

(19)



**Евразийское
патентное
ведомство**

(11) **046542**(13) **B1**(12) **ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ**

(45) Дата публикации и выдачи патента
2024.03.26

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

(21) Номер заявки
202092929

(22) Дата подачи заявки
2019.06.04

(54) **РНК-УПРАВЛЯЕМЫЕ НУКЛЕАЗЫ И АКТИВНЫЕ ФРАГМЕНТЫ, И ИХ ВАРИАНТЫ И СПОСОБЫ ПРИМЕНЕНИЯ**

(31) **62/680,845; 62/680,846; 62/680,853;
62/680,859; 62/680,862; 62/680,863;
62/686,901; 62/805,041; 62/805,045**

(32) **2018.06.05; 2018.06.05; 2018.06.05;
2018.06.05; 2018.06.05; 2018.06.05;
2018.06.19; 2019.02.13; 2019.02.13**

(33) **US**

(43) **2021.05.28**

(86) **PCT/US2019/035373**

(87) **WO 2019/236566 2019.12.12**

(71)(73) Заявитель и патентовладелец:
**ЛАЙФЭДИТ ТЕРАПЬЮТИКС, ИНК.
(US)**

(72) Изобретатель:
**Боуэн Тайсон Д., Элич Тедд Д., Кроли
Александра Брайнер, Баррангу
Родольф, 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)**

(56) **WO-A1-2017155714
WO-A1-2013176772
WO-A1-2017155717
WO-A1-2016186946
WO-A2-2015071474
WO-A1-2017222773**

**FUGUO JIANG ET AL: "CRISPR-Cas9
Structures and Mechanisms", ANNUAL REVIEW
OF BIOPHYSICS, vol. 46, no. 1, 30 March
2017 (2017-03-30), XP055362997, ISSN: 1936-122X,
DOI: 10.1146/annurev-biophys-062215-010822 see
abstract and pages 508-516**

**PAUL BG VAN ERP ET AL: "The history and
market impact of CRISPR RNA-guided nucleases",
CURRENT OPINION IN VIROLOGY, vol. 12, 1
June 2015 (2015-06-01), pages 85-90, XP055402267,
United Kingdom ISSN: 1879-6257, DOI: 10.1016/
j.coviro.2015.03.011 the whole document**

(57) Предлагаются композиции и способы связывания с представляющей интерес целевой последовательностью. Композиции находят применение при расщеплении или модификации представляющей интерес целевой последовательности, визуализации представляющей интерес целевой последовательности и модификации экспрессии представляющей интерес последовательности. Композиции включают полипептиды нуклеаз, управляемых РНК, CRISPR РНК, трансактивирующие CRISPR РНК, управляющие РНК и молекулы нуклеиновых кислот, кодирующие их. Также предлагаются векторы и клетки-хозяева, содержащие молекулы нуклеиновых кислот. Кроме того, предусмотрены системы CRISPR для связывания представляющей интерес целевой последовательности, при этом система CRISPR включает полипептид нуклеазы, управляемой РНК, и одну или несколько управляющих РНК.

B1**046542****046542****B1**

Область техники, к которой относится изобретение

Настоящее изобретение относится к области молекулярной биологии и редактирования генов.

Предпосылки создания изобретения

Целевое редактирование или модификация генома быстро становится важным инструментом фундаментальных и прикладных исследований. Первоначальные методы включали конструирование нуклеаз, таких как мегануклеазы, слитые белки с цинковыми пальцами или TALEN, что требовало создания химерных нуклеаз с сконструированными, программируемыми, специфичными для последовательности ДНК-связывающими доменами, специфичными для каждой конкретной целевой последовательности. РНК-управляемые нуклеазы, такие как Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-ассоциированные (cas) белки бактериальной системы CRISPR-cas, позволяют нацеливаться на определенные последовательности путем образования комплекса нуклеаз с управляющей РНК, которая специфически гибридизуется с конкретной последовательностью-мишенью. Производство специфичных для мишени управляющих РНК менее затратно и более эффективно, чем создание химерных нуклеаз для каждой целевой последовательности. Такие РНК-управляемые нуклеазы можно использовать для редактирования геномов путем введения специфичного для последовательности двух-цепочечного разрыва, который восстанавливается с помощью подверженного ошибкам негомологичного соединения концов (NHEJ), чтобы ввести мутацию в конкретное место генома. Кроме того, гетерологичная ДНК может быть введена в геномный сайт посредством гомологически направленной репарации.

Краткое изложение сущности изобретения

Предлагаются композиции и способы связывания представляющей интерес целевой последовательности. Такие композиции находят применение при расщеплении или модификации представляющей интерес целевой последовательности, для обнаружения представляющей интерес целевой последовательности и модификации экспрессии представляющей интерес последовательности. Композиции включают полипептиды РНК-управляемой нуклеазы (RGN), CRISPR-РНК (сгРНК), трансактивирующие CRISPR-РНК (tracrРНК), управляющие РНК (gРНК), молекулы нуклеиновых кислот, кодирующие их, а также векторы и клетки-хозяева, содержащие молекулы нуклеиновых кислот. Также предусмотрены системы CRISPR для связывания представляющей интерес целевой последовательности, при этом система CRISPR включает РНК-управляемый полипептид нуклеазы и одну или несколько управляющих РНК. Таким образом, предложенные в настоящем изобретении способы предназначены для связывания представляющей интерес целевой последовательности и, в некоторых вариантах осуществления изобретения, для расщепления или модификации представляющей интерес последовательности-мишени. Представляющая интерес целевая последовательность может быть модифицирована, например, в результате негомологичного соединения концов или гомологически направленной репарации с введенной донорской последовательностью.

Подробное описание

Специалисты в данной области, к которой относятся эти изобретения, имея возможность ознакомиться с приведенными ниже в данном документе описанием и соответствующими рисунками, сможет представить много модификаций и другие варианты осуществления изобретений, описанных в данном изобретении. Следовательно, следует понимать, что осуществления изобретения не должны ограничиваться конкретными представленными вариантами осуществления изобретения и, что модификации и другие варианты осуществления изобретения предназначены для включения в объем прилагаемых вариантов осуществления изобретения. Хотя здесь используются специфические термины, они используются только в общем и описательном смысле, а не в целях ограничения.

I. Обзор.

РНК-управляемые нуклеазы (RGN) позволяют целенаправленно манипулировать одним сайтом в геноме и полезны в контексте применения генного нацеливания для терапевтических и исследовательских целей. У множества организмов, включая млекопитающих, РНК-управляемые нуклеазы использовались для геномной инженерии, например, путем стимуляции негомологичного соединения концов и гомологичной рекомбинации. Композиции и способы, описанные в данном документе, применимы для создания одно- или двух-цепочечных разрывов в полинуклеотидах, модификации полинуклеотидов, обнаружения конкретного сайта в полинуклеотиде или модификации экспрессии конкретного гена.

Описанные здесь РНК-управляемые нуклеазы могут изменять экспрессию гена путем модификации целевой последовательности. В конкретных вариантах осуществления изобретения РНК-управляемые нуклеазы направляются в целевую последовательность с помощью управляющей РНК (gРНК), как части кластерной системы РНК-управляемых нуклеаз с регулярными интервалами коротких палиндромных повторов (CRISPR). Управляющие РНК образуют комплекс с РНК-управляемыми нуклеазами, чтобы направлять РНК-управляемую нуклеазу на связывание с целевой последовательностью и в некоторых вариантах осуществления изобретения на то чтобы вносить одно- или двух-цепочечный разрыв в целевую последовательность. После того, как целевая последовательность была расщеплена, разрыв может быть восстановлен таким образом, что последовательность ДНК целевой последовательности будет изменена в процессе восстановления. Таким образом, в настоящем документе представлены способы использования РНК-управляемых нуклеаз для модификации целевой последовательности в ДНК клеток-

хозяев. Например, РНК-управляемые нуклеазы могут использоваться для модификации целевой последовательности в геномном локусе эукариотических или прокариотических клеток.

II. РНК-управляемые нуклеазы.

В предлагаемом изобретении представлены РНК-управляемые нуклеазы. Термин "РНК-управляемая нуклеаза" (RGN) относится к полипептиду, который связывает конкретную целевую нуклеотидную последовательность специфическим для последовательности образом, и направляется к целевой нуклеотидной последовательности с помощью молекулы управляющей РНК, которая образует комплекс с полипептидом и гибридизуется с целевой последовательностью. Хотя РНК-управляемая нуклеаза способна расщеплять целевую последовательность при связывании, термин "РНК-управляемая нуклеаза" также включает РНК-управляемые нуклеазы с неактивной нуклеазой, которые способны связываться с целевой последовательностью, но не расщеплять ее. Расщепление целевой последовательности РНК-управляемой нуклеазой может привести к одно- или двух-цепочечному разрыву. РНК-управляемые нуклеазы, способные расщеплять только одну цепь двух-цепочечной молекулы нуклеиновой кислоты, называются здесь никазами.

Представленные в предлагаемом изобретении РНК-управляемые нуклеазы включают APG05083.1, APG07433.1, APG07513.1, APG08290.1, APG05459.1, APG04583.1 и APG1688.1, РНК-управляемые нуклеазы, аминокислотные последовательности которых представлены далее в SEQ ID NO: 1, 11, 19, 27, 36, 45 или 54, соответственно, и их активные фрагменты или варианты, которые сохраняют способность связываться с целевой нуклеотидной последовательностью РНК-управляемым специфическим для последовательности образом. В некоторых вариантах осуществления изобретения активный вариант RGN APG05083.1, APG07433.1, APG07513.1, APG08290.1, APG05459.1, APG04583.1 и APG1688.1 способен расщеплять одно- или двух-цепочечную целевую последовательность. В некоторых вариантах осуществления изобретения активный фрагмент или вариант RGN APG05083.1, APG07433.1, APG07513.1, APG08290.1, APG05459.1, APG04583.1 или APG1688.1 включает аминокислотную последовательность, 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 идентичную последовательности с аминокислотной последовательностью, представленной в SEQ ID NO: 1, 11, 19, 27, 36, 45 или 54. В некоторых вариантах осуществления изобретения активный фрагмент APG05083.1, APG07433.1, APG07513.1, APG08290.1, APG05459.1, APG04583.1 или APG1688.1 RGN состоит по меньшей мере из 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050 или более смежных аминокислотных остатков аминокислотной последовательности, представленной в SEQ ID NO: 1, 11, 19, 27, 36, 45 или 54. РНК-управляемые нуклеазы, представленные в предлагаемом изобретении, могут содержать по меньшей мере один домен нуклеазы (например, ДНКазы, домен РНКазы) и по меньшей мере один распознавания РНК и/или РНК-связывающий домен для взаимодействия с управляющими РНК. Дополнительные домены, которые могут быть обнаружены в РНК-управляемых нуклеазах, представленных в настоящем документе, включают ДНК-связывающие домены, домены геликазы, домены межбелкового взаимодействия и домены димеризации, но не ограничиваются ими. В определенных вариантах осуществления изобретения нуклеазы, управляемые РНК, представленные в настоящем документе, могут состоять по меньшей мере на 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% из одного или нескольких из ДНК-связывающих доменов, доменов геликазы, доменов межбелкового взаимодействия и доменов димеризации.

Целевая нуклеотидная последовательность связывается с РНК-управляемой нуклеазой, представленной в предлагаемом изобретении, и гибридизуется с управляющей РНК, связанной с РНК-управляемой нуклеазой. Затем целевая последовательность может быть впоследствии расщеплена РНК-управляемой нуклеазой, если полипептид обладает нуклеазной активностью. Термины "расщеплять" или "расщепление" относятся к гидролизу по меньшей мере одной фосфодиэфирной связи в основной цепи целевой нуклеотидной последовательности, что может приводить к одно-цепочечным или двух-цепочечным разрывам в целевой последовательности. Описанные здесь 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 изобретения представленные RGN могут расщеплять нуклеотиды целевой последовательности в любом положении полинуклеотида и, таким образом, функционировать как эндонуклеаза, так и как экзонуклеаза. Расщепление целевого полинуклеотида описанной в данном изобретении RGN, может приводить к ступенчатым разрывам или тупым концам.

Описанные в данном изобретении РНК-управляемые нуклеазы могут быть последовательностями дикого типа, происходящими от видов бактерий или архей. Кроме того, РНК-управляемые нуклеазы могут быть вариантами или фрагментами полипептидов дикого типа. RGN дикого типа можно модифицировать, например, для изменения нуклеазной активности или специфичности РАМ (мотив, примыкающий к протоспейсеру). В некоторых вариантах осуществления изобретения РНК-управляемые нуклеазы не встречаются в природе.

В некоторых вариантах осуществления изобретения РНК-управляемая нуклеаза действует как никаза, расщепляя только одну цепь целевой нуклеотидной последовательности. Такие РНК-управляемые

нуклеазы имеют один функционирующий нуклеазный домен. В некоторых вариантах осуществления изобретения были изменены дополнительные нуклеазные домены, так что нуклеазная активность была снижена или устранена.

В других вариантах осуществления изобретения РНК-управляемая нуклеаза полностью лишена нуклеазной активности или демонстрирует пониженную нуклеазную активность и упоминается в данном описании изобретения как "неактивная нуклеаза". Любой способ известный в данной области для введения мутаций в аминокислотную последовательность, такой как PCR-опосредованный мутагенез и сайт-направленный мутагенез, может быть использован для создания нуклеазы или RGN с неактивной нуклеазой. См. например, U.S. Publ. № 2014/0068797 и патент США № 9,790,490, каждый из которых полностью включен в описание заявленного изобретения в качестве ссылки.

РНК-управляемые нуклеазы, лишённые нуклеазной активности, можно использовать для доставки слитого полипептида, полинуклеотида или небольшой молекулы в определенное место генома. В некоторых из вариантов осуществления изобретения полипептид RGN или управляющая РНК могут быть слиты с детектируемой меткой, чтобы сделать возможным обнаружение конкретной последовательности. В качестве неограничивающего примера RGN с неактивной нуклеазой может быть слита с детектируемой меткой (например, флуоресцентным белком) и нацелена на конкретную последовательность, ассоциированную с заболеванием, чтобы сделать возможным обнаружение ассоциированной с заболеванием последовательности.

Кроме того, RGN с неактивной нуклеазой могут быть нацелены на определенные участки генома, чтобы изменить экспрессию желаемой последовательности. В некоторых вариантах осуществления изобретения связывание РНК-управляемой не активной нуклеазы с целевой последовательностью приводит к подавлению экспрессии целевой последовательности или гена под транскрипционным контролем целевой последовательности путем вмешательства в связывание РНК-полимеразы или факторов транскрипции в целевой области генома. В других вариантах осуществления изобретения RGN (например, RGN с неактивной нуклеазой) или ее комплексная управляющая РНК дополнительно содержит модулятор экспрессии, который после связывания с целевой последовательностью служит либо для подавления, либо для активации экспрессии целевой последовательности или гена под транскрипционным контролем с помощью целевой последовательности. В некоторых вариантах осуществления изобретения модулятор экспрессии модулирует экспрессию целевой последовательности или регулируемого гена посредством эпигенетических механизмов.

В других вариантах осуществления изобретения RGN с неактивной нуклеазой или RGN, обладающие только никазной активностью, могут быть нацелены на определенные участки генома для модификации последовательности целевого полинуклеотида посредством слияния с полипептидом, редактирующим основания, например, с полипептидом дезаминазы или активным вариантом или фрагментом, которые дезаминируют нуклеотидное основание, что приводит к превращению одного нуклеотидного основания в другое. Полипептид, редактирующий основания, может быть слит с RGN на ее N-конце или C-конце. Кроме того, редактирующий основания полипептид может быть слит с RGN через пептидный линкер. Неограничивающий пример полипептида дезаминазы, который можно использовать для таких композиций и способов, включает цитидиндезаминазу или редактор оснований - аденозиндезаминазу, описанный у Gaudelli и др. (2017) Nature № 551: 464-471, U.S. Publ. №№ 2017/0121693 и 2018/0073012 и InteRNKtional Publ. № WO/2018/027078, каждый из этих источников полностью включен в описание предлагаемого изобретения в качестве ссылки.

РНК-управляемые нуклеазы, слитые с полипептидом или доменом, могут быть разделены или соединены линкером. Используемый здесь термин "линкер" относится к химической группе или молекуле, связывающей две молекулы или части, например, связывающий домен и домен расщепления нуклеазы. В некоторых вариантах осуществления изобретения линкер соединяет домен связывания gРНК РНК управляемой нуклеазы и редактирующий основания полипептид, такой как дезаминаза. В некоторых вариантах осуществления изобретения линкер соединяет RGN с неактивной нуклеазой и дезаминазой. Обычно линкер располагается между двумя группами, молекулами или другими фрагментами, или фланкируется ими и соединяется с каждой из них ковалентной связью, соединяя их таким образом. В некоторых вариантах осуществления изобретения линкер представляет собой аминокислоту или множество аминокислот (например, пептид или белок). В некоторых вариантах осуществления изобретения линкер представляет собой органическую молекулу, группу, полимер или химический фрагмент. В некоторых вариантах осуществления изобретения линкер имеет длину 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. Также рассматриваются более длинные или более короткие линкеры.

Описанные в данном изобретении РНК-управляемые нуклеазы могут содержать по меньшей мере один сигнал ядерной локализации (NLS) для усиления доставки RGN к ядру клетки. Сигналы ядерной локализации известны в данной области и обычно содержат ряд основных аминокислот (см., например, Lange и др., J. Biol. Chem. (2007) № 282: сс. 5101-5105). В определенных вариантах осуществления изобретения RGN содержит 2, 3, 4, 5, 6 или более сигналов ядерной локализации. Сигнал ядерной локализации

ции может быть гетерологичным NLS. Неограничивающими примерами сигналов ядерной локализации, пригодных для приведенных здесь RGN, являются сигналы ядерной локализации большого Т-антигена SV40, нуклеоплазмина и с-Мус (см., например, Ray и др. (2015) *Bioconjug Chem* № 26 (6): сс.1004-7). В определенных вариантах осуществления изобретения RGN содержит последовательность NLS, представленную в SEQ ID NO: 67. RGN может содержать одну или несколько последовательностей NLS на своем 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.

Другие сигнальные последовательности локализации, известные в данной области, которые сосредотачивают полипептиды в конкретном субклеточном местоположении(ях), также могут быть использованы для нацеливания RGN, включая, помимо прочего, последовательности пластидной локализации, последовательности митохондриальной локализации последовательности сигналов двойного нацеливания, которые нацелены как на пластиду, так и на митохондрии (см., например, Nassoury and Morse (2005) *Biochim Biophys Acta* № 1743: сс. 5-19; Kunze and Berger (2015) *Front Physiol* dx.doi.org/10.3389/fphys.2015.00259; Herrmann and Neupert (2003) *IUBMB Life* № 55: сс. 219-225; Soil (2002) *Curr Opin Plant Biol* № 5: сс. 529-535; Carrie and Small (2013) *Biochim Biophys Acta* № 1833: сс. 253-259; Carrie и др. (2009) *FEBS J* № 276: сс. 1187-1195; Silva-Filho (2003) *Curr Opin Plant Biol* № 6: сс. 589-595; Peeters and Small (2001) *Biochim Biophys Acta* № 1541: сс. 54-63; Murcha и др. (2014) *J Exp Bot* № 65: сс. 6301-6335; Mackenzie (2005) *Trends Cell Biol* № 15: сс. 548-554; Glaser и др. (1998) *Plant Mol Biol* № 38: сс. 311-338).

В определенных вариантах осуществления изобретения описанные РНК-управляемые нуклеазы содержат по меньшей мере один проникающий в клетку домен, который способствует клеточному поглощению RGN. Проникающие в клетки домены известны в данной области и обычно включают участки положительно заряженных аминокислотных остатков (т.е. поликатионные проникающие в клетки домены), чередующиеся полярные аминокислотные остатки и неполярные аминокислотные остатки (т.е. амфипатические проникающие в клетки домены) или гидрофобные аминокислотные остатки (т.е. гидрофобные проникающие в клетки домены) (см., например, Milletti F. (2012) *Drug Discov Today* № 17: сс. 850-860). Неограничивающим примером домена, проникающего в клетки, является трансактивирующий активатор транскрипции (ТАТ) из вируса иммунодефицита человека 1.

Сигнал ядерной локализации, сигнал пластидной локализации, сигнал митохондриальной локализации, сигнал двойного нацеливания и/или проникающий в клетку домен могут располагаться на аминоконце (N-конце), карбоксильном конце (С-конце) или во внутреннем положении РНК-управляемой нуклеазы.

Представленные в данном изобретении 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 располагаться на N-конце, С-конце или во внутреннем местоположении нуклеазы, управляемой РНК. В некоторых вариантах осуществления изобретения компонент RGN слитого белка представляет собой RGN с неактивной нуклеазой.

В некоторых вариантах осуществления изобретения слитый белок RGN содержит домен расщепления, который представляет собой любой домен, который способен расщеплять полинуклеотид (т.е. РНК, ДНК или гибрид РНК/ДНК), и включает эндонуклеазы рестрикции и самоуправляемые эндонуклеазы, такие как эндонуклеазы типа IIS (например, FokI), но не ограничиваются ими (см., например, Belfort и др. (1997) *Nucleic Acids Res.* № 25: сс. 3379-3388; Linn и др. (под редакцией) *Nucacess*, Издательство Cold Spring Harbor Laboratory Press, 1993).

В других вариантах осуществления изобретения слитый белок RGN содержит домен дезаминазы, который дезаминирует нуклеотидное основание, что приводит к превращению одного нуклеотидного основания в другое, и включает цитидиндезаминазу или редактор оснований аденозиндезаминазы, но не ограничивается ими (см., например, Gaudelli и др. (2017) *Nature* № 551: сс. 464-471, U.S. Publ. Nos. 2017/0121693 и 2018/0073012, патент США № 9840699 и *IntePHKtional Publ. No. WO/2018/027078*).

В других вариантах осуществления изобретения эффекторный домен слитого белка RGN может быть доменом модулятора экспрессии, который является доменом, служащим либо для усиления, либо для подавления транскрипции. Домен модулятора экспрессии может быть доменом эпигенетической модификации, доменом-репрессором транскрипции или доменом активации транскрипции.

В некоторых вариантах осуществления изобретения модулятор экспрессии слитого белка RGN содержит домен эпигенетической модификации, который ковалентно модифицирует ДНК или гистоновые белки для изменения структуры гистонов и/или хромосомной структуры без изменения последовательности ДНК, что приводит к изменениям в экспрессии генов (т.е. повышающая или понижающая регуляция). Неограничивающие примеры эпигенетических модификаций включают ацетилирование или метилирование остатков лизина, метилирование аргинина, фосфорилирование серина и треонина, и убиквитинирование лизина и сумоилирование гистоновых белков, а также метилирование и гидроксиметилирование остатков цитозина в ДНК. Неограничивающие примеры доменов эпигенетической модификации включают домены гистонацетилтрансферазы, домены гистондеацетилазы, домены гистонметилтрансферазы, домены гистондеметилазы, домены ДНК-метилтрансферазы и домены ДНК-

деметилазы.

В некоторых вариантах осуществления 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 уменьшения или прекращения транскрипции по меньшей мере одного гена. Домены-репрессоры транскрипции известны в данной области и включают Sp1-подобные репрессоры, IκB и домены-боксы, связанные с Krüppel (KRAB), но не ограничиваются ими.

В еще одних вариантах осуществления 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, помимо прочего, домен активации VP16 вируса простого герпеса и домен активации NFAT.

Описанные в настоящем изобретении полипептиды RGN могут содержать детектируемую метку или метку очистки. Детектируемая метка или метка очистки может быть присоединена на N-конце, C-конце или во внутреннем положении РНК-управляемой нуклеазы, прямо или с помощью линкерного пептида. В некоторых вариантах осуществления изобретения компонент RGN слитого белка представляет собой RGN с неактивной нуклеазой. В других вариантах осуществления изобретения компонент RGN слитого белка представляет собой RGN с никазной активностью.

Детектируемая метка - это молекула, которую можно визуализировать или наблюдать любым образом. Детектируемая метка может быть слита с RGN как слитый белок (например, флуоресцентный белок) или может представлять собой небольшую молекулу, конъюгированную с полипептидом RGN, которая может быть обнаружена визуальными или другими способами. Детектируемые метки, которые могут быть слиты с описанными в данном изобретении RGN в качестве слитого белка, включают любой обнаруживаемый белковый домен, включая флуоресцентный белок или белковый домен, который может быть обнаружен с помощью специфического антитела, но не ограничиваются ими. Неограничивающие примеры флуоресцентных белков включают зеленые флуоресцентные белки (например, GFP, EGFP, ZsGreen1) и желтые флуоресцентные белки (например, YFP, EYFP, ZsYellow1). Неограничивающие примеры детектируемых меток малых молекул включают радиоактивные метки, такие как ³H и ³⁵S.

Полипептиды RGN могут также содержать метку очистки, которая представляет собой любую молекулу, которая может использоваться для выделения белка или слитого белка из смеси (например, биологического образца, культуральной среды). Неограничивающие примеры меток очистки включают биотин, пус, белок, связывающий мальтозу (МБП) и глутатион-S-трансферазу (GST).

II. Управляющая РНК.

В настоящем изобретении представлены управляющие РНК и полинуклеотиды, кодирующие их. Термин "управляющая РНК" относится к нуклеотидной последовательности, имеющей достаточную комплементарность с целевой нуклеотидной последовательностью для гибридизации с целевой последовательностью и прямого специфичного для последовательности связывания ассоциированной РНК-управляемой нуклеазы с целевой нуклеотидной последовательностью. Таким образом, соответствующая управляющая РНК RGN представляет собой одну или несколько молекул РНК (обычно одна или две), которые могут связываться с RGN и управлять RGN для связывания с конкретной целевой нуклеотидной последовательностью, и в тех случаях, когда RGN содержит никазу или нуклеазную активность, также расщеплять целевую нуклеотидную последовательность. Обычно управляющая РНК включает CRISPR-РНК (сгРНК) и трансактивирующую CRISPR-РНК (tracrРНК). Нативные управляющие РНК, которые включают как сгРНК, так и tracrРНК, обычно содержат две отдельные молекулы РНК, которые гибридизуются друг с другом через повторяющуюся последовательность сгРНК и антиповторную последовательность tracrРНК.

Нативные последовательности прямых повторов в матрице CRISPR обычно имеют длину от 28 до 37 пар оснований, хотя длина может варьироваться примерно от 23 до 55 пар оснований. Последовательности спейсеров в матрице CRISPR обычно имеют длину примерно от 32 до 38 п.о., хотя длина может составлять примерно от 21 п.о. до 72 п.о. Каждая матрица CRISPR обычно содержит менее 50 единиц спейсерной последовательности CRISPR повторов. CRISPR транскрибируются как часть длинного транскрипта, называемого первичным транскриптом CRISPR, который составляет большую часть матрицы CRISPR. Первичный транскрипт CRISPR расщепляется белками Cas с образованием сгРНК или, в некоторых случаях, с образованием pre-сгРНК, которые далее процессируются дополнительными белками Cas в зрелые сгРНК. Зрелые сгРНК содержат спейсерную последовательность и повторяющуюся последовательность CRISPR. В некоторых вариантах осуществления изобретения, в которых pre-сгРНК процессируются в зрелые (или процессированные) сгРНК, созревание включает удаление примерно от одного до шести или более 5', 3' или 5' и 3' нуклеотидов. В целях редактирования генома или нацеливания на конкретную представляющую интерес нуклеотидную целевую последовательность нуклеотиды, которые удаляются во время созревания молекулы pre-сгРНК, не являются необходимыми для создания или конструирования управляющей РНК.

PHK CRISPR (сгPHK) содержит последовательность спейсера и последовательность CRISPR повторов. "Спейсерная последовательность" представляет собой нуклеотидную последовательность, которая непосредственно гибридизуется с представляющей интерес целевой нуклеотидной последовательностью. Спейсерная последовательность сконструирована так, чтобы быть полностью или частично комплементарной с представляющей интерес целевой последовательностью. В различных вариантах осуществления изобретения спейсерная последовательность может содержать примерно от 8 нуклеотидов до 30 нуклеотидов или более. Например, спейсерная последовательность может быть примерно 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 или более нуклеотидов в длину. В некоторых вариантах осуществления изобретения спейсерная последовательность составляет от 10 до 26 нуклеотидов в длину или от 12 до 30 нуклеотидов в длину. В конкретных вариантах осуществления изобретения спейсерная последовательность составляет около 30 нуклеотидов в длину. В некоторых вариантах осуществления изобретения степень комплементарности между спейсерной последовательностью и соответствующей ей целевой последовательностью при оптимальном выравнивании с использованием подходящего алгоритма выравнивания составляет примерно или более чем 50%, 60%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или более. В определенных вариантах осуществления изобретения спейсерная последовательность не имеет вторичной структуры, которую можно предсказать, используя любой подходящий алгоритм укладки полинуклеотидов, известный в данной области, включая, *mFold*, но, не ограничиваясь им (см., например, Zuker and Stiegler (1981) *Nucleic Acids Res.* № 9: сс. 133-148) и PHKfold (см., например, Gruber и др. (2008) *Cell* № 106 (1): сс.23-24).

Белки RGN могут иметь различную чувствительность к несовпадениям между спейсерной последовательностью в гPHK и ее целевой последовательностью, что влияет на эффективность расщепления. Как обсуждалось в примере 5, RGN APG05459.1 имеет необычную чувствительность к несоответствиям между последовательностью спейсера и целевой последовательностью, расширяясь по меньшей мере на 15 5'-концевых нуклеотидов сайта PAM. Таким образом, APG05459.1 может более точно (т.е. конкретно) нацеливаться на определенные последовательности с большей точностью, чем другие RGN, с меньшей чувствительностью к несоответствиям между последовательностью спейсера и целевой последовательностью.

Последовательность PHK CRISPR повторов состоит из нуклеотидной последовательности, которая включает область с достаточной комплементарностью для гибридизации с *tracrPHK*. В различных вариантах осуществления изобретения последовательность PHK CRISPR повторов может содержать примерно от 8 до примерно 30 нуклеотидов или более. Например, последовательность CRISPR повторов 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 некоторых вариантах осуществления изобретения последовательность CRISPR повторов составляет примерно 21 нуклеотид. В некоторых вариантах осуществления изобретения степень комплементарности между последовательностью CRISPR повторов и соответствующей ей последовательностью *tracrPHK* при оптимальном выравнивании с использованием подходящего алгоритма выравнивания составляет примерно или более 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 идентичности нуклеотидной последовательности, представленной в SEQ ID NOs: 2, 12, 20, 28, 37, 46, или 55. В некоторых вариантах осуществления изобретения активный фрагмент последовательности CRISPR повторов дикого типа содержит по меньшей мере 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, или 20 смежных нуклеотидов нуклеотидной последовательности, представленной в SEQ ID NO: 2, 12, 20, 28, 37, 46 или 55.

В некоторых вариантах осуществления изобретения сгPHK не встречается в природе. В некоторых вариантах осуществления изобретения конкретная последовательность CRISPR повторов не связана со сконструированной в природе спейсерной последовательностью и последовательность CRISPR повторов считается гетерологичной спейсерной последовательности. В некоторых вариантах осуществления изобретения спейсерная последовательность представляет собой сконструированную последовательность, которая не встречается в природе.

Трансактивирующая молекула PHK CRISPR или *tracrPHK* содержит нуклеотидную последовательность, имеющую область с достаточной комплементарностью для гибридизации с последовательностью CRISPR повторов сгPHK, которая упоминается здесь как антиповторная область. В некоторых вариантах осуществления изобретения молекула *tracrPHK* дополнительно содержит область с вторичной структурой (например, стебель-петля) или образует вторичную структуру при гибридизации с соответствующей сгPHK. В определенных вариантах осуществления изобретения область *tracrPHK*, которая полностью или частично комплементарна последовательности повторов CRISPR, находится на 5'-конце молекулы, а 3'-конец *tracrPHK* содержит вторичную структуру. Эта область вторичной структуры обычно состоит из

нескольких шпилечных структур, включая шпильку нексуса, которая находится рядом с антиповторной последовательностью. Шпилька нексуса часто имеет консервативную нуклеотидную последовательность в основании стержня шпильки с мотивом UNANNG, UNANNU или UNANNA (SEQ ID NO: 68, 557 и 558, соответственно), обнаруженным во многих шпильках нексуса в *tracrPHK*. Часто терминальные шпильки находятся на 3'-конце *tracrPHK*, которые могут различаться по структуре и количеству, но часто содержат GC-обогащенную Rho независимую транскрипционную шпильку терминатор, за которой следует цепочка из U на 3'-конце. См., например, Briner и др. (2014) *Molecular Cell* № 56:cc. 333-339, Briner and Barrangou (2016) *Cold Spring Harb Protoc*; doi: 10.1101/pdb.top090902 и U.S. Publication No. № 2017/0275648, которые полностью включены в настоящее изобретение в качестве ссылки.

В различных вариантах осуществления изобретения антиповторный участок *tracrPHK*, который полностью или частично комплементарен последовательности CRISPR повторов, содержит примерно от примерно 8 нуклеотидов до примерно 30 нуклеотидов или более. Например, область сопряжения оснований между антиповторной последовательностью *tracrPHK* и последовательностью CRISPR повторов 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 изобретения антиповторный участок *tracrPHK*, который полностью или частично комплементарен последовательности CRISPR повторов, имеет длину около 20 нуклеотидов. В некоторых вариантах осуществления изобретения степень комплементарности между повторяющейся последовательностью CRISPR и соответствующей ей антиповторной последовательностью *tracrPHK* при оптимальном выравнивании с использованием подходящего алгоритма выравнивания составляет примерно или более чем 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 изобретения целая *tracrPHK* может содержать от примерно 60 до примерно 140 нуклеотидов. Например, *tracrPHK* может быть 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 определенных вариантах осуществления изобретения *tracrPHK* имеет длину 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 некоторых вариантах осуществления изобретения *tracrPHK* имеет длину около 85 нуклеотидов.

В отдельных вариантах осуществления изобретения *tracrPHK* содержит нуклеотидную последовательность SEQ ID NO: 3, 13, 21, 29, 38, 47 или 56, или ее активный вариант или фрагмент, которая при включении в управляющую РНК способна управлять последовательно-специфическим связыванием ассоциированной РНК-управляемой нуклеазы, представленной в предлагаемом изобретении, с представляющей интерес целевой последовательностью. В некоторых вариантах осуществления изобретения активный вариант последовательности *tracrPHK* дикого типа содержит нуклеотидную последовательность 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 идентичную последовательности с нуклеотидной последовательностью, представленной в SEQ ID NO: 3, 13, 21, 29, 38, 47 или 56. В некоторых вариантах осуществления изобретения активный фрагмент последовательности *tracrPHK* дикого типа содержит по меньшей мере 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80 или более смежных нуклеотидов нуклеотидной последовательности, представленной в SEQ ID NO: 3, 13, 21, 29, 38, 47 или 56.

Две полинуклеотидные последовательности можно рассматривать по существу как комплементарные, если две последовательности гибридизуются друг с другом в строгих условиях. Аналогичным образом, считается, что RGN связывается с конкретной целевой последовательностью в рамках специфического для последовательности способа, если управляющая РНК, связанная с RGN, связывается с целевой последовательностью в строгих условиях. Под "строгими условиями" или "строгими условиями гибридизации" подразумеваются условия, при которых две полинуклеотидные последовательности будут гибридизоваться друг с другом в значительно большей степени, чем с другими последовательностями (например, по меньшей мере в 2 раза выше, чем фоновая). Строгие условия зависят от последовательности и будут разными в разных обстоятельствах. Как правило, строгими условиями будут те, в которых концентрация соли составляет примерно менее 1,5 М иона Na, обычно от 0,01 до 1,0 М концентрации иона Na (или других солей) при pH от 7,0 до 8,3, и температура 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). Строгие условия также могут быть достигнуты путем добавления дестабилизирующих агентов, таких как формамид. Примеры условий низкой строгости включают гибридизацию с буферным раствором, содержащим от 30 до 35% формамида, 1 М NaCl, 1% SDS (додецилсульфат натрия) при 37°C и отмывку 1-2×SSC (20×SSC=3,0 М NaCl/0,3 М тринатрийцитрат) при 50-

55°C. Примеры условий умеренной строгости включают гибридизацию в 40-45% формамиде, 1,0 M NaCl, 1% SDS при 37°C и отмывку в 0,5-1×SSC при 55-60°C. Примеры условий высокой строгости включают гибридизацию в 50% формамиде, 1 M NaCl, 1% SDS при 37°C и отмывку в 0,1×SSC при 60-65°C. Опционно промывочные буферы могут содержать примерно от 0,1% до 1% SDS. Продолжительность гибридизации обычно составляет менее 24 ч, обычно примерно от 4 до примерно 12 ч. Продолжительность отмывания должна составлять по крайней мере время, достаточное для достижения равновесия.

T_m - это температура (при определенной ионной силе и pH), при которой 50% комплементарной целевой последовательности гибридизуются с идеально подобранной последовательностью. Для гибридов ДНК-ДНК T_m может быть аппроксимирована уравнением Meinkoth and Wahl (1984) Anal. Biochem. № 138: сс. 267-284: $T_m = 81,5^\circ\text{C} + 16,6(\log M) + 0,41(\% \text{ GC}) - 0,61(\% \text{ форм.}) - 500/L$; где M - молярность одновалентных катионов, % GC - процентное содержание гуанозиновых и цитозиновых нуклеотидов в ДНК, % форм. - Процентное содержание формамида в растворе для гибридизации, а L - длина гибрида в парах оснований. Обычно строгие условия выбираются так, чтобы температура была примерно на 5°C ниже, чем температура плавления (T_m) для конкретной последовательности и ее дополнения при определенной ионной силе и pH. Однако в чрезвычайно строгих условиях можно использовать гибридизацию и/или отмывку при температуре на 1, 2, 3 или 4°C ниже, чем температура плавления (T_m); при умеренно строгих условиях можно использовать гибридизацию и/или отмывку при температуре на 6, 7, 8, 9 или 10°C ниже, чем температура плавления (T_m); в условиях низкой строгости можно использовать гибридизацию и/или отмывку при температуре на 11, 12, 13, 14, 15 или 20°C ниже, чем температура плавления (T_m). Используя уравнение, композиции для гибридизации и промывки, а также желаемую T_m , рядовые специалисты поймут, что, по сути, описаны вариации в строгости гибридизации и/или растворов промывки. Подробное руководство по гибридизации нуклеиновых кислот можно найти у Tijssen (1993) Laboratory Techniques in Biochemistry and Molecular Biology - Hybridization with Nucleic Acid Probes, Часть I, Глава 2 (Издательство Elsevier, Нью-Йорк); и Ausubel и др., ред. (1995) Current Protocols in Molecular Biology, Глава 2 (Издательство Greene Publishing and Wiley-Interscience, Нью-Йорк). См. Sambrook и др. (1989) Molecular Cloning: A Laboratory Manual (2-е изд., Издательство Cold Spring Harbor Laboratory Press, Plainview, Нью-Йорк).

Управляющая РНК может быть отдельной управляющей РНК или системой двойной управляющей РНК. Отдельная управляющая РНК содержит сгРНК и трасРНК на одной молекуле РНК, тогда как система двойной управляющей РНК включает сгРНК и трасРНК, присутствующие на двух разных молекулах РНК, гибридизованных друг с другом через по крайней мере часть последовательности CRISPR повторов сгРНК и по крайней мере часть трасРНК, которая может быть полностью или частично комплементарной последовательности CRISPR повторов сгРНК. В некоторых вариантах осуществления изобретения, где управляющая РНК представляет собой отдельную управляющую РНК, сгРНК и трасРНК разделены линкерной нуклеотидной последовательностью. В общем, линкерная нуклеотидная последовательность - это последовательность, которая не включает комплементарные основания, чтобы избежать образования вторичной структуры внутри или содержащей нуклеотиды линкерной нуклеотидной последовательности. В некоторых вариантах осуществления изобретения линкерная нуклеотидная последовательность между сгРНК и трасРНК составляет в длину 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 конкретных вариантах осуществления изобретения линкерная нуклеотидная последовательность одиночной управляющей РНК имеет длину, по меньшей мере, 4 нуклеотида. В некоторых вариантах осуществления изобретения линкерная нуклеотидная последовательность представляет собой нуклеотидную последовательность, представленную в SEQ ID NO: 63, 64 или 65. В других вариантах осуществления изобретения линкерная нуклеотидная последовательность имеет длину по меньшей мере 6 нуклеотидов. В некоторых вариантах осуществления изобретения линкерная нуклеотидная последовательность представляет собой нуклеотидную последовательность, представленную в SEQ ID NO: 65.

Отдельная управляющая РНК или двойная управляющая РНК могут быть синтезированы химическим путем или посредством транскрипции *in vitro*. Исследования определения специфического связывания последовательностей между RGN и управляющей РНК известны в данной области и включают, помимо прочего, анализы связывания *in vitro* между экспрессируемым RGN и управляющей РНК, которые могут быть помечены детектируемой меткой (например, биотином) и используются в анализе обнаружения преципитации (pull-down), в котором комплекс управляющая РНК: RGN захватывается с помощью детектируемой метки (например, с помощью гранул стрептавидина). Контрольная управляющая РНК с неродственной последовательностью или структурой управляющей РНК может использоваться в качестве отрицательного контроля для неспецифического связывания RGN с РНК. В некоторых вариантах осуществления изобретения управляющая РНК представляет собой SEQ ID NO: 10, 18, 26, 35, 44, 53 или 62, где спейсерная последовательность может быть любой последовательностью и обозначена, как поли-N последовательность.

Как описано в примере 8, некоторые RGN в настоящем изобретении могут совместно использовать

определенные управляющие РНК. Каждая APG05083.1, APG07433.1, APG07513.1 и APG08290.1 могут функционировать с использованием управляющих РНК, содержащих сгРНК, которая включает нуклеотидную последовательность SEQ ID NO: 2, 12, 20 или 28, с соответствующей тасгРНК, содержащей нуклеотидную последовательность SEQ ID NO: 3, 13, 21 или 29, соответственно. Кроме того, каждая из APG04583.1 и APG01688.1 может функционировать с использованием управляющих РНК, содержащих сгРНК, которая включает нуклеотидную последовательность SEQ ID NO: 46 или 55, с соответствующей тасгРНК, содержащей нуклеотидную последовательность SEQ 47 или 56, соответственно.

В некоторых вариантах осуществления изобретения управляющая РНК может быть введена в клетку-мишень, органеллу или эмбрион в виде молекулы РНК. Управляющая РНК может быть транскрибирована *in vitro* или синтезирована химическим путем. В других вариантах осуществления изобретения нуклеотидная последовательность, кодирующая управляющую РНК, вводится в клетку, органеллу или эмбрион. В некоторых из этих вариантов осуществления изобретения нуклеотидная последовательность, кодирующая управляющую РНК, функционально связана с промотором (например, промотором РНК-полимеразы III). Промотор может быть нативным или гетерологичным нуклеотидной последовательности, кодирующей управляющую РНК.

В различных вариантах осуществления изобретения управляющая РНК может быть введена в клетку-мишень, органеллу или эмбрион в виде рибонуклеопротеинового комплекса, как описано в данном изобретении, где управляющая РНК связана с полипептидом РНК-управляющей нуклеазы.

Управляющая РНК направляет связанную РНК-управляемую нуклеазу к конкретной представляющей интерес целевой нуклеотидной последовательности посредством гибридизации управляющей РНК с целевой нуклеотидной последовательностью. Целевая нуклеотидная последовательность может содержать ДНК, РНК или их комбинацию и может быть одно-цепочечной или двух-цепочечной. Целевая нуклеотидная последовательность может представлять собой геномную ДНК (т.е. хромосомную ДНК), плазмидную ДНК или молекулу РНК (например, информационную РНК, рибосомную РНК, транспортную РНК, микро РНК, малую интерферирующую РНК). Целевая нуклеотидная последовательность может быть связана с (и в некоторых вариантах осуществления изобретения расщеплена) нуклеазой, управляемой РНК, *in vitro* или в клетке. Хромосомная последовательность, на которую нацелена RGN, может быть ядерной, плазмидной или митохондриальной хромосомной последовательностью. В некоторых вариантах осуществления изобретения целевая нуклеотидная последовательность уникальна для целевого генома.

Целевая нуклеотидная последовательность находится рядом с мотивом, примыкающим к протоспейсеру (PAM). Мотив, прилегающий к протоспейсеру, обычно находится в пределах примерно от 1 до примерно 10 нуклеотидов от целевой нуклеотидной последовательности, включая примерно 1, примерно 2, примерно 3, примерно 4, примерно 5, примерно 6, примерно 7, примерно 8, примерно 9 или примерно 10 нуклеотидов из нуклеотидной последовательности-мишени. PAM может быть 5' или 3' целевой последовательности. В некоторых вариантах осуществления изобретения PAM представляет собой 3' последовательности для представленных здесь RGN. Обычно PAM представляет собой консенсусную последовательность примерно из 3-4 нуклеотидов, но в конкретных вариантах осуществления изобретения может иметь длину 2, 3, 4, 5, 6, 7, 8, 9 или более нуклеотидов. В различных вариантах осуществления изобретения последовательность PAM, распознаваемая представленными здесь RGN, включает консенсусную последовательность, представленную как SEQ ID NO: 6, 32, 41, 50 или 59. Неограничивающие иллюстративные последовательности PAM представляют собой нуклеотидные последовательности, представленные как SEQ ID NO: 7, 69, 70, 71 и 72.

В определенных вариантах осуществления изобретения РНК-управляемая нуклеаза, имеющая SEQ ID NO: 1, 11, 19, 27, 36, 45 или 54, или ее активный вариант или фрагмент, связывает, соответственно, целевую нуклеотидную последовательность, соседнюю с последовательностью PAM, обозначенную как SEQ ID NO: 6, 32, 41, 50, 59 или 7. В некоторых вариантах осуществления изобретения RGN связывается с управляющей последовательностью, содержащей повторяющуюся последовательность CRISPR, представленную как SEQ ID NO: 2, 12, 20, 28, 37, 46 или 55, соответственно, или их активный вариант или фрагмент, и последовательность тасгРНК, указанная в SEQ ID NO: 3, 13, 21, 29, 38, 47 или 56, соответственно, или активный вариант или фрагмент из них. Системы RGN дополнительно описаны в примере 1 и таблице 1 настоящего описания.

В данной области хорошо известно, что на специфичность последовательности PAM для данного фермента нуклеазы влияет концентрация фермента (см., например, Karvelis и др. (2015) *Genome Biol* № 16: сс. 253), которая может быть изменена путем изменения промотора, используемого для экспрессии RGN, или количества рибонуклеопротеинового комплекса, доставленного в клетку, органеллу или эмбрион.

После распознавания соответствующей последовательности PAM, RGN может расщеплять целевую нуклеотидную последовательность в конкретном сайте расщепления. В контексте настоящего описания сайт расщепления состоит из двух конкретных нуклеотидов в целевой нуклеотидной последовательности, между которыми нуклеотидная последовательность расщепляется 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 нуклеотиды PAM в 5'

или 3' направлениях. В некоторых вариантах осуществления изобретения сайт расщепления может составлять более 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 или 20 нуклеотидов от РАМ в 5' или 3' направлениях. В некоторых вариантах осуществления изобретения сайт расщепления находится на расстоянии 4 нуклеотидов от РАМ. В других вариантах осуществления изобретения сайт расщепления находится на расстоянии по меньшей мере 15 нуклеотидов от РАМ. Поскольку RGN могут расщеплять целевую нуклеотидную последовательность, приводя к расположению концов в шахматном порядке, в некоторых вариантах осуществления изобретения сайт расщепления определяется на основе расстояния двух нуклеотидов от РАМ на положительной (+) цепи полинуклеотида и расстояния двух нуклеотидов от РАМ на отрицательной (-) цепи полинуклеотида.

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

В настоящем описании представлены полинуклеотиды, описанные в настоящем изобретении РНК CRISPR, tracrРНК и/или sgРНК и полинуклеотиды, содержащие нуклеотидную последовательность, кодирующую в настоящем изобретении РНК-управляемые нуклеазы, CRISPR-РНК, tracrРНК и/или sgРНК. Описанные в предлагаемом изобретении полинуклеотиды включают те, которые содержат или кодируют повторяющуюся последовательность CRISPR, содержащую нуклеотидную последовательность SEQ ID NO: 2, 12, 20, 28, 37, 46 или 55, или ее активный вариант или фрагмент, который, когда он содержится в управляющей РНК, способен управлять специфичным для последовательности связыванием соответствующей РНК-управляемой нуклеазой с представляющей интерес целевой последовательностью. Также описаны полинуклеотиды, включающие в себя или кодирующие tracrРНК, содержащую нуклеотидную последовательность SEQ ID NO: 3, 13, 21, 29, 38, 47 или 56, или ее активный вариант или фрагмент, который при включении в управляющую РНК способен направлять специфичное для последовательности связывание соответствующей РНК-управляемой нуклеазы с представляющей интерес целевой последовательностью. Также предлагаются полинуклеотиды, которые кодируют РНК-управляемую нуклеазу, содержащую аминокислотную последовательность, обозначенную как SEQ ID NO: 1, 11, 19, 27, 36, 45 или 54, и ее активные фрагменты или варианты, которые сохраняют способность связываться с целевой нуклеотидной последовательностью специфичным для последовательности способом, управляемым РНК.

Использование термина "полинуклеотид" не предназначено для ограничения настоящего описания изобретения полинуклеотидами, содержащими ДНК. Специалисты в данной области поймут, что полинуклеотиды могут 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), химеры ПНК-ДНК, заблокированные нуклеиновые кислоты (LNA) и последовательности, связанные фосфотиоратом. Описанные в настоящем изобретении полинуклеотиды также охватывают все формы последовательностей, включая, но, не ограничиваясь ими, одно-цепочечные формы, двух-цепочечные формы, гибриды ДНК-РНК, триплексные структуры, структуры типа стержень и петля и т.п.

Молекулы нуклеиновых кислот, кодирующие RGN, могут быть оптимизированы по кодоновому составу для экспрессии в представляющем интерес организме. "Оптимизированная по кодоновому составу" кодирующая последовательность представляет собой полинуклеотидную кодирующую последовательность, частота использования кодонов которой разработана таким образом, чтобы имитировать частоту предпочтительного использования кодонов или условия транскрипции определенной клетки-хозяина. Экспрессия в определенной клетке-хозяине или организме усиливается в результате изменения одного или нескольких кодонов на уровне нуклеиновой кислоты таким образом, что транслируемая аминокислотная последовательность не изменяется. Молекулы нуклеиновой кислоты могут быть оптимизированы по кодоновому составу полностью или частично. Таблицы кодонов и другие ссылки, содержащие информацию о предпочтениях для широкого диапазона организмов доступны в данной области техники (см., например, Campbell and Gowri (1990) *Plant Physiol.* № 92: сс. 1-11 для обсуждения использования предпочтительных для растений кодонов). В данной области доступны способы синтеза предпочтительных для растений генов. См., например, патенты США №№ 5380831 и 5436391, а также Муггау и др. (1989) *Nucleic Acids Res.* №17: сс. 477-498, включенные в данное описание изобретения в качестве ссылки.

Полинуклеотиды, кодирующие RGN, crРНК, tracrРНК и/или sgРНК, представленные в настоящем документе, могут быть входить в кассеты экспрессии для экспрессии *in vitro* или экспрессии в представляющей интерес клетке, органелле, эмбрионе или организме. Кассета будет включать 5'- и 3'-регуляторные последовательности, функционально связанные с полинуклеотидом, кодирующим RGN, crРНК, tracrРНК и/или sgРНК, представленными в данном изобретении, которые обеспечивают экспрессию полинуклеотида. Кассета может дополнительно содержать по меньшей мере один дополнительный ген или генетический элемент, ко-трансформируемый в организм. Если включены дополнительные гены или элементы, компоненты функционально связаны. Термин "функционально связанный" предназначен для обозначения функциональной связи между двумя или более элементами. Например, функциональная связь между промотором и интересующей кодирующей областью (например, областью, кодирующей RGN, crРНК, tracrРНК и/или sgРНК) является функциональной связью, которая обеспечивает экспрес-

сию интересующей кодирующей области. Функционально связанные элементы могут быть смежными или несмежными. При использовании для обозначения соединения двух областей, кодирующих белок, термин "функционально связанный" означает, что кодирующие области находятся в одной и той же рамке считывания. Кроме того, дополнительный ген(ы) или элемент(ы) могут быть предоставлены на нескольких кассетах экспрессии. Например, нуклеотидная последовательность, кодирующая описанную здесь RGN, может присутствовать на одной кассете экспрессии, тогда как нуклеотидная последовательность, кодирующая *сr*РНК, *tracr*РНК или полную управляющую РНК, может находиться на отдельной кассете экспрессии. Такая кассета экспрессии снабжена множеством сайтов рестрикции и/или сайтов рекомбинации для вставки полинуклеотидов, находящихся под регуляцией транскрипции регуляторных областей. Кассета экспрессии может дополнительно содержать ген селективного маркера.

Кассета экспрессии будет включать в 5'-3'направлении транскрипции: область инициации транскрипции (и, в некоторых вариантах осуществления изобретения, трансляции) (т.е. промотор), RGN, *сr*РНК, *tracr*РНК и/или *sg*РНК-кодирующий полинуклеотид, представленный в изобретении, и область терминации транскрипции (а в некоторых вариантах осуществления изобретения, область терминации трансляции), функциональную в представляющем интерес организм. Промоторы, описанные в изобретении, способны направлять или управлять экспрессией кодирующей последовательности в клетке-хозяине.

Регуляторные области (например, промоторы, регуляторные области транскрипции и области терминации трансляции) могут быть эндогенными или гетерологичными по отношению к клетке-хозяину или друг к другу. Используемый здесь термин "гетерологичный" по отношению к последовательности представляет собой последовательность, которая происходит от чужеродного вида или, если от того же вида, существенно модифицирована по сравнению с ее природной формой по составу и/или геномному локусу в результате преднамеренного вмешательства человека. В данном контексте химерный ген содержит кодирующую последовательность, функционально связанную с областью инициации транскрипции, которая является гетерологичной кодирующей последовательности.

Подходящие области терминации доступны из T1-плазмиды *A. tumefaciens*, такие как области терминации октопинсинтазы и нопалинсинтазы. См. также Guerineau и др. (1991) *Mol. Genet. Genet.* № 262: сс. 141-144; Proudfoot (1991) *Cell* № 64: сс. 671-674; Sanfaçon и др. (1991) *Genes Dev.* № 5: сс. 141-149; Mogen и др. (1990) *Plant Cell* № 2: сс. 1261-1272; Munroe и др. (1990) *Gene* № 91: сс. 151-158; Ballas и др. (1989) *Nucleic Acids Res.* № 17: сс. 7891-7903; и Joshi и др. (1987) *Nucleic Acids Res.* № 15: сс. 9627-9639.

Дополнительные регуляторные сигналы включают, но не ограничиваются ими, сайты начала инициации транскрипции, операторы, активаторы, энхансеры, другие регуляторные элементы, сайты связывания рибосом, кодон инициации, сигналы терминации и тому подобное. См., например, патент США No. №№ 5039523 и 4853331; EPO 0480762A2; Sambrook и др. (1992) *Molecular Cloning: A Laboratory Manual*, под ред. Maniatis и др. (Издательство Cold Spring Harbor Laboratory Press, Cold Spring Harbor, Нью-Йорк), далее "Sambrook 11"; Davis и др., ред. (1980) *Advanced Bacterial Genetics* (Издательство Cold Spring Harbor Laboratory Press), Издательство Cold Spring Harbor, Нью-Йорк, и цитируемые здесь ссылки.

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

В практике изобретения можно использовать ряд промоторов. Промоторы могут быть выбраны в зависимости от желаемого результата. Нуклеиновые кислоты могут быть объединены с конститутивными, индуцибельными, специфическими для стадии роста, специфическими для клеточного типа, предпочтительными для ткани, тканеспецифическими или другими промоторами для экспрессии в интересующем организме. См., например, промоторы, изложенные в WO 99/43838 и в патентах США № 8,575,425; 7,790,846; 8 147 856; 8,586,832; 7,772,369; 7,534,939; 6072050; 5,659,026; 5,608,149; 5,608,144; 5,604,121; 5,569,597; 5,466,785; 5,399,680; 5 268 463; 5,608,142; и 6 177 611, в данном документе включенные в качестве ссылки.

Для экспрессии в растениях конститутивные промоторы также включают промотор 35S CaMV (Odell и др. (1985) *Nature* № 313: сс. 810-812); актина риса (McElroy и др. (1990) *Plant Cell* № 2: сс. 163-171); убиквитина (Christensen и др. (1989) *Plant Mol. Biol.* № 12: сс. 619-632 и Christensen и др. (1992) *Plant Mol. Biol.* № 18: сс. 675-689); pEMU (Last и др. (1991) *Theor. Appl. Genet.* № 81: сс. 581-588); и MAS (Velten и др. (1984) *EMBO J.* № 3: сс. 2723-2730).

Примерами индуцибельных промоторов являются промотор *Adh1*, который индуцируется гипоксией или Холодовым стрессом, промотор *Hsp70*, который индуцируется тепловым стрессом, промотор *PPDK* и промотор пепкарбоксилазы, которые индуцируются светом. Также полезны промоторы, которые являются химически индуцируемыми, такие как промотор *In2-2*, индуцируемый антидотом (патент США № 5364780), промотор *Axig1*, который индуцируется ауксином и специфичен для тапетума, но также

активен в каллусе (PCT US01 /22169 который), стероид-чувствительные промоторы (см., например, промотор ERE, который индуцируется эстрогеном, и промотор, индуцируемый глюкокортикоидами, описанные у Schena и др. (1991) Proc. Natl. Acad. Sci. USA № 88: 10421-10425 и McNellis и др. (1998) Plant J. №14 (2): сс. 247-257)), а так же промоторы, индуцируемые и репресслируемые тетрациклином (см., например, Gatz и др. (1991) Mol. Gen. Genet. № 227: сс. 229-237 и патенты США №№ 5814618 и 5789156, которые включены в описание изобретения в качестве ссылки).

Тканеспецифические или предпочтительные для тканей промоторы можно использовать для нацеливания на экспрессию экспрессирующей конструкции в конкретной ткани. В некоторых вариантах осуществления изобретения тканеспецифические или предпочтительные для тканей промоторы активны в растительной ткани. Примеры промоторов, находящихся под контролем развития растений, включают промоторы, которые инициируют транскрипцию преимущественно в определенных тканях, таких как листья, корни, плоды, семена или цветы. "Тканеспецифический" промотор - это промотор, который инициирует транскрипцию только в определенных тканях. В отличие от конститутивной экспрессии генов, тканеспецифическая экспрессия является результатом нескольких взаимодействующих уровней генной регуляции. По существу, промоторы из гомологичных или близкородственных видов растений могут быть предпочтительны при использовании для достижения эффективной и надежной экспрессии трансгенов в конкретных тканях. В некоторых вариантах осуществления изобретения экспрессия включает предпочтительный для ткани промотор. "Предпочтительный для ткани" промотор представляет собой промотор, который предпочтительно инициирует транскрипцию, но не обязательно полностью или только в определенных тканях.

В некоторых вариантах осуществления изобретения молекулы нуклеиновой кислоты, кодирующие RGN, *cr*PHK и/или *tracr*PHK, содержат промотор, специфичный для определенного типа клеток. Промотор "специфичный для определенного типа клетки" представляет собой промотор, 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 активными промоторы, специфичные для определенного типа клеток, функциональные в растениях, включают, например, клетки BETL, сосудистые клетки в корнях, листьях, клетки стебля и ствольные клетки. Молекулы нуклеиновой кислоты могут также включать промоторы, предпочтительные для определенного типа клеток. Промотор, "предпочтительный для определенного типа клеток", представляет собой промотор, который в первую очередь управляет экспрессией в основном, но не обязательно полностью или только в определенных типах клеток в одном или нескольких органах. Некоторые примеры растительных клеток, в которых промоторы, предпочтительные для определенного типа клеток, функциональные в растениях, могут быть преимущественно активными, включают, например, клетки BETL, сосудистые клетки в корнях, листьях, клетки стебля и ствольные клетки.

Последовательности нуклеиновых кислот, кодирующие RGN, *cr*PHK, *tracr*PHK и/или *sg*PHK, могут быть функционально связаны с промоторной последовательностью, которая распознается фаговой РНК-полимеразой, например, для синтеза мРНК *in vitro*. В таких вариантах осуществления РНК, транскрибируемая *in vitro*, может быть очищена для использования в описанных в данном документе способах. Например, промоторная последовательность может быть промоторной последовательностью T7, T3 или SP6 или вариантом промоторной последовательности T7, T3 или SP6. В таких вариантах осуществления изобретения экспрессированный белок и/или РНК могут быть очищены для использования в способах модификации генома, описанных в предлагаемом изобретении.

В некоторых вариантах осуществления изобретения полинуклеотид, кодирующий RGN, *cr*PHK, *tracr*PHK и/или *sg*PHK, также может быть связан с сигналом полиаденилирования (например, сигналом полиА SV40 и другими сигналами, функциональными в растениях) и/или по меньшей мере с одной последовательностью терминации транскрипции. Кроме того, последовательность, кодирующая RGN, также может быть связана с последовательностью(ями), кодирующей по меньшей мере один сигнал ядерной локализации, по меньшей мере один проникающий в клетки домен и/или по меньшей мере один сигнальный пептид, способный доставлять белки в определенные субклеточные местоположения, как описано в другом месте описания.

Полинуклеотид, кодирующий RGN, *cr*PHK, *tracr*PHK и/или *sg*PHK, может присутствовать в векторе или нескольких векторах. "Вектор" относится к полинуклеотидной композиции для переноса, доставки или введения нуклеиновой кислоты в клетку-хозяина. Подходящие векторы включают плазмидные векторы, фагомиды, 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 т.п. Дополнительную информацию можно найти в "Current Protocols in Molecular Biology" Ausubel и др., Издательство John Wiley & Sons, Нью-Йорк, 2003 или "Molecular Cloning: A Laboratory Manual" Издательство Sambrook & Russell, Cold Spring Harbor Press, Cold Spring Harbor, Нью-Йорк, 3-е издание, 2001.

Вектор также может содержать ген селективируемого маркера для отбора трансформированных клеток. Гены селективируемых маркеров используются для отбора трансформированных клеток или тканей. Гены-маркеры включают гены, кодирующие устойчивость к антибиотикам, такие как гены, кодирующие неомизинфосфотрансферазу II (NEO) и гигромицинфосфотрансферазу (HPT), а также гены, придающие устойчивость к гербицидным соединениям, таким как глюофосинатаммоний, бромоксинил, имидазолиноны и 2,4-оксидхлорфенат. (2,4-Д).

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

Как указано, экспрессионные конструкции, содержащие нуклеотидные последовательности, кодирующие RGN, сгРНК, тсгРНК и/или sgРНК, могут использоваться для трансформации представляющих интерес организмов. Способы трансформации включают введение нуклеотидной конструкции в интересующий организм. Под "введением" подразумевается введение нуклеотидной конструкции в клетку-хозяина таким образом, чтобы конструкция получила доступ внутрь клетки-хозяина. Способы изобретения не требуют особого способа введения нуклеотидной конструкции в организм-хозяина, кроме такого, при котором нуклеотидная конструкция получает доступ внутрь по меньшей мере одной клетки организма-хозяина. Клетка-хозяин может быть эукариотической или прокариотической клеткой. В конкретных вариантах осуществления изобретения эукариотическая клетка-хозяин представляет собой растительную клетку, клетку млекопитающего или клетку насекомого. Способы введения нуклеотидных конструкций в растения и другие клетки-хозяева известны в данной области техники и включают методы стабильной трансформации, методы временной трансформации и методы, опосредованные вирусами, но не ограничиваются ими.

В результате этих методов получают трансформированный организм, такой как растение, включая целое растение, а также органы растения (например, листья, стебли, корни и т.д.), семена, клетки растений, ростки, зародыши и их потомство. Клетки растений могут быть дифференцированными или недифференцированными (например, каллус, клетки суспензионной культуры, протопласты, клетки листа, клетки корня, клетки флоэмы, пыльца).

"Трансгенные организмы", или "трансформированные организмы", или "стабильно трансформированные" организмы, или клетки, или ткани относятся к организмам, которые включили или интегрировали полинуклеотид, кодирующий RGN, сгРНК и/или тсгРНК, описанные в предлагаемом изобретении. Признано, что другие экзогенные или эндогенные последовательности нуклеиновых кислот или фрагменты ДНК также могут быть включены в клетку-хозяина. Агробактериальная-и-биолистически опосредованная трансформация остаются двумя преимущественно используемыми подходами для трансформации растительных клеток. Однако трансформация клетки-хозяина может быть осуществлена путем инфицирования, трансфекции, микроинъекции, электропорации, микропроекции, биолистики или бомбардировки частицами, электропорации, осуществлена с помощью диоксида кремния/углеродных волокон, она может быть опосредована ультразвуком, опосредована ПЭГ, осуществлена ко-осаждением с фосфатом кальция, осуществлена поликатион-DMSO, DEAE декстрановой процедурой, а также может быть опосредованной вирусами, опосредованной липосомами и т.п. Опосредованное вирусами введение полинуклеотида, кодирующего RGN, сгРНК и/или тсгРНК, включает опосредованное ретровирусом, лентивирусом, аденовирусом и аденоассоциированным вирусом введение и экспрессию, а также использование Колимовирусов (Caulimoviruses), Герминивирусов (Geminiviruses) и РНК вирусов растений.

Протоколы трансформации, а также протоколы введения полипептидов или полинуклеотидных последовательностей в растения могут варьироваться в зависимости от типа клетки-хозяина (например, клетка однодольного или двудольного растения), на которую нацелена трансформация. Способы трансформации известны в данной области и включают те, которые изложены в патентах США №№ 8,575,425; 7,692,068; 8,802,934; 7,541,517, каждый из которых включен в данное изобретение в качестве ссылки. См. также: Rakoczy-Trojanowska, M. (2002) *Cell Mol Biol Lett.* № 7: 849-858; Jones и др. (2005) *Plant Methods* № 1: с. 5; Rivera и др. (2012) *Physics of Life Reviews* № 9: сс. 308-345; Bartlett и др. (2008) *Plant Methods* № 4: сс. 1-12; Bates, G.W. (1999) *Methods in Molecular Biology* № 111: сс. 359-366; Binns and Thomashow (1988) *Annual Reviews in Microbiology* № 42: сс. 575-606; Christou, P. (1992) *The Plant Journal* № 2: сс. 275-281; Christou, P. (1995) *Euphytica* № 85: сс. 13-27; Tzfira и др. (2004) *TRENDS in Genetics* № 20: сс. 375-383; Yao и др. (2006) *Journal of Experimental Botany* № 57: сс. 3737-3746; Zupan and Zambryski (1995) *Plant Physiology* № 107: сс. 1041-1047; Jones и др. (2005) *Plant Methods* № 1: с. 5.

Трансформация может привести к стабильному или временному включению нуклеиновой кислоты в клетку. "Стабильная трансформация" означает, что нуклеотидная конструкция, введенная в хозяйскую

клетку, интегрируется в геном хозяйской клетки и способна наследоваться ее потомками. "Временная трансформация" означает, что полинуклеотид вводится в клетку-хозяина и не интегрируется в геном клетки-хозяина.

Способы трансформации хлоропластов известны в данной области техники. См., например, Svab и др. (1990) Proc. Natl. Acad. Sci. USA № 87: сс. 8526-8530; Svab and Maliga (1993) Proc. Natl. Acad. Sci. USA № 90: сс. 913-917; Svab and Maliga (1993) EMBO J. № 12: сс. 601-606. Этот метод основан на доставке ДНК, содержащей селективируемый маркер, и нацеливании ДНК на пластидный геном посредством гомологичной рекомбинации. Кроме того, трансформация пластид может быть осуществлена путем трансактивации молчащего трансгена, переносимого пластидами, с помощью экспрессии предпочтительной для ткани кодируемой ядром и направленной на пластиду РНК-полимеразы. О такой системе сообщалось у McBride и др. (1994) Proc. Natl. Acad. Sci. США № 91: сс. 7301-7305.

Трансформированные клетки можно вырастить в трансгенный организм, такой как растение, обычными способами. См., например, McCormick и др. (1986) Plant Cell Reports № 5: сс. 81-84. Затем эти растения можно выращивать и опылять одним и тем же трансформированным штаммом или разными штаммами, и идентифицировать полученный гибрид, имеющий конститутивную экспрессию желаемой фенотипической характеристики. Можно выращивать два или более поколений, чтобы гарантировать, что экспрессия желаемой фенотипической характеристики стабильно сохраняется и наследуется, а затем собирать семена, чтобы гарантировать проявление желаемой фенотипической характеристики. Таким образом, настоящее изобретение обеспечивает трансформированные семена (также называемые "трансгенные семена"), содержащие нуклеотидную конструкцию, описанную в данном изобретении, например кассету экспрессии, описанную в данном изобретении, стабильно встроенную в их геном.

Кроме того, трансформированные клетки могут быть введены в организм. Эти клетки могли бы происходить из организма, в котором клетки трансформируются методом *ex vivo*.

Приведенные в данном изобретении последовательности можно использовать для трансформации любых видов растений, включая, помимо прочего, однодольные и двудольные. Примеры представляющих интерес растений включают, но не ограничиваются ими, кукурузу (маис), сорго, пшеницу, подсолнечник, помидоры, крестоцветные, перец, картофель, хлопок, рис, сою, сахарную свеклу, сахарный тростник, табак, ячмень и масличный рапс, Brassica sp., люцерна, рожь, просо, сафлор, арахис, сладкий картофель, кассая, кофе, кокос, ананас, цитрусовые деревья, какао, чай, банан, авокадо, инжир, гуава, манго, оливки, папайя, кешью, макадамия, миндаль, овес, овощи, декоративные растения и хвойные деревья.

Овощи включают, помимо прочего, помидоры, салат, стручковую фасоль, фасоль лима, горох и представителей рода *Cucumis*, такие как огурец, дыня и мускусная дыня. Декоративные растения включают, помимо прочего, азалию, гортензию, гибискус, розы, тюльпаны, нарциссы, петунии, гвоздики, пуансеттию и хризантемы. Предпочтительно, растения по настоящему изобретению представляют собой культурные растения (например, кукурузу, сорго, пшеницу, подсолнечник, томат, крестоцветные, перец, картофель, хлопок, рис, сою, сахарную свеклу, сахарный тростник, табак, ячмень, масличный рапс и т.д.).

Используемый в данном изобретении термин "растение" включает растительные клетки, протопласты растений, культуры тканей растительных клеток, из которых можно регенерировать растения, каллусы растений, сгустки растений и растительные клетки, которые не повреждены в растениях или частях растений, таких как зародыши, пыльца, семяпочки, семена, листья, цветы, ветви, плоды, ядра, колосья, початки, шелуха, стебли, корни, кончики корней, пыльники и т.п. Под зерном подразумеваются зрелые семена, произведенные коммерческими производителями для целей, не связанных с выращиванием или воспроизводством вида. Потомство, варианты и мутанты регенерированных растений также включены в объем изобретения при условии, что эти части содержат введенные полинуклеотиды. Кроме того, предоставляется обработанный растительный продукт или побочный продукт, который сохраняет последовательности, описанные в данном изобретении, включая, например, соевые продукты.

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

Полинуклеотиды, кодирующие RGN, crPHK и/или tracrPHK, можно использовать для трансформации любых эукариотических видов, включая, помимо прочего, животных (например, млекопитающих, насекомых, рыб, птиц и рептилий), грибы, амёбы, водоросли и дрожжи.

Для введения нуклеиновых кислот в клетки млекопитающих или целевые ткани можно использовать обычные методы переноса генов на вирусной и невирусной основе. Такие методы можно использовать для введения нуклеиновых кислот, кодирующих компоненты системы CRISPR, в клетки в культуре или в организме-хозяине. Системы доставки невирусных векторов включают плазмиды ДНК, РНК (например, транскрипт вектора, описанного в данном документе), голую нуклеиновую кислоту и нуклеиновую кислоту в комплексе с носителем для доставки, таким как липосома. Системы доставки вирусных векторов включают ДНК и РНК вирусы, которые после доставки в клетку имеют либо эписомальный, либо интегрированный геномы. Для обзора процедур генной терапии см. Anderson, Science № 256: сс.

808-813 (1992); Nabel и Feigner, TIBTECH 11: 211-217 (1993); Mitani и Caskey, TIBTECH 11: 162-166 (1993); Dillon, TIBTECH 11: 167-175 (1993); Miller, Nature № 357: сс. 455-460 (1992); Van Brunt, Biotechnology № 6 (10): сс. 1149-1154 (1988); Vigne, Восстановительная неврология и неврология, № 8: сс. 35-36 (1995); Kremer и Perricaudet, British Medical Bulletin № 51 (1): сс. 31-44 (1995); Haddada и др., в Current Topics in Microbiology and Immunology, под ред. Doerfler and Bohm (1995); и Yu и др., Gene Therapy № 1: сс. 13-26 (1994).

Способы невирусной доставки нуклеиновых кислот включают липофекцию, нуклеофекцию, микроинъекцию, биолистику, виросомы, липосомы, иммунолипосомы, поликатион или липид: конъюгаты нуклеиновой кислоты, голую ДНК, искусственные вирионы и усиленный агентом захват ДНК. Липофекция описана, например, в патентах США No. №№ 5049386, 4946787; и 4897355), а реагенты для липофекции продаются коммерчески (например, Transfectam™ и Lipofectin™). Катионные и нейтральные липиды, которые подходят для эффективной липофекции полинуклеотидов с распознаванием рецепторов, включают липиды Feigner, WO 91/17424; WO 91/16024. Доставка может осуществляться в клетки (например, введение *in vitro* или *ex vivo*) или в целевые ткани (например, введение *in vivo*). Получение комплексов липид: нуклеиновая кислота, включая целевые липосомы, такие как иммунолипидные комплексы, хорошо известно специалисту в данной области (см., например, Crystal, Science № 270: сс. 404-410 (1995); Blaese и др., Cancer Gene Ther. № 2: сс. 291-297 (1995); Behr и др., Bioconjugate Chem. № 5: сс. 382-389 (1994); Remy и др., Bioconjugate Chem. № 5: сс. 647-654 (1994); Gao и др., Gene Therapy № 2: сс. 710-722 (1995); Ahmad и др., Cancer Res. № 52: сс. 4817-4820 (1992); патенты США 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, и 4,946,787).

Использование систем РНК или ДНК на основе вирусов для доставки нуклеиновых кислот пользуется преимуществом перед высокоразвитыми процессами для нацеливания вируса на определенные клетки в организме и доставки вирусной полезной нагрузки к ядру. Вирусные векторы можно вводить непосредственно пациентам (*in vivo*) или их можно использовать для обработки клеток *in vitro*, а модифицированные клетки могут опционно вводиться пациентам (*ex vivo*). Обычные вирусные системы могут включать ретровирусные, лентивирусные, аденовирусные, аденоассоциированные векторы и векторы вируса простого герпеса. Интеграция в геном хозяина возможна с помощью методов переноса генов ретровируса, лентивируса и аденоассоциированного вируса, что часто приводит к долговременной экспрессии встроенного трансгена. Кроме того, высокая эффективность трансдукции наблюдалась во многих различных типах клеток и целевых тканях.

Тропизм ретровируса можно изменить путем включения белков чужой оболочки, увеличивая потенциальную целевую популяцию целевых клеток. Лентивирусные векторы - это ретровирусные векторы, которые способны трансдуктировать или инфицировать неделящиеся клетки и обычно продуцируют высокие титры вируса. Следовательно, выбор системы переноса ретровирусного гена будет зависеть от целевой ткани. Ретровирусные векторы состоят из цис-действующих длинных концевых повторов с упаковочной способностью до 6-10 т.п.н. чужеродной последовательности. Минимальных цис-действующих LTR достаточно для репликации и упаковки векторов, которые затем используются для интеграции терапевтического гена в целевую клетку для обеспечения постоянной экспрессии трансгена. Широко используются ретровирусные векторы включают векторы на основе вируса лейкемии мышей (MuLV), вируса лейкемии обезьян гиббонов (GaLV), вируса иммунодефицита обезьян (SIV), вируса иммунодефицита человека (ВИЧ) и их комбинаций (см., например, Buchscher и др., J. Viral. № 66: сс. -2739 (1992); Johann и др., J. Viral. № 66: сс. 1635-1640 (1992); Sommerfelt и др., Viral. №176: сс. 58-59 (1990); Wilson и др., J. Viral. №63: сс. 2374-2378 (1989); Miller и др., 1. Viral. № 65: сс. 2220-2224 (1991); PCT/US94/05700).

В прикладных задачах, где предпочтительна временная экспрессия, можно использовать системы на основе аденовирусов. Векторы на основе аденовирусов обладают очень высокой эффективностью трансдукции во многих типах клеток и не требуют деления клеток. С такими векторами были получены высокие титр и уровни экспрессии. Этот вектор может быть произведен в больших количествах в относительно простой системе. Векторы аденоассоциированного вируса ("AAV") также могут быть использованы для трансдукции клеток целевыми нуклеиновыми кислотами, например, при производстве нуклеиновых кислот и пептидов *in vitro*, а также для процедур генной терапии *in vivo* и *ex vivo* (см., например, West и др., Virology № 160: сс. 38-47 (1987); патент США № 4797368; WO 93/24641; Katin, Human Gene Therapy № 5: сс. 793-801 (1994); Muzyczka, 1. Clin. Invest. № 94: сс. 1351 (1994). Конструирование рекомбинантных векторов AAV описано в ряде публикаций, включая патент США №5 173 414; Tratschin и др., Mol. Cell. Biol. № 5: сс. 3251-3260 (1985); Tratschin и др., Mol. Cell. Biol. № 4: сс. 2072-2081 (1984); Hermonat и Muzyczka, PNAS № 81: сс. 6466-6470 (1984); и Samulski и др., 1. Viral. № 63: сс. 03822-3828 (1989) Упаковочные клетки обычно используются для образования вирусных частиц, способных инфицировать клетку-хозяина. Такие клетки включают клетки 293, которые упаковывают аденовирус, и клетки ψJ2 или клетки PA317, которые упаковывают ретровирус.

Вирусные векторы, используемые в генной терапии, обычно получают путем получения линии клеток, которая упаковывает вектор на основе нуклеиновой кислоты в вирусную частицу. Векторы обычно содержат минимальные вирусные последовательности, необходимые для упаковки и последующей инте-

грации в хозяина, другие вирусные последовательности заменяются кассетой экспрессии для экспрессируемого(ых) полинуклеотида(ов). Утраченные вирусные функции обычно передаются упаковочной транс-клеточной линии. 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 помещена в клеточную линию, которая содержит вспомогательную плазмиду, кодирующую другие гены AAV, а именно гер и сар, но без последовательностей ITR.

Клеточная линия также может быть инфицирована аденовирусом в качестве помощника. Хелперные вирусные промоторы способствуют репликации вектора AAV и экспрессии генов AAV из хелперной плазмиды. Хелперная плазида не упакована в значительных количествах из-за отсутствия последовательностей ITR. Загрязнение аденовирусом можно снизить, например, с помощью тепловой обработки, к которой аденовирус более чувствителен, чем AAV. Дополнительные методы доставки нуклеиновых кислот в клетки известны специалистам в данной области. См., например, US20030087817, включенный в данном документе в качестве ссылки.

В некоторых вариантах осуществления изобретения клетка-хозяин временно или непременно трансфицируется одним или несколькими описанными здесь векторами. В некоторых вариантах осуществления изобретения клетка трансфицируется, как это обычно происходит у субъекта. В некоторых вариантах осуществления изобретения трансфицированная клетка берется у субъекта. В некоторых вариантах осуществления изобретения клетка получена из клеток, взятых у субъекта, таких как клеточная линия. В данной области известно большое количество клеточных линий для тканевых культур. Примеры клеточных линий включают, но не ограничиваются,

C8161, CCRF-CEM, MOLT,

mIMCD-3, NHDF, HeLaS3, Huhl, Huh4, Huh7, HUVEC, HASMC, HEKn, HEKa, MiaPaCell, Panel, PC-3, TFl, CTLL-2, CIR, Rat6, CVI, RPTE, AIO, T24, 182, A375, ARH-77, Calul, SW480, SW620, SKOV3, SK-UT, CaCo2, P388DI, SEM-K2, WENI-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, BS-C-1

эпителиальный почечный эпителий обезьяны, фибробласт эмбриона мыши BALB/3T3, фибробласты плода человека 3T3, 3T3-L1, 132-d5; 10.1 фибробласты мыши,

293-T, 3T3, 721, 9L, A2780, A2780ADR, A2780cis, A172, A20, A253, A431, A-549, ALC, B16, B35, клетки BCP-I, 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, lurkat, клетки IY, клетки K562, Ku812, KCL22, KGI, KYOI, 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, линии клеток OPCN / OPCT, Peer, PNT-1A / PNT 2, RenCa, RIN-5F, RMA / RMAS, клетки Saos-2, Sf-9, SkBr3, T2, T-47D, T84, линия клеток THPI, U373, U87, U937, VCaP, клетки Vero, WM39, WT-49, X63, YAC-1, YAR

и их трансгенные разновидности. Клеточные линии доступны из множества источников, известных специалистам в данной области (см., например, Американскую коллекцию типовых культур (ATCC), Манассас, Вирджиния).

В некоторых вариантах осуществления изобретения клетка, трансфицированная одним или несколькими описанными здесь векторами, используется для создания новой линии клеток, содержащей одну или несколько последовательностей, полученных из вектора. В некоторых вариантах осуществления изобретения клетка, временно трансфицированная компонентами системы CRISPR, как описано в настоящем изобретении (например, временной трансфекцией одного или нескольких векторов или трансфекцией РНК), и модифицированная посредством активности комплекса CRISPR, используется для

создания новой клеточной линии, включающей клетки, содержащие модификацию, но не имеющие какой-либо другой экзогенной последовательности. В некоторых вариантах осуществления изобретения клетки, временно или непременно трансфицированные одним или несколькими векторами, описанными в предлагаемом изобретении, или клеточные линии, полученные из таких клеток, используются для оценки одного или нескольких тестируемых соединений.

В некоторых вариантах осуществления изобретения один или несколько описанных здесь векторов используются для получения трансгенного растения или трансгенного животного, не являющегося человеком. В некоторых вариантах реализации трансгенное животное представляет собой млекопитающее, такое как мышь, крыса или кролик.

IV. Варианты и фрагменты полипептидов и полинуклеотидов.

В настоящем описании изобретения представлены активные варианты и фрагменты природной (т.е. дикого типа) РНК-управляемой нуклеазы, аминокислотная последовательность которой представлена в SEQ ID NO: 1, 11, 19, 27, 36, 45 или 54, а также активные варианты и фрагменты встречающихся в природе повторов CRISPR, такие как последовательность, обозначенная как SEQ ID NO: 2, 12, 20, 28, 37, 46 или 55, и активный вариант и фрагменты встречающейся в природе *tracrP*РНК, такие как последовательность, обозначенная как SEQ ID NO: 3, 13, 21, 29, 38, 47 или 56, и полинуклеотиды, кодирующие их.

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

Фрагменты и варианты встречающихся в природе полипептидов RGN, таких как полипептиды, описываемые в предлагаемом изобретении, будут сохранять специфичную для последовательности управляемую РНК ДНК-связывающую активность. В конкретных вариантах осуществления изобретения фрагменты и варианты встречающихся в природе полипептидов RGN (одно-цепочечные или двух-цепочечные), таких как описанные в данном документе, будут сохранять нуклеазную активность.

Фрагменты и варианты встречающихся в природе повторов CRISPR, таких как описанные в настоящем документе, сохраняют свою способность, как часть управляющей РНК (включающей *tracrP*РНК), связывать с и направлять РНК-управляемую нуклеазу (в комплексе с управляющей РНК) к целевой нуклеотидной последовательности специфичным для последовательности образом.

Фрагменты и варианты встречающихся в природе *tracrP*РНК, таких как описаны в предлагаемом изобретении, сохраняют способность, как часть управляющей РНК (включая CRISPR-РНК), направлять РНК-управляемую нуклеазу (в комплексе с управляющей РНК) к целевой нуклеотидной последовательности специфическим для последовательности образом.

Термин "фрагмент" относится к части полинуклеотидной или полипептидной последовательности, представленной в изобретении. "Фрагменты" или "биологически активные части" включают полинуклеотиды, содержащие достаточное количество смежных нуклеотидов для сохранения биологической активности (т.е. связывание и направление RGN специфическим для последовательности образом к целевой нуклеотидной последовательности, если она включена в управляющую РНК). "Фрагменты" или "биологически активные части" включают полипептиды, содержащие достаточное количество смежных аминокислотных остатков для сохранения биологической активности (т.е. связывания с целевой нуклеотидной последовательностью специфическим для последовательности образом при образовании комплекса с управляющей РНК). Фрагменты белков RGN включают те, которые короче полноразмерных последовательностей из-за использования расположенного ниже альтернативного стартового сайта. 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 смежных аминокислотных остатков SEQ ID NO: 1, 11, 19, 27, 36, 45 или 54. Такие биологически активные части могут быть получены рекомбинантными методами и оцениваются в отношении специфичной для последовательности РНК управляемой ДНК-связывающей активности. Биологически активный фрагмент последовательности CRISPR повторов может содержать по меньшей мере 8 смежных аминокислот SEQ ID NO: 2, 12, 20, 28, 37, 46 или 55. Биологически активная часть последовательности CRISPR повторов может быть полинуклеотидом, который включает, например, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 или 20 смежных нуклеотидов SEQ ID NO: 2, 12, 20, 28, 37, 46 или 55. Биологически активная часть *tracrP*РНК может быть полинуклеотидом, который включает, 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 SEQ ID NO: 3, 13, 21, 29, 38, 47 или 56.

В общем, "варианты" предназначены для обозначения практически одинаковых последовательностей. Для полинуклеотидов вариант включает делецию и/или вставку одного или нескольких нуклеотидов в одном или нескольких внутренних сайтах нативного полинуклеотида и/или замену одного или нескольких нуклеотидов в одном или нескольких сайтах в нативном полинуклеотиде. Используемый здесь термин "нативный" или "дикого типа" полинуклеотид или полипептид включает 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 интересующего 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, как описано ниже. Варианты полинуклеотидов также включают полинуклеотиды синтетического происхождения, такие как полинуклеотиды, полученные, например, с использованием сайт-направленного мутагенеза, но которые все еще кодируют представляющий интерес полипептид или полинуклеотид. Как правило, варианты конкретного полинуклеотида, описанные в предлагаемом изобретении, будут иметь по меньшей мере около 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или более идентичности последовательности с этим конкретным полинуклеотидом, как определено программами выравнивания последовательностей и параметрами, описанными в другом месте предлагаемого изобретения.

Варианты конкретного полинуклеотида, описанного в данном изобретении (то есть эталонного полинуклеотида), также можно оценивать путем сравнения процентной идентичности последовательностей между полипептидом, кодируемым вариантом полинуклеотида, и полипептидом, кодируемым эталонным полинуклеотидом. Процент идентичности последовательностей между любыми двумя полипептидами можно рассчитать с использованием программ выравнивания последовательностей и параметров, описанных в другом месте в настоящем документе. Если любую данную пару полинуклеотидов, описанных в предлагаемом изобретении, оценивают путем сравнения процентной идентичности последовательностей, разделяемой двумя полипептидами, 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 идентичности последовательностей.

В конкретных вариантах осуществления изобретения описанные в настоящем документе полинуклеотиды кодируют РНК-управляемый нуклеазный полипептид, содержащий аминокислотную последовательность, имеющую по меньшей мере 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 идентичности с аминокислотной последовательностью SEQ ID NO: 1, 11, 19, 27, 36, 45 или 54.

Биологически активный вариант полипептида RGN по изобретению может отличаться всего лишь примерно на 1-15, всего лишь примерно на 1-10, всего лишь на 6-10, всего лишь на 5, всего лишь на 4, всего лишь на 3, всего лишь на 2 аминокислотных остатка или всего лишь на 1 аминокислотный остаток. В конкретных вариантах осуществления изобретения полипептиды могут содержать усечение на N-конце или на C-конце, которое может иметь, по меньшей мере, делецию 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 аминокислотных остатков или более с N- или C-конца полипептида.

В некоторых вариантах осуществления изобретения полинуклеотиды содержат или кодируют повторы CRISPR, содержащий нуклеотидную последовательность, имеющую по меньшей мере 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 идентичности с нуклеотидной последовательностью, представленной в SEQ ID NO: 2, 12, 20, 28, 37, 46 или 55.

Описанные в настоящем изобретении полинуклеотиды могут содержать или кодировать *tracrRNA*, содержащую нуклеотидную последовательность, имеющую по меньшей мере 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 идентичности нуклеотидной последовательности, представленной в SEQ ID NO: 3, 13, 21, 29, 38, 47 или 56.

Биологически активные варианты повтора CRISPR или *tracrRNA* по изобретению могут отличаться всего лишь на 1-15 нуклеотидов, всего лишь на 1-10, всего лишь на 6-10, всего лишь на 5, всего лишь на 4, всего лишь на 3, всего лишь на 2 или всего лишь на 1 нуклеотид. В конкретных вариантах осуществления изобретения полинуклеотиды могут содержать 5' или 3' усечение, которое может включать по меньшей мере делецию 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80 нуклеотидов или более от 5' или 3' конца полинуклеотида.

Признано, что в полипептиды RGN, CRISPR повторы и *tracrRNA*, представленные в предлагаемом изобретении, могут быть внесены модификации, создавая варианты белков и полинуклеотидов. Изменения, разработанные человеком, могут быть внесены путем применения методов сайт-направленного мутагенеза. Кроме того, также могут быть идентифицированы нативные, пока неизвестные или еще не идентифицированные полинуклеотиды и/или полипептиды, структурно и/или функционально связанные с последовательностями, представленными в данном документе, которые входят в рамки настоящего изобретения. Консервативные аминокислотные замены могут быть осуществлены в неконсервативных областях, которые не изменяют функцию белков RGN. В качестве альтернативы могут быть внесены модификации, улучшающие активность RGN.

Варианты полинуклеотидов и белков также включают последовательности и белки, полученные в результате мутагенной и рекомбиногенной процедуры, такой как перетасовка ДНК. С помощью такой

процедуры одним или несколькими различными белками RGN, описанными в данном документе (например, SEQ ID NO: 1, 11, 19, 27, 36, 45 или 54), манипулируют для создания нового белка RGN, обладающего желаемыми свойствами. Таким образом, библиотеки рекомбинантных полинуклеотидов генерируются из популяции полинуклеотидов с родственными последовательностями, содержащими участки последовательностей, которые имеют существенную идентичность последовательностей и могут быть гомологично рекомбинированы *in vitro* или *in vivo*. Например, используя этот подход, мотивы последовательностей, кодирующие представляющий интерес домен, могут быть перетасованы между последовательностями RGN, представленными в предлагаемом изобретении, и другими известными генами RGN для получения нового гена, кодирующего белок с улучшенным представляющим интерес свойством, таким как увеличенный K_m в случае фермента. Стратегии такой перетасовки ДНК известны в данной области техники. См., например, Stemmer (1994) Proc. Natl. Acad. Sci. USA 91: 10747-10751; Stemmer (1994) Nature 370: 389-391; Cramer и др. (1997) Nature Biotech. 15: 436-438; Мур и др. (1997) J. Mol. Биол. 272: 336-347; Zhang и др. (1997) Proc. Natl. Acad. Sci. USA 94: 4504-4509; Cramer и др. (1998) Nature 391:288-291; и патенты США №№ 5,605,793 и 5,837,458. "Перетасованная" нуклеиновая кислота представляет собой нуклеиновую кислоту, полученную с помощью процедуры перетасовки, такой как любая процедура перетасовки, изложенная в предлагаемом изобретении. Перетасованные нуклеиновые кислоты получают путем рекомбинации (физически или виртуально) двух или более нуклеиновых кислот (или строки символов), например, искусственным и опционно рекурсивным образом. Обычно в процессах перетасовки используются одна или несколько стадий скрининга для идентификации представляющих интерес нуклеиновых кислот; этот этап скрининга можно проводить до или после любого этапа рекомбинации. В некоторых (но не во всех) вариантах осуществления изобретения перетасовка желательно выполнить в несколько циклов рекомбинации перед выбором, чтобы увеличить разнообразие пула, подлежащего скринингу. Общий процесс рекомбинации и отбора необязательно повторяется рекурсивно. В зависимости от контекста перетасовка может относиться к общему процессу рекомбинации и отбора или, альтернативно, может просто относиться к рекомбинационным частям всего процесса.

Используемые в предлагаемом изобретении термины "идентичность последовательностей" или "идентичность" в контексте двух полинуклеотидных или полипептидных последовательностей относятся к остаткам в двух последовательностях, которые являются одинаковыми при выравнивании для максимального соответствия в указанном окне сравнения. Когда процент идентичности последовательностей используется по отношению к белкам, признается, что положения остатков, которые не идентичны, часто отличаются консервативными аминокислотными заменами, когда аминокислотные остатки заменены другими аминокислотными остатками с аналогичными химическими свойствами (например, заряд или гидрофобность) и, следовательно, не меняют функциональных свойств молекулы. Когда последовательности различаются консервативными заменами, процент идентичности последовательностей может быть увеличен для корректировки консервативного характера замены. Считается, что последовательности, которые отличаются такими консервативными заменами, имеют "подобные последовательности" или "подобие". Способ выполнения этой регулировки хорошо известен специалистам в данной области. Обычно это включает оценку консервативной замены как частичного, а не полного несоответствия, тем самым увеличивая процентную идентичность последовательностей. Таким образом, например, когда идентичной аминокислоте присваивается оценка 1, а неконсервативной замене присваивается оценка ноль, консервативной замене присваивается оценка от нуля до 1. Вычисляется оценка консервативных замен, например, по реализованной программе PC/GENE (Intelligenetics, Mountain View, California).

Используемый в предлагаемом изобретении термин "процент идентичности последовательностей" означает значение, определенное путем сравнения двух оптимально выровненных последовательностей в окне сравнения, при этом часть полинуклеотидной последовательности в окне сравнения может содержать вставки или делеции (т.е. пробелы) по сравнению с эталонной последовательностью (которая не содержит вставок или делеций) для оптимального выравнивания двух последовательностей. Процент рассчитывается путем определения количества положений, в которых идентичное основание нуклеиновой кислоты или аминокислотный остаток встречается в обеих последовательностях, чтобы получить количество совпадающих положений, путем деления количества совпадающих положений на общее количество положений в окне сравнения и умножения результата на 100, чтобы получить процент идентичности последовательностей.

Если не указано иное, значения идентичности/подобия последовательностей, представленные в настоящем изобретении, относятся к значению, полученному с использованием GAP Версия 10 с использованием следующих параметров: % идентичности и % подобия для нуклеотидной последовательности с использованием GAP Weight 50 и Length Weight 3 и nwsgapdna.cmp скоринговой матрицы; % идентичности и % подобия для аминокислотной последовательности с использованием GAP Weight 8 и Length Weight 2 и скоринговой матрицы BLOSUM 62 или любой эквивалентной программы. Под "эквивалентной программой" подразумевается любая программа сравнения последовательностей, которая для любых двух рассматриваемых последовательностей генерирует выравнивание, имеющее идентичные совпадения нуклеотидных или аминокислотных остатков и идентичный процент идентичности последовательностей по сравнению с соответствующим выравниванием, созданным GAP Версией 10.

Две последовательности "оптимально выровнены", когда они выровнены для оценки сходства с использованием определенной матрицы аминокислотных замен (например, BLOSUM 62), штрафа за существование пробела и штрафа за удлинение пробела, чтобы достичь наивысшего возможного результата для этой пары последовательностей. Матрицы аминокислотных замен и их использование для количественной оценки сходства между двумя последовательностями хорошо известны в данной области техники и описаны, например, у Dayhoff и др. (1978) "Модель эволюционного изменения белков". В "Атласе последовательности и структуры белков" Том 5, Прил. 3 (под ред. М. О. Дэйхоффа), сс. 345-352. *Natl. Biomed. Res. Found.*, Вашингтон, округ Колумбия, и Henikoff и др. (1992) *Proc. Natl. Acad. Sci. США* № 89: сс. 10915-10919. Матрица BLOSUM 62 часто используется в качестве скоринговой матрицы замен по умолчанию в протоколах выравнивания последовательностей. Штраф за существование пробела накладывается за введение одного аминокислотного пробела в одну из выровненных последовательностей, а штраф за удлинение пробела накладывается за каждую дополнительную пустую аминокислотную позицию, вставленную в уже открытый пробел. Выравнивание определяется положениями аминокислот каждой последовательности, в которых выравнивание начинается и заканчивается, и, необязательно, вставкой пробела или множества пробелов в одну или обе последовательности, чтобы получить максимально возможную оценку. Хотя оптимальное выравнивание и подсчет очков можно выполнить вручную, этот процесс облегчается использованием алгоритма выравнивания, реализованного на компьютере, например BLAST 2.0 с промежутками, описанного Altschul и др. (1997) *Nucleic Acids Res.* № 25: сс. 3389-3402 и доступен для общественности на веб-сайте Национального центра биотехнологической информации (www.ncbi.nlm.nih.gov). Оптимальные сопоставления, включая множественные сопоставления, могут быть получены с использованием, например, PSI-BLAST, доступного на www.ncbi.nlm.nih.gov и описанного Altschul и др. (1997) *Nucleic Acids Res.* № 25: сс. 3389-3402.

Что касается аминокислотной последовательности, которая оптимально выровнена с эталонной последовательностью, в ней аминокислотный остаток "соответствует" положению в эталонной последовательности, с которой этот остаток спарен при выравнивании. "Положение" обозначается числом, которое последовательно идентифицирует каждую аминокислоту в эталонной последовательности на основании ее положения относительно N-конца. Из-за делеций, вставок, усечений, слияний и т.д., которые необходимо учитывать при определении оптимального выравнивания, в целом количество аминокислотных остатков в тестируемой последовательности, определяемое простым подсчетом от N-конца, не обязательно будет то же, что и номер соответствующей позиции в эталонной последовательности. Например, в случае, когда есть делеция в выровненной тестируемой последовательности, то в ней не будет аминокислоты, которая соответствует положению в эталонной последовательности на сайте делеций. Если имеется вставка в выровненной эталонной последовательности, эта вставка не будет соответствовать какому-либо положению аминокислоты в эталонной последовательности. В случае усечения или слияния могут быть отрезки аминокислот либо в эталонной, либо в выровненной последовательности, которые не соответствуют какой-либо аминокислоте в соответствующей последовательности.

V. Антитела.

Антитела к полипептидам RGN или рибонуклеопротеинам, содержащим полипептиды RGN, описанные в предлагаемом изобретении, включая те, которые имеют аминокислотную последовательность, указанную как SEQ ID NO: 1, 11, 19, 27, 36, 45 или 54, или их активные варианты или фрагменты, также включены. Способы получения антител хорошо известны в данной области см., например, Harlow и Lane (1988) *Antibodies: A Laboratory Manual*, Издательство Cold Spring Harbor Laboratory, Издательство Cold Spring Harbor, Нью-Йорк и патент США № 4 196265). Эти антитела можно использовать в наборах для обнаружения и выделения полипептидов или рибонуклеопротеинов RGN. Таким образом, в данном описании представлены наборы, содержащие антитела, которые специфически связываются с полипептидами или рибонуклеопротеинами, описанными в данном документе, включая, например, полипептиды, имеющие последовательность SEQ ID NO: 1, 11, 19, 27, 36, 45 или 54.

VI. Системы и комплексы рибонуклеопротеинов для связывания целевой представляющей интерес последовательности и способы их создания.

В предлагаемом описании изобретения представлена система для связывания целевой представляющей интерес последовательности-мишени, где система содержит по меньшей мере одну управляющую РНК или нуклеотидную последовательность, кодирующую ее, и по меньшей мере одну РНК-управляемую нуклеазу или нуклеотидную последовательность, кодирующую ее. Управляющая РНК гибридизуется с целевой представляющей интерес последовательностью, а также образует комплекс с полипептидом RGN, тем самым направляя полипептид RGN на связывание с целевой последовательностью. В некоторых вариантах осуществления изобретения RGN содержит аминокислотную последовательность SEQ ID NO: 1, 11, 19, 27, 36, 45 или 54 или ее активный вариант или фрагмент. В различных вариантах осуществления изобретения управляющая РНК содержит последовательность CRISPR повторов, содержащую нуклеотидную последовательность SEQ ID NO: 2, 12, 20, 28, 37, 46 или 55, или ее активный вариант или фрагмент. В конкретных вариантах осуществления изобретения управляющая РНК включает *trac*РНК, содержащую нуклеотидную последовательность SEQ ID NO: 3, 13, 21, 29, 38, 47 или 56, или ее активный вариант или фрагмент. Управляющая РНК системы может быть одиночной управляющей РНК

или двойной управляющей РНК. В конкретных вариантах осуществления изобретения система содержит РНК-управляемую нуклеазу, которая является гетерологичной по отношению к управляющей РНК, при этом RGN и gРНК не образуют комплексов в природе.

Система связывания целевой представляющей интерес последовательности, представленная в настоящем изобретении, может представлять собой комплекс рибонуклеопротеина, который состоит по меньшей мере из одной молекулы РНК, связанной по меньшей мере с одним белком. Комплексы рибонуклеопротеидов, представленные в настоящем документе, содержат по меньшей мере одну управляющую РНК в качестве компонента РНК и управляемую РНК нуклеазу, в качестве белкового компонента. Такие рибонуклеопротеиновые комплексы могут быть выделены из клетки или организма, которые естественным образом экспрессируют полипептид RGN, и были сконструированы для экспрессии конкретной управляющей РНК, которая специфична для целевой представляющей интерес последовательности. Кроме того, рибонуклеопротеиновый комплекс можно выделить путем очистки из клетки или организма, трансформированного полинуклеотидами, кодирующими полипептид RGN и управляющую РНК, и культивировать в условиях, допускающих экспрессию полипептида RGN и управляющей РНК. Таким образом, предложены способы получения полипептида RGN или комплекса рибонуклеопротеина RGN. Такие способы включают культивирование клетки, содержащей нуклеотидную последовательность, кодирующую полипептид RGN, и в некоторых вариантах осуществления изобретения нуклеотидную последовательность, кодирующую управляющую РНК, в условиях, в которых экспрессируется полипептид RGN (и в некоторых вариантах осуществления изобретения управляющую РНК). Затем полипептид RGN или рибонуклеопротеин RGN можно очистить из лизата культивируемых клеток.

Способы очистки полипептида RGN или комплекса рибонуклеопротеина RGN из лизата биологического образца известны в данной области (например, вытеснительная и/или аффинная хроматография, 2D-PAGE, HPLC, обращенно-фазовая хроматография, иммунопреципитация). В конкретных способах полипептид RGN получают рекомбинантно, и он содержит метку очистки, чтобы способствовать его очистке, включая, помимо прочего, глутатион-S-трансферазу (GST), хитинсвязывающий белок (CBP), мальтозосвязывающий белок, тиоредоксин (TRX), поли (NANP), тег тандемной аффинной очистки (TAP), мус, 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, 6×His, 10×His, белок-носитель карбоксила биотина (BCCP) и кальмодулин. Обычно меченый полипептид RGN или комплекс рибонуклеопротеина RGN очищают с использованием аффинной хроматографии с иммобилизованным металлом. Следует понимать, что можно использовать другие аналогичные методы, известные в данной области, включая другие формы хроматографии или, например, иммунопреципитацию, либо по отдельности, либо в комбинации.

"Выделенный" или "очищенный" полипептид или его биологически активная часть по существу не содержат компонентов, которые обычно сопровождают полипептид или взаимодействуют с ним, как это обнаруживается в его естественной среде. Таким образом, выделенный или очищенный полипептид практически не содержит другого клеточного материала или культуральной среды при получении рекомбинантными методами или практически не содержит химических предшественников или других химических веществ при химическом синтезе. Белок, который практически не содержит клеточного материала, включает препараты белка, содержащие менее примерно 30%, 20%, 10%, 5% или 1% (на сухой вес) загрязняющего белка. Когда белок, описанный в настоящем изобретении, или его биологически активная часть продуцируется рекомбинантно, оптимально культуральная среда представляет примерно менее, чем 30%, 20%, 10%, 5% или 1% (на сухой вес) химических предшественников или небелковых веществ, или представляющих интерес химических веществ.

Конкретные способы, представленные в данном документе для связывания и/или расщепления представляющей интерес целевой последовательности, включают использование собранного *in vitro* комплекса рибонуклеопротеина RGN. Сборка рибонуклеопротеинового комплекса RGN *in vitro* может быть осуществлена с использованием любого метода, известного в данной области, в котором полипептид RGN контактирует с управляющей РНК в условиях, позволяющих полипептиду RGN связываться с управляющей РНК. Используемые здесь термины "контактировать", "контактирующие", "контактированные" относятся к размещению компонентов желаемой реакции вместе в условиях, подходящих для проведения желаемой реакции. Полипептид RGN может быть очищен из биологического образца, клеточного лизата или культуральной среды, полученной путем трансляции *in vitro* или синтезированной химическим путем. Управляющая РНК может быть очищена из биологического образца, клеточного лизата или культуральной среды, транскрибирована *in vitro* или химически синтезирована. Полипептид RGN и управляющая РНК могут быть приведены в контакт в растворе (например, забуференном физиологическом растворе), чтобы обеспечить сборку RGN рибонуклеопротеинового комплекса *in vitro*.

VII. Способы связывания, расщепления или изменения целевой последовательности.

В предлагаемом описании изобретения представлены способы связывания, расщепления и/или модификации представляющей интерес целевой нуклеотидной последовательности. Способы включают доставку системы, содержащей по меньшей мере одну управляющую РНК или полинуклеотид, кодирующий ее, и по меньшей мере один полипептид RGN или полинуклеотид, кодирующий то же самое в целевой последовательности или клетке, органелле или эмбрионе, содержащих целевую последователь-

ность. В некоторых вариантах осуществления изобретения RGN содержит аминокислотную последовательность SEQ ID NO: 1, 11, 19, 27, 36, 45 или 54 или ее активный вариант, или фрагмент. В различных вариантах осуществления изобретения управляющая РНК содержит последовательность CRISPR повторов, содержащую нуклеотидную последовательность SEQ ID NO: 2, 12, 20, 28, 37, 46 или 55, или ее активный вариант, или фрагмент. В конкретных вариантах осуществления изобретения управляющая РНК содержит *tracr*РНК, содержащую нуклеотидную последовательность SEQ ID NO: 3, 13, 21, 29, 38, 47 или 56, или ее активный вариант, или фрагмент. Управляющая РНК системы может быть одиночной или двойной. RGN система может содержать неактивную нуклеазу, обладать никазой активностью или может быть гибридным полипептидом. В некоторых вариантах осуществления изобретения слитый полипептид содержит редактирующий основание полипептид, например цитидиндезаминазу или аденозиндезаминазу. В конкретных вариантах осуществления изобретения RGN и/или управляющая РНК является гетерологичной клетке, органелле или эмбриону, в которые введены RGN и/или управляющая РНК (или полинуклеотид(ы)), кодирующие по меньшей мере одну из RGN и управляющей РНК).

В тех вариантах осуществления изобретения, в которых способ включает доставку полинуклеотида, кодирующего управляющую РНК и/или полипептид RGN, в клетку или эмбрион затем можно культивировать в условиях, в которых экспрессируется управляющая РНК и/или полипептид RGN. В различных вариантах осуществления изобретения способ включает контактирование целевой последовательности с рибонуклеопротеидным комплексом RGN. Рибонуклеопротеиновый комплекс RGN может включать RGN с неактивной нуклеазой или обладает никазой активностью. В некоторых вариантах осуществления изобретения RGN рибонуклеопротеинового комплекса представляет собой слитый полипептид, содержащий полипептид, редактирующий основание. В некоторых вариантах осуществления изобретения способ включает введение в клетку, органеллу или эмбрион комплекса рибонуклеопротеина RGN, содержащего целевую последовательность. Рибонуклеопротеидный комплекс RGN может быть очищен из биологического образца, получен рекомбинантно и впоследствии очищен или собран *in vitro*, как описано в данном изобретении. В тех вариантах осуществления изобретения, где рибонуклеопротеиновый комплекс RGN, который контактирует с целевой последовательностью или клеточной органеллой, или эмбрионом, был собран *in vitro*, способ может дополнительно включать сборку комплекса *in vitro* перед контактом с целевой последовательностью, клеткой, органеллой, или эмбрионом.

Очищенный или собранный *in vitro* рибонуклеопротеидный комплекс RGN может быть введен в клетку, органеллу или эмбрион с использованием любого метода, известного в данной области, включая электропорацию, но, не ограничиваясь ею. Кроме того, полипептид и/или полинуклеотид RGN, кодирующий или содержащий управляющую РНК, может быть введен в клетку, органеллу или эмбрион с использованием любого метода, известного в данной области (например, методом электропорации).

При доставке или контакте с целевой последовательностью или клеткой, органеллой или эмбрионом, содержащими целевую последовательность, управляющая РНК направляет RGN на связывание с целевой последовательностью специфическим для последовательности образом. В тех вариантах осуществления изобретения, где RGN обладает нуклеазной активностью, полипептид RGN расщепляет интересующую целевую последовательность при связывании. Целевая последовательность впоследствии может быть модифицирована с помощью механизмов эндогенной репарации, таких как негомологичное соединение концов или гомологично направленная репарация предоставленным донорным полинуклеотидом.

Способы измерения связывания полипептида RGN с целевой последовательностью известны в данной области и включают анализы иммунопреципитации хроматина, анализы сдвига подвижности геля, анализы тесты на связывание ДНК, анализы репортеров, анализы захвата и обнаружения на микропланшетах. Аналогичным образом, способы измерения расщепления или модификации целевой последовательности известны в данной области и включают анализы расщепления *in vitro* или *in vivo*, в которых расщепление подтверждается с помощью ПЦР, секвенирования или гель-электрофорезом с или без присоединения соответствующей метки (например, радиоизотопа, флуоресцентного вещества) к целевой последовательности для облегчения обнаружения продуктов разложения. В качестве альтернативы может быть использован анализ реакции экспоненциальной амплификации, инициируемой расщеплением (NTEXPAR) (см., например, Zhang и др. (2016) *Chem. Sci.* № 7: сс. 4951-4957). Расщепление *in vivo* можно оценить с помощью анализа Surveyor (Guschin и др. (2010) *Methods Mol Biol* № 649: сс. 247-256).

В некоторых вариантах осуществления изобретения способы включают использование одного типа RGN в комплексе с более чем одной управляющей РНК. Более чем одна управляющая РНК может быть нацелена на разные области одного гена или на несколько генов.

В тех вариантах осуществления изобретения, в которых донорный полинуклеотид не предусмотрен, двухцепочечный разрыв, введенный полипептидом RGN, может быть восстановлен с помощью процесса репарации негомологичным концевым соединением (NHEJ). Из-за склонности к ошибкам NHEJ восстановление двухцепочечного разрыва может привести к модификации целевой последовательности. В данном контексте "модификация" в отношении молекулы нуклеиновой кислоты относится к изменению нуклеотидной последовательности молекулы нуклеиновой кислоты, которое может представлять собой делецию, вставку или замену одного или нескольких нуклеотидов или их комбинацию. Модификация целевой последовательности может привести к экспрессии измененного белкового продукта или инактив-

вазии кодирующей последовательности.

В тех вариантах осуществления изобретения, где присутствует донорный полинуклеотид, донорная последовательность в донорном полинуклеотиде может быть интегрирована в целевую нуклеотидную последовательность или заменена ей в ходе репарации введенного двухцепочечного разрыва, что приводит к введению экзогенной донорной последовательности. Таким образом, донорный полинуклеотид содержит донорную последовательность, которую желательно ввести в интересующую целевую последовательность. В некоторых вариантах осуществления изобретения донорная последовательность изменяет исходную целевую нуклеотидную последовательность так, что новая интегрированная донорная последовательность не будет распознаваться и расщепляться RGN. Интеграция донорной последовательности может быть усилена включением в донорный полинуклеотид фланкирующих последовательностей, которые в значительной степени идентичны последовательностям, фланкирующим целевую нуклеотидную последовательность, что делает возможным процесс восстановления, гомологично направленным. В тех вариантах осуществления изобретения, в которых полипептид RGN вводит двухцепочечные разрывы в шахматном порядке, донорный полинуклеотид может содержать донорскую последовательность, фланкированную совместимыми выступами, что делает возможным прямое лигирование донорной последовательности с расщепленной целевой нуклеотидной последовательностью, содержащей выступы, в процессе негомологичного восстановления при репарации двухцепочечного разрыва.

В тех вариантах осуществления изобретения, в которых способ включает использование RGN, которая представляет собой никазу (т.е. она способна расщеплять только одну цепь двухцепочечного полинуклеотида), способ может включать введение двух никакз RGN, которые нацелены на идентичные или перекрывающиеся целевые последовательности и расщепляют разные цепи полинуклеотида. Например, никакза RGN, которая расщепляет только положительную (+) цепь двухцепочечного полинуклеотида, может быть введена вместе со второй никакзой RGN, которая расщепляет только отрицательную (-) цепь двухцепочечного полинуклеотида.

В различных вариантах осуществления изобретения предложен способ связывания целевой нуклеотидной последовательности и обнаружения целевой последовательности, при этом способ включает введение в клетку, органеллу или эмбрион по меньшей мере одной управляющей РНК или полинуклеотида, кодирующего ее, и по меньшей мере одной RGN, полипептида или полинуклеотида, кодирующего ее, экспрессируя управляющую РНК и/или полипептид RGN (если введены кодирующие последовательности), где полипептид RGN представляет собой RGN с неактивной нуклеазой и дополнительно содержит детектируемую метку, а способ дополнительно включает обнаружение детектируемой метки. Обнаруживаемая метка может быть слита с RGN в виде слитого белка (например, флуоресцентного белка) или может представлять собой небольшую молекулу, конъюгированную или включенную в полипептид RGN, которую можно обнаружить визуальными или другими способами.

В настоящем описании изобретения также представлены способы модуляции экспрессии целевой последовательности или интересующего гена при регуляции целевой последовательности. Способы включают введение в клетку, органеллу или эмбрион по меньшей мере одной управляющей РНК или полинуклеотида, кодирующего ее, и по меньшей мере одного полипептида RGN или полинуклеотида, кодирующего ее, экспрессируя управляющую РНК и/или полипептид RGN, описанные в предлагаемом изобретении, (если введены кодирующие последовательности), где полипептид RGN представляет собой RGN с неактивной нуклеазой. В некоторых из этих вариантов осуществления изобретения RGN с неактивной нуклеазой представляет собой слитый белок, содержащий домен модулятора экспрессии (т.е. домен эпигенетической модификации, домен активации транскрипции или домен репрессора транскрипции), как описано в предлагаемом изобретении.

В предлагаемом описании изобретения также представлены способы связывания и/или модификации интересующей целевой нуклеотидной последовательности. Способы включают доставку системы, содержащей по меньшей мере одну управляющую РНК или полинуклеотид, кодирующий ее, и по меньшей мере один слитый полипептид, который содержит RGN, предлагаемую в настоящем изобретении, и полипептид, редактирующий основание, например цитидиндезаминазу или аденозиндезаминазу, или полинуклеотид, кодирующий слитый полипептид с целевой последовательностью или клеткой, органеллой или эмбрионом, содержащим целевую последовательность.

Обычный специалист в данной области поймет, что любой из описанных здесь способов может быть использован для нацеливания на единственную целевую последовательность или множественные целевые последовательности. Таким образом, способы включают использование одного полипептида RGN в комбинации с множеством отдельных управляющих РНК, которые могут нацеливаться на множество различных последовательностей в пределах одного гена и/или нескольких генов. В настоящее изобретение также включены способы, в которых несколько отдельных управляющих РНК вводятся в комбинации с несколькими отдельными полипептидами RGN. Эти управляющие РНК и управляющие полипептидные системы РНК/RGN могут нацеливаться на несколько различных последовательностей в пределах одного гена и/или нескольких генов.

В одном аспекте изобретение представляет наборы, содержащие любой один или несколько элементов, описанных в приведенных выше способах и композициях. В некоторых вариантах осуществле-

ния изобретения набор включает векторную систему и инструкции по использованию набора. В некоторых вариантах осуществления изобретения векторная система содержит (а) первый регуляторный элемент, функционально связанный с последовательностью *tracr*, и один или несколько сайтов вставки для вставки управляющей последовательности перед последовательностью *tracr*, причем при экспрессии управляющая последовательность направляет специфическое для последовательности связывание комплекса CRISPR с целевой последовательностью в эукариотической клетке, при этом комплекс CRISPR включает фермент CRIS PR в комплексе с (1) управляющей последовательностью, которая гибридизуется с целевой последовательностью и (2) последовательностью *tracr*, которая гибридизована с *tracr* последовательностью; и/или (б) второй регуляторный элемент, функционально связанный с последовательностью кодирующей фермент, которая кодирует указанный фермент CRISPR, содержащий последовательность ядерной локализации. Элементы могут быть предоставлены индивидуально или в комбинациях в любом подходящем контейнере, таком как пузырек, флакон или пробирка.

В некоторых вариантах осуществления изобретения набор включает инструкции на одном или нескольких языках. В некоторых вариантах осуществления изобретения набор включает один или несколько реагентов для применения в способе с использованием одного или нескольких элементов, описанных в данном документе. Реагенты могут быть предоставлены в любом подходящем контейнере. Например, набор может включать один или несколько буферов для реакции или хранения. Реагенты могут быть предоставлены в форме, пригодной для использования в конкретном анализе, или в форме, которая требует добавления одного или нескольких других компонентов перед использованием (например, в виде концентрата или в лиофилизированной форме). Буфер может быть любым буфером, включая, помимо прочего, буфер карбоната натрия, буфер бикарбоната натрия, боратный буфер, буфер Tris, буфер MOPS, буфер HEPES и их комбинации. В некоторых вариантах осуществления изобретения буфер является щелочным. В некоторых вариантах осуществления изобретения буфер имеет pH примерно от 7 до 10.

В некоторых вариантах осуществления изобретения набор содержит один или несколько олигонуклеотидов, соответствующих управляющей последовательности для вставки в вектор, для того чтобы функционально связать управляющую последовательность и регуляторный элемент. В некоторых вариантах осуществления изобретения набор включает полинуклеотид-матрицу для гомологичной рекомбинации. В одном аспекте изобретение предоставляет способы использования одного или нескольких элементов системы CRISPR. Комплекс CRISPR, предлагаемый в настоящем изобретении, обеспечивает эффективное средство модификации целевого полинуклеотида. Комплекс CRISPR согласно в настоящему изобретению имеет широкое применение, включая модификацию (например, удаление, вставку, транслокацию, инактивацию, активацию) целевого полинуклеотида во множестве типов клеток. Как таковой комплекс CRISPR, предлагаемый в настоящем изобретении, имеет широкий спектр применения, например, в генной терапии, скрининге лекарственных средств, диагностике заболеваний и прогнозировании. Типичный комплекс CRISPR включает фермент CRISPR в комплексе с управляющей последовательностью, гибридизированной с целевой последовательностью в целевой полинуклеотиде.

VIII. Целевые полинуклеотиды.

В одном аспекте изобретение относится к способам модификации целевого полинуклеотида в эукариотической клетке, которая может быть *in vivo*, *ex vivo* или *in vitro*. В некоторых вариантах осуществления изобретения способ предполагает взятие образцов клетки или популяции клеток человека или животного, не являющегося человеком, или растения (включая микроводоросли), и модификацию клетки или клеток. Культивирование может происходить на любой стадии *ex vivo*. Клетку или клетки можно даже повторно ввести в растение (включая микроводоросли) или животное, не являющееся человеком.

Используя естественную изменчивость, селекционеры комбинируют наиболее полезные гены нужных качеств, таких как урожайность, качество, однородность, выносливость и устойчивость к вредителям. Эти нужные качества также включают рост, предпочтительную длину светового дня, требования к температуре, дату начала цветения или репродуктивного развития, содержание жирных кислот, устойчивость к насекомым, устойчивость к болезням, устойчивость к нематодам, устойчивость к грибам, устойчивость к гербицидам, устойчивость к различным факторам окружающей среды, включая засуху, жару, влажность, холод, ветер и неблагоприятные почвенные условия, включая высокую засоленность. Источники этих полезных генов включают местные или чужеродные сорта, наследственные сорта, диких родственников растений и индуцированные мутации, например, обработку растительного материала мутагенными агентами. С помощью предлагаемого изобретения селекционерам растений предоставляется новый инструмент для индукции мутаций. Соответственно, специалист в данной области может анализировать геном на предмет источников полезных генов, а в сортах, имеющих нужные характеристики или признаки, использовать настоящее изобретение для индукции количества полезных генов с большей точностью, чем предыдущие мутагенные агенты, и, следовательно, ускорять и улучшать программы селекции растений.

Целевой полинуклеотид системы RGN может быть любым полинуклеотидом, эндогенным или экзогенным по отношению к эукариотической клетке. Например, целевой полинуклеотид может быть полинуклеотидом, находящимся в ядре эукариотической клетки. Целевой полинуклеотид может представлять собой последовательность, кодирующую продукт гена (например, белок), или не кодирующую последо-

вательность (например, регуляторный полинуклеотид или избыточную ДНК). Не ограничиваясь теорией, полагают, что целевая последовательность должна быть связана с PAM (мотив, примыкающий к протоспейсеру), то есть короткой последовательностью, распознаваемой комплексом CRISPR. Требования к точной последовательности и длине для PAM различаются в зависимости от используемого фермента CRISPR, но PAM обычно представляет собой последовательности из 2-5 пар оснований, смежные с протоспейсером (то есть с целевой последовательностью).

Целевой полинуклеотид комплекса CRISPR может включать ряд связанных с заболеванием генов и полинуклеотидов, а также гены и полинуклеотиды, связанные с биохимическим путем передачи сигналов. Примеры целевых полинуклеотидов включают последовательность, связанную с сигнальным биохимическим путем, например, ген или полинуклеотид, связанный с сигнальным биохимическим путем. Примеры целевых полинуклеотидов включают ген или полинуклеотид, связанный с заболеванием. "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 пораженных болезнью тканей, по сравнению со здоровыми контрольными тканями или клетками. Это может быть ген, который экспрессируется на аномально высоком уровне; это может быть ген, который экспрессирован на аномально низком уровне, где измененная экспрессия коррелирует с возникновением и/или прогрессированием заболевания. Связанный с заболеванием ген также относится к гену, обладающему мутацией(-ями) или генетической вариацией, которые непосредственно ответственны или находятся в неравновесном сцеплении с геном(-ами), который ответственен за этиологию заболевания (например, причинную мутацию). Транскрибированные или транслированные продукты могут быть известными или неизвестными, а также могут находиться на нормальном или ненормальном уровне. Примеры связанных с заболеванием генов и полинуклеотидов доступны в Институте генетической медицины МакКьюзика-Натанса Университета Джона Хопкинса (Балтимор, штат Мэриленд) и в Национальном центре биотехнологической информации Национальной медицинской библиотеки (Бетесда, штат Мэриленд), а также доступны в World Wide Web.

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 устранения причинной мутации. Один из подходов заключается в создании слитого белка между RGN (предпочтительно неактивным или никакным вариантом RGN) и ферментом, редактирующим основание, или активным доменом фермента, редактирующего основание, например, цитидиндезаминазы или редактора оснований аденозиндезаминазы (патент США № 9840699, включенный в данный документ в качестве ссылки). В некоторых вариантах осуществления изобретения способы включают молекулу ДНК, контактирующую с (а) гибридным белком, содержащим RGN, предлагаемую в настоящем изобретении, и полипептид, редактирующий основание, такой как дезаминаза; и (б) gPНК, нацеленную на слитый белок (а) на целевую нуклеотидную последовательность цепи ДНК, где молекула ДНК контактирует со слитым белком и gPНК в эффективном количестве и в условиях, подходящих для дезаминирования нуклеотидного основания. В некоторых вариантах осуществления изобретения целевая последовательность ДНК включает последовательность, связанную с заболеванием или нарушением, и где дезаминирование нуклеотидного основания приводит к последовательности, которая не связана с заболеванием или нарушением. В некоторых вариантах осуществления изобретения целевая последовательность ДНК находится в аллеле культурного растения, где конкретная аллель интересующего признака приводит к получению растения меньшей агрономической ценности. В результате дезаминирования нуклеотидного основания образуется аллель, улучшающий признак и повышающий агрономическую ценность растения.

В некоторых вариантах осуществления изобретения последовательность ДНК включает точечную мутацию T→C или A→G, связанную с заболеванием или нарушением, и при этом дезаминирование мутантного основания C или G приводит к последовательности, которая не связана с заболеванием или нарушением. В некоторых вариантах осуществления изобретения дезаминирование исправляет точечную мутацию в последовательности, связанной с заболеванием или нарушением.

В некоторых вариантах осуществления 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 кодируемого белка. В некоторых вариантах осуществления изобретения контактирование осуществляется *in vivo* у субъекта, подверженного заболеванию или расстройству, имеющего это заболевание, или у субъекта, у которого оно диагностировано. В некоторых вариантах реализации изобретения заболевание или нарушение представляет собой заболевание, связанное с точечной мутацией или одноосновной мутацией в геноме. В некоторых вариантах осуществления изобретения заболевание представляет собой генетическое заболевание, рак, нарушение обмена веществ или лизосомную болезнь.

Дополнительные примеры локусов, которые являются причиной определенных генетических заболеваний, в частности локусов, на которые можно легко нацелиться с помощью RGN или слитых белков редактор оснований-RGN, предлагаемых в настоящем изобретении, можно найти в примере 9 и соответствующей табл. 12.

Синдром Гурлера.

Примером генетически наследуемого заболевания, которое может быть исправлено с использованием подхода, основанного на использовании слитого белка RGN-редактор оснований, предлагаемого в настоящем изобретении, является синдром Гурлера. Синдром Гурлера, также известный как MPS-1, является результатом дефицита α -L-идуронидазы (IDUA), что приводит к лизосомной болезни, характеризующейся на молекулярном уровне накоплением дерматансульфата и гепарансульфата в лизосомах. Это заболевание обычно является наследственным генетическим заболеванием, вызванным мутациями гена IDUA, кодирующего α -L-идуронидазу. Распространенными мутациями IDUA являются W402X и Q70X, обе нонсенс-мутации, приводящие к преждевременному прекращению трансляции. Такие мутации хорошо поддаются лечению с помощью подходов точного редактирования генома (PGE), поскольку реверсия одного нуклеотида, например, с помощью подхода редактирования оснований, восстановит кодирующую последовательность дикого типа и приведет к экспрессии белка, контролируемой эндогенными регуляторными механизмами генетического локуса. Кроме того, поскольку известно, что гетерозиготы бессимптомны, терапия PGE, направленная на одну из этих мутаций, будет полезна для значительной части пациентов с этим заболеванием, так как необходимо исправить только одну из мутировавших аллелей (Bunge и др. (1994), *Hum. Mol. Genet.* № 3 (6): сс. 861-866), включенная в данный документ в качестве ссылки).

Современные методы лечения синдрома Гурлера включают заместительную ферментную терапию и трансплантацию костного мозга (Vellodi и др. (1997) *Arch. Dis. Child.* № 76 (2): сс. 92-99; Peters и др. (1998) *Blood* № 91 (7): сс. 2601-2608, включенные в данный документ в качестве ссылки). Хотя заместительная ферментативная терапия оказала сильное влияние на выживаемость и качество жизни пациентов с синдромом Гурлера, этот подход требует дорогостоящих и длительных еженедельных инфузий. Дополнительные подходы включают доставку гена IDUA в вектор экспрессии или встраивание гена в локус высокой экспрессией, такой как локус сывороточного альбумина (патент США № 9956247, включенный в настоящее описание изобретения в качестве ссылки). Однако эти подходы не восстанавливают исходный локус IDUA до правильной кодирующей последовательности. Стратегия редактирования генома будет иметь ряд преимуществ, в первую очередь то, что регуляция экспрессии генов будет контролироваться естественными механизмами, присутствующими у здоровых людей. Кроме того, использование базового редактирования не требует разрыва двухцепочечной ДНК, что может привести к большим хромосомным перестройкам, гибели клеток или онкогенности из-за нарушения механизмов подавления опухоли. Возможное описание способа исправления причинной мутации этого заболевания представлено в примере 10. Описанные способы являются примером общей стратегии, направленной на использование слитых белков RGN-редактор оснований, предложенной в настоящем изобретении, для нацеливания и коррекции определенных вызывающих заболевание мутаций в геноме человека. Следует принять во внимание, что аналогичные подходы к целевым заболеваниям, таких как описаны в табл. 12, также могут быть применены. Также будет понятно, что аналогичные подходы к нацеливанию на вызывающие заболевание мутации у других видов, особенно обычных домашних животных или домашнего скота, также могут быть применены с использованием RGN, предлагаемой в настоящем изобретении. Обычные домашние животные и домашний скот включают собак, кошек, лошадей, свиней, коров, овец, кур, ослов, змей, хорьков, рыбу, включая лосося, и креветок.

Атаксия Фридрейха.

RGN, предлагаемая в настоящем изобретении, также может быть полезной в терапевтических подходах для лечения заболеваний человека, где причинная мутация более сложная. Например, некоторые заболевания, такие как атаксия Фридрейха и болезнь Хантингтона, являются результатом значительного увеличения количества повторов трех нуклеотидного мотива в определенной области гена, что влияет на способность экспрессируемого белка функционировать или экспрессироваться. Атаксия Фридрейха (FRDA) - это аутосомно-рецессивное заболевание, приводящее к прогрессирующей дегенерации нервной ткани в спинном мозге. Снижение уровня белка фратаксина (FXN) в митохондриях вызывает окислительные повреждения и дефицит железа на клеточном уровне. Снижение экспрессии FXN было связано с экспансией триплета GAA в интроне 1 соматического и зародышевого гена FXN. У пациентов с FRDA повторение GAA часто состоит из более чем 70, иногда даже из более чем 1000 (чаще всего 600-900) триплетов, тогда как у здоровых людей имеется около 40 повторов или меньше (Pandolfo и др. (2012) *Handbook of Clinical Neurology* № 103: сс. 275-294; Campuzano и др. (1996) *Science* № 271: сс. 1423-1427; Pandolfo (2002) *Adv. Exp. Med. Biol.* № 516: сс. 99-118; все включены в предлагаемое описание изобретения в качестве ссылки).

Экспансия последовательности тринуклеотидного повтора, вызывающая атаксию Фридрейха (FRDA), происходит в определенном генетическом локусе внутри гена FXN, который называется областью нестабильности FRDA. РНК-управляемые нуклеазы (RGN) могут использоваться для удаления области нестабильности в клетках пациентов с FRDA. Этот подход требует использования 1) последовательности RGN и управляющей РНК, которая может быть запрограммирована для нацеливания на аллель в геноме человека; и 2) способа доставки RGN и управляющей последовательности. Многие нуклеазы,

используемые для редактирования генома, такие как обычно используемая нуклеаза Cas9 из *S. pyogenes* (SpCas9), слишком велики, чтобы их можно было упаковывать в аденоассоциированные вирусные (AAV) векторы, особенно с учетом длины гена SpCas9 и управляющей РНК в дополнение к другим генетическим элементам, необходимым для функциональных каскадов экспрессии. Это затрудняет подход с использованием SpCas9.

Компактные РНК-управляемые нуклеазы, предлагаемые в настоящем изобретении, в частности APG07433.1 и APG08290.1, уникально хорошо подходят для удаления области нестабильности FRDA. Каждая RGN имеет, как необходимое условие РНК, которая находится в непосредственной близости от области нестабильности FRDA. Кроме того, каждая из этих RGN может быть упакована в вектор AAV вместе с управляющей РНК. Упаковка двух управляющих РНК, вероятно, потребует второго вектора, но этот подход все же выгодно отличается от того, который требует более крупной нуклеазы, такой как SpCas9, которая может потребовать разделения последовательности белка между двумя векторами. Возможное описание способа исправления причинной мутации этого заболевания представлено в примере 11. Описанные способы охватывают стратегию с использованием RGN, предлагаемой в настоящем изобретении, в которой удаляется область геномной нестабильности. Такая стратегия применима к другим заболеваниям и расстройствам, имеющим аналогичную генетическую основу, таким как болезнь Хантингтона. Аналогичные стратегии с использованием RGN, предлагаемой в настоящем изобретении, также могут быть применимы к аналогичным заболеваниям и расстройствам у животных, не относящихся к человеку, имеющих агрономическое или экономическое значение, включая собак, кошек, лошадей, свиней, коров, овец, кур, ослов, змей, хорьков, рыб, включая лосось и креветки.

Гемоглобинопатии.

RGN, раскрытые в настоящем описании, также могут вводить разрушающие мутации, 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 серповидно-клеточную анемию и талассемии.

У взрослых людей гемоглобин представляет собой гетеротетрамер, содержащий две альфа (α)-подобные цепи глобина и две бета (β)-подобные цепи глобина и 4 гемовые группы. У взрослых тетрамер $\alpha_2\beta_2$ называется гемоглобином А (HbA) или взрослым гемоглобином. Обычно альфа- и бета-глобиновые цепи синтезируются в соотношении приблизительно 1:1, и это соотношение кажется критическим с точки зрения стабилизации гемоглобина и эритроцитов (RBC). У развивающегося плода образуется другая форма гемоглобина, гемоглобин плода (HbF), который имеет более высокое сродство связывания с кислородом, чем гемоглобин А, так что кислород может доставляться в систему ребенка через кровотоки матери. Гемоглобин плода также содержит две цепи α -глобина, но вместо цепей β -глобина взрослого человека он имеет две цепи гамма (γ)-глобина плода (т.е. гемоглобин плода $\alpha_2\gamma_2$). Регулирование перехода от продукции гамма-глобина к бета-глобину является довольно сложным и в первую очередь включает подавление транскрипции гамма-глобина с одновременным усилением транскрипции бета-глобина. Примерно на 30 неделе беременности синтез гамма-глобина у плода начинает падать, 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 и во взрослом возрасте (примерно 1-3% от общего гемоглобина). У большинства пациентов с гемоглобинопатией гены, кодирующие гамма-глобин, остаются, но 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) (GAG в GTG на уровне ДНК), где образующийся гемоглобин обозначается как "гемоглобин S" или "HbS". В условиях пониженного содержания кислорода молекулы HbS агрегируют и образуют волокнистые осадки. Эти агрегаты вызывают аномалию или "серповидность" эритроцитов, что приводит к потере гибкости клеток. Серповидные эритроциты больше не могут протискиваться в капиллярные русла и могут вызывать вазо-окклюзионный криз у пациентов с серповидными клетками. Кроме того, серповидные эритроциты более хрупкие, чем нормальные эритроциты, и имеют тенденцию к гемолизу, что в конечном итоге приводит к анемии у пациента.

Лечение и ведение пациентов с серповидно-клеточной анемией - это пожизненная проблема, включающая лечение антибиотиками, обезболивание и переливание крови во время острых приступов. Один из подходов - использование гидроксимочевины, которая частично оказывает свое действие за счет увеличения производства гамма-глобина. Однако долгосрочные побочные эффекты хронической терапии гидроксимочевинной все еще неизвестны, и лечение дает нежелательные побочные эффекты и может иметь различную эффективность у разных пациентов. Несмотря на повышение эффективности лечения серповидными клетками, ожидаемая продолжительность жизни пациентов все еще находится в диапазоне от середины до конца 50-ти лет, и связанные с этим заболевания оказывают сильное влияние на качество жизни пациента.

Талассемии (альфа-талассемия и бета-талассемия) также являются заболеваниями, связанными с гемоглобином, и обычно являются причиной снижения экспрессии цепей глобина. Это может происхо-

дуть из-за мутаций в регуляторных областях генов или из-за мутации в последовательности, кодирующей глобин, которая приводит к снижению экспрессии или снижению уровней функционального белка глобина. Лечение талассемии обычно включает переливание крови и хелатирование железа. Трансплантация костного мозга также используется для лечения людей с тяжелой формой талассемии, если можно найти подходящего донора, но эта процедура может иметь значительные риски.

Один из подходов, который был предложен для лечения как SCD, так и бета-талассемии, заключается в увеличении экспрессии гамма-глобина так, чтобы HbF функционально замещал аберрантный гемоглобин взрослого человека. Как упоминалось выше, лечение пациентов с SCD гидроксимочевинной считается успешным отчасти из-за ее влияния на увеличение экспрессии гамма-глобина (DeSimone (1982) Proc Nat'l Acad Sci USA № 79 (14): сс. 4428-31; Ley и др. (1982) N. Engl. J. Medicine, № 307: сс. 1469- 1475; Ley и др. (1983) Blood № 62: сс. 370-380; Constantoulakis и др. (1988) Blood № 72 (6): сс. 1961-1967, все включены в предлагаемое описание изобретения в качестве ссылки). Увеличение экспрессии HbF включает идентификацию генов, продукты которых играют роль в регуляции экспрессии гамма-глобина. Один из таких генов - BCL11A. BCL11A кодирует белок цинкового пальца, который экспрессируется во взрослых клетках-предшественниках эритроидов, и подавление его экспрессии приводит к увеличению экспрессии гамма-глобина (Sankaran и др. (2008) Science № 322: с. 1839, включено в предлагаемое описание изобретения в качестве ссылки). Было предложено использование ингибирующей РНК, нацеленной на ген BCL11A (например, патент США 2011/0182867, включенный в настоящее описание в качестве ссылки), но эта технология имеет несколько потенциальных недостатков, в том числе невозможность полного нокаута и доставки таких РНК может быть проблематичным, а РНК должны присутствовать постоянно, что требует многократного лечения в течение всей жизни.

RGN, предлагаемые в настоящем изобретении, можно использовать для нацеливания на область энхансера BCL11A, чтобы нарушить экспрессию BCL11A, тем самым увеличивая экспрессию гамма-глобина. Это целевое нарушение может быть достигнуто за счет негомологичного соединения концов (NHEJ), при котором RGN, предлагаемая в настоящем изобретении, нацеливается на конкретную последовательность в области энхансера BCL11A, делает двух-цепочечный разрыв, и аппарат клетки восстанавливает разрыв, обычно одновременно внося вредные мутации. Подобно тому, что описано для других мишеней заболевания, RGN, предлагаемые в настоящем изобретении, имеет преимущества перед другими известными RGN из-за их относительно небольшого размера, что позволяет упаковывать кассеты экспрессии для RGN и ее управляющей РНК в один вектор AAV для доставки *in vivo*. Возможное описание этого способа представлено в примере 12. Подобные стратегии с использованием RGN, предлагаемой в настоящем изобретении, также могут быть применимы к аналогичным заболеваниям и расстройствам, как у людей, так и у животных, не относящихся к человеку, которые имеют агрономическое или экономическое значение.

IX. Клетки, содержащие генетическую модификацию полинуклеотидов.

В настоящем изобретении представлены клетки и организмы, содержащие представляющую интерес целевую последовательность, которая была модифицирована с использованием процесса, опосредованного RGN, sgРНК и/или tracrРНК, как описано в предлагаемом изобретении. В некоторых из вариантов осуществления изобретения RGN содержит аминокислотную последовательность SEQ ID NO: 1, 11, 19, 27, 36, 45 или 54 или ее активный вариант или фрагмент. В других вариантах осуществления изобретения управляющая РНК содержит последовательность CRISPR повторов, включающую в себя нуклеотидную последовательность SEQ ID NO: 2, 12, 20, 28, 37, 46 или 55, или ее активный вариант или фрагмент. В конкретных вариантах осуществления изобретения управляющая РНК содержит tracrРНК, содержащую нуклеотидную последовательность SEQ ID NO: 3, 13, 21, 29, 38, 47 или 56, или ее активный вариант или фрагмент. Управляющая РНК системы может быть одиночной управляющей РНК или двойной управляющей РНК.

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

Хромосомная модификация клетки, организма, органеллы или эмбриона может приводить к измененной экспрессии (повышающая или понижающая регуляция), 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 модифицированная клетка, организм, органелла или эмбрион называется "нокаутом". Фенотип нокаута может быть результатом делеционной мутации (т.е. делеции по меньшей мере одного нуклеотида), инсерционной мутации (т.е. вставки по меньшей мере одного нуклеотида) или нонсенс-мутации (т.е. замены по меньшей мере одного нуклеотида так, что вводится стоп-кодон).

Кроме того, хромосомная модификация клетки, организма, органеллы или эмбриона может привести к "нокауту", который является результатом хромосомной интеграции нуклеотидной последовательности

сти, кодирующей белок. В некоторых из приведенных вариантов осуществления изобретения кодирующая последовательность интегрирована в хромосому таким образом, что хромосомная последовательность, кодирующая белок дикого типа, инактивирована, но экзогенно введенный белок экспрессируется.

В других вариантах осуществления изобретения хромосомная модификация приводит к продукции вариантного белкового продукта. Экспрессированный вариантный белковый продукт может иметь по меньшей мере одну аминокислотную замену и/или вставку, или делецию по меньшей мере, одной аминокислоты. Вариант белкового продукта, кодируемый измененной хромосомной последовательностью, может проявлять измененные характеристики или активности по сравнению с белком дикого типа, включая, помимо прочего, измененную ферментативную активность или специфичность к субстрату.

В ряде других вариантов осуществления изобретения хромосомная модификация может приводить к измененной схеме экспрессии белка. В качестве неограничивающего примера, хромосомные изменения в регуляторных областях, контролирующей экспрессию белкового продукта, могут привести к сверхэкспрессии или подавлению белкового продукта или к измененной тканевой или временной схеме экспрессии.

Артикль "a" и "an" используются в данном документе для обозначения одного или нескольких (то есть по меньшей мере одного) грамматического объекта статьи. Например, "полипептид" означает один или несколько полипептидов.

Все публикации и заявки на патенты, упомянутые в описании, указывают на уровень специалистов в области, к которой относится это изобретение. Все публикации и заявки на патенты включены сюда посредством ссылки в той же степени, как если бы каждая отдельная публикация или патентная заявка была специально и индивидуально приведена в нем для включения посредством ссылки.

Хотя вышеизложенное изобретение было описано довольно подробно с помощью иллюстраций и примеров с целью ясности понимания, будет очевидно, что определенные изменения и модификации могут быть применены на практике в пределах объема прилагаемых вариантов осуществления изобретения.

Неограничивающие варианты осуществления изобретения включают:

1. Молекулу нуклеиновой кислоты, содержащую полинуклеотид, кодирующий полипептид РНК-управляемой нуклеазы (RGN), где указанный полинуклеотид содержит нуклеотидную последовательность, кодирующую полипептид RGN, содержащий аминокислотную последовательность, имеющую по меньшей мере 95% идентичность с последовательностью SEQ ID NO: 1, 11, 19, 27, 36, 45 или 54,

где указанный полипептид RGN связывает последовательность целевой ДНК специфическим для РНК-управляемой последовательности образом при связывании с управляющей РНК (gРНК), способной гибридизоваться с указанной последовательностью целевой ДНК, и

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

2. Молекулу нуклеиновой кислоты, описанную в варианте осуществления изобретения 1, где указанный полипептид RGN способен расщеплять указанную последовательность Целевой ДНК при связывании.

3. Молекулу нуклеиновой кислоты по п.2, где расщепление указанным полипептидом RGN приводит к образованию двухцепочечного разрыва.

4. Молекулу нуклеиновой кислоты по п.2, в которой расщепление указанным полипептидом RGN приводит к одноцепочечному разрыву.

5. Молекулу нуклеиновой кислоты по любому из пп.1-4, где полипептид RGN функционально слит с полипептидом, редактирующим основание.

6. Молекулу нуклеиновой кислоты по любому из пп.1-5, где полипептид RGN содержит один или несколько сигналов ядерной локализации.

7. Молекулу нуклеиновой кислоты по любому из пп.1-6, где полипептид RGN имеет кодон, оптимизированный для экспрессии в эукариотической клетке.

8. Молекулу нуклеиновой кислоты по любому из пп.1-7, в которой указанная последовательность целевая ДНК расположена рядом с мотивом, примыкающим к протоспейсеру (PAM).

9. Вектор, содержащий молекулу нуклеиновой кислоты по любому из пп.1-8.

10. Вектор по п.9, дополнительно содержащий по меньшей мере одну нуклеотидную последовательность, кодирующую указанную gРНК, способную гибридизоваться с указанной последовательностью целевой ДНК.

11. Вектор по п.10, где указанная gРНК представляет собой единственную управляющую РНК.

12. Вектор по п.10, где указанная gРНК представляет собой двойную управляющую РНК.

13. Вектор по любому из пп.10-12, в котором управляющая РНК содержит CRISPR РНК, содержащую повторяющуюся последовательность CRISPR, имеющую по меньшей мере 95% идентичность с последовательностью SEQ ID NO: 2, 12, 20, 28, 37, 46, или 55.

14. Вектор по любому из пп.10-13, где управляющая РНК включает tracrРНК, по меньшей мере на 95% идентичную последовательности SEQ ID NO: 3, 13, 21, 29, 38, 47 или 56.

15. Клетку, содержащую молекулу нуклеиновой кислоты по любому из пп.1-8 или вектор по пп.9-14.

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

17. Способ получения полипептида RGN, включающий введение в клетку молекулы гетерологичной нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую полипептид РНК-управляемой нуклеазы (RGN), содержащий аминокислотную последовательность по меньшей мере на 95% идентичную с последовательности SEQ ID NO: 1, 11, 19, 27, 36, 45 или 54,

где указанный полипептид RGN связывает последовательность целевой ДНК специфически для РНК-управляемой последовательности образом, при связывании с управляющей РНК (gРНК), способной гибридизоваться с указанной последовательностью целевой ДНК и культивирование указанной клетки в условиях, в которых экспрессируется полипептид RGN.

18. Способ п.16 или 17, дополнительно включающий очистку указанного полипептида RGN.

19. Способ по п.16 или 17, где указанная клетка дополнительно экспрессирует одну или несколько управляющих РНК, которые связываются с указанным полипептидом RGN с образованием комплекса рибонуклеопротеина RGN.

20. Способ по п.19, дополнительно включающий очистку указанного комплекса рибонуклеопротеина RGN.

21. Молекулу нуклеиновой кислоты, содержащую полинуклеотид, кодирующий CRISPR РНК (сгРНК), где указанная сгРНК содержит спейсерную последовательность и последовательность CRISPR повторов, причем указанная последовательность CRISPR повторов включает нуклеотидную последовательность, имеющую по меньшей мере 95% идентичность последовательности SEQ ID NO: 2, 12, 20, 28, 37, 46 или 55,

где управляющая РНК, содержащая:

а) указанную сгРНК; и

б) трансктивирующую РНК CRISPR (trасгРНК), гибридизованную с указанной повторяющейся последовательностью CRISPR указанной сгРНК; которая способна гибридизоваться с последовательностью целевой ДНК специфическим для последовательности образом через спейсерную последовательность указанной сгРНК, когда указанная управляющая РНК связана с полипептидом РНК-управляемой нуклеазы (RGN), и

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

22. Вектор, содержащий молекулу нуклеиновой кислоты по п.21.

23. Вектор по п.22, где указанный вектор дополнительно содержит полинуклеотид, кодирующий указанную trасгРНК.

24. Вектор по п.23, где указанная trасгРНК содержит нуклеотидную последовательность, по меньшей мере на 95% идентичную последовательности SEQ ID NO: 3, 13, 21, 29, 38, 47 или 56.

25. Вектор по п.23 или 24, где указанный полинуклеотид, кодирующий указанную сгРНК, и указанный полинуклеотид, кодирующий указанную trасгРНК, функционально связаны с одним и тем же промотором и кодируются, как единственная управляющая РНК.

26. Вектор по п.23 или 24, где указанный полинуклеотид, кодирующий указанную сгРНК, и указанный полинуклеотид, кодирующий указанную trасгРНК, функционально связаны с отдельными промоторами.

27. Вектор по любому из пп.22-26, где указанный вектор дополнительно содержит полинуклеотид, кодирующий указанный полипептид RGN, где указанный полипептид RGN содержит аминокислотную последовательность, по меньшей мере на 95% идентичную последовательности SEQ ID NO: 1, 11, 19, 27, 36, 45 или 54.

28. Молекулу нуклеиновой кислоты, содержащую полинуклеотид, кодирующий трансктивирующую РНК CRISPR (trасгРНК), содержащую нуклеотидную последовательность, по меньшей мере на 95% идентичную последовательности с SEQ ID NO: 3, 13, 21, 29, 38, 47 или 56;

где управляющая РНК, содержащая:

а) указанную trасгРНК; и

б) сгРНК, содержащая спейсерную последовательность и последовательность CRISPR повторов, где указанная trасгРНК гибридизуется с указанной последовательностью CRISPR повторов указанной сгРНК;

которая способна гибридизоваться с последовательностью целевой ДНК специфическим для последовательности образом через спейсерную последовательность указанной сгРНК, когда указанная управляющая РНК связана с полипептидом РНК-управляемой нуклеазы (RGN), и

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

29. Вектор, содержащий молекулу нуклеиновой кислоты по п.28.

30. Вектор по п.29, где указанный вектор дополнительно содержит полинуклеотид, кодирующий указанную сгРНК.

31. Вектор по п.30, где последовательность CRISPR повторов указанной сгРНК включает нуклеотидную последовательность, по меньшей мере на 95% идентичную последовательности SEQ ID NO: 2, 12, 20, 28, 37, 46 или 55.

32. Вектор по п.30 или 31, где указанный полинуклеотид, кодирующий указанную *cr*РНК, и указанный полинуклеотид, кодирующий указанную *tracr*РНК, функционально связаны с одним и тем же промотором и кодируются как единственная управляющая РНК.

33. Вектор по п.30 или 31, где указанный полинуклеотид, кодирующий указанную *cr*РНК, и указанный полинуклеотид, кодирующий указанную *tracr*РНК, функционально связаны с отдельными промоторами.

34. Вектор по любому из пп.29-33, где указанный вектор дополнительно содержит полинуклеотид, кодирующий указанный полипептид RGN, где указанный полипептид RGN содержит аминокислотную последовательность, по меньшей мере на 95% идентичную последовательности с SEQ ID NO: 1, 11, 19, 27, 36, 45 или 54.

35. Систему связывания целевой последовательности ДНК, где указанная система включает:

а) одну или несколько управляющих РНК, способных гибридизоваться с указанной последовательностью целевой ДНК или одну или несколько нуклеотидных последовательностей, кодирующих одну или несколько направляющих РНК (gРНК); и

б) полипептид РНК-управляемой нуклеазы (RGN), содержащий аминокислотную последовательность, по меньшей мере на 95% идентичную последовательности SEQ ID NO: 1, 11, 19, 27, 36, 45 или 54, или нуклеотидную последовательность, кодирующую полипептид RGN,

где каждая указанная нуклеотидная последовательность, кодирующая одну или несколько управляющих РНК и кодирующая полипептид RGN, функционально связаны с промотором, гетерологичным указанной нуклеотидной последовательности,

где одна или несколько управляющих РНК гибридизуются с последовательностью целевой ДНК и

где одна или несколько управляющих РНК образуют комплекс с полипептидом RGN, тем самым заставляя указанный полипептид RGN связываться с указанной последовательностью целевой ДНК.

36. Систему п.35, в которой указанная gРНК представляет собой единственную управляющую РНК (sgРНК).

37. Систему по п.35, в которой указанная gРНК представляет собой двойную управляющую РНК.

38. Систему по любому из пп.35-37, где указанная gРНК содержит последовательность CRISPR повторов, содержащую нуклеотидную последовательность, по меньшей мере на 95% идентичную последовательности SEQ ID NO: 2, 12, 20, 28, 37, 46 или 55.

39. Систему по любому из пп.35-38, в которой указанная gРНК включает *tracr*РНК, содержащую нуклеотидную последовательность, по меньшей мере на 95% идентичную последовательности с SEQ ID NO: 3, 13, 21, 29, 38, 47 или 56.

40. Систему по любому из пп.35-39, в которой указанная последовательность целевой ДНК расположена рядом с мотивом, примыкающим к протоспейсеру (PAM).

41. Систему по любому из пп.35-40, в которой целевая последовательность ДНК находится внутри клетки.

42. Систему по п.41, в которой клетка является эукариотической клеткой.

43. Систему по п.42, где эукариотическая клетка представляет собой растительную клетку.

44. Систему по п.42, в которой эукариотическая клетка представляет собой клетку млекопитающего.

45. Систему по п.42, в которой эукариотическая клетка представляет собой клетку насекомого.

46. Систему по п.41, в которой клетка представляет собой прокариотическую клетку.

47. Систему по любому из пп.35-46, где при транскрибировании одна или несколько управляющих РНК гибридизуются с последовательностью целевой ДНК, а управляющая РНК образует комплекс с полипептидом RGN, который вызывает расщепление последовательности целевой ДНК.

48. Систему по п.47, в которой расщепление приводит к двухцепочечному разрыву.

49. Систему по п.47, в которой расщепление указанным полипептидом RGN приводит к одноцепочечному разрыву.

50. Систему по любому из пп.35-49, в которой полипептид RGN функционально связан с полипептидом, редактирующим основание.

51. Систему по любому из пп.35-50, где полипептид RGN содержит один или несколько сигналов ядерной локализации.

52. Систему по любому из пп.35-51, где полипептид RGN имеет кодон, оптимизированный для экспрессии в эукариотической клетке.

53. Систему по любому из пп.35-52, в которой нуклеотидные последовательности, кодирующие одну или несколько управляющих РНК, и нуклеотидная последовательность, кодирующая полипептид RGN, расположены в одном векторе.

54. Систему по любому из пп.35-53, где указанная система дополнительно содержит один или несколько донорных полинуклеотидов или одну или несколько нуклеотидных последовательностей, кодирующих один или несколько донорных полинуклеотидов.

55. Способ связывания последовательности целевой ДНК, включающий доставку системы по любому из пп.35-54 к указанной последовательности Целевой ДНК или к клетке, содержащей последовательность Целевой ДНК.

56. Способ по п.55, где указанный полипептид RGN или указанная управляющая РНК дополнительно содержит детектируемую метку, что позволяет обнаруживать указанную последовательность целевой ДНК.

57. Способ по п.55, где указанная управляющая РНК или указанный полипептид RGN дополнительно содержат модулятор экспрессии, тем самым модулируя экспрессию указанной целевой последовательности ДНК или гена, находящегося под транскрипционным контролем указанной последовательности целевой ДНК.

58. Способ расщепления или модификации последовательности целевой ДНК, включающий доставку системы по любому из пп.35-54 в указанную последовательность целевой ДНК или клетку, содержащую последовательность целевой ДНК.

59. Способ по п.58, где указанная модифицированная последовательность целевой ДНК включает вставку гетерологичной ДНК в последовательность целевой ДНК.

60. Способ по п.58, где указанная модифицированная последовательность целевой ДНК включает делецию по меньшей мере одного нуклеотида из последовательности Целевой ДНК.

61. Способ по п.58, где указанная модифицированная последовательность целевой ДНК содержит мутацию по меньшей мере одного нуклеотида в последовательности целевой ДНК.

62. Способ связывания целевой последовательности ДНК, включающий:

а) сборку рибонуклеотидного комплекса с РНК-управляемой нуклеазой (RGN) *in vitro* путем комбинирования:

I) одной или нескольких управляющих РНК, способных гибридизоваться с последовательностью целевой ДНК; и

II) полипептида RGN, содержащего аминокислотную последовательность, по меньшей мере на 95% идентичности SEQ ID NO: 1, 11, 19, 27, 36, 45 или 54 в условиях, подходящих для образования рибонуклеотидного комплекса RGN и

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

где одна или несколько управляющих РНК гибридизуются с целевой последовательностью ДНК, тем самым направляя указанный полипептид RGN на связывание с указанной целевой последовательностью ДНК.

63. Способ по п.62, где указанный полипептид RGN или указанная управляющая РНК дополнительно содержит детектируемую метку, что позволяет обнаруживать указанную последовательность целевую ДНК.

64. Способ по п.62, где указанная управляющая РНК или указанный полипептид RGN дополнительно содержит модулятор экспрессии, что позволяет модулировать экспрессию указанной последовательности целевой ДНК.

65. Способ расщепления и/или модификации целевой последовательности ДНК, включающий введение молекулы ДНК в контакт с:

а) полипептидом РНК-управляемой нуклеазы (RGN), где указанная RGN содержит аминокислотную последовательность, по меньшей мере на 95% идентичности SEQ ID NO: 1, 11, 19, 27, 36, 45 или 54; и

б) одной или несколькими управляющими РНК, способными нацеливать RGN из (а) на последовательность целевой ДНК,

где одна или несколько управляющих РНК гибридизуются с целевой последовательностью ДНК, тем самым направляя указанный полипептид RGN на связывание с указанной целевой последовательностью ДНК, и происходит расщепление и/или модификация указанной целевой последовательности ДНК.

66. Способ по п.65, где указанная модифицированная последовательность целевой ДНК включает вставку гетерологичной ДНК в последовательность целевой ДНК.

67. Способ по п.65, где указанная модифицированная последовательность целевой ДНК включает делецию по меньшей мере одного нуклеотида из последовательности целевой ДНК.

68. Способ по п.65, где указанная модифицированная последовательность целевой ДНК включает мутацию по меньшей мере одного нуклеотида в последовательности целевой ДНК.

69. Способ по любому из пп.62-68, где указанная gРНК представляет собой единичную управляющую РНК (sgРНК).

70. Способ по любому из пп.62-68, где указанная gРНК представляет собой двойную управляющую РНК.

71. Способ по любому из пп.62-70, где указанная gРНК содержит последовательность CRISPR повторов, содержащую нуклеотидную последовательность, по меньшей мере на 95% идентичности SEQ ID NO: 2, 12, 20, 28, 37, 46 или 55.

72. Способ по любому из пп.62-71, где указанная gРНК включает tracrРНК, содержащую нуклеотидную последовательность, по меньшей мере на 95% идентичности SEQ ID NO: 3, 13, 21, 29, 38, 47 или 56.

73. Способ по любому из пп.62-72, в котором указанная последовательность целевой ДНК распо-

ложена рядом с мотивом, примыкающим к протоспейсеру (PAM).

74. Способ по любому из пп.55-73, где целевая последовательность ДНК находится внутри клетки.
75. Способ по п.74, где клетка является эукариотической клеткой.
76. Способ по п.75, где эукариотическая клетка представляет собой растительную клетку.
77. Способ по п.75, где эукариотическая клетка представляет собой клетку млекопитающего.
78. Способ по п.75, в котором эукариотическая клетка представляет собой клетку насекомого.
79. Способ по п.74, где клетка является прокариотической клеткой.
80. Способ по любому из пп.74-79, дополнительно включающий культивирование клетки в условиях, в которых полипептид RGN экспрессируется и расщепляет последовательность целевой ДНК с получением модифицированной последовательности ДНК; и выбор клетки, содержащей указанную модифицированную последовательность ДНК.
81. Клетку, содержащую модифицированную последовательность целевой ДНК по п.80.
82. Клетку по п.81, где клетка является эукариотической клеткой.
83. Клетку по п.82, где эукариотическая клетка представляет собой растительную клетку.
84. Растение, содержащее клетку по п.83.
85. Семя, содержащее клетку по п.83.
86. Клетку по п.82, где эукариотическая клетка представляет собой клетку млекопитающего.
87. Клетку по п.82, где эукариотическая клетка представляет собой клетку насекомого.
88. Клетку по п.81, где клетка представляет собой прокариотическую клетку.
89. Способ получения генетически модифицированной клетки с коррекцией причинной мутации генетически наследуемого заболевания, включающий введение в клетку:
 - а) полипептида РНК-управляемой нуклеазы (RGN), где полипептид RGN содержит аминокислотную последовательность, по меньшей мере на 95% идентичную последовательности SEQ ID NO: 1, 11, 19, 27, 36, 45 или 54, или полинуклеотид, кодирующий указанный полипептид RGN, при этом указанный полинуклеотид, кодирующий полипептид RGN, функционально связан с промотором для обеспечения экспрессии полипептида RGN в клетке; и
 - б) управляющую РНК (gРНК), где gРНК содержит последовательность CRISPR повторов, содержащую нуклеотидную последовательность, по меньшей мере на 95% идентичную последовательности SEQ ID NO: 2, 12, 20, 28, 37, 46 или 55, или полинуклеотид, кодирующий указанную gРНК, при этом указанный полинуклеотид, кодирующий gРНК, функционально связан с промотором для обеспечения экспрессии gРНК в клетке посредством чего RGN и gРНК нацелены на геномное местоположение причинной мутации и модифицируют геномную последовательность для удаления причинной мутации.
90. Способ по п.89, где RGN слита с полипептидом, который обладает активностью редактирования оснований.
91. Способ по п.90, где полипептид с активностью редактирования оснований представляет собой цитидиндезаминазу или аденозиндезаминазу.
92. Способ по п.89, где клетка представляет собой клетку животного.
93. Способ по п.89, где клетка представляет собой клетку млекопитающего.
94. Способ по п.92, где клетка получена от собаки, кошки, мыши, крысы, кролика, лошади, коровы, свиньи или человека.
95. Способ по п.92, где генетически наследуемое заболевание представляет собой заболевание, указанное в табл. 12.
96. Способ по п.92, где генетически наследуемое заболевание представляет собой синдром Гурлера.
97. Способ по п.96, где gРНК дополнительно содержит спейсерную последовательность, которая нацелена на SEQ ID NO: 453, 454 или 455.
98. Способ получения генетически модифицированной клетки с делецией в вызывающей заболевание геномной области нестабильности, включающий введение в клетку:
 - а) полипептида РНК-управляемой нуклеазы (RGN), где полипептид RGN содержит аминокислотную последовательность, по меньшей мере на 95% идентичную последовательности SEQ ID NO: 1, 11, 19, 27, 36, 45 или 54, или полинуклеотид, кодирующий указанный полипептид RGN, при этом указанный полинуклеотид, кодирующий полипептид RGN, функционально связан с промотором для обеспечения экспрессии полипептида RGN в клетке и
 - б) управляющую РНК (gРНК), где эта gРНК содержит последовательность CRISPR повторов, содержащую нуклеотидную последовательность, по меньшей мере на 95% идентичную последовательности SEQ ID NO: 2, 12, 20, 28, 37, 46 или 55, или полинуклеотид, кодирующий указанную gРНК, причем указанный полинуклеотид, кодирующий gРНК, функционально связан с промотором, обеспечивающим экспрессию gРНК в клетке, и, кроме того, где gРНК содержит спейсерную последовательность, которая нацелена на 5'-фланг геномной области нестабильности и
 - в) вторую направляющую РНК (gРНК), где gРНК содержит последовательность CRISPR повторов, содержащую нуклеотидную последовательность, по меньшей мере на 95% идентичную последовательности SEQ ID NO: 2, 12, 20, 28, 37, 46 или 55, или полинуклеотид, кодирующий указанную gРНК, при этом указанный полинуклеотид, кодирующий gРНК, функционально связан с промотором для обеспече-

ния экспрессии gPHK в клетке, и, кроме того, указанная вторая gPHK содержит спейсерную последовательность, нацеленную на 3'-конец геномной области нестабильности, посредством чего RGN и две gPHK нацелены на нестабильную область генома и по меньшей мере часть нестабильной области генома удаляется.

99. Способ по п.98, где клетка представляет собой клетку животного.

100. Способ по п.98, где клетка представляет собой клетку млекопитающего.

101. Способ по п.100, где клетка получена от собаки, кошки, мыши, крысы, кролика, лошади, коровы, свиньи или человека.

102. Способ по п.99, где генетически наследуемым заболеванием является атаксия Фридриха или болезнь Хантингтона.

103. Способ по п.102, где первая gPHK дополнительно содержит спейсерную последовательность, которая нацелена на SEQ ID NO: 468, 469 или 470.

104. Способ по п.103, где вторая gPHK дополнительно содержит спейсерную последовательность, нацеленную на SEQ ID NO: 471.

105. Способ получения генетически модифицированной гематопозитической клетки-предшественника млекопитающего, имеющей пониженную экспрессию gPHK и белка BCL11A, при этом способ включает введение в выделенную гематопозитическую клетку-предшественницу человека:

а) полипептида РНК-управляемой нуклеазы (RGN), где полипептид RGN содержит аминокислотную последовательность, по меньшей мере на 95% идентичную последовательности SEQ ID NO: 1, 11, 19, 27, 36, 45 или 54, или полинуклеотид, кодирующий указанный полипептид RGN, при этом указанный полинуклеотид, кодирующий полипептид RGN, функционально связан с промотором для обеспечения экспрессии полипептида RGN в клетке и

б) управляющей РНК (gPHK), где gPHK содержит последовательность CRISPR повторов, содержащую нуклеотидную последовательность, по меньшей мере на 95% идентичную последовательности SEQ ID NO: 2, 12, 20, 28, 37, 46 или 55, или полинуклеотид, кодирующий указанную gPHK, при этом указанный полинуклеотид, кодирующий gPHK, функционально связан с промотором для обеспечения экспрессии gPHK в клетке, посредством чего RGN и gPHK экспрессируются в клетке и расщепляются в области энхансера BCL11A, что приводит к генетической модификации гематопозитической клетки-предшественника человека и снижению экспрессии mPHK и/или белка BCL11A.

106. Способ по п.105, где gPHK дополнительно содержит спейсерную последовательность, которая нацелена на SEQ ID NO: 473, 474, 475, 476, 477 или 478.

107. Систему связывания последовательности целевой ДНК; указанная система включает:

а) одну или несколько управляющих РНК, способных гибридизоваться с указанной последовательностью целевой ДНК или одну, или несколько нуклеотидных последовательностей, кодирующих одну или несколько управляющих РНК (gPHK), и

б) полипептид РНК-управляемой нуклеазы (RGN), содержащий аминокислотную последовательность, по меньшей мере на 95% идентичную последовательности SEQ ID NO: 1, 11, 19, 27, 36, 45 или 54, где одна или несколько управляющих РНК гибридизуются с последовательностью целевой ДНК и где одна или несколько управляющих РНК образуют комплекс с полипептидом RGN, тем самым направляя указанный полипептид RGN на связывание с указанной последовательностью целевой ДНК.

108. Систему по п.107, в которой указанный полипептид RGN имеет мертвую нуклеазу или функционирует как никаза.

109. Систему по п.107 или п.108, в которой указанный полипептид RGN функционально слит с полипептидом, редактирующим основание.

110. Систему по п.109, в которой полипептид, редактирующий основание, представляет собой дезаминазу.

111. Систему по п.110, в которой дезаминаза представляет собой цитидиндезаминазу или аденозиндезаминазу.

Следующие ниже примеры предлагаются в качестве иллюстрации, а не в качестве ограничения.

Экспериментальная часть.

Пример 1. Идентификация РНК-управляемой нуклеазы.

Было идентифицировано семь различных CRISPR-связанных РНК-управляемых нуклеаз (RGN), которые описаны ниже в табл. 1. В табл. 1 представлены названия каждой RGN, ее аминокислотная последовательность, источник, из которого она была получена, а также процессированные последовательности sgPHK и tracrPHK. В табл. 1 также представлена общая последовательность единой управляющей РНК (sgPHK), где поли-N указывает расположение спейсерной последовательности, которая определяет последовательность-мишень нуклеиновой кислоты sgPHK. Системы RGN APG систем APG05083.1, APG07433.1, APG08290.1 и APG08290.1 имели консервативную последовательность в основании стержня шпильки tracrPHK, UNANNG (SEQ ID NO: 68). Для системы AP05459.1 последовательность в том же месте - UNANNU (SEQ ID NO: 557). Для систем APG04583.1 и APG01688.1 последовательность представляет собой UNANNA (SEQ ID NO: 558).

Сводка SEQ ID и систем, связанных с CRISPR

RGN ID	SEQ ID NO.	Источник	crPHK повтор seq (SEQ ID NO.)	tracrPHK (SEQ ID NO.)	sgPHK (SEQ ID NO.)
APG05083.1	1	Bacillus sp.	2	3	10
APG07433.1	11	Bacillus sp.	12	13	18
APG07513.1	19	Bacillus sp.	20	21	26
APG08290.1	27	Bacillus sp.	28	29	35
APG05459.1	36	Enterococcus sp.	37	38	44
APG04583.1	45	Enterococcus sp.	46	47	53
APG01688.1	54	Empedobacter sp.	55	56	62

Пример 2: Идентификация управляющей РНК и конструирование sgPHK.

Культуры бактерий, которые изначально экспрессируют исследуемую систему нуклеаз, управляемую РНК, выращивали до средней логарифмической фазы (OD600~0,600), осаждали и быстро замораживали. РНК выделяли из осадка с использованием набора для выделения miPHK mirVANA (Life Technologies, Карлсбад, Калифорния), и библиотеки секвенирования получали из выделенной РНК с использованием набора для подготовки библиотеки малых NEBNext РНК (NEB, Беверли, Массачусетс). Препарат библиотеки фракционировали на 6%-ном полиакриламидном геле на фракции 2 размеров, соответствующие видам РНК 18-65nt (нуклеотидов) и 90-200nt (нуклеотидов), для обнаружения crPHKs и tracrPHKs, соответственно. Глубокое секвенирование (40 п.о. с парным концом для меньшей фракции и 80 п.о. с парным концом для большей фракции) было выполнено на приборе Next Seq 500 (High Output kit) поставщиком услуг (MoGene, Сент-Луис, Миссури). Чтения были качественно обрезаны с помощью Cutadapt и сопоставлены с эталонными геномами с помощью Bowtie2. Пользовательская программа РНКseq была написана на языке python для обнаружения транскриптов crPHK и tracrPHK. Границы обработанной crPHK определяли по охвату последовательностей массива спейсеров нативных повторов. Антиповторную часть tracrPHK идентифицировали с использованием разрешающих параметров BLASTn. Глубина секвенирования РНК подтвердила границы обработанной tracrPHK путем идентификации транскрипта, содержащего антиповтор. Ручная настройка РНК была выполнена с использованием предсказания вторичной структуры NUPACK, программного обеспечения для сворачивания РНК. Кассеты sgPHK получали путем синтеза ДНК и обычно конструировали следующим образом (5'→3'): спейсерная последовательность 20-30 п.о. - обработанная повторяющаяся часть crPHK - некомплементарный линкер 4 п.о. (AAAG; SEQ ID NO: 63) - процессированная tracrPHK. Также можно использовать другие некомплементарные линкеры длиной 4 п.о., например GAAA (SEQ ID NO: 64) или ACUU (SEQ ID NO: 65). В некоторых случаях может использоваться нуклеотидный линкер длиной 6 п.о., например CAAAGG (SEQ ID NO: 66). Для анализов in vitro sgPHK синтезировали транскрипцией in vitro кассеты sgPHK с помощью набора GeneArt™ Precision gPHK Synthesis Kit (ThermoFisher). Обработанные последовательности crPHK и tracrPHK для каждого из полипептидов RGN идентифицированы и представлены в табл. 1. См. ниже sgPHKs, сконструированные для РАМ библиотек 1 и 2.

Пример 3: Определение требований РАМ для каждой RGN.

Требования РАМ для каждого RGN определяли с использованием анализа истощения РАМ, по существу адаптированного из Kleinstiver и др. (2015) Nature 523: сс. 481-485 и Zetsche и др. (2015) Cell 163: сс. 759-771. Вкратце, в основной цепи pUC18 (ampR) были созданы две библиотеки плазмид (L1 и L2), каждая из которых содержала отдельную последовательность протоспейсера (мишень) размером 30 п.о., фланкированную 8 случайными нуклеотидами (т.е. область РАМ). Целевая последовательность и фланкирующая область РАМ библиотеки 1 и библиотеки 2 для каждой RGN представлены в табл. 2.

Библиотеки были отдельно электропорированы в клетки E. coli BL21 (DE3), несущие векторы экспрессии pRSF-1b, содержащие RGN, предлагаемую в данном изобретении (кодоны оптимизированы для E. coli) вместе с родственной sgPHK, содержащей спейсерную последовательность, соответствующую протоспейсеру в L1 или L2. Достаточное количество библиотечной плазмиды использовали в реакции трансформации для получения >10⁶ КОЕ. Как RGN, так и sgPHK в скелете pRSF-1b находились под контролем промоторов T7. Реакционной смеси давали возможность восстановиться в течение 1 ч, после чего ее разбавляли средой LB, содержащей карбенициллин и канамицин, и выращивали в течение ночи. На следующий день смесь разводили в самоиндуцирующейся среде Overnight Express™ Instant TB (Millipore Sigma), чтобы обеспечить экспрессию RGN и sgPHK, и выращивали в течение дополнительных 4 или 20 ч, после чего клетки раскручивали и выделяли плазмидную ДНК с помощью набора Mini-prep (Qiagen, Джермантаун, Мэриленд). В присутствии соответствующей sgPHK плазмиды, содержащие РАМ, которая распознается RGN, будут расщепляться, что приведет к их удалению из популяции. Плазмиды, содержащие РАМ, которые не распознаются RGN или которые трансформируются в бактерии, не содержащие

подходящей sgPHK, будут выживать и реплицироваться. Области PAM и протоспейсеров нерасщепленных плазмид были амплифицированы с помощью ПЦР и подготовлены для секвенирования в соответствии с опубликованными протоколами (руководство по подготовке 16s-метагеномной библиотеки 15044223B, Illumina, Сан-Диего, Калифорния). Глубокое секвенирование (чтение с одного конца 80 пар оснований) было выполнено на MiSeq (Illumina) поставщиком услуг (MoGene, Сент-Луис, Миссури). Как правило, на ампликон получали 1-4 млн считываний. Области PAM были извлечены, подсчитаны и нормализованы к общему количеству считываний для каждого образца. PAM, которые приводят к расщеплению плазмиды, были идентифицированы по тому, что они недостаточно представлены по сравнению с контролями (т.е. когда библиотека трансформируется в *E. coli*, содержащую RGN, но не имеющую подходящей sgPHK). Чтобы представить требования PAM для нового RGN, коэффициенты истощения (частота в выборке/частота в контроле) для всех последовательностей в рассматриваемой области были преобразованы в значения обогащения с преобразованием $-\log$ по основанию 2. Достаточные PAM были определены, как PAM со значениями обогащения $>2,3$ (что соответствует коэффициентам истощения $< \sim 0,2$). PAM, превышающие этот порог, в обеих библиотеках были собраны и использованы для создания веб-логотипов, которые, например, могут быть созданы с помощью веб-службы в Интернете, известной как "weblogo". Последовательности PAM были идентифицированы и запротоколированы, когда имелась согласованная картина у основных обогащенных PAM. PAM (имеющая коэффициент обогащения (EF) $>2,3$) для каждой RGN представлен в табл. 2. Для некоторых RGN также были идентифицированы неограничивающие примерные PAM (имеющие EF $>3,3$). Для APG005083.1 примерная PAM представляет собой NNRNCC (SEQ ID NO: 69). Для APG007433.1 примерная PAM представляет собой NNNNCCR (SEQ ID NO: 70). Для APG007513.1 примерная PAM представляет собой NNRNCC (SEQ ID NO: 71). Для APG001688.1 примерная PAM представляет собой NNRANC (SEQ ID NO: 72).

Таблица 2

Определение PAM

RGN ID	sgPHK L1 (SEQ ID NO.)	sgPHK L2 (SEQ ID NO.)	PAM (SEQ ID NO.)	Целевая последовательность и область PAM библиотеки плазмид 1 (SEQ ID NO.)	Целевая последовательность и область PAM библиотеки плазмид 2 (SEQ ID NO.)
APG05083.1	4	5	6	8	9
APG07433.1	13	14	6	16	17
APG07513.1	21	22	6	24	25
APG08290.1	29	30	32	33	34
APG05459.1	37	38	41	42	43
APG04583.1	45	46	50	51	52
APG01688.1	53	54	59	60	61

Пример 4. Определение расщепления.

Сайты расщепления определяли по реакциям расщепления *in vitro* с использованием RNP (рибонуклеопротеинов). Плазмиды экспрессии, содержащие RGN, слитую с меткой His6 или His10, были сконструированы и трансформированы в штаммы BL21 (DE3) *E. coli*. Экспрессию проводили с использованием самоиндуцирующей среды или среды с индукцией IPTG. После лизиса и осветления белки очищали аффинной хроматографией с иммобилизованным металлом.

Рибонуклеопротеиновые комплексы (содержащие нуклеазу и sgPHK или дуплекс sgPHK и tracrPHK) были сформированы путем инкубации нуклеазы и PHK в буферном растворе в течение 20 мин при комнатной температуре. Комплекс перенесли в пробирку, содержащую буфер для расщепления и амплифицированную мишень ПЦР, называемую "Последовательность 1". Последовательность 1 включает нуклеотидную последовательность (SEQ ID NO: 73), непосредственно связанную своим 3'-концом с соответствующей последовательностью PAM для каждой RGN. Каждую RGN в виде рибонуклеопротеидного комплекса инкубировали с ее соответствующим целевым полинуклеотидом при 25°C (APG04583.1) или 37°C (все остальные) в течение 30 или 60 мин (только для APG05459.1 и APG01688.1). Реакцию расщепления инактивировали нагреванием и проводили на агарозном геле. Полосы продукта расщепления вырезали из геля и секвенировали с использованием секвенирования по Сэнгеру. Сайты расщепления идентифицировали путем сопоставления результатов секвенирования с ожидаемой последовательностью продукта ПЦР. Результаты показаны в табл. 3. Как показано в табл. 3, RGN APG007433.1 также может давать тупой разрез с другой целевой последовательностью.

Сайт расщепления для последовательности 2 (SEQ ID NO: 559, функционально слитый на своем 3'-конце с последовательностью PAM для RGN APG0733.1) определяли с помощью следующего подхода для нуклеазы APG07433.1. После расщепления очищенные в геле продукты ДНК обрабатывали набором для восстановления концов ДНК (Thermo Scientific K0771), лигировали в линеаризованный вектор с тупыми концами, и полученную кольцевую ДНК трансформировали в компетентные клетки *E. coli*. Сту-

пенчатый разрез с выступом 5' может привести к обнаружению перекрывающихся последовательностей в клонах от обоих продуктов расщепления. 3' выступ приведет к отсутствию последовательности, а тупой разрез приведет к обнаружению всей исходной последовательности без перекрытия. Этот эксперимент также подтвердил результаты описанного выше метода для последовательности 1 - большинство клонов были обнаружены, как происходящие из разреза с 5' перекрытием, поэтому не считается, что обнаружение тупого разреза является артефактом при использовании этого способа.

Таблица 3

Сайты расщепления RGN

Нуклеаза	Последовательность 1		Выступ	Последовательность 2		Выступ
	Расстояние от PAM			Расстояние от PAM		
	NTS вырезанный сайт	TS вырезанный сайт		NTS вырезанный сайт	TS вырезанный сайт	
APG07433.1	4	3	1 nt, 5'	3	3	Нет
APG08290.1	4	3	1 nt, 5'	Не определено		
APG05459.1	3	3	None	Не определено		
APG04583.1	3	3	None	Не определено		
APG01688.1	3	3	None	Не определено		

NTS=нецелевая нить;

TS=целевая нить.

Пример 5. Анализ чувствительности к несоответствию.

Плазмиды были сконструированы и получены с последовательностью-мишенью (SEQ ID NO: 73), находящейся сразу за 5'-концом от подходящего мотива PAM для оцениваемой нуклеазы. Также были созданы одиночные несовпадающие последовательности с измененной последовательностью в указанном положении (табл. 4). Комплексы РНП очищенной нуклеазы (APG08290.1 или APG05459.1) и управляющей РНК были сформированы и инкубированы с ПЦР-амплифицированной линейной ДНК из сконструированных плазмид. После инкубации в течение определенного периода времени и инактивации нуклеазами, образцы анализировали электрофорезом в агарозном геле для определения оставшейся фракции линейного продукта ПЦР. Процент отщепленной интактной полосы показан в табл. 5 для несоответствий в каждом положении.

Таблица 4

Последовательности, протестированные для анализа чувствительности к несоответствию для APG08290.1 и APG05459.1

Последовательность протоспейсера	SEQ ID NO.	Несоответствие
GAGCGGACAGCAGCTTCCSTATATCTCGTAC	73	Нет
GAGCGGACAGCAGCTTCCSTATATCTCGTAG	74	1
GAGCGGACAGCAGCTTCCSTATATCTCGTTC	75	2
GAGCGGACAGCAGCTTCCSTATATCTCGAAC	76	3
GAGCGGACAGCAGCTTCCSTATATCTCCTAC	77	4
GAGCGGACAGCAGCTTCCSTATATCTGGTAC	78	5
GAGCGGACAGCAGCTTCCSTATATCACGTAC	79	6
GAGCGGACAGCAGCTTCCSTATATGTCGTAC	80	7
GAGCGGACAGCAGCTTCCSTATAAATCTCGTAC	81	8
GAGCGGACAGCAGCTTCCSTATTTCTCGTAC	82	9
GAGCGGACAGCAGCTTCCSTAAATCTCGTAC	83	10
GAGCGGACAGCAGCTTCCSTTTATCTCGTAC	84	11
GAGCGGACAGCAGCTTCCAAATATCTCGTAC	85	12
GAGCGGACAGCAGCTTCCGTATATCTCGTAC	86	13
GAGCGGACAGCAGCTTGGTATATCTCGTAC	87	14
GAGCGGACAGCAGCTACSTATATCTCGTAC	88	15
GAGCGGACAGCAGCATCSTATATCTCGTAC	89	16
GAGCGGACAGCAGGTTCCSTATATCTCGTAC	90	17
GAGCGGACAGCACCTTCCSTATATCTCGTAC	91	18
GAGCGGACAGCTGCTTCCSTATATCTCGTAC	92	19
GAGCGGACAGGAGCTTCCSTATATCTCGTAC	93	20
GAGCGGACACCAGCTTCCSTATATCTCGTAC	94	21
GAGCGGACTGCAGCTTCCSTATATCTCGTAC	95	22
GAGCGGAGAGCAGCTTCCSTATATCTCGTAC	96	23
GAGCGGTCAGCAGCTTCCSTATATCTCGTAC	97	24
GAGCGCACAGCAGCTTCCSTATATCTCGTAC	98	25
GAGCCGACAGCAGCTTCCSTATATCTCGTAC	99	26
GAGGGGACAGCAGCTTCCSTATATCTCGTAC	100	27
GACCGGACAGCAGCTTCCSTATATCTCGTAC	101	28
GTGCGGACAGCAGCTTCCSTATATCTCGTAC	102	29
CAGCGGACAGCAGCTTCCSTATATCTCGTAC	103	30

Таблица 5

Чувствительность к несоответствию для RGN APG08290.1 и RGN APG05459.1

Положение несоответствия	% расщепления	
	APG08290.1	APG05459.1
Несовместимая РАМ, несоответствия нет	0	0
	95	67
1	0	0
2	0	74
3	73	3
4	0	0
5	0	0
6	31	30
7	0	12
8	0	51
9	0	0
10	75	52
11	77	5
12	79	62
13	28	18
14	8	5
15	90	6
16	85	5
17	81	4
18	100	0
19	100	0
20	100	2
21	100	30
22	100	48
23	100	40
24	100	45
25	100	29
26	100	33
27	100	73
28	100	46
29	100	59
30	100	57

Аналогичный эксперимент по чувствительности к несоответствию был проведен для RGN APG07433.1. Этот эксперимент был аналогичен описанному выше, за исключением того, что альтернативное основание было введено в управляющую РНК, а не в целевую ДНК. Последовательности ДНК для синтеза несовпадающей sgРНК показаны в табл. 6. Результаты анализа чувствительности к несоответствию показаны в табл. 7.

Таблица 6

Последовательности, протестированные для анализа чувствительности к несоответствию для RGN APG07433.1

ДНК-матрица для синтеза sgРНК	SEQ ID NO.	Положение несоответствия
GAGCGGACAGCAGCTTCSTATATCTCGTAC	73	Нет
GAGCGGACAGCAGCTTCSTATATCTCGTAT	104	1
GAGCGGACAGCAGCTTCSTATATCTCGTGC	105	2
GAGCGGACAGCAGCTTCSTATATCTCGCAC	106	3
GAGCGGACAGCAGCTTCSTATATCTCATAC	107	4
GAGCGGACAGCAGCTTCSTATATCTTGTAC	108	5
GAGCGGACAGCAGCTTCSTATATCCCCTAC	109	6

GAGCGGACAGCAGCTTCCTATATTTTCGTAC	110	7
GAGCGGACAGCAGCTTCCTATACCTTCGTAC	111	8
GAGCGGACAGCAGCTTCCTATGTCTCGTAC	112	9
GAGCGGACAGCAGCTTCCTACATCTCGTAC	113	10
GAGCGGACAGCAGCTTCCTGTATCTCGTAC	114	11
GAGCGGACAGCAGCTTCCCATATCTCGTAC	115	12
GAGCGGACAGCAGCTTCTTATATCTCGTAC	116	13
GAGCGGACAGCAGCTTCTATATCTCGTAC	117	14
GAGCGGACAGCAGCTCCCTATATCTCGTAC	118	15
GAGCGGACAGCAGCCTCCTATATCTCGTAC	119	16
GAGCGGACAGCAGTTTCCTATATCTCGTAC	120	17
GAGCGGACAGCAACTTCCTATATCTCGTAC	121	18
GAGCGGACAGCGGCTTCCTATATCTCGTAC	122	19
GAGCGGACAGTAGCTTCCTATATCTCGTAC	123	20
GAGCGGACAACAGCTTCCTATATCTCGTAC	124	21
GAGCGGACGGCAGCTTCCTATATCTCGTAC	125	22
GAGCGGATAGCAGCTTCCTATATCTCGTAC	126	23

Таблица 7

Чувствительность к несоответствию для RGN APG07433.1

Положение несоответствия	% расщепления
	APG07433.1
Нет несоответствия	86
1	6
2	21
3	-2
4	1
5	-1
6	0
7	7
8	24
9	14
10	-1
11	72
12	44
13	54
14	60
16	65
17	76
18	84
19	86
20	83
21	83
22	93
23	80

RGN APG07433.1 и APG08290.1 проявляют значительную чувствительность к несоответствию в положениях 1-10 5' от PAM за некоторыми исключениями (табл. 5 и табл. 7). RGN APG05459.1 также чувствительна к несоответствиям в этой области, но ее способность расщеплять днДНК также в значительной степени отменяется из-за несоответствий, удаленных от сайта PAM (табл. 5). Общее количество сайтов со значительным влиянием на то, происходит ли расщепление, составляет по меньшей мере 15 положений в спейсерной последовательности. Это выгодно отличается от других инструментов редактирования генома, таких как хорошо изученная нуклеаза Cas9 из *S. pyogenes*, которая обычно чувствительна к 10-13 п.о. (Hsu и др., Nat Biotechnol (2013) № 31 (9): сс. 827-832). Кроме того, многие из критических сайтов, отменяющих опосредованное RGN APG05459.1 расщепление, очень далеки от последовательности PAM, особенно в диапазоне 13-20 п.о., где многие другие нуклеазы проявляют небольшую чувствительность к ошибочным спариваниям. Это свойство может быть чрезвычайно полезным для нацеливания на генетические локусы, которые имеют близкое сходство последовательностей с другими сайтами в интересующем организме.

Пример 6: Демонстрация активности редактирования генов в клетках млекопитающих.

Кассеты экспрессии RGN были получены и введены в векторы для экспрессии млекопитающих. Каждая RGN была оптимизирована по кодонам для экспрессии в человеке (SEQ ID NO: 127-133) и функционально слита на 5'-конце с последовательностью ядерной локализации SV40 (NLS; SEQ ID NO 134) и с тегами 3×FLAG (SEQ ID NO: 135) и функционально слита на 3'-конце с последовательностями NLS нуклеоплазмина (SEQ ID NO: 136). Каждая кассета экспрессии находилась под контролем промотора цитомегаловируса (CMV) (SEQ ID NO: 137). В данной области техники известно, что энхансер транскрипции CMV (SEQ ID NO: 138) также может быть включен в конструкции, содержащие промотор CMV. Были получены конструкции экспрессии управляющей РНК, кодирующие одиночную гРНК, каждая под контролем промотора U6 человеческой РНК-полимеразы III (SEQ ID NO. 139), которые были введены в вектор pTwist High Copy Amp. Последовательности целевых последовательностей для каждой ориентации приведены в табл. 9.

Описанные выше конструкции вводили в клетки млекопитающих. За день до трансфекции 1×10^5 клеток HEK293T/лунку (Sigma) высевали в 24-луночные чашки в среде Игла, модифицированной Дульбекко (DMEM) плюс 10% (по объему) фетальной телячьей сыворотки (Gibco) и 1% пенициллина-стрептомицин (Gibco). На следующий день, когда конfluence клеток составляла 50-60%, 500 нг плазмиды экспрессии RGN плюс 500 нг плазмиды экспрессии одиночной гРНК были котрансфицированы с использованием 1,5 мкл липофектамина 3000 (Thermo Scientific) на лунку в соответствии с инструкциями производителя. После 48 ч выращивания общую геномную ДНК собирали с использованием набора для выделения геномной ДНК (Machery-Nagel) в соответствии с инструкциями производителя.

Затем анализировали общую геномную ДНК, чтобы определить скорость редактирования для каждой RGN для каждой целевой геномной. Сначала были получены олигонуклеотиды для использования в ПЦР-амплификации и последующего анализа амплифицированного целевого сайта генома. Используемые олигонуклеотидные последовательности перечислены в табл. 8.1-8.5.

Все реакции ПЦР проводили с использованием 10 мкл ДНК-полимеразы 2X Master Mix Phusion High-Fidelity (Thermo Scientific) в 20 мкл реакционной смеси, включая 0,5 мкМ каждого праймера. Большие области генома, охватывающие каждый целевой ген, сначала амплифицировали с использованием праймеров PCR # 1, используя программу: 98°C, 1 мин; 30 циклов [98°C, 10 с; 62°C, 15 с; 72°C, 5 мин]; 72°C, 5 мин; 12°C, постоянно. Затем один микролитр этой реакции ПЦР был дополнительно амплифицирован с использованием праймеров, специфичных для каждой ориентации а (праймеры ПЦР №2), с использованием программы: 98°C, 1 мин; 35 циклов [98°C, 10 с; 67°C, 15 с; 72°C, 30 с]; 72°C, 5 мин; 12°C, постоянно. Праймеры для ПЦР №2 включают выступающие последовательности транспозазного адаптера Nextera Read 1 и Read 2 для секвенирования Illumina.

Таблица 8.1

Олигонуклеотиды для определения активности редактирования генов в клетках млекопитающих, ПЦР №1

Описание	Последовательность	SEQ ID NO
RelA FWD	5'-CTT AGT TTC ACC GCA GGT TCT A-3'	479
RelA REV	5'-CTG TGC ACT CAA CAC TGA TCT A-3'	480
AurkB FWD	5'-CCC AGC CCT AGG TTG TTT ATT-3'	481
AurkB REV	5'-CTG GCT ACA TCT TCC TTG ACT AC-3'	482
HPRT1 FWD	5'-GTG GCA GAA GCA GTG AGT AA-3'	483
HPRT1 REV	5'-TCC CAT CTA GGC ACT AGG TAA A-3'	484

Таблица 8.2

Олигонуклеотиды для обнаружения активности редактирования генов в клетках млекопитающих, ПЦР № 2 для APG05083.1, APG07433.1, APG07513.1 и APG08290.1

Описание	SEQ ID NO.
FWD_Управление 134, Управление 135, Управление 136, Управление 137	485
REV_Управление 134, Управление 135, Управление 136, Управление 137	486
FWD_Управление 138, Управление 139, Управление 140, Управление 141	487
REV_Управление 138, Управление 139, Управление 140, Управление 141	488
REV_Управление 142, Управление 143, Управление 144, Управление 145	489
FWD_Управление 142, Управление 143, Управление 144, Управление 145	490
REV_Управление 164, Управление 165, Управление 166, Управление 167	491
FWD_Управление 164, Управление 165, Управление 166, Управление 167	492
REV_Управление 168, Управление 169, Управление 170, Управление 171	493
FWD_Управление 168, Управление 169, Управление 170, Управление 171	494
REV_Управление 172, Управление 173, Управление 174, Управление 175	495
FWD_Управление 172, Управление 173, Управление 174, Управление 175	496
REV_Управление 185, Управление 186, Управление 187, Управление 188	497
FWD_Управление 185, Управление 186, Управление 187, Управление 188	498
REV_Управление 189, Управление 190, Управление 191, Управление 192	499
FWD_Управление 189, Управление 190, Управление 191, Управление 192	500
REV_Управление 193, Управление 194, Управление 195, Управление 196	501
FWD_Управление 193, Управление 194, Управление 195, Управление 196	502

Таблица 8.3

Олигонуклеотиды для обнаружения активности редактирования генов в клетках млекопитающих, ПЦР №2 для APG005459.1

Описание	SEQ ID NO.
FWD_Управление 146	503
REV_Управление 146	504
FWD_Управление 147	505
REV_Управление 147	506
REV_Управление 148	507
FWD_Управление 148	508
REV_Управление 176	509
FWD_Управление 176	510
REV_Управление 177	511
FWD_Управление 177	512
REV_Управление 209	513
FWD_Управление 209	514
REV_Управление 197	515
FWD_Управление 197	516
REV_Управление 198	517
FWD_Управление 198	518
REV_Управление 199	519
FWD_Управление 199	520

Таблица 8.3

Олигонуклеотиды для обнаружения активности редактирования генов в клетках млекопитающих, ПЦР
№2 для APG005459.1

Описание	SEQ ID NO.
FWD_Управление 146	503
REV_Управление 146	504
FWD_Управление 147	505
REV_Управление 147	506
REV_Управление 148	507
FWD_Управление 148	508
REV_Управление 176	509
FWD_Управление 176	510
REV_Управление 177	511
FWD_Управление 177	512
REV_Управление 209	513
FWD_Управление 209	514
REV_Управление 197	515
FWD_Управление 197	516
REV_Управление 198	517
FWD_Управление 198	518
REV_Управление 199	519
FWD_Управление 199	520

Таблица 8.4

Олигонуклеотиды для обнаружения активности редактирования генов в клетках млекопитающих, ПЦР
№2 для APG004583.1

Описание	SEQ ID NO.
FWD_Управление 149	521
REV_Управление 149	522
FWD_Управление 150	523
REV_Управление 150	524
REV_Управление 151	525
FWD_Управление 151	526
REV_Управление 179	527
FWD_Управление 179	528
REV_Управление 180	529
FWD_Управление 180	530
REV_Управление 181	531
FWD_Управление 181	532
REV_Управление 200	533
FWD_Управление 200	534
REV_Управление 201	535
FWD_Управление 201	536
REV_Управление 202	537
FWD_Управление 202	538

Олигонуклеотиды для обнаружения активности редактирования генов в клетках млекопитающих. ПЦР №2 для APG01988.1

Описание	SEQ ID NO.
FWD_Управление 152	539
REV_Управление 152	540
FWD_Управление 153	541
REV_Управление 153	542
FWD_Управление 154	543
REV_Управление 154	544
FWD_Управление 182	545
REV_Управление 182	546
FWD_Управление 183	547
REV_Управление 183	548
FWD_Управление 184	549
REV_Управление 184	550
FWD_Управление 203	551
REV_Управление 203	552
FWD_Управление 204	553
REV_Управление 204	554
FWD_Управление 205	555
REV_Управление 205	556

Очищенную геномную ДНК подвергали ПЦР № 1 и ПЦР № 2, как указано выше. После второй амплификации ПЦР ДНК очищали с использованием набора для очистки ПЦР (Zymo) в соответствии с инструкциями производителя и элюировали водой. 200-500 нг очищенного продукта ПЦР №2 объединяли с 2 мкл 10× буфера NEB 2 и водой в 20 мкл реакции и отжигали с образованием гетеродуплексной ДНК с использованием программы: 95°C, 5 мин; 95-85°C, охлаждение со скоростью 2°C/c; 85-25°C, охлаждение со скоростью 0,1°C/c; 12°C, постоянно. После отжига 5 мкл ДНК удаляли в качестве контроля без фермента, добавляли 1 мкл эндонуклеазы I T7 (NEB) и реакционную смесь инкубировали при 37°C в течение 1 ч. После инкубации добавляли загрузочный краситель 5x FlashGel (Lonza) и по 5 мкл каждой реакции и контроли анализировали с помощью 2,2% агарозного FlashGel (Lonza) с использованием гелеэлектрофореза. После визуализации геля процент негомологичных концевых соединений (NHEJ) определяли с использованием следующего уравнения: % событий NHEJ=100×[1-(1-расщепленная фракция)^(1/2)], где (расщепленная фракция) определяется как: (плотность расщепленных продуктов)/(плотность расщепленных продуктов+нерасщепленная родительская полоса).

Для некоторых образцов SURVEYOR® использовался для анализа результатов после экспрессии в клетках млекопитающих. Клетки инкубировали при 37°C в течение 72 ч после трансфекции перед экстракцией геномной ДНК. Геномную ДНК экстрагировали с помощью раствора для экстракции ДНК QuickExtract (Episcenter) в соответствии с протоколом производителя. Геномную область, фланкирующую сайт-мишень RGN, амплифицировали с помощью ПЦР, и продукты очищали с использованием колонки QiaQuick Spin (Qiagen) в соответствии с протоколом производителя. Всего 200-500 нг очищенных продуктов ПЦР смешивали с 1 мкл 10×Taq ДНК-полимеразного ПЦР-буфера (Enzymatics) и сверхчистой воды до конечного объема 10 мкл и подвергали процессу повторного отжига для образования гетеродуплекса: 95°C в течение 10 мин, от 95°C до 85°C с линейным изменением со скоростью -2°C/c, от 85°C до 25°C со скоростью -0,25°C/c и выдержкой 25°C в течение 1 мин.

После повторного отжига продукты обрабатывали нуклеазой SURVEYOR® и энхансером SURVEYOR® S (Integrated DNA Technologies) в соответствии с протоколом, рекомендованным производителем, и анализировали на 4-20% полиакриламидных гелях Novex TBE (Life Technologies). Гели окрашивали ДНК-красителем SYBR Gold (Life Technologies) в течение 10 мин и отображали с помощью системы визуализации гелей Gel Doc (Bio-rad). Количественная оценка была основана на относительной интенсивности полос. Процент INDEL определяли по формуле: $100 \times (1 - (1 - (b+c)/(a+b+c))^{1/2})$, где a - интегрированная интенсивность нерасщепленного продукта ПЦР, a, b и c представляют собой интегрированные интенсивности каждого продукта расщепления.

Кроме того, продукты из ПЦР № 2, содержащие выступающие последовательности Illumina, преобразовали в препарат, представляющий собой библиотеку, с последующим протоколом метагеномного секвенирования библиотеки Illumina 16S. Глубокое секвенирование было выполнено на платформе Illumina Mi-Seq поставщиком услуг (MOGene). Обычно на ампликон генерируется 200 000 из 250 п.о. спаренных концевых считываний (2×100000 считываний). Считывания были проанализированы с помощью

CRISPResso (Pinello и др., 2016 Nature Biotech, № 34: сс. 695-697) для расчета скорости редактирования. Выходные выравнивания были подобраны вручную для подтверждения сайтов вставки и делеции, а также для идентификации сайтов микрогомологии в сайтах рекомбинации. Скорость редактирования показана в табл. 9. Все эксперименты проводились на клетках человека. "Целевая последовательность" - это целевая последовательность в пределах целевого гена. Для каждой целевой последовательности управляющая РНК включала комплементарную целевую последовательность РНК и соответствующую sgРНК в зависимости от используемой RGN. Выбранная разбивка экспериментов по управляющей РНК показана в табл. 10.1-10.9.

Таблица 9

Общие показатели редактирования

RGN	Управление РНК ID	Последовательность-мишень (SEQ ID NO.)	Ген-мишень	Общая скорость редактирования в образце	Скорость делеции в образце	Скорость вставки в образце
APG05083.1	189	140	RelA	6.9%		100%
APG05083.1	185	141	RelA	8.2%	79.9%	20.1%
APG05083.1	168	142	HPRT1	11.3%	36.3%	72.4%
APG07433.1	135	143	AurkB	1.7%	88.3%	11.7%
APG07433.1	139	144	AurkB	3.32%	84.3%	15.6%
APG07433.1	143	145	AurkB	2.2%	35.1%	64.9%
APG07433.1	190	146	RelA	60.5%	94.8%	5.2%
APG07433.1	194	147	RelA	6.2%		100%
APG07433.1	165	148	HPRT1	3.5%	68.0%	32.0%
APG07433.1	169	149	HPRT1	18.1%	30.3%	69.7%
APG07433.1	173	150	HPRT1	26.6%	91.9%	10.0%
APG07513.1	144	151	AurkB	2.4%	59.1%	40.9%
APG07513.1	136	152	AurkB	0.9%	80.5%	19.5%
APG08290.1	145	153	AurkB	14.18%	75.85%	24.15%
APG08290.1	188	154	RelA	21.40%	99.05%	50.05%
APG08290.1	192	155	RelA	28.98%	42.05%	57.95%
APG08290.1	196	156	RelA	13.27%	91.80%	8.20%
APG08290.1	167	157	HPRT1	14.14%	65.98%	34.02%
APG08290.1	171	158	HPRT1	48.23%	58.26%	41.74%
APG08290.1	175	159	HPRT1	13.60%	74.18%	25.82%
APG05459.1	197	160	RelA	12.95%	92.16%	7.84%
APG05459.1	199	161	RelA	5.19%	100%	
APG05459.1	146	162	AurkB	1.12%	61.50%	38.50%
APG05459.1	148	163	AurkB	0.78%	49.47%	50.53%
APG05459.1	176	164	HPRT1	6.20%	48.91%	51.09%
APG05459.1	177	165	HPRT1	9.00%	9.33%	90.67%
APG05459.1	209	166	HPRT1	2.50%		100%
APG04583.1	151	167	AurkB	0.0%		
APG01688.1	152	168	AurkB	0.0%		

Особые вставки и делеции для соответствующих управляющих показаны в табл. 10.1-10.7. В этих таблицах целевая последовательность обозначена жирными заглавными буквами. 8-членные области PAM подчеркнуты двойным подчеркиванием, а основные распознаваемые нуклеотиды выделены жирным шрифтом. Вставки обозначаются строчными буквами. Удаления обозначаются тире (---). Местоположение INDEL вычисляется от проксимального края PAM целевой последовательности, причем край является местоположением 0. Местоположение является положительным (+), если местоположение находится на целевой стороне края; расположение отрицательное (-), если расположение находится на стороне PAM ребра.

Таблица 10.1

Специальные вставки и делеции для управления 139 с использованием RGN APG07433.1

Управление 139 (SEQ ID NO: 144)	№ Считывания	% Считывания	% INDEL	Тип	Положение INDEL	Размер
CCTGGGTGTGAGGCTGGGCCATTAACCTCTC C	82540	95.562				
CCTG-----AAAACCTCTCC	170	0.199	23.16	Делеция	-6	19
CCTGGGTGTGA- GCTGGGCCATTAACCTCTCC	132	0.155	17.98	Делеция	+1	1
C-----CTGGGCCATTAACCTCTCC	107	0.125	14.57	Делеция	-9	12
CCTGG-----CTCTCC	101	0.118	13.76	Делеция	-5	23
C-----GGGCCATTAACCTCTCC	61	0.071	8.31	Делеция	-9	14
CCTGGGTGTGAGGcagacCTGGGCCATTAAC CTCTCC	49	0.057	6.67	Вставка	+3	6
CCTGGGTGTGAGggaagctgacgtccttccatggctgctgc ctgtgtgccaccGCTGGGCCATTAACCTCTCC	44	0.051	5.99	Вставка	+2	45
CCTGGGTGTGA-cCTGGGCCATTAACCTCTCC	39	0.045	5.31	Делеция & Мутация	+1	1
CCTGGGTGTGAGGaCTGGGCCATTAACCTCT CC	31	0.036	4.22	Вставка	+3	1

Таблица 10.2

Особые вставки и делеции для управления 143 с использованием RGN APG07433.1

Управление 143 (SEQ ID NO: 145)	№ Считывания	% Считывания	% INDEL	Тип	Положение INDEL	Размер
AGTTGGCAGATGCTCTAATGTACTGCCATGGG AA	84043	99.646				
AGTTGGCAGATGC--- AATGTACTGCCATGGGAA	126	0.149	42.281	Делеция	+3	3
AGTTGGCAGATGCTCTAATGTACTGCCATGGGAA	81	0.096	27.181	Делеция	+3	4
AGTTGGCAGATGCTCTAATGTACTGCCATGGGAA	42	0.049	14.093	Делеция	+4	3
AGTTGGCAGATGC-- TAATGTACTGCCATGGGAA	34	0.040	11.409	Делеция	+3	2
AGTTGGCAGATGCTCTAATGTACTGCCATGGGAA	8	0.009	2.684	Делеция & Мутация	+4	3
AGTTGGCAGATGCTCT- ATGTACTGCCATGGGAA	7	0.008	2.348	Делеция	+6	1

Таблица 10.3

Особые вставки и делеции для управления 190 с использованием RGN APG07433.1

Управление 190 (SEQ ID NO: 146)	№ Считывания	% Считывания	% INDEL	Тип	Положение INDEL	Размер
CGACCTGAATGCTGTGC----- -----	64040	55.46	91.70	Делеция	-164	170
-----GGCGCTCTGGCTTCATTCAATC						
CGACCTGAATGCTGTGCGGCTCTGCTTCCAGGT GA	45619	39.51		WT		
CGACCTGAATGCTGTGCGGCaTCTGCTTCCAGGT GA	3620	3.13	5.18	Insertion	+3	1
CGACCTGAATG-----CTGCTTCCAGGTGA	1110	0.96	1.58	Делеция	+2	10
CGACCTGAATGCT-----TCTGCTTCCAGGTGA	858	0.74	1.22	Делеция	+3	7
CGACCTGAA-----TGCTTCCAGGTGA	206	0.17	0.29	Делеция	+1	13

Таблица 10.4

Особые вставки и делеции для управления 194 с использованием RGN APG07433.1

Управление 194 (SEQ ID NO: 147)	№ Считывания	% Считывания	% INDEL	Тип	Положение INDEL	Размер
GCGTGGGGACTACGACCTGAATGCTGTGCGGCTCT	96635	97.318				
GCG-----ACCTGAATGCTGTGCGGCTCT	1194	1.202	44.836	Делеция	-9	11
GCGTGGGGACTACGA-----GCTGTGCGGCTCT	547	0.550	20.540	Делеция	+3	7
GCGTGGGGA-----CTGAATGCTGTGCGGCTCT	473	0.476	17.761	Делеция	-3	7
GCGTGGGGACT---CCTGAATGCTGTGCGGCTCT	270	0.271	10.138	Делеция	-1	4
GCGTGGGGACTACGAaCCTGAaTGCTGTGCGGCTCT	88	0.088	3.304	Вставка	+3	1
GCGTGGGGACTACGA-----ATGCTGTGCGGCTCT	41	0.041	1.539	Делеция	+3	5
GCGTGGGGACTAC---CTGAATGCTGTGCGGCTCT	31	0.031	1.164	Делеция	+2	3
GCG-----TGAATGCTGTGCGGCTCT	9	0.009	0.337	Делеция	-9	14
GCG-----ACCTGAcTGCTGTGCGGCTCT	5	0.005	0.187	Делеция & Мутация	-9	11
GCGTGGGGACTACG-CCTGAATGCTGTGCGGCTCT	5	0.005	0.187	Делеция	+2	1

Таблица 10.5

Специальные вставки и делеции для управления 145 с использованием RGN APG08290.1

Управление 145 (SEQ ID NO: 153)	№ Считывания	% Считывания	% INDEL	Тип	Положение INDEL	Размер
ATGGAGGAGTTGGCAGATGCTCTAATGTACTGC CATGGGAAG	62618	95.889				
ATGGAGGAGTTGGCAGATGC— TAATGTACTGCC ATGGGAAG	976	1.494	36.363	Делеция	+3	2
ATGGAGGAGTTGGCAGATG-----TACTGC CATGGGAAG	319	0.488	11.885	Делеция	+2	8
ATG-----TACTGC CATGGGAAG	168	0.257	6.259	Делеция	-14	24

ATGGAGGAGTTGG-----TGTACTGC CATGGGAAG	157	0.240	5.849	Делеция	-4	12
ATGGAGGAGTTGGCAGATGCTCTaAATGTACTG CCATGGGAAG	147	0.225	5.476	Вставка	+6	1
ATGGAGGAGTTGGCAGATGCTctTCTAATGTACT TGCCATGGGAAG	123	0.188	4.582	Вставка	+2	3
ATGGAGGAGTTGGCAGATGccCTCTAATGTACT GCCATGGGAAG	110	0.168	4.098	Вставка	+2	2
ATGGAGGAGTTGGCAGAT-----AATGTACTGC CATGGGAAGAAG	103	0.157	3.837	Делеция	+1	5
ATGG-----cGTACTGC CATGGGAAGAAG	96	0.147	3.57	Делеция & Мутация	-7	21
ATGGAGGAGTTGGCAGATGCTTCTAATGTACTG CCATGGGAAGAAG	85	0.130	3.166	Вставка	+3	1
ATGGAGGAGTTGGCA-----TCTGC CATGGGAAGAAG	84	0.128	3.129	Делеция	-2	13
ATGGAGGAGTTGGCAGATGCT---AATGTACTGC CATGGGAAGAAG	79	0.120	2.943	Делеция	+3	3
ATGGAGGAGTTGGCAGATGCc aaactgaaaaac aaatcaagcactcttattgagtgctggcgatc cccgacgccacggcgcaaaccttatcataga aaCTCTAATGTACTGCCATGGGAAG	58	0.0884	2.160	Вставка	+3	81
ATGGAGGAGTTGGCAGATGCTgcttatatagac ctcccaccgtacacgctaccgcccatttCTA ATGTACTGCCATGGGAAG	53	0.081	1.974	Вставка	+3	42
ATGGAGGAGTTG-----TCTAATGTACTGC CATGGGAAG	47	0.071	1.751	Делеция	-5	8
-----CTGC CATGGGAAGAAG	26	0.039	0.968	Делеция		
ATGGAGGAGTTGGCAGATGCTgcttatatagac tacagaaccgtggcgagatgtggatcaaggat gcTCTAATGTACTGCCATGGGAAG	21	0.032	0.782	Вставка	+3	48
ATGGAGGAGTTGGCAGATGCTAATGTACTGC CATGGGAAG	14	0.021	0.521	Делеция	+3	1
ATGGAGGAGTTGGCAGATGCTgctcatgcttt ttccgctcgtgctggacttgcctcagttctctg gccagctcgTCTAATGTACTGCCATGGGAAG	10	0.015	0.372	Вставка	+3	55
ATGGAGGAGTTGGCAGATGCTCT-ATGTACTGC CATGGGAAG	8	0.012	0.29	Делеция	+6	1

Таблица 10.6

Специальные вставки и делеции для управления 188 с использованием RGN APG08290.1

Управление 188 (SEQ ID NO: 154)	№ Считывания	% Считывания	% INDEL	Тип	Положение INDEL	Размер
CAGGGACAGTGC GCATCTCCCTGGTCACCAAG	59686	97.000				
CAGGGACA-----GTCACCAAG	1286	2.089	69.664	Делеция	0	15
CAGGGACAGTGC GCATCTC-CTGGTCACCAAG	473	0.768	25.622	Делеция	+3	1
CAGGGACAGTGC GCATCT--CTGGTCACCAAG	57	0.092	3.087	Делеция	+3	2
CAGGGACAGTGC GCATCTCCiCTGGTCACCAAG	11	0.017	0.595	Делеция	+3	1
CAGGGACAGTGC GCATC---CTGGTCACCAAG	7	0.011	0.379	Делеция	+3	3
CAGGGAC-----GGTCACCAAG	7	0.011	0.379	Делеция	+2	15
CGGGACAGgGC GCATCTC-CTGGTCACCAAG	5	0.008	0.270	Делеция & Мутация	+3	1

Таблица 10.7

Специальные вставки и делеции для управления 192 с использованием RGN APG08290.1

Управление 192 (SEQ ID NO: 155)	№ Считывания	% Считывания	% INDEL	Тип	Положение INDEL	Размер
CGACCTGAATGCTGTGCGGCTCTGCTTCCAGG	62352	95.658				
CGACCTGAATGCTGTGCGGCaTCTGCTTCCAGG	1262	1.936	44.593	Вставка	+3	1
CGACCTGAATGCTGTGCGGCtTCTGCTTCCAGG	842	1.291	29.752	Вставка	+3	1
CGACCTGAATGCTGTG----TCTGCTTCCAGG	686	1.052	24.240	Делеция	+3	4
CGACCTGcATGCTGTGCGGCaTCTGCTTCCAGG	18	0.027	0.636	Вставка & Мутация	+3	1
CGACCTGcATGCTGTGCGGCtTCTGCTTCCAGG	11	0.016	0.388	Вставка & Мутация	+3	1
CGACCTGcATGCTGTG----TCTGCTTCCAGG	6	0.009	0.212	Вставка & Мутация	+3	4
CGACCTGAATGCTGTGCGCaTCTGCTTCCAGG	5	0.007	0.176	Вставка & Мутация	+3	2

Таблица 10.8

Конкретные вставки и делеции для управления 196 с использованием RGN APG08290.1

Управление 196 (SEQ ID NO: 156)	№ Считывания	% Считывания	% INDEL	Тип	Положение INDEL	Размер
TGGGGACTACGACCTGAATGCTGTGCGGCTCT	37206	93.073				
TGGGGACTACGA-----ATGCTGTGCGGCTCT	1288	3.222	46.514	Делеция	+3	5
TGGGGACTACGAagcaggcagaagtatgcaaacgcatcatcattCCTGAATGCTGTGCGGCTCT	881	2.203	31.816	Вставка	+3	34
TGGGGACTACGAagaagcgatagaaggccatgcctgcgaatcgggagcggCCTGAATGCTGTGCGGCTCT	302	0.755	10.906	Вставка	+3	40
TGGGGACTACGAtgactegctgcgctcggtcgttcggctgcccgcgagcggtatcagctcactcaaaaggcgtaatacggCCTGAATGCTGTGCGGCTCT	272	0.680	9.823	Вставка	+3	67
TGGG-----GTGCGGCTCT	13	0.032	0.4694	Делеция	-5	18
TGGGGACTACGAC-----TGCTGTGCGGCTCT	13	0.032	0.469	Делеция	+4	5

Таблица 10.9

Конкретные вставки и делеции для управления 190 с использованием RGN APG07433.1

Управление 190 (SEQ ID NO: 146)	№ Считывания	% Считывания	% INDEL	Тип	Положение INDEL	Размер
CGACCTGAATGCTGTGC----- ----- ----- GGCGCTCTGGCTTCATTCAATC	64040	55.46	91.70	Делеция	-164	170
CGACCTGAATGCTGTGCGGCTCTGCTTCCAGGTGA	45619	39.51		WT		
CGACCTGAATGCTGTGCGGCaTCTGCTTCCAGGTGA	3620	3.13	5.18	Вставка	+3	1
CGACCTGAATG-----CTGCTTCCAGGTGA	1110	0.96	1.58	Делеция	+2	10
CGACCTGAATGCT-----TCTGCTTCCAGGTGA	858	0.74	1.22	Делеция	+3	7
CGACCTGAA-----TGCTTCCAGGTGA	206	0.17	0.29	Делеция	+1	13

Пример 7: Демонстрация активности редактирования генов в растительных клетках.

РНК-управляемая нуклеазная активность RGN, представленная в данном изобретении, демонстрируется в растительных клетках с использованием протоколов, адаптированных по Li, и др. (2013) Nat. Biotech. № 31: сс. 688-691. Вкратце, оптимизированные версии кодонов растений каждой RGN (SEQ ID NO: 169-182), содержащей N-концевой сигнал ядерной локализации SV40, клонируются за сильным конститутивным промотором 35S в векторе временной трансформации. sgРНК, нацеленные на один или несколько сайтов в гене PDS растения, фланкирующую подходящую последовательность PAM, клонируются за промотором U6 растения во втором векторе временной экспрессии. Экспрессионные векторы

вводят в протопласты мезофилла *Nicotiana benthamiana* с использованием ПЭГ-опосредованной трансформации. Трансформированные протопласты инкубируют в темноте до 36 ч. Геномную ДНК выделяют из протопластов с помощью набора DNeasy Plant Mini (Qiagen). Геномную область, фланкирующую сайт-мишень RGN, амплифицируют с помощью ПЦР, и продукты очищают с использованием колонки QiaQuick Spin (Qiagen) в соответствии с протоколом производителя. Всего 200-500 нг очищенных продуктов ПЦР смешивают с 1 мкл 10×Taq ДНК-полимеразного ПЦР-буфера (Enzymatics) и сверхчистой воды до конечного объема 10 мкл и подвергают процессу повторного отжига для образования гетеродуплекса: 95°C в течение 10 мин, от 95°C до 85°C с линейным изменением со скоростью -2°C/c, от 85°C до 25°C со скоростью -0,25°C/c и выдержкой 25°C в течение 1 мин. После повторного отжига продукты обрабатывают нуклеазой SURVEYOR и энхансером SURVEYOR S (Integrated DNA Technologies) в соответствии с протоколом, рекомендованным производителем, и анализируют на 4-20% полиакриламидных гелях Novex TBE (Life Technologies). Гели окрашивают красителем ДНК SYBR Gold (Life Technologies) в течение 10 мин и визуализируют с помощью системы визуализации геля Gel Doc (Bio-rad). Количественная оценка основана на относительной интенсивности полос. Процент Indel определяется по формуле $100 \times (1 - (1 - (b+c)/(a+b+c))^{1/2})$, где a - интегрированная интенсивность непереваренного продукта ПЦР, b и c - интегрированные интенсивности каждого продукта расщепления.

Кроме того, продукты ПЦР, полученные из целевой геномной последовательности, могут быть подвергнуты ПЦР, аналогичной описанной в примере 6, так что продукты ПЦР содержат выступающие последовательности Illumina и могут быть преобразованы в библиотеки и подвергнуты глубокому секвенированию. Этот метод позволяет определять скорость редактирования, как показано в табл. 9.

Пример 8: Перекрестная совместимость управления.

Для определения перекрестной совместимости управляющих РНК между RGN был проведен эксперимент по интерференции двух плазмид (Esvelt и др. (2013), Nat. Methods № 10 (11): сс. 1116-21). Первая плазида содержала RGN с несколькими мишенями, содержащими определенные РАМ на скелете устойчивом к канамицину. Эти плазмиды были трансформированы в *E.coli* BL21, и трансформированные штаммы были сделаны химически компетентными. Затем вводили вторую плазмиду, содержащую управляющую РНК на скелете устойчивости к ампициллину. Клетки высевали на среду, содержащую оба антибиотика. Если RGN может использовать указатель на второй плазмиде, плазида устойчивости к канамицину расщепляется и линейризуется, что приводит к небольшому образованию колоний или их отсутствию. Если RGN не может использовать управление на второй плазмиде, плазида устойчивости к канамицину не расщепляется, что приводит к образованию большого количества колоний. Управляющие РНК для *Streptococcus pyogenes* Cas9 (SpyCas9) и *Staphylococcus aureus* Cas9 (SauCas9) также использовались для определения перекрестной совместимости с этими управляющими РНК.

Для расчета процента истощения, количество колоний для каждой управляющей трансформации сравнивается с эффективностью трансформации с использованием положительного контроля. Основываясь на этом сравнении, если RGN может использовать управляющую, процент истощения должен быть равен 0, поскольку никакие колонии не могут выжить. Если RGN не может использовать управляющую, процент истощения должен быть равен 1, поскольку все плазмиды остаются нетронутыми. Результаты показаны в табл. 11 ниже, "sg" означает управляющую РНК для перечисленных RGN.

Таблица 11

Анализ перекрестной совместимости						
	APG05083.1	APG07513.1	APG08290.1	APG05459.1	APG04583.1	APG01688.1
sgAPG05083.1	0	0	0	0.21	1	0.74
sgAPG07433.1	0	0	0	0.16	0.78	0.33
sgAPG07513.1	0	0.01	0	0.32	0.97	0.64
sgAPG05459.1	0.24	0.53	0.26	0.09	1	0.49
sgAPG04583.1	0.74	0.8	0.36	0.21	0	0
sgAPG01688.1	0.12	0.26	0.18	0.43	0	0
sgSauCas9	1	0.23	0.27	0.53	0.51	0.92
sg Spy	0.16	0.27	0.32	0.06	1	1

Как показано в табл. 11, существует четыре группы ортогональных систем. RGN могут распознавать управляющие из других систем в своих группах, но не могут использовать управляющие из других групп. Первая группа содержит APG05083.1, APG07433.1, APG07513.1 и APG08290.1. Вторая группа - это SpyCas9 и APG05459.1. Третья группа содержит APG04583.1 и APG01688.1. Четвертая группа содержит SauCas9.

Пример 9: Идентификация целевого заболевания.

База данных клинических вариантов была получена из базы данных NCBI ClinVar, которая доступна в Интернете на веб-сайте NCBI ClinVar. Патогенные одиночные нуклеотидные полиморфизмы (SNP)

были идентифицированы из этого списка. Используя информацию о геномном локусе, были идентифицированы CRISPR мишени в области, перекрывающейся и окружающей каждый SNP. Выбор SNP, которые могут быть скорректированы с использованием базового редактирования в комбинации с RGN в рамках настоящего изобретения для нацеливания на причинную мутацию, приведен в табл. 12. В табл. 12 приведен только одно альтернативное название каждого заболевания. "RS #" соответствует номеру доступа RS в базе данных SNP на веб-сайте NCBI. AlleleID соответствует регистрационному номеру причинного аллеля, а регистрационный номер хромосомы также предоставляет справочную информацию о доступе, которую можно найти на веб-сайте NCBI. В табл. 12 также представлена информация о последовательности геномных мишеней, подходящая для RGN, перечисленных для каждого заболевания. Информация о целевой последовательности также содержит последовательность протоспейсера для продукции необходимой sgРНК для соответствующей RGN, представленной в данном изобретении.

Таблица 12

Мишени заболевания для RGN, представленные в предложенном изобретении

Заболевание	RS# (dbSNP)	RGN	Причинная мутация	Аллель ID	Присоединение хромосомы	Символ гена	Последовательность-мишень (SEQ ID NO)
Синдром атаксии-телеангиэктазии	1137887	APG04583.1	G>A	18083	NC_000011.10,NC_000011.9	ATM	197
Дефицит анил-КоА-дегидрогеназы с очень длинной цепью	2309689	APG05459.1	G>A	33868	NC_000017.10,NC_000017.11	ACADVL	198
Нарушение физиологии Т-клеток	3218716	APG01688.1	G>A	52071	NC_000014.8,NC_000014.9	MYH7	199
Сердечно-сосудистый фенотип	5742905	APG05083, APG07433.1, APG07513.1, APG08290.1	T>C	15159	NC_000021.8,NC_000021.9	CBS	200
Дефицит 3-оксо-5-альфа-стероид-дальта-4-дегидрогеназы	9332964	APG04583.1	G>A	18390	NC_000002.11,NC_000002.12	SRD5A2	201
Острый миелоидный лейкоз	11540652	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	27395	NC_000017.10,NC_000017.11	TP53	202
Острый миелоидный лейкоз	11540652	APG05459.1	G>A	27395	NC_000017.10,NC_000017.11	TP53	203
Кожная злокачественная меланома 3	11547328	APG05459.1	C>T	31967	NC_000012.11,NC_000012.12	CDK4	204
Дефицит альфа-1-антитрипсина	28929474	APG05459.1	G>A	33006	NC_000014.8,NC_000014.9	SERPINA1	205
Болезнь Шарко-Мари-Тута, тип 2	28933093	APG05459.1	G>A	29543	NC_000001.10,NC_000001.11	LMNA	206
Наследственный предрасполагающий к раку синдром	28934578	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	27413	NC_000017.10,NC_000017.11	TP53	207
Наследственный предрасполагающий к раку синдром	28934578	APG01688.1	G>A	27413	NC_000017.10,NC_000017.11	TP53	208
Наследственный предрасполагающий к раку синдром	28934872	APG05459.1	G>A	27436	NC_000016.9,NC_000016.10	TSC2	209
Синдром Бругада	28937316	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	24408	NC_000003.11,NC_000003.12	SCN5A	210
Синдром Бругада	28937318	APG05459.1	G>A	24429	NC_000003.11,NC_000003.12	SCN5A	211
ГРАЦИЛЬНЫЙ синдром	28937590	APG05459.1	A>G	21206	NC_000002.11,NC_000002.12	BCS1L	212
Синдром усиленного s-конуса	28937873	APG05459.1	G>A	20571	NC_000015.9,NC_000015.10	NR2E3	213
Болезнь Шарко-Мари-Тута, тип 2	28940293	APG05083, APG07433.1, APG07513.1, APG08290.1	T>C	17309	NC_000001.10,NC_000001.11	MFN2	214
Болезнь Шарко-Мари-Тута, тип 2	28940293	APG05459.1	T>C	17309	NC_000001.10,NC_000001.11	MFN2	215
Арилсульфатаза a, аллель a	28940893	APG05459.1	C>T	18091	NC_000022.10,NC_000022.11	ARSA	216
Наследственная гиперхолестеринемия	28942078	APG05459.1	G>A	18733	NC_000019.9,NC_000019.10	LDLR	217
Семейная гиперхолестеринемия	28942079	APG05459.1	G>A	18734	NC_000019.9,NC_000019.10	LDLR	218
ПАБК ГЕМОГЛОБИН АРЛИНГТОН	33930165	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	30165	NC_000011.9,NC_000011.10	HBB	219
Семейная гипертрофическая кардиомиопатия 1	36211715	APG05459.1	G>A	29159	NC_000014.8,NC_000014.9	MYH7	220
Сердечно-сосудистый фенотип	36211723	APG05459.1	G>A	45266	NC_000011.9,NC_000011.10	MYBPC3	221
Сердечно-сосудистый фенотип	36211723	APG01688.1	G>A	45266	NC_000011.9,NC_000011.10	MYBPC3	222
Синдром Бругада	45546039	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	48043	NC_000003.11,NC_000003.12	SCN5A	223
Синдром Бругада	45546039	APG01688.1	G>A	48043	NC_000003.11,NC_000003.12	SCN5A	224
Наследственный предрасполагающий к раку синдром	55863639	APG05459.1	G>A	176641	NC_000017.10,NC_000017.11	TP53	225

Дефицит бутирил-КоА дегидрогеназы	57443665	APG05459.1	T>C	18867	NC_000012.11,NC_000012.12	ACADS	226
Дефицит бутирил-КоА дегидрогеназы	57443665	APG01688.1	T>C	18867	NC_000012.11,NC_000012.12	ACADS	227
Доброкачественная лопаточно-перинсальная мышечная дистрофия с кардиомиопатией	59332535	APG05459.1	G>A	77828	NC_000001.10,NC_000001.11	LMNA	228
Доброкачественная лопаточно-перинсальная мышечная дистрофия с кардиомиопатией	60458016	APG05459.1	G>A	29564	NC_000001.10,NC_000001.11	LMNA	229
Конусно-стержневая дистрофия 6	61750173	APG05459.1	G>A	24396	NC_000017.10,NC_000017.11	GUCY2D	230
Конусно-стержневая дистрофия 6	61750173	APG01688.1	G>A	24396	NC_000017.10,NC_000017.11	GUCY2D	231
Болезнь Штаргардта 1	61750641	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	105317	NC_000001.10,NC_000001.11	ABCA4	232
Врожденный амавроз Лебера 2	61751276	APG05459.1	G>A	104715	NC_000001.10,NC_000001.11	RPE65	233
Конусно-стержневая дистрофия 3	61751407	APG05459.1	G>A	105292	NC_000001.10,NC_000001.11	ABCA4	234
Несиндромный окулокожный альбинизм	61754375	APG05459.1	G>A	18835	NC_000011.9,NC_000011.10	TYR	235
Фенилкетонурия	62508646	APG05459.1	T>C	15654	NC_000012.11,NC_000012.12	PAH	236
Фенилкетонурия	62516101	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	15658	NC_000012.11,NC_000012.12	PAH	237
Рак груди и яичников, семейный 1	62625303	APG05459.1	C>T	68931	NC_000017.10,NC_000017.11	BRCA1	238
Гиперфенилаланиемия, неприкосновенная	62644499	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	15656	NC_000012.11,NC_000012.12	PAH	239
Гиперфенилаланиемия, неприкосновенная	62644499	APG05459.1	G>A	15656	NC_000012.11,NC_000012.12	PAH	240
Наследственный	63750217	APG05459.1	G>A	32138	NC_000003.11,NC_000003.12	MLH1	241
предрасполагающий к раку синдром							
Наследственный предрасполагающий к раку синдром	63750741	APG05083, APG07433.1, APG07513.1, APG08290.1	T>C	94663	NC_000002.11,NC_000002.12	MSH6	242
Наследственный предрасполагающий к раку синдром	63750809	APG05459.1	T>C	95480	NC_000003.11,NC_000003.12	MLH1	243
Наследственный предрасполагающий к раку синдром	63751657	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	95331	NC_000003.11,NC_000003.12	MLH1	244
Наследственный предрасполагающий к раку синдром	63751711	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	95792	NC_000003.11,NC_000003.12	MLH1	245
Наследственный предрасполагающий к раку синдром	63751711	APG01688.1	G>A	95792	NC_000003.11,NC_000003.12	MLH1	246
Дисгенезия переднего сегмента 6	72549387	APG05459.1	G>A	22776	NC_000002.11,NC_000002.12	CYP11B1	247
Синдром Бругада	72549410	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	78547	NC_000003.11,NC_000003.12	SCN5A	248
Синдром Бругада	72549410	APG05459.1	G>A	78547	NC_000003.11,NC_000003.12	SCN5A	249
Дефицит орнитинкарбамоилтрансферазы	72554308	APG01688.1	G>A	26053	NC_000023.10,NC_000023.11	OTC	250
Несовершенный остеогенез I типа	72645321	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	414022	NC_000017.10,NC_000017.11	COL1A1	251
Несовершенный остеогенез I типа	72645321	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	414022	NC_000017.10,NC_000017.11	COL1A1	252

Запор	74799832	APG05083, APG07433.1, APG07513.1, APG08290.1	T>C	28958	NC_000010.10,NC_000010.11	RET	253
Дефицит дофамин-бета-гидроксилазы	74853476	APG05083, APG07433.1, APG07513.1, APG08290.1	T>C	16789	NC_000009.11,NC_000009.12	DBH	254
Муковисцидоз	75096551	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	33858	NC_000007.13,NC_000007.14	CFTR	255
Фенилкетонурия	75193786	APG01688.1	T>C	15675	NC_000012.11,NC_000012.12	PAH	256
Дефицит УДФ-глюкозо-гексозо-1-фосфатуридилтрансферазы	75391579	APG05459.1	A>G	18653	NC_000009.11,NC_000009.12	GALT	257
Амилоидная кардиомиопатия, связанная с транстиретином	76992529	APG05459.1	G>A	28465	NC_000018.9,NC_000018.10	TTR	258
Синдром дефицита углеводов гликопротеинов типа I	80338707	APG01688.1	G>A	22758	NC_000016.9,NC_000016.10	PMM2	259
Метахроматическая лейкодистрофия	80338815	APG01688.1	G>A	18090	NC_000022.10,NC_000022.11	ARSA	260
Синдром Смита-Лемли-Опица	80338857	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	34128	NC_000011.9,NC_000011.10	DHCR7	261
Глухота аутосомно-рецессивная 1А	80338940	APG05459.1	G>A	32068	NC_000013.10,NC_000013.11	GJB2	262
Врожденное омфалоцеле	80338945	APG05083, APG07433.1, APG07513.1, APG08290.1	T>C	32055	NC_000013.10,NC_000013.11	GJB2	263
Врожденное омфалоцеле	80338945	APG05083, APG07433.1, APG07513.1, APG08290.1	T>C	32055	NC_000013.10,NC_000013.11	GJB2	264
Врожденное омфалоцеле	80338945	APG05459.1	T>C	32055	NC_000013.10,NC_000013.11	GJB2	265
Врожденное омфалоцеле	80338945	APG05459.1	T>C	32055	NC_000013.10,NC_000013.11	GJB2	266
Врожденная миотония, аутосомно-доминантная форма	80356701	APG05083, APG07433.1, APG07513.1, APG08290.1	T>C	33902	NC_000007.13,NC_000007.14	CLCN1	267
Рак груди и яичников, семейный 1	80356914	APG05459.1	G>A	70276	NC_000017.10,NC_000017.11	BRCA1	268
Рак груди и / или яичников	80356962	APG05459.1	G>A	70247	NC_000017.10,NC_000017.11	BRCA1	269
Рак груди и яичников, семейный 1	80357212	APG05459.1	G>A	70255	NC_000017.10,NC_000017.11	BRCA1	270
Рак груди и яичников, семейный 1	80357281	APG05459.1	T>C	70177	NC_000017.10,NC_000017.11	BRCA1	271
Рак груди и яичников, семейный 1	80357307	APG05459.1	G>A	70275	NC_000017.10,NC_000017.11	BRCA1	272
Рак груди и яичников, семейный 1	80357352	APG05459.1	C>T	69958	NC_000017.10,NC_000017.11	BRCA1	273
Рак груди и яичников, семейный 1	80358145	APG05459.1	G>A	46229	NC_000017.10,NC_000017.11	BRCA1	274
Врожденные генетические заболевания	80358259	APG05083, APG07433.1, APG07513.1, APG08290.1	T>C	18006	NC_000018.9,NC_000018.10	NPC1	275
Рак груди и яичников, семейный 2	80358543	APG05459.1	G>A	131539	NC_000013.10,NC_000013.11	BRCA2	276
Рак груди и яичников, семейный 2	80358544	APG05459.1	G>A	46368	NC_000013.10,NC_000013.11	BRCA2	277
Рак груди и яичников, семейный 2	80358997	APG05459.1	G>A	67062	NC_000013.10,NC_000013.11	BRCA2	278
Рак груди и / или яичников	80359003	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	67069	NC_000013.10,NC_000013.11	BRCA2	279
Рак груди и яичников, семейный 2	80359004	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	46672	NC_000013.10,NC_000013.11	BRCA2	280
Рак груди и яичников,	80359071	APG05459.1	G>A	67203	NC_000013.10,NC_000013.11	BRCA2	281

семейный 2							
Рак груди и яичников, семейный 2	80359112	APG05459.1	C>T	67292	NC_000013.10,NC_000013.11	BRCA2	282
Рак груди и яичников, семейный 2	80359115	APG05459.1	C>T	67294	NC_000013.10,NC_000013.11	BRCA2	283
Синдром Смита-Лемли-Опица	104886033	APG05459.1	A>G	21833	NC_000011.9,NC_000011.10	DHCR7	284
Синдром Альпорта 1, X-сцепленный рецессивный	104886142	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	35796	NC_000023.10,NC_000023.11	COL4A5	285
Острая нейропатическая болезнь Гоше	104886460	APG05459.1	G>A	99352	NC_000001.10,NC_000001.11	GBA	286
Дефицит гонадотропина	104893836	APG05459.1	A>G	31062	NC_000004.11,NC_000004.12	GNRHR	287
Дистальный артрогрипоз типа 1A	104894129	APG05459.1	G>A	27501	NC_000009.11,NC_000009.12	TPM2	288
Дистальный артрогрипоз типа 1A	104894129	APG05459.1	G>A	27501	NC_000009.11,NC_000009.12	TPM2	289
Наследственный предрасполагающий к раку синдром	104894261	APG05459.1	C>T	31727	NC_000011.9,NC_000011.10	MEN1	290
Врожденные генетические заболевания	104894313	APG05459.1	C>T	18816	NC_000011.9,NC_000011.10	TYR	291
Смерть в раннем взрослом возрасте	104894368	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	29104	NC_000012.11,NC_000012.12	MYL2	292
Смерть в раннем взрослом возрасте	104894368	APG05459.1	G>A	29104	NC_000012.11,NC_000012.12	MYL2	293
Тяжелая аутосомно-рецессивная мышечная дистрофия детского возраста - североафриканский тип	104894423	APG05459.1	G>A	17048	NC_000013.10,NC_000013.11,NC_000013.9	SGCG	294
Сердечно-сосудистый фенотип	104894503	APG05459.1	G>A	27495	NC_000015.9,NC_000015.10	TPM1	295
Синдром дефицита углеводов гликопротеинов типа I	104894525	APG01688.1	G>A	22747	NC_000016.9,NC_000016.10	PMM2	296
Болезнь Шарко-Мари-Тута, тип I	104894621	APG05459.1	C>T	23472	NC_000017.10,NC_000017.11	PMP22	297
Врожденные генетические заболевания	104894635	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	20146	NC_000017.10,NC_000017.11	SGSH	298
Врожденные генетические заболевания	104894635	APG05459.1	G>A	20146	NC_000017.10,NC_000017.11	SGSH	299
Семейная средиземноморская лихорадка	104895097	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	17588	NC_000016.9,NC_000016.10	MEFV	300
Глухота аутосомно-рецессивная 2	111033178	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	52388	NC_000011.9,NC_000011.10	MYO7A	301
Глухота аутосомно-рецессивная 2	111033178	APG01688.1	G>A	52388	NC_000011.9,NC_000011.10	MYO7A	302
Глухота, X-сцепленная 2	111033299	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	53902	NC_000013.10,NC_000013.11	GJB2	303
Увеличенный вестибулярный акведук	111033305	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	52666	NC_000007.13,NC_000007.14	SLC26A4	304
Врожденное сенсорное нарушение слуха	111033364	APG05459.1, APG01688.1	G>A	17396	NC_000001.10,NC_000001.11	USH2A	305
Дефицит УДФ-глюкозо-6-фосфатуредилтрансферазы	111033728	APG05459.1	T>C	36556	NC_000009.11,NC_000009.12	GALT	306
Дефицит ацил-КоА-дегидрогеназы с очень длинной цепью	112406105	APG05459.1	G>A	200333	NC_000017.10,NC_000017.11	ACADVL	307
Сердечно-сосудистый фенотип	112645512	APG05459.1	C>T	178700	NC_000015.10,NC_000015.9	FBN1	308
Дефицит пируваткиназы эритроцитов	113403872	APG05459.1	G>A	16550	NC_000001.10,NC_000001.11	PKLR	309

Аутосомно-доминантная прогрессирующая наружная офтальмоплегия с делециями митохондриальной ДНК I	113994095	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	28535	NC_000015.9,NC_000015.10	POLG	310
Дефицит ацил-КоА-дегидрогеназы с очень длинной цепью	113994167	APG05459.1	T>C	33877	NC_000017.10,NC_000017.11	ACADVL	311
Цистиноз глазной нефропатический	113994205	APG05459.1	G>A	19482	NC_000017.10,NC_000017.11	CTNS	312
Дефицит пируваткиназы эритроцитов	116100695	APG05459.1	C>T	16552	NC_000001.10,NC_000001.11	PKLR	313
Дистальная миопатия по типу Татеямы	116840778	APG01688.1	G>A	23322	NC_000003.11,NC_000003.12	CAV3;SSUH2	314
Злокачественная гипертермия, восприимчивость к, I	118192122	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	76888	NC_000019.9,NC_000019.10	RYR1	315
Злокачественная гипертермия, восприимчивость к, I	118192122	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	76888	NC_000019.9,NC_000019.10	RYR1	316
Миопатия, центральное ядро	118192158	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	76835	NC_000019.9,NC_000019.10	RYR1	317
Миопатия, центральное ядро	118192158	APG05459.1	G>A	76835	NC_000019.9,NC_000019.10	RYR1	318
Миопатия, центральное ядро	118192158	APG01688.1	G>A	76835	NC_000019.9,NC_000019.10	RYR1	319
Злокачественная гипертермия, восприимчивость к, I	118192170	APG05083, APG07433.1, APG07513.1, APG08290.1	T>C	28014	NC_000019.9,NC_000019.10	RYR1	320
Цероидный липофуциноз нейрональный 2	119455954	APG05459.1	G>A	17681	NC_000011.9,NC_000011.10	TPP1	321
Цероидный липофуциноз нейрональный 2	119455954	APG01688.1	G>A	17681	NC_000011.9,NC_000011.10	TPP1	322
Глутаровая ацидурия, тип I	120074135	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	18010	NC_000018.9,NC_000018.10	NPC1	323
Заболевания, связанные с CAPN3	121434372	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	17127	NC_000019.9,NC_000019.10	GCDH	324
Заболевания, связанные с CAPN3	121434548	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	32661	NC_000015.9,NC_000015.10	CAPN3;POMT1	325
Болезнь накопления гликогена, тип II	121434548	APG05459.1	G>A	32661	NC_000015.9,NC_000015.10	CAPN3;POMT1	326
Несиндромный окулокожный альбинизм	121907943	APG05459.1	C>T	19073	NC_000017.10,NC_000017.11	GAA	327
Семейная гиперхолестеринемия	121908011	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	18814	NC_000011.9,NC_000011.10	TYR	328
Семейная гиперхолестеринемия	121908033	APG05459.1	G>A	18765	NC_000019.9,NC_000019.10	LDLR	329
Глухота аутосомно-рецессивная 7	121908039	APG05459.1	G>A	18778	NC_000019.9,NC_000019.10	LDLR	330
Хронический детский неврологический, кожный и суставной синдром	121908073	APG05459.1	C>T	19142	NC_000009.11,NC_000009.12	TMC1	331
Врожденная мышечная дистрофия по типу Эйхсфельда	121908153	APG05459.1	G>A	19416	NC_000001.10,NC_000001.11	NLRP3	332
Врожденные генетические заболевания	121908185	APG05459.1	G>A	19531	NC_000001.10,NC_000001.11	SELENON	333
Гиперкалисический периодический паралич I типа	121908192	APG05459.1	G>A	23730	NC_000016.9,NC_000016.10	GFER	334
Миопатия с тельцами включения 2	121908557	APG05459.1	G>A	20958	NC_000017.10,NC_000017.11	SCN4A	335
Тяжелый иммунодефицит APG05083.1, APG07433.1,	121908627	APG05459.1	G>A	21067	NC_000009.11,NC_000009.12	GNE	336

APG07513.1, APG08290.1 из-за дефицита ADA								
Тяжелый иммунодефицит APG05083.1, APG07433.1, APG07513.1, APG08290.1 из-за дефицита ADA	121908716	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	16996	NC_000020.10, NC_000020.11	ADA	337	
Глутаровая ацидурия, тип 1	121908739	APG05083, APG07433.1, APG07513.1, APG08290.1	T>C	17004	NC_000020.10, NC_000020.11	ADA	338	
Адренокортикальная карцинома в педиатрии	121908987	APG05459.1	G>A	21885	NC_000007.13, NC_000007.14	PRKAG2	339	
Дефицит фумаразы	121909019	APG05459.1	G>A	22197	NC_000007.13, NC_000007.14	CFTR	340	
Аденокарцинома простаты	121909036	APG05459.1	T>C	22247	NC_000007.13, NC_000007.14	CFTR	341	
Семейная гипертрофическая кардиомиопатия 1	121912664	APG01688.1	G>A	27418	NC_000017.10, NC_000017.11	TP53	342	
Гипофосфатазия взрослых	121913123	APG05459.1	G>A	31275	NC_000001.10, NC_000001.11	FH	343	
Гипофосфатазия взрослых	121913272	APG05459.1	T>C	40610	NC_000003.11, NC_000003.12	PIK3CA	344	
Гипофосфатазия взрослых	121913638	APG05459.1	G>A	29144	NC_000014.8, NC_000014.9	MYH7	345	
Врожденные генетические заболевания	121918007	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	28709	NC_000001.10, NC_000001.11	ALPL	346	
Врожденные генетические заболевания	121918007	APG01688.1	G>A	28709	NC_000001.10, NC_000001.11	ALPL	347	
Синдром Крузона	121918007	APG01688.1	G>A	28709	NC_000001.10, NC_000001.11	ALPL	348	
Дефицит пропионил-КоА карбоксилазы	121918166	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	15994	NC_000015.9, NC_000015.10	OCA2	349	
Сердечно-сосудистый фенотип	121918243	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	16464	NC_000001.10, NC_000001.11	MMACHC	350	
Мультиплексный дизостоз	121918505	APG05459.1	T>C	28329	NC_000010.10, NC_000010.11	FGFR2	351	
Множественный дефицит сульфатазы	121964961	APG05459.1	A>G	27057	NC_000003.11, NC_000003.12	PCCB	352	
Бифункциональная недостаточность пероксисомальных ферментов	121964962	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	15156	NC_000021.8, NC_000021.9	CBS	353	
Бифункциональная недостаточность пероксисомальных ферментов	121965019	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	26947	NC_000004.11, NC_000004.12	IDUA	354	
Наследственный предрасполагающий к раку синдром	137852850	APG05459.1	T>C	17711	NC_000003.11, NC_000003.12	SUMF1	355	
Сердечно-сосудистый фенотип	137853096	APG05459.1	G>A	22694	NC_000005.9, NC_000005.10	HSD17B4	356	
Сердечно-сосудистый фенотип	137853096	APG01688.1	G>A	22694	NC_000005.9, NC_000005.10	HSD17B4	357	
Конечностно-поясная мышечная дистрофия 2L типа	137853293	APG05459.1	C>T	28112	NC_000013.10, NC_000013.11	RB1	358	
Адренокортикальная карцинома в педиатрии	137854478	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	31496	NC_000015.9, NC_000015.10	FBN1	359	
Дефицит фумаразы	137854478	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	31496	NC_000015.9, NC_000015.10	FBN1	360	
Аденокарцинома простаты	137854529	APG05459.1	C>T	17205	NC_000011.9, NC_000011.10	ANO5	361	
Семейная гиперхолестеринемия	137929307	APG01688.1	G>A	171217	NC_000019.9, NC_000019.10	LDLR	362	
Синдром плавучей гавани	141659620	APG05459.1	G>A	21858	NC_000016.9, NC_000016.10	SPG7	363	
Врожденный синдром удлиненного интервала QT	142761835	APG05459.1	G>A	177782	NC_000015.9, NC_000015.10	IVD	364	
Врожденный синдром удлиненного интервала QT	145787161	APG05459.1	G>A	18783	NC_000019.10, NC_000019.9	LDLR	365	
Синдром Андерсена Тавила	146015592	APG05459.1	G>A	46845	NC_000003.11, NC_000003.12	BTD	366	
Сердечно-сосудистый фенотип	146015592	APG05459.1	G>A	46845	NC_000003.11, NC_000003.12	BTD	367	

Семейная гиперхолестеринемия	150726175	APG01688.1	G>A	45795	NC_000001.10,NC_000001.11	NMNAT1	368
Сердечно-сосудистый фенотип		APG05083, APG07433.1, APG07513.1, APG08290.1					
	151344623		G>A	24127	NC_000011.9,NC_000011.10	ABCC8	369
Опухоль стромы желудочно-кишечного тракта		APG05083, APG07433.1, APG07513.1, APG08290.1					
	180177122		G>A	132185	NC_000016.10,NC_000016.9	PALB2	370
Врожденный дискератоз	180177366	APG05459.1	G>A	71322	NC_000008.10,NC_000008.11	VPS13B	371
Врожденный дискератоз		APG05083, APG07433.1, APG07513.1, APG08290.1					
	187830361		T>C	45267	NC_000011.9,NC_000011.10	MYBPC3	372
Болезнь накопления гликогена IIIa	193922103	APG05459.1	A>G	44370	NC_000013.10,NC_000013.11	ATP7B	373
Дилатационная кардиомиопатия 1DD		APG05083, APG07433.1, APG07513.1, APG08290.1					
	193922110		G>A	44393	NC_000013.10,NC_000013.11	ATP7B	374
Почечный дефект транспорта карнитина	193922110	APG05459.1	G>A	44393	NC_000013.10,NC_000013.11	ATP7B	375
Синдром Барайцера-Винтера 1	193922566	APG05459.1	G>A	45113	NC_000019.9,NC_000019.10	LDLR	376
Дефицит ацил-КоА-дегидрогеназы с очень длинной цепью	193922566	APG05459.1	G>A	45113	NC_000019.9,NC_000019.10	LDLR	377
Семейная гиперхолестеринемия	199469464	APG05459.1	C>T	39865	NC_000016.9,NC_000016.10	SRCAP	378
Конечностно-поясная мышечная дистрофия 2A типа		APG05083, APG07433.1, APG07513.1, APG08290.1					
	199472712		G>A	67758	NC_000011.9,NC_000011.10	KCNQ1	379
Семейная гиперхолестеринемия	199472712	APG05459.1	G>A	67758	NC_000011.9,NC_000011.10	KCNQ1	380
Аневризма аорты семейная грудная 6	199473384	APG01688.1	G>A	78481	NC_000017.10,NC_000017.11	KCNJ2	381
Семейная гиперхолестеринемия		APG05083, APG07433.1, APG07513.1, APG08290.1					
	199473460		T>C	67776	NC_000011.9,NC_000011.10	KCNQ1	382
Синдром плавучей гавани	200238879	APG05459.1	T>C	18777	NC_000019.9,NC_000019.10	LDLR	383
Врожденный синдром удлинённого интервала QT		APG05083, APG07433.1, APG07513.1, APG08290.1					
	200411226		G>A	174776	NC_000011.9,NC_000011.10	MYBPC3	384
Врожденный синдром удлинённого интервала QT	201286421	APG05459.1	C>T	50215	NC_000001.10,NC_000001.11	SDHC	385
Синдром Андерсена Тавила	201540674	APG05459.1	G>A	51186	NC_000020.10,NC_000020.11	RTEL1	386
Сердечно-сосудистый фенотип	201540674	APG01688.1	G>A	51186	NC_000020.10,NC_000020.11	RTEL1	387
Семейная гиперхолестеринемия	267606640	APG04583.1	G>A	16147	NC_000001.10,NC_000001.11	AGL	388
Сердечно-сосудистый фенотип		APG05083, APG07433.1, APG07513.1, APG08290.1					
	267607004		G>A	15310	NC_000010.10,NC_000010.11	RBM20	389
Опухоль стромы желудочно-кишечного тракта	267607054	APG05459.1	C>T	21466	NC_000005.9,NC_000005.10	SLC22A5	390
Врожденный дискератоз		APG05083, APG07433.1, APG07513.1, APG08290.1					
	281875534		G>A	38553	NC_000007.13,NC_000007.14	ACTB	391
Врожденный дискератоз		APG05083, APG07433.1, APG07513.1, APG08290.1					
	369560930		G>A	98197	NC_000017.10,NC_000017.11	ACADVL	392
Болезнь накопления гликогена IIIa	373822756	APG05459.1	A>G	181233	NC_000019.9,NC_000019.10	LDLR	393
Дилатационная кардиомиопатия 1DD	376107921	APG05459.1	G>A	213634	NC_000015.9,NC_000015.10	CAPN3	394
Почечный дефект транспорта карнитина		APG05083, APG07433.1, APG07513.1, APG08290.1					
	376459828		G>A	198012	NC_000019.10,NC_000019.9	LDLR	395

Синдром Барайцера-Винтера 1	387906592	APG05459.1	G>A	38552	NC_000010.10,NC_000010.11	ACTA2	396
Акромикрическая дисплазия	387906623	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	38652	NC_000015.9,NC_000015.10	FBN1	397
Болезнь Шарко-Мари-Тута 2С типа	387906905	APG01688.1	G>A	39430	NC_000012.11,NC_000012.12	TRPV4	398
Рак груди и яичников, семейный 2	397507389	APG01688.1	G>A	46666	NC_000013.10,NC_000013.11	BRCA2	399
Рак груди и яичников, семейный 1	397509284	APG05459.1	G>A	70248	NC_000017.10,NC_000017.11	BRCA1	400
Болезнь Шарко-Мари-Тута тип 2С	397514494	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	48018	NC_000012.11,NC_000012.12	TRPV4	401
Болезнь Шарко-Мари-Тута тип 2С	397514494	APG01688.1	G>A	48018	NC_000012.11,NC_000012.12	TRPV4	402
Наследственный предрасполагающий к раку синдром	397514495	APG05459.1	G>A	152034	NC_000017.10,NC_000017.11	TP53	403
Ранняя детская эпилептическая энцефалопатия	397514581	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	48359	NC_000020.10,NC_000020.11	KCNQ2	404
Ранняя детская эпилептическая энцефалопатия	397514581	APG05459.1	G>A	48359	NC_000020.10,NC_000020.11	KCNQ2	405
Ранняя детская эпилептическая энцефалопатия	397514581	APG01688.1	G>A	48359	NC_000020.10,NC_000020.11	KCNQ2	406
Акромикрическая дисплазия	397515757	APG05459.1	G>A	51454	NC_000015.9,NC_000015.10	FBN1	407
Гипертрофическая кардиомиопатия	397515982	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	51820	NC_000011.9,NC_000011.10	MYBPC3	408
Сердечно-сосудистый фенотип	397516031	APG04583.1	G>A	51898	NC_000011.9,NC_000011.10	MYBPC3	409
Сердечно-сосудистый фенотип	397516074	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	51962	NC_000011.9,NC_000011.10	MYBPC3	410
Сердечно-сосудистый фенотип	397516083	APG01688.1	G>A	51977	NC_000011.9,NC_000011.10	MYBPC3	411
Семейная гипертрофическая кардиомиопатия 1	397516269	APG05083, APG07433.1, APG07513.1, APG08290.1	T>C	52276	NC_000014.8,NC_000014.9	MYH7	412
Синдром атаксии-телеангиэктазии	397517889	APG05459.1	C>T	57195	NC_000001.10,NC_000001.11	LMNA	413
Семейный рак груди	398123172	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	415590	NC_000017.10,NC_000017.11	GAA	414
Конечно-поясная мышечная дистрофия 2А типа	587776576	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	18532	NC_000011.10,NC_000011.9	WT1	415
Наследственный предрасполагающий к раку синдром	587776954	APG05459.1	A>G	51108	NC_000012.11,NC_000012.12	C12orf57	416
Асимметричная гипертрофия перегородки	587779826	APG05459.1	T>C	132814	NC_000011.10,NC_000011.9	A1M	417
Наследственный предрасполагающий к раку синдром	587780226	APG05459.1	C>T	133611	NC_000017.10,NC_000017.11	BRIP1	418
Семейная гипертрофическая кардиомиопатия 2	587780290	APG01688.1	G>A	134019	NC_000015.9,NC_000015.10	CAPN3	419
Семейная гипертрофическая кардиомиопатия 2	587781462	APG05459.1	C>T	150772	NC_000002.11,NC_000002.12	MSH6	420
Семейная гипертрофическая кардиомиопатия 2	587782958	APG01688.1	G>A	165560	NC_000011.10,NC_000011.9	MYBPC3	421
Эритроцитоз, семейный, 2	587783050	APG05459.1	G>A	166264	NC_000016.10,NC_000016.9	CDH1	422
Смерть в младенчестве	727504247	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	172354	NC_000001.10,NC_000001.11	TNNT2	423

Мышечные заболевания	727504247	APG01688.1	G>A	172354	NC_000001.10,NC_000001.11	TNNT2	424
Семейная гиперхолестеринемия	727504247	APG01688.1	G>A	172354	NC_000001.10,NC_000001.11	TNNT2	425
Семейный рак груди	730882035	APG01688.1	G>A	180121	NC_000003.12,NC_000003.11	VHL	426
Наследственный предрасполагающий к раку синдром	730882246	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	181441	NC_000014.9,NC_000014.8	ISCA2	427
Семейная гиперхолестеринемия	751995154	APG05459.1	G>A	200340	NC_000017.10,NC_000017.11	ACADVL	428
Семейная гиперхолестеринемия	756039188	APG04583.1	G>A	243266	NC_000019.9,NC_000019.10	LDLR	429
Семейная гиперхолестеринемия	761494650	APG05459.1	C>T	185659	NC_000022.10,NC_000022.11	CHEK2	430
Миопатия с тельцами включения 2	762307622	APG01688.1	G>A	232266	NC_000001.10,NC_000001.11	MUTYH	431
Синдром атаксин-телеангиэктазии	769370816	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	228176	NC_000019.10,NC_000019.9	LDLR	432
Синдром атаксин-телеангиэктазии	775092314	APG05083, APG07433.1, APG07513.1, APG08290.1	T>C	228197	NC_000019.9,NC_000019.10	LDLR	433
Семейный рак груди	775924858	APG05083, APG07433.1, APG07513.1, APG08290.1	G>A	246116	NC_000019.9,NC_000019.10	LDLR	434
Конечно-поясная мышечная дистрофия 2А типа	779694939	APG04583.1	T>C	214934	NC_000009.12,NC_000009.11	GNE	435
Наследственный предрасполагающий к раку синдром	780619951	APG05459.1	C>T	212851	NC_000011.10,NC_000011.9	ATM	436
Доброкачественные семейные неонатально-младенческие судороги	794727152	APG04583.1	G>A	191718	NC_000002.11,NC_000002.12	SCN2A	437
Синдром Марфана / Синдром Лойса-Дитца / Семейные аневризмы и расслоения грудной аорты	794728228	APG05459.1	C>T	197690	NC_000015.10,NC_000015.9	FBN1	438
Дилатационная кардиомиопатия 1G	869320740	APG01688.1	T>C	136355	NC_000002.11,NC_000002.12	TTN	439
Семейная гиперхолестеринемия	875989911	APG05459.1	G>A	228151	NC_000019.9,NC_000019.10	LDLR	440
Рак груди и яичников, семейный 2	876657678	APG05459.1	C>T	230443	NC_000013.10,NC_000013.11	BRCA2	441
Семейная гиперхолестеринемия	879254600	APG05459.1	G>A	245669	NC_000019.10,NC_000019.9	LDLR	442
Семейная гиперхолестеринемия	879254803	APG05083, APG07433.1, APG07513.1, APG08290.1	T>C	246008	NC_000019.10,NC_000019.9	LDLR	443
Семейная гиперхолестеринемия	879254803	APG01688.1	T>C	246008	NC_000019.10,NC_000019.9	LDLR	444
Семейная гиперхолестеринемия	879254849	APG01688.1	T>C	246074	NC_000019.10,NC_000019.9	LDLR	445
Семейный рак груди	1057517585	APG01688.1	G>A	358911	NC_000016.10,NC_000016.9	PALB2	446
Наследственная геморрагическая телеангиэктазия 2 типа	1057517944	APG05459.1	C>T	360048	NC_000012.11,NC_000012.12	ACVRL1	447

Пример 10: Целевые мутации, ответственные за синдром Гурлера.

Далее описывается потенциальное лечение синдрома Гурлера, также называемого MPS-1, с использованием системы редактирования оснований, направленной на РНК, которая исправляет мутацию, ответственную за синдром Гурлера, у значительной части пациентов с этим заболеванием. В этом подходе используется базовый слитый белок для редактирования, который управляется РНК и который может быть упакован в один вектор AAV для доставки в широкий спектр типов тканей. В зависимости от точных регуляторных элементов и используемого домена редактора базового блока также можно сконструировать единый вектор, который кодирует как базовый слитый белок редактирования, так и единственную управляющую РНК для нацеливания на пораженный локус. Пример 10.1: Определение RGN с идеальной РАМ Генетическое заболевание MPS-1 представляет собой нарушение лизосомального хранения, характеризующуюся на молекулярном уровне накоплением дерматансульфата и гепарансульфата в лизосомах. Это заболевание обычно является наследственным генетическим заболеванием, вызванным мутациями в гене IDUA (эталонная последовательность NCBI NG_008103.1), который кодирует α -L-идуронидазу. Заболевание является результатом дефицита α -L-идуронидазы. Наиболее частыми мутациями IDUA, обнаруженными в исследованиях на лицах северо-европейского происхождения, являются W402X и Q70X, обе нонсенс мутации, приводящие к преждевременному прекращению трансляции (Bunge и др. (1994), Hum. Mol. Genet, 3 (6): сс. 861- 866, включено здесь в качестве ссылки). Реверсирование одного нуклеотида может восстановить кодирующую последовательность дикого типа и привести к экспрессии белка, контролируемой эндогенными регуляторными механизмами генетического локуса.

Мутация W402X гена Idua человека является причиной высокой доли случаев MPS-1H. Базовые редакторы могут нацеливаться на узкое окно последовательности по отношению к сайту связывания компонента протоспейсера управляющей РНК, и, таким образом, наличие последовательности РАМ на определенном расстоянии от целевого локуса является важным для успеха стратегии. Учитывая ограничения, что целевая мутация должна находиться на открытой нецелевой цепи (NTS) во время взаимодействия с базовым редактирующим белком и что след домена RGN будет блокировать доступ к области рядом с РАМ, доступный локус может составлять 10-30 п.н. РАМ. Чтобы избежать редактирования и мута-

гене́за других близлежащих оснований аденозина в этом окне, проверяются различные линкеры. Идеальное окно составляет 12-16 п.о. PAM.

Последовательность PAM, совместимая с APG07433.1 и APG08290.1, легко обнаруживается в генетическом локусе и в пределах идеального окна редактирования основания, как определено выше. Эти нуклеазы имеют PAM-последовательность NNNNCC (SEQ ID NO: 6) и NNRNCC (SEQ ID NO: 32), соответственно, и компактны по размеру, что потенциально обеспечивает доставку с помощью одного вектора AAV. Такой подход к доставке дает множество преимуществ по сравнению с другими, например доступ к широкому спектру тканей (печень, мышцы, ЦНС), а также хорошо зарекомендовавший себя профиль безопасности и технологии производства.

Cas9 из *S. pyogenes* (SpyCas9) требует PAM-последовательности NGG (SEQ ID NO: 448), которая присутствует рядом с локусом W402X, но размер SpyCas9 предотвращает упаковку гена, кодирующего гибридный белок базового редактирующего домена и нуклеазу SpyCas9, в единый вектор AAV и, таким образом, лишает этот подход вышеупомянутых преимуществ. Включение последовательности, кодирующей управляющую РНК, в этот вектор было бы еще менее осуществимо, даже если потребуются значительные технологические усовершенствования, которые уменьшат размер регуляторных элементов гена или увеличат пределы упаковки векторов AAV. Хотя может использоваться стратегия двойной доставки (например, Ryu и др., (2018), *Nat. Biotechnol.*, 36 (6): сс. 536-539, включенная здесь в качестве ссылки), что значительно усложняет производство и увеличивает стоимость. Кроме того, двойная доставка вирусного вектора значительно снижает эффективность генной коррекции, поскольку для успешного редактирования в данной клетке требуется инфицирование обоими векторами и сборка гибридного белка в клетке.

Обычно используемый ортолог Cas9 из *S. aureus* (SauCas9) значительно меньше по размеру по сравнению с SpyCas9, но имеет более сложную потребность в PAM - NGRRT (SEQ ID NO: 449). Эта последовательность, однако, находится за пределами диапазона, который, как ожидается, будет полезен для базового редактирования каузативного локуса.

Пример 10.2. Конструкции слияния RGN и последовательностей sgРНК.

Последовательность ДНК, кодирующая слитый белок со следующими доменами, получают с использованием стандартных методов молекулярной биологии: 1) домен RGN с мутациями, которые инактивируют активность расщепления ДНК ("мертвый" или "никаза"); 2) аденозиндезаминазы, полезная для редактирования базы. Все конструкции, описанные в таблице ниже, содержат слитый белок с активным доменом редактирования основания, в этом примере ADAT (SEQ ID NO: 450), функционально слитый с N-терминальным концом RGN APG08290.1. В данной области известно, что слитый белок также может быть получен с ферментом, редактирующим основание, на C-конце RGN. Кроме того, RGN и редактор оснований слитого белка обычно разделены аминокислотной последовательностью линкера. В данной области известно, что длина стандартных линкеров находится в диапазоне 15-30 аминокислот. Кроме того, в данной области известно, что определенные слитые белки между RGN и ферментом, редактирующим основание, например цитидиндезаминазой, также могут содержать по меньшей мере один домен ингибитора урацилгликозилазы (UGI), который может повысить эффективность редактирования оснований (США Патент № 10167457, включенный здесь в качестве ссылки). Следовательно, слитый белок может содержать APG08290.1, фермент, модифицирующий основание, и по меньшей мере один UGI.

Таблица 13

Конструкции для редактирования РНК-нацеленных оснований

Seq ID No.	Конструкция	RGN	Мертвая (D) или Никазная (N)	Редактор оснований	Линкер
451	Nuc-ADAT-линкер-dAPG08290.1-линкер-SV40	APG08290.1	D	ADAT	XTEN1
452	Nuc-ADAT-XTEN1-nAPG08290.1-линкер-SV40	APG08290.1	N	ADAT	XTEN1

Доступные сайты редактирования RGN определяются последовательностью PAM. При объединении RGN с базовым редактирующим доменом целевой остаток для редактирования должен находиться на нецелевой цепи (NTS), поскольку NTS является одноцепочечным, в то время как RGN связана с локусом. Оценка количества нуклеаз и соответствующих управляющих РНК позволяет выбрать наиболее подходящий инструмент редактирования генов для этого конкретного локуса. Несколько потенциальных последовательностей PAM, на которые могут быть нацелены описанные выше конструкции в гене *Idua* человека, находятся в непосредственной близости от мутантного нуклеотида, ответственного за мутацию W402X. Последовательность, кодирующая транскрипт управляющей РНК, содержащий 1) "спейсер", комплементарный некодирующей цепи ДНК в локусе заболевания, а 2) также производится последовательность РНК, необходимая для ассоциации управляющей РНК с RGN. Полезные управляющие последовательности РНК (sgРНК) показаны в табл. 14 ниже. Эти управляющие последовательности РНК можно оценить по их эффективности в направлении указанных выше редакторов оснований к интересующей

му локусу.

Таблица 14

Последовательность управляющих РНК

Последовательность целевой геномной последовательности	SEQ ID NO.	Кодирующая последовательность sgРНК (SEQ ID NO.)
5'-GGAGCAGCTCTAGGCCGAAGTGTCTG-3'	453	456
5'-TAGGCCGAAGTGTCTGCGAGCCGGGA-3'	454	457
5'-GCTCTAGGCCGAAGTGTCTGCGAGGCC-3'	455	458

Пример 10.3. Анализ активности в клетках пациентов с болезнью Гурлера.

Для проверки стратегии генотипа и оценки описанных выше конструкций используются фибробласты от пациентов с болезнью Гурлера. Конструируют вектор, содержащий соответствующие промоторы перед кодирующей последовательностью слитого белка и кодирующей последовательностью sgРНК для их экспрессии в клетках человека, аналогично векторам, описанным в примере 5. Известно, что промоторы и другие элементы ДНК (например, энхансеры или терминаторы) либо обладают высокими уровнями экспрессии в клетках человека, либо могут специфически хорошо экспрессироваться в клетках фибробластов. Вектор трансфицируется в фибробласты с использованием стандартных методик, например трансфекции, подобно тому, это описано в примере 6. Кроме того, можно использовать электропорацию. Клетки культивируют 1-3 дня. Геномную ДНК (g ДНК) выделяют стандартными методами. Эффективность редактирования определяется путем проведения анализа генотипирования кПЦР и/или секвенирования следующего поколения очищенной g ДНК, как описано ниже.

В анализе qPCR Taqman™ используются зонды, специфичные для аллелей дикого типа и мутантных аллелей. Эти зонды содержат флуорофоры, которые определяются по их спектральным свойствам возбуждения и/или излучения с использованием прибора qPCR. Набор для генотипирования, содержащий праймеры и зонды для ПЦР, может быть получен коммерчески (т.е. с помощью анализа генотипирования SNP Thermo Fisher Taqman™, ID C_2786275310 для SNP ID rs121965019) или может быть разработан. Пример разработанного набора праймеров и зондов показан в табл. 15.

Таблица 15

Праймеры и зонды RT-PCR

Описание	Последовательность	SEQ ID NO.
Праймер для прямой амплификации	5'-GACTCCTTCACCAAG-3'	459
Праймер для обратной амплификации	5'-GTAGATCAGCACCG-3'	460
Зонд дикого типа	5'-CTCTGGGCCGAAGT-3'	461
Зонд W402X	5'-CTCTAGGCCGAAGT-3'	462

После эксперимента по редактированию g ДНК подвергают анализу qPCR с использованием стандартных методов и праймеров и зондов, описанных выше. Ожидаемые результаты показаны в табл. 16. Эта система *in vitro* может быть использована для целесообразной оценки конструкций и выбора одной конструкции с высокой эффективностью редактирования для дальнейших исследований. Системы будут оцениваться в сравнении с клетками с мутацией W402X и без нее, и предпочтительно с некоторыми мутациями, которые являются гетерозиготными по этой мутации. Значения Ct будут сравниваться либо с эталонным геном, либо с общей амплификацией локуса с использованием красителя, такого как Sybr green.

Таблица 16

Ожидаемые результаты qPCR

Генотип	Трансфекция редактором оснований	Ожидаемые результаты ПЦР
Idua ^{WT/WT}	Нет	Гомозиготный WT
Idua ^{WT/W402X}	Нет	Гетерозиготный: 50% WT, 50% W402X
Idua ^{W402X/W402X}	Нет	Гомозиготный W402X
Idua ^{W402X/W402X}	Да	Вариабельный

Ткани также можно анализировать с помощью секвенирования следующего поколения. Могут быть использованы сайты связывания праймеров, такие как показаны ниже (табл. 17), или другие подходящие сайты связывания праймеров, которые может идентифицировать специалист в данной области. После ПЦР-амплификации продукты, содержащие выступающие последовательности Illumina Nextera XT, проходят подготовку библиотеки в соответствии с протоколом метагеномного секвенирования библиотеки Illumina 16S. Глубокое секвенирование выполняется на платформе Illumina Mi-Seq. Обычно для каждого ампликона генерируется 200000 из 250 пар оснований считываний (2×100000 считываний). Считывания анализируются с помощью CRISPResso (Pinello и др., 2016) для расчета скорости редактирования. Выходные выравнивания подбираются вручную для подтверждения сайтов вставки и делеции, а также для идентификации сайтов микргомологии в сайтах рекомбинации.

Таблица 17

Сайты связывания праймеров NGS

Направление	Последовательность	SEQ ID NO.
Вперед	5'-ACTTCCTCCAGCC-3'	463
Назад	5'-GAACCCCGGCTTA-3'	464

Вестерн-блоттинг клеточного лизата трансфицированных клеток и контрольных клеток с использованием антитела против IDUA выполняется для проверки экспрессии полноразмерного белка, а анализ активности фермента на клеточном лизате с использованием субстрата 4-метилумбеллиферил-альдуронида подтверждает, что фермент является каталитически активный (Hopwood и др., Clin.Chim. Acta (1979), 92 (2): сс. 257-265, включено в описание изобретения в качестве ссылки). Эти эксперименты проводят по сравнению с исходной линией клеток IduaW402X/W402X (без трансфекции), линией клеток IduaW402X/W402X, трансфицированной конструкцией для редактирования оснований и случайной направляющей последовательностью и линией клеток, экспрессирующей IDUA дикого типа.

Пример 10.4: Подтверждение лечения заболевания на мышинной модели.

Чтобы проверить эффективность этого терапевтического подхода, используется модель мыши с нонсенс-мутацией аналогичной аминокислоты. Линия мышей несет мутацию W392X в своем гене Idua (Gene ID: 15932), которая соответствует гомологичной мутации у пациентов с синдромом Гурлера (Bunge и др., (1994), Hum. Mol. Genet. 3 (6): сс. 861- 866, включено в описание изобретения в качестве ссылки). Этот локус содержит отдельную нуклеотидную последовательность по сравнению с таковой у человека, в которой отсутствует последовательность РАМ, необходимая для коррекции с помощью редакторов оснований, описанных в предыдущих примерах, и, таким образом, требует разработки отдельного слитого белка для выполнения нуклеотидной коррекции. Облегчение заболевания у этого животного может подтвердить терапевтический подход к исправлению мутации в тканях, доступных для вектора доставки генов.

Мыши, гомозиготные по этой мутации, обладают рядом фенотипических характеристик, аналогичных тем, которые находят у пациентов с синдромом Гурлера. Слитый белок, редактирующий основание и RGN, как описано выше (табл. 13), вместе с управляющей РНК последовательностью включается в вектор экспрессии, который обеспечивает экспрессию белка и транскрипцию РНК у мышей. План исследования показан ниже в табл. 18. Исследование включает группы, которым вводили высокую дозу вектора экспрессии, включающего слитый белок, редактирующий основание, и РНК управляющую последовательность, низкую дозу того же вектора экспрессии - контроль, который является мышинной моделью, обработанной вектором экспрессии, который не содержит слитого белка для редактирования основания или управляющей РНК, и второй контроль, который представляет собой мышью дикого типа, обработанную тем же пустым вектором.

Таблица 18

Эксперимент по редактированию генома на мышинной модели

Группа	Линия мышей	N	Лечение
1	Idua-W392X ¹	≥ 5	Низкие дозы вектора
2	Idua-W392X	≥ 5	Высокие дозы вектора
3	Idua-W392X	≥ 5	Способ доставки
4	129/Sv (WT)	5	Способ доставки

Конечные точки для оценки включают массу тела, экскрецию GAG с мочой, ферментативную активность IDUA в сыворотке, активность IDUA в представляющих интерес тканях, патологию тканей, генотипирование интересующих тканей для проверки коррекции SNP, а также поведенческую и неврологическую оценку. Поскольку некоторые конечные точки являются терминальными, могут быть добавлены дополнительные группы для оценки, например, тканевой патологии и активности тканевых IDUA до окончания исследования. Дополнительные примеры конечных точек можно найти в опубликованных статьях, раскрывающих модели животных с синдромом Гурлера (Shull и др. (1994), Proc. Natl. Acad. Sci. USA, № 91 (26): сс. 12937-12941; Wang и др. (2010), Mol. Genet. Metab., № 99 (1): сс. 62-71; Hartung и др. (2004), Mol. Ther., № 9 (6): сс. 866-875; Liu и др. (2005), Mol. Ther., № 11 (1): сс. 35-47; Clarke и др. (1997), Hum. Mol. Genet. № 6 (4): сс. 503-511; все они включены в настоящее описание изобретения в качестве ссылки).

Один из возможных векторов доставки использует аденоассоциированный вирус (AAV). Создан вектор, содержащий последовательность, кодирующую слитый белок редактор оснований dRGN (например, SEQ ID NO: 452), которой предшествует энхансер CMV (SEQ ID NO: 138) и промотор (SEQ ID NO: 137) или другая подходящая комбинация энхансера и промотора), необязательно последовательность Козака, и которая функционально слитая на 3'-конце с последовательностью терминатора и последовательностью полиаденилирования, такой как минимальная последовательность, описанная у Levitt, N.;

Briggs, D.; Gil, A.; Proudfoot, N. J. Определение эффективного синтетического поли (А) сайта. Genes Dev. 1989, 3 (7), сс. 1019-1025. Вектор может дополнительно содержать кассету экспрессии, кодирующую единственную управляющую РНК, функционально связанную своим 5'-концом с промотором U6 человека (SEQ ID NO: 139), или другой промотор, подходящий для продукции малых некодирующих РНК, и дополнительно содержащий последовательности инвертированных концевых повторов (ITR), необходимые и хорошо известные в данной области для упаковки в капсид AAV. Производство и упаковка вирусов осуществляется стандартными методами, такими как те, что описаны в патенте США № 9,587,250, включенном в настоящее описание изобретения в качестве ссылки.

Другие возможные вирусные векторы включают аденовирусные и лентивирусные векторы, которые обычно используются и могут содержать аналогичные элементы с различными возможностями упаковки и требованиями. Также можно использовать невирусные методы доставки, такие как мРНК и сгРНК, инкапсулированные липидными наночастицами (Cullis, PR и Allen, TM (2013), Adv. Drug Deliv. Rev. 65 (1): сс. 36-48; Finn и др. (2018), Cell Rep. 22 (9): сс. 2227-2235, оба включены в качестве ссылки) гидродинамическая инъекция плазмидной ДНК (Suda T and Liu D, 2007) Mol. Ther. 15 (12): сс. 2063-2069, включена сюда в качестве ссылки), или рибонуклеопротеиновые комплексы sgРНК и связанные с наночастицами золота (Lee, K.; Conboy, M.; Park, HM; Jiang, F.; Kim, HJ; Dewitt, MA; Mackley, VA; Chang, K.; Rao, A.; Skinner, C; и др. Доставка наночастиц рибонуклеопротеина Cas9 и донорской ДНК in vivo индуцирует гомологически направленную репарацию ДНК. Nat. Biomed. Eng. 2017, 1 (11), сс. 889-90).

Пример 10.5: Коррекция заболевания на мышинной модели с гуманизированным локусом.

Чтобы оценить эффективность идентичной конструкции редактора оснований, которая будет использоваться для терапии человека, необходима модель мыши, в которой нуклеотиды рядом с W392 изменены, чтобы соответствовать последовательности у человека вокруг W402. Это может быть достигнуто с помощью различных методов, включая использование шаблона RGN и HDR для вырезания и замены локуса в эмбрионах мыши.

Из-за высокой степени консервативности аминокислот большинство нуклеотидов в локусе мыши могут быть изменены на нуклеотиды последовательности человека с молчащими мутациями, как показано в табл. 19. Происходят только изменения оснований, приводящие к измененной кодирующей последовательности в конечном сконструированном геноме мыши после введенного стоп-кодона.

Таблица 19

Нуклеотидные мутации для создания локуса гуманизированной мыши

Характеристика	Человек (W402X)		Мышь (W392X)		Гуманизированная мышь	
	Нуклеотид (SEQ ID NO: 465)	Закодировано AA	Нуклеотид (SEQ ID NO: 466)	Закодировано AA	Нуклеотид (SEQ ID NO: 467)	Закодировано AA
Протоспейсер	G	E	A	G	G	G
	G	E	G	E	G	E
	A		A			
	G		A			
	C	Q	C	Q	C	Q
	A		A			
	G		A			
	C	L	C	L	C	L
	T		T			
	C		C			
	T	СТОП	T	СТОП	T	СТОП
	A		A			
	G		G			
	G	A	G	A	G	A
	C		C			
	C		C			
	G	E	G	E	G	E
	A		A			
	A		G			
	G	V	G	V	G	V
T	T					
G	C					
T	S	T	S	T	S	
C		C				
G		A				
РАМ, не критично	C	Q	A	K	C	Q
	A		A			
	G		G			
РАМ, критично	G	A	G	A	G	A
	C		C			
	C		T			

После конструирования этой линии мышей будут проведены аналогичные эксперименты, как описано в примере 10.4.

Пример 11. Целевые мутации, ответственные за атаксию Фридрейха.

Экспансия последовательности тринуклеотидного повтора, вызывающая атаксию Фридрейха (FRDA), происходит в определенном генетическом локусе внутри гена FXN, который называется областью нестабильности FRDA. РНК-управляемые нуклеазы (RGN) могут использоваться для удаления области нестабильности в клетках пациентов с FRDA. Этот подход требует 1) последовательности RGN и управляющей РНК, которая может быть запрограммирована для нацеливания на аллель в геноме человека; и 2) подход к доставке RGN и управляющей последовательности. Многие нуклеазы, используемые для редактирования генома, такие как обычно используемая нуклеаза Cas9 из *S. pyogenes* (SpCas9), слишком велики, чтобы их можно было упаковывать в аденоассоциированные вирусные (AAV) векторы, особенно с учетом длины гена SpCas9 и управляющей РНК в дополнение к другим генетическим элементам, необходимым для функциональных каскадов экспрессии. Это делает маловероятным жизнеспособность подхода с использованием SpCas9.

Компактные РНК-управляемые нуклеазы, описанные в изобретении, в частности APG07433.1 и APG08290.1, уникально хорошо подходят для удаления области нестабильности FRDA. Каждая RGN имеет нужную РНК, которая находится в непосредственной близости от области нестабильности FRDA. Кроме того, каждая из этих RGN может быть упакована в вектор AAV вместе с управляющей РНК. Упаковка двух управляющих РНК, вероятно, потребует второго вектора, но этот подход все же выгодно отличается от того, что потребовалось бы от более крупной нуклеазы, такой как SpCas9, которая потребовала бы расщепления последовательности белка между двумя векторами.

В табл. 20 показано расположение геномных последовательностей-мишеней, подходящих для нацеливания APG07433.1 или APG08290.1 на 5' и 3' фланги области нестабильности FRDA. Оказавшись в локусе, RGN будет вырезать область нестабильности FA. Исечение области можно проверить с помощью секвенирования локуса Illumina.

Таблица 20

Геномные последовательности-мишени для систем RGN

Управление No.	Расположение относительно области нестабильности FRDA	Целевая последовательность генома	SEQ ID NO.
1	5'	ATCACCTGAGGTCCGGAGTTCAAGA	468
2	5'	GTCTTGAACCTCCGGACCTCAGGTGA	469
3	5'	TGAACTCCGGACCTCAGGTGATCCA	470
4	3'	GAAAAGTTAGCCGGGCGTGGTGTCTG	471

Пример 12. Целевые мутации, ответственные за заболевания серповидных клеток.

Целевые последовательности в области энхансера BCL11A (SEQ ID NO: 472) могут обеспечивать механизм увеличения количества гемоглобулина плода (HbF) для лечения или облегчения симптомов заболеваний серповидных клеток. Например, полногеномные исследования ассоциации выявили набор генетических вариаций BCL11A, которые связаны с повышенными уровнями HbF. Эти вариации представляют собой набор SNP, обнаруженных в некодирующих областях BCL11A, которые функционируют как специфичные для определенных стадий, ограниченные клонами энхансерные области. Дальнейшее исследование показало, что этот энхансер BCL11A необходим в эритроидных клетках для экспрессии BCL11A (Bauer и др., (2013) Science № 343: сс. 253-257, включена в описание изобретения в качестве ссылки). Область энхансера была обнаружена в интроне 2 гена BCL11A, и были идентифицированы три области гиперчувствительности к ДНКазе I (часто указывающие на состояние хроматина, связанное с регуляторным потенциалом) в интроне 2. Эти три области были идентифицированы как "+62", "+58" и "+55" в соответствии с расстоянием в килобазах от сайта начала транскрипции BCL11A. Эти энхансерные области составляют примерно 350 (+55); 550 (+58); и 350 (+62) нуклеотидов в длину (Bauer и др., 2013).

Пример 12.1. Определение предпочтительных систем RGN.

Здесь мы описываем 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, который более эффективен в клетках млекопитающих. Кроме того, этот подход использует нуклеазу достаточно маленького размера, которая может быть упакована в один вектор AAV для доставки *in vivo*.

Мотив энхансера GATA1 в области энхансера BCL11A человека (SEQ ID NO: 472) является идеальной мишенью для разрушения с использованием нуклеаз, управляемых РНК (RGN), для снижения экспрессии BCL11A с одновременной повторной экспрессией HbF в эритроцитах взрослого человека (Wu и др. (2019) Nat Med 387: с. 2554). Несколько последовательностей РНК, совместимых с APG07433.1 и APG08290.1, легко обнаруживаются в генетическом локусе, окружающем этот сайт GATA1. Эти нуклеазы имеют последовательность РНК 5'-NNNNCC-3' (SEQ ID NO: 6) и компактны по размеру, что потенциально позволяет их доставку вместе с соответствующей управляющей РНК в один AAV или аденовирусный вектор. Этот подход к доставке дает множество преимуществ по сравнению с другими, таких как доступ к гемопоэтическим стволовым клеткам и хорошо установленный профиль

безопасности и способы получения.

Обычно используемая нуклеаза Cas9 из *S. pyogenes* (SpyCas9) требует последовательностей PAM 5'-NGG-3' (SEQ ID NO: 448), некоторые из которых присутствуют рядом с мотивом GATA1. Однако размер SpyCas9 не позволяет упаковку в один AAV или аденовирусный вектор и, таким образом, лишает этот подход вышеупомянутых преимуществ. Хотя можно использовать стратегию двойной доставки, это значительно усложнит производство и повысит стоимость. Кроме того, двойная доставка вирусного вектора значительно снижает эффективность коррекции генов, поскольку для успешного редактирования данной клетки требуется инфицирование обоими векторами.

Получают кассету экспрессии, кодирующую оптимизированный по кодонам человека APG07433.1 (SEQ ID NO: 128) или APG08290.1 (SEQ ID NO: 130), аналогично способу, описанному в примере 6. Кассеты экспрессии, которые экспрессируют управляющие РНК для RGN APG07433.1 и APG08290.1 также производятся. Эти управляющие РНК включают 1) последовательность протоспейсера, которая комплементарна не кодирующей или кодирующей цепи ДНК в локусе энхансера BCL11A (целевая последовательность), и 2) последовательность РНК, необходимую для ассоциации направляющей РНК с RGN (SEQ ID NO: 18 для APG07433.1 и SEQ ID NO: 35 для APG08290.1). Поскольку несколько потенциальных последовательностей PAM для нацеливания с помощью APG07433.1 или APG08290.1 окружают мотив энхансера BCL11A GATA1, создается несколько потенциальных направляющих РНК-конструкций для определения наилучшей последовательности протоспейсера, которая обеспечивает надежное расщепление и опосредованное NHEJ нарушение последовательности энхансера BCL11A GATA1. Целевые геномные последовательности в таблице, представленной ниже (табл. 21) оцениваются для направления RGN в этот локус.

Таблица 21

Целевые последовательности для локуса энхансера BCL11A GATA1

Управление	Нуклеаза	Целевая последовательность генома	Целевая SEQ ID NO.
1	APG07433.1	GCACTAGACTAGCTTCAAAGTTGTAG	473
2	APG07433.1	CCTAATCAGAGGCCAAACCCTTCCTG	474
3	APG07433.1	CAAGCTAACAGTTGCTTTTATCACAG	475
4	APG08290.1	GCACTAGACTAGCTTCAAAGTTGTAG	476
5	APG08290.1	CCTAATCAGAGGCCAAACCCTTCCTG	477
6	APG08290.1	CAAGCTAACAGTTGCTTTTATCACAG	478

Для оценки эффективности, с которой APG07433.1 или APG08290.1 генерирует вставки или делеции, которые нарушают область энхансера BCL11A, используются линии клеток человека, такие как клетки эмбриональной почки человека (клетки HEK). Получают ДНК-вектор, включающий в себя кассету экспрессии RGN (например, как описано в примере 6). Также получают отдельный вектор, содержащий кассету экспрессии с кодирующей последовательностью для последовательности управляющей РНК из табл. 21. Такая кассета экспрессии может дополнительно содержать промотор U6 РНК-полимеразы III человека (SEQ ID NO: 139), как описано в примере 6. Кроме того, можно использовать один вектор, содержащий кассеты экспрессии как RGN, так и управляющей РНК. Вектор вводят в клетки HEK с использованием стандартных методик, таких как описанные в примере 6, и клетки культивируют в течение 1 -3 дней. После этого периода культивирования выделяют геномную ДНК и определяют частоту вставок или делеций с использованием расщепления эндонуклеазой I T7 и/или прямого секвенирования ДНК, как описано в примере 6.

Область ДНК, охватывающую целевой участок BCL11A, амплифицируют с помощью ПЦР с праймерами, содержащими выступающие последовательности Illumina Nextera XT. Эти ПЦР-ампликоны либо исследуют на образование NHEJ с использованием расщепления эндонуклеазой I T7, либо подвергают подготовке библиотеки в соответствии с протоколом метагеномного секвенирования библиотеки Illumina 16S или аналогичной подготовкой библиотеки секвенирования следующего поколения (NGS). После глубокого секвенирования генерированные чтения анализируются CRISPResso для расчета скорости редактирования. Выходные выравнивания подбираются вручную для подтверждения сайтов вставки и делеции. Этот анализ определяет предпочтительную RGN и соответствующую предпочтительную управляющую РНК (sgРНК). Анализ может привести к тому, что как APG07433.1, так и APG08290.1 будут в равной степени предпочтительны. Кроме того, анализ может определить, что существует более одной предпочтительной управляющей РНК или что все целевые геномные последовательности в табл. 21 являются одинаково предпочтительными.

Пример 12.2. Анализ экспрессии гемоглобина плода.

В этом примере генерированные APG07433.1 или APG08290.1 вставки или делеции, нарушающие область энхансера BCL11A, анализируются на экспрессию гемоглобина плода. Используют CD34+гемопозитические стволовые клетки (HSC) здорового донора человека. Эти HSC культивируют, и вектор (ы), содержащий кассеты экспрессии с кодирующими областями предпочтительной RGN и предпочтительной sgРНК, вводят с использованием способов, аналогичных тем, которые описаны в примере

11.1. После электропорации эти клетки дифференцируются *in vitro* в эритроциты с использованием установленных протоколов (например, Giarratana et al. (2004) Nat Biotechnology 23: 69-74, включенный в настоящее описание в качестве ссылки). Затем измеряют экспрессию HbF с помощью вестерн-блоттинга с антителом против HbF человека или количественно определяют с помощью высокоэффективной жидкостной хроматографии (ВЭЖХ). Ожидается, что успешное разрушение локуса энхансера BCL11A приведет к увеличению продукции HbF по сравнению с HSC, электропорированными только с RGN, но без проводника.

Пример 12.3: Анализ уменьшения образования серповидных клеток.

В этом примере генерированные APG07433.1 или APG08290.1 вставки или делеции, нарушающие область энхансера BCL11A, анализируются с точки зрения снижения образования серповидных клеток. Используются донорские CD34+гемопозитические стволовые клетки (HSC) от пациентов, страдающих серповидно-клеточной анемией. Эти HSC культивируют, и вектор(ы), содержащий(и) кассеты экспрессии с кодирующими областями предпочтительной RGN и предпочтительной sgPHK, вводят с использованием способов, аналогичных тем, которые описаны в примере 11.1. После электропорации эти клетки дифференцируются *in vitro* в эритроциты с использованием установленных протоколов (Giarratana и др. (2004) Nat Biotechnology № 23: сс. 69-74). Затем измеряют экспрессию HbF с помощью вестерн-блоттинга с антителом против HbF человека или количественно определяют с помощью высокоэффективной жидкостной хроматографии (ВЭЖХ). Ожидается, что успешное разрушение локуса энхансера BCL11A приведет к увеличению продукции HbF по сравнению с HSC, электропорированными только с RGN, но без проводника.

В этих дифференцированных эритроцитах образование серповидных клеток индуцируется добавлением метабисульфита. Количество серповидных и нормальных эритроцитов подсчитывают с помощью микроскопа. Ожидается, что количество серповидных клеток меньше в клетках, обработанных APG07433.1 или APG08290.1 плюс sgPHK, чем в клетках, необработанных или обработанных только RGN.

Пример 12.4. Подтверждение лечения заболевания на мышинной модели.

Для оценки эффективности использования APG07433.1 или APG08290.1, разрушающих локус BCL11A, используются подходящие гуманизированные мышинные модели с серповидно-клеточной анемией. Кассеты экспрессии, кодирующие предпочтительную RGN и предпочтительную sgPHK, упаковывают в векторы AAV или аденовирусные векторы. В частности, аденовирус типа Ad5/35 эффективен в отношении HSC. Выбрана подходящая модель мыши, содержащая гуманизированный локус HBB с аллелями серповидных клеток, такая как B6; FVB-Tg (LCR-HBA2, LCR-HBB * E26K) 53Hhb/J или B6.Cg-Hbattm1Paz Hbbtm1Tow Tg (HBA-HBBs) 41Paz/HhbJ. Этим мышам вводят гранулоцитарный колониестимулирующий фактор отдельно или в комбинации с плериксафором для мобилизации HSC в кровотоки. Затем внутривенно вводят AAV или аденовирусы, несущие RGN и управляющую плазмиду, и мышам дают возможность восстановиться в течение недели. Кровь, полученная от этих мышей, тестируется в тесте серпинга на серповидные клетки *in vitro* с использованием метабисульфита, и за мышами наблюдают в долгосрочном плане для отслеживания показателей смертности и кроветворной функции. Ожидается, что лечение AAV или аденовирусами, несущими RGN и управляющую PHK, снизит серповидность, смертность и улучшит гематопозитическую функцию по сравнению с мышами, получавшими лечение вирусами, не имеющими обеих кассет экспрессии, или вирусами, несущими только кассету экспрессии RGN.

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Молекула нуклеиновой кислоты, содержащая полинуклеотид, кодирующий полипептид РНК-управляемой нуклеазы (RGN), где указанный полинуклеотид содержит нуклеотидную последовательность, кодирующую полипептид RGN, содержащую аминокислотную последовательность, по меньшей мере на 95% идентичную последовательности SEQ ID NO: 11 или 27;

где указанный полипептид RGN связывает последовательность целевой ДНК специфическим для РНК-управляемой последовательности образом при связывании с управляющей РНК (gPHK), способной гибридизоваться с указанной последовательностью целевой ДНК.

2. Молекула нуклеиновой кислоты, содержащая полинуклеотид, кодирующий полипептид РНК-управляемой нуклеазы (RGN), где указанный полинуклеотид содержит нуклеотидную последовательность, кодирующую полипептид RGN, содержащую аминокислотную последовательность, по меньшей мере на 95% идентичную последовательности SEQ ID NO: 11 или 27.

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

4. Молекула нуклеиновой кислоты по любому из пп.1-3, где указанный полипептид RGN не является активной нуклеазой или функционирует как никаза.

5. Молекула нуклеиновой кислоты по п.4, где полипептид RGN функционально слит с полипептидом, редактирующим основание.

6. Вектор, содержащий молекулу нуклеиновой кислоты по любому из пп.1-5.

7. Вектор по п.6, в котором указанный вектор дополнительно содержит по меньшей мере одну нуклеотидную последовательность, кодирующую указанную управляющую РНК, и где управляющая РНК

содержит РНК CRISPR, содержащую повторяющуюся последовательность CRISPR, по меньшей мере на 95% идентичную последовательности SEQ ID NO: 12 или 28.

8. Вектор по п.6 или 7, отличающийся тем, что управляющая РНК содержит tracrPНК по меньшей мере на 95% идентичную последовательности SEQ ID NO: 13 или 29.

9. Клетка, содержащая молекулу нуклеиновой кислоты по любому из пп.1-5 или вектор по любому из пп.6-8.

10. Полипептид РНК-управляемой нуклеазы (RGN), содержащий аминокислотную последовательность, по меньшей мере на 95% идентичную последовательности SEQ ID NO: 11 или 27.

11. Полипептид RGN по п.10, где указанный RGN полипептид содержит аминокислотные последовательности SEQ ID NO: 11 или 27.

12. Система связывания последовательности целевой ДНК, которая включает в себя:

а) одну или несколько управляющих РНК, способных гибридизоваться с указанной последовательностью целевой ДНК или с одной или несколькими нуклеотидными последовательностями, кодирующими одну или несколько управляющих РНК (gРНК) и

б) полипептид РНК-управляемой нуклеазы (RGN), содержащий аминокислотную последовательность, по меньшей мере на 95% идентичную последовательности SEQ ID NO: 11 или 27, или нуклеотидную последовательность, кодирующую полипептид RGN,

где одна или несколько управляющих РНК гибридизуются с последовательностью целевой ДНК и

где одна или несколько управляющих РНК образуют комплекс с полипептидом RGN, тем самым заставляя указанный полипептид RGN связываться с указанной последовательностью целевой ДНК.

13. Система по п.12, где по меньшей мере одна указанная нуклеотидная последовательность, кодирующая одну или более управляющих РНК, и указанная нуклеотидная последовательность, кодирующая полипептид RGN, каждая функционально связана с промотором, гетерологичным к указанной нуклеотидной последовательности.

14. Система по п.12 или 13, в которой целевая последовательность ДНК находится внутри эукариотической клетки.

15. Система по любому из пп.12-14, в которой указанный полипептид RGN является не активной нуклеазой или функционирует как никаза, и где полипептид RGN функционально связан с полипептидом, редактирующим основание.

16. Система по любому из пп.12-15, отличающаяся тем, что указанная система дополнительно включает один или несколько донорных полинуклеотидов или одну или несколько нуклеотидных последовательностей, кодирующих один или несколько донорных полинуклеотидов, причем каждая из указанных нуклеотидных последовательностей, кодирующих один или несколько донорных полинуклеотидов, функционально связана с промотором, гетерологичным для каждой указанной нуклеотидной последовательности.

17. Применение системы по любому из пп.12-16 в качестве лекарственного средства.

18. Применение системы по любому из пп.12-16 в способе лечения генетически наследуемых болезней, где указанный способ содержит доставку системы к указанной последовательности целевой ДНК или клетке, содержащей последовательность целевой ДНК.

19. Способ связывания последовательности целевой ДНК, включающий доставку системы согласно любому из вариантов осуществления изобретения по пп.12-16 к указанной последовательности целевой ДНК или к клетке, содержащей последовательность целевой ДНК.

20. Способ расщепления и/или модификации последовательности целевой ДНК, включающий приведение последовательности целевой ДНК в контакт с:

а) полипептидом РНК-управляемой нуклеазы (RGN), где указанная RGN содержит аминокислотную последовательность по меньшей мере на 95% идентичную последовательности SEQ ID NO: 11 или 27; и

б) одной или несколькими управляющими РНК, способными нацеливать RGN (а) на последовательность целевой ДНК,

где одна или несколько управляющих РНК гибридизуются с целевой последовательностью ДНК, тем самым направляя указанный полипептид RGN на связывание с указанной целевой последовательностью ДНК, и таким образом происходит расщепление и/или модификация указанной целевой последовательности ДНК.

21. Способ по п.20, в котором указанная модифицированная последовательность целевой ДНК включает вставку гетерологичной ДНК в последовательность целевой ДНК.

22. Способ по п.20, где указанная модифицированная последовательность целевой ДНК включает делецию по меньшей мере одного нуклеотида из последовательности целевой ДНК.

23. Способ по п.20, где указанная модифицированная последовательность целевой ДНК содержит мутацию по меньшей мере одного нуклеотида в последовательности целевой ДНК.

24. Способ по любому из пп.20-23, отличающийся тем, что последовательность целевая ДНК находится внутри клетки.

25. Способ по п.24, в котором клетка является эукариотической клеткой.

26. Способ по п.24 или 25, дополнительно включающий культивирование клетки в условиях, в ко-

торых полипептид RGN экспрессируется и расщепляет последовательность целевой ДНК с получением модифицированной последовательности ДНК и отбор клетки, содержащей указанную модифицированную последовательность ДНК.

