего изобретения предусматривают наборы, содержащие какой-либо один или несколько любых элементов, описанных в приведенных выше методах и композициях. В некоторых вариантах осуществления настоящего изобретения набор содержит векторную систему и инструкции по применению набора. В некоторых вариантах осуществления настоящего изобретения векторная система содержит (а) первый регуляторный элемент, функционально связанный с последовательностью ДНК, кодирующей последовательность sgРНК, и один или несколько сайтов для инсерции направляющей последовательности выше по цепи от кодируемой последовательности sgРНК, причем при экспрессии направляющая последовательность направляет специфичное к последовательности связывание

комплекса CRISPR с последовательностью-мишенью в эукариотической клетке, при этом комплекс CRISPR содержит фермент CRISPR в комплексе с полинуклеотидом направляющей РНК; и/или (б) второй регуляторный элемент, функционально связанный с кодирующей фермент последовательностью, кодирующей указанный фермент CRISPR, содержащей последовательность ядерной локализации. Элементы могут быть предусмотрены по отдельности или в комбинациях и могут содержаться в любом подходящем контейнере, таком как флакон, пузырек или пробирка.

В некоторых вариантах осуществления настоящего изобретения набор включает инструкции на одном или нескольких языках. В некоторых вариантах осуществления настоящего изобретения набор включает один или несколько реагентов для использования в процессе, использующем один или несколько элементов, описанных в настоящем изобретении. Реагенты могут быть в любом подходящем контейнере. Например, набор может содержать один или несколько реакционных буферов или буферов для хранения. Реагенты могут поставляться в форме, пригодной для использования в конкретном анализе, или в форме, требующей добавления одного или нескольких других компонентов перед использованием (например, в виде концентрата или лиофилизированной формы). Буфер может быть любым буфером, включая, но не ограничиваясь этим, карбонатно-натриевый буфер, бикарбонатно-натриевый буфер, боратный буфер, трис-буфер, MOPS-буфер, HEPES-буфер и их комбинации. В некоторых вариантах осуществления настоящего изобретения буфер является щелочным. В некоторых вариантах осуществления настоящего изобретения буфер имеет рН от примерно 7 до примерно 10.

В некоторых вариантах осуществления настоящего изобретения набор содержит один или несколько олигонуклеотидов, соответствующих направляющей последовательности, для инсерции в вектор таким образом, чтобы функционально связать направляющую последовательность и регуляторный элемент. В некоторых вариантах осуществления настоящего изобретения набор содержит полинуклеотид-матрицу гомологичной рекомбинации. В одном объекте настоящего изобретения предусматривают способы применения одного или нескольких элементов системы CRISPR. Комплекс CRISPR по настоящему изобретению обеспечивает эффективное средство для модификации полинуклеотида-мишени. Комплекс CRISPR по

настоящему изобретению имеет широкое применение, включая модификации (например, делеции, инсерции, транслокации, инактивации, активации, редактирование оснований) полинуклеотида-мишени в клетках множества типов. Таким образом, комплекс CRISPR по настоящему изобретению имеет широкий спектр применений, например, в генной терапии, скрининге лекарств, диагностике заболеваний и прогнозировании. Типовой комплекс CRISPR содержит фермент CRISPR в комплексе с направляющей последовательностью, гибридованной с последовательностью-мишенью в полинуклеотиде-мишени.

10 VIII. Целевые полинуклеотиды

В одном объекте настоящего изобретения предусматривают способы модификации полинуклеотида-мишени в эукариотической клетке, которые могут осуществляться *in vivo*, *ex vivo* или *in vitro*. В некоторых вариантах осуществления настоящего изобретения способ включает получение образца клетки, или популяции клеток человека, животного или растения (включая микроводоросли), и модификацию клетки или клеток. Культивирование может происходить на любой стадии *ex vivo*. Клетка или клетки могут быть также повторно введены в животное (но не в человека) или растение (включая микроводоросли).

20 Используя естественную изменчивость, растениеводы комбинируют наиболее полезные гены для получения нужных качеств, таких как урожайность, качество, однородность, выносливость и устойчивость к вредителям. Эти требуемые качества также включают рост, предпочтительную длину дня, требования к температуре, срок начала цветения или плодоношения, содержание 25 жирных кислот, устойчивость к насекомым, болезням, устойчивость к нематодам, устойчивость к грибным поражениям, устойчивость к гербицидам, устойчивость к различным факторам окружающей среды, включая засуху, жару, влажные, холодные, ветровые и неблагоприятные почвенные условия, включая высокую засоленность. Источниками этих полезных генов являются местные или 30 чужеродные сорта, сохранившиеся старые сорта, дикие родственные формы растений и индуцированные мутации, например, обработка растительного материала мутагенами. С помощью настоящего изобретения селекционеры получают новый инструмент для индуцирования мутаций. Соответственно, специалист в данной области может проанализировать геном на наличие

5 полезных генов, и в сортах, обладающих желаемыми свойствами или признаками, использовать настоящее изобретение для индукции роста полезных генов с большей точностью, чем предыдущие мутагенные агенты, и, следовательно, ускорить и улучшить программы усовершенствования сортов растений.

Полинуклеотид-мишень системы RGN может быть любым полинуклеотидом, эндогенным или экзогенным по отношению к эукариотической клетке. Например, полинуклеотид-мишень может быть полинуклеотидом, находящимся в ядре эукариотической клетки.

10 Полинуклеотид-мишень может представлять собой последовательность, кодирующую продукт гена (например, белок), или некодирующую последовательность (например, регуляторный полинуклеотид или мусорную ДНК). Не ограничиваясь какой-либо теорией, полагают, что последовательность-мишень должна быть связана с PAM (мотивом, смежным с протоспейсером), то есть с короткой последовательностью, распознаваемая комплексом CRISPR.

15 Требования к точной последовательности и длине PAM различаются в зависимости от используемого фермента CRISPR, но PAM обычно представляют собой последовательности из 2–5 пар оснований, примыкающих к протоспейсеру (то есть к последовательности-мишени).

20 Полинуклеотид-мишень комплекса CRISPR может включать ряд генов и полинуклеотидов, связанных с заболеванием, а также гены и полинуклеотиды, связанные с сигнальным биохимическим путем. Примеры полинуклеотидов-мишеней включают последовательность, связанную с биохимическим путем передачи сигнала, например, ген или полинуклеотид, связанный с биохимическим путем передачи сигнала. Примеры полинуклеотидов-мишеней включают ген или полинуклеотид, ассоциированный с заболеванием.

25 «Связанный с заболеванием» ген или полинуклеотид относится к любому гену или полинуклеотиду, который дает продукты транскрипции или трансляции на аномальном уровне или в аномальной форме в клетках, полученных из тканей, пораженных заболеванием, по сравнению с контрольными не пораженными заболеваниями тканями или клетками. Это может быть ген, который экспрессируется на аномально высоком уровне; это может быть ген, который экспрессируется на аномально низком уровне, где измененная экспрессия коррелирует с возникновением и/или прогрессированием заболевания. Ген,

30

ассоциированный с заболеванием, также относится к гену, обладающему мутацией (мутациями) или генетической вариацией, которая непосредственно ответственна или находится в неравновесном сцеплении с геном (генами), ответственными за этиологию заболевания (например, причинную мутацию).

5 Транскрибированные или транслированные продукты могут быть известны или неизвестны и, кроме того, могут находиться на нормальном или аномальном уровне. Примеры генов и полинуклеотидов, ассоциированных с заболеванием, доступны в Институте генетической медицины МакКьюсика-Натанса
10 Университета Джона Хопкинса (Балтимор, Мэриленд) и в Национальном центре биотехнологической информации Национальной медицинской библиотеки (Бетезда, Мэриленд), доступны на сайте всемирного интернета.

Хотя системы CRISPR особенно полезны из-за их относительной простоты нацеливания на представляющие интерес геномные последовательности, все еще остается вопрос о том, что RGN может сделать для
15 устранения причинной мутации. Один из подходов заключается в получении слитого белка между RGN (предпочтительно неактивным или никасным вариантом RGN) и ферментом, редактирующим основания, или активным доменом фермента, редактирующего основания, таким как редактор оснований цитидиндезаминаза или аденозиндезаминаза, (патент US 9840699, включенный в
20 настоящее описание в качестве ссылки). В некоторых вариантах осуществления настоящего изобретения способы включают приведение молекулы ДНК в контакт со (а) слитым белком, содержащим RGN по настоящему изобретению, и полипептидом, редактирующим основания, таким как дезаминаза; и (б) gРНК, нацеливающая слитый белок (а) на целевую нуклеотидную последовательность
25 цепи ДНК; при этом молекула ДНК контактирует со слитым белком и gРНК в эффективном количестве и в условиях, подходящих для дезаминирования нуклеотидного основания. В некоторых вариантах осуществления настоящего изобретения последовательность ДНК-мишени содержит последовательность, связанную с заболеванием или нарушением, и где дезаминирование
30 нуклеотидного основания приводит к последовательности, не связанной с заболеванием или нарушением. В некоторых вариантах осуществления настоящего изобретения целевая последовательность ДНК находится в аллеле культурного растения, где конкретный аллель представляющего интерес признака приводит к растению с меньшей агрономической ценностью.

Дезаминирование нуклеотидного основания приводит к аллелю, улучшающему признак и повышающему агротехническую ценность растения.

В некоторых вариантах осуществления настоящего изобретения последовательность ДНК содержит точечную мутацию $T \rightarrow C$ или $A \rightarrow G$,
5 связанную с заболеванием или нарушением, и при этом дезаминирование мутантного основания C или G приводит к образованию последовательности, не связанной с заболеванием или нарушением. В некоторых вариантах осуществления настоящего изобретения дезаминирование корректирует точечную мутацию в последовательности, связанной с заболеванием или
10 расстройством.

В некоторых вариантах осуществления настоящего изобретения последовательность, связанная с заболеванием или нарушением, кодирует белок, и при этом дезаминирование вводит стоп-кодон в последовательность, связанную с заболеванием или нарушением, что приводит к усечению
15 кодируемого белка. В некоторых вариантах осуществления настоящего изобретения контакт осуществляют *in vivo* у субъекта, предрасположенного к заболеванию, болящему или с диагностированным заболеванием или нарушением. В некоторых вариантах осуществления настоящего изобретения заболевание или расстройство представляет собой заболевание, связанное с
20 точечной мутацией или мутацией одного основания в геноме. В некоторых вариантах осуществления настоящего изобретения заболевание представляет собой генетическое заболевание, рак, метаболическое заболевание или лизосомную болезнь накопления.

25 Модификация причинных мутаций с помощью редактирования оснований

Примером наследственного заболевания, которое можно скорректировать с помощью подхода, основанного на слитом белке-редакторе на основе RGN по настоящему изобретению, является синдром Гурлера. Синдром Гурлера, также известный как MPS-1, является результатом дефицита α -L-идуронидазы (IDUA),
30 приводящего к лизосомной болезни накопления, характеризующейся на молекулярном уровне накоплением дерматансульфата и гепарансульфата в лизосомах. Это заболевание, как правило, является наследственным генетическим заболеванием, вызванным мутациями в гене IDUA, кодирующем α -L-идуронидазу. Распространенными мутациями IDUA являются W402X и

Q70X, обе бессмысленные мутации приводят к преждевременному прекращению трансляции. Такие мутации хорошо устраняются с помощью подходов точного редактирования генома (PGE – precise genome editing), поскольку реверсия одного нуклеотида, например, с помощью подхода по редактированию оснований, может восстановить кодирующую последовательность дикого типа и привести к экспрессии белка, контролируемой эндогенными регуляторными механизмами генетического локуса. Кроме того, поскольку известно, что гетерозиготы бессимптомны, терапия PGE, нацеленная на одну из этих мутаций, будет полезна для значительной части пациентов с этим заболеванием, поскольку необходимо скорректировать только один из мутировавших аллелей (Bunge с соавт., *Hum. Mol. Genet.*, 1994, 3(6):861-866; публикация включена в настоящее изобретение в виде ссылки).

Современные методы лечения синдрома Гурлера включают заместительную ферментную терапию и трансплантацию костного мозга (Vellodi с соавт., *Arch. Dis. Child.*, 1997, 76(2):92-99; Peters с соавт., *Blood*, 1998, 91(7):2601-2608, публикация включена в настоящее изобретение в виде ссылки). Хотя заместительная терапия ферментами оказала значительное влияние на выживаемость и качество жизни пациентов с синдромом Гурлера, этот подход требует дорогостоящих и трудоемких еженедельных инфузий. Дополнительные подходы включают доставку гена IDUA на векторе экспрессии или инсерцию гена в локус с высокой экспрессией, такой как локус сывороточного альбумина (US 9956247, патент включен в настоящее изобретение в виде ссылки). Однако эти подходы не восстанавливают исходный локус IDUA до правильной кодирующей последовательности. Стратегия редактирования генома будет иметь ряд преимуществ и наиболее значимые заключаются в том, что регуляция экспрессии генов может контролироваться естественными механизмами, присутствующими у здоровых людей. Кроме того, использование редактирования оснований не обязательно вызывает двухцепочечные разрывы ДНК, которые могут привести к большим хромосомным перестройкам, гибели клеток или онкогенности из-за нарушения механизмов подавления опухоли. Общая стратегия может быть направлена на использование слитых белков-редакторов на основе RGN по настоящему изобретению для нацеливания и коррекции определенных вызывающих заболевание мутаций в геноме человека. Следует понимать, что также могут применяться аналогичные подходы к

целевым заболеваниями, которые могут быть исправлены путем редактирования оснований. Кроме того, следует понимать, что аналогичные подходы к нацеливанию на мутации, вызывающие заболевание, у других видов, в частности у обычных домашних животных или домашнего скота, также могут быть применены с использованием RGN по настоящему изобретению. Обычные домашние животные и домашний скот включают собак, кошек, лошадей, свиней, коров, овец, кур, ослов, змей, хорьков и рыб, включая лосося и креветок.

Модификация причинных мутаций с помощью направленных делеций

RGN по настоящему изобретению также могут быть использованы в терапевтических целях для лечения людей, если причинная мутация является более сложной. Например, некоторые заболевания, такие как атаксия Фридрейха и болезнь Хантингтона, являются результатом значительного увеличения количества повторов трехнуклеотидного мотива в определенной области гена, что влияет на способность экспрессируемого белка функционировать или экспрессироваться. Атаксия Фридрейха (FRDA – Friedreich’s Ataxia) это аутосомное рецессивное заболевание, приводящее к прогрессирующей дегенерации нервной ткани в спинном мозге. Снижение уровня белка фратаксина (FXN – frataxin) в митохондриях вызывает окислительные повреждения и дефицит железа на клеточном уровне. Пониженная экспрессия FXN связана с экспансией триплета GAA в интроне 1 соматического и зародышевого гена FXN. У пациентов с FRDA повтор GAA часто состоит из более чем 70, а иногда даже более чем 1000 (чаще всего 600-900) триплетов, тогда как у здоровых людей имеется около 40 повторов или меньше (Pandolfo с соавт., *Handbook of Clinical Neurology*, 2012, 103: 275-294; Campuzano с соавт., *Science* 1996, 271: 1423-1427; Pandolfo, *Adv. Exp. Med. Biol.*, 2002, 516: 99-118; эти публикации включены в настоящее описание в виде ссылок).

Расширение последовательности тринуклеотидных повторов, вызывающее атаксию Фридрейха (FRDA), происходит в определенном генетическом локусе в гене FXN, называемом областью нестабильности FRDA. РНК-направляемые нуклеазы (RGN) можно использовать для вырезания области нестабильности в клетках пациентов с FRDA. Этот подход требует 1) последовательности RGN и направляющей РНК, которые можно запрограммировать для нацеливания на аллель в геноме человека; и 2) подход по доставке для RGN и направляющей

последовательности. Многие нуклеазы, используемые для редактирования генома, такие как обычно используемая нуклеаза Cas9 из *S. pyogenes* (SpCas9), слишком велики, чтобы их можно было упаковать в векторы аденоассоциированных вирусов (AAV), особенно с учетом длины гена SpCas9 и направляющей РНК в дополнение к другим генетическим элементам, необходимым для функциональных кассет экспрессии. Это усложняет подход с использованием SpCas9.

Определенные РНК-направляемые нуклеазы по настоящему изобретению хорошо подходят для упаковки в вектор AAV вместе с направляющей РНК. Для упаковки двух направляющих РНК, вероятно, потребуется второй вектор, но этот подход все же выгодно отличается от того, что потребовалось бы для более крупной нуклеазы, такой как SpCas9, которая может потребовать разделения последовательности белка между двумя векторами. Настоящее изобретение охватывает стратегию использования RGN по настоящему изобретению, в которой удаляют область геномной нестабильности. Такая стратегия применима к другим заболеваниям и нарушениям, имеющим сходную генетическую основу, например к болезни Гентингтона. Аналогичные стратегии с использованием RGN по настоящему изобретению также могут быть применимы к аналогичным заболеваниям и нарушениям у животных, имеющих агрономическое или экономическое значение, включая собак, кошек, лошадей, свиней, коров, овец, кур, ослов, змей, хорьков и рыб, в том числе лососей, а также креветок.

Модификация причинных мутаций с помощью направленного мутагенеза

RGN по настоящему изобретению также могут быть направлены на введение разрушительных мутаций, которые могут привести к положительному эффекту. Генетические дефекты в генах, кодирующих гемоглобин, особенно в цепи бета-глобина (ген HBB), могут быть причиной ряда заболеваний, известных как гемоглобинопатии, включая серповидноклеточную анемию и талассемии.

У взрослых людей гемоглобин представляет собой гетеротетрамер, состоящий из двух альфа-(α)-подобных глобиновых цепей и двух бета-(β)-подобных глобиновых цепей и 4 групп гемов. У взрослых тетрамер $\alpha_2\beta_2$ называется гемоглобином А (HbA) или гемоглобином взрослых. Как правило, цепи альфа- и бета-глобина синтезируются примерно в соотношении 1:1, и это соотношение, по-видимому, является критическим с точки зрения стабилизации гемоглобина и эритроцитов (красных кровяных телец – RBC). У развивающегося

плода вырабатывается другая форма гемоглобина, фетальный гемоглобин (HbF), который имеет более высокое сродство к кислороду, чем гемоглобин А, причем кислород может доставляться в организм ребенка через кровоток матери.

Фетальный гемоглобин также содержит две цепи α -глобина, но вместо взрослых цепей β -глобина имеет две цепи фетального гамма-(γ)-глобина (т.е. фетальный гемоглобин – это $\alpha_2\gamma_2$). Регуляция переключения с выработки гамма-глобина на бета-глобин довольно сложна и в первую очередь включает в себя понижение транскрипции гамма-глобина с одновременным усилением транскрипции бета-глобина. Примерно на 30-й неделе беременности синтез гамма-глобина у плода начинает снижаться, а производство бета-глобина увеличивается. Примерно к 10-месячному возрасту гемоглобин новорожденного почти полностью состоит из $\alpha_2\beta_2$, хотя некоторое количество HbF сохраняется во взрослом возрасте (примерно 1–3% от общего гемоглобина). У большинства пациенток с гемоглобинопатиями гены, кодирующие гамма-глобин, остаются, но их экспрессия относительно низкая из-за нормальной репрессии генов, происходящей во время родов, как описано выше.

Серповидно клеточная анемия вызывается мутацией V6E в гене β -глобина (HBB) (от GAG до GTG на уровне ДНК), где образующийся гемоглобин обозначается как «гемоглобин S» или «HbS». В условиях пониженного содержания кислорода молекулы HbS агрегируют и образуют фиброзные осадки. Эти агрегаты вызывают аномалии или «серповидность» эритроцитов, что приводит к потере гибкости клеток. Серповидные эритроциты больше не могут протискиваться в капиллярное русло и могут привести к вазоокклюзионному кризу у пациентов с серповидно-клеточной анемией. Кроме того, серповидные эритроциты более хрупкие, чем нормальные эритроциты, и склонны к гемолизу, что в конечном итоге приводит к анемии у пациента.

Лечение и ведение пациентов с серповидно-клеточной анемией — это пожизненное лечение, включающее лечение антибиотиками, обезболивание и переливание крови во время острых эпизодов. Одним из подходов является использование гидроксимочевины, действие которой частично проявляется в виде увеличения продукции гамма-глобина. Однако долгосрочные побочные эффекты хронической терапии гидроксимочевиной до сих пор неизвестны, и лечение дает нежелательные побочные эффекты и может иметь различную эффективность от пациента к пациенту. Несмотря на повышение эффективности

лечения серповидно-клеточной анемии, ожидаемая продолжительность жизни пациентов по-прежнему составляет лишь от 55 до 59 лет, а связанные с этим заболевания расстройства оказывают выраженное влияние на качество жизни пациентов.

5 Талассемии (альфа-талассемия и бета-талассемия) также являются заболеваниями, связанными с гемоглобином, и обычно обусловлены пониженной экспрессией глобиновых цепей. Это может произойти из-за мутаций в регуляторных областях генов или из-за мутации в последовательности, кодирующей глобин, что приводит к снижению экспрессии или к пониженным
10 уровням функционального белка глобина. Лечение талассемии обычно включает переливание крови и хелатирование железа. Трансплантацию костного мозга также используют для лечения людей с тяжелой формой талассемии, если можно найти подходящего донора, но эта процедура может иметь значительный риск.

Один подход, который был предложен для лечения как SCD, так и бета-
15 талассемии, заключается в увеличении экспрессии гамма-глобина таким образом, чтобы HbF функционально замещал aberrantный гемоглобин взрослых. Как упоминалось выше, лечение пациентов с SCD гидроксимочевинной считается успешным отчасти благодаря ее влиянию на увеличение экспрессии гамма-глобина (DeSimone, *Proc Nat'l Acad Sci USA*, 1982, 79(14):4428-31; Ley с соавт., *N. Engl. J. Medicine*, 1982, 307: 1469-1475; Ley с соавт., *Blood*, 1983, 62:370-380; Constantoulakis с соавт., *Blood*, 1988, 72(6):1961-1967; сущность каждой
20 публикации включена в настоящее изобретение в виде ссылки). Увеличение экспрессии HbF включает идентификацию генов, продукты которых играют роль в регуляции экспрессии гамма-глобина. Одним из таких генов является BCL11A.
25 Ген BCL11A кодирует белок цинкового пальца, который экспрессируется в эритроидных клетках-предшественниках у взрослых, и подавление его экспрессии приводит к увеличению экспрессии гамма-глобина (Sankaran с соавт., *Science*, 2008, 322:1839, публикация включена в настоящее изобретение в виде ссылки). Было предложено использование ингибиторной РНК, нацеленной
30 на ген BCL11A (например, публикация патента US 2011/0182867, включенная в настоящее изобретение в качестве ссылки), но эта технология имеет несколько потенциальных недостатков, в том числе невозможность полного нокдауна доставки таких РНК, который может быть проблематичным, и РНК должны

присутствовать постоянно, что требует многократного лечения на протяжении всей жизни.

RGN по настоящему изобретению можно использовать для нацеливания на область энхансера BCL11A, чтобы нарушить экспрессию BCL11A, тем самым увеличивая экспрессию гамма-глобина. Это целевое нарушение может быть достигнуто путем негомологичного соединения концов (NHEJ – non-homologous end joining), посредством чего RGN по настоящему изобретению нацеливается на конкретную последовательность в области энхансера BCL11A, производит двухцепочечный разрыв, а механизм клетки восстанавливает разрыв, как правило, одновременно внедряя вредные мутации. Подобно тому, что описано для других мишеней заболевания, RGN по настоящему изобретению могут иметь преимущества перед другими известными RGN из-за их относительно небольшого размера, что позволяет упаковывать экспрессионные кассеты для RGN и её направляющей РНК в единый вектор AAV для доставки *in vivo*. Аналогичные стратегии с использованием RGN по настоящему изобретению также могут быть применимы к аналогичным заболеваниям и нарушениям как у людей, так и у животных, помимо человека, имеющих агрономическое или экономическое значение.

20 IX. Клетки, включающие полинуклеотидную генетическую модификацию

В настоящем документе представлены клетки и организмы, содержащие представляющую интерес последовательность-мишень, которая была модифицирована с использованием процесса, опосредованного RGN, crРНК и/или tracrРНК, как описано в настоящем изобретении. В некоторых из этих вариантов осуществления настоящего изобретения RGN аминокислотную последовательность SEQ ID NO: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 117, 137 или 235 или их активный вариант или фрагмент. В различных вариантах осуществления настоящего изобретения направляющая РНК содержит повторяющуюся последовательность CRISPR, включающую нуклеотидную последовательность SEQ ID NO: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 118, 240, 273 или 287 их активный вариант или фрагмент. В определенных вариантах осуществления настоящего изобретения направляющая РНК включает tracrРНК, включающую нуклеотидную последовательность SEQ ID NO: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112, 119, 241, 274

или 286 или их активный вариант или фрагмент. Направляющая РНК системы может быть однонаправляющей РНК или двунаправляющей РНК.

5 Модифицированные клетки могут быть эукариотическими (например, клетки млекопитающих, растений, насекомых) или прокариотическими. Также предусматривают органеллы и эмбрионы, содержащие по меньшей мере одну нуклеотидную последовательность, которая была модифицирована с использованием RGN, crРНК и/или tracrРНК, как описано в настоящем изобретении. Генетически модифицированные клетки, организмы, органеллы и эмбрионы могут быть гетерозиготными или гомозиготными по модифицированной нуклеотидной последовательности.

10 Хромосомная модификация клетки, организма, органеллы или эмбриона может привести к измененной экспрессии (регуляция, направленная на повышение или понижение), инактивации или экспрессии измененного белкового продукта или интегрированной последовательности. В тех случаях, когда хромосомная модификация приводит либо к инактивации гена, либо к экспрессии нефункционального белкового продукта, генетически модифицированную клетку, организм, органеллу или эмбрион называют подвергшимися «нокауту». Нокаутный фенотип может быть результатом мутации по типу делеции (т.е., делеции по меньшей мере одного нуклеотида), мутации по типу инсерции (т.е., инсерции хотя бы одного нуклеотида) или нонсенс-мутации (т.е., заменой по меньшей мере одного нуклеотида таким образом, что вводят стоп-кодон).

15 В другом варианте хромосомная модификация клетки, организма, органеллы или эмбриона может вызвать «нокин», который является результатом хромосомной интеграции нуклеотидной последовательности, кодирующей белок. В некоторых из этих вариантов осуществления настоящего изобретения кодирующая последовательность интегрируется в хромосому таким образом, что хромосомная последовательность, кодирующая белок дикого типа, инактивируется, но экзогенно введенный белок экспрессируется.

20 В других вариантах осуществления настоящего изобретения хромосомная модификация приводит к выработке варианта белкового продукта. Продукт экспрессированного варианта белка может иметь замену по меньшей мере одной аминокислоты и/или добавление или делецию по меньшей мере одной аминокислоты. Продукт в виде варианта белка, кодируемый измененной

хромосомной последовательностью, может проявлять измененные характеристики или активность по сравнению с белком дикого типа, включая, помимо прочего, измененную ферментативную активность или субстратную специфичность.

5 В еще одних вариантах осуществления настоящего изобретения хромосомная модификация может привести к изменению паттерна экспрессии белка. В качестве примера, не ограничивающего рамок охвата настоящего изобретения, хромосомные изменения в регуляторных областях, контролирующих экспрессию белкового продукта, могут приводить к
10 сверхэкспрессии или подавлению экспрессии белкового продукта или к измененному паттерну ткани или времени экспрессии.

Х. Наборы или способы обнаружения целевой ДНК или одноцепочечной ДНК

15 Некоторые RGN (например, APG09106.1 и APG09748, представленные как SEQ ID NO: 54 и 137) могут беспорядочно расщеплять нецелевую одноцепочечную ДНК (оцДНК) после активации путем обнаружения ДНК-мишени. Таким образом, в настоящем изобретении предусмотрены композиции и способы обнаружения целевой ДНК (двухцепочечной или одноцепочечной) в
20 образце. В некоторых вариантах осуществления настоящего изобретения желаемая мишень может существовать в виде РНК, такой как геном или часть генома РНК-вируса, такого как, например, коронавирус. В некоторых вариантах осуществления настоящего изобретения коронавирус может быть SARS-подобным коронавирусом. В дополнительных вариантах осуществления
25 настоящего изобретения коронавирус может представлять собой SARS-CoV-2, SARS-CoV или SARS-подобный коронавирус летучих мышей, такой как bat-SL-CoVZC45 (номер регистрации MG772933). В вариантах осуществления настоящего изобретения, если мишень существует в виде РНК, мишень может быть подвергнута обратной транскрипции в молекулу ДНК, на которую RGN
30 может эффективно нацеливаться. За обратной транскрипцией может следовать этап амплификации, такой как методы ОТ-ПЦР, известные в данной области, которые включают термоциклирование, или могут быть изотермические методы, такие как изотермическая амплификация, опосредованная петлей обратной

транскрипции (OT-LAMP – reverse transcription loop-mediated isothermal amplification) (Notomi с соавт., *Nucleic Acids Res*, 2000, 28: E63).

Эти композиции и способы включают использование детекторной одноцепочечной ДНК, которая не гибридизируется с направляющей РНК и не является целевой одноцепочечной ДНК. В некоторых вариантах осуществления настоящего изобретения детекторная оцДНК содержит детектируемую метку, которая обеспечивает детектируемый сигнал после расщепления детекторной оцДНК. Примером, не ограничивающим области охвата настоящего изобретения, является детекторная оцДНК, содержащая пару флуорофор/гаситель, где флуорофор не флуоресцирует, если детекторная оцДНК является целой (т.е., нерасщепленной), поскольку его сигнал подавляется присутствием гасителя в непосредственной близости. Расщепление детекторной одноцепочечной ДНК приводит к удалению гасителя, после чего можно обнаружить флуоресцентную метку. Примерами, не ограничивающими области охвата настоящего изобретения, флуоресцентных меток или красителей, являются Cy5, флуоресцеин (например, FAM, 6 FAM, 5(6) FAM, FITC), Cy3, красители Alexa Fluor® и Texas Red. Примерами, не ограничивающими области охвата настоящего изобретения, являются Iowa Black®FQ, Iowa Black® RQ, гаситель Qx1, гаситель АТТ0 и краситель QSY. В некоторых вариантах осуществления настоящего изобретения детекторная оцДНК содержит второй гаситель, такой как внутренний гаситель, например ZEN™, ТАО™ и Black Hole Quencher®, который может снизить фон и усилить обнаружение сигнала.

В других вариантах осуществления настоящего изобретения детекторная оцДНК содержит обнаруживаемую метку, которая обеспечивает детектируемый сигнал до расщепления детекторной оцДНК, а расщепление оцДНК ингибирует или предотвращает обнаружение сигнала. Примером, не ограничивающим области охвата настоящего изобретения, является детекторная одноцепочечная ДНК, которая содержит пару флуоресцентного резонансного переноса энергии (FRET – fluorescence resonance energy transfer). FRET – это процесс, при котором происходит передача энергии без излучения от возбужденного состояния первого (донорного) флуорофора ко второму (акцепторному) флуорофору, находящемуся в непосредственной близости. Спектр излучения донорного флуорофора перекрывается со спектром возбуждения акцепторного флуорофора. Таким образом, акцепторный флуорофор будет флуоресцировать, если

детекторная оцДНК будет целой (т.е., нерасщепленной), и акцепторный флуорофор перестанет флуоресцировать, если детекторная оцДНК будет расщеплена, поскольку донорный и акцепторный флуорофоры больше не будут находиться в непосредственной близости друг от друга. Донорные и акцепторные флуорофоры FRET известны в данной области и включают, но не ограничиваются ими, голубой флуоресцентный белок (CFP - cyan fluorescent protein)/зеленый флуоресцентный белок (GFP – green fluorescent protein), Cy3/Cy5 и GFP/желтый флуоресцентный белок (YFP – yellow fluorescent protein).

В некоторых вариантах осуществления настоящего изобретения детекторная оцДНК имеет длину от примерно 2 нуклеотидов до примерно 30 нуклеотидов, включая, но не ограничиваясь ими, примерно 2, примерно 3, примерно 4, примерно 5, примерно 6, примерно 7, примерно 8, примерно 9, примерно 10, примерно 11, примерно 12, примерно 13, примерно 14, примерно 15, примерно 16, примерно 17, примерно 18, примерно 19, примерно 20, примерно 21, примерно 22, примерно 23, примерно 24, примерно 25 нуклеотидов, примерно 26 нуклеотидов, примерно 27 нуклеотидов, примерно 28 нуклеотидов, примерно 29 нуклеотидов и примерно 30 нуклеотидов.

Методы обнаружения ДНК-мишени молекулы ДНК включают контактирование образца с RGN, направляющей РНК, способной гибридизироваться с RGN, и последовательностью ДНК-мишени в молекуле ДНК, а также детекторной одноцепочечной ДНК (оцДНК), которая не гибридизуется с направляющей РНК с последующим измерением детектируемого сигнала, продуцируемого расщеплением оцДНК под действием RGN, тем самым обнаруживая последовательность ДНК-мишени молекулы ДНК.

В некоторых вариантах осуществления настоящего изобретения метод может включать стадию амплификации молекул нуклеиновой кислоты в образце либо до контакта с RGN и направляющей РНК, либо одновременно с ним. В некоторых из этих вариантов осуществления настоящего изобретения специфические последовательности, с которыми будет гибридизироваться направляющая РНК, могут быть амплифицированы для повышения чувствительности метода обнаружения.

Образец, в котором ДНК-мишень может быть обнаружена с использованием этих композиций и способов, включающих детекторную одноцепочечную ДНК, включает любой образец, содержащий или

предположительно содержащий нуклеиновую кислоту (например, молекулу ДНК или РНК). Образец может быть получен из любого источника, включая синтетическую комбинацию очищенных нуклеиновых кислот или биологический образец, такой как экстракты мазка из дыхательных путей (например, мазок из носоглотки), клеточный лизат, образец пациента, клетки, ткани, слюна, кровь, сыворотка, плазма, моча, аспират, образцы биопсии, спинномозговая жидкость или микроорганизмы (например, бактерии, вирусы).

Контактирование образца с RGN, направляющей РНК и детекторной одноцепочечной ДНК может включать контактирование *in vitro*, *ex vivo* или *in vivo*. В некоторых вариантах осуществления настоящего изобретения детекторная оцДНК и/или RGN и/или направляющая РНК иммобилизуется, например, на устройстве латерального потока, при этом образец контактирует с иммобилизованной детекторной оцДНК и/или RGN и/или направляющей РНК. В некоторых вариантах осуществления настоящего изобретения антитела против фрагментов антигена на детекторной оцДНК иммобилизуют, например, на устройстве латерального потока таким образом, который позволяет отличить расщепленную детекторную оцДНК от интактной детекторной оцДНК.

В некоторых вариантах осуществления настоящего изобретения способы могут дополнительно включать определение количества ДНК-мишени, присутствующей в образце. Измерение детектируемого сигнала в исследуемом образце можно сравнить с эталонным измерением (например, измерением эталонного образца или его серии, содержащей известное количество ДНК-мишени).

К примерам, не ограничивающим области охвата настоящего изобретения, относят обнаружение однонуклеотидного полиморфизма (SNP – single-nucleotide polymorphism), скрининг рака, обнаружение бактериальной инфекции, обнаружение устойчивости к антибиотикам и обнаружение вирусной инфекции.

Детектируемый сигнал, полученный при расщеплении одноцепочечной ДНК под действием RGN, может быть измерен с использованием любого подходящего метода, известного в данной области техники, включая, наряду с другими, измерение флуоресцентного сигнала, визуальный анализ полос на геле, колориметрическое изменение и наличие или отсутствие электрического сигнала.

Настоящее изобретение предусматривает наборы для обнаружения ДНК-мишени молекулы ДНК в образце, причем набор содержит полипептид RGN, направляющую РНК, способную гибридизироваться с RGN, и последовательность ДНК-мишени в молекуле ДНК, а также детекторную одноцепочечную ДНК, не гибридизирующуюся с направляющей РНК.

Также в настоящем изобретении предусмотрены способы расщепления одноцепочечных ДНК путем контактирования популяции нуклеиновых кислот, где популяция включает последовательность ДНК-мишени молекулы ДНК и множество нецелевых оцДНК, с RGN и направляющей РНК, способной гибридизироваться с RGN и последовательность ДНК-мишени.

В настоящем описании применение единственного числа может подразумевать также множественное значение. Например, «полипептид» может на самом деле означать один или много полипептидов.

Все публикации и патентные заявки, упомянутые в настоящем описании, относятся к той же области техники, к которой относится настоящее изобретение. Все публикации и патентные заявки включены в настоящее изобретение в виде ссылок.

Несмотря на то, что вышеизложенное изобретение было довольно подробно описано с помощью иллюстраций и примеров для ясности изложения сути, очевидно, что изменения и модификации могут быть осуществлены в пределах рассматриваемых вариантов осуществления настоящего изобретения.

К вариантам осуществления настоящего изобретения, не ограничивающим области его охвата, относятся:

1. Молекула нуклеиновой кислоты, включающая полинуклеотид, который кодирует полипептид РНК-направляемой нуклеазы (RGN), содержащий аминокислотную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, или 117;

где указанный полипептид RGN способен связываться с последовательностью ДНК-мишени молекулы ДНК специфическим для РНК-направляющей последовательности образом при связывании с направляющей РНК (gРНК), способной гибридизироваться с указанной последовательностью ДНК-мишени, и

где указанный полинуклеотид, кодирующий полипептид RGN, функционально связан с промотором, гетерологичным указанному полинуклеотиду.

5 2. Молекула нуклеиновой кислоты по п. 1, отличающаяся тем, что указанный полипептид RGN способен расщеплять указанную последовательность ДНК-мишени при связывании.

3. Молекула нуклеиновой кислоты по п. 2, , отличающаяся тем, что указанный полипептид RGN способен производить двухцепочечный разрыв.

10 4. Молекула нуклеиновой кислоты по п. 2,, отличающаяся тем, что указанный полипептид RGN способен производить одноцепочечный разрыв.

5. Молекула нуклеиновой кислоты по п. 2, отличающаяся тем, что указанный полипептид RGN является инактивированной нуклеазой.

15 6. Молекула нуклеиновой кислоты по п.п. 1-5, отличающаяся тем, что указанный полипептид RGN функционально слит с полипептидом с отредактированными основаниями.

7. Молекула нуклеиновой кислоты по п. 6, отличающаяся тем, что редактирующий основания полипептид является дезаминазой.

8. Молекула нуклеиновой кислоты по п. 7, отличающаяся тем, что дезаминаза является цитидиновой дезаминазой или аденозиновой дезаминазой.

20 9. Молекула нуклеиновой кислоты по п.п. 1-8, отличающаяся тем, что полипептид RGN содержит один или несколько сигналов ядерной локализации.

10. Молекула нуклеиновой кислоты по п.п. 1-9, отличающаяся тем, что полипептид RGN оптимизирован по кодонам для экспрессии в эукариотических клетках.

25 11. Молекула нуклеиновой кислоты по п.п. 1-10, отличающаяся тем, что указанная последовательность ДНК-мишени расположена рядом с мотивом, прилегающим к протоспейсеру (PAM – protospacer adjacent motif).

12. Вектор, включающий молекулу нуклеиновой кислоты по п.п. 1-11.

30 13. Вектор по п. 12, дополнительно включающий по меньшей мере одну нуклеотидную последовательность, кодирующую указанную gРНК, способную гибридизироваться с указанной последовательностью ДНК-мишени.

14. Вектор по п. 13, отличающийся тем, что направляющая РНК содержит РНК CRISPR, включающую повторяющуюся последовательность, которая по

меньшей мере на 95% идентична SEQ ID NO: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111 или 118.

15. Вектор по п.п. 13-14, отличающийся тем, что указанная gРНК содержит tracrРНК.

5 16. Вектор по п. 15, отличающийся тем, что tracrРНК по меньшей мере на 95% идентична SEQ ID NO: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112 или 119.

17. Вектор по п.п. 15-16, отличающийся тем, что указанная gРНК является однонаправляющей РНК.

10 18. Вектор по п.п. 15-16, отличающийся тем, что указанная gРНК является двунаправляющей РНК.

19. Клетка, содержащая молекулу нуклеиновой кислоты по п.п. 1-11 или вектор по п.п. 12-18.

15 20. Способ получения полипептида RGN, включающий культивирование клетки по п. 18 в условиях, в которых экспрессируется полипептид RGN.

21. Способ получения полипептида RGN, включающий введение в клетку молекулы гетерологичной нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую полипептид РНК-направляемой нуклеазы (RGN), содержащей аминокислотную последовательность, которая по меньшей
20 мере на 95% идентична SEQ ID NO: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110 или 117;

причем указанный полипептид RGN способен связываться с последовательностью ДНК-мишени молекулы ДНК в РНК-направляемой последовательности специфическим образом, когда связывание с направляющей РНК (gРНК) способно гибридизироваться с указанной последовательностью ДНК-мишени;
25

и культивирование указанной клетки в условиях, в которых экспрессируется полипептид RGN.

30 22. Способ по п.п. 20-21, дополнительно включающий очистку указанного полипептида RGN.

23. Способ по п.п. 20-21, отличающийся тем, что указанная клетка дополнительно экспрессирует одну или несколько направляющих РНК, которые связываются с указанным полипептидом RGN с образованием рибонуклеопротеинового комплекса RGN.

24. Способ по п. 23, дополнительно включающий очистку указанного рибонуклеопротеинового комплекса RGN.

25. Молекула нуклеиновой кислоты, включающая полинуклеотид, кодирующий РНК CRISPR (сгРНК), причем указанная сгРНК содержит
5 спейсерную последовательность и последовательность повтора CRISPR, при этом указанная последовательность повтора CRISPR содержит нуклеотидную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111 или 118;

где направляющая РНК включает:

а) указанную сгРНК; и необязательно,

б) трансактивирующую CRISPR РНК (tracrРНК), способную гибридизироваться с указанной последовательностью повторов CRISPR
указанной сгРНК;

способная гибридизироваться с последовательностью ДНК-мишени
15 молекулы ДНК специфичным для последовательности образом через спейсерную последовательность указанной сгРНК, когда указанная направляющая РНК связана с полипептидом РНК-направляемой нуклеазы (RGN), и

где указанный полинуклеотид, кодирующий сгРНК, функционально связан
20 с промотором, гетерологичным указанному полинуклеотиду.

26. Вектор, включающий молекулу нуклеиновой кислоты по п. 25.

27. Вектор по п. 26, отличающийся тем, что указанный вектор дополнительно содержит полинуклеотид, кодирующий указанную tracrРНК.

28. Вектор по п. 27, отличающийся тем, что указанная tracrРНК содержит
25 нуклеотидную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112 или 119.

29. Вектор по п.п. 27-28, отличающийся тем, что указанный полинуклеотид, кодирующий указанную сгРНК, и указанный полинуклеотид, кодирующий указанную tracrРНК, функционально связаны с одним и тем же промотором и кодируются как однонаправляющая РНК.
30

30. Вектор по п.п. 27-28, отличающийся тем, что указанный полинуклеотид, кодирующий указанную сгРНК, и указанный полинуклеотид, кодирующий указанную tracrРНК, функционально связаны с отдельными промоторами.

31. Вектор по п.п. 26-30, отличающийся тем, что указанный вектор дополнительно содержит полинуклеотид, кодирующий указанный полипептид RGN, причем указанный полипептид RGN содержит аминокислотную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110 или 117.

32. Молекула нуклеиновой кислоты, содержащая полинуклеотид, кодирующий транскрибируемую РНК CRISPR (tracrPНК), содержащую нуклеотидную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112 или 119;

причем направляющая РНК содержит:

а) указанную tracrPНК; и

б) crPНК, содержащую последовательность спейсера и последовательность повтора CRISPR, где указанная tracrPНК способна гибридизироваться с указанной последовательностью повтора CRISPR указанной crPНК;

способная гибридизироваться с последовательностью ДНК-мишени молекулы ДНК специфичным для последовательности образом через спейсерную последовательность указанной crPНК, когда указанная направляющая РНК связана с полипептидом РНК-направляемой нуклеазы (RGN), и

где указанный полинуклеотид, кодирующий tracrPНК, функционально связан с промотором, гетерологичным указанному полинуклеотиду.

33. Вектор, включающий молекулу нуклеиновой кислоты по п. 32.

34. Вектор по п. 33, отличающийся тем, что указанный вектор дополнительно содержит полинуклеотид, кодирующий указанную crPНК.

35. Вектор по п. 34, отличающийся тем, что повторяющаяся последовательность CRISPR указанной crPНК содержит нуклеотидную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111 или 118.

36. Вектор по п.п. 34-35, отличающийся тем, что указанный полинуклеотид, кодирующий указанную crPНК, и указанный полинуклеотид, кодирующий указанную tracrPНК, функционально связаны с одним и тем же промотором и кодируются как однонаправляющая РНК.

37. Вектор по п.п. 34-35, отличающийся тем, что указанный полинуклеотид, кодирующий указанную sgРНК, и указанный полинуклеотид, кодирующий указанную tracrРНК, функционально связаны с отдельными промоторами.

5 38. Вектор по п.п. 33-37, отличающийся тем, что указанный вектор дополнительно содержит полинуклеотид, кодирующий указанный полипептид RGN, где указанный полипептид RGN содержит аминокислотную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110 или 117.

10 39. Система для связывания целевой последовательности ДНК молекулы ДНК, содержащая:

а) одну или несколько направляющих РНК, способных гибридизироваться с указанной целевой последовательностью ДНК, или один или несколько полинуклеотидов, содержащих нуклеотидные последовательности, кодирующие одну или несколько направляющих РНК (gРНК); а также

15 б) полипептид РНК-направляющей нуклеазы (RGN), включающий аминокислотную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110 или 117; или полинуклеотид, содержащий нуклеотидную последовательность, кодирующую полипептид RGN;

20 где указанные нуклеотидные последовательности, кодирующие одну или несколько направляющих РНК и кодирующие полипептид RGN, каждая из которых функционально связана с промотором, гетерологичным указанной нуклеотидной последовательности; а также

25 где одна или несколько направляющих РНК способны образовывать комплекс с полипептидом RGN, чтобы направить указанный полипептид RGN на связывание с указанной последовательностью ДНК-мишени молекулы ДНК.

40. Система по п. 39, отличающаяся тем, что указанная gРНК содержит повторяющуюся последовательность CRISPR, включающую нуклеотидную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111 или 118.

41. Система по п.п. 39-40, отличающаяся тем, что указанная gРНК содержит tracrРНК.

42. Система по п. 41, отличающаяся тем, что указанная tracrRNA содержит нуклеотидную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112 или 119.

5 43. Система по п.п. 41-42, отличающаяся тем, что указанная gRNA является а однонаправляющей РНК (sgRNA).

44. Система по п.п. 41-42, отличающаяся тем, что указанная gRNA является а двунаправляющей РНК.

10 45. Система по п.п. 39-44, отличающаяся тем, что последовательность ДНК-мишени непосредственно примыкает к мотиву, примыкающему к протоспейсеру (PAM).

46. Система по п.п. 39-45, отличающаяся тем, что последовательность ДНК-мишени находится внутри клетки.

47. Система по п. 46, отличающаяся тем, что клетка является эукариотической клеткой.

15 48. Система по п. 47, отличающаяся тем, что эукариотическая клетка является растительной клеткой.

49. Система по п. 47, отличающаяся тем, что эукариотическая клетка является клеткой млекопитающего.

20 50. Система по п. 47, отличающаяся тем, что эукариотическая клетка является клеткой насекомого.

51. Система по п. 46, отличающаяся тем, что клетка является прокариотической клеткой.

25 52. Система по п.п. 39-51, отличающаяся тем, что при транскрипции одна или несколько направляющих РНК способны гибридизироваться с последовательностью ДНК-мишени, и направляющая РНК способна образовывать комплекс с полипептидом RGN для прямого расщепления последовательности ДНК-мишени.

53. Система по п. 52, отличающаяся тем, что указанный полипептид RGN способен создавать двухцепочечные разрывы.

30 54. Система по п. 52, отличающаяся тем, что указанный полипептид RGN способен создавать одноцепочечные разрывы.

55. Система по п. 52, отличающаяся тем, что указанный полипептид RGN является неактивной нуклеазой.

56. Система по п.п. 39-55, отличающаяся тем, что указанный полипептид RGN функционально связан с полипептидом, редактирующим основания.

57. Система по п. 56, отличающаяся тем, что полипептид, редактирующий основания, является дезаминазой.

5 58. Система по п. 57, отличающаяся тем, что дезаминаза является цитидиндезаминазой или аденозиндезаминазой.

59. Система по п.п. 39-58, отличающаяся тем, что полипептид RGN содержит один или несколько сигналов ядерной локализации.

10 60. Система по п.п. 39-59, отличающаяся тем, что полипептид RGN оптимизирован по кодонам для экспрессии в эукариотической клетке.

61. Система по п.п. 39-60, отличающаяся тем, что полинуклеотиды, содержащие нуклеотидные последовательности, кодирующие одну или несколько направляющих РНК, и полинуклеотид, содержащий нуклеотидные последовательности, кодирующие полипептид RGN, расположены в одном
15 векторе.

62. Система по п.п. 39-61, отличающаяся тем, что указанная система дополнительно содержит один или несколько донорных полинуклеотидов или одну или несколько нуклеотидных последовательностей, кодирующих один или несколько донорных полинуклеотидов.

20 63. Способ связывания последовательности ДНК-мишени молекулы ДНК, включающий доставку системы в соответствии с п.п. 39-62 к указанной последовательности ДНК-мишени или клетке, содержащей последовательность ДНК-мишени.

25 64. Способ по п. 63, отличающийся тем, что указанный полипептид RGN или указанная направляющая РНК дополнительно содержат выявляемую метку, тем самым допуская обнаружение указанной последовательности ДНК-мишени.

65. Способ по п. 63, отличающийся тем, что указанная направляющая РНК или указанный полипептид RGN дополнительно содержит модулятор экспрессии, тем самым модулируя экспрессию указанной последовательности
30 ДНК-мишени или гена, находящегося под контролем транскрипции указанной последовательностью ДНК-мишени.

66. Способ расщепления или модификации последовательности ДНК-мишени молекулы ДНК, включающий доставку системы по п.п. 39-62 в указанную последовательность ДНК-мишени или клетку, содержащую эту

молекулу ДНК, и расщепление или модификация указанной последовательности ДНК-мишени.

5 67. Способ по п. 66, отличающийся тем, что указанная модифицированная последовательность ДНК-мишени включает инсерцию гетерологичной ДНК в последовательность ДНК-мишени.

68. Способ по п. 66, отличающийся тем, что указанная модифицированная последовательность ДНК-мишени включает делецию по меньшей мере одного нуклеотида в последовательности ДНК-мишени.

10 69. Способ по п. 66, отличающийся тем, что указанная модифицированная последовательность ДНК-мишени включает мутацию по меньшей мере одного нуклеотида в последовательности ДНК-мишени.

70. Способ связывания последовательности ДНК-мишени молекулы ДНК, включающий:

15 а) сборку рибонуклеотидного комплекса РНК-направляемой нуклеазы (RGN) *in vitro* путем объединения:

і) одной или нескольких направляющих РНК, способных гибридизироваться с последовательностью ДНК-мишени; и

20 іі) полипептида RGN, включающего аминокислотную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110 или 117 в условиях, пригодных для образования рибонуклеотидного комплекса RGN; и

25 б) приведение в контакт указанной последовательности ДНК-мишени или клетки, содержащей указанную последовательность ДНК-мишени, с собранным *in vitro* рибонуклеотидным комплексом RGN;

причем одна или несколько направляющих РНК гибридизируются с последовательностью ДНК-мишени, тем самым направляя указанный полипептид RGN на связывание с указанной последовательностью ДНК-мишени.

30 71. Способ по п. 70, отличающийся тем, что указанный полипептид RGN или указанная направляющая РНК дополнительно содержит выявляемую метку, что позволяет обнаруживать указанную последовательность ДНК-мишени.

72. Способ по п. 70, отличающийся тем, что указанная направляющая РНК или указанный полипептид RGN дополнительно содержит модулятор

экспрессии, что позволяет модулировать экспрессию указанной последовательности ДНК-мишени.

73. Способ расщепления и/или модификации последовательности ДНК-мишени молекулы ДНК, включающий приведение молекулы ДНК в контакт с:

5 а) полипептидом РНК-направляющей нуклеазы (RGN), причем указанная RGN содержит аминокислотную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110 или 117; и

10 б) одной или несколькими направляющими РНК, способными нацеливать RGN (а) на последовательность ДНК-мишени;

причем одна или несколько направляющих РНК гибридизируются с последовательностью ДНК-мишени, тем самым направляя указанный полипептид RGN на связывание с указанной последовательностью ДНК-мишени, и происходит расщепление и/или модификация указанной последовательности ДНК-мишени.

74. Способ по п. 73, отличающийся тем, что расщепление указанным полипептидом RGN приводит к образованию двухцепочечного разрыва.

75. Способ по п. 73, отличающийся тем, что расщепление указанным полипептидом RGN приводит к образованию одноцепочечного разрыва.

20 76. Способ по п. 73, отличающийся тем, что указанный полипептид RGN является неактивной нуклеазой.

77. Способ по п.п. 73-76, отличающийся тем, что указанный полипептид RGN функционально связан с полипептидом, редактирующим основания.

25 78. Способ по п. 77, отличающийся тем, что указанный полипептид, редактирующий основания, включает дезаминазу.

79. Способ по п. 78, отличающийся тем, что указанная дезаминаза представляет собой цитидиндезаминазу или аденозиндезаминазу.

30 80. Способ по п.п. 73-79, отличающийся тем, что указанная модифицированная последовательность ДНК-мишени включает инсерцию гетерологичной ДНК в последовательность ДНК-мишени.

81. Способ по п.п. 73-79, отличающийся тем, что указанная модифицированная последовательность ДНК-мишени включает делецию по меньшей мере одного нуклеотида в последовательности ДНК-мишени.

82. Способ по п.п. 73-79, отличающийся тем, что указанная модифицированная последовательность ДНК-мишени включает мутацию по меньшей мере одного нуклеотида в последовательности ДНК-мишени.

5 83. Способ по п.п. 70-82, отличающийся тем, что указанная gРНК включает повторяющуюся последовательность CRISPR, содержащую нуклеотидную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111 или 118.

84. Способ по п.п. 70-83, отличающийся тем, что указанная gРНК включает tracrРНК.

10 85. Способ по п. 84, отличающийся тем, что указанная tracrРНК содержит нуклеотидную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112 или 119.

86. Способ по п.п. 84-85, отличающийся тем, что указанная gРНК является однонаправляющей РНК (sgРНК).

15 87. Способ по п.п. 84-85, отличающийся тем, что указанная gРРНК является двунаправляющей РНК.

88. Способ по п.п. 63-87, отличающийся тем, что указанная последовательность ДНК-мишени примыкает к мотиву, в свою очередь примыкающему к протоспейсеру (РАМ – protospacer adjacent motif).

20 89. Способ по п.п. 63-88, отличающийся тем, что ДНК-мишень находится внутри клетки.

90. Способ по п. 89, отличающийся тем, что клетка является эукариотической клеткой.

25 91. Способ по п. 90, отличающийся тем, что эукариотическая клетка является растительной клеткой.

92. Способ по п. 90, отличающийся тем, что эукариотическая клетка является клеткой млекопитающих.

93. Способ по п. 90, отличающийся тем, что эукариотическая клетка является клеткой насекомого.

30 94. Способ по п. 89, отличающийся тем, что клетка является прокариотической клеткой.

95. Способ по п.п. 89-94, дополнительно включающий культивирование клетки в условиях, в которых полипептид RGN экспрессируется и расщепляет последовательность ДНК-мишени с получением молекулы ДНК, содержащей

модифицированную последовательность ДНК; и выбор клетки, содержащей указанную модифицированную последовательность ДНК-мишени.

96. Клетка, содержащая модифицированную последовательность ДНК-мишени в соответствии со способом по п. 95.

5 97. Клетка по п. 96, отличающаяся тем, что клетка является эукариотической клеткой.

98. Клетка по п. 97, отличающаяся тем, что эукариотическая клетка является растительной клеткой.

99. Растение, содержащее клетку по п. 98.

10 100. Семя, содержащее клетку по п. 98.

101. Клетка по п. 97, отличающаяся тем, что эукариотическая клетка является клеткой млекопитающего.

102. Клетка по п. 98, отличающаяся тем, что эукариотическая клетка является клеткой насекомого.

15 103. Клетка по п. 96, отличающаяся тем, что клетка является прокариотической клеткой.

104. Способ получения генетически модифицированной клетки с коррекцией причинной мутации генетически наследуемого заболевания, включающий введение в клетку:

20 а) полипептида РНК-направляемой нуклеазы (RGN), причем полипептид RGN содержит аминокислотную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110 или 117; или полинуклеотида, кодирующего указанный полипептид RGN, причем указанный полинуклеотид, кодирующий полипептид RGN,
25 функционально связан с промотором для обеспечения экспрессии полипептида RGN в клетке; и

б) направляющей РНК (gРНК), причем gРНК содержит повторяющуюся последовательность CRISPR, содержащую нуклеотидную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 2, 10, 17, 24, 31, 39, 47,
30 55, 62, 70, 76, 83, 90, 96, 104, 111 или 118, или полинуклеотид, кодирующий указанную gРНК, причем указанный полинуклеотид, кодирующий gРНК, функционально связан с промотором для обеспечения экспрессии полипептида gРНК в клетке,

при этом RGN и gPHK нацелены на местоположение в геноме причинной мутации и модифицируют геномную последовательность, чтобы удалить причинную мутацию.

5 105. Способ по п. 104, отличающийся тем, что RGN функционально связана с полипептидом, редактирующим основания.

106. Способ по п. 105, отличающийся тем, что полипептидом, редактирующим основания, является дезаминаза.

107. Способ по п. 106, отличающийся тем, что дезаминаза является цитидиндезаминазой или аденозиндезаминазой.

10 108. Способ по п.п. 104-107, отличающийся тем, что клетка является клеткой животного.

109. Способ по п.п. 104-107, отличающийся тем, что клетка является клеткой млекопитающего.

15 110. Способ по п.п. 108, отличающийся тем, что клетка происходит от собаки, кошки, мыши, крысы, кролика, лошади, коровы, свиньи или человека.

111. Способ по п. 108, отличающийся тем, что наследственное генетическое заболевание вызывается полиморфизмом одного нуклеотида.

112. Способ по п. 111, отличающийся тем, что наследственное генетическое заболевание является синдромом Гурлера.

20 113. Способ по п. 112, отличающийся тем, что gPHK дополнительно содержит спейсерную последовательность, нацеленную на область, проксимальную к причинному однонуклеотидному полиморфизму.

25 114. Способ получения генетически модифицированной клетки с делецией в вызывающей заболевание области нестабильности генома, включающий введение в клетку:

а) полипептида РНК-направляемой нуклеазы (RGN), причем полипептид RGN содержит аминокислотную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110 или 117; или полинуклеотида, кодирующего указанный полипептид RGN, причем указанный полинуклеотид, кодирующий полипептид RGN, функционально связан с промотором для обеспечения экспрессии полипептида RGN в клетке; и

б) направляющей РНК (gPHK), причем gPHK содержит повторяющуюся последовательность CRISPR, содержащую нуклеотидную последовательность,

которая по меньшей мере на 95% идентична SEQ ID NO: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111 или 118, или полинуклеотида, кодирующего указанную gРНК, причем указанный полинуклеотид, кодирующий gРНК, функционально связан с промотором для обеспечения экспрессии gРНК в клетке, и, кроме того, gРНК содержит спейсерную последовательность, нацеленную на 5'-фланг геномной области нестабильности; и

в) второй направляющей РНК (gРНК), причем gРНК содержит повторяющуюся последовательность CRISPR, содержащую нуклеотидную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111 или 118, или полинуклеотида, кодирующего указанную gРНК, причем указанный полинуклеотид, кодирующий gРНК, функционально связан с промотором для обеспечения экспрессии второй gРНК в клетке, и, кроме того, указанная вторая gРНК содержит спейсерную последовательность, нацеленную на 3'-фланг геномной области нестабильности;

причем RGN и две gРНК нацелены на область нестабильности в геноме и по меньшей мере часть области нестабильности в геноме удаляют.

115. Способ по п. 114, отличающийся тем, что клетка является клеткой животного.

116. Способ по п. 114, отличающийся тем, что клетка является клеткой млекопитающего.

117. Способ по п. 115, отличающийся тем, что клетка происходит от собаки, кошки, мыши, крысы, кролика, лошади, коровы, свиньи или человека.

118. Способ по п. 115, отличающийся тем, что наследственное генетическое заболевание является атаксией Фридрейха или болезнью Гентингтона.

119. Способ по п. 118, отличающийся тем, что первая gРНК дополнительно содержит спейсерную последовательность, нацеленную на область внутри нестабильной области генома или вблизи нее.

120. Способ по п. 118, отличающийся тем, что вторая gРНК дополнительно содержит спейсерную последовательность, нацеленную на область внутри нестабильной области генома или вблизи нее.

121. Способ получения генетически модифицированной гематопозитической клетки-предшественницы млекопитающего со сниженной

экспрессией мРНК и белка BCL11A, включающий введение в выделенную гематопозитическую клетку-предшественницу человека:

а) полипептида РНК-направляемой нуклеазы (RGN), где полипептид RGN содержит аминокислотную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110 или 117; или полинуклеотида, кодирующего указанный полипептид RGN, причем указанный полинуклеотид, кодирующий полипептид RGN, функционально связан с промотором для обеспечения экспрессии полипептида RGN в клетке; и

б) направляющей РНК (gРНК), причем gРНК содержит повторяющуюся последовательность CRISPR, содержащую нуклеотидную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111 или 118, или полинуклеотида, кодирующего указанную gРНК, причем указанный полинуклеотид, кодирующий gРНК, функционально связан с промотором для обеспечения экспрессии gРНК в клетке,

причем RGN и gРНК экспрессируются в клетке и расщепляются в области энхансера BCL11A, что приводит к генетической модификации гематопозитической клетки-предшественница человека и снижению экспрессии мРНК и/или белка BCL11A.

122. Способ по п. 121, отличающийся тем, что gРНК дополнительно содержит спейсерную последовательность, которая нацелена на область внутри области энхансера BCL11A или вблизи нее.

123. Способ по п.п. 104-122, отличающийся тем, что направляющая РНК содержит tracrРНК, включающую нуклеотидную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112 или 119.

124. Система по связыванию целевой последовательности ДНК-мишени молекулы ДНК, содержащая:

а) одну или несколько направляющих РНК, способных гибридизироваться с указанной последовательностью ДНК-мишени, или одним или несколькими полинуклеотидами, содержащими одну или несколько нуклеотидных последовательностей, кодирующих одну или несколько направляющих РНК (gРНК); и

б) полипептид РНК-направляемой нуклеазы (RGN), содержащий аминокислотную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110 или 117;

5 причем одна или несколько направляющих РНК способны гибридизироваться с последовательностью ДНК-мишени, и

 причем одна или несколько направляющих РНК способны образовывать комплекс с полипептидом RGN, чтобы направить указанный полипептид RGN на связывание с указанной последовательностью ДНК-мишени молекулы ДНК.

10 125. Система по п. 124, отличающаяся тем, что указанный полипептид RGN является неактивной нуклеазой или способен функционировать в качестве никазы.

 126. Система по п.п. 124 или 125, отличающаяся тем, что указанный полипептид RGN функционально слит с полипептидом, редактирующим основания.

15 127. Система по п. 126, отличающаяся тем, что полипептид, редактирующий основания, представляет собой дезаминазу.

 128. Система по п. 127, отличающаяся тем, что дезаминаза является цитидиндезаминазой или аденозиндезаминазой.

20 129. Способ обнаружения последовательности ДНК-мишени молекулы ДНК в образце, включающий:

а) контактирование образца с:

 i) полипептидом РНК-направляемой нуклеазы (RGN), содержащим аминокислотную последовательность, которая по меньшей мере на 25 95% идентична SEQ ID NO: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 117 или 137, причем указанный полипептид RGN способен связывать указанную последовательность ДНК-мишени молекулы ДНК специфическим для РНК-направляемой последовательности образом при связывании с направляемой РНК, способной гибридизироваться с указанной последовательностью ДНК-мишени;

 ii) указанной направляющей РНК; и

 iii) детекторной одноцепочечной ДНК (оцДНК), которая не гибридизуется с направляющей РНК; и

б) измерение детектируемого сигнала, создаваемого путем расщепления детекторной оцДНК с помощью RGN, тем самым обнаруживая ДНК-мишень.

130. Способ по п. 129, отличающийся тем, что указанный образец содержит молекулы ДНК из лизата клеток.

5 131. Способ по п. 129, отличающийся тем, что указанный образец содержит клетки.

132. Способ по п. 131, отличающийся тем, что указанные клетки представляют собой эукариотические клетки.

10 133. Способ по п. 129, отличающийся тем, что молекулу ДНК, содержащую последовательность ДНК-мишени, получают путем обратной транскрипции молекулы-матрицы РНК, присутствующей в образце, содержащем РНК.

134. Способ по п. 133, отличающийся тем, что молекула-матрица РНК является РНК-содержащим вирусом.

15 135. Способ по п. 134, отличающийся тем, что РНК-содержащий вирус является коронавирусом.

136. Способ по п. 135, отличающийся тем, что коронавирус является SARS-подобным коронавирусом летучих мышей, SARS-CoV или SARS-CoV-2.

20 137. Способ по п.п. 133-136, отличающийся тем, что образец, содержащий РНК, получают из образца, содержащего клетки.

138. Способ по п.п. 129-137, отличающийся тем, что указанная детекторная оцДНК содержит пару флуорофор/гаситель.

25 139. Способ по п.п. 129-137, отличающийся тем, что указанная детекторная оцДНК содержит пару флуоресцентного резонансного переноса энергии (FRET – fluorescence resonance energy transfer).

30 140. Способ по п.п. 129-139, отличающийся тем, что указанная направляющая РНК содержит повторяющуюся последовательность CRISPR, состоящую из нуклеотидной последовательности, которая по меньшей мере на 95% идентична SEQ ID NO: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 118 или 273.

141. Способ по п.п. 129-140, отличающийся тем, что указанная направляющая РНК содержит tracrРНК.

142. Способ по п. 141, отличающийся тем, что указанная tracrРНК содержит нуклеотидную последовательность, которая по меньшей мере на 95%

идентична SEQ ID NO: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112, 119 или 274.

143. Способ по п.п. 141-142, отличающийся тем, что указанная направляющая РНК является однонаправляющей РНК (sgРНК).

5 144. Способ по п.п. 141-142, отличающийся тем, что указанная направляющая РНК является двунаправляющей РНК.

145. Способ по п.п. 129-144, отличающийся тем, что указанный способ дополнительно содержит амплификацию нуклеиновых кислот в образце до стадии контактирования а) или одновременно с ней.

10 146. Способ по п. 145, отличающийся тем, что амплификацию молекулы ДНК осуществляют путем обратной транскрипции молекулы РНК.

147. Набор для обнаружения последовательности ДНК-мишени молекулы ДНК в образце, включающий:

а) полипептид РНК-направляющей нуклеазы (RGN), включающий
15 аминокислотную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 117 или 137, причем указанный полипептид RGN способен связывать указанную последовательность ДНК-мишени молекулы ДНК специфическим для последовательности РНК-направляемым образом, способной гибридизироваться с указанной последовательностью ДНК-мишени;

б) указанную направляющую РНК; и

20 в) детекторную одноцепочечную ДНК (оцДНК), которая не гибридизуется с направляющей РНК.

148. Набор по п. 147, отличающийся тем, что указанная одноцепочечная
25 ДНК детектора содержит пару флуорофор/гаситель.

149. Набор по п. 147, отличающийся тем, что указанная детекторная оцДНК содержит пару флуоресцентного резонансного переноса энергии (FRET – fluorescence resonance energy transfer).

150. Набор по п.п. 147-149, отличающийся тем, что указанная
30 направляющая РНК содержит повторяющуюся последовательность CRISPR, включающую нуклеотидную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 118 или 273.

151. Набор по п.п. 147-150, отличающийся тем, что указанная направляющая РНК включает tracrРНК.

152. Набор по п. 151, отличающийся тем, что указанная tracrРНК содержит нуклеотидную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112, 119 или 274.

153. Набор по п.п. 151-152, отличающийся тем, что указанная направляющая РНК является однонаправляющей РНК (sgРНК).

154. Набор по п.п. 151-152, отличающийся тем, что указанная направляющая РНК является двунаправляющей РНК.

155. Способ расщепления одноцепочечных ДНК, включающий контактирование популяции нуклеиновых кислот, причем указанная популяция содержит молекулу ДНК, содержащую целевую последовательность ДНК и множество нецелевых оцДНК, с:

а) полипептидом РНК-направляемой нуклеазы (RGN), включающим аминокислотную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 117 или 137, причем указанный полипептид RGN способен связывать указанную последовательность ДНК-мишени специфическим для РНК-направляемой последовательностью образом при связывании с направляющей РНК, способной гибридизироваться с указанной последовательностью ДНК-мишени; и

б) указанной направляющей РНК;

при этом полипептид RGN расщепляет нецелевые оцДНК из указанного множества.

156. Способ по п. 155, отличающийся тем, что указанная популяция нуклеиновых кислот находится в лизате клеток.

157. Способ по п.п. 155-156, отличающийся тем, что указанная направляющая РНК содержит повторяющуюся последовательность CRISPR, включающую нуклеотидную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 118 или 273.

158. Способ по п.п. 155-157, отличающийся тем, что указанная направляющая РНК содержит tracrРНК.

159. Способ по п. 158, отличающийся тем, что указанная tracrPHK содержит нуклеотидную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112, 119 или 274.

5 160. Способ по п.п. 158-159, отличающийся тем, что указанная направляющая РНК является однонаправляющей РНК (sgPHK).

161. Способ по п.п. 158-159, отличающийся тем, что указанная направляющая РНК является двунаправляющей РНК (dgPHK).

10 162. Молекула нуклеиновой кислоты, включающая полинуклеотид, кодирующий CRISPR РНК (crPHK), причем указанная crPHK содержит спейсерную последовательность и повторяющуюся последовательность CRISPR, где указанная повторяющаяся последовательность CRISPR содержит нуклеотидную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 240.

15 где направляющая РНК включает:

а) указанную crPHK; и необязательно,

б) трансактивирующую CRISPR РНК (tracrPHK), способную гибридизироваться с указанной повторяющейся последовательностью CRISPR указанной crPHK;

20 способная гибридизироваться с последовательностью ДНК-мишени молекулы ДНК специфичным для последовательности образом через спейсерную последовательность указанной crPHK, когда указанная направляющая РНК связана с полипептидом РНК-направляемой нуклеазы (RGN),

25 и где указанный полинуклеотид, кодирующий crPHK, функционально связан с промотором, гетерологичным указанному полинуклеотиду.

163. Вектор, включающий молекулу по п. 162.

164. Вектор по п. 163, отличающийся тем, что указанный вектор дополнительно содержит полинуклеотид, кодирующий указанную tracrPHK.

30 165. Вектор по п. 164, отличающийся тем, что указанная tracrPHK содержит нуклеотидную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 241.

166. Вектор по п.п. 164-165, отличающийся тем, что указанный полинуклеотид, кодирующий указанную crPHK, и указанный полинуклеотид,

кодирующий указанную tracrPHK, функционально связаны с одним и тем же промотором и кодируются как единая направляющая РНК.

167. Вектор по п.п. 164-165, отличающийся тем, что указанный полинуклеотид, кодирующий указанную crPHK, и указанный полинуклеотид, кодирующий указанную tracrPHK, функционально связаны с отдельными промоторами.

168. Вектор по п.п. 163-167, отличающийся тем, что указанный вектор дополнительно содержит полинуклеотид, кодирующий указанный полипептид RGN, причем указанный полипептид RGN содержит аминокислотную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 235.

169. Молекула нуклеиновой кислоты, содержащая полинуклеотид, кодирующий транскрибируемую CRISPR РНК (tracrPHK), включающую нуклеотидную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 241;

где направляющая РНК содержит:

а) указанную tracrPHK; и

б) crPHK, содержащую спейсерную последовательность и последовательность повтора CRISPR, причем указанная tracrPHK способна гибридизироваться с указанной последовательностью повтора CRISPR указанной crPHK;

способная гибридизоваться с последовательностью ДНК-мишени молекулы ДНК специфичным для последовательности образом через спейсерную последовательность указанной crPHK, когда указанная направляющая РНК связана с полипептидом РНК-направляемой нуклеазы (RGN), и

где указанный полинуклеотид, кодирующий tracrPHK, функционально связан с промотором, гетерологичным указанному полинуклеотиду.

170. Вектор, содержащий молекулу нуклеиновой кислоты по п. 169.

171. Вектор по п. 170, отличающийся тем, что указанный вектор дополнительно содержит полинуклеотид, кодирующий указанную crPHK.

172. Вектор по п. 171, отличающийся тем, что повторяющаяся последовательность CRISPR указанной crPHK содержит нуклеотидную

последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 240.

173. Вектор по п.п. 171-172, отличающийся тем, что указанный полинуклеотид, кодирующий указанную crPHK, и указанный полинуклеотид, кодирующий указанную tracrPHK, функционально связаны с одним и тем же промотором и кодируются как единая направляющая РНК.

174. Вектор по п.п. 171-172, отличающийся тем, что указанный полинуклеотид, кодирующий указанную crPHK, и указанный полинуклеотид, кодирующий указанную tracrPHK, функционально связаны с отдельными промоторами.

175. Вектор по п.п. 170-174, отличающийся тем, что указанный вектор дополнительно содержит полинуклеотид, кодирующий указанный полипептид RGN, где указанный полипептид RGN содержит аминокислотную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 235.

Приводимые ниже примеры предназначены для иллюстрирования настоящего изобретения, но не ограничивают рамок его охвата.

Примеры

Пример 1. Идентификация РНК-направляющих нуклеаз

Идентифицировано семнадцать различных CRISPR-ассоциированных РНК-направляемых нуклеаз (RGN), которые описаны в табл. 1 ниже. В табл. 1 приведены обозначения каждой RGN, ее аминокислотная последовательность, источник, из которого она была получена, и процессированные последовательности crPHK и tracrPHK (методы идентификации см. в примере 2). В табл. 1 также представлена общая последовательность однонаправленная РНК (sgPHK), где поли-N указывает расположение спейсерной последовательности, которая определяет последовательность-мишень нуклеиновой кислоты sgPHK. Системы RGN APG05733.1, APG06207.1, APG01647.1, APG08032.1, APG02675.1, APG01405.1, APG06250.1, APG04293.1 и APG01308.1 имеют консервативную последовательность в основании стебля шпильки tracrPHK UNANNA (SEQ ID NO: 8). Для APG05712.1 последовательность в том же месте представляет собой CNANNG (SEQ ID NO: 37). Для APG01658.1 последовательность в том же месте представляет собой CNANNU (SEQ ID NO: 45). Для систем RGN APG06498.1 и APG06877.1 консервативной последовательностью в основании стебля шпильки

tracrPHK является UNANNG (SEQ ID NO: 53). Для APG09882.1 и APG06646.1 последовательность в том же месте представляет собой UNANNC (SEQ ID NO: 68). Для APG09053.1 последовательность в том же месте представляет собой CNANNU (SEQ ID NO: 102).

5 Таблица 1. Краткое описание SEQ ID и CRISPR-ассоциированных систем

RGN ID	SEQ ID NO	Источник	Повторяющаяся последовательность crPHK (SEQ ID NO)	tracrPHK (SEQ ID NO)	sgPHK (SEQ ID NO)
APG05733.1	1	<i>Bacillus</i> sp.	2	3	4
APG06207.1	9	<i>Chryseobacterium</i> sp.	10	11	12
APG01647.1	16	<i>Sphingobacterium</i> sp.	17	18	19
APG08032.1	23	<i>Chryseobacterium</i> sp.	24	25	26
APG05712.1	30	<i>Bruceella</i> sp.	31	32	33
APG01658.1	38	<i>Staphylococcus</i> sp.	39	40	41
APG06498.1	46	<i>Bacillus</i> sp.	47	48	49
APG09106.1	54	<i>Brevibacillus</i> sp.	55	56	57
APG09882.1	61	<i>Enterococcus</i> sp.	62	63	64
APG02675.1	69	<i>Sphingobacterium</i> sp.	70	71	72
APG01405.1	75	<i>Chryseobacterium</i> sp.	76	77	78
APG06250.1	82	<i>Chryseobacterium</i> sp.	83	84	85
APG06877.1	89	<i>Bacillus</i> sp.	90	91	92
APG09053.1	95	<i>Rhizobium</i> sp.	96	97	98
APG04293.1	103	<i>Myroides</i> sp.	104	105	106
APG01308.1	110	<i>Chryseobacterium</i> sp.	111	112	113
APG06646.1	117	<i>Bacillus</i> sp.	118	119	120

Пример 2. Идентификация направляющей РНК и конструкции sgPHK

10 Культуры бактерий, которые нативно экспрессируют исследуемую РНК-направляемую нуклеазную систему, выращивают до средней логарифмической фазы (ОП600 примерно ~0,600), осаждают и быстро замораживают. РНК
15 выделяют из осадков с помощью набора для выделения миРНК miRVANA (фирма Life Technologies, Карлсбад, Калифорния) и библиотеки секвенирования готовят из выделенной РНК с помощью набора для подготовки библиотеки малых РНК NEBNext (фирма NEB, Беве́рли, Массачусетс). Препарат библиотеки
фракционируют в 6% полиакриламидном геле для захвата видов РНК размером менее 200 нуклеотидов для обнаружения crPHK и tracrPHK, соответственно. Глубокое секвенирование (парный конец 75 п.н.) выполняют с помощью Next Seq 500 (комплект высокой производительности) от поставщика фирмы MoGene,

Сент-Луис, Миссури. Прочтения обрабатывают качественно по программе Cutadapt и картируют по эталонным геномам с помощью команды Bowtie2. На языке питон был разработан собственный источник информации РНКseq для обнаружения транскриптов sgРНК и tracrРНК. Границы процессированных sgРНК определяют по покрытию последовательностей массива нативных повторяющихся спейсеров. Анти-повторяющаяся часть tracrРНК идентифицирована с использованием рекомендованных параметров BLASTn. Глубина секвенирования РНК подтверждает границы процессируемой tracrРНК путем идентификации транскрипта, содержащего анти-повтор. Ручное курирование РНК выполняют с использованием прогноза по вторичной структуре с помощью NUPACK, программного обеспечения для укладки РНК. Кассеты sgРНК получают путем синтеза ДНК и обычно разрабатывают следующим образом (5'→3'): спейсерная последовательность длиной 20-30 п.о., функционально связанная с 3-конца с частью процессированного повтора sgРНК, функционально связанной с 4 п.о. некомплементарного линкера (AAAG; SEQ ID NO: 123), функционально связанного с его 3'-конца с процессированной tracrРНК. Также можно использовать другие некомплементарные линкеры длиной 4 п.н.

Для анализов *in vitro* sgРНК синтезируют путем транскрипции *in vitro* кассет sgРНК с помощью набора для синтеза gРНК GeneArt™ Precision (фирма ThermoFisher). Процессированные последовательности sgРНК и tracrРНК для каждого из полипептидов RGN идентифицированы и представлены в табл. 1. См. ниже sgРНК, сконструированные для библиотек PAM 1 и 2.

Пример 3. Определение потребностей PAM для каждой RGN

Потребности PAM для каждой RGN определяют, используя метод истощения PAM, особенно адаптированный на основе публикаций Kleinstiver с соавт., *Nature* 2015, 523:481-485, и Zetsche с соавт., *Cell*, 2015, 163:759-771. Вкратце, две плазмидные библиотеки (L1 и L2) создают в каркасе плазмиды Puc18 (ampR), причем каждая из них содержит отдельную последовательность протоспейсера (мишень) из 30 п.н., окруженную 8 случайными нуклеотидами (т.е., область PAM). Целевая последовательность и фланкирующая область PAM библиотеки 1 и библиотеки 2 для каждого RGN указаны ниже в табл. 2.

Библиотеки отдельно путем электропорации внедряют в клетки *E. coli* BL21(DE3), несущие векторы экспрессии Prsf-1b, содержащие RGN по

настоящему изобретению (оптимизация по кодонам для *E. coli*) вместе с родственной sgРНК, содержащей спейсерную последовательность, соответствующую протоспейсеру в L1 или L2. Достаточное количество библиотечной плазмиды используют в реакции трансформации для получения >10⁶ КОЕ. Как RGN, так и sgРНК в каркасе вектора Prsf-1b находятся под контролем промоторов фага T7. Реакции трансформации восстанавливают в течение 1 ч, после чего разводят в среде LB, содержащей карбенициллин и канамицин, и выращивают в течение ночи. На следующий день смесь разводили в самоиндуцирующейся среде Overnight Express™ Instant TB (фирма Millipore Sigma), чтобы обеспечить экспрессию RGN и sgРНК, и выращивают в течение дополнительных 4 или 20 ч, после чего клетки осаждают центрифугированием и выделяют плазмидную ДНК с использованием набора Mini-prep kit (фирма Qiagen, Джермантаун, Мэриленд). В присутствии соответствующих sgРНК плазмиды, содержащие РАМ, распознаваемый RGN, расщепляются, что приводит к их удалению из популяции. Плазмиды, содержащие РАМ, не распознаваемые RGN или трансформированные в бактерии, не содержащие соответствующей sgРНК, выживают и реплицируются. РАМ и протоспейсерные области нерасщепленных плазмид амплифицируют методом ПЦР и готовят для секвенирования в соответствии с опубликованными протоколами (руководство по подготовке 16s-метагеномной библиотеки 15044223В, фирма Illumina, Сан-Диего, Калифорния).

Глубокое секвенирование (считывание с одного конца 75 п.н.) выполняют на секвенаторе MiSeq (фирма Illumina) поставщиком услуг (фирма MoGene, Сент-Луис, Миссури). Обычно на ампликон приходится 1-4 млн прочтений. Области РАМ выделяют, подсчитывают и нормализуют по общему количеству прочтений для каждого образца. РАМ, приводящие к расщеплению плазмиды, идентифицируют как не выявленные по сравнению с контролем (т.е., когда библиотека трансформируется в *E. coli*, содержащей RGN, но не имеющей соответствующей sgРНК). Чтобы представить потребности РАМ для новой RGN, коэффициенты истощения (частота в образце/частота в контроле) для всех последовательностей в рассматриваемой области преобразуют в значения обогащения с преобразованием по логарифму с основанием 2. Достаточные РАМ были определены как те, у которых значения обогащения > 2,3 (что соответствует коэффициентам истощения < ~ 0,2). РАМ выше этого порога в

обеих библиотеках собирают и используют для создания веб-логотипов, которые, например, могут быть созданы с помощью веб-службы в интернете, известной как «weblogo». Последовательности РАМ идентифицируют и сообщают, когда в наиболее обогащенных РАМ наблюдают постоянную картину. Консенсусные последовательности РАМ (с фактором обогащения (EF – enrichment factor) >2,3) для каждой RGN представлены в табл. 2. Ориентация РАМ также указана в табл. 2. В настоящем описании нуклеазы APG06646.1 и APG04293.1 не обладают РАМ-взаимодействующими доменами. Результаты в табл. 2 показывают, что они также не обладают типичным качеством РАМ, а именно 2-5 нуклеотидами. Было показано, что для расщепления нуклеазам APG06646.1 и APG04293.1 требуются одиночные нуклеотиды.

Таблица 2. РАМ или РАМ-подобное определение

RGN ID	sgPHK L1 (SEQ ID NO)	sgPHK L2 (SEQ ID NO)	РАМ (SEQ ID NO)	Ориентация РАМ
APG05733.1	5	6	7	5'-мишень-РАМ-3'
APG06207.1	13	14	15	5'-мишень-РАМ-3'
APG01647.1	20	21	22	5'-мишень-РАМ-3'
APG08032.1	27	28	29	5'-мишень-РАМ-3'
APG05712.1	34	35	36	5'-мишень-РАМ-3'
APG01658.1	42	43	44	5'-мишень-РАМ-3'
APG06498.1	50	51	52	5'-мишень-РАМ-3'
APG09106.1	58	59	60	5'-РАМ-мишень-3'
APG09882.1	65	66	67	5'-мишень-РАМ-3'
APG02675.1	73	74	15	5'-мишень-РАМ-3'
APG01405.1	79	80	81	5'-мишень-РАМ-3'
APG06250.1	86	87	88	5'-мишень-РАМ-3'
APG06877.1	93	94	7	5'-мишень-РАМ-3'
APG09053.1	99	100	101	5'-мишень-РАМ-3'
APG04293.1	107	108	109	5'-мишень-РАМ-3'
APG01308.1	114	115	116	5'-мишень-РАМ-3'
APG06646.1	121	122	109	5'-мишень-РАМ-3'

Пример 4. Конструирование направляющей РНК для повышения нуклеазной активности

4.1 RGN APG09748 и APG09106.1

Для RGN APG09748 (указан как SEQ ID NO: 137, а повторяющаяся последовательность sgPHK APG09748, последовательность tracrPHK и общая последовательность sgPHK указаны как SEQ ID No: 273, 274 и 275,

соответственно; все последовательности описаны в международной заявке на патент PCT/US2019/068079, который полностью включен в настоящее описание в качестве ссылки) и APG09106.1, которые имеют очень высокую идентичность последовательностей и имеют одинаковые РАМ, прогнозы по сворачиванию РНК используют для определения областей в направляющей РНК, которые могут быть изменены для оптимизации активности нуклеаз. Стабильность пары оснований *cg*РНК:*tracr*РНК в области повтор:антиповтор увеличена за счет укорочения области повтор:антиповтор, добавления пар оснований G-C и удаления неоднозначных пар G-U. «Оптимизированные» направляющие варианты тестируют и сравнивают с *g*РНК дикого типа с использованием RGN APG09748 в анализах расщепления *in vitro*.

Для получения RGN для образования RNP конструируют экспрессирующие плазмиды, содержащие RGN, слитые с С-концевой меткой His6 (SEQ ID NO: 276) или His10 (SEQ ID NO: 277), и трансформируют в штаммы BL21 (DE3) *E. coli*. Экспрессию проводят с использованием среды Magic Media (фирма Thermo Fisher) с добавлением 50 мкг/мл канамицина. После лизиса и осветления белок очищают с помощью аффинной хроматографии с иммобилизованным металлом и количественно определяют с использованием набора для количественного анализа белка Qubit (фирма Thermo Fisher) или с помощью УФ-видимого излучения с использованием рассчитанного коэффициента экстинкции.

Рибонуклеопротеин (РНП) получают путем инкубации очищенной RGN с *sg*РНК в соотношении ~2:1 в течение 20 мин при комнатной температуре. Для реакций расщепления *in vitro* РНП инкубируют с плазмидами или линейной *ds*ДНК, содержащими целевой протоспейсер, фланкированный предпочтительной последовательностью РАМ, в течение > 30 минут при комнатной температуре. Протестированы две последовательности-мишени нуклеиновой кислоты в локусе TRAC, TRAC11 (SEQ ID NO: 278) и TRAC14 (SEQ ID NO: 279). Исследуют *g*РНК как на целевую активность с правильной последовательностью-мишенью нуклеиновой кислоты (например, *g*РНК имеет спейсерную последовательность TRAC11, а анализируемой мишенью является TRAC11), так и без правильной последовательности-мишени нуклеиновой кислоты (например, *g*РНК имеет спейсерную последовательность TRAC11, а анализируемая целью является TRAC14). Активность, определяемую

расщеплением плазмиды, оценивают с помощью электрофореза в агарозном геле. Результаты представлены в табл. 3.

Варианты направляющих последовательностей представлены как SEQ ID NO: 280-283 и снабжены спейсерными последовательностями. В этих направляющих последовательностях используют некомплементарный нуклеотидный линкер AAAA (SEQ ID NO: 284). Оптимизированная gPHK (SEQ ID NO: 285; поли-N указывает расположение спейсерной последовательности) с повышенным связыванием повтор: антиповтор имеет оптимизированные компоненты tracrPHK (SEQ ID NO: 286) и оптимизированную crPHK (SEQ ID NO: 287). Оптимизированный вариант направляющей способен расщеплять два локуса, где ранее не было обнаружено расщепления с использованием направляющей РНК дикого типа. За счет оптимизации гибридизации в области повтор: антиповтор расщепление *in vitro* APG09748 увеличилось с 0% расщепления до 100% расщепления для множества мишеней в локусе TRAC.

Таблица 3. Эффективность редактирования APG09748 со сконструированными направляющими вариантами

Вариант gPHK (SEQ ID NO.)	Сконструированная направляющая	Исследуемая мишень	Гель 1 – нагрузка 2 мкл		Гель 2 – нагрузка 1 мкл	
			% интактных	% расщепленных	% интактных	% расщепленных
280	Оптимизированная	TRAC11	68	32	57	43
280	Оптимизированная	TRAC14	100	0	100	0
281	Оптимизированная	TRAC11	100	0	100	0
281	Оптимизированная	TRAC14	70	30	69	31
282	Дикий тип	TRAC11	100	0	100	0
282	Дикий тип	TRAC14	100	0	100	0
283	Дикий тип	TRAC11	100	0	100	0
283	Дикий тип	TRAC14	100	0	100	0
Нет		TRAC11	100	0	100	0
Нет		TRAC14	100	0	100	0

Были разработаны и протестированы дополнительные оптимизированные варианты gPHK. Кроме того, были протестированы различные длины спейсерной последовательности, чтобы определить, как длина спейсера может влиять на эффективность расщепления. За пределами спейсерной последовательности в

этом анализе sgРНК называется «основой». В табл. 4 они обозначены как «дикий тип» (SEQ ID NO: 288, последовательность дикого типа) и три оптимизированные sgРНК: V1 (SEQ ID NO: 289), V2 (SEQ ID NO: 290) и V3 (SEQ ID NO: 291). Все эти последовательности имеют поли-N для обозначения 5 местоположения спейсерной последовательности. Направляющиеся экспрессируются в виде sgРНК с помощью транскрипции *in vitro* (IVT). По отношению к основе sgРНК дикого типа, V1 идентичен на 87,8%, V2 – на 92,4%, а V3 – на 85,5%. Также были получены и испытаны синтетические дуплексы tracrРНК:crРНК («синтетические»), представляющие РНК с двойной 10 направленностью, но в остальном сходные с указанными выше оптимизированными sgРНК дикого типа.

Для такого набора анализов используют RGN APG09106.1; в остальном способы реакций расщепления *in vitro* аналогичны описанным выше. Целевыми последовательностями нуклеиновых кислот являются мишень 1 (SEQ ID NO: 15 292) и мишень 2 (SEQ ID NO: 293). Результаты представлены в табл. 4.

Таблица 4. Эффективность редактирования APG09106.1 со сконструированными направляющими вариантами

Источник РНК	Мишень	Длина спейсера	Основа	Спейсер SEQ ID NO.	Расщепление (%)
Синтетический	2	18	Дикий тип	294	12,3
Синтетический	1	20	Дикий тип	295	0
Синтетический	2	20	Дикий тип	296	55,0
Синтетический	1	25	Дикий тип	297	0
Синтетический	2	25	Дикий тип	298	61,4
IVT	2	25	V1	299	1,1
IVT	2	25	V2	300	0,9
IVT	2	25	V3	301	0,7
IVT	2	20	V3	302	21,0
IVT	1	25	V3	303	2,0

4.2 RGN APG07433.1

20 Прогнозы по сворачиванию РНК используют для определения областей в направляющей РНК для APG07433.1 (указанной как SEQ ID NO: 235 и описанной в заявке US2019/0367949 и международной заявке WO2019/236566), каждая из которых полностью включена в настоящее изобретение в виде 25 ссылки), которые можно изменить для оптимизации активности нуклеазы и укорочения направляющей РНК для упаковки в вирусные векторы. Две области

идентифицированы как потенциальные места для внесения изменений:
 спаривание повтор:антиповтор между crPHK и областью tracrPHK и концевых
 шпилек в tracrPHK. Участок повтор:антиповтор укорочен до 7 (APG07433.1-7
 п.н., SEQ ID No: 238 и 239 п.н.), 13 (APG07433.1-13 п.н., SEQ ID No: 250 и 251
 5 п.н.) и 15 пар оснований в длину (APG07433.1-15 п.н., SEQ ID No: 242 и 243
 п.н.). Кроме того, был испытан четвертый вариант, в котором была изменена
 последовательность области повтор:антиповтор и уменьшена длина спаривания
 до 11 пар оснований (APG07433.1-11bp-syn, SEQ ID No: 240 и 241 п.н.), с
 введением менее выступающей РНК по сравнению с направляющей дикого типа.
 10 Также были укорочены концевые шпильки в tracrPHK, включая сокращение
 нативной последовательности до 40 (APG07433.1-40ntTHP, SEQ ID No: 254 и
 255) и 42 (APG07433.1-42ntTHP, SEQ ID No: 244 и 245) нуклеотидов в длину.
 Измененные структуры стебель-петля сконструированы таким образом, чтобы
 сократить их до 35 (APG07433.1-35ntTHP-syn, SEQ ID No: 248 и 249) и 39
 15 (APG07433.1-39ntTHP-syn, SEQ ID No: 246 и 247) нуклеотидов. Одна
 направляющая объединила укороченный участок повтор:антиповтор из 13 пар
 оснований и укороченную концевую шпильку из 42 нуклеотидов (APG07433.1-
 13bp42ntTH, SEQ ID No: 252 и 253). Также проверено влияние длины спейсера
 на расщепление. Кроме того, определяют в нуклеотидах длины спейсеров 25
 20 (APG07433.1-11bp-syn, SEQ ID No: 236 и 237; последовательность спейсера
 представлена как SEQ ID NO: 271) и 18 (APG07433.1-native[18], SEQ ID No: 256
 и 257; последовательность спейсера указана как SEQ ID NO: 272) в основе
 направляющей РНК дикого типа.

Следующие двунаправленные РНК получают отжигом crPHK и tracrPHK
 25 путем получения раствора, содержащего 20 мкМ crPHK и 10 мкМ tracrPHK в
 буфере для отжига (фирма Synthego), с последующим прогревом до 78°C в
 течение 10 мин, а затем при 37°C в течение 30 мин. РНП формируют путем
 инкубации с очищенным белком APG07433.1 в концентрации 0,5 мкМ и
 двунаправленной РНК в концентрации 1 мкМ в фосфатно-солевом буфере в
 30 течение 20 мин. Направляющие РНК, использованные в этом эксперименте,
 показаны в табл. 5.

Таблица 5. Двунаправленные РНК

Название двунаправленных gPHK	crPHK SEQ ID NO	tracrPHK SEQ ID NO
APG07433.1-native[25]	236	237

Название двунаправленных gPHK	crPHK SEQ ID NO	tracrPHK SEQ ID NO
APG07433.1-7bp	238	23
APG07433.1-11bp-syn	240	241
APG07433.1-15bp	242	243
APG07433.1-42ntTHP	244	245
APG07433.1-39ntTHP-syn	246	247
APG07433.1-35ntTHP-syn	248	249
APG07433.1-13bp	250	251
APG07433.1-13bp42ntTH	252	253
APG07433.1-40ntTHP	254	255
APG07433.1-native[18]	256	257
APG07433.1cr-13bp tr-35ntTHP	258	259

Эти РНП инкубируют с продуктом ПЦР, содержащим соответствующую целевую последовательность (SEQ ID NO: 260), полученную путем амплификации с использованием праймеров, меченных FAM и Cy3, для

5 облегчения точного и чувствительного количественного определения продукта расщепления. Реакции содержат 250 нМ РНП, полученного, как указано выше, и 150 нМ продукта ПЦР, меченного FAM и Cy3, в буфере 1X Cutsmart (фирма New England Biolabs). Реакцию осуществляют в течение 15 мин при 37°C и

10 заканчивают добавлением РНКазы А до 0,1 мг/мл и EDTA до 45 мМ. Погасшую реакцию нагревают при 50°C в течение 30 мин и 95°C в течение 5 мин. Затем образцы анализируют с помощью электрофореза в нативном акриламидном 5% геле TBE (фирма Bio-rad). Их визуализируют с помощью сканера ChemiDoc MP. Каждый из двух красителей можно использовать для количественного

15 определения; каждый образец выполняли в двух повторностях. 9 Результаты этого эксперимента показаны в табл. 6.

Таблица 6. Результаты расщепления

РНП	Расщепление (%)				Общее среднее
	Канал Cy3		Канал FAM		
	Повторности				
	1	2	1	2	
Только нуклеаза (без направляющей)	0	0	0	0	0
APG07433.1-native[25]	66,41	61,16	68,81	56,85	63,3
APG07433.1-7bp	0	3,38	2,83	0	1,6
APG07433.1-11bp-syn	66,33	76,38	66,73	68	69,4
APG07433.1-15bp	50,7	65,61	54,62	59,95	57,7
APG07433.1-42ntTHP	44,81	50,59	48,84	51,77	49,0
APG07433.1-39ntTHP-syn	19,16	32,88	24,91	33,43	27,6
APG07433.1-35ntTHP-syn	26,77	38,61	32,09	40,2	34,4

РНП	Расщепление (%)				Общее среднее
	Канал Су3		Канал FАM		
	Повторности				
APG07433.1-13bp	44,36	33,03	42,93	35,22	38,9
APG07433.1-13bp42ntTНP	35,91	36,74	33,84	39,2	36,4
APG07433.1-40ntTНP	44,52	57,14	48,34	57,84	52,0
APG07433.1-native[18]	11,1	9,76	12,93	13,24	11,8
APG07433.1cr-13bp tr-35ntTНP	10,66	12,25	12,65	13,54	12,3

Этот анализ показывает, что нативная основа превосходит большинство укороченных вариантов и вариант, содержащий укороченную целевую последовательность (APG07433.1-native[18], SEQ ID NO: 256 и 257). Из укороченных последовательностей скелета наиболее высоким уровнем расщепления является последовательность, обозначаемая APG07433.1-11bp-syn, которая содержит целевую последовательность длиной 25 нуклеотидов (спейсер; SEQ ID NO: 271); последовательность crPHK, представляемая SEQ ID NO: 240; последовательность tracrPHK, представляемая SEQ ID NO: 241. Такой вариант направляющей включает измененную петлю стебля повтор:антиповтор со сконструированной выпуклостью на стебле.

Пример 5. Демонстрация активности редактирования генов в клетках млекопитающих

Кассеты экспрессии RGN получают и вводят в векторы для экспрессии в млекопитающих. Каждая RGN APG09748, APG09106.1, APG05712.1, APG01658.1, APG05733.1, APG06498.1, APG06646.1, APG09882.1, APG01405.1 и APG01308.1 оптимизирована по кодонам для экспрессии в млекопитающих (SEQ ID NO: 304, 305 и 411-418, соответственно), и экспрессированные белки функционально слиты на N-конце с последовательностью ядерной локализации NLS вируса SV40 (NLS– nuclear localization sequence; SEQ ID NO: 125) и метками 3Xflag (SEQ ID NO : 126), а также функционально слиты на C-конце с последовательностями NLS нуклеоплазмина (SEQ ID NO: 127). Используют две копии последовательности NLS, функционально слитые в тандеме. Каждая кассета экспрессии находится под контролем промотора цитомегаловируса (CMV) (SEQ ID NO: 306). В данной области техники известно, что энхансер транскрипции CMV (SEQ ID NO: 307) также может быть включен в конструкции, содержащие промотор CMV. Конструкции экспрессии направляющей РНК, каждая из которых кодирует одну gPHK под контролем

промотора U6 РНК-полимеразы III человека (SEQ ID NO: 308), получают и вводят в вектор экспрессии. Направляющие целевые области выбранных генов, включающие RelA, AurkB, GAPDH, LINC01509, HBB, CFTR, HPRT1, TRA, EMX1 и VEGFA, показаны в табл. 7. Для РНК-направляемой нуклеазы APG09106.1 специфические остатки мутированы для увеличения нуклеазной активностью белка, в частности остаток T849 в APG09106 мутирован и заменен на аргинин (SEQ ID NO: 309). Такая точечная мутация увеличила степень редактирования в клетках млекопитающих.

Описанные выше конструкции вводят в клетки млекопитающих. За сутки до трансфекции 1×10^5 клеток HEK293T (фирма Sigma) высевают в 24-луночные планшеты в среду Игла в модификации Дульбекко (DMEM) с добавлением 10 об.% эмбриональной сыворотки быка (фирма Gibco) и 1% препарата пенициллин-стрептомицин (фирма Gibco). На следующий день, когда клетки достигают 50-60% слияния, 500 нг экспрессирующей RGN плазмиды плюс 500 нг плазмиды gРНК совместно трансфецируют с использованием 1,5 мкл липофектамина 3000 (фирма Thermo Scientific) на лунку по инструкции производителя. Через 48 ч роста собирают общую геномную ДНК с использованием набора для выделения геномной ДНК (фирма Machery-Nagel) в соответствии с инструкциями производителя.

Анализируют общую геномную ДНК для определения степени редактирования целевого гена. Получены олигонуклеотиды для использования в ПЦР-амплификации и последующем анализе амплифицированного геномного сайта-мишени (SEQ ID NO: 310 и 311). Все реакции ПЦР проводят с использованием 10 мкл реактива 2X Master Mix Phusion High-Fidelity ДНК-полимеразы (фирма Thermo Scientific) в 20 мкл реакции, включая 0,5 мкМ каждого праймера. Большие геномные области, охватывающие каждый целевой ген, сначала амплифицируют с использованием праймеров ПЦР#1 (SEQ ID No: 310 и 311) по программе: 98°C, 1 мин; 30 циклов [98°C, 10 сек; 62°C, 15 сек; 72°C, 5 мин]; 72°C, 5 мин; 12°C постоянно.

Затем один микролитр этой ПЦР-реакции дополнительно амплифицируют с использованием праймеров, специфичных для каждой направляющей (праймеры ПЦР#2; SEQ ID NO: 365-370), используя программу: 98°C, 1 мин; 35 циклов [98°C, 10 сек; 67°C, 15 сек; 72°C, 30 сек]; 72°C, 5 мин; 12°C постоянно.

Праймеры для ПЦР#2 включают выступающие последовательности адаптера транспозазы Read 1 и Read 2 Nextera для секвенирования Illumina.

После второй ПЦР-амплификации ДНК очищают с помощью набора для очистки ПЦР (фирма Zymo) в соответствии с инструкциями производителя и элюируют водой. 200-500 нг очищенного продукта ПЦР#2 объединяют с 2 мкл буфера 10X NEB Buffer 2 и водой в 20 мкл реакции и подвергают отжигу с образованием гетеродуплексной ДНК по программе: 95°C, 5 мин; 95-85°C, охлаждение со скоростью 2°C/сек; 85-25°C, охлаждение со скоростью 0,1°C/сек.; 12°C постоянно. После отжига 5 мкл ДНК удаляют в качестве контроля без фермента, добавляют 1 мкл эндонуклеазы I фага T7 (NEB) и реакционную смесь инкубируют при 37°C в течение 1 ч. После инкубации добавляют 5х загрузочного краситель FlashGel (фирма Lonza) и по 5 мкл каждой реакционной смеси и контролей, анализируют с помощью гель-электрофореза 2,2% агарозного FlashGel (фирма Lonza). После визуализации геля определяют процент негомологичного соединения концов (NHEJ – non-homologous end joining) с помощью следующего уравнения: % событий NHEJ = 100 x [1-(1-расщепленная фракция)^(1/2)], где (расщепленная фракция) определяют как: (плотность расщепленных продуктов)/(плотность расщепленных продуктов + полоса нерасщепленного исходного образца).

Для некоторых образцов применяют программу SURVEYOR® для анализа результатов после экспрессии в клетках млекопитающих. Клетки инкубируют при 37°C в течение 72 ч после трансфекции перед выделением геномной ДНК. Геномную ДНК выделяют с помощью раствора для экстракции ДНК QuickExtract (фирма Epicentre) в соответствии с протоколом производителя. Геномную область, примыкающую к целевому сайту RGN, амплифицируют с помощью ПЦР, полученные продукты очищают с использованием спиновой колонки QiaQuick (фирма Qiagen) в соответствии с протоколом производителя. Всего 200–500 нг очищенных продуктов ПЦР смешивают с 1 мкл буфера 10 × Taq DNA Polymerase PCR buffer (фирма Enzymatics) и сверхчистой воды до конечного объема 10 мкл и подвергают повторному отжигу для обеспечения образования гетеродуплекса: 95°C в течение 10 мин, от 95°C до 85°C линейно со скоростью -2°C/сек, от 85°C до 25°C со скоростью -0,25°C/сек и выдерживают при 25°C в течение 1 мин.

После повторного выравнивания продукты обрабатывают нуклеазой SURVEYOR® и энхансером SURVEYOR® S (фирма Integrated DNA Technologies) по рекомендованному производителем протоколу и анализируют в 4-20% полиакриламидных гелях Novex TBE (фирма Life Technologies). Гели окрашивают красителем SYBR Gold DNA (фирма Life Technologies) в течение 10 мин и визуализируют с помощью системы визуализации гелей Gel Doc (фирма Bio-rad). Количественная оценка основана на относительной интенсивности полос. Процент индел определяют по формуле $100 \times (1 - (1 - (b+c)/(a+b+c))^{1/2})$, где a – интегральная интенсивность нерасщепленного продукта ПЦР, b и c – интегральные интенсивности каждого продукта расщепления.

Кроме того, продукты ПЦР#2, содержащие выступающие последовательности Illumina, используют для получения библиотеки в соответствии с протоколом Illumina 16S Metagenomic Sequencing Library. Глубокое секвенирование выполняют на платформе Illumina Mi-Seq поставщиком услуг (фирма MOGene). Обычно 200000 прочтений парных концов по 250 п.н. (2×100000 прочтений) получают в расчете на один ампликон. Прочтения анализируют с помощью программы CRISPResso (Pinello с соавт., *Nature Biotech*, 2016? 34:695-697) для расчета степени редактирования. Выходные данные по выравниванию подбирают вручную, чтобы подтвердить сайты инсерции и делеции, а также идентифицировать сайты микргомологии в местах рекомбинации. Общие показатели редактирования, а также доли делеций и инсерций в каждом образце, показаны в табл. 7. Все эксперименты проводят на клетках человека. «Мишенью» является последовательность-мишень в целевом гене. Для каждой последовательности-мишени направляющая РНК включает комплементарную РНК последовательности спейсера и соответствующую sgРНК, зависящую от используемой RGN. Выбранный анализ экспериментов по направляющим РНК показан в табл. 8.1 и 8.2.

Таблица 7. Общие показатели редактирования генных мишеней

RGN	Генные мишени	Направляющая РНК (SEQ ID NO.)	Мишень (SEQ ID NO.)	Доля общего редактирования	Доля делеций	Доля инсерций

RGN	Генные мишени	Направляющая РНК (SEQ ID NO.)	Мишень (SEQ ID NO.)	Доля общего редактирования	Доля делеций	Доля инсерций
APG09106.1	AurkB	316	314	0,55%	100%	
APG09106.1	AurkB	312	315	0,60%	54%	46%
APG09106.1 T849R	AurkB	316	314	2,97%	98%	2,00%
APG09106.1 T849R	AurkB	312	315	2,36%	100%	
APG05712.1	RelA	419	482	0,36%	72,33%	27,67%
APG05712.1	AurkB	420	483	1,39%	25,94%	75,12%
APG05712.1	RelA	421	484	1,04%	32,10%	70,33%
APG05712.1	RelA	422	485	0,28%	64,13%	35,87%
APG01658.1	RelA	423	486	10,03%	70,66%	29,33%
APG01658.1	GAPDH	424	487	5,60%	76,96%	25,97%
APG01658.1	LINC01509	425	488	9,87%	75,10%	26,61%
APG01658.1	HBB	426	489	0,42%	63,86%	36,13%
APG01658.1	CFTR	427	490	6,06%	75,42%	27,86%
APG01658.1	AurkB	428	491	13,07%	88,02%	13,14%
APG05733.1	HPRT1	429	492	0,36%		100%
APG05733.1	AurkB	430	493	5,01%	86,40%	13,60%
APG05733.1	RelA	431	494	0,76%		100%
APG05733.1	HPRT1	432	495	0,17%	38,42%	61,58%
APG05733.1	RelA	433	496	0,96%	77,52%	22,48%
APG06498.1	TRA	434	497	35,35%	92,25%	8,29%
APG06498.1	TRA	435	498	21,04%	84,87%	16,34%
APG06498.1	TRA	436	499	10,16%	89,44%	10,69%
APG06498.1	TRA	437	500	13,21%	87,24%	15,00%
APG06498.1	TRA	438	501	13,56%	86,73%	13,66%
APG06498.1	TRA	439	502	9,62%	95,85%	19,90%
APG06498.1	TRA	440	503	1,51%	75,89%	24%
APG06498.1	TRA	441	504	2,63%	88,81%	11,76%
APG06498.1	TRA	442	505	4,27%	52,09%	47,90%
APG06646.1	TRA	443	497	0,33%	0%	100%
APG06646.1	VEGFA	444	506	4,52%	40,83%	63,79%
APG06646.1	AurkB	445	507	0,66%	69,38%	30,62%
APG06646.1	AurkB	446	508	6,55%	79,97%	20,37%
APG06646.1	TRA	447	509	2,47%	73,40%	40,81%
APG06646.1	TRA	448	510	0,14%	34,46%	65,55%

RGN	Генные мишени	Направляющая РНК (SEQ ID NO.)	Мишень (SEQ ID NO.)	Доля общего редактирования	Доля делеций	Доля инсерций
APG06646.1	TRA	449	511	0,61%	0%	100%
APG06646.1	AurkB	450	493	7,94%	67,95%	32,40%
APG06646.1	TRA	451	512	7,03%	86,95%	14,03%
APG06646.1	TRA	452	513	0,50%	21,80%	78,10%
APG06646.1	TRA	453	514	3,53%	86,42%	16,28%
APG06646.1	RelA	454	515	6,22%	80,39%	45,34%
APG06646.1	TRA	455	516	0,38%	0%	100%
APG06646.1	TRA	456	517	0,34%	66,76%	48,22%
APG06646.1	RelA	457	518	3,78%	75,25%	67,90%
APG06646.1	TRA	458	519	0,42%	69,61%	30,39%
APG06646.1	TRA	459	520	0,08%	100,00%	0,00%
APG06646.1	AurkB	460	521	1,09%	30,18%	69,81%
APG06646.1	AurkB	461	522	0,28%	86,60%	21,78%
APG06646.1	AurkB	462	523	10,92%	78,87%	22,05%
APG06646.1	TRA	463	504	0,08%	0%	100%
APG06646.1	TRA	464	505	2,84%	38,26%	61,47%
APG06646.1	TRA	465	524	0,43%	71,67%	28,33%
APG06646.1	TRA	466	525	0,03%	0,00%	100,00%
APG06646.1	TRA	467	526	0,08%	100%	0%
APG09882.1	RelA	468	482	0,32%	100,00%	0,00%
APG09882.1	EMX1	469	527	12,84%	82,46%	17,79%
APG09882.1	VEGFA	470	528	14,76%	92,56%	7,77%
APG09882.1	TRA	471	529	14,66%	94,19%	7,33%
APG09882.1	TRA	472	530	7,84%	95,07%	5,75%
APG09882.1	VEGFA	473	531	24,45%	88,96%	11,57%
APG09882.1	TRA	474	532	14,43%	90,24%	10,96%
APG01405.1	RelA	475	482	0,06%	0%	100%
APG01405.1	AurkB	476	533	6,81%	80,58%	20,27%
APG01405.1	AurkB	477	534	0,54%	33,83%	66,16%
APG01405.1	HPRT1	478	535	1,07%	42,14%	57,87%
APG01405.1	RelA	479	484	0,55%	2,07%	97,93%
APG01308.1	HPRT1	480	536	1,19%	80,57%	19,43%
APG01308.1	AurkB	481	537	0,90%	94,97%	6,46%

Конкретные инсерции и делеции для соответствующих направляющих показаны в табл. 8.1 и 8.2. В этих таблицах целевая последовательность обозначена жирным шрифтом заглавными буквами. Восьмимерные области РАМ подчеркнуты двойными линиями, а основные распознанные нуклеотиды выделены жирным шрифтом. Инсерции обозначают строчными буквами. Удаления обозначены тире (---). Местоположение INDEL вычисляются по проксимальному краю РАМ целевой последовательности, при этом край является местоположением 0. Местоположение является положительным (+), если оно находится на целевой стороне края; местоположение отрицательное (-), если местоположение находится на стороне края РАМ.

Таблица 8.1. Специфические инсерции и делеции для направляющей для направляющей 831 с применением RGN APG09106.1

Отредактированная целевая последовательность	# прочтения	Прочтения (%)	INDEL (%)	Тип	Локализация INDEL	Размер
GTCTGATTGCCTGTCGTTGCCCTCCAGATC ATGGAGGAGTTGGCAGA (дикий тип; SEQ ID NO: 316)	92294	99,40				
GTCTGATTGCCTGTCGTTGCCCTCCCA ---- ----AGGAGTTGGCAGA (SEQ ID NO: 317)	263	0,28	54,22	Делеция	+19	8
GTCTGATTGCCTGTCGTTGCC cctaagtgtat taagcattgtctcagagattttGGAGGAGTTG GCAGA (SEQ ID NO: 318)	222	0,24	45,77	Инсерция	+13	20

Таблица 8.2. Специфические инсерции и делеции для направляющей 831 с применением APG09106.1 T849R

Отредактированные последовательности-мишени	# прочтения	Прочтения (%)	Индел (%)	Тип	Локализация индел	Размер
GTCTGATTGCCTGTCGTTGCCCTCCAGATC ATGGAGGAGTTGGCAGA (wild type; SEQ ID NO: 316)	189881	97,64				
GTCTGATTGCCTGTCGTTGCC ----- -TGGAGGAGTTGGCAGA (SEQ ID NO: 319)	602	0,309	13,129	Делеция	+14	10

Отредактированные последовательности-мишени	# прочтения	Прочтения (%)	Индел (%)	Тип	Локализация индел	Размер
<u>GTCTGATTGCCTGTCGTTGCCCTCCAGATC</u> -GGAGGAGTTGGCAGA (SEQ ID NO: 320)	394	0,202	8,593	Делеция	+23	2
<u>GTCTGATTGCCTGTCGTTGCCCTCCAGAT-</u> ----AGGAGTTGGCAGA (SEQ ID NO: 321)	399	0,205	8,702	Делеция	+22	5
<u>GTCTGATTGCCTGTCGTTGCCaTC</u> ----- -TG-GGAGTTGGCAGA (SEQ ID NO: 322)	379	0,194	8,266	Делеция и Мутация	+16	10
<u>GTCTGATTGCCTGTCGTTGCCCTC</u> ----- -TGGAGGAGTTGGCAGA (SEQ ID NO: 323)	350	0,179	7,633	Делеция	+16	8
<u>GTCTGAT</u> ----- -TGGAGGAGTTGGCAGA (SEQ ID NO: 324)	309	0,158	6,739	Делеция	-1	26
<u>GTCTGATTGCCTGTCGTTGCCCTC</u> ----- --GGAGGAGTTGGCAGA (SEQ ID NO: 325)	280	0,143	6,106	Делеция	+16	9
<u>GTCTGATTGCCTGTCGTTGCCCTCC</u> ----- -Aggaggagttggcaga (SEQ ID NO: 326)	274	0,140	5,976	Делеция и Мутация	+17	7
<u>GTCTGATTGCCTGTCGTTGCC</u> ----- ----GGAGTTGGCAGA (SEQ ID NO: 327)	251	0,129	5,474	Делеция	+13	15
<u>GTCTGATTGCCTGTCGTTGCC</u> ----- ATCATGGAGGAGTTGGCAGA (SEQ ID NO: 328)	250	0,128	5,452	Делеция	+13	7
<u>GTCTGATTGCCTGTCGTTGCCCTC</u> ----- CATGGAGGAGTTGGCAGA (SEQ ID NO: 329)	231	0,118	5,038	Делеция	+16	6
<u>GTCTGATTGCCTGTCGTTGCCCTCCSA</u> ---- -----GТАCT (SEQ ID NO: 330)	218	0,112	4,754	Делеция	+19	30
<u>GTCTGATTGCCTGTCGTTGCCCC</u> ----- aATCtTGGAGGAGTTGGCAGA (SEQ ID NO: 331)	206	0,105	4,492	Делеция и Мутация	+14	5
<u>GTCTGATTGCCTGTCGTTGCC</u> ----- TgggATGGAGGAGTTGGCAGA (SEQ ID NO: 332)	162	0,083	3,533	Делеция и Мутация	+13	8
<u>GTCTGATTGCCTGTCGTTGCCCTC</u> ----- -----AGTTGGCAGA (SEQ ID NO: 333)	158	0,081	3,446	Делеция	+16	14
<u>GTCTGATTGCCTGTCGTTGCCCC</u> ----- TCATGGAGGAGTTGGCAGA (SEQ ID NO: 334)	122	0,062	2,660	Делеция	+14	7

Пример 6. Проявление активности по геному редактированию в клетках растений

РНК-направляемая активность нуклеаз RGN по настоящему изобретению продемонстрирована в растительных клетках с использованием протоколов, адаптированных из публикации Li, с соавт. (*Nat. Biotech.*, 2013, 31:688-691). Вкратце, оптимизированная по кодонам растительная версия RGN по настоящему изобретению (SEQ ID NO: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110 или 117), функционально связанная с последовательностью нуклеиновой кислоты, кодирующей N-концевой сигнал ядерной локализации вируса SV40, клонируют с сильным конститутивным промотором 35S в векторе временной трансформации. Молекулы sgРНК, нацеленные на один или несколько сайтов в гене PDS растения, фланкируют соответствующую последовательность PAM, клонируют с промотором U6 растения во втором векторе временной экспрессии. Векторы экспрессии вводят в протопласты мезофилла *Nicotiana benthamiana* с помощью ПЭГ-опосредованной трансформации. Трансформированные протопласты инкубируют в темноте в течение до 36 ч. Геномную ДНК выделяют из протопластов с помощью набора Dneasy Plant Mini Kit (фирма Qiagen). Геномную область, фланкирующую целевой сайт RGN, амплифицируют с помощью ПЦР, продукты реакции очищают с использованием спиновой колонки QiaQuick (фирма Qiagen) в соответствии с протоколом производителя. Всего 200-500 нг очищенных продуктов ПЦР смешивают с 1 мкл буфера 10× Taq DNA Polymerase PCR (фирма Enzymatics) и сверхчистой воды до конечного объема 10 мкл и подвергают повторному отжигу для обеспечения образования гетеродуплекса: 95°C в течение 10 мин, от 95°C до 85°C линейно со скоростью -2°C/сек, от 85°C до 25°C со скоростью -0,25°C/сек и выдерживают при 25°C в течение 1 мин.

После повторного выравнивания продукты обрабатывают нуклеазой SURVEYOR и энхансером SURVEYOR S (фирма Integrated DNA Technologies) по рекомендованному производителем протоколу и анализируют в 4-20% полиакриламидных гелях Novex TBE (фирма Life Technologies). Гели окрашивают ДНК-красителем SYBR Gold (фирма Life Technologies) в течение 10 мин и визуализируют с помощью системы визуализации гелей Gel Doc (фирма Bio-rad). Количественная оценка основана на относительной интенсивности полос. Процент индел определяют по формуле $100 \times (1 - (1 - (b+c)/(a+b+c))^{1/2})$, где a

представляет собой интегральную интенсивность нерасщепленного продукта ПЦР, b и c представляют собой интегральные интенсивности каждого продукта расщепления.

Пример 7. Идентификация мишеней при заболеваниях

5 База данных клинических вариантов получена из базы данных NCBI ClinVar, которая доступна во всемирной паутине на веб-сайте NCBI ClinVar. Патогенные однонуклеотидные полиморфизмы (SNP – Single Nucleotide Polymorphisms) идентифицированы из этого списка. Используя информацию о геномном локусе, идентифицируют мишени CRISPR в области,

10 перекрывающейся и окружающей каждый SNP. Выбор SNP, которые можно скорректировать с помощью редактирования оснований в сочетании с системой RGN, включающей APG09748 или APG09106.1, для нацеливания на причинную мутацию («Cas1 Mut.»), приведен в табл. 9.1. Выбор SNP, которые можно скорректировать с помощью редактирования оснований в сочетании с системой

15 RGN, включающей APG06646.1 или APG01658.1, для нацеливания на причинную мутацию («Cas1 Mut.»), приведен в табл. 9.2. В обеих таблицах 9.1 и 9.2 указано только одно название для каждой болезни. Знак «RS#» соответствует регистрационному номеру RS в базе данных SNP на веб-сайте NCBI. В номер регистрации хромосомы заложена справочная информация, которую можно

20 найти на веб-сайте NCBI. В табл. 9.1 и 9.2 также показана информация о целевой последовательности генома, подходящей для систем RGN APG09748 или APG09106.1 (в табл. 9.1), или APG06646.1 или APG01658.1 (в табл. 9.2) для каждого заболевания. Информация о последовательности-мишени также предоставляет последовательность протоспейсера для получения необходимой

25 sgРНК для соответствующей RGN по настоящему изобретению.

Таблица 9.1. Мишени при заболеваниях для APG09748 и APG09106.1

Заболевания	RS#	Причинная мутация	Номер регистрации хромосомы	Обозначение гена	Мишень (SEQ ID NO.)
Болезнь Штаргардта 1	1800728	A>G	NC_000001.10,NC_000001.11	ABCA4	313

Заболевания	RS#	Причинная мутация	Номер регистрации хромосомы	Обозначение гена	Милпечь (SEQ ID NO.)
Болезнь накопления гликогена 1А типа	1801175	C>T	NC_000017.10,NC_000017.11	G6PC	335
Тяжелый комбинированный иммунодефицит	3218716	C>T	NC_000014.8,NC_000014.9	MYH7	336
Фенилкетонурия	5030858	G>A	NC_000012.11,NC_000012.12	PAH	337
Гиперфенилаланинемия	5030860	T>C	NC_000012.11,NC_000012.12	PAH	338
Дефицит альфа-1-антитрипсина	28929474	C>T	NC_000014.8,NC_000014.9	SERPINA1	339
Заболевания, связанные с MECР2	28934906	G>A	NC_000023.10,NC_000023.11	MECР2	340
Миопатия с тельцами включения 2	28937594	A>G	NC_000009.11,NC_000009.12	ГNE	341
Болезнь фон Виллебранда	41276738	C>T	NC_000012.11,NC_000012.12	VWF	342
Рак молочной железы и/или яичников	41293455	G>A	NC_000017.10,NC_000017.11	BRCA1	343
Заболевания, связанные с MECР2	61750240	G>A	NC_000023.10,NC_000023.11	MECР2	344
Расстройство, связанное с MEFV	61752717	T>C	NC_000016.9,NC_000016.10	MEFV	345
Рак молочной железы и/или яичников	62625307	G>A	NC_000017.10,NC_000017.11	BRCA1	346
Рак молочной железы и/или яичников	62625308	G>A	NC_000017.10,NC_000017.11	BRCA1	347
Наследственный предрасполагающий к раку синдром	63749795	C>T	NC_000003.11,NC_000003.12	MLH1	348
Наследственный предрасполагающий к раку синдром	63749849	C>T	NC_000002.11,NC_000002.12	MSH2	349
Наследственный предрасполагающий к раку синдром	63750636	C>T	NC_000002.11,NC_000002.12	MSH2	350
Дефицит карнитинпальмитоилтрансферазы II	74315294	C>T	NC_000001.10,NC_000001.11	CPT2	351
Муковисцидоз	74597325	C>T	NC_000007.13,NC_000007.14	CFTR	352
Дефицит УДФ-глюкозо-гексозо-1-фосфат уридилтрансферазы	75391579	A>G	NC_000009.11,NC_000009.12	GALT	353
Муковисцидоз	75527207	G>A	NC_000007.13,NC_000007.14	CFTR	354
Амилоидогенный транстретиновый амилоидоз	76992529	G>A	NC_000018.9,NC_000018.10	TTR	355
Муковисцидоз	77010898	G>A	NC_000007.13,NC_000007.14	CFTR	356
Метахроматическая лейкодистрофия	80338815	C>T	NC_000022.10,NC_000022.11	ARSA	357
Синдром Каудена 3	80338844	C>T	NC_000011.9,NC_000011.10	SDHD	358
Синдром Смита-Лемли-Опица	80338853	G>A	NC_000011.9,NC_000011.10	DHCR7	359
Рак молочной железы и/или яичников	80356962	C>T	NC_000017.10,NC_000017.11	BRCA1	360
Рак молочной железы и/или яичников	80357123	G>A	NC_000017.10,NC_000017.11	BRCA1	361

Заболевания	RS#	Причинная мутация	Номер регистрации хромосомы	Обозначение гена	Мишень (SEQ ID NO.)
Врожденные генетические заболевания	80358259	A>G	NC_000018.9,NC_000018.10	NPC1	362
Рак молочной железы и/или яичников	80359212	C>T	NC_000013.10,NC_000013.11	BRCA2	363
Анемия Фанкони	104886457	G>A	NC_000009.11,NC_000009.12	FANCC	364
Заболевания, связанные с SLC26A2	104893915	C>T	NC_000005.9,NC_000005.10	SLC26A2	365
Кардиомиопатия	104894368	C>T	NC_000012.11,NC_000012.12	MYL2	366
Глухота, сцепленная с X-хромосомой	104894396	C>T	NC_000013.10,NC_000013.11	GJB2	367
Врожденные генетические заболевания	104894635	C>T	NC_000017.10,NC_000017.11	SGSH	368
Семейная средиземноморская лихорадка	104895097	C>T	NC_000016.9,NC_000016.10	MEFV	369
Семейная дисавтономия	111033171	A>G	NC_000009.11,NC_000009.12	ELP1	370
Синдром Швахмана	113993993	A>G	NC_000007.13,NC_000007.14	SBDS	371
Заболевания, связанные с RYR1	118192172	C>T	NC_000019.9,NC_000019.10	RYR1	372
Цероидный липофусциноз нейронов 2	119455955	G>A	NC_000011.9,NC_000011.10	TPP1	373
Дефицит ацилкофермента А дегидрогеназы со средней длиной цепи	121434274	G>A	NC_000001.10,NC_000001.11	ACADM	374
Первичная гипероксалурия	121908529	G>A	NC_000002.11,NC_000002.12	AGXT	375
Кардиомиопатия	121908987	C>T	NC_000007.13,NC_000007.14	PRKAG2	376
Синдром Каудена	121909219	C>T	NC_000010.10,NC_000010.11	PTEN	377
Кардиомиопатия	121913628	C>T	NC_000014.8,NC_000014.9	MYH7	378
Врожденные генетические заболевания	121918243	G>A	NC_000001.10,NC_000001.11	MMACHC	379
Расстройство, связанное с RTPN11	121918457	C>T	NC_000012.11,NC_000012.12	PTPN11	380
Ювенильный миело monocитарный лейкоз	121918462	C>T	NC_000012.11,NC_000012.12	PTPN11	381
Ювенильный миело monocитарный лейкоз	121918466	A>G	NC_000012.11,NC_000012.12	PTPN11	382
Мукополисахаридоз I типа	121965020	C>T	NC_000004.11,NC_000004.12	IDUA	383
Цероидный липофусциноз нейронов 1	137852700	G>A	NC_000001.10,NC_000001.11	PPT1	384
СНЕК2-связанная предрасположенность к раку	137853007	G>A	NC_000022.10,NC_000022.11	CHEK2	385
Колоректальный рак	137854568	C>T	NC_000005.9,NC_000005.10	APC	386
Семейная гиперхолестеринемия	137929307	G>A	NC_000019.9,NC_000019.10	LDLR	387
Кардио-фацио-кожный синдром	180177035	T>C	NC_000007.13,NC_000007.14	BRAF	388
Семейный рак молочной железы	180177083	G>A	NC_000016.10,NC_000016.9	PALB2	389
Заболевания, связанные с MYBPC3	200411226	C>T	NC_000011.9,NC_000011.10	MYBPC3	390

Заболевания	RS#	Причинная мутация	Номер регистрации хромосомы	Обозначение гена	Мишень (SEQ ID NO.)
Заболевания, связанные с RYR1	200563280	C>T	NC_000019.9,NC_000019.10	RYR1	391
Заболевания, связанные с MYBPC3	387907267	G>A	NC_000011.9,NC_000011.10	MYBPC3	392
Десмоидная болезнь, наследственная	397515734	C>T	NC_000005.9,NC_000005.10	APC	393
Синдром Марфана/синдром Лойса-Дитца/семейные аневризмы и расслоения грудной аорты	397515757	C>T	NC_000015.9,NC_000015.10	FBN1	394
Иммунодефицит 14	397518423	G>A	NC_000001.10,NC_000001.11	PIK3CD	395
Врожденные генетические заболевания	398123009	C>T	NC_000011.9,NC_000011.10	PACS1	396
Лимфома лимфобластного лейкоза В, без подтипа ICD-O	529008617	G>A	NC_000001.10,NC_000001.11	MUTYH	397
Семейный рак молочной железы	587780021	G>A	NC_000002.11,NC_000002.12	BARD1	398
Семейная гиперхолестеринемия	746118995	C>T	NC_000019.9,NC_000019.10	LDLR	399
Семейная гиперхолестеринемия	769370816	G>A	NC_000019.10,NC_000019.9	LDLR	400

Таблица 9.2. Мишени при заболеваниях для APG6646.1 или APG01658.1

Заболевания	RGN	RS#	CasI Mut.	Номер регистрации хромосомы	Обозначение гена	Мишень (SEQ ID NO)
Болезнь Штаргардта 1	APG01658.1	1800553	C>T	NC_000001.10,NC_000001.11	ABCA4	538
Болезнь Штаргардта 1	APG06646.1	1800553	C>T	NC_000001.10,NC_000001.11	ABCA4	539
Наследственный предрасполагающий к раку синдром	APG01658.1	5030804	A>G	NC_000003.11,NC_000003.12	VHL	540
Наследственный предрасполагающий к раку синдром	APG06646.1	5030804	A>G	NC_000003.11,NC_000003.12	VHL	541
Фенилкетонурия	APG01658.1	5030851	G>A	NC_000012.11,NC_000012.12	PAH	542
Фенилкетонурия	APG06646.1	5030851	G>A	NC_000012.11,NC_000012.12	PAH	543
Меланома и миелома	APG01658.1	11547328	G>A	NC_000012.11,NC_000012.12	CDK4	544

Заболевания	RGN	RS#	CasI Mut.	Номер регистрации хромосомы	Обозначение гена	Милень (SEQ ID NO)
Меланома и миелома	APG06646.1	11547328	G>A	NC_000012.11,NC_000012.12	CDK4	545
Семейная гиперхолестеринемия	APG01658.1	11547917	C>T	NC_000019.9,NC_000019.10	LDLR	546
Семейная гиперхолестеринемия	APG06646.1	11547917	C>T	NC_000019.9,NC_000019.10	LDLR	547
Болезнь Шарко-Мари-Тута	APG01658.1	61444459	G>A	NC_000001.10,NC_000001.11	LMNA	548
Болезнь Шарко-Мари-Тута	APG06646.1	61444459	G>A	NC_000001.10,NC_000001.11	LMNA	549
Синдром Ангельмана	APG01658.1	61751362	G>A	NC_000023.10,NC_000023.11	MECP2	550
Синдром Ангельмана	APG06646.1	61751362	G>A	NC_000023.10,NC_000023.11	MECP2	551
Фенилкетонурия	APG01658.1	62514895	C>T	NC_000012.11,NC_000012.12	PAH	552
Фенилкетонурия	APG06646.1	62514895	C>T	NC_000012.11,NC_000012.12	PAH	553
Фенилкетонурия	APG01658.1	62516152	C>T	NC_000012.11,NC_000012.12	PAH	554
Фенилкетонурия	APG06646.1	62516152	C>T	NC_000012.11,NC_000012.12	PAH	555
Синдром Линча	APG01658.1	63749795	C>T	NC_000003.11,NC_000003.12	MLH1	556
Синдром Линча	APG06646.1	63749795	C>T	NC_000003.11,NC_000003.12	MLH1	557
Наследственный рак	APG01658.1	63749843	C>T	NC_000002.11,NC_000002.12	MSH6	558
Наследственный рак	APG06646.1	63749843	C>T	NC_000002.11,NC_000002.12	MSH6	559
Наследственный предрасполагающий к раку синдром	APG01658.1	63749849	C>T	NC_000002.11,NC_000002.12	MSH2	560
Наследственный предрасполагающий к раку синдром	APG06646.1	63749849	C>T	NC_000002.11,NC_000002.12	MSH2	561
Наследственный предрасполагающий к раку синдром	APG01658.1	63750508	C>T	NC_000002.11,NC_000002.12	MSH2	562
Наследственный предрасполагающий к раку синдром	APG06646.1	63750508	C>T	NC_000002.11,NC_000002.12	MSH2	563
Наследственный предрасполагающий к раку синдром	APG01658.1	63750636	C>T	NC_000002.11,NC_000002.12	MSH2	564
Наследственный предрасполагающий к раку синдром	APG06646.1	63750636	C>T	NC_000002.11,NC_000002.12	MSH2	565
Наследственный предрасполагающий к раку синдром	APG06646.1	63750871	G>A	NC_000007.13,NC_000007.14	PMS2	566
Синдром Линча	APG01658.1	63751194	C>T	NC_000003.11,NC_000003.12	MLH1	567

Заболевания	RGN	RS#	CasI Mut.	Номер регистрации хромосомы	Обозначение гена	Милень (SEQ ID NO)
Синдром Линча	APG06646.1	63751194	C>T	NC_000003.11,NC_000003.12	MLH1	568
Синдром Линча	APG01658.1	63751711	G>A	NC_000003.11,NC_000003.12	MLH1	569
Синдром Линча	APG06646.1	63751711	G>A	NC_000003.11,NC_000003.12	MLH1	570
Муковисцидоз	APG01658.1	75039782	C>T	NC_000007.13,NC_000007.14	CFTR	571
Муковисцидоз	APG06646.1	75039782	C>T	NC_000007.13,NC_000007.14	CFTR	572
Глухота	APG01658.1	76434661	C>T	NC_000013.10,NC_000013.11	GJB2	573
Глухота	APG06646.1	76434661	C>T	NC_000013.10,NC_000013.11	GJB2	574
Муковисцидоз	APG01658.1	76713772	G>A	NC_000007.13,NC_000007.14	CFTR	575
Муковисцидоз	APG06646.1	76713772	G>A	NC_000007.13,NC_000007.14	CFTR	576
Кардиомиопатия	APG01658.1	76992529	G>A	NC_000018.9,NC_000018.10	TTR	577
Кардиомиопатия	APG06646.1	76992529	G>A	NC_000018.9,NC_000018.10	TTR	578
Муковисцидоз	APG01658.1	77010898	G>A	NC_000007.13,NC_000007.14	CFTR	579
Муковисцидоз	APG06646.1	77010898	G>A	NC_000007.13,NC_000007.14	CFTR	580
Муковисцидоз	APG01658.1	80224560	G>A	NC_000007.13,NC_000007.14	CFTR	581
Муковисцидоз	APG06646.1	80224560	G>A	NC_000007.13,NC_000007.14	CFTR	582
Синдром Каудена	APG01658.1	80338844	C>T	NC_000011.9,NC_000011.10	SDHD	583
Синдром Каудена	APG06646.1	80338844	C>T	NC_000011.9,NC_000011.10	SDHD	584
Заболевания, связанные с SLC26A4	APG01658.1	80338849	G>A	NC_000007.13,NC_000007.14	SLC26A4	585
Заболевания, связанные с SLC26A4	APG06646.1	80338849	G>A	NC_000007.13,NC_000007.14	SLC26A4	586
Рак молочной железы-яичников	APG01658.1	80356885	C>T	NC_000017.10,NC_000017.11	BRCA1	587
Рак молочной железы-яичников	APG06646.1	80356885	C>T	NC_000017.10,NC_000017.11	BRCA1	588
Рак молочной железы-яичников	APG01658.1	80358557	C>T	NC_000013.10,NC_000013.11	BRCA2	589
Рак молочной железы-яичников	APG06646.1	80358557	C>T	NC_000013.10,NC_000013.11	BRCA2	590
Рак молочной железы-яичников	APG01658.1	80358650	G>A	NC_000013.10,NC_000013.11	BRCA2	591
Рак молочной железы-яичников	APG06646.1	80358650	G>A	NC_000013.10,NC_000013.11	BRCA2	592
Рак молочной железы-яичников	APG01658.1	80358659	C>T	NC_000013.10,NC_000013.11	BRCA2	593

Заболевания	RGN	RS#	CasI Mut.	Номер регистрации хромосомы	Обозначение гена	Милень (SEQ ID NO)
Рак молочной железы-яичников	APG06646.1	80358659	C>T	NC_000013.10,NC_000013.11	BRCA2	594
Рак молочной железы-яичников	APG06646.1	80358663	C>T	NC_000013.10,NC_000013.11	BRCA2	595
Рак молочной железы-яичников	APG01658.1	80358871	G>A	NC_000013.10,NC_000013.11	BRCA2	596
Рак молочной железы-яичников	APG06646.1	80358871	G>A	NC_000013.10,NC_000013.11	BRCA2	597
Рак молочной железы-яичников	APG01658.1	80358920	C>T	NC_000013.10,NC_000013.11	BRCA2	598
Рак молочной железы-яичников	APG06646.1	80358920	C>T	NC_000013.10,NC_000013.11	BRCA2	599
Рак молочной железы-яичников	APG01658.1	80358972	C>T	NC_000013.10,NC_000013.11	BRCA2	600
Рак молочной железы-яичников	APG06646.1	80358972	C>T	NC_000013.10,NC_000013.11	BRCA2	601
Рак молочной железы-яичников	APG01658.1	80358981	C>T	NC_000013.10,NC_000013.11	BRCA2	602
Рак молочной железы-яичников	APG06646.1	80358981	C>T	NC_000013.10,NC_000013.11	BRCA2	603
Рак молочной железы-яичников	APG01658.1	80359003	G>A	NC_000013.10,NC_000013.11	BRCA2	604
Рак молочной железы-яичников	APG06646.1	80359003	G>A	NC_000013.10,NC_000013.11	BRCA2	605
Рак молочной железы-яичников	APG01658.1	80359004	G>A	NC_000013.10,NC_000013.11	BRCA2	606
Рак молочной железы-яичников	APG06646.1	80359004	G>A	NC_000013.10,NC_000013.11	BRCA2	607
Рак молочной железы-яичников	APG01658.1	80359013	G>A	NC_000013.10,NC_000013.11	BRCA2	608
Рак молочной железы-яичников	APG06646.1	80359013	G>A	NC_000013.10,NC_000013.11	BRCA2	609
Рак молочной железы-яичников	APG01658.1	80359027	G>A	NC_000013.10,NC_000013.11	BRCA2	610
Рак молочной железы-яичников	APG06646.1	80359027	G>A	NC_000013.10,NC_000013.11	BRCA2	611
Рак молочной железы-яичников	APG01658.1	80359071	G>A	NC_000013.10,NC_000013.11	BRCA2	612
Рак молочной железы-яичников	APG06646.1	80359071	G>A	NC_000013.10,NC_000013.11	BRCA2	613
Рак молочной железы-яичников	APG01658.1	80359115	C>T	NC_000013.10,NC_000013.11	BRCA2	614
Рак молочной железы-яичников	APG06646.1	80359115	C>T	NC_000013.10,NC_000013.11	BRCA2	615
Рак молочной железы-яичников	APG01658.1	80359140	C>T	NC_000013.10,NC_000013.11	BRCA2	616
Рак молочной железы-яичников	APG06646.1	80359140	C>T	NC_000013.10,NC_000013.11	BRCA2	617
Рак молочной железы-яичников	APG06646.1	80359159	C>T	NC_000013.10,NC_000013.11	BRCA2	618
Рак молочной железы-яичников	APG01658.1	80359180	C>T	NC_000013.10,NC_000013.11	BRCA2	619

Заболевания	RGN	RS#	CasI Mut.	Номер регистрации хромосомы	Обозначение гена	Милень (SEQ ID NO)
Рак молочной железы-яичников	APG06646.1	80359180	C>T	NC_000013.10,NC_000013.11	BRCA2	620
Рак молочной железы-яичников	APG01658.1	81002846	G>A	NC_000013.10,NC_000013.11	BRCA2	621
Рак молочной железы-яичников	APG06646.1	81002846	G>A	NC_000013.10,NC_000013.11	BRCA2	622
Рак молочной железы-яичников	APG06646.1	81002862	A>G	NC_000013.10,NC_000013.11	BRCA2	623
Анемия Фанкони	APG01658.1	104886457	G>A	NC_000009.11,NC_000009.12	FANCC	624
Анемия Фанкони	APG06646.1	104886457	G>A	NC_000009.11,NC_000009.12	FANCC	625
Ахондрогенез	APG01658.1	104893915	C>T	NC_000005.9,NC_000005.10	SLC26A2	626
Ахондрогенез	APG06646.1	104893915	C>T	NC_000005.9,NC_000005.10	SLC26A2	627
Заболевания, связанные с RYR1	APG01658.1	118192172	C>T	NC_000019.9,NC_000019.10	RYR1	628
Заболевания, связанные с RYR1	APG06646.1	118192172	C>T	NC_000019.9,NC_000019.10	RYR1	629
СНЕК2-связанная предрасположенность к раку	APG01658.1	137853007	G>A	NC_000022.10,NC_000022.11	CHEK2	630
СНЕК2-связанная предрасположенность к раку	APG06646.1	137853007	G>A	NC_000022.10,NC_000022.11	CHEK2	631
нейрофиброматоз	APG06646.1	137854550	A>G	NC_000017.10,NC_000017.11	NF1	632
нейрофиброматоз	APG01658.1	137854556	G>A	NC_000017.10,NC_000017.11	NF1	633
нейрофиброматоз	APG06646.1	137854556	G>A	NC_000017.10,NC_000017.11	NF1	634
Миопатия	APG01658.1	193922390	C>T	NC_000014.8,NC_000014.9	MYH7	635
Миопатия	APG06646.1	193922390	C>T	NC_000014.8,NC_000014.9	MYH7	636
Синдром удлиненного интервала QT	APG01658.1	199473428	C>T	NC_000007.13,NC_000007.14	KCNH2	637
Синдром удлиненного интервала QT	APG06646.1	199473428	C>T	NC_000007.13,NC_000007.14	KCNH2	638
Фенилкетонурия	APG01658.1	199475575	G>A	NC_000012.11,NC_000012.12	PAH	639
Фенилкетонурия	APG06646.1	199475575	G>A	NC_000012.11,NC_000012.12	PAH	640
Фенилкетонурия	APG01658.1	199475679	C>T	NC_000012.11,NC_000012.12	PAH	641
Фенилкетонурия	APG06646.1	199475679	C>T	NC_000012.11,NC_000012.12	PAH	642
Заболевания, связанные с MYBP3	APG01658.1	200411226	C>T	NC_000011.9,NC_000011.10	MYBP3	643
Заболевания, связанные с MYBP3	APG06646.1	200411226	C>T	NC_000011.9,NC_000011.10	MYBP3	644
Миопатия	APG06646.1	267606908	T>C	NC_000014.8,NC_000014.9	MYH7	645

Заболевания	RGN	RS#	CasI Mut.	Номер регистрации хромосомы	Обозначение гена	Милень (SEQ ID NO)
Кардиомиопатия	APG01658.1	267607003	C>T	NC_000010.10,NC_000010.11	RBM20	646
Кардиомиопатия	APG06646.1	267607003	C>T	NC_000010.10,NC_000010.11	RBM20	647
Нунан синдром	APG01658.1	267607048	A>G	NC_000010.10,NC_000010.11	SHOC2	648
Нунан синдром	APG06646.1	267607048	A>G	NC_000010.10,NC_000010.11	SHOC2	649
Семейная гиперхолестеринемия	APG01658.1	267607213	G>A	NC_000019.9,NC_000019.10	LDLR	650
Семейная гиперхолестеринемия	APG06646.1	267607213	G>A	NC_000019.9,NC_000019.10	LDLR	651
Синдром Линча	APG01658.1	267607789	G>A	NC_000003.11,NC_000003.12	MLH1	652
Синдром Линча	APG06646.1	267607789	G>A	NC_000003.11,NC_000003.12	MLH1	653
Синдром Линча	APG01658.1	267607845	G>A	NC_000003.11,NC_000003.12	MLH1	654
Синдром Линча	APG06646.1	267607845	G>A	NC_000003.11,NC_000003.12	MLH1	655
Наследственный предрасполагающий к раку синдром	APG01658.1	267608098	A>G	NC_000002.11,NC_000002.12	MSH6	656
Наследственный предрасполагающий к раку синдром	APG06646.1	267608098	A>G	NC_000002.11,NC_000002.12	MSH6	657
Заболевания, связанные с МУВРС3	APG01658.1	387906397	A>G	NC_000011.9,NC_000011.10	MYBP3	658
Заболевания, связанные с МУВРС3	APG06646.1	387906397	A>G	NC_000011.9,NC_000011.10	MYBP3	659
Рак молочной железы-яичников	APG01658.1	387906843	G>A	NC_000017.10,NC_000017.11	RAD51D	660
Рак молочной железы-яичников	APG06646.1	387906843	G>A	NC_000017.10,NC_000017.11	RAD51D	661
Заболевания, связанные с МУВРС3	APG01658.1	387907267	G>A	NC_000011.9,NC_000011.10	MYBP3	662
Заболевания, связанные с МУВРС3	APG06646.1	387907267	G>A	NC_000011.9,NC_000011.10	MYBP3	663
Рак молочной железы-яичников	APG01658.1	397507389	G>A	NC_000013.10,NC_000013.11	BRCA2	664
Рак молочной железы-яичников	APG06646.1	397507389	G>A	NC_000013.10,NC_000013.11	BRCA2	665
Рак молочной железы-яичников	APG01658.1	397507404	G>A	NC_000013.10,NC_000013.11	BRCA2	666
Рак молочной железы-яичников	APG06646.1	397507404	G>A	NC_000013.10,NC_000013.11	BRCA2	667
Лейкоз	APG01658.1	397507545	G>A	NC_000012.11,NC_000012.12	PTPN11	668
Лейкоз	APG06646.1	397507545	G>A	NC_000012.11,NC_000012.12	PTPN11	669
Лейкоз	APG01658.1	397507547	A>G	NC_000012.11,NC_000012.12	PTPN11	670

Заболевания	RGN	RS#	CasI Mut.	Номер регистрации хромосомы	Обозначение гена	Милень (SEQ ID NO)
Лейкоз	APG06646.1	397507547	A>G	NC_000012.11,NC_000012.12	PTPN11	671
Рак молочной железы-яичников	APG01658.1	397507922	G>A	NC_000013.10,NC_000013.11	BRCA2	672
Рак молочной железы-яичников	APG06646.1	397507922	G>A	NC_000013.10,NC_000013.11	BRCA2	673
Синдром Марфана	APG01658.1	397515757	C>T	NC_000015.9,NC_000015.10	FBN1	674
Синдром Марфана	APG06646.1	397515757	C>T	NC_000015.9,NC_000015.10	FBN1	675
Заболевания, связанные с МУВРС3	APG01658.1	397516074	C>T	NC_000011.9,NC_000011.10	MYBP3	676
Заболевания, связанные с МУВРС3	APG06646.1	397516074	C>T	NC_000011.9,NC_000011.10	MYBP3	677
Кардиомиопатия	APG01658.1	397516354	C>T	NC_000019.9,NC_000019.10	TNNI3	678
Кардиомиопатия	APG06646.1	397516354	C>T	NC_000019.9,NC_000019.10	TNNI3	679
Иммунодефицит	APG01658.1	397518423	G>A	NC_000001.10,NC_000001.11	PIK3CD	680
Иммунодефицит	APG06646.1	397518423	G>A	NC_000001.10,NC_000001.11	PIK3CD	681
В лимфобластный лейкоз лимфома	APG01658.1	529008617	G>A	NC_000001.10,NC_000001.11	MUTYH	682
В лимфобластный лейкоз лимфома	APG06646.1	529008617	G>A	NC_000001.10,NC_000001.11	MUTYH	683
Семейная гиперхолестеринемия	APG01658.1	570942190	C>T	NC_000019.9,NC_000019.10	LDLR	684
Семейная гиперхолестеринемия	APG06646.1	570942190	C>T	NC_000019.9,NC_000019.10	LDLR	685
Синдром атаксии-телеангиэктазии	APG01658.1	587776551	G>A	NC_000011.10,NC_000011.9	ATM	686
Синдром атаксии-телеангиэктазии	APG06646.1	587776551	G>A	NC_000011.10,NC_000011.9	ATM	687
Аденоидно-кистозная карцинома	APG01658.1	587778720	C>T	NC_000017.10,NC_000017.11	TP53	688
Аденоидно-кистозная карцинома	APG06646.1	587778720	C>T	NC_000017.10,NC_000017.11	TP53	689
Наследственный предрасполагающий к раку синдром	APG01658.1	587779075	C>T	NC_000002.11,NC_000002.12	MSH2	690
Наследственный предрасполагающий к раку синдром	APG06646.1	587779075	C>T	NC_000002.11,NC_000002.12	MSH2	691
Наследственный предрасполагающий к раку синдром	APG06646.1	587779333	T>C	NC_000007.13,NC_000007.14	PMS2	692
Синдром атаксии-телеангиэктазии	APG01658.1	587779866	A>G	NC_000011.10,NC_000011.9	ATM	693
Синдром атаксии-телеангиэктазии	APG06646.1	587779866	A>G	NC_000011.10,NC_000011.9	ATM	694
Рак молочной железы	APG01658.1	587780021	G>A	NC_000002.11,NC_000002.12	BARD1	695

Заболевания	RGN	RS#	CasI Mut.	Номер регистрации хромосомы	Обозначение гена	Милль (SEQ ID NO)
Рак молочной железы	APG06646.1	587780021	G>A	NC_000002.11,NC_000002.12	BARD1	696
Рак молочной железы-яичников	APG01658.1	587781629	G>A	NC_000013.10,NC_000013.11	BRCA2	697
Рак молочной железы-яичников	APG06646.1	587781629	G>A	NC_000013.10,NC_000013.11	BRCA2	698
Наследственный предрасполагающий к раку синдром	APG01658.1	587782144	C>T	NC_000017.10,NC_000017.11	TP53	699
Наследственный предрасполагающий к раку синдром	APG06646.1	587782144	C>T	NC_000017.10,NC_000017.11	TP53	700
Синдром Марфана	APG01658.1	727503054	A>G	NC_000015.9,NC_000015.10	FBN1	701
Синдром Марфана	APG06646.1	727503054	A>G	NC_000015.9,NC_000015.10	FBN1	702
Нунан синдром	APG01658.1	727503110	T>C	NC_000012.11,NC_000012.12	KRAS	703
Нунан синдром	APG06646.1	727503110	T>C	NC_000012.11,NC_000012.12	KRAS	704
Семейная гиперхолестеринемия	APG01658.1	746118995	C>T	NC_000019.9,NC_000019.10	LDLR	705
Семейная гиперхолестеринемия	APG06646.1	746118995	C>T	NC_000019.9,NC_000019.10	LDLR	706
Семейная гиперхолестеринемия	APG01658.1	748944640	G>A	NC_000019.9,NC_000019.10	LDLR	707
Семейная гиперхолестеринемия	APG06646.1	748944640	G>A	NC_000019.9,NC_000019.10	LDLR	708
Семейная гиперхолестеринемия	APG01658.1	765696008	G>A	NC_000019.10,NC_000019.9	LDLR	709
Семейная гиперхолестеринемия	APG06646.1	765696008	G>A	NC_000019.10,NC_000019.9	LDLR	710
Семейная гиперхолестеринемия	APG01658.1	769370816	G>A	NC_000019.10,NC_000019.9	LDLR	711
Семейная гиперхолестеринемия	APG06646.1	769370816	G>A	NC_000019.10,NC_000019.9	LDLR	712
Семейная гиперхолестеринемия	APG01658.1	769737896	C>T	NC_000019.10,NC_000019.9	LDLR	713
Семейная гиперхолестеринемия	APG06646.1	769737896	C>T	NC_000019.10,NC_000019.9	LDLR	714
Семейная гиперхолестеринемия	APG01658.1	771019366	A>G	NC_000019.10,NC_000019.9	LDLR	715
Семейная гиперхолестеринемия	APG06646.1	771019366	A>G	NC_000019.10,NC_000019.9	LDLR	716
Параганглиомы	APG01658.1	772551056	C>T	NC_000001.11,NC_000001.10	SDHB	717
Параганглиомы	APG06646.1	772551056	C>T	NC_000001.11,NC_000001.10	SDHB	718
Наследственный рак	APG01658.1	786201042	C>T	NC_000002.12,NC_000002.11	MSH6	719
Наследственный рак	APG06646.1	786201042	C>T	NC_000002.12,NC_000002.11	MSH6	720

Заболевания	RGN	RS#	CasI Mut.	Номер регистрации хромосомы	Обозначение гена	Милень (SEQ ID NO)
Семейная гиперхолестеринемия	APG01658.1	875989907	G>A	NC_000019.9,NC_000019.10	LDLR	721
Семейная гиперхолестеринемия	APG06646.1	875989907	G>A	NC_000019.9,NC_000019.10	LDLR	722
Семейная гиперхолестеринемия	APG01658.1	879254797	G>A	NC_000019.10,NC_000019.9	LDLR	723
Семейная гиперхолестеринемия	APG06646.1	879254797	G>A	NC_000019.10,NC_000019.9	LDLR	724
Семейная гиперхолестеринемия	APG01658.1	879254871	C>T	NC_000019.10,NC_000019.9	LDLR	725
Семейная гиперхолестеринемия	APG06646.1	879254871	C>T	NC_000019.10,NC_000019.9	LDLR	726
Семейная гиперхолестеринемия	APG01658.1	879255000	T>C	NC_000019.10,NC_000019.9	LDLR	727
Семейная гиперхолестеринемия	APG06646.1	879255000	T>C	NC_000019.10,NC_000019.9	LDLR	728

Пример 8. Нацеливание на мутации, ответственные за атаксию Фридрейха

Расширение последовательности тринуклеотидных повторов, вызывающее атаксию Фридрейха (FRDA – Friedreich’s Ataxia), происходит в определенном генетическом локусе в гене FXN, называемом областью нестабильности FRDA. Нуклеазы, направляемые РНК (RGN), можно использовать для вырезания области нестабильности в клетках пациентов с FRDA. Этот подход требует 1) последовательности RGN и направляющей РНК, которые можно запрограммировать для нацеливания на аллель в геноме человека; и 2) подхода к доставке RGN и направляющей последовательности. Многие нуклеазы, используемые для редактирования генома, такие как обычно используемая нуклеаза Cas9 из *S. pyogenes* (SpCas9), слишком велики, чтобы их можно было упаковать в векторы аденоассоциированных вирусов (AAV – adeno-associated viral), особенно с учетом длины гена SpCas9 и направляющей РНК в дополнение к другим генетическим элементам, необходимым для кассет функциональной экспрессии. Это делает жизнеспособный подход с использованием SpCas9 маловероятным.

Компактные РНК-направляемые нуклеазы по настоящему изобретению, такие как APG09748, APG09106.1 и APG06646.1, однозначно хорошо подходят

для удаления области нестабильности FRDA. Для каждой RGN нужен мотив РАМ, который находится в непосредственной близости от области нестабильности FRDA. Кроме того, каждая из таких RGN может быть упакована в вектор AAV вместе с направляющей РНК. Для упаковки двух направляющих РНК, вероятно, потребуется второй вектор, но этот подход все же выгодно отличается от того, что потребовалось бы для более крупной нуклеазы, такой как SpCas9, которая потребовала бы расщепления последовательности белка между двумя векторами.

В табл. 10 показано расположение геномных последовательностей-мишеней, пригодных для нацеливания APG09748, APG09106.1 или APG06646.1 на 5'- и 3'-стороны области нестабильности FRDA, а также последовательности sgРНК для геномных мишеней. Оказавшись в локусе, RGN может вырезать область нестабильности FA. Иссечение области можно проверить путем секвенирования локуса с помощью Illumina.

Таблица 10. Геномные последовательности-мишени для систем RGN

Положение относительно области нестабильности FRDA	Геномные последовательности-мишени для APG09748 или APG09106.1 (SEQ ID NO.)	sgРНК для APG09748 или APG09106.1 (SEQ ID NO.)	Геномные последовательности-мишени для APG06646.1 (SEQ ID NO.)	sgРНК для APG06646.1 (SEQ ID NO.)
5'	401	405	729	733
5'	402	406	730	734
3'	403	407	731	735
3'	404	408	732	736

Пример 9. Нацеливание на мутации, ответственные за серповидноклеточную анемию

Нацеливание последовательностей в области энхансера BCL11A (SEQ ID NO: 220) могут обеспечивать механизм повышения фетального гемоглобулина (HbF) для излечения или облегчения симптомов серповидно-клеточной анемии. Например, в результате широких геномных исследований выявили набор генетических вариаций в BCL11A, которые связаны с повышенными уровнями HbF. Эти вариации представляют собой набор SNP, обнаруженных в некодирующих областях BCL11A, которые функционируют как специфические в отношении стадии, области энхансера, ограниченные происхождением. Дальнейшее исследование показало, что энхансер BCL11A необходим в

эритроидных клетках для экспрессии BCL11A (публикация Bauer с соавт., *Science*, 2013, 343:253-257, включена в настоящее описание в качестве ссылки). Эхансерная область обнаружена в интроне 2 гена BCL11A, и в интроне 2 были идентифицированы три области гиперчувствительности к ДНКазе I (часто свидетельствующие о состоянии хроматина, связанном с регуляторным потенциалом). Эти три области были идентифицированы как «+62», «+58» и «+55» в соответствии с расстоянием в килобазах от сайта начала транскрипции BCL11A. Такие эхансерные области состоят примерно из 350 (+55); 550 (+58); и 350 (+62) нуклеотидов в длину (Bauer с соавт., 2013).

10 Пример 9.1. Идентификация предпочтительных систем RGN

В настоящем изобретении описывают возможное лечение бета-гемоглобинопатий с использованием системы RGN, которая нарушает связывание BCL11A с его сайтом связывания в локусе HBB, который является геном, ответственным за образование бета-глобина в гемоглобине взрослых. Этот подход использует NHEJ, что более эффективно в клетках млекопитающих. Кроме того, в этом подходе используют нуклеазу достаточно малого размера, которую можно упаковать в единый вектор AAV для доставки *in vivo*.

Эхансерный мотив *gata1* в эхансерной области BCL11A человека (SEQ ID NO: 220) является идеальной мишенью для разрушения с использованием РНК-направляемых нуклеаз (RGN) для снижения экспрессии BCL11A с одновременной реэкспрессией HbF в эритроцитах взрослого человека (Wu с соавт. (2019) *Nat Med* 387:2554). Несколько последовательностей PAM, совместимых с APG09748 или APG09106.1, легко обнаруживают в генетическом локусе, окружающем этот сайт GATA1. Эти нуклеазы содержат последовательность PAM (5'-DTTN-3'; SEQ ID NO: 60) и имеют компактный размер, что потенциально позволяет доставлять их вместе с соответствующей направляющей РНК в одном AAV или аденовирусном векторе. Помимо своего размера APG06646.1 имеет минимальные требования к PAM (SEQ ID NO: 109), что делает его хорошо подходящим для такого методического подхода. Этот подход к доставке дает множество преимуществ по сравнению с другими, такими как доступ к гемопоэтическим стволовым клеткам и хорошо зарекомендовавший себя профиль безопасности и методы получения.

Обычно используемая нуклеаза Cas9 из *S. pyogenes* (SpyCas9) требует последовательности PAM 5'-NGG-3' (SEQ ID NO: 101), некоторые из которых

расположены рядом с мотивом GATA1. Однако размер нуклеазы SpyCas9 не позволяет упаковывать её в единый вектор AAV или вектор аденовируса и, таким образом, нивелирует вышеупомянутые преимущества этого подхода. Хотя можно использовать стратегию двойной доставки, это значительно усложняет и удорожает процесс. Кроме того, доставка двойного вирусного вектора значительно снижает эффективность коррекции генов, поскольку для успешного редактирования в данной клетке требуется заражение обоими векторами.

Получают экспрессирующую кассету, кодирующую оптимизированную по кодонам человека APG09748 (SEQ ID NO: 409), APG09106.1 (SEQ ID NO: 410) или APG06646.1 (SEQ ID NO: 415), подобно описанному в примере 5. Также получают экспрессирующие кассеты, которые экспрессируют направляющие РНК для RGN APG09748, APG09106.1 или APG06646.1. Эти направляющие РНК включают: 1) последовательность протоспейсера, которая комплементарна либо некодирующей, либо кодирующей цепи ДНК в локусе энхансера BCL11A (последовательность-мишень) и 2) последовательность РНК, необходимую для ассоциации направляющей РНК с RGN. Поскольку несколько потенциальных последовательностей PAM для нацеливания с помощью каждой RGN окружают мотив энхансера BCL11A GATA1, получают несколько потенциальных конструкций направляющих РНК для определения наилучшей последовательности протоспейсера, которая обеспечивает надежное расщепление и опосредованное NHEJ нарушение последовательности энхансера BCL11A GATA1. Целевые геномные последовательности в табл. 11 оценивают с использованием sgРНК, представленной в табл. 11.

Таблица 11. Последовательности-мишени для локуса энхансера BCL11A GATA1 с использованием APG06646.1

Геномные последовательности-мишени для APG09748 или APG0916.1 (SEQ ID NO.)	sgРНК для APG09748 или APG0916.1 (SEQ ID NO.)	Геномные последовательности-мишени для APG06646.1 (SEQ ID NO.)	sgРНК для APG06646.1 (SEQ ID NO.)
221	224	737	740
222	269	738	741
223	270	739	742

Для оценки эффективности, с которой APG09748, APG09106.1 или APG06646.1 производят инсерции или делеции, разрушающие область энхансера

BCL11A, используют клеточные линии человека, такие как эмбриональные клетки почки человека (клетки НЕК). Получают вектор ДНК, содержащий кассету экспрессии RGN (например, согласно описанию в примере 5). Также получают отдельный вектор, содержащий кассету экспрессии, включающую кодирующую последовательность для направляющей последовательности РНК из табл. 11. Такая экспрессирующая кассета может дополнительно содержать промотор РНК-полимеразы III U6 человека (SEQ ID NO: 308), как описано в примере 5. В другом варианте можно использовать один вектор, содержащий экспрессирующие кассеты как RGN, так и направляющей РНК. Вектор вводят в клетки НЕК с использованием стандартных методик, таких как описанные в примере 5, и клетки культивируют в течение 1-3 дней. После этого периода культивирования выделяют геномную ДНК и определяют частоту инсерций или делеций с использованием расщепления эндонуклеазой I фага T7 и/или прямого секвенирования ДНК, как описано в примере 5.

Область ДНК, охватывающую целевую область BCL11A, амплифицируют с помощью ПЦР с праймерами, содержащими выступающие последовательности Illumina Nextera XT. Эти ампликоны ПЦР либо исследуют на предмет образования NHEJ с помощью расщепления эндонуклеазой I T7, либо подвергают подготовке библиотеки в соответствии с протоколом библиотеки метагеномного секвенирования Illumina 16S или аналогичной подготовкой библиотеки для секвенирования следующего поколения (NGS – Next Generation Sequencing). После глубокого секвенирования сгенерированные прочтения анализируют с помощью программы CRISPResso для расчета степени редактирования. Выравнивание выходных данных проверяют вручную для подтверждения сайтов инсерций и делеций. Этот анализ идентифицирует предпочтительную RGN и соответствующую предпочтительную направляющую РНК (sgРНК). Анализ может привести к тому, что либо APG09748, APG09106.1 или APG06646.1 будут в равной степени предпочтительными, либо одна RGN будет наиболее предпочтительной. Кроме того, анализ может определить, что существует более одной предпочтительной направляющей РНК или что все целевые геномные последовательности в табл. 17 являются одинаково предпочтительными.

Пример 9.2. Анализ экспрессии фетального гемоглобина

В этом примере APG09748, APG09106.1 или APG06646.1 создают инсерции или делеции, разрушающие энхансерную область BCL11A, для анализа экспрессии фетального гемоглобина. Используют CD34⁺ гемопоэтические стволовые клетки (HSC – hematopoietic stem cells) здоровых людей-доноров.

5 HSC культивируют и вводят вектор (векторы), содержащие кассеты экспрессии, включающие кодирующие области предпочтительной RGN и предпочтительной sgРНК, с использованием способов, аналогичных описанным в примере 5. В другом варианте можно использовать электропорацию. После электропорации клетки дифференцируют *in vitro* в эритроциты с использованием установленных

10 протоколов (например, описанных в работе Giarratana с соавт., *Nat Biotechnology*, 2004, 23:69-74, включенной в настоящее изобретение в виде ссылки). Экспрессию HbF затем измеряют с помощью вестерн-блоттинга с антителом против HbF человека или количественно определяют с помощью высокоэффективной жидкостной хроматографии (ВЭЖХ). Ожидают, что

15 успешное разрушение локуса энхансера BCL11A приведет к увеличению продукции HbF по сравнению с HSC, подвергнутыми электропорации только RGN, но без направляющей.

Пример 9.3. Анализ пониженного образования серповидных клеток

В этом примере APG09748, APG09106.1 или APG06646.1 создают инсерции

20 или делеции, разрушающие энхансерную область BCL11A, исследуемую на снижение образования серповидных клеток. Используют CD34⁺ гемопоэтические стволовые клетки (HSC – hematopoietic stem cells) от пациентов с серповидно клеточной анемией. HSC культивируют и вводят вектор (векторы), содержащие кассеты экспрессии, включающие кодирующие области предпочтительной RGN

25 и предпочтительной sgРНК, с использованием способов, аналогичных описанным в примере 5. В другом варианте можно использовать электропорацию. После электропорации клетки дифференцируют *in vitro* в эритроциты с использованием установленных протоколов (например, описанных в работе Giarratana с соавт., *Nat Biotechnology*, 2004, 23:69-74, включенной в

30 настоящее изобретение в виде ссылки). Экспрессию HbF затем измеряют с помощью вестерн-блоттинга с антителом против HbF человека или количественно определяют с помощью высокоэффективной жидкостной хроматографии (ВЭЖХ). Ожидают, что успешное разрушение локуса энхансера

BCL11A приведет к увеличению выработки HbF по сравнению с HSC, подвергнутыми электропорации только RGN, но без направляющей.

Образование серповидных клеток индуцируется в этих дифференцированных эритроцитах добавлением метабисульфита. Количество серповидных и нормальных эритроцитов подсчитывают с помощью микроскопии. Ожидают, что количество серповидных клеток меньше в клетках, обработанных APG09748, APG09106.1 или APG06646.1 плюс sgPНК, чем в необработанных клетках или обработанных только RGN.

Пример 9.4. Анализ лечения заболевания на мышинной модели

Для оценки эффективности разрушения локуса BCL11A ферментами APG09748, APG09106.1 или APG06646.1, применяют приемлемую модель гуманизированных мышей с серповидно клеточной анемией. Кассеты экспрессии, кодирующие предпочтительную RGN и предпочтительную sgPНК, упаковывают в векторы AAV или аденовирусные векторы. В частности, аденовирус типа Ad5/35 эффективен в отношении HSC. Выбирают подходящую модель мыши, содержащую гуманизированный локус HBB с аллелями серповидных клеток, такую как B6; FVB-Tg(LCR-HBA2,LCR-HBB*E26K)53Hhb/J или B6.Cg-Hbatm1Paz Hbbtm1Tow Tg(HBA-HBBs)41Paz/HhbJ. Этим мышам вводят гранулоцитарный колониестимулирующий фактор отдельно или в комбинации с плериксафором для мобилизации HSC в кровообращение. Затем внутривенно вводят AAV или аденовирусы, несущие RGN и направляющую плазмиду, и мышам дают выздороветь в течение недели. Кровь, полученную от этих мышей, тестируют для определения серповидных клеток *in vitro* с использованием метабисульфита, и за мышами длительно наблюдают для фиксирования показателей смертности и оценке гемопоэтической функции. Ожидают, что лечение AAV или аденовирусами, несущими RGN и направляющую PНК, уменьшит серповидность, смертность и улучшит гемопоэтическую функцию по сравнению с мышами, обработанными вирусами, лишенными обеих кассет экспрессии, или вирусами, несущими только кассету экспрессии RGN.

Пример 10. Активность по редактированию оснований в клетках млекопитающих

Кассету экспрессии, которая вырабатывает слитый белок цитидиндезаминаза-RGN, конструируют следующим образом. Кодирующая

последовательность для RGN (APG06646.1, описанная в настоящем изобретении, и APG08290.1, описанная в публикации заявки US 2019/0367949 и публикации международной заявки WO2019/236566, каждая из которых включена в настоящее изобретение в виде ссылки) оптимизирована по кодонам для экспрессии у млекопитающих и мутирована так, чтобы функционировать в качестве никазы (SEQ ID NO: 128 и 262, соответственно). Эту кодирующую последовательность вводят в кассету экспрессии, которая вырабатывает слитый белок, содержащий NLS на своем N-конце (SEQ ID NO: 125), функционально связанный на своем C-конце с меткой 3Xflag (SEQ ID NO: 126), функционально связанную на своем C-конце с цитидиндезаминазой (SEQ ID NO: 129-132, все из которых описаны в международной заявке PCT/US 2019/068079, сущность которой в полной мере включена в настоящее изобретение посредством ссылки), функционально связанной своим C-концом с аминокислотным линкером (SEQ ID NO: 133), функционально связанным своим C-концом с RGN-никазой, функционально связанной своим C-концом со второй NLS (SEQ ID NO: 127). Каждую из этих кассет экспрессии вводят в вектор pTwist CMV (фирма Twist Bioscience), способный управлять экспрессией слитого белка в клетках млекопитающих. Также получают отдельные векторы, экспрессирующие направляющие РНК под контролем промотора U6 человека в клетках млекопитающих. Эти направляющие РНК (SEQ ID NO: 134-136 для Narg06646.1 и SEQ ID NO: 263-265 для Narg08290.1) способны направлять слитые белки дезаминазы-Nrgn или саму RGN к целевой геномной последовательности для редактирования оснований или редактирования гена, соответственно.

Трансфецируют совместно 500 нг плазмид экспрессии цитидиндезаминазы RGN или стандартной плазмиды экспрессии RGN и 500 нг плазмид экспрессии направляющей РНК в клетки HEK293FT при 75-90% слиянии клеток в 24-луночных планшетах с использованием реагента Lipofectamine 2000 (фирма Life Technologies). Life). Затем клетки инкубируют при 37°C в течение 72 ч. Затем экстрагируют геномную ДНК, используя NucleoSpin 96 Tissue (фирма Macherey-Nagel) в соответствии с протоколом производителя. Геномную область, фланкирующую сайт-мишень направляющей РНК, амплифицируют с помощью ПЦР, и продукты очищают с помощью ZR-96 DNA Clean and Concentrator (фирма Zymo Research) в соответствии с протоколом производителя. Очищенные продукты ПЦР затем отправляют на секвенирование следующего поколения на

Illumina MiSeq (2x250). Результаты анализируют на предмет образования индел или специфической мутации цитозина.

В табл. 12 ниже показано редактирование цитидинового основания и образование индел для каждой комбинации конструкции и направляющей.

5 Интересно, что при сравнении активности RGN APG06646.1 с активностью цитидиндезаминазы-Napg06646.1 с использованием той же направляющей РНК
10 наблюдают до 20 раз более высокое редактирование цитидиновых оснований, чем редактирование гена. Эти результаты демонстрируют, что RGN, которая имеет низкую нуклеазную активность в конкретном сайте-мишени, все же может
15 быть эффективным редактором оснований в этом сайте. Кроме того, поскольку образование инделей в приложениях для редактирования оснований часто является нежелательным результатом, RGN, которые имеют низкую нуклеазную активность на сайте, могут быть предпочтительными для приложений по редактированию оснований.

Таблица 12. Редактирование оснований цитидина и редактирование генов RGN и цитидиндезаминазы-Nrgn

RGN или дезаминаза-RGN	Направляющая РНК	% общего числа прочтений с редактированием оснований цитидина	% общего числа прочтений с формированием индел
ARM05-Napg06646.1	SGN000775	8,02	0,82
ARM06CTD-Napg06646.1	SGN000775	0	0
ARM08-Napg06646.1	SGN000775	5,15	0,46
ARM11-Napg06646.1	SGN000775	0,66	0
APG06646.1	SGN000775	0	0,47
ARM05-Napg06646.1	SGN000777	2,5	0,08
ARM06CTD-Napg06646.1	SGN000777	0,53	0
ARM08-Napg06646.1	SGN000777	8,12	0,25
ARM11-Napg06646.1	SGN000777	3,52	1,04
APG06646.1	SGN000777	0	1,44
ARM05-Napg06646.1	SGN000781	13,51	0,27
ARM06CTD-Napg06646.1	SGN000781	2,69	0,13
ARM08-Napg06646.1	SGN000781	12,97	0,12
ARM11-Napg06646.1	SGN000781	4,7	0,14
APG06646.1	SGN000781	0	0,65
ARM05-Napg08290.1	SGN000143	1,35	7,04

RGN или дезаминаза-RGN	Направляющая РНК	% общего числа прочтений с редактированием оснований цитидина	% общего числа прочтений с формированием индел
ARM06CTD-Napg08290.1	SGN000143	1,36	3,17
ARM08-Napg08290.1	SGN000143	2,59	10,98
ARM11-Napg08290.1	SGN000143	4,78	5,77
APG08290.1	SGN000143	0	6,21
ARM05-Napg08290.1	SGN000169	4,26	12,87
ARM06CTD-Napg08290.1	SGN000169	1,63	8,29
ARM08-Napg08290.1	SGN000169	6,99	14,27
ARM11-Napg08290.1	SGN000169	4,99	6,39
APG08290.1	SGN000169	0	11,85
ARM05-Napg08290.1	SGN000173	3,26	5,51
ARM06CTD-Napg08290.1	SGN000173	0,72	0,96
ARM08-Napg08290.1	SGN000173	6,6	6,06
ARM11-Napg08290.1	SGN000173	2,83	4,82
APG08290.1	SGN000173	0	2,58

Пример 11. Расщепление транс оцДНК

11.1 Исследование условий для расщепления транс ДНК

Очищенную нуклеазу APG09748 (представленную как SEQ ID NO: 137 и описанная в международной заявке PCT/US2019/068079, которая полностью включена в настоящее изобретение в качестве ссылки) инкубируют с 5 однонаправляющей РНК (sgРНК) в буфере Cutsmart (фирма New England Biolabs B7204S) при конечной концентрации либо 50 нМ нуклеазы и 100 нМ sgРНК, либо 200 нМ нуклеазы и 400 нМ sgРНК в течение 10 мин. Эти растворы РНП 10 добавляют к растворам оцДНК – целевой или несовместимый отрицательной контрольной оцДНК – в конечной концентрации 10 нМ и репортерным зондам при конечной концентрации 250 нМ в буфере 1,5X Cutsmart (фирма New England Biolabs B7204S). Репортерные зонды (ТВ0125 и ТВ0089, представленные как SEQ ID NO: 138 и 139, соответственно) содержат флуоресцентный краситель на 15 5'-конце (56-FAM для ТВ0125 и Cy5 для ТВ0089), гаситель на 3'-конце (3IABkFQ для ТВ0125 и 3IABRQSp для ТВ0089) и необязательно внутренний гаситель (внутренний гаситель ZEN присутствует только на ТВ0125). Расщепление репортерного зонда приводит к отсутствию гашения флуоресцентного красителя

и, таким образом, к увеличению флуоресцентного сигнала. Для мониторинга интенсивности флуоресценции 10 мкл каждой реакции инкубируют в 384-луночном микропланшете Corning малого объема при 30°C в микропланшетном ридере (фирма CLARIOstar Plus).

5 Изучен ряд условий, чтобы определить подходящие параметры данного анализа. Для того, чтобы определить, есть ли влияние дизайна зонда гашения или характеристик флуорофора, два таких репортера включают в виде смеси в каждую реакцию. Они содержатся в той же равной концентрации в любой
10 данной реакции. Во всех случаях контрольная или целевая концентрация оцДНК (LE201 или LE205, представленные как SEQ ID No: 140 и 141, соответственно) составляют 10 нМ. Названия РНП обозначают нуклеазу и мишень, как указано в табл. 13 ниже.

Таблица 13. Рибонуклеопротеиновые комплексы

РНП	Нуклеаза	sgРНК	Предполагаемая мишень
APG09748.1	APG09748	27sg.1 (SEQ ID NO: 144)	LE201 (SEQ ID NO: 140)
APG09748.2	APG09748	27sg.2 (SEQ ID NO: 145)	LE205 (SEQ ID NO: 141)

15 Результаты показаны в табл. 14 ниже.

11.2 Расщепление транс-ДНК нуклеазой APG09748 и влияние очистки на неспецифическую активность

Очищенную нуклеазу APG09748 инкубируют с однонаправленной РНК (sgРНК) в буфере 1X Cutsmart (фирма New England Biolabs B7204S) в конечной
20 концентрации 200 нМ нуклеазы и 400 нМ sgРНК в течение 10 мин при 37°C.

Затем эти растворы РНП добавляют к растворам оцДНК – мишени несоответствующего отрицательного контроля оцДНК – в конечной концентрации 10 нМ и репортерного зонда (TP0003) в конечной концентрации 250 нМ в буфере 1.5X Cutsmart (фирма New England Biolabs B7204S).

25 Репортерный зонд содержит флуоресцентный краситель на 5'-конце и гаситель на 3'-конце. Расщепление репортерного зонда приводит к отрицанию гашения флуоресцентного красителя и, таким образом, к увеличению флуоресцентного сигнала. Для мониторинга интенсивности флуоресценции по 10 мкл каждой реакции инкубируют в 384-луночном микропланшете Corning при 37°C в ридере
30 для микропланшетов (фирма CLARIOstar Plus).

Инкубация с последовательностями-мишенями приводит к значительному увеличению интенсивности флуоресценции как функции времени по сравнению

с отрицательным контролем. Скорость расщепления определяют как наклон линейной части флуоресценции относительно времени как показано в табл. 15.

Таблица 15. Результаты анализа по расщеплению транс-ДНК (ОЕФ – относительные единицы флуоресценции)

Последовательность субстрата оцДНК	Наклон (ОЕФ/мин)
Несоответствие (LE111; SEQ ID NO: 142)	549
Соответствие (LE113; SEQ ID NO: 143)	7724

5

Эти данные показывают дифференциацию целевой последовательности относительно отрицательного контроля с помощью РНП, которые четко обнаруживаются с помощью расщепленного репортерного зонда.

11.3 Определение РАМ, используя параллельную библиотеку плазмидной ДНК

10

Олигонуклеотиды (LE00680 и LE00688, представленные как SEQ ID NO: 266 и 267, соответственно), содержащие целевую последовательность, которой предшествует вырожденная последовательность РАМ длиной 5 нуклеотидов (NNNNN), отжигают и клонируют в плазмиде Puc19 двойного расщепления, используя набор NEBuilder HiFi DNA Assembly Master Mix (фирма New England Biolabs). Каждая колония, полученная в результате трансформации этой реакции, соответствует последовательности клональной плазмидной ДНК таким образом, что препараты плазмидной ДНК из культур, происходящих из отдельных колоний, являются уникальными препаратами плазмид, взятыми из исходной библиотеки. Препараты плазмид получают из выборки в 96 колоний. Эти препараты по отдельности подвергают секвенированию по Сэнгеру для проверки последовательности РАМ. Очищенную APG09748 инкубируют с sgРНК (27sg.2) в буфере 1X Cutsmart (фирма New England Biolabs B7204S) в конечной концентрации 200 нМ нуклеазы и 400 нМ sgРНК в течение 20 мин при комнатной температуре.

15

20

25

Эти растворы РНП добавляют до конечной концентрации 100 нМ к растворам плазмидной ДНК-мишени в конечной концентрации 8,3 нМ и репортеры ТВ0125 и ТВ0089 (представленные как SEQ ID NO: 138 и 139, соответственно) в концентрации 250 нМ и 50 нМ, соответственно, в буфере 1.5X Cutsmart (фирма New England Biolabs B7204S). Для мониторинга интенсивности флуоресценции по 10 мкл каждой реакции инкубируют в 384-луночном

30

микропланшете Corning при 37°C в ридере для микропланшетов (фирма CLARIOstar Plus).

Концентрация плазмиды в небольшом количестве образцов была либо ниже целевого значения для нормализации концентрации, либо объема раствора было
 5 недостаточно для доставки предполагаемого целевого количества в реакционную лунку. Эти образцы не исключают из анализа, и поэтому ожидают, что они будут способствовать возникновению ошибки из-за несоответствия между дубликатами. Образцы, в которых используют плазмиды, определенные после секвенирования по Сэнгеру как имеющие перекрестное загрязнение
 10 (множественные следы, очевидные в области PAM) или изменения в последовательности-мишени, удаляют из анализа, и их результаты не показаны ниже. Результаты анализа показаны в табл. 16 ниже в порядке убывания наклона в канале FAM.

Таблица 16. Последовательности PAM (ОЕФ – относительные единицы
 15 флуоресценции)

PAM	SEQ ID NO	Наклон (ОЕФ/мин)	
		Канал Cy5	Канал FAM
АТАТГ	147	111,52	1728,13
АТАТГ	147	102,42	1612,86
ТАТТГ	148	137,15	846,12
ТАТТГ	148	148,61	799,82
СГТТС	149	151,06	744,18
СГТТС	149	134,61	735,35
ТТТТТ	150	167,71	731,11
ТТТТТ	150	149,83	711,78
СААТС	151	120,84	675,71
ААТСТ	152	134,67	622,52
ААТСТ	152	134,1	620,2
СААТС	151	96,94	608,58
АААТТ	153	152,6	607,24
АААТТ	153	150,57	591,77
САТСС	154	97,34	588,7
АGАТА	155	214,69	579,56
ТАТТС	156	124,22	570,81
АТТТТ	157	134,72	551,91
АGАТА	155	198,45	544,81
ТАТТС	156	108,57	538,18
ТСТТТ	158	141,66	532,87

PAM	SEQ ID NO	Наклон (ОЕФ/мин)	
		Канал Су5	Канал FАM
ATTTT	157	107,9	526,75
AGATC	159	106,93	497,71
AATAA	160	78,94	484,67
TCTTT	158	117,19	484,13
GATCC	161	112,85	479,91
GATCC	161	105,04	474,94
AGATC	159	106,5	473,98
ATTCA	162	115,03	473,24
AATAA	160	77,43	465,75
AAAAT	163	78,3	458,23
TTATC	164	102,14	456,73
TATCA	165	106,45	448,78
TCACT	166	167,02	429,36
CTTAT	167	82,9	418,38
TATTG	148	116,08	417,94
TCACT	166	160,29	415,64
AAAAT	163	56,43	411,26
TGATA	168	89,02	406,14
TGATA	168	85,99	395,57
TACTA	169	90,61	391,06
ATTTT	157	113,18	382,59
TACTA	169	74,33	364,56
ATTTT	157	106,72	364,02
TATTG	148	93,95	341,36
CTTAT	167	76,12	336,8
ATATA	170	83,39	326,86
ATATA	170	72,98	318,2
TATCA	165	65,41	317,53
GACTC	171	56,56	302,2
GCTCA	172	22,1	293,68
TAACC	173	56,08	283,99
CGTGT	174	66,57	268,89
TAACC	173	56,55	265,45
TTTCG	175	59,55	262,91
CGTGT	174	50,71	262,37
TTTCG	175	61,42	261,65
CAGTA	176	51,81	240,09
GCTCA	172	19,08	228,08
CAGTA	176	51,09	220,25
CCCTT	177	17,83	136,69

PAM	SEQ ID NO	Наклон (ОЕФ/мин)	
		Канал Су5	Канал FАМ
ССТАТ	178	40,4	125,81
ТСССТ	179	39,93	124,6
АGGТТ	180	40,96	117,18
АGGТТ	180	40,81	114,79
ТААСG	181	32,53	94,68
ААСАG	182	8,19	91,72
ССТАТ	178	34,51	87,57
ТGААС	183	10,22	86,38
ТААGТ	184	19,35	85,12
ТСССТ	179	26,67	83,02
ТААСG	181	27,62	79,17
ТААСG	181	28,89	78,55
ТААСG	181	22,23	78,36
СТGТА	185	23,27	77,06
АGGGG	186	30,8	74,4
GCTGT	187	28,3	71,54
АТСТА	188	18,73	69,15
GCTGT	187	25,61	69,06
ТТАТС	164	22,8	68,13
ААСАG	182	23,98	67,27
ТСАGG	189	17,25	67,04
АТGТС	190	13,91	65,93
ТGСТА	191	26,33	65,73
ТАААА	192	17,32	65,43
АТСТА	188	16,88	65,3
СТGТА	185	21,81	64,15
GACTC	171	24,05	64,02
АТТСА	162	25,04	63,87
ТGААС	183	24,53	63,64
СССТТ	177	21,35	61,95
АТGТС	190	12,99	60,6
АТСАG	193	22,32	59,63
АТСАG	193	18,55	56,58
СGGGG	194	27,86	56,28
ТАААА	192	17,72	56,03
ТТСТС	195	25,7	55,83
GTCAT	196	23,07	55,76
GCGTG	197	17,36	55,12
GCGTG	197	22,02	55,02
GCAAT	198	15,47	54,91

PAM	SEQ ID NO	Наклон (ОЕФ/мин)	
		Канал Су5	Канал FAM
TTCTC	199	21,89	54,89
GTCAT	196	17,22	54,34
GCAAT	198	17,36	53,32
AGCAT	200	22,47	52,23
GTCAA	201	14,14	52,12
AGAGC	202	19,62	51,79
AGAGC	202	15,99	50,37
GTGGG	203	22,38	50,1
CTGAA	204	21,24	49,45
CTGAT	205	12,05	49,32
ATGAA	206	17,93	49,15
ACCTT	207	19,66	49,12
ACCTT	207	22,74	49,12
GTCAA	201	13	48,21
AACCA	208	17,81	48,04
AACCA	208	15,75	47,87
GTAAG	209	16,3	47,85
TAGGA	210	18,68	47,73
CCGAA	211	20,24	47,58
AGCAT	200	22,95	47,52
CGGGG	194	20,94	47,41
GTCAT	196	15,17	47,24
CTGAA	204	20,4	46,36
GTGGG	203	19,98	46,25
TACGA	212	16,79	46,18
TGGCT	213	14,81	45,75
GTCAT	196	15,55	45,51
ATGAA	206	19,11	45,09
TAGGA	210	21,65	44,94
TTCTC	199	21,44	44,76
CCAAG	214	16,98	44,61
GTAAG	209	15,35	44,59
TGCTT	215	15,79	44,54
CTAAG	216	12,72	44,28
TGCTA	191	16	44,14
TACGA	212	16,07	44,06
TCGAT	217	22,28	43,96
TTCTC	199	15,94	42,55
CTGAT	205	16,13	42,27
TCAGG	189	11,38	41,9

PAM	SEQ ID NO	Наклон (ОЕФ/мин)	
		Канал Су5	Канал FAM
ССАAG	214	11,28	41,45
AGGGG	186	16,04	41,44
TCGAT	217	18,68	40,5
CCGAA	211	15,39	39,09
СТАAA	216	10,72	30,37
GТАAC	218	4,36	9,51
TGCTT	215	-0,32	-0,7
ТАAGT	184	-0,57	-1,3
САТСС	154	0,47	-1,95
TGGCT	213	-0,03	-2,2
GТАAC	219	-0,46	-2,3

Последовательности с наивысшим наклоном, по-видимому, соответствуют предсказанному PAM (DTTN, представленному как SEQ ID NO: 60), определенному с помощью анализа истощения плазмиды, ранее описанному в международной заявке PCT/US2019/068079, сущность которой полностью включена в настоящее описание в виде ссылки. В частности, определено сильное предпочтение «Т» в положении двух нуклеотидов 5' мишени. Неожиданно наиболее активный сайт PAM, обнаруженный в этом анализе (ATATG, SEQ ID NO: 147), не полностью соответствует консенсусу DTTN (SEQ ID NO: 60), что предполагает некоторый уровень гибкости этого сайта узнавания.

11.4. Активность расщепления транс-оцДНК в присутствии ПЦР-амплифицированной ДНК

Амплифицированные ПЦР мишени получают из геномной ДНК (ампликоны TRAC и VEGF, представленные в виде SEQ ID NO: 225 и 226, соответственно), используя соответствующие праймеры. Мишень VEGF (представленная как SEQ ID NO: 227) амплифицирована ПЦР из геномной ДНК клеток НЕК293Т с применением праймеров LE573 и LE578 (представленные как SEQ ID NO: 228 и 229, соответственно). Мишень TRAC (представленная как SEQ ID NO: 230) является амплифицированной ПЦР с применением LE257 and LE258 (представленные как SEQ ID NO: 231 и 232, соответственно).

РНП получают путем инкубирования нуклеазы APG09106.1, описанной в настоящем изобретении, и sgРНК в количестве 0,5 мкМ и 1 мкМ,

соответственно, в буфере 1X NEBuffer 2 (фирма New England Biolabs) и инкубируют при комнатной температуре в течение 20 мин.

Таблица 17. Рибонуклеопротеиновые комплексы

РНП	Нуклеаза	Направляющая РНК	Предполагаемая мишень
APG09106.2	APG09106.1	27sg.2 (SEQ ID NO: 145)	LET126 ампликон Рандомизированный ампликон РАМ
APG09106.836	APG09106.1	27sg.836 (SEQ ID NO: 233)	VEGF ампликон
APG09106.838	APG09106.1	27sg.838 (SEQ ID NO: 234)	TRAC ампликон

5 Реакцию расщепления проводят в буфере 1,5X NEBuffer 2 с 1,5 мкМ олигонуклеотидного репортера оцДНК с меткой 5' TEX615, гасителем 3' Iowa Black FQ и 100 нМ соответствующего продукта ПЦР. Расщепление репортерного зонда приводит к отмене гашения флуоресцентного красителя и, таким образом, к увеличению флуоресцентного сигнала. Для мониторинга интенсивности флуоресценции 10 мкл каждой реакции инкубируют в 384-луночном микропланшете Corning малого объема при 37°C в ридере для микропланшетов (фирма CLARIOstar Plus). Результаты кинетического анализа показаны в табл. 18.

15 Таблица 18. Результаты исследования по расщеплению транс-ДНК (ОЕФ – относительные единицы флуоресценции)

РНП	Концентрация мишени (нМ)	Мишень	Наклон (ОЕФ/мин)
28.836	30	TRAC ампликон	1204
28.838	30	TRAC ампликон	8436
28.838	30	VEGF ампликон	766
28.836	30	VEGF ампликон	3484

20 Эти результаты указывают на специфическую активацию активности по расщеплению транс-оцДНК в присутствии ПЦР-амплифицированной ДНК из различных источников. Активность зависит от концентрации ПЦР-амплифицированного субстрата.

11.5 Применение расщепления оцДНК в качестве диагностикума

Благодаря способности этих нуклеаз генерировать оптически выявляемый сигнал в присутствии последовательности ДНК-мишени их можно внедрять в

диагностические устройства для обнаружения генетического заболевания или агентов инфекционных заболеваний, таких как бактерии, вирусы или грибки.

5 Диагностическая процедура может включать выделение или амплификацию нуклеиновых кислот из тестируемого образца. Также может оказаться целесообразным использовать некоторые образцы без выделения или очистки нуклеиновых кислот, поскольку они могут присутствовать в образце в достаточно больших количествах, чтобы их можно было обнаружить без амплификации (например, ПЦР) или без материалов, мешающих обнаружению или выработке сигналов.

10 РНП, сформированные, как описано в других примерах, затем можно было бы подвергнуть воздействию образца (или процессированного образца, как описано в предыдущем параграфе) вместе с репортером, таким как олигонуклеотиды оцДНК, модифицированные по флуорофору-гасителю, использованные в предыдущих примерах, или каким-либо другим видом
15 субстрата оцДНК, который производит видимый или иным образом легко обнаруживаемый сигнал при расщеплении. При использовании олигонуклеотидов ДНК, конъюгированных с флуорофором-гасителем (как в ранее описанных примерах), их можно обнаружить с помощью флуориметра, как описано в предыдущих примерах. Чтобы упростить обнаружение, вместо
20 кинетических анализов, описанных выше, можно проводить анализ конечной точки, что означает, что анализы можно проводить в течение фиксированного времени и считывать по прошествии этого времени относительно положительного и отрицательного контроля.

25 Эти реагенты также могут быть интегрированы в устройство для тестирования с боковым потоком, которое позволяет обнаруживать данный возбудитель болезни или конкретную последовательность нуклеиновой кислоты (например, больной аллель у индивидуума) с очень небольшим набором инструментов. В этом анализе репортер оцДНК может быть конъюгирован с несколькими молекулами, подходящими для захвата антителами или аффинными
30 реагентами, такими как флуоресцеин, биотин и/или дигоксигенин.

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Молекула нуклеиновой кислоты, включающая полинуклеотид, кодирующий полипептид РНК-направляемой нуклеазы (RGN), где указанный полинуклеотид включает нуклеотидную последовательность, кодирующую полипептид RGN, выбранный из группы, состоящей из:

а) полипептида RGN, включающего аминокислотную последовательность, которая по меньшей мере на 90% идентична SEQ ID NO: 61, 30, 75, 1, 9, 16, 23, 38, 46, 69, 82, 89, 95, 103 или 110; и

б) полипептида RGN, включающего аминокислотную последовательность, которая по меньшей мере на 99% идентична SEQ ID NO: 54.

2. Молекула нуклеиновой кислоты по п. 1, в которой указанный полипептид RGN способен связываться с последовательностью ДНК-мишени молекулы ДНК специфичным для РНК-направляющей последовательности образом при связывании с направляющей РНК (gРНК), способной гибридизироваться с указанной последовательностью ДНК-мишени.

3. Молекула нуклеиновой кислоты по п. 1 или п.2, в которой указанный полинуклеотид, кодирующий полипептид RGN, функционально связан с промотором, гетерологичным указанному полинуклеотиду.

4. Молекула нуклеиновой кислоты по любому из п.п. 1-3, в которой указанный полипептид RGN, включает аминокислотную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 61, 30, 75, 1, 9, 16, 23, 38, 46, 69, 82, 89, 95, 103 или 110.

5. Молекула нуклеиновой кислоты по любому из п.п. 1-4, в которой указанный полипептид RGN, включает аминокислотную последовательность SEQ ID NO: 61, 30, 75, 1, 9, 16, 23, 38, 46, 69, 82, 89, 95, 103, 110 или 54.

6. Молекула нуклеиновой кислоты по любому из п.п. 1-4, в которой указанный полипептид RGN является неактивной нуклеазой или способен функционировать в качестве нуклеазы.

5 7. Молекула нуклеиновой кислоты по любому из п.п. 1-6, в которой полипептид RGN функционально слит с полипептидом с отредактированными основаниями.

10 8. Молекула нуклеиновой кислоты по п. 7, в которой указанный полипептид с отредактированными основаниями представляет собой дезаминазу.

9. Молекула нуклеиновой кислоты по любому из п.п. 1-6, в которой указанный полипептид RGN функционально слит с эффекторным доменом.

15 10. Молекула нуклеиновой кислоты по п. 9, в которой указанный эффекторный домен, включает домен расщепления, домен дезаминазы или домен модулятора экспрессии.

20 11. Молекула нуклеиновой кислоты по п. 9 или п. 10, в которой указанный эффекторный домен функционально слит на N-конце, C-конце или во внутренней локализации, указанного полипептида RGN.

25 12. Вектор, включающий молекулу нуклеиновой кислоты по любому из п.п. 1-11.

30 13. Вектор по п. 12, дополнительно включающий по меньшей мере одну нуклеотидную последовательность, кодирующую указанную направляющую РНК, способную гибридизироваться с указанной последовательностью ДНК-мишени, и в котором направляющая РНК содержит РНК CRISPR, включающую повторяющуюся CRISPR, включающую нуклеотидную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 62, 31, 76, 2, 10, 17, 24, 39, 47, 55, 70, 83, 90, 96, 104 или 111.

14. Вектор по п. 13, в котором указанная gРНК включает tracrРНК, включающую нуклеотидную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 63, 32, 77, 3, 11, 18, 25, 40, 48, 56, 71, 84, 91, 97, 105 или 112.

5

15. Клетка, содержащая молекулу нуклеиновой кислоты по любому из п.п. 1-11 или вектор по любому из п.п. 12-14.

10 16. Система для связывания последовательности ДНК-мишени молекулы ДНК, причем указанная система содержит:

а) одну или несколько направляющих РНК (gРНК), способных гибридизироваться с указанной последовательностью ДНК-мишени, или одним или несколькими полинуклеотидами, содержащими нуклеотидные последовательности, кодирующие одну или несколько направляющих gРНК; и

15 б) полипептид РНК-направляемой нуклеазы (RGN), содержащий аминокислотную последовательность, выбранную из группы, состоящей из:

i) аминокислотной последовательности, которая по меньшей мере на 95% идентична SEQ ID NO: 61, 30, 75, 1, 9, 16, 23, 38, 46, 69, 82, 89, 95, 103 или 110; и

20 ii) аминокислотной последовательности, которая по меньшей мере на 99% идентична SEQ ID NO: 54;

или полинуклеотид, включающего нуклеотидную последовательность, которая кодирует полипептид RGN.

25 17. Система по п. 16, в которой одна или несколько направляющих gРНК способны образовывать комплекс с полипептидом RGN, чтобы направить указанный полипептид RGN на связывание с указанной последовательностью ДНК-мишени молекулы ДНК.

30 18. Система по п. 16 или п. 17, в которой по меньшей мере одна из указанных нуклеотидных последовательностей, кодирующих одну или несколько направляющих gРНК и кодирующих полипептид RGN, функционально связаны с промотором, гетерологичным указанной нуклеотидной последовательности.

19. Система по любому из п.п. 16-18, в которой последовательность ДНК-мишени находится внутри клетки.

20. Система по любому из п.п. 16-19, в которой указанный полипептид RGN, включает аминокислотную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 61, 30, 75, 1, 9, 16, 23, 38, 46, 69, 82, 89, 95, 103 или 110.

21. Система по любому из п.п. 16-20, в которой указанный полипептид RGN, включает аминокислотную последовательность SEQ ID NO: 61, 30, 75, 1, 9, 16, 23, 38, 46, 69, 82, 89, 95, 103, 110 или 54.

22. Система по любому из п.п. 16-20, в которой указанный полипептид RGN является неактивной нуклеазой или способен функционировать в качестве никазы.

23. Система по любому из п.п. 16-22, в которой полипептид RGN функционально слит с полипептидом с отредактированными основаниями.

24. Система по п. 23, в которой указанный полипептид с отредактированными основаниями представляет собой дезаминазу.

25. Система по любому из п.п. 16-22, в которой указанный полипептид RGN функционально слит с эффекторным доменом.

26. Система по п. 25, в которой указанный эффекторный домен, включает домен расщепления, домен дезаминазы или домен модулятора экспрессии.

27. Система по п. 25 или п. 26, в которой указанный эффекторный домен функционально слит на N-конце, C-конце или во внутренней локализации, указанного полипептида RGN.

28. Система по любому из п.п. 16-27, в которой указанная система дополнительно содержит один или несколько донорных полинуклеотидов или

одну или несколько нуклеотидных последовательностей, кодирующих один или несколько донорных полинуклеотидов.

29. Способ связывания последовательности ДНК-мишени молекулы ДНК, включающий доставку системы по любому из п.п. 16-28 к указанной последовательности ДНК-мишени или клетке, содержащей последовательность ДНК-мишени.

30. Способ расщепления или модификации последовательности ДНК-мишени молекулы ДНК, включающий доставку системы по любому из п.п. 16-28 в указанную последовательность ДНК-мишени или клетку, содержащую молекулу ДНК, и происходит расщепление или модификация указанной последовательности ДНК-мишени.

31. Способ связывания последовательности ДНК-мишени молекулы ДНК, включающий:

а) сборку *in vitro* рибонуклеотидного комплекса РНК-направляющей нуклеазы (RGN) путем комбинирования:

і) одной или нескольких направляющих РНК, способных гибридизироваться с последовательностью ДНК-мишени; и

іі) полипептида RGN, содержащего аминокислотную последовательность, выбранную из группы, состоящей из:

А) аминокислотной последовательности, которая по меньшей мере на 90% идентична SEQ ID NO: 61, 30, 75, 1, 9, 16, 23, 38, 46, 69, 82, 89, 95, 103 или 110; и

Б) аминокислотной последовательности, которая по меньшей мере на 99% идентична SEQ ID NO: 54;

в условиях, пригодных для образования рибонуклеотидного комплекса RGN; и

б) контактирование указанной последовательности ДНК-мишени или клетки, содержащей указанную последовательность ДНК-мишени, с собранным *in vitro* рибонуклеотидным комплексом RGN;

причем одна или несколько направляющих РНК гибридизируются с последовательностью ДНК-мишени, тем самым направляя указанный полипептид RGN на связывание с указанной последовательностью ДНК-мишени.

5 32. Способ расщепления и/или модификации последовательности ДНК-мишени молекулы ДНК, включающий контактирование молекулы ДНК с:

а) полипептидом РНК-направляемой нуклеазы (RGN), где указанная RGN содержит аминокислотную последовательность, выбранную из группы, состоящей из:

10 i) аминокислотной последовательности, которая по меньшей мере на 90% идентична SEQ ID NO: 61, 30, 75, 1, 9, 16, 23, 38, 46, 69, 82, 89, 95, 103 или 110; и

ii) аминокислотной последовательности, которая по меньшей мере на 99% идентична SEQ ID NO: 54;

15 б) одной или несколькими направляющими РНК, способными нацеливать RGN по подпункту (а) на последовательность ДНК-мишени;

причем одна или несколько направляющих РНК гибридизируются с последовательностью ДНК-мишени, тем самым направляя указанный полипептид RGN на связывание с указанной последовательностью ДНК-мишени и вызывая расщепление и/или модификацию указанной последовательности ДНК-мишени.

20 33. Способ по п. 32, в котором указанная модифицированная последовательность ДНК-мишени включает делецию или мутацию по меньшей мере одного нуклеотида в последовательности ДНК-мишени.

30 34. Способ по п. 32 или п. 33, в котором указанный полипептид RGN, включает аминокислотную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 61, 30, 75, 1, 9, 16, 23, 38, 46, 69, 82, 89, 95, 103 или 110.

35. Способ по любому из п.п. 32-34, в котором указанный полипептид RGN, включает аминокислотную последовательность SEQ ID NO: 61, 30, 75, 1, 9, 16, 23, 38, 46, 69, 82, 89, 95, 103, 110 или 54.

36. Способ по любому из п.п. 32-34, в котором указанный полипептид RGN является неактивной нуклеазой или способен функционировать в качестве
никазы.

5

37. Способ по любому из п.п. 32-36, в котором полипептид RGN функционально слит с полипептидом с отредактированными основаниями.

10 38. Способ по п. 32, в котором в котором указанная модифицированная последовательность ДНК-мишени включает инсерцию гетерологичной ДНК в последовательность ДНК-мишени.

39. Способ по любому из п.п. 32-38, в котором последовательность ДНК-мишени находится внутри клетки.

15

40. Способ по п. 39, дополнительно включающий культивирование клетки в условиях, в которых полипептид RGN экспрессируется и расщепляет последовательность ДНК-мишени с получением молекулы ДНК, содержащей модифицированную последовательность ДНК; и выбор клетки, содержащей
20 указанную модифицированную последовательность ДНК-мишени.

41. Клетка, содержащая модифицированную последовательность ДНК-мишени по способу по п. 40.

25 42. Полипептид РНК-направляемой нуклеазы (RGN), включающий аминокислотную последовательность, выбранную из группы, состоящей из:

а) полипептида RGN, включающего аминокислотную последовательность, которая по меньшей мере на 90% идентична SEQ ID NO: 61, 30, 75, 1, 9, 16, 23, 38, 46, 69, 82, 89, 95, 103 или 110; и

30 б) полипептида RGN, включающего аминокислотную последовательность, которая по меньшей мере на 99% идентична SEQ ID NO: 54.

43. Полипептид RGN по п. 42, в котором указанный полипептид RGN, включает аминокислотную последовательность, которая по меньшей мере на

95% идентична SEQ ID NO: 61, 30, 75, 1, 9, 16, 23, 38, 46, 69, 82, 89, 95, 103 или 110.

5 44. Полипептид RGN по п. 42 или п. 43, в котором указанный полипептид RGN, включает аминокислотную последовательность SEQ ID NO: 61, 30, 75, 1, 9, 16, 23, 38, 46, 69, 82, 89, 95, 103, 110 или 54.

10 45. Полипептид RGN по п. 42 или п. 43, в котором указанный полипептид RGN является неактивной нуклеазой или способен функционировать в качестве нуклеазы.

46. Полипептид RGN по любому из п.п. 42-45, в котором полипептид RGN функционально слит с полипептидом с отредактированными основаниями.

15 47. Полипептид RGN по п. 46, в котором указанный полипептид с отредактированными основаниями представляет собой дезаминазу.

48. Полипептид RGN по любому из п.п. 42-45, в котором указанный полипептид RGN функционально слит с эффекторным доменом.

20 49. Полипептид RGN по п. 48, в котором указанный эффекторный домен, включает домен расщепления, домен дезаминазы или домен модулятора экспрессии.

25 50. Полипептид RGN по п. 48 или п. 49, в котором указанный эффекторный домен функционально слит на N-конце, C-конце или во внутренней локализации, указанного полипептида RGN.

51. Слитый полипептид, включающий

30 А) полипептид РНК-направляемой нуклеазы (RGN), включающий аминокислотную последовательность, которая по меньшей мере на 90% идентична SEQ ID NO: 117, 54, 61, 30, 75, 1, 9, 16, 23, 38, 46, 69, 82, 89, 95, 103 или 110; и

Б) гетерологичный полипептид.

52. Слитый полипептид по п. 51, в котором указанный полипептид RGN, включает аминокислотную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 117, 54, 61, 30, 75, 1, 9, 16, 23, 38, 46, 69, 82, 89, 95, 103 или 110.
53. Слитый полипептид по п. 51 или п. 52, в котором указанный полипептид RGN, включает аминокислотную последовательность SEQ ID NO: 117, 54, 61, 30, 75, 1, 9, 16, 23, 38, 46, 69, 82, 89, 95, 103 или 110.
54. Слитый полипептид по п. 51 или п. 52, в котором указанный полипептид RGN является неактивной нуклеазой или способен функционировать в качестве нуклеазы.
55. Слитый полипептид по любому из п.п. 51-54, в котором гетерологичный полипептид представляет собой полипептидом с отредактированными основаниями.
56. Слитый полипептид по п. 55, в котором указанный полипептид с отредактированными основаниями представляет собой дезаминазу.
57. Слитый полипептид по любому из п.п. 51-54, в котором указанный гетерологичный полипептид представляет собой эффекторный домен.
58. Слитый полипептид по п. 57, в котором указанный эффекторный домен, включает домен расщепления, домен дезаминазы или домен модулятора экспрессии.
59. Слитый полипептид по п. 57 или п. 58, в котором указанный эффекторный домен функционально слит на N-конце, C-конце или во внутренней локализации, указанного полипептида RGN.
60. Молекула нуклеиновой кислоты, содержащая полинуклеотид, который кодирует CRISPR РНК (сгРНК), в которой указанная сгРНК включает

спейсерную последовательность и повторяющуюся последовательность CRISPR, причем указанная повторяющаяся последовательность CRISPR содержит нуклеотидную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 118, 31, 76, 2, 10, 17, 24, 39, 47, 55, 62, 70, 83, 90, 96, 104 или 111.

5

61. Молекула нуклеиновой кислоты по п. 60, в которой направляющая РНК содержит:

а) указанную sgРНК; и необязательно

б) транс-активирующую CRISPR РНК (tracrРНК), способную

10 гибридизироваться с указанной повторяющейся последовательностью CRISPR указанной sgРНК;

способная гибридизироваться с последовательностью ДНК-мишени молекулы ДНК специфичным для последовательности образом через спейсерную последовательность указанной sgРНК, когда указанная
15 направляющая РНК связана с полипептидом РНК-направляемой нуклеазы (RGN).

15

62. Молекула нуклеиновой кислоты по п. 60 или 61, в которой указанный полинуклеотид, кодирующий sgРНК, функционально связан с промотором,
20 гетерологичным указанному полинуклеотиду.

20

63. Вектор, включающий молекулу нуклеиновой кислоты любому из п.п. 60-62.

25

64. Вектор по п. 63, где указанный вектор дополнительно содержит полинуклеотид, кодирующий указанную tracrРНК, причем указанная tracrРНК содержит нуклеотидную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 119, 32, 77, 3, 11, 18, 25, 40, 48, 56, 63, 71, 84, 91, 97, 105 или 112.

30

65. Молекула нуклеиновой кислоты, содержащая полинуклеотид, кодирующий транс-активирующую РНК CRISPR (tracrРНК), включающую последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 119, 32, 77, 3, 11, 18, 25, 40, 48, 56, 63, 71, 84, 91, 97, 105 или 112.

66. Молекула нуклеиновой кислоты по п. 65, в которой направляющая РНК содержит:

а) указанную tracrРНК; и

5 б) crРНК, содержащую спейсерную последовательность и повторяющуюся последовательность CRISPR, где указанная tracrРНК способна гибридизироваться с указанной повторяющейся последовательностью CRISPR указанной crРНК;

10 способная гибридизоваться с последовательностью ДНК-мишени молекулы ДНК специфичным для последовательности образом через спейсерную последовательность указанной crРНК, где указанная направляющая РНК связана с полипептидом РНК-направляемой нуклеазы (RGN).

15 67. Молекула нуклеиновой кислоты по п. 65 или 66, в которой указанный полинуклеотид, кодирующий tracrРНК, функционально связан с промотором, гетерологичным указанному полинуклеотиду.

20 68. Вектор, включающий молекулу нуклеиновой кислоты любому из п.п. 65-67.

25 69. Вектор по п. 68, где указанный вектор дополнительно включает полинуклеотид, кодирующий указанную crРНК, и где повторяющаяся последовательность CRISPR указанной crРНК содержит нуклеотидную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 118, 31, 76, 2, 10, 17, 24, 39, 47, 55, 62, 70, 83, 90, 96, 104 или 111.

70. Способ обнаружения последовательности ДНК-мишени молекулы ДНК в образце, включающий:

а) контактирование образца с:

30 i) полипептидом РНК-направляющей нуклеазы (RGN), включающим аминокислотную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 54 или 137, причем указанный полипептид RGN способен связывать указанную последовательность ДНК-мишени молекулы ДНК специфичным для

РНК-направляющей последовательности образом, при связывании с направляющей РНК, способной гибридизироваться с указанной последовательностью ДНК-мишени;

ii) указанной направляющей РНК; и

5 iii) детекторной одноцепочечной ДНК (оцДНК), не гибридизирующейся с направляющей РНК; и

б) измерение детектируемого сигнала, вырабатываемого расщеплением детекторной оцДНК под действием RGN, тем самым обнаруживая ДНК-мишень.

10 71. Набор для обнаружения последовательности ДНК-мишени в молекуле ДНК в образце, причем указанный набор включает:

а) полипептид РНК-направляемой нуклеазы (RGN), включающий аминокислотную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 54 или 137, причем указанный полипептид RGN способен связывать указанную последовательность ДНК-мишени молекулы ДНК специфичным для РНК-направляющей последовательности образом, при связывании с направляющей РНК, способной гибридизироваться с указанной последовательностью ДНК-мишени;

б) указанную направляющую РНК; и

20 в) детекторную одноцепочечную ДНК (оцДНК), не гибридизирующуюся с направляющей РНК.

72. Способ по п. 50 или набор по п. 71, в котором указанная детекторная оцДНК содержит пару флуорофор/гаситель или пару резонансного переноса энергии флуоресценции (FRET).

73. Способ расщепления одноцепочечных ДНК, включающий контактирование популяции нуклеиновых кислот, причем указанная популяция содержит молекулу ДНК, включающую последовательность ДНК-мишени и множество нецелевых оцДНК:

30 а) с полипептидом РНК-направляемой нуклеазы (RGN), включающим аминокислотную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 54 или 137, причем указанный полипептид RGN способен связывать указанную последовательность ДНК-мишени специфичным для РНК-

направляющей последовательности образом при связывании с направляющей РНК, способной гибридизироваться с указанной последовательностью ДНК-мишени; и

б) с указанной направляющей РНК;

5 где полипептид RGN расщепляет нецелевую оцДНК указанного множества.

74. Молекула нуклеиновой кислоты, содержащая полинуклеотид, кодирующий РНК CRISPR (сгРНК), причем указанная сгРНК содержит спейсерную последовательность и повторяющуюся последовательность CRISPR, где указанная повторяющаяся последовательность CRISPR содержит нуклеотидную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 240;

где направляющая РНК включает:

а) указанную сгРНК; и необязательно

15 б) трансактивирующую РНК CRISPR (trасгРНК), способную гибридизироваться с указанной повторяющейся последовательностью CRISPR указанной сгРНК;

способная гибридизоваться с последовательностью ДНК-мишени молекулы ДНК специфичным для последовательности образом через спейсерную последовательность указанной сгРНК, когда указанная направляющая РНК связана с полипептидом РНК-направляемой нуклеазы (RGN), и

20 где указанный полинуклеотид, кодирующий сгРНК, функционально связан с промотором, гетерологичным указанному полинуклеотиду.

25 75. Молекула нуклеиновой кислоты, содержащая полинуклеотид, кодирующий транс-активирующую РНК CRISPR (trасгРНК), содержащую нуклеотидную последовательность, которая по меньшей мере на 95% идентична SEQ ID NO: 241;

причем направляющая РНК содержит:

30 а) указанную trасгРНК; и

б) сгРНК, содержащую спейсерную последовательность и повторяющуюся последовательность CRISPR, где указанная trасгРНК способна гибридизироваться с указанной повторяющейся последовательностью CRISPR указанной сгРНК;

способная гибридизироваться с последовательностью ДНК-мишени молекулы ДНК специфичным для последовательности образом через спейсерную последовательность указанной сгРНК, когда указанная направляющая РНК связана с полипептидом РНК-направляемой нуклеазы (RGN), и

где указанный полинуклеотид, кодирующий tracrРНК, функционально связан с промотором, гетерологичным указанному полинуклеотиду.