

(19)



**Евразийское
патентное
ведомство**

(11) **038321**(13) **B1**

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ

(45) Дата публикации и выдачи патента
2021.08.09

(51) Int. Cl. *C12N 9/22* (2006.01)
C12N 15/62 (2006.01)

(21) Номер заявки
201790968

(22) Дата подачи заявки
2015.11.03

**(54) ОПОСРЕДУЕМАЯ ПЕПТИДОМ ДОСТАВКА НАПРАВЛЯЕМОЙ РНК
ЭНДОНУКЛЕАЗЫ В КЛЕТКИ**

(31) 62/075,999

WO-A2-2014165825

(32) 2014.11.06

WO-A1-2013176772

(33) US

(43) 2018.01.31

(86) PCT/US2015/058760

(87) WO 2016/073433 2016.05.12

(71)(73) Заявитель и патентовладелец:
**Е.И. ДЮПОН ДЕ НЕМУР ЭНД
КОМПАНИ (US)**

ROBERTA MARCHIONE ET AL.: "ZEBRA cell-penetrating peptide as an efficient delivery system in *Candida albicans*", BIOTECHNOLOGY JOURNAL, vol. 9, no. 8, 12 August 2014 (2014-08-12), pages 1088-1094, XP055243411, DE ISSN: 1860-6768, DOI: 10.1002/biot.201300505 the whole document

(72) Изобретатель:
**Фриш Райан Л., Фань Сяочунь, Хонг
Сеунг-Пио, Джексон Этель Ноланд
(US)**

NEKHOTIAEVA N. ET AL.: "Cell entry and antimicrobial properties of eukaryotic cell-penetrating peptides", THE FASEB JOURNAL, FEDERATION OF AMERICAN SOCIETIES FOR EXPERIMENTAL BIOLOGY, BETHESDA; US, vol. 18, 1 February 2004 (2004-02-01), pages 394-396, XP003024027, ISSN: 0892-6638 the whole document

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

J. E. DICARLO ET AL.: "Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems", NUCLEIC ACIDS RESEARCH, vol. 41, no. 7, 4 March 2013 (2013-03-04), pages 4336-4343, XP055086617, ISSN: 0305-1048, DOI: 10.1093/nar/gkt135 abstract

(56) S. RAMAKRISHNA ET AL.: "Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA", GENOME RESEARCH, vol. 24, no. 6, 2 April 2014 (2014-04-02), pages 1020-1027, XP055128944, ISSN: 1088-9051, DOI: 10.1101/gr.171264.113 the whole document
WO-A1-2014065596
WO-A1-2014089290

WO-A1-2015086795

DATABASE WPI Week 201535 Thomson Scientific, London, GB; AN 2015-20198F XP002753224, -& CN 104 328 138 A (SHANGHAI DIDA BIOLOGICAL TECHNOLOGY CO) 4 February 2015 (2015-02-04) abstract

(57) Раскрыта композиция, которая содержит по меньшей мере один белковый компонент направляемой РНК эндонуклеазы (RGEN) и по меньшей мере один проникающий в клетку пептид (CPP), где белковый компонент RGEN и CPP ковалентно или нековалентно связаны друг с другом в комплексе белок RGEN-CPP. Комплекс белок RGEN-CPP может пересекать (i) клеточную мембрану или (ii) клеточную стенку и клеточную мембрану клетки. Белковый компонент RGEN из комплекса белок RGEN-CPP в определенных вариантах осуществления может быть ассоциирован с подходящим РНК-компонентом с обеспечением RGEN, способной к специфическому нацеливанию на ДНК. Также раскрыты композиции, содержащие по меньшей мере один белковый компонент из комплекса направляющий полинуклеотид/эндонуклеаза Cas и по меньшей мере один CPP, а также способы доставки белков RGEN в микробные клетки, а также способы нацеливания на ДНК с помощью RGEN.

B1**038321****038321 B1**

Заявка испрашивает приоритет предварительной заявки на патент США № 62/075999, поданной 6 ноября 2014 года, которая включена в данный документ в своем полном объеме посредством ссылки.

Область техники, к которой относится изобретение

Настоящее изобретение относится к области молекулярной биологии. В частности, настоящее изобретение относится к доставке белковых компонентов направляемых РНК эндонуклеаз в клетки с применением проникающих в клетку пептидов.

Ссылка на перечень последовательностей, предоставленный в электронном виде

Официальная копия данного перечня последовательностей предоставлена в электронном виде с помощью EFS-Web как перечень последовательностей в файле формата ASCII с названием "20151013_CL6273PCT_SequenceListing_ST25.txt", созданном 13 октября 2015 года и имеющем размер 384 килобайта, и подана одновременно с описанием. Перечень последовательностей, содержащийся в данном документе в формате ASCII, является частью описания и включен в данный документ посредством ссылки во всей своей полноте.

Предпосылки изобретения

Путем к пониманию функции гена в пределах организма является подавление его экспрессии. Подавления экспрессии гена можно достичь, например, путем разрыва или делеции последовательности ДНК гена, что приводит к "нокауту" гена (Austin et al., Nat. Genetics 36:921-924). Нокауты генов обычно выполнялись посредством гомологичной рекомбинация (HR), методики, применимой для широкого спектра организмов от бактерий до млекопитающих. Другой путь изучения функции гена может заключаться в "нокине" гена, который также обычно проводят с помощью HR. При HR, применяемой для нацеливания на ген (обеспечение нокаута или нокина), может использоваться присутствие экзогенно вносимой ДНК, характеризующейся гомологией с целевым сайтом ("донорная ДНК").

Было показано, что HR для нацеливания на ген усиливается, если подвергнутый нацеливанию сайт ДНК содержит двухнитевой разрыв (Rudin et al., Genetics 122:519-534; Smih et al., Nucl. Acids Res. 23:5012-5019). Следовательно, были разработаны стратегии введения двухнитевых разрывов, чтобы облегчить опосредуемое HR нацеливание на ДНК. Например, для расщепления специфических сайтов ДНК были сконструированы нуклеазы с цинковыми пальцами, что приводило к повышенным уровням HR в сайте в случае присутствия донорной ДНК (Bibikova et al., Science 300:764; Bibikova et al., Mol. Cell. Biol. 21:289-297). Аналогично были также разработаны искусственные мегануклеазы (хоминг-эндонуклеазы) и подобные активаторам транскрипции эффекторные (TALE) нуклеазы для применения в опосредуемом HR нацеливании на ДНК (Epinat et al., Nucleic Acids Res. 31: 2952-2962; Miller et al., Nat. Biotech. 29:143-148).

Локусы, кодирующие системы расщепления ДНК CRISPR (короткие палиндромные повторы, регулярно расположенные группами) были обнаружены только в геномах приблизительно 40% бактерий и в геномах большинства архей (Horvath and Barrangou, Science 327:167-170; Karginov and Hannon, Mol. Cell 37:7-19). В частности, была разработана CRISPR-ассоциированная (Cas) направляемая РНК эндонуклеаза (RGEN), Cas9, из системы CRIPSR типа II в качестве средства для введения сайт-специфических разрывов в нить ДНК, которые стимулируют HR с донорной ДНК (предварительная заявка на патент США № 61/868706, поданная 22 августа 2013 года). Последовательность РНК-компонента Cas9 можно конструировать таким образом, чтобы Cas9 распознавала и расщепляла ДНК, содержащую (i) последовательность, комплементарную части РНК-компонента, и (ii) последовательность прилегающего к протоспейсеру мотива (PAM).

Нативные комплексы Cas9/РНК содержат две последовательности РНК, РНК CRISPR (crRNA) и транскрибирующую РНК CRISPR (tracrRNA). crRNA содержит, в направлении от 5' к 3', уникальную последовательность, комплементарную целевому сайту ДНК, и часть последовательности, кодируемую повторяющимся участком из локуса CRISPR, из которого данная crRNA была получена. tracrRNA содержит, в направлении от 5' к 3', последовательность, которая подвергается отжигу с повторяющимся участком crRNA, и часть, содержащую структуру "стебель-петля". Недавние работы привели к разработке направляющих РНК (gRNA), которые представляют собой химерные последовательности, содержащие, в направлении от 5' к 3', crRNA, связанную с tracrRNA (предварительная заявка на патент США № 61/868706, поданная 22 августа 2013 года).

Белковый компонент и РНК-компонент для проведения опосредуемого Cas9 целенаправленного воздействия на ДНК в клетке были получены в некоторых исследованиях с помощью стратегий экспрессии рекомбинантных ДНК. Например, белок Cas9 был экспрессирован в клетках с применением систем экспрессии на основе нуклеиновых кислот. Способы экспрессии РНК-компонентов, таких как gRNA, в определенных типах клеток включали применение промоторов РНК-полимеразы III (Pol III), которые обеспечивают возможность транскрипции РНК с точно определенными немодифицированными 5'- и 3'-концами (DiCarlo et al., Nucleic Acids Res. 41: 4336-4343; Ma et al., Mol. Ther. Nucleic Acids 3:e161). Эти методики экспрессии белка и РНК были применены в клетках нескольких разных видов, включая клетки маиса и сои (предварительная заявка на патент США № 61/868706, поданная 22 августа 2013 года), а также клетки человека, мыши, данио-рерио, *Trichoderma* и *Saccharomyces cerevisiae*.

Несмотря на данные успехи, все еще представляют интерес другие средства для введения белкового

компонента и РНК-компонента в клетку, такую как микробная клетка, чтобы опосредовать опосредуемое Cas9 целенаправленное воздействие на ДНК.

Краткое описание настоящего изобретения

В одном варианте осуществления настоящее изобретение относится к композиции, содержащей по меньшей мере один белковый компонент направляемой РНК эндонуклеазы (RGEN) и по меньшей мере один проникающий в клетку пептид (CPP), где белковый компонент и CPP ковалентно или нековалентно связаны друг с другом в комплекс белок RGEN-CPP, и комплекс белок RGEN-CPP может пересекать (i) клеточную мембрану или (ii) клеточную стенку и клеточную мембрану микробной клетки.

Во втором варианте осуществления белковый компонент RGEN ассоциирован по меньшей мере с одним РНК-компонентом, который содержит последовательность, комплементарную последовательности целевого сайта на хромосоме или эписоме в микробной клетке, где RGEN может связываться с последовательностью целевого сайта и необязательно расщеплять одну или обе нити ДНК в последовательности целевого сайта. В третьем варианте осуществления РНК-компонент содержит направляющую РНК (gRNA), содержащую РНК CRISPR (crRNA), функционально связанную с транскрибирующей РНК CRISPR (tracrRNA). В четвертом варианте осуществления RGEN может расщеплять одну или обе нити ДНК в последовательности целевого сайта.

В пятом варианте осуществления RGEN содержит аминокислотную последовательность CRISPR-ассоциированного (Cas) белка-9 (Cas9).

В шестом варианте осуществления белковый компонент RGEN и CPP связаны ковалентно.

В седьмом варианте осуществления белковый компонент RGEN и CPP связаны нековалентно.

В восьмом варианте осуществления CPP является катионным или амфипатическим.

В девятом варианте осуществления CPP включает в себя (i) CPP из трансактиваторного белка Zebra вируса Эпштейна-Барр, (ii) CPP с 6 или более смежными аргининовыми остатками, (iii) CPP транспор-тан-10 (TP10) или (iv) CPP из белка кадгерин сосудистого эндотелия.

В десятом варианте осуществления комплекс белок RGEN-CPP может пересекать клеточную стенку и клеточную мембрану микробной клетки.

Одиннадцатый вариант осуществления относится к микробной клетке, содержащей композицию, раскрытую в данном документе.

Двенадцатый вариант осуществления относится к способу доставки белкового компонента направляемой РНК эндонуклеазы (RGEN) в микробную клетку. Данный способ включает приведение микробной клетки в контакт с композицией, содержащей белковый компонент RGEN и по меньшей мере один проникающий в клетку пептид (CPP), где белковый компонент RGEN и CPP ковалентно или нековалентно связаны друг с другом в комплекс белок RGEN-CPP. В результате данной стадии приведения в контакт комплекс белок RGEN-CPP может пересекать (i) клеточную мембрану или (ii) клеточную стенку и клеточную мембрану микробной клетки, и тем самым проникать в микробную клетку.

В тринадцатом варианте осуществления применительно к способу (i) композиция дополнительно содержит по меньшей мере один РНК-компонент, который ассоциирован с белковым компонентом RGEN, или (ii) микробная клетка содержит РНК-компонент, при этом РНК-компонент ассоциирует с белковым компонентом RGEN после того как, комплекс белок RGEN-CPP попадет в микробную клетку; где РНК-компонент в (i) или (ii) содержит последовательность, комплементарную последовательности целевого сайта на хромосоме или эписоме в микробной клетке, и где RGEN может связываться с последовательностью целевого сайта и необязательно расщеплять одну или обе нити ДНК в последовательности целевого сайта. В четырнадцатом варианте осуществления RGEN может расщеплять одну или обе нити ДНК в последовательности целевого сайта. В пятнадцатом варианте осуществления микробная клетка дополнительно содержит донорный полинуклеотид, содержащий по меньшей мере одну последовательность, гомологичную последовательности в последовательности целевого сайта или вблизи нее, где донорный полинуклеотид интегрируется в последовательность целевого сайта или вблизи нее с помощью гомологичной рекомбинации.

Шестнадцатый вариант осуществления относится к полинуклеотидной последовательности, содержащей нуклеотидную последовательность, кодирующую белок слияния белок RGEN-CPP, который содержит белковый компонент направляемой РНК эндонуклеазы (RGEN) и по меньшей мере один проникающий в клетку пептид (CPP), где необязательно нуклеотидная последовательность функционально связана с промоторной последовательностью.

Семнадцатый вариант осуществления относится к способу получения белка слияния белок RGEN-CPP. Данный способ включает: (a) обеспечение полинуклеотидной последовательности, содержащей нуклеотидную последовательность, кодирующую белок слияния белок RGEN-CPP, который содержит белковый компонент направляемой РНК эндонуклеазы (RGEN) и по меньшей мере один проникающий в клетку пептид (CPP), где необязательно нуклеотидная последовательность функционально связана с промоторной последовательностью; (b) экспрессию белка слияния белок RGEN-CPP с полинуклеотидной последовательности, в результате чего получают белок слияния белок RGEN-CPP, где экспрессию необязательно проводят в микробной клетке; и (c) необязательно выделение белка слияния белок RGEN-CPP, полученного на стадии (b).

Восемнадцатый вариант осуществления относится к композиции, содержащей по меньшей мере один белковый компонент комплекса направляющий полинуклеотид/эндонуклеаза Cas и по меньшей мере один проникающий в клетку пептид (CPP), где белковый компонент и CPP ковалентно или нековалентно связаны друг с другом в комплекс направляющий полинуклеотид/эндонуклеаза Cas-CPP, где комплекс направляющий полинуклеотид/эндонуклеаза Cas-CPP может пересекать (i) клеточную мембрану или (ii) клеточную стенку и клеточную мембрану микробной клетки.

Девятнадцатый вариант осуществления относится к способу модификации целевого сайта в геноме микробной клетки. Данный способ включает обеспечение направляющего полинуклеотида, проникающего в клетку пептида (CPP) и эндонуклеазы Cas для микробной клетки, где направляющий полинуклеотид, эндонуклеаза Cas и CPP ковалентно или нековалентно связаны друг с другом в комплекс направляющий полинуклеотид/эндонуклеаза Cas-CPP и где комплекс направляющий полинуклеотид/эндонуклеаза Cas-CPP может пересекать (i) клеточную мембрану или (ii) клеточную стенку и клеточную мембрану микробной клетки.

Краткое описание графических материалов и последовательностей

Фиг. 1 - плазмида pZUFcas9 (SEQ ID NO: 6) содержит кассету экспрессии Cas9, кодон-оптимизированную для *Yarrowia*, изложенную в SEQ ID NO: 5. Точки начала репликации (ARS 18, fl ori, ColE1) отмечены перекрестной штриховкой, а селектируемые маркеры (Ura3, Amp) обозначены серым цветом. См. пример 1.

Фиг. 2A - плазмида pBAD/HisB (SEQ ID NO: 10) для экспрессии гетерологичных белков в *E. coli*. Промотор pBAD обозначен белым цветом. Точка начала репликации обозначена перекрестной штриховкой. См. пример 1.

Фиг. 2B - плазмида pRF48 (SEQ ID NO: 11) для экспрессии Cas9-NLS ("Cas9" на фигуре) в *E. coli*. Точка начала репликации обозначена перекрестной штриховкой. См. пример 1.

Фиг. 3A - плазмида pRF144 (SEQ ID NO: 20) для экспрессии слияния 6xHis-Zebra CPP-Cas9-NLS в *E. coli*. Точка начала репликации обозначена перекрестной штриховкой. См. пример 1.

Фиг. 3B - плазмида pRF145 (SEQ ID NO: 21) для экспрессии слияния 6xHis-PolyR CPP-Cas9-NLS в *E. coli*. Точка начала репликации обозначена перекрестной штриховкой. См. пример 1.

Фиг. 3C - плазмида pRF146 (SEQ ID NO: 22) для экспрессии слияния 6xHis-TP10 CPP-Cas9-NLS в *E. coli*. Точка начала репликации обозначена перекрестной штриховкой. См. пример 1.

Фиг. 3D - плазмида pRF162 (SEQ ID NO: 23) для экспрессии слияния 6xHis-pVEC CPP-Cas9-NLS в *E. coli*. Точка начала репликации обозначена перекрестной штриховкой. См. пример 1.

Фиг. 4 - разделение с помощью SDS-PAGE фракций очистки 6xHis-Zebra-Cas9-NLS. Обозначены лизаты, промывки, фракции элюирования и стандарты молекулярной массы. См. пример 1.

Фиг. 5 - структурная модель одиночного направляющего полинуклеотида, такого как одиночная направляющая РНК (sgRNA). Вариабельный нацеливающий (VT) домен показан серым цветом. Домен распознавания эндонуклеазы Cas9 (CER) показан черным цветом.

Фиг. 6 - *in vitro* транскрипция RGR sgRNA (для целенаправленного воздействия на локус Can1-1) с матрицы, полученной из плазмиды pRF46 (SEQ ID NO: 30). В реакциях *in vitro* транскрипции, которые инкубировали в течение 2, 4, 6 и 18 ч, продуцировались аналогичные уровни sgRNA. Также продуцировались продукты автокаталитического расщепления под действием рибозимов. См. пример 2.

Фиг. 7 - анализ *in vitro* расщепления с применением Zebra CPP-Cas9 в комплексе с sgRNA, специфической в отношении целевого сайта Can1-1. В каждую реакцию смесь вводили ДНК-полинуклеотид (982 п. о.), содержащий целевой сайт Can1-1. Каждую реакцию смесь подвергали разделению с помощью электрофореза на 1,2% геле. В реакционных смесях "только мишень", "только sgRNA", "только Zebra-Cas9" и "только Zebra-Cas9 (2xFT)" (FT, замораживание-оттаивание) не происходило расщепление целевого полинуклеотида. В реакционных смесях "Zebra-Cas9/sgRNA", "Zebra-Cas9/sgRNA (2xFT)" и "Cas9/sgRNA" (Cas9 дикого типа) целевой полинуклеотид расщеплялся специфическим образом, на что указывают полученные продукты расщепления. См. пример 3.

Фиг. 8 - измерение эффективности целенаправленного воздействия на геном у *Zebra* CPP-Cas9 (не ассоциирован с sgRNA) и комплексов Zebra CPP-Cas9/gRNA после приведения их в контакт с клетками *Yarrowia lipolytica*. Конечная концентрация используемого в отдельности Zebra-Cas9 составляла 5 мкМ, при этом в комплексах с sgRNA использовали различные конечные концентрации (1-5 мкМ) Zebra CPP-Cas9. Частота мутаций приведена как доля колоний дрожжей (выращиваемых на неселективной среде после приведения клеток в контакт либо с Zebra CPP-Cas9, либо Zebra CPP-Cas9/gRNA), которые засчитывались как устойчивые к канаванину после переноса на среду, содержащую канаванин. См. пример 4.

Фиг. 9 - пример анализа геля PAGE при очистке CPP-dsRED. 12,5% гель для PAGE окрашивали красителем Simply blue. Дорожка 1: стандарт молекулярной массы, дорожка 2: очищенный от примесей экстракт клеток tp10-dsREDexpress, дорожка 3: очищенный от примесей экстракт клеток после обработки гранулами tp10-dsREDexpress, дорожка 4: конечный раствор белка tp10-dsREDexpress, дорожка 5: очищенный от примесей экстракт клеток MPG-dsREDexpress, дорожка 6: очищенный от примесей экстракт клеток после обработки гранулами MPG-dsREDexpress, дорожка 7: конечный раствор белка MPG-dsREDexpress.

Таблица 1. Краткое описание последовательностей нуклеиновых кислот и белков под номерами SEQ ID

Описание	Нуклеиновая кислота с SEQ ID NO	Белок с SEQ ID NO
Открытая рамка считывания Cas9 <i>Streptococcus pyogenes</i> , кодон-оптимизированная для экспрессии в <i>Y. lipolytica</i>	1 (4107 оснований)	
Открытая рамка считывания Cas9 <i>Streptococcus pyogenes</i> , включающая C-терминальный линкер и NLS SV40 ("Cas9-NLS"), кодон-оптимизированная для экспрессии в <i>Y. lipolytica</i>	2 (4140 оснований)	3 (1379 ак)
Промотор FBAl <i>Y. lipolytica</i>	4 (543 основания)	
Кассета экспрессии Cas9-NLS (промотор FBAl и открытая рамка считывания Cas9-NLS)	5 (4683 оснований)	
Плазмида p2UFCas9	6 (10706 оснований)	
Прямой праймер для ПЦР Cas9-NLS	7 (35 оснований)	
Обратный праймер для ПЦР Cas9-NLS	8 (31 основание)	
Продукт ПЦР EcoRI-Cas9-NLS-HinDIII	9 (4166 оснований)	
Плазмида pBAD/HisB	10 (4092 основания)	
Плазмида pRF48	11 (8237 оснований)	
Проникающий в клетку пептид (CPP) Zebra из трансактиваторного белка Zebra вируса Эпштейна-Барр		12 (54 ак)
pVEC CPP, из белка кадгерина эндотелия мыши		13 (18 ак)
TR10 CPP, из белка нейропептида галанина		14 (21 ак)
Полиаргининовый (PolyR) CPP		15 (17 ак)
NcoI-6xHis-Zebra CPP-EcoRI	16 (194 основания)	
NcoI-6xHis-pVEC CPP-EcoRI	17 (86 оснований)	
NcoI-6xHis-TR10 CPP-EcoRI	18 (95 оснований)	
NcoI-6xHis-PolyR CPP-EcoRI	19 (83 основания)	
Плазмида pRF144, кодирующая белок слияния Zebra CPP-Cas9	20 (8294 основания)	
Плазмида pRF145, кодирующая белок слияния PolyR CPP-Cas9	21 (8183 основания)	
Плазмида pRF146, кодирующая белок слияния TR10 CPP-Cas9	22 (8195 оснований)	
Плазмида pRF162, кодирующая белок слияния pVEC CPP-Cas9	23 (8186 оснований)	
Домен распознавания эндонуклеазы Cas9 (CER) из gRNA.	24 (80 оснований)	
Целевой сайт Can1-1 <i>Y. lipolytica</i> или в качестве альтернативы ДНК, кодирующая переменный нацеливающий на Can1-1 домен gRNA.	25 (20 оснований)	
Рибозим типа hammerhead (HH)	26 (43 основания)	
Рибозим HDV	27 (68 оснований)	
Кассета экспрессии pre-sgRNA, HH-sgRNA-HDV (RGR), или в качестве альтернативы кассета экспрессии "RGR" (для нацеливания на локус Can1-1)	28 (211 оснований)	
Промотор РНК-полимеразы T7	29 (20 оснований)	
Плазмида pRF46	30 (2875 оснований)	
Прямой праймер T7	31	

	(20 оснований)	
	32	
Обратный праймер gRNArev1	(20 оснований)	
	33	
Праймер IV-up	(21 основание)	
	34	
Праймер IV-down	(20 оснований)	
Последовательность ДНК из анализа расщепления Cas1	35 (982 основания)	
Последовательность РНК, образующая петлю (GAAA)	36 (4 основания)	
Последовательность РНК, образующая петлю (CAAA)	37 (4 основания)	
Последовательность РНК, образующая петлю (AAAG)	38 (4 основания)	
		39 (1434 ак)
Белок слияния Zebra CPP-Cas9-NLS		40 (1397 ак)
Белок слияния PolyR CPP-Cas9-NLS		41 (1401 ак)
Белок слияния TP10 CPP-Cas9-NLS		42 (1398 ак)
Белок слияния pVEC CPP-Cas9-NLS		
Пример целевого сайта: последовательности PAM для Cas9	43 (23 основания)	
Последовательность PAM NGG	44 (3 основания)	
Последовательность PAM NNAGAA	45 (6 оснований)	
Последовательность PAM NNAGAAW	46 (7 оснований)	
Последовательность PAM NGGNG	47 (5 оснований)	
Последовательность PAM NNNNGATT	48 (8 оснований)	
Последовательность PAM NAAAAC	49 (6 оснований)	
Последовательность PAM NG	50 (2 основания)	
Парная последовательность для tracrRNA, пример 1	51 (22 основания)	
Парная последовательность для tracrRNA, пример 2	52 (15 оснований)	
Парная последовательность для tracrRNA, пример 3	53 (12 оснований)	
Парная последовательность для tracrRNA, пример 4	54 (13 оснований)	
TracrRNA, пример 1	55 (60 оснований)	
TracrRNA, пример 2	56 (45 оснований)	
TracrRNA, пример 3	57 (32 основания)	
TracrRNA, пример 4	58 (85 оснований)	
TracrRNA, пример 5	59 (77 оснований)	
TracrRNA, пример 6	60 (65 оснований)	
gRNA, пример 1	61 (131 основание)	
gRNA, пример 2	62 (117 оснований)	
gRNA, пример 3	63 (104 основания)	
gRNA, пример 4	64 (99 оснований)	
gRNA, пример 5	65 (81 основание)	
gRNA, пример 6	66 (68 оснований)	
gRNA, пример 7	67 (100 оснований)	
СРР, полученный из Tat (GRKKRRQRRR)		68 (10 ак)
СРР, полученный из Tat (RKKRRQRRR)		69 (9 ак)
СРР, полученный из Tat (RKKRRQRR)		70 (8 ак)
СРР пенетратина (RQIKIWFQNRRMKWKK)		71 (16 ак)

Полиаргининовый CPP (THRLPFRRRRR)		72 (11 ак)
Полиаргининовый CPP (GGRFRARRRRR)		73 (11 ак)
pVEC CPP (более короткий вариант), из белка кадгерина эндотелия мыши		74 (17 ак)
CPP, содержащий (KFF) ₃ K		75 (10 ак)
MAP-пептид CPP		76 (18 ак)
CPP (RRQRRTSKLMKR)		77 (12 ак)
CPP (KALAWKAKLAKALAKALAKLAKALAKLAKCEA)		78 (33 ак)
Богатый пролином повтор в CPP, VHLPPP		79 (6 ак)
Богатый пролином повтор в CPP, VHRPPP		80 (6 ак)
MPG-пептид CPP		81 (27 ак)
Per-1-пептид CPP		82 (21 ак)
hCT CPP, пример 1		83 (24 ак)
hCT CPP, пример 2		84 (18 ак)
dsRED с his-меткой		85
dsRED, кодон-оптимизированный для E. coli	86	
pBAD/HisB	87	
pRF161	88	
TAT		89
TLM		90
MPG1		91
pep1		92
CFFKDEL		93
his-TAT, оптимизированный для E. coli	94	
his-TLM, оптимизированный для E. coli	95	
his-MPG1, оптимизированный для E. coli,	96	
his-pep1, оптимизированный для E. coli	97	
his-CFFKDEL, оптимизированный для E. coli	98	
pRF224	99	
pRF214	100	
pRF213	101	
pRF217	102	
pRF216	103	
Олигонуклеотид 36	104	
His-Zebra PCR	105	
His-tp10 PCR	106	
His-pVEC PCR	107	
pRF144	108	
pRF162	109	
pRF146	110	
Олигонуклеотид 153	111	
pRF186	112	
pRF192	113	
pRF190	114	
his-CFFKDEL-Cas9		115
his-MPG1-Cas9		116
pRF48	117	
pRF243	118	
pRF238	119	
Ген galK	120	
Ген galE		121
Ген galT		122
CER-домен 1	123	
ДНК, кодирующая CER, для ПЦР	124	
pRF291	125	

Прямой праймер для CER	126	
Универсальный обратный праймер	127	
Универсальный прямой праймер T7	128	
Прямой праймер galk2-1	129	
Обратный праймер galk2-1	130	
Матрица для in vitro транскрипции sgRNA galk2-1	131	
Промотор T7	132	
ДНК, кодирующая переменный нацеливающий домен для воздействия на galk2-1	133	
Целевой сайт galk2-1	134	
galk2-1 sgRNA	135	
his-MFG1-dsREDexpress;		136
pVEC-dsREDexpress		137
CFFKDEL-dsREDexpress		138
TLM-dsREDexpress		139
Zebra-dsREDexpress		140
pep1-dsREDexpress		141
tp10-dsREDexpress		142
Zebra-Cas9		143
pVEC-Cas9		144

Подробное описание изобретения

Раскрытия всех процитированных документов, относящихся к патентной и непатентной литературе, включены в данный документ посредством ссылки во всей своей полноте.

Используемый в данном документе термин "изобретение" или "раскрытое изобретение" не предназначен для ограничения, напротив, в целом применим к любому из изобретений, определенных в формуле изобретения или описанных в данном документе. Эти термины используются в данном документе взаимозаменяемо.

Термином "клетка" согласно данному документу называется любой тип клетки, как, например, прокариотическая или эукариотическая клетка. Эукариотическая клетка имеет ядро и другие окруженные мембранами структуры (органеллы), в то время как у прокариотической клетки ядро отсутствует. В определенных вариантах осуществления клетка может быть клеткой млекопитающего или клеткой организма, не относящегося к млекопитающим. Клетки организма, не относящегося к млекопитающим, могут быть эукариотическими или прокариотическими. Например, клеткой организма, не относящегося к млекопитающим, согласно данному документу может называться микробная клетка или клетка многоклеточного организма, не относящегося к млекопитающим, такого как растение, насекомое, нематода, вид птицы, амфибия, рептилия или рыба.

Микробной клеткой согласно данному документу может называться, например, грибная клетка (например, дрожжевая клетка), прокариотическая клетка, клетка протистов (например, клетка водорослей), клетка эвгленовых, клетка страменопилов или клетка оомицетов. Прокариотической клеткой согласно данному документу может называться, например, бактериальная клетка или клетка архей. Грибные клетки (например, дрожжевые клетки), клетки протистов (например, клетки водорослей), клетки эвгленовых, клетки страменопилов и клетки оомицетов являются примерами эукариотических микробных клеток. Эукариотическая микробная клетка имеет ядро и другие окруженные мембранами структуры (органеллы), в то время как у прокариотической клетки ядро отсутствует.

Термином "дрожжи" согласно данному документу называют виды грибов, которые существуют преимущественно в одноклеточной форме. В качестве альтернативы дрожжи могут называться "дрожжевыми клетками". Дрожжи согласно данному документу могут быть охарактеризованы, например, как традиционные виды дрожжей или нетрадиционные виды дрожжей.

Термином "традиционные виды дрожжей" ("модельные дрожжи") согласно данному документу в целом называют виды дрожжей *Saccharomyces* или *Schizosaccharomyces*. В определенных вариантах осуществления традиционные виды дрожжей представляют собой дрожжи, в которых происходят процессы репарации ДНК путем гомологичной рекомбинации (HR), а не процессы репарации, опосредуемые негомологичным соединением концов (NHEJ).

Термином "нетрадиционные виды дрожжей" согласно данному документу называют любые дрожжи, которые не являются видами дрожжей *Saccharomyces* или *Schizosaccharomyces*. Нетрадиционные виды дрожжей описаны в *Non-Conventional Yeasts in Genetics, Biochemistry and Biotechnology: Practical Protocols* (K. Wolf, K.D. Breunig, G. Barth, Eds., Springer-Verlag, Berlin, Germany, 2003) и Spencer et al. (*Appl. Microbiol. Biotechnol.* 58:147-156), которые включены в данный документ посредством ссылки. В определенных вариантах осуществления нетрадиционные виды дрожжей могут дополнительно (или в качестве альтернативы) представлять собой дрожжи, в которых происходят процессы репарации ДНК

путем NHEJ, а не процессы репарации, опосредуемые HR. Определение термина нетрадиционные виды дрожжей по этому принципу предпочтительности NHEJ в сравнении с HR - дополнительно раскрыто в Chen et al. (PLoS ONE 8:e57952), который включен в данный документ посредством ссылки. Предпочтительными нетрадиционными видами дрожжей согласно данному документу являются дрожжи из рода *Yarrowia* (например, *Yarrowia lipolytica*).

Термином "растение" согласно данному документу называют растения целиком, органы растений, ткани растений, растительные клетки, семена или их потомство. Растительные клетки включают в себя без ограничения клетки из семян, суспензионных культур, зародышей, участков меристемы, каллюсной ткани, листьев, корней, побегов, гаметофитов, спорофитов, пыльцы и микроспор. Части растений включают в себя дифференцированные и недифференцированные ткани, в том числе без ограничений корни, стебли, побеги, листья, пыльцу, семена, опухолевую ткань и различные формы клеток и культур (например, отдельные клетки, протопласты, зародыши и каллюсную ткань). Растительная ткань может присутствовать в растении или в органе растения, ткани или культуре клеток. Термином "орган растения" называют растительную ткань или группу тканей, которые составляют морфологически и функционально отдельную часть растения. В данном документе термином "геном" называют полный комплект генетического материала (гены и некодирующие последовательности), который присутствует в каждой клетке организма, или вируса, или органеллы; и/или полный набор хромосом, наследуемый как (гаплоидная) единица от одного родителя. "Потомство" включает любое последующее поколение растения.

Трансгенное растение включает в себя, например, растение, которое содержит в своем геноме гетерологичный полинуклеотид, введенный с помощью стадии трансформации. Гетерологичный полинуклеотид может быть стабильно интегрирован в геном, так что полинуклеотид передается следующим поколениям. Гетерологичный полинуклеотид может быть интегрирован в геном в отдельности или как часть рекомбинантной ДНК-конструкции. Трансгенное растение также может содержать в своем геноме более одного гетерологичного полинуклеотида. Каждый гетерологичный полинуклеотид может придавать трансгенному растению свой признак. Гетерологичный полинуклеотид может включать последовательность, которая происходит из чужеродного вида или, если она происходит из того же вида, может быть существенно модифицирована по сравнению со своей нативной формой. Трансгенный растительный материал может включать любую клетку, клеточную линию, каллюс, ткань, часть растения или растение, генотип которых был изменен за счет присутствия гетерологичной нуклеиновой кислоты, в том числе трансгенных объектов, которые изначально были изменены таким образом, а также тех, которые были получены путем половых скрещиваний или бесполого размножения из исходного трансгенного объекта. Подразумевается, что изменения растительного генома (хромосомного или внехромосомного) с помощью традиционных способов селекции растений, процедур редактирования генома, описанных в данном документе, которые не приводят к вставке чужеродного полинуклеотида, или с помощью встречающихся в природе событий, таких как случайное перекрестное опыление, инфекция нереккомбинантного вируса, трансформация нереккомбинантной бактерией, нереккомбинантная транспозиция или спонтанная мутация, не рассматриваются как образование трансгенных объектов.

Фертильное растение представляет собой растение, которое производит жизнеспособные мужские и женские гаметы и является самоопыляемым. Такое самоопыляемое растение может произвести растение-потомка без участия гаметы и генетического материала, содержащегося в ней, от любого другого растения. Растения с мужской стерильностью включают в себя растения, которые не производят мужские гаметы, которые являются жизнеспособными или, иными словами, способными к оплодотворению. Растения с женской стерильностью включают в себя растения, которые не производят женские гаметы, которые являются жизнеспособными или, иными словами, способными к оплодотворению. Понятно, что растения с мужской стерильностью и растения с женской стерильностью, могут характеризоваться женской фертильностью и мужской фертильностью соответственно. Также понятно, что растение с мужской фертильностью (но с женской стерильностью) может производить жизнеспособное потомство при скрещивании с растением с женской фертильностью, и что растение с женской фертильностью (но мужской стерильностью) может производить жизнеспособное потомство при скрещивании с растением с мужской фертильностью.

Термином "направляемая РНК эндонуклеаза" (RGEN) согласно данному документу называют комплекс, содержащий по меньшей мере один CRISPR (короткие палиндромные повторы, регулярно расположенные группами)-ассоциированный (Cas) белок и по меньшей мере один РНК-компонент. Термины "белковый компонент из RGEN" и "белковый компонент RGEN" используются в данном документе взаимозаменяемо и относятся к белку Cas, который представляет собой эндонуклеазный компонент RGEN или образует его часть. В определенных вариантах осуществления белковый компонент может представлять собой полную эндонуклеазу (например, Cas9); причем такой белковый компонент в качестве альтернативы может называться "эндонуклеазным компонентом" RGEN. RGEN согласно данному документу, как правило, характеризуется специфической активностью нацеливания на ДНК, при условии ее ассоциации по меньшей мере с одним РНК-компонентом.

Вкратце, РНК-компонент RGEN содержит последовательность, которая комплементарна ДНК-последовательности в последовательности целевого сайта. На основе этой комплементарности RGEN

может специфически распознавать и расщеплять конкретную последовательность целевого сайта ДНК. RGEN согласно данному документу может содержать белок(белки) Cas и подходящий(подходящие) РНК-компонент(компоненты) любой из четырех известных систем CRISPR (Horvath and Barrangou, Science 327:167-170), таких как система CRISPR типа I, II или III. В предпочтительных вариантах осуществления RGEN содержит эндонуклеазу Cas9 (система CRISPR II) и по меньшей мере один РНК-компонент (например, crRNA и tracrRNA или gRNA).

Термином "CRISPR" (короткие палиндромные повторы, регулярно расположенные группами) называют определенные генетические локусы, кодирующие факторы системы расщепления ДНК класса I, II или III, например, используемые бактериальными клетками или клетками архей для разрушения чужеродной ДНК (Horvath and Barrangou, Science 327:167-170). Компоненты систем CRISPR используются в данном документе гетерологичным образом для нацеливания на ДНК в клетках.

Термины "система CRISPR типа II" и "система CRISPR-Cas типа II" используются в данном документе взаимозаменяемо и относятся к системе расщепления ДНК, использующей эндонуклеазу Cas9 в комплексе по меньшей мере с одним РНК-компонентом. Например, Cas9 может находиться в комплексе с РНК CRISPR (crRNA) и транскрибирующей РНК CRISPR (tracrRNA). В другом примере Cas9 может находиться в комплексе с направляющей РНК. Таким образом, crRNA, tracrRNA и направляющая РНК представляют собой неограничивающие примеры РНК-компонентов согласно данному документу.

В данном документе термином CRISPR-ассоциированная ("Cas") эндонуклеаза называют белок Cas, кодируемый геном Cas. В определенных вариантах осуществления эндонуклеаза Cas, находясь в комплексе с подходящим РНК-компонентом, способна расщеплять всю или часть специфической целевой последовательности ДНК. Например, она может быть способна ввести одно- или двухнитевой разрыв в специфическую целевую последовательность ДНК; в качестве альтернативы ее можно охарактеризовать как обладающую способностью расщеплять одну или обе нити специфической целевой последовательности ДНК. Эндонуклеаза Cas может раскручивать ДНК-дуплекс в месте целевой последовательности, и она расщепляет по меньшей мере одну нить ДНК, что опосредуется распознаванием целевой последовательности с помощью crRNA или направляющей РНК, которые находятся в комплексе с Cas. Такое распознавание и разрезание целевой последовательности эндонуклеазой Cas, как правило, происходит, если на 3'-конце целевой последовательности ДНК или вблизи него расположен правильный прилегающий к протоспейсеру мотив (PAM). В качестве альтернативы белок Cas согласно данному документу может не обладать активностью расщепления или внесения однонитевого разрыва ДНК, но может продолжать специфически связываться с целевой последовательностью ДНК, находясь в комплексе с подходящим РНК-компонентом. В данном документе предпочтительный белок Cas представляет собой Cas9.

В способах, раскрытых в данном документе, может применяться любая направляемая эндонуклеаза. Такие эндонуклеазы включают в себя без ограничений эндонуклеазы Cas9 и Cpf1. К настоящему времени было описано множество эндонуклеаз, которые могут распознавать специфические последовательности PAM (см., например, заявку на патент США 62/162377, поданную 15 мая 2015 г., и 62/162353, поданную 15 мая 2015 года, и Zetsche B et al. 2015. Cell 163, 1013) и расщеплять целевую ДНК в определенных положениях. Понятно, что на основании способов и вариантов осуществления, описанных в данном документе, в которых используется система направляемой Cas, в данной ситуации можно приспособить эти способы, чтобы в них использовалась любая система направляемой эндонуклеазы.

"Cas9" (ранее называемая как Cas5, Csn1 или Csx12) согласно данному документу называют эндонуклеазу Cas из системы CRISPR типа II, которая образует комплекс с crRNA и tracrRNA или с направляющей РНК для специфического распознавания и расщепления всей или части целевой последовательности ДНК. Белок Cas9 содержит домен нуклеазы RuvC и домен нуклеазы HNH (H-N-H), каждый из которых расщепляет одну нить ДНК в целевой последовательности (согласованное действие обоих доменов приводит к двухнитевому расщеплению ДНК, тогда как активность одного домена приводит к однонитевому разрыву). В целом, домен RuvC содержит субдомены I, II и III, при этом домен I расположен возле N-конца Cas9, а субдомены II и III расположены в средней части белка, фланкируя HNH-домен (Hsu et al, Cell 157:1262-1278). "Аро-Cas9" называют Cas9, которая не находится в комплексе с РНК-компонентом. Аро-Cas9 может связывать ДНК, но осуществляет это неспецифическим образом и не способна расщеплять ДНК (Sternberg et al., Nature 507:62-67).

Термином "РНК-компонент" согласно данному документу называют РНК-компонент RGEN, содержащий последовательность рибонуклеиновой кислоты, которая комплементарна нити целевой последовательности ДНК. Данную комплементарную последовательность называют в данном документе "направляющей последовательностью" или последовательностью "вариабельного нацеливающего домена" (фиг. 5). Примеры подходящих РНК-компонентов согласно данному документу включают в себя crRNA и направляющую РНК. В определенных вариантах осуществления РНК-компоненты (например, направляющая РНК в отдельности, crRNA+tracrRNA) могут придавать RGEN способность к специфическому нацеливанию на ДНК.

Термином "РНК CRISPR" (crRNA) согласно данному документу называют последовательность РНК, которая образует комплекс с одним или несколькими белками Cas (например, Cas9) и обеспечивает свойство специфичности связывания ДНК у данного комплекса. crRNA обеспечивает специфичность

связывания ДНК, поскольку она содержит "направляющую последовательность" ("вариабельный нацеливающий домен" [VT]), которая комплементарна нити целевой последовательности ДНК. crRNA дополнительно содержит "повторяющуюся последовательность" ("парную последовательность для tracrRNA"), кодируемую повторяющимся участком локуса CRISPR, из которого получена crRNA. Повторяющаяся последовательность crRNA может подвергаться отжигу с последовательностью на 5'-конце tracrRNA. crRNA в нативных системах CRISPR образуется из "pre-crRNA", транскрибируемой с локуса CRISPR. Pre-crRNA содержит спейсерные участки и повторяющиеся участки; при этом спейсерные участки содержат уникальную последовательность, комплементарную последовательности целевого сайта ДНК. В нативных системах pre-crRNA процессируется во множество различных crRNA, каждая с направляющей последовательностью вместе с частью, представляющей собой повторяющуюся последовательность. В системах CRISPR crRNA используется, например, для обеспечения специфичности нацеливания на ДНК.

Термином "транскрибирующая РНК CRISPR" (tracrRNA) согласно данному документу называют некодирующую РНК, используемую в системах CRISPR типа II и содержащую, в направлении от 5' к 3', (i) последовательность, которая подвергается отжигу с повторяющимся участком crRNA из CRISPR типа II и (ii) часть, содержащую структуру "стебель-петля" (Deltcheva et al., Nature 471:602-607).

Термины "направляющая РНК" (gRNA) и "одиночная направляющая РНК" (sgRNA) используются в данном документе взаимозаменяемо. gRNA согласно данному документу может называться химерная последовательность, содержащая crRNA, функционально связанную с tracrRNA. В качестве альтернативы gRNA может называться, например, синтетическое слияние crRNA и tracrRNA. gRNA также можно охарактеризовать как имеющую направляющую последовательность (вариабельный нацеливающий домен), за которой следует домен распознавания эндонуклеазы Cas (CER). CER-домен может содержать парную последовательность для tracrRNA, за которой следует последовательность tracrRNA.

Вместо РНК-компонента может необязательно применяться "ДНК CRISPR" (crDNA). crDNA имеет последовательность ДНК, соответствующую последовательности crRNA, раскрытой в данном документе. crDNA может применяться с tracrRNA в комплексе crDNA/tracrRNA, который, в свою очередь, может быть ассоциирован с белковым компонентом RGEN. В заявке на патент США № 61/953090 раскрыта crDNA и способы ее применения в нацеливании на ДНК, опосредованном RGEN. Подразумевается, что в данном документе любое раскрытие, которое касается crRNA, можно аналогичным образом применять к использованию crDNA соответственно. Таким образом, в данном документе в случае вариантов осуществления, включающих crDNA, "направляемая РНК эндонуклеаза" (RGEN) может вместо этого называться комплексом, содержащим по меньшей мере один белок Cas и по меньшей мере одну crDNA.

Используемый в данном документе термин "направляющий полинуклеотид" относится к полинуклеотидной последовательности, которая может образовывать комплекс с эндонуклеазой Cas и обеспечивает способность эндонуклеазы Cas распознавать и необязательно расщеплять целевой сайт ДНК. Направляющий полинуклеотид может представлять собой одиночную молекулу или димерную молекулу. Направляющая полинуклеотидная последовательность может представлять собой последовательность РНК, последовательность ДНК или их комбинацию (последовательность из комбинации РНК-ДНК). Необязательно, направляющий полинуклеотид может содержать по меньшей мере один нуклеотид, фосфодиэфирную связь или модификацию связи, такие как закрытая нуклеиновая кислота (LNA), 5-метил-dC, 2,6-диаминопурин, 2'-фтор А, 2'-фтор U, 2'-О-метил РНК, фосфоротиоатная связь, связь с молекулой холестерина, связь с молекулой полиэтиленгликоля, связь с молекулой спейсера 18 (цепь гексаэтиленгликоля) или 5'-3' ковалентную связь, приводящую к образованию кольца. Направляющий полинуклеотид, который содержит исключительно рибонуклеиновые кислоты, также называется "направляющей РНК".

Направляющий полинуклеотид может представлять собой димерную молекулу (также называемый дуплексным направляющим полинуклеотидом), содержащую первый домен нуклеотидной последовательности (называемый вариабельным нацеливающим доменом или VT-доменом), который является комплементарным нуклеотидной последовательности в целевой ДНК, и второй домен нуклеотидной последовательности (называемый доменом распознавания эндонуклеазы Cas или CER-доменом), который взаимодействует с полипептидом эндонуклеазы Cas. CER-домен направляющего полинуклеотида в виде димерной молекулы содержит две отдельные молекулы, которые гибридизируются вдоль участка комплементарности. Две отдельные молекулы могут представлять собой последовательности РНК, ДНК и/или комбинации РНК-ДНК. В некоторых вариантах осуществления первую молекулу из дуплексного направляющего полинуклеотида, содержащую VT-домен, связанный с CER-доменом ("cr-нуклеотид"), называют "crDNA" (если она состоит из непрерывного отрезка из ДНК-нуклеотидов), или "crRNA" (если она состоит из непрерывного отрезка из РНК-нуклеотидов), или "crDNA-RNA" (если она состоит из комбинации ДНК- и РНК-нуклеотидов). В некоторых вариантах осуществления вторую молекулу дуплексного направляющего полинуклеотида, содержащую CER-домен, называют "tracrRNA" (если она состоит из непрерывного отрезка из РНК-нуклеотидов), или "tracrDNA" (если она состоит из непрерывного отрезка из ДНК-нуклеотидов), или "tracrDNA-RNA" (если она состоит из комбинации ДНК- и РНК-нуклеотидов).

Направляющий полинуклеотид также может представлять собой одиночную молекулу, содержа-

щую первый домен нуклеотидной последовательности (называемый дариабельным нацеливающим доменом или VT-доменом), который является комплементарным нуклеотидной последовательности в целевой ДНК, и второй домен нуклеотидной последовательности (называемый доменом распознавания эндонуклеазы Cas или CER-доменом), который взаимодействует с полипептидом эндонуклеазы Cas. Под "доменом" имеют в виду непрерывный отрезок из нуклеотидов, который может представлять собой последовательность РНК, ДНК и/или комбинации РНК-ДНК. VT-домен и/или CER-домен одиночного направляющего полинуклеотида могут содержать последовательность РНК, последовательность ДНК или последовательность комбинации РНК-ДНК. В некоторых вариантах осуществления одиночный направляющий полинуклеотид содержит а сг-нуклеотид (содержащий VT-домен, связанный с CER-доменом), связанный с тгасг-нуклеотидом (содержащим CER-домен), где под связью подразумевают нуклеотидную последовательность, содержащую последовательность РНК, последовательность ДНК или последовательность комбинации РНК-ДНК. Одиночный направляющий полинуклеотид, состоящий из последовательностей сг-нуклеотида и тгасг-нуклеотида, может называться "одиночной направляющей РНК" (когда он состоит из непрерывного отрезка из РНК-нуклеотидов), или "одиночной направляющей ДНК" (когда он состоит из непрерывного отрезка из ДНК-нуклеотидов), или "одиночной направляющей РНК-ДНК" (когда он состоит из комбинации РНК- и ДНК-нуклеотидов).

Таким образом, в определенных вариантах осуществления направляющий полинуклеотид и эндонуклеаза Cas типа II могут образовывать комплекс друг с другом (называемый "комплекс направляющий полинуклеотид/эндонуклеаза Cas" или также называемый "система из направляющего полинуклеотида/эндонуклеазы Cas"), где комплекс направляющий полинуклеотид/эндонуклеаза Cas может управлять целенаправленным воздействием эндонуклеазы Cas на целевой сайт в геноме клетки (например, растительной клетки), необязательно обеспечивая способность эндонуклеазы Cas вводить одно- или двухнитевой разрыв в целевой сайт в геноме. Комплекс направляющий полинуклеотид/эндонуклеаза Cas может быть связан по меньшей мере с одним СРР, при этом такой комплекс способен связываться и необязательно создавать одно- или двухнитевой разрыв в целевом сайте клетки (например, растительной клетки).

Термины "вариабельный нацеливающий домен" или "VT-домен" используются в данном документе взаимозаменяемо, и так называют нуклеотидную последовательность, которая комплементарна одной нити (нуклеотидной последовательности) целевого сайта из двухнитевой ДНК. Процент комплементарности между первым доменом нуклеотидной последовательности (VT-доменом) и целевой последовательностью может составлять по меньшей мере 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 63%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или 100%. Длина вариабельного нацеливающего домена может составлять по меньшей мере 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 или 30 нуклеотидов. В некоторых вариантах осуществления вариабельный нацеливающий домен содержит непрерывный отрезок из 12-30 нуклеотидов. Вариабельный нацеливающий домен может состоять из последовательности ДНК, последовательности РНК, модифицированной последовательности ДНК, модифицированной последовательности РНК (см., например, модификации, описанные в данном документе) или любой их комбинации.

Термин "домен распознавания эндонуклеазы Cas" или "CER-домен" направляющего полинуклеотида используются в данном документе взаимозаменяемо и относятся к нуклеотидной последовательности (такой как второй домен нуклеотидной последовательности направляющего полинуклеотида), которая взаимодействует с полипептидом эндонуклеазы Cas. CER-домен может состоять из последовательности ДНК, последовательности РНК, модифицированной последовательности ДНК, модифицированной последовательности РНК (см., например, модификации, описанные в данном документе) или любой их комбинации.

Термины "целевой сайт", "целевая последовательность", "целевая ДНК", "целевая последовательность ДНК", "целевой локус", "протоспейсер" и т. п. используются в данном документе взаимозаменяемо. Последовательностью целевого сайта называют полинуклеотидную последовательность на хромосоме, эписоме или любой другой молекуле ДНК в геноме клетки, которую RGEN согласно данному документу может распознавать, с которой она может связываться и необязательно вносить одонитевой разрыв или расщеплять. Целевой сайт может представлять собой (i) эндогенный/нативный сайт в клетке, (ii) гетерологичный в отношении клетки и, следовательно, не встречающийся в геноме в естественном состоянии, или (iii) обнаруживаемый в местоположении в геноме, гетерологичном относительно того, где он встречается в нативном состоянии.

Длина последовательности целевого сайта согласно данному документу составляет по меньшей мере 13 нуклеотидов, и она имеет нить, которая в достаточной степени комплементарна направляющей последовательности (сгRNA или гRNA), чтобы иметь возможность гибридизироваться с направляющей последовательностью и управлять специфичным к последовательности связыванием белка Cas или комплекса белка Cas с целевой последовательностью (если подходящий РАМ прилепает к целевой последовательности в определенных вариантах осуществления). Сайт расщепления/одонитевого разрыва (применительно к Cas с эндонуклеолитической активностью или активностью образования одонитевого

разрыва) может находиться в пределах целевой последовательности (например, при применении Cas9), или сайт расщепления/однонитевого разрыва может находиться за пределами целевой последовательности (например, при применении Cas9, слитой с доменом гетерологичной эндонуклеазы, такой как полученный из фермента FokI). Также возможно, что последовательность целевого сайта будет связываться RGEN, у которой отсутствует активность расщепления или внесения однонитевого разрыва.

"Искусственным целевым сайтом" или "искусственной целевой последовательностью" согласно данному документу называют целевую последовательность, которая была введена в геном клетки. В некоторых вариантах осуществления последовательность искусственной целевой последовательности может быть идентична последовательности нативной целевой последовательности в геноме клетки, но располагаться в другом положении (гетерологичное положение) в геноме или отличаться от нативной целевой последовательности, если расположена в том же положении в геноме клетки.

"Эписомой" согласно данному документу называют молекулу ДНК, которая может существовать в клетке автономно (может реплицироваться и переходить в дочерние клетки) независимо от хромосом клетки. Эписомная ДНК может быть либо нативной, либо гетерологичной в отношении клетки. Примеры нативных эписом согласно данному документу включают в себя митохондриальную ДНК (mtDNA) и ДНК хлоропластов. Примеры гетерологичных эписом согласно данному документу включают в себя плазмиды и искусственные хромосомы дрожжей (YAC).

"Прилегающим к протоспейсеру мотивом" (PAM) согласно данному документу называют короткую последовательность, которая распознается RGEN согласно данному документу. Последовательность и длина PAM согласно данному документу могут отличаться в зависимости от применяемого белка Cas или комплекса белка Cas, но, как правило, он составляет, например, 2, 3, 4, 5, 6, 7 или 8 нуклеотидов в длину.

Термины "5'-кэп" и "7-метилгуанилатный (m⁷G) кэп" используются в данном документе взаимозаменяемо. 7-метилгуанилатный остаток расположен на 5'-конце РНК, транскрибированной с помощью РНК-полимеразы II (Pol II) в эукариотах. Кэпированная РНК согласно данному документу имеет 5'-кэп, тогда как некэпированная РНК не имеет такой кэп.

Термины "некэпированная", "не имеющая 5'-кэп" и т. п. используются в данном документе взаимозаменяемо, чтобы обозначить РНК, у которой отсутствует 5'-кэп и необязательно имеется, например, 5'-гидроксильная группа вместо 5'-кэпа. Некэпированная РНК может лучше накапливаться в ядре после транскрипции, тогда как 5'-кэпированная РНК подвергается экспорту из ядра.

Термины "рибозим", "фермент-рибонуклеиновая кислота" и "саморасщепляющийся рибозим" используются в данном документе взаимозаменяемо. Рибозимом называют одну или несколько последовательностей РНК, формирующих вторичную, третичную и/или четвертичную структуру(структуры), которые могут расщеплять РНК в специфическом сайте, в частности в цис-сайте относительно последовательности рибозима (т.е. являются автокаталитическими или саморасщепляющимися). Общая природа нуклеолитической природы рибозимов была описана (например, Lilley, Biochem. Soc. Trans. 39:641-646). "Рибозим типа hammerhead" (HHR) согласно данному документу может содержать небольшой каталитический мотив РНК, составленный из трех стеблей со спаренными основаниями и ядра из высоко консервативных, некоплементарных нуклеотидов, которые принимают участие в катализе. В Pley et al. (Nature 372:68-74) и Hammann et al. (RNA 18:871-885), которые включены в данный документ посредством ссылки, раскрыта структура и активность рибозима типа hammerhead. Рибозим типа hammerhead согласно данному документу может содержать, например, "минимальную последовательность типа hammerhead", как раскрыто в Scott et al. (Cell 81:991-1002, включенном в данный документ посредством ссылки).

Термины "нацеливание", "нацеливание на ген", "нацеливание на ДНК", "редактирование", "редактирование генов" и "редактирование ДНК" используются в данном документе взаимозаменяемо. Нацеливание на ДНК согласно данному документу может представлять собой специфическое введение вставки/делеции, нокаута или нокина в определенную последовательность, расположенную в хромосоме или эписоме клетки. В целом, нацеливание на ДНК согласно данному документу может осуществляться путем расщепления одной или обеих нитей в специфической последовательности ДНК в клетке с помощью белка Cas, ассоциированного с подходящим РНК-компонентом. Такое расщепление ДНК, в случае двунитевого разрыва (DSB), может стимулировать процессы NHEJ, которые могут приводить к образованию вставки/делеции в целевом сайте. Также, независимо от того, приводит ли расщепление к однонитевому разрыву (SSB) или DSB, могут стимулироваться процессы HR, если подходящий донорный полинуклеотид ДНК обеспечивается в сайте однонитевого разрыва или расщепления ДНК. Такой процесс HR можно применять для введения нокаута или нокина в целевой сайт, в зависимости от последовательности донорного полинуклеотида ДНК. В качестве альтернативы нацеливанием на ДНК согласно данному документу может называться специфическая ассоциация комплекса Cas/РНК-компонент согласно данному документу с целевой последовательностью ДНК, при этом белок Cas разрезает или не разрезает нить ДНК (в зависимости от статуса эндонуклеолитических доменов белка Cas).

Термином "вставка/делеция" согласно данному документу называют вставку или делецию нуклеотидного основания или оснований в целевой последовательности ДНК в хромосоме или эписоме. Такая вставка или делеция может состоять, например, из 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, или более оснований. В опре-

деленных вариантах осуществления вставка/делеция может быть даже более длинной, по меньшей мере составлять приблизительно 20, 30, 40, 50, 60, 70, 80, 90, или 100 оснований. Если вставку/делецию вводят в пределах открытой рамки считывания (ORF) гена, зачастую вставка/делеция нарушает экспрессию белка дикого типа, кодируемого ORF, создавая мутацию "сдвига рамки".

Термины "нокаут", "нокаут гена" и "генный нокаут" используются в данном документе взаимозаменяемо. Нокауту соответствует последовательность ДНК в клетке согласно данному документу, которая стала частично или полностью неработоспособной из-за нацеливания белка Cas; причем такая последовательность ДНК до нокаута, например, могла кодировать аминокислотную последовательность или могла обладать регуляторной функцией (например, промоторной). К нокауту может приводить вставка/делеция (за счет NHEJ, простимулированного расщеплением, опосредуемым Cas) или специфическое удаление последовательности (за счет HR, простимулированной расщеплением или образованием однонитевого разрыва, опосредуемыми Cas, когда также применяется подходящий донорный полинуклеотид ДНК), что приводит к снижению или полному нарушению функции последовательности в сайте нацеливания, примыкающем к нему или вблизи него. Подвергнутая нокауту полинуклеотидная последовательность ДНК согласно данному документу, в качестве альтернативы, охарактеризована, например, как частично или полностью разрушенная или подвергнутая понижающей регуляции.

Термины "нокин", "нокин гена" и "генетический нокин" используются в данном документе взаимозаменяемо. Нокину соответствует замещение или вставка последовательности ДНК в специфическую последовательность ДНК в клетке путем нацеливания с помощью белка Cas (путем HR, простимулированной расщеплением или образованием однонитевого разрыва, опосредованными Cas, когда также применяется подходящий донорный полинуклеотид ДНК). Примерами нокинов являются специфическая вставка гетерологичной последовательности, кодирующей аминокислоту, в кодирующий участок гена или специфическая вставка элемента регуляции транскрипции в локус гена.

Термины "донорный полинуклеотид", "донорная ДНК", "нацеливающий полинуклеотид" и "нацеливающая ДНК" используются в данном документе взаимозаменяемо. Донорным полинуклеотидом называют последовательность ДНК, содержащую по меньшей мере одну последовательность, которая гомологична последовательности в целевом сайте ДНК (например, последовательности, на которую специфически целенаправленно воздействует белок Cas согласно данному документу) или вблизи него. Подходящий донорный полинуклеотид может подвергаться HR с целевым сайтом ДНК, если целевой сайт содержит SSB или DSB (которые могут быть введены с помощью определенных белков Cas согласно данному документу, ассоциированных с соответствующим РНК-компонентом). "Гомологичная последовательность" в пределах донорного полинуклеотида согласно данному документу, например, может содержать или состоять из последовательности по меньшей мере из приблизительно 25 нуклеотидов, например, которая на 100% идентична последовательности в целевом сайте или вблизи него, или по меньшей мере на приблизительно 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, или 99% идентична последовательности в целевом сайте или вблизи него.

В определенных вариантах осуществления донорный полинуклеотид ДНК может иметь две гомологичные последовательности, разделенные последовательностью (или парой оснований), которая является гетерологичной относительно последовательности в целевом сайте. Эти две гомологичные последовательности такого донорного полинуклеотида могут называться "плечами гомологии", которые фланкируют гетерологичную последовательность. HR между целевым сайтом и донорным полинуклеотидом с двумя "плечами гомологии", как правило, приводит к замещению последовательности в целевом сайте на гетерологичную последовательность донорного полинуклеотида (последовательность целевого сайта, расположенная между последовательностями ДНК, гомологичными "плечам гомологии" донорного полинуклеотида, замещается гетерологичной последовательностью донорного полинуклеотида). В донорном полинуклеотиде с двумя "плечами гомологии" данные "плечи" могут быть разделены 1 или несколькими нуклеотидами (т. е. длина гетерологичной последовательности в донорном полинуклеотиде может составлять по меньшей мере 1 нуклеотид). Различные процедуры HR, которые можно осуществлять в клетке согласно данному документу, раскрыты, например, в *DNA Recombination: Methods and Protocols: 1st Edition* (H. Tsubouchi, Ed., Springer-Verlag, New York, 2011), который включен в данный документ посредством ссылки.

Термины "проникающий в клетку пептид" (CPP) и "домен белковой трансдукции" (PTD) используются в данном документе взаимозаменяемо. CPP называют пептид, как правило, длиной приблизительно 5-60 аминокислотных остатков, который может облегчать захват клеткой молекулярного груза, в частности, одного или нескольких белковых компонентов RGEN согласно данному документу (например, белка Cas9). Такой белковый груз может быть ассоциирован с одним или несколькими CPP за счет ковалентной или нековалентной связи. В определенных вариантах осуществления CPP также может быть охарактеризован как способный облегчить перемещение молекулярного груза через или пересечение им одного или нескольких из липидного бислоя, мицеллы, клеточной мембраны, мембраны органеллы, мембраны везикулы или клеточной стенки. В определенных вариантах осуществления CPP согласно данному документу может быть катионным, амфипатическим или гидрофобным. Примеры CPP, примененных в данном документе, и дополнительное описание CPP в целом раскрыты в Schmidt et al. (*FEBS Lett.* 584:1806-

1813), Holm et al. (Nature Protocols 1:1001-1005), Yandek et al. (Biophys. J. 92:2434-2444), Morris et al. (Nat. Biotechnol. 19:1173-1176), и публикации заявки на патент США № 2014/0068797, все из которых включены в данный документ посредством ссылки.

"Катионным" или "поликатионным" CPP согласно данному документу называют CPP с высоким относительным содержанием (по меньшей мере 60%) положительно заряженных аминокислот, таких как лизин (K), аргинин (R) и/или гистидин (H).

"Амфипатическим" или "амфифильным" CPP согласно данному документу называют CPP с аминокислотной последовательностью, содержащей перемежающийся паттерн из полярных/заряженных остатков и неполярных, гидрофобных остатков. В качестве альтернативы амфипатический CPP может быть охарактеризован как обладающий как гидрофильными, так и липофильными свойствами.

"Гидрофобный" или "липофильный" CPP согласно данному документу содержит преимущественно или исключительно неполярные остатки с низким суммарным зарядом и/или гидрофобными аминокислотными группами.

Термины "ковалентно связанный", "ковалентно прикрепленный", "ковалентно ассоциированный", "ковалентная связь", "ковалентное взаимодействие" и т.п. используются в данном документе взаимозаменяемо. Ковалентная связь согласно данному документу, например, может быть образована за счет пептидной связи(связей) или химической перекрестной сшивки(сшивок). Ковалентная связь может быть непосредственной, например, в случае непосредственной ковалентной связи (непосредственного связывания) между белковым компонентом RGEN и CPP (например, имеется химическая связь [обобществление электронов] между атомом белкового компонента RGEN и атомом CPP). В качестве альтернативы ковалентная связь может быть опосредованной, например, если белковый компонент RGEN и CPP связаны друг с другом посредством по меньшей мере одного промежуточного фактора. Такой промежуточный фактор или группа промежуточных факторов, которые сами по себе ковалентно связаны вместе, ковалентно связаны с белковым компонентом RGEN и CPP. Таким образом, промежуточный фактор или их группа могут быть охарактеризованы как мостик между белковым компонентом RGEN и CPP.

Термины "белок слияния", "белковое слияние", "химерный белок" и т. п. используются в данном документе взаимозаменяемо. Белок слияния согласно данному документу содержит по меньшей мере две различные (гетерологичные) аминокислотные последовательности, связанные вместе в пределах одного полипептида. Белки слияния, как правило, получают с помощью способов генной инженерии, в которых последовательности ДНК, кодирующие разные аминокислотные последовательности, соединяют вместе, чтобы они кодировали один белок, содержащий разные аминокислотные последовательности. Примеры белков слияния согласно данному документу включают в себя слияния белок RGEN-CPP (аминокислотная последовательность белка RGEN слита с аминокислотными последовательностями одного или нескольких CPP).

Термины "нековалентно связанный", "нековалентно прикрепленный", "нековалентно ассоциированный", "нековалентная связь", "нековалентное взаимодействие" и т. п. используются в данном документе взаимозаменяемо. Нековалентная связь согласно данному документу называют взаимодействие между атомами, при котором электроны не обобществляются. Данный тип взаимодействия является более слабым, чем ковалентная связь. Гидрофобные взаимодействия представляют пример нековалентной связи, которая может существовать между белковым компонентом RGEN и одним или несколькими CPP. Другие примеры нековалентных связей, которые могут быть применимы в данном документе, включают в себя электростатические силы (например, ионную, водородную связь) и Ван-дер-Ваальсовы силы (лондоновские дисперсионные силы).

Используемым в данном документе "комплексом белок RGEN-CPP" называют комплекс между белковым компонентом RGEN и по меньшей мере одним CPP, при этом RGEN и CPP взаимодействуют посредством ковалентной или нековалентной связи. Оба компонента RGEN и CPP в данном комплексе, как правило, сохраняют всю свою соответствующую активность/функцию, раскрываемую в данном документе, или некоторую ее часть (например, по меньшей мере 50%). Например, в варианте осуществления, в котором белковый компонент RGEN представляет собой Cas9, Cas9 в комплексе Cas9-CPP способна к ассоциации с подходящим РНК-компонентом (например, gRNA) и нацеливанию комплекса Cas9-CPP на целевой сайт ДНК в клетке.

Термины "пересекать", "перемещаться", "преодолевать", "проходить через" и т.п. используются в данном документе взаимозаменяемо.

Термины "клеточная мембрана", "плазматическая мембрана" и "цитоплазматическая мембрана" используются в данном документе взаимозаменяемо и обозначают биологическую мембрану, которая отделяет внутреннее содержимое клетки от ее внешней среды. Клеточная мембрана, как правило, содержит фосфолипидный бислой с белками, встроенными в него. Помимо ряда прочих функций клеточная мембрана может служить в качестве поверхности для прикрепления внеклеточных структур, таких как клеточная стенка или структуры гликокаликса. Подробная информация, касающаяся липидных бислоев клеточной мембраны, представлена в Molecular Biology of the Cell. 4th Edition (B. Alberts et al., Eds., Garland Science, New York, 2002), который включен в данный документ посредством ссылки.

Термином "клеточная стенка" согласно данному документу называют прочный, гибкий (но иногда

довольно жесткий) слой, который окружает некоторые типы клеток организма, не относящегося к млекопитающим (например, бактерий, растений, водорослей, грибов, таких как дрожжи). Она расположена снаружи клеточной мембраны и обеспечивает структурную опору и защиту для клеток. В определенных вариантах осуществления главной функцией клеточной стенки является содействие поддержанию осмотического давления в клетке. Стенки грибных клеток (например, дрожжевой клетки) обычно содержат хитин, а стенки клеток водорослей обычно содержат гликопротеины и полисахариды. Стенки растительных клеток обычно содержат преимущественно полисахариды с небольшими количествами других компонентов (например, сложных эфиров фенола, структурных белков). "Первичная клеточная стенка" и/или "вторичная клеточная стенка" могут применяться для характеристики стенки растительной клетки, при этом вторичная стенка расположена внутри первичной стенки. Главным компонентом вторичных стенок является лигнин. Клеточные стенки бактерий обычно в качестве основного составляющего содержат пептидогликан. В определенных аспектах, как например у бактерий, клеточная стенка может дополнительно содержать в своем наружном слое гликокаликс, который обычно представляет собой слой из полисахаридов.

Термином "домен лейциновой застежки" согласно данному документу называют домен димеризации, характеризующийся присутствием лейцинового остатка через каждые семь остатков в отрезке из примерно 35 остатков. Домены лейциновой застежки образуют димеры, удерживаемые вместе с помощью альфа-спиральной суперспирали. Суперспираль имеет 3,5 остатка на оборот, что означает, что каждый седьмой остаток занимает эквивалентное положение с точки зрения оси спирали. Регулярное расположение лейцинов внутри суперспирали стабилизирует структуру посредством гидрофобных и Ван-дер-Ваальсовых взаимодействий.

Термины "процент по объему", "объемный процент", "об. %" и "об./об. %" используются в данном документе взаимозаменяемо. Процент по объему растворенного вещества в растворе может быть определен с помощью формулы: $[(\text{объем растворенного вещества})/(\text{объем раствора})] \times 100\%$.

Термины "процент по весу", "весовая процентная доля (вес. %)" и "процентная доля вес-вес (вес./вес. %)" используются в данном документе взаимозаменяемо. Процент по весу относится к проценту вещества на весовой основе, когда оно содержится в композиции, смеси или растворе.

Термины "полинуклеотид", "полинуклеотидная последовательность" и "последовательность нуклеиновой кислоты" используются в данном документе взаимозаменяемо. Эти термины охватывают нуклеотидные последовательности и т. п. Полинуклеотид может представлять собой полимер ДНК или РНК, которые являются одно- или двухнитевыми, необязательно содержат синтетические, не встречающиеся в природе или измененные нуклеотидные основания. Полинуклеотид может состоять из одного или нескольких сегментов кДНК, геномной ДНК, синтетической ДНК или их смесей. Нуклеотиды (рибонуклеотиды или дезоксирибонуклеотиды) могут называться с помощью следующих однобуквенных обозначений: "А" применительно к аденилату или дезоксиаденилату (в случае РНК или ДНК соответственно), "С" применительно к цитидилату или дезоцицитидилату (в случае РНК или ДНК соответственно), "G" применительно к гуанилату или дезоксигуанилату (в случае РНК или ДНК соответственно), "U" применительно к уридилату (в случае РНК), "T" применительно к дезокситимидилату (в случае ДНК), "R" применительно к пуринам (А или G), "Y" применительно к пиримидинам (С или Т), "K" применительно к G или Т, "H" применительно к А, или С, или Т, "I" применительно к инозину, "W" применительно к А или Т и "N" применительно к любому нуклеотиду (например, N может представлять собой А, С, Т или G при ссылке на последовательность ДНК; N может представлять собой А, С U или G при ссылке на последовательность РНК). Любая последовательность РНК (например, crRNA, tracrRNA, gRNA), раскрытая в данном документе, может быть закодирована с помощью соответствующей последовательности ДНК.

Используемым в данном документе термином "выделенная" называют полинуклеотидную или полипептидную молекулу, которая была полностью или частично очищена из ее нативного источника. В некоторых случаях выделенная полинуклеотидная или полипептидная молекула является частью большей композиции, буферной системы или смеси реагентов. Например, выделенная полинуклеотидная или полипептидная молекула может содержаться в клетке или организме гетерологичным образом. Композиции согласно данному документу, содержащие белковый компонент RGEN и проникающий в клетку пептид, могут считаться выделенными композициями. Эти композиции содержат гетерологичные компоненты и не встречаются в природе.

Используемым в данном документе термином "ген" называют полинуклеотидную последовательность ДНК, с которой экспрессируется РНК (РНК транскрибируется с полинуклеотидной последовательности ДНК) с кодирующего участка, при этом РНК может представлять собой матричную РНК (кодирующую белок) или РНК, не кодирующую белок (например, crRNA, tracrRNA или gRNA согласно данному документу). Геном может называться кодирующий участок сам по себе, или он может включать регуляторные последовательности, расположенные выше и/или ниже кодирующего участка (например, промоторы, 5'-нетранслируемые участки, 3'-участки терминации транскрипции). В качестве альтернативы кодирующий участок, который кодирует белок, может называться в данном документе "открытой рамкой считывания" (ORF). Геном, который является "нативным" или "эндогенным" называет ген, который встречается в природе со своими собственными регуляторными последовательностями; причем та-

кой ген находится в своем естественном положении в геноме клетки-хозяина. "Химерным" геном называют любой ген, который не является нативным геном, содержащий регуляторные и кодирующие последовательности, которые не встречаются в природе вместе (т. е. регуляторные и кодирующие участки, которые гетерологичны относительно друг друга). Соответственно, химерный ген может содержать регуляторные последовательности и кодирующие последовательности, которые получены из различных источников, или регуляторные последовательности и кодирующие последовательности, которые получены из одного и того же источника, но расположены в порядке, отличающемся от встречающегося в природе. "Чужеродным" или "гетерологичным" геном называют ген, который вводят в организм-хозяин с помощью переноса генов.

Чужеродные/гетерологичные гены могут предусматривать нативные гены, помещенные в организм, не являющийся для них нативным, нативные гены, введенные в новое положение в нативном хозяине, или химерные гены. В определенных вариантах осуществления, раскрытых в данном документе, полинуклеотидные последовательности являются гетерологичными. "Трансген" представляет собой ген, который был введен в геном с помощью процедуры доставки генов (например, трансформации). "Кодон-оптимизированная" открытая рамка считывания характеризуется частотой использования кодонов, разработанной для имитации частоты предпочтительного использования кодонов в клетке-хозяине.

"Мутированный ген" представляет собой ген, который был изменен за счет вмешательства человека. Такой "мутированный ген" имеет последовательность, которая отличается от последовательности соответствующего немутированного гена по меньшей мере добавлением, делецией или заменой одного нуклеотида. В определенных вариантах осуществления настоящего раскрытия мутированный ген содержит изменение, которое получено с помощью применения системы направляющий полинуклеотид/эндонуклеаза Cas, раскрытой в данном документе. Мутированное растение представляет собой растение, содержащее по меньшей мере один мутированный ген.

"Ненативная" аминокислотная последовательность или полинуклеотидная последовательность, содержащиеся в клетке или в организме согласно данному документу, не присутствуют в нативном (естественном) аналоге такой клетки или организма.

Используемыми в данном документе "регуляторными последовательностями" называют нуклеотидные последовательности, расположенные выше сайта инициации транскрипции (например, промотор), 5'-нетранслируемые участки и 3'-некодирующие участки гена и последовательности, которые могут оказывать влияние на транскрипцию, процессинг или стабильность, или трансляцию РНК, транскрибированной с гена. Регуляторные последовательности согласно данному документу могут включать промоторы, энхансеры, сайленсеры, 5'-нетранслируемые лидерные последовательности, интроны, последовательности распознавания полиаденилирования, сайты процессинга РНК, сайты связывания с эффектором, структуры типа "стебель-петля" и другие элементы, вовлеченные в регуляцию генной экспрессии. Один или несколько регуляторных элементов согласно данному документу могут быть гетерологичными относительно кодирующего участка.

Используемым в данном документе термином "промотор" называют последовательность ДНК, которая способна осуществлять контроль транскрипции РНК с гена. Обычно промоторная последовательность находится в гене выше сайта инициации транскрипции. Промоторы могут быть получены целиком из нативного гена или могут состоять из разных элементов, полученных из разных промоторов, обнаруживаемых в природе, или даже содержать сегменты синтетической ДНК. Промоторы, которые вызывают экспрессию гена в клетке в большинстве случаев при всех условиях, обычно называют "конститутивными промоторами". Один или несколько промоторов согласно данному документу могут быть гетерологичными относительно кодирующего участка.

Используемым в данном документе "сильным промотором" называют промотор, который может приводить к относительно большому количеству продуктивных инициации транскрипции за единицу времени, и/или он представляет собой промотор, приводящий к более высокому уровню транскрипции гена, чем средний уровень транскрипции генов в клетке.

Растительный промотор представляет собой промотор, способный к инициации транскрипции в растительной клетке; обзор растительных промоторов см. в Potenza et al., (2004) *In Vitro Cell Dev Biol* 40:1-22. Конститутивные промоторы включают в себя, например, коровый промотор промотора Rsyn7 и другие конститутивные промоторы, раскрытые в WO99/43838 и патенте США № 6072050; коровый промотор CaMV 35S (Odell et al., (1985) *Nature* 313:810-2); актиновый промотор риса (McElroy et al., (1990) *Plant Cell* 2:163-71); убиквитиновый промотор (Christensen et al., (1989) *Plant Mol Biol* 12:619-32; Christensen et al., (1992) *Plant Mol Biol* 18:675-89); pEMU (Last et al., (1991) *Theor Appl Genet* 81:581-8); MAS (Velten et al., (1984) *EMBO J* 3:2723-30); промотор ALS (патент США № 5659026) и т. п. Другие конститутивные промоторы описаны, например, в патентах США №№ 5608149; 5608144; 5604121; 5569597; 5466785; 5399680; 5268463; 5608142 и 6177611. В некоторых примерах может применяться индуцируемый промотор. Индуцируемые патогеном промоторы, которые индуцируются после заражения патогеном, включают в себя без ограничений промоторы, регулирующие экспрессию белков PR, белков SAR, бета-1,3-глюканазы и т.д.

Регулируемые химическими веществами промоторы можно применять для модуляции экспрессии

гена в растении посредством применения экзогенного химического регулятора. Промотор может представлять собой индуцируемый химическим веществом промотор, если применение химического вещества индуцирует экспрессию гена, или репрессируемый химическим веществом промотор, если применение химического вещества подавляет экспрессию гена. Индуцируемые химическими веществами промоторы включают в себя без ограничений промотор In2-2 маиса, активируемый антитодами к бензолсульфонамидным гербицидам (De Veylder et al., (1997) *Plant Cell Physiol* 38:568-77), промотор GST маиса (GST-II-27, WO93/01294), активируемый гидрофобными электрофильными соединениями, применяемыми в качестве дозыходовых гербицидов, и промотор PR-1a табака (Ono et al., (2004) *Biosci Biotechnol Biochem* 68:803-7), активируемый салициловой кислотой. Другие регулируемые химическими веществами промоторы включают в себя чувствительные к стероидам промоторы (см., например, индуцируемый глюкокортикоидами промотор (Scheda et al., (1991) *Proc. Natl. Acad. Sci. USA* 88:10421-5; McNellis et al., (1998) *Plant J* 14:247-257); и индуцируемые тетрациклином и репрессируемые тетрациклином промоторы (Gatz et al., (1991) *Mol Gen Genet* 227:229-37; патенты США №№ 5814618 и 5789156).

Промоторы, активные преимущественно в определенной ткани, можно использовать для целенаправленного усиления экспрессии в пределах конкретной растительной ткани. Промоторы, активные преимущественно в определенной ткани, включают в себя, например, описанные в Kawamata et al., (1997) *Plant Cell Physiol* 38:792-803; Hansen et al., (1997) *Mol Gen Genet* 254:337-43; Russell et al., (1997) *Transgenic Res* 6:157-68; Rinehart et al., (1996) *Plant Physiol* 112:1331-41; Van Camp et al., (1996) *Plant Physiol* 112:525-35; Canevascini et al., (1996) *Plant Physiol* 112:513-524; Lam, (1994) *Results Probl Cell Differ* 20:181-96; и Guevara-Garcia et al., (1993) *Plant J* 4:495-505. Промоторы, активные преимущественно в листьях, включают в себя, например, описанные в Yamamoto et al., (1997) *Plant J* 12:255-65; Kwon et al., (1994) *Plant Physiol* 105:357-67; Yamamoto et al., (1994) *Plant Cell Physiol* 35:773-8; Gotor et al., (1993) *Plant J* 3:509-18; Orozco et al., (1993) *Plant Mol Biol* 23:1129-38; Matsuoka et al., (1993) *Proc. Natl. Acad. Sci. USA* 90:9586-90; Simpson et al., (1958) *EMBO J* 4:2723-9; Timko et al., (1988) *Nature* 318:57-8.

Промоторы, активные преимущественно в корнях, включают в себя, например, описанные в Hire et al., (1992) *Plant Mol Biol* 20:207-18 (ген, специфической в отношении корня глутаминсинтазы сои); Miao et al., (1991) *Plant Cell* 3:11-22 (цитозольная глутаминсинтаза (GS)); Keller and Baumgartner, (1991) *Plant Cell* 3:1051-61 (специфический в отношении корня контрольный элемент в гене GRP 1.8 фасоли); Sanger et al., (1990) *Plant Mol Biol* 14:433-43 (специфический в отношении корня промотор маннопинсинтазы (MAS) *A. tumefaciens*); Bogusz et al., (1990) *Plant Cell* 2:633-41 (специфические в отношении корня промоторы, выделенные из *Parasponia andersonii* и *Trema tomentosa*); Leach and Aoyagi, (1991) *Plant Sci* 79:69-76 (гены *rolC* и *rolD* *A. rhizogenes*, индуцирующие разрастание корней); Teeri et al., (1989) *EMBO J* 8:343-50 (индуцируемые ранением гены TR1' и TR2' *Agrobacterium*); промотор гена VfENOD-GRP3 (Kuster et al., (1995) *Plant Mol Biol* 29:759-72); и промотор *rolB* (Capana et al., (1994) *Plant Mol Biol* 25:681-91); промотор гена фазеолина (Murai et al., (1983) *Science* 23:476-82; Sengopta-Gopalen et al., (1988) *Proc. Natl. Acad. Sci. USA* 82:3320-4). См. также патенты США №№ 5837876; 5750386; 5633363; 5459252; 5401836; 5110732 и 5023179.

Промоторы, активные преимущественно в семенах, включают в себя как специфические в отношении семян промоторы, активные во время развития семян, так и промоторы "прорастания семян", активные во время прорастания семян. См. Thompson et al., (1989) *BioEssays* 10:108. Промоторы, активные преимущественно в семенах, включают в себя без ограничения *Cim1* (индуцируемый цитокинином транскрипт); CZ19B1 (19 кДа зеин маиса) и *milps* (миоинозитол-1-фосфатсинтаза); (WO00/11177 и патент США № 6225529). В случае двудольных растений промоторы, активные преимущественно в семенах, включают в себя без ограничения промоторы гена бета-фазеолина бобовых, гена напина, гена бета-конглицинина, гена лектина сои, гена круциферина и т. п. В случае однодольных растений промоторы, активные преимущественно в семенах, включают в себя без ограничений промоторы гена 15 кДа зеина маиса, 22 кДа зеина, 27 кДа гамма-зеина, *waxy*, *shrunk1*, *shrunk2*, глобулина 1, олеозина и *nucl*. См. также WO00/12733, в которой раскрыты промоторы, активные преимущественно в семенах, из генов END1 и END2.

Используемыми в данном документе терминами "3'-некодирующая последовательность", "терминатор транскрипции" и "терминатор" называют последовательности ДНК, расположенные ниже кодирующей последовательности. Они включают в себя последовательности распознавания полиаденилирования и другие последовательности, кодирующие регуляторные сигналы, способные воздействовать на процессинг мРНК или экспрессию гена.

Используемым в данном документе термином "кассета" называют промотор, функционально связанный с последовательностью ДНК, кодирующей РНК, кодирующую белок, или РНК, не кодирующую белок.

Кассета необязательно может быть функционально связана с 3'-некодирующей последовательностью.

Терминами "выше" и "ниже", используемыми в данном документе в отношении полинуклеотидов, называют "в направлении 5' от" или "в направлении 3' от" соответственно.

Используемым в данном документе термином "экспрессия" называют (i) транскрипцию РНК (на-

пример, мРНК или РНК, не кодирующей белок, такой как crRNA, tracrRNA или gRNA) с кодирующего участка или (ii) трансляцию полипептида с мРНК.

При применении для описания экспрессии гена или полинуклеотидной последовательности термины "понижающая регуляция", "нарушение", "подавление", "инактивация" и "сайленсинг" используются в данном документе взаимозаменяемо для обозначения случаев, при которых транскрипция полинуклеотидной последовательности снижена или не происходит. Это приводит к снижению или отсутствию образования РНК-транскриптов за счет полинуклеотидной последовательности, что приводит к снижению или отсутствию экспрессии белка, получаемого за счет полинуклеотидной последовательности (если ген содержит ORF). В качестве альтернативы к понижающей регуляцией могут относиться случаи, при которых снижена или отсутствует трансляция белка за счет транскриптов, полученных с помощью полинуклеотидной последовательности. В качестве альтернативы, однако, понижающая регуляция может относиться к случаям, при которых белок, экспрессируемый за счет полинуклеотидной последовательности, характеризуется сниженной активностью. Снижение в случае любого из вышеупомянутых процессов (транскрипция, трансляция, активность белка) в клетке может составлять приблизительно 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% или 100% относительно транскрипции, трансляции или активности белка соответствующей контрольной клетки. Понижающая регуляция может быть, например, результатом события нацеливания, раскрытого в данном документе (например, вставка/делеция, нокаут).

Термины "контрольная клетка" и "соответствующая контрольная клетка" используются в данном документе взаимозаменяемо и могут применяться при ссылке на клетку (т.е. "экспериментальную клетку"), в которой была произведена модификация (например, сверхэкспрессия полинуклеотида, понижающая регуляция полинуклеотида). Контрольная клетка может представлять собой любую клетку, которая не имеет конкретной модификации экспериментальной клетки или в которой такая модификация не экспрессируется. Таким образом, контрольная клетка может представлять собой нетрансформированную клетку дикого типа или может представлять собой генетически трансформированную клетку, в которой не происходит экспрессия генетической трансформации. Например, контрольная клетка может представлять собой непосредственную материнскую клетку экспериментальной клетки, при этом непосредственная материнская клетка не имеет конкретной модификации, которая имеется у экспериментальной клетки. В качестве альтернативы контрольная клетка может представлять собой материнскую клетку экспериментальной клетки, которая отделена одним или несколькими поколениями. В качестве альтернативы, однако, контрольная клетка может представлять собой сестринскую клетку экспериментальной клетки, при этом сестринская клетка не содержит конкретной модификации, которая присутствует в экспериментальной клетке.

Используемый в данном документе термин "повышенный" может относиться к количеству или активности, которые по меньшей мере на приблизительно 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 50%, 100% или 200% превышают количество или активность, с которой сравнивают повышенное количество или активность. Термины "повышенный", "возросший", "усиленный", "большой чем" и "улучшенный" используются в данном документе взаимозаменяемо. В данном документе термин "повышенный" может применяться, например, чтобы охарактеризовать экспрессию полинуклеотида, кодирующего белок, при этом "повышенная экспрессия" может также означать "сверхэкспрессию".

Используемым в данном документе термином "функционально связанный" называют ассоциацию двух или более последовательностей нуклеиновых кислот, так что на функцию одной влияет другая. Например, промотор функционально связан с кодирующей последовательностью, когда он способен воздействовать на экспрессию данной кодирующей последовательности. То есть, кодирующая последовательность находится под транскрипционным контролем промотора. Например, кодирующие последовательности могут быть функционально связаны с регуляторными последовательностями. Также, например, crRNA может быть функционально связана (слита) с tracrRNA согласно данному документу, так что парная последовательность для tracrRNA из crRNA подвергается отжигу с 5'-последовательностью tracrRNA. Такая функциональная связь может предусматривать соответствующую последовательность, образующую петлю, такую как GAAA (SEQ ID NO: 36), CAAA (SEQ ID NO: 37) или AAAG (SEQ ID NO: 38). Также, например, RGEN может быть функционально связана (слита) с одним или несколькими CPP.

Используемым в данном документе термином "рекомбинантный" называют искусственную комбинацию двух в иных случаях разделенных сегментов последовательности, например, полученную с помощью химического синтеза или с помощью манипуляции с выделенными сегментами нуклеиновых кислот посредством методик генной инженерии.

Способы получения рекомбинантных конструкций/векторов согласно данному документу (например, полинуклеотида ДНК, кодирующего кассету с РНК-компонентом согласно данному документу, или полинуклеотида ДНК, кодирующего белок Cas или белок слияния Cas-CPP согласно данному документу) можно осуществлять в соответствии со стандартными методами рекомбинантной ДНК и молекулярного клонирования, описанными в J. Sambrook and D. Russell (Molecular Cloning: A Laboratory Manual, 3rd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001); T.J. Silhavy et al. (Experiments with Gene Fusions, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1984); и F.M. Ausubel et al.

(Short Protocols in Molecular Biology, 5th Ed. Current Protocols, John Wiley and Sons, Inc., NY, 2002), например.

Используемым в данном документе термином "трансформация" называют перенос молекулы нуклеиновой кислоты в организм-хозяин или клетку-хозяин с помощью любого способа. Молекула нуклеиновой кислоты, которая была трансформирована в организм/клетку, может представлять собой молекулу, которая реплицируется автономно в организме/клетке, или которая интегрируется в геном организма/клетки, или которая временно находится в клетке без репликации или интеграции. В данном документе раскрыты неограничивающие примеры молекул нуклеиновой кислоты, подходящих для трансформации, такие как плазмиды и линейные молекулы ДНК.

"Трансгенное растение" согласно данному документу включает, например, растение, которое содержит в своем геноме гетерологичный полинуклеотид, введенный с помощью стадии трансформации. Гетерологичный полинуклеотид может быть стабильно интегрирован в геном, так что полинуклеотид передается следующим поколениям. Гетерологичный полинуклеотид может быть интегрирован в геном в отдельности или как часть рекомбинантной ДНК-конструкции. Трансгенное растение также может содержать в своем геноме более одного гетерологичного полинуклеотида. Каждый гетерологичный полинуклеотид может придавать трансгенному растению свой признак. Трансгенный растительный материал может включать любую клетку, клеточную линию, каллус, ткань, часть растения или растение, генотип которых был изменен за счет присутствия гетерологичной нуклеиновой кислоты, в том числе трансгенных объектов, которые изначально были изменены таким образом, а также тех, которые были получены путем половых скрещиваний или бесполого размножения из исходного трансгенного объекта. Изменения генома (хромосомного или внехромосомного) с помощью традиционных способов селекции растений, процедур редактирования генома, которые не приводят к вставке чужеродного полинуклеотида, или с помощью встречающихся в природе событий, таких как случайное перекрестное опыление, инфекция нереккомбинантного вируса, трансформация нереккомбинантной бактерией, нереккомбинантная транспозиция или спонтанная мутация, не рассматриваются как образование трансгенных объектов.

"Фенотипический маркер" представляет собой поддающийся скринингу или селективируемый маркер, который включает визуальные маркеры и селективируемые маркеры, независимо от того, является ли он маркером, используемым в положительной или отрицательной селекции. Можно применять любой фенотипический маркер. В частности, селективируемый или поддающийся скринингу маркер включает в себя сегмент ДНК, который позволяет идентифицировать или провести положительную или отрицательную селекцию молекулы или клетки, которая его содержит, часто в конкретных условиях. Эти маркеры могут отвечать за активность, такую как без ограничения выработка РНК, пептида или белка, или могут обеспечивать сайт связывания для РНК, пептидов, белков, неорганических и органических соединений или композиций и т.п.

Примеры селективируемых маркеров включают в себя без ограничения сегменты ДНК, которые содержат сайты для ферментов рестрикции; сегменты ДНК, которые кодируют продукты, обеспечивающие устойчивость к соединениям, в иных случаях токсичным, в том числе антибиотикам, таким как спектиномицин, ампициллин, канамицин, тетрациклин, Basta, неомицинофосфотрансферазу II (NEO) и гигромицинофосфотрансферазу (HPT); сегменты ДНК, которые кодируют продукты, в иных случаях отсутствующих в клетке-реципиенте (например, гены тРНК, маркеры ауксотрофности); сегменты ДНК, кодирующие продукты, которые можно легко идентифицировать (например, фенотипические маркеры, такие как бета-галактозидаза, GUS; флуоресцентные белки, такие как зеленый (GFP), голубой (CFP), желтый (YFP), красный (RFP) флуоресцентные белки и белки клеточной поверхности); образование новых сайтов связывания праймеров для ПЦР (например, расположение рядом двух последовательностей ДНК, ранее рядом не располагавшихся), включение последовательностей ДНК, не подвергавшихся воздействию или подвергавшихся воздействию рестриктазы или другого модифицирующего ДНК фермента, химического вещества и т. д.; и включение последовательностей ДНК, необходимых для специфической модификации (например, метилирования), что обеспечивает возможность идентификации.

Дополнительные селективируемые маркеры включают в себя гены, которые придают устойчивость к гербицидным соединениям, таким как глюофосинат аммония, бромксинил, имидазолиноны и 2,4-дихлорфеноксиацетат (2,4-D). См., например, Yarranton, (1992) *Curr Opin Biotech* 3:506-11; Christopherson et al., (1992) *Proc. Natl. Acad. Sci. USA* 89:6314-8; Yao et al., (1992) *Cell* 71:63-72; Reznikoff, (1992) *Mol Microbiol* 6:2419-22; Hu et al., (1987) *Cell* 48:555-66; Brown et al., (1987) *Cell* 49:603-12; Figge et al., (1988) *Cell* 52:713-22; Deuschle et al., (1989) *Proc. Natl. Acad. Sci. USA* 86:5400-4; Fuerst et al., (1989) *Proc. Natl. Acad. Sci. USA* 86:2549-53; Deuschle et al., (1990) *Science* 248:480-3; Gossen, (1993) Диссертация на соискание научной степени Доктора философии, Гейдельбергский университет; Reines et al., (1993) *Proc. Natl. Acad. Sci. USA* 90:1917-21; Labow et al., (1990) *Mol Cell Biol* 10:3343-56; Zambretti et al., (1992) *Proc. Natl. Acad. Sci. USA* 89:3952-6; Baim et al., (1991) *Proc. Natl. Acad. Sci. USA* 88:5072-6; Wyborski et al., (1991) *Nucleic Acids Res* 19:4647-53; Hillen and Wissman, (1989) *Topics Mol Struc Biol* 10:143-62; Degenkolb et al., (1991) *Antimicrob Agents Chemother* 35:1591-5; Kleinschmidt et al., (1988) *Biochemistry* 27:1094-104; Bonin, (1993) Диссертация на соискание научной степени Доктора философии, Гейдельбергский университет; Gossen et al., (1992) *Proc. Natl. Acad. Sci. USA* 89:5547-51; Oliva et al., (1992) *Antimicrob Agents Chemother*

36:913-9; Hlavka et al., (1985) Handbook of Experimental Pharmacology, Vol. 78 (Springer-Verlag, Berlin); Gill et al., (1988) Nature 334:721-4.

Используемыми в данном документе терминами "идентичность последовательности" или "идентичность" применительно к полинуклеотидной или полипептидной последовательностям называют остатки нуклеиновой кислоты или аминокислотные остатки в двух последовательностях, которые являются одинаковыми, в случае выравнивания для максимального соответствия в пределах определенного окна сравнения. Таким образом, "процентной долей идентичности последовательностей" или "процентом идентичности" называют значение, определяемое путем сравнения двух последовательностей при оптимальном выравнивании в окне сравнения, при этом часть полинуклеотидной или полипептидной последовательности в окне сравнения может содержать добавления или делеции (т. е. гэпы) по сравнению с эталонной последовательностью (которая не содержит добавлений или делеций), чтобы достичь оптимального выравнивания двух последовательности. Процентная доля рассчитывается путем определения числа положений, в которых в обеих последовательностях находится идентичное основание нуклеиновой кислоты или аминокислотный остаток, с получением числа совпадающих положений, деления числа совпадающих положений на общее число положений в окне сравнения и умножения результатов на 100 с получением процентной доли идентичности последовательностей. Будет понятно, что при подсчете идентичности последовательностей между последовательностью ДНК и последовательностью РНК остатки Т в последовательности ДНК выравниваются и считаются "идентичными" с остатками U в последовательности РНК. С целью определения процента комплементарности первого и второго полинуклеотидов можно воспользоваться, например, определением (i) процента идентичности между первым полинуклеотидом и последовательностью, комплементарной второму полинуклеотиду (или наоборот), и/или (ii) процентной доли оснований в первом и втором полинуклеотидах, которые будут формировать канонические пары оснований по Уотсону-Крику.

Алгоритм средства поиска основного локального выравнивания (BLAST), который доступен онлайн на веб-сайте Национального центра биотехнологической информации (NCBI), можно применять, например, для измерения процента идентичности между или у двух или более полинуклеотидных последовательностей (алгоритм BLASTN) или полипептидных последовательностей (алгоритм BLASTP), раскрытых в данном документе. В качестве альтернативы процент идентичности между последовательностями можно определять с помощью алгоритма Clustal (например, ClustalW или ClustalV). Для множественных выравниваний с применением способа выравнивания Clustal значения по умолчанию могут соответствовать GAP PENALTY=10 и GAP LENGTH PENALTY=10. Параметры по умолчанию для парных выравниваний и расчета процента идентичности белковых последовательностей с применением способа Clustal могут представлять собой KTUPLE=1, GAP PENALTY=3, WINDOW=5 и DIAGONALS SAVED=5. Для нуклеиновых кислот эти параметры могут представлять собой KTUPLE=2, GAP PENALTY=5, WINDOW=4 и DIAGONALS SAVED=4. В качестве альтернативы, однако, процент идентичности последовательностей можно определять с помощью алгоритма EMBOSS (например, needle) с параметрами, такими как GAP OPEN=10, GAP EXTEND=0,5, END GAP PENALTY=false, END GAP OPEN=10, END GAP EXTEND=0,5 с применением матрицы BLOSUM (например, BLOSUM62).

В данном документе первая последовательность, которая "комплементарна" второй последовательности, в качестве альтернативы может называться как находящаяся в "антисмысловой" ориентации относительно второй последовательности.

Различные аминокислотные последовательности полипептидов и полинуклеотидные последовательности раскрыты в данном документе в качестве признаков некоторых вариантов осуществления раскрытого изобретения. Можно использовать варианты этих последовательностей, которые по меньшей мере на 70-85%, 85-90% или 90%-95% идентичны последовательностям, раскрытым в данном документе. В качестве альтернативы вариант аминокислотной последовательности или полинуклеотидной последовательности может быть по меньшей мере на 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% или 99% идентичен последовательности, раскрытой в данном документе. Вариант аминокислотной последовательности или полинуклеотидной последовательности характеризуется равносильной функцией/активностью с раскрытой последовательностью, или они составляют по меньшей мере приблизительно 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% или 99% функции/активности раскрытой последовательности.

Все аминокислотные остатки, раскрытые в данном документе в каждом аминокислотном положении белков Cas9 согласно данному документу, приведены в качестве примеров. Учитывая тот факт, что признаки структуры и/или заряда у некоторых аминокислот подобны друг другу (т. е. являются консервативными), аминокислота в каждом положении Cas9 может быть такой, как предусмотрено в раскрытых последовательностях, или замещена консервативным аминокислотным остатком ("консервативная аминокислотная замена") следующим образом.

1. Следующие небольшие алифатические, неполярные или слабополярные остатки могут замещать друг друга: Ala (A), Ser (S), Thr (T), Pro (P), Gly (G);

2. Следующие полярные, отрицательно заряженные остатки и их амиды могут замещать друг друга:

Asp (D), Asn (N), Glu (E), Gln (Q);

3. Следующие полярные, положительно заряженные остатки могут замещать друг друга: His (H), Arg (R), Lys (K);

4. Следующие алифатические, неполярные остатки могут замещать друг друга: Ala (A), Leu (L), Ile (I), Val (V), Cys (C), Met (M); и

5. Следующие крупные ароматические остатки могут замещать друг друга: Phe (F), Trp (Y), Trp (W).

В некоторых работах был достигнут прогресс в экспрессии белка и РНК-компонентов в клетках для осуществления нацеливания на ДНК, опосредуемого RGEN (например, предварительные заявки на патент США №№ 61/868706 и 62/036652). Такие стратегии, как правило, подразумевали экспрессию рекомбинантной ДНК в целевых клетках. Все еще представляют интерес дополнительные средства для введения белкового компонента и РНК-компонента в клетку, чтобы опосредовать нацеливание на ДНК, опосредуемое RGEN.

Варианты осуществления раскрытого изобретения относятся к композиции, содержащей по меньшей мере один белковый компонент направляемой РНК эндонуклеазы (RGEN) и по меньшей мере один проникающий в клетку пептид (CPP), где белковый компонент RGEN и CPP ковалентно или нековалентно связаны друг с другом в комплекс белок RGEN-CPP. Комплекс белок RGEN-CPP может пересекать (i) клеточную мембрану или (ii) клеточную стенку и клеточную мембрану клетки.

Необходимо отметить, что некоторые варианты осуществления раскрытого изобретения можно применять для доставки RGEN, уже ассоциированного (предварительно ассоциированного) с РНК-компонентом, в клетку. В таких вариантах осуществления может исключаться необходимость в доставке ДНК-конструкции в клетки для экспрессии РНК-компонента RGEN, тем самым устраняются любые потенциально нежелательные эффекты введения экзогенной ДНК в клетки. В то же время раскрытое изобретение является гибким, поскольку в некоторых других вариантах осуществления РНК-компонент может обеспечиваться (например, экспрессироваться) в клетке, в которую доставляется комплекс белок RGEN-CPP. РНК-компонент, обеспечиваемый таким образом, может вступать в ассоциацию с белковым компонентом RGEN после доставки/проникновения комплекса белок RGEN-CPP в клетку. Независимо от способа доставки РНК-компонента комплекс белок RGEN-CPP согласно данному документу способен вступать в ассоциацию с РНК-компонентом, образуя комплекс RGEN-CPP, который может целенаправленно воздействовать на специфическую последовательность ДНК в клетке. Таким образом, раскрытое изобретение обладает значительной гибкостью в обеспечении клеток RGEN для осуществления нацеливания на ДНК, опосредуемого RGEN.

Композиции, раскрытые в определенных вариантах осуществления, содержат по меньшей мере один белковый компонент RGEN. RGEN согласно данному документу называют комплекс, содержащий по меньшей мере один белок Cas и по меньшей мере один РНК-компонент. Таким образом, белковым компонентом RGEN может называться белок Cas, такой как Cas9. Примеры подходящих белков Cas включают в себя одну или несколько эндонуклеаз Cas из систем CRISPR типа I, II, или III (Bhaya et al., *Аппи. Rev. Genet.* 45:273-297, включен в данный документ посредством ссылки). Белок Cas из CRISPR типа I может представлять собой, например, белок Cas3 или Cas4. Белок Cas из CRISPR типа II может представлять собой, например, белок Cas9. Белок Cas из CRISPR типа III может представлять собой, например, белок Cas10. Белок Cas9 применяют в некоторых предпочтительных вариантах осуществления. В определенных вариантах осуществления белок Cas может представлять собой белок бактерий или архей. Белки Cas из CRISPR типов I-III согласно данному документу, как правило, происходят из прокариотических организмов; белки Cas типа I и III могут быть получены из видов бактерий и архей, в то время как белки Cas типа II (т. е. Cas9) могут быть получены, например, из видов бактерий. В других вариантах осуществления подходящие белки Cas включают в себя один или несколько из Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9, Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, их гомологов или их модифицированных вариантов.

В других аспектах раскрытого изобретения белок Cas согласно данному документу может происходить из любого из следующих родов: *Aeropyrum*, *Pyrobaculum*, *Sulfolobus*, *Archaeoglobus*, *Haloarcula*, *Methanobacterium*, *Methanococcus*, *Methanosarcina*, *Methanopyrus*, *Pyrococcus*, *Picrophilus*, *Thermioplasmia*, *Corynebacterium*, *Mycobacterium*, *Streptomyces*, *Aquifex*, *Porphyromonas*, *Chlorobium*, *Thermus*, *Bacillus*, *Listeria*, *Staphylococcus*, *Clostridium*, *Thermoanaerobacter*, *Mycoplasma*, *Fusobacterium*, *Azarcus*, *Chromobacterium*, *Neisseria*, *Nitrosomonas*, *Desulfovibrio*, *Geobacter*, *Myrococcus*, *Campylobacter*, *Wolinella*, *Acinetobacter*, *Erwinia*, *Escherichia*, *Legionella*, *Methylococcus*, *Pasteurella*, *Photobacterium*, *Salmonella*, *Xanthomonas*, *Yersinia*, *Streptococcus*, *Treponema*, *Francisella*, или *Thermotoga*. В качестве альтернативы белок Cas согласно данному документу может кодировать, например, любая из SEQ ID NO: 462-465, 467-472, 474-477, 479-487, 489-492, 494-497, 499-503, 505-508, 510-516 или 517-521, раскрытая в публикации заявки на патент США № 2010/0093617, которая включена в данный документ посредством ссылки.

Белковый компонент RGEN может содержать, например, аминокислотную последовательность

Cas9. RGEN, содержащая данный тип белкового компонента, как правило, может быть охарактеризована как имеющая Cas9 в качестве эндонуклеазного компонента RGEN. Аминокислотная последовательность белка Cas9 согласно данному документу, а также некоторых других белков Cas согласно данному документу, может быть получена, например, из видов *Streptococcus* (например, *S. pyogenes*, *S. pneumoniae*, *S. thermophilus*, *S. agalactiae*, *S. parasanguinis*, *S. oralis*, *S. salivarius*, *S. macacae*, *S. dysgalactiae*, *S. anginosus*, *S. constellatus*, *S. pseudoporcinus*, *S. mutans*), *Listeria* (например, *L. innocua*), *Spiroplasma* (например, *S. apis*, *S. syrrhodicola*), *Peptostreptococcaceae*, *Atopobium*, *Porphyromonas* (например, *P. catoniae*), *Prevotella* (например, *P. intermedia*), *Veillonella*, *Treponema* (например, *T. socranskii*, *T. denticola*), *Carnocytophaga*, *Finegoldia* (например, *F. magna*), *Coriobacteriaceae* (например, *C. bacterium*), *Olsenella* (например, *O. profusa*), *Haemophilus* (например, *H. sputorum*, *H. pittmaniae*), *Pasteurella* (например, *P. bettyae*), *Olivibacter* (например, *O. sitiensis*), *Epilithonimonas* (например, *E. tenax*), *Mesononia* (например, *M. mobilis*), *Lactobacillus*, *Bacillus* (например, *B. cereus*), *Aquimarina* (например, *A. muelleri*), *Chryseobacterium* (например, *C. palustre*), *Bacteroides* (например, *B. graminisolvens*), *Neisseria* (например, *N. meningitidis*), *Francisella* (например, *F. novicida*), или *Flavobacterium* (например, *F. frigidarium*, *F. soli*). В определенных аспектах данного документа предпочтительным является Cas9 из *S. pyogenes*. В качестве другого примера белок Cas9 может представлять собой любой из белков Cas9, раскрытых в Chylinski et al. (RNA Biology 10:726-737), который включен в данный документ посредством ссылки.

Соответственно, последовательность белка Cas9 согласно данному документу может включать, например, любую из аминокислотных последовательностей Cas9, раскрытых под следующими №№ доступа в GenBank G3ECR1 (*S. thermophilus*), WP_026709422, WP_027202655, WP_027318179, WP_027347504, WP_027376815, WP_027414302, WP_027821588, WP_027886314, WP_027963583, WP_028123848, WP_028298935, Q03JI6 (*S. thermophilus*), EGP66723, EGS38969, EGV05092, EHI65578 (*S. pseudoporcinus*), EIC75614 (*S. oralis*), EID22027 (*S. constellatus*), EIJ69711, EJP22331 (*S. oralis*), EJP26004 (*S. anginosus*), EJP30321, EPZ44001 (*S. pyogenes*), EPZ46028 (*S. pyogenes*), EQL78043 (*S. pyogenes*), EQL78548 (*S. pyogenes*), ERL10511, ERL12345, ERL19088 (*S. pyogenes*), ESA57807 (*S. pyogenes*), ESA59254 (*S. pyogenes*), ESU85303 (*S. pyogenes*), ETS96804, UC75522, EGR87316 (*S. dysgalactiae*), EGS33732, EGV01468 (*S. oralis*), EHI52063 (*S. macacae*), EID26207 (*S. oralis*), EID33364, EIG27013 (*S. parasanguinis*), EIJ37476, EJO19166 (*Streptococcus* sp. BS35b), EJU16049, EJU32481, YP_006298249, ERF61304, ERK04546, ETJ95568 (*S. agalactiae*), TS89875, ETS90967 (*Streptococcus* sp. SR4), ETS92439, EUB27844 (*Streptococcus* sp. BS21), AFJ08616, EUC82735 (*Streptococcus* sp. CM6), EWC92088, EWC94390, EJP25691, YP_008027038, YP_008868573, AGM26527, ANK22391, ANH36273, Q927P4, G3ECR1, или Q99ZW2 (*S. pyogenes*), которые включены посредством ссылки. Может применяться вариант любой из этих последовательностей белка Cas9, но он должен обладать активностью специфического связывания и необязательно активностью расщепления или внесения одонитевого разрыва в ДНК, если он ассоциирован с РНК-компонентом согласно данному документу. Такой вариант может содержать аминокислотную последовательность, которая по меньшей мере на приблизительно 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% или 99% идентична аминокислотной последовательности эталонного Cas9.

В качестве альтернативы белок Cas9 согласно данному документу может быть закодирован, например, в любой из SEQ ID NOs: 462 (*S. thermophilus*), 474 (*S. thermophilus*), 489 (*S. agalactiae*), 494 (*S. agalactiae*), 499 (*S. mutans*), 505 (*S. pyogenes*), или 518 (*S. pyogenes*), раскрытых в публикации заявки на патент США № 2010/0093617 (включена в данный документ посредством ссылки). В качестве альтернативы, однако, белок Cas9 согласно данному документу может содержать, например, аминокислотную последовательность из SEQ ID NO: 3 или остатки 1-1368, 2-1368, или 2-1379, of SEQ ID NO: 3. В качестве альтернативы, однако, белок Cas9 может содержать аминокислотную последовательность, которая, например, по меньшей мере на приблизительно 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% или 99% идентична любой из вышеизложенных аминокислотных последовательностей. Такой вариант белка Cas9 должен обладать активностью специфического связывания и необязательно активностью расщепления или внесения одонитевого разрыва в ДНК, если он ассоциирован с РНК-компонентом согласно данному документу.

Белок Cas, применяемый в данном документе (например, Cas9), может происходить из того же вида, из которого получают РНК-компонент (компоненты), или он может происходить из другого вида. Например, RGEN, содержащая белок Cas9, полученный из вида *Streptococcus* (например, *S. pyogenes* или *S. thermophilus*), может находиться в комплексе по меньшей мере с одним РНК-компонентом, имеющим последовательность (например, повторяющаяся последовательность crRNA, последовательность tracrRNA), полученную из того же вида *Streptococcus*. В качестве альтернативы белок Cas, применяемый в данном документе (например, Cas9), может происходить не из того вида, из которого получен(получены) РНК-компонент(компоненты) (белок Cas и РНК-компонент (компоненты) могут быть гетерологичными относительно друг друга); причем такие RGEN с гетерологичными Cas/РНК-компонентом должны характеризоваться активностью нацеливания на ДНК.

Определение активности связывания и/или эндонуклеолитической активности белка Cas согласно данному документу в отношении специфической целевой последовательности ДНК можно проводить с

помощью любого подходящего анализа, известного из уровня техники, как например, раскрытого в патенте США № 8697359, который раскрыт в данном документе посредством ссылки. Определение можно проводить, например, с помощью экспрессии белка Cas и подходящего РНК-компонента в клетке и затем проверки предсказанного целевого сайта ДНК на присутствие вставки/делеции (белок Cas в данном конкретном анализе, как правило, будет обладать полной эндонуклеолитической активностью [активностью расщепления двойной нити]). Проверку на наличие изменения/модификации (например, вставки/делеции) в предсказанном целевом сайте можно осуществлять, например, с помощью способа секвенирования ДНК или делая вывод об образовании изменения/модификации с помощью анализа в отношении потери функции у целевой последовательности. В другом примере активность белка Cas можно определить с помощью проведения экспрессии белка Cas и подходящего РНК-компонента в клетке, в которой обеспечили донорную ДНК, содержащую последовательность, гомологичную последовательности в целевом сайте или вблизи него. Присутствие последовательности донорной ДНК в целевом сайте (такое как можно предсказать по успешной HR между донорной и целевой последовательностями) будет указывать, что нацеливание произошло. В еще одном варианте осуществления активность белка Cas можно определить с применением *in vitro* анализа, в котором белок Cas и подходящий РНК-компонент смешивают вместе с полинуклеотидом ДНК, содержащим подходящую целевую последовательность. Данный анализ можно применять для обнаружения связывания (например, задержка электрофоретического сдвига в геле) у белков Cas, не обладающих активностью расщепления, или расщепления под действием белков Cas, которые обладают достаточной эндонуклеолитической активностью.

В определенных аспектах белок Cas согласно данному документу, такой как Cas9, может дополнительно содержать гетерологичную последовательность ядерной локализации (NLS). Гетерологичная аминокислотная последовательность NLS согласно данному документу может характеризоваться достаточной эффективностью, например, чтобы привести к накоплению белка Cas или комплекса белок Cas - CPP в обнаруживаемом количестве в ядре клетки согласно данному документу. NLS содержит одну (одинарный) или несколько (например, двойной) коротких последовательностей (например, имеющих от 2 до 20 остатков) из основных, положительно заряженных остатков (например, лизина и/или аргинина) и может быть расположен в любом месте аминокислотной последовательности Cas, но так, чтобы он располагался на поверхности белка. NLS может быть функционально связан, например, с N-концом или C-концом белка Cas согласно данному документу. С белком Cas могут быть связаны две или более последовательности NLS, например, такие как на обоих N- и C-концах белка Cas. Неограничивающие примеры последовательностей NLS, подходящих для применения в данном документе, включают в себя раскрытые в патентах США №№ 6660830 и 730957 6 (например, табл. 1 в них), оба из которых включены в данный документ посредством ссылки. Другой пример NLS, применимого в данном документе, включает аминокислотные остатки 1373-1379 из SEQ ID NO: 3. Белок Cas, раскрываемый в данном документе, например, может быть слит с CPP (пример белка Cas, ковалентно связанного с CPP). Будет понятно, что такой белок слияния Cas-CPP также может содержать NLS, описываемый выше. Также будет понятно, что в вариантах осуществления, в которых белок Cas слит с аминокислотной последовательностью, нацеливающейся на другую органеллу (например, митохондрию), такой белок Cas, как правило, не будет содержать NLS.

В определенных вариантах осуществления белок Cas и его соответственный РНК-компонент (например, crRNA), который управляет специфическим в отношении ДНК нацеливанием, осуществляемым белком Cas, может быть гетерологичным в отношении клетки, в частности, клетки, не являющейся прокариотической. Гетерологичная природа данных компонентов RGEN связана с тем, что белки Cas и их соответственные РНК-компоненты, как известно, встречаются только у прокариот (бактерий и архей).

В некоторых вариантах осуществления белок Cas представляет собой часть белка слияния, содержащего один или несколько гетерологичных белковых доменов (например, 1, 2, 3 или более доменов в дополнение к белку Cas). Эти варианты осуществления могут охватывать, например, белок Cas, который ковалентно связан с CPP и одной или несколькими дополнительными гетерологичными аминокислотными последовательностями. Другие варианты осуществления могут охватывать, например, белок Cas, который ковалентно связан с одной или несколькими дополнительными гетерологичными аминокислотными последовательностями, не включающими CPP (в таких вариантах осуществления CPP будет нековалентно связан с белком слияния на основе Cas). Белок слияния, содержащий белок Cas, может содержать любую дополнительную белковую последовательность и необязательно линкерную последовательность между любыми двумя доменами, как например, между Cas и первым гетерологичным доменом. Примеры белковых доменов, которые могут быть слиты с белком Cas согласно данному документу, включают в себя без ограничений эпитопные метки (например, гистидин [His, полигистидин], V5, FLAG, гемагглютинин вируса гриппа [HA], мус, VSV-G, тиоредоксин [Ttx]), репортеры (например, глутатион-S-трансферазу [GST], пероксидазу хрена [HRP], хлорамфеникол-ацетилтрансферазу [CAT], бета-галактозидазу, бета-глюкуронидазу [GUS], люциферазу, зеленый флуоресцентный белок [GFP], HcRed, DsRed, голубой флуоресцентный белок [CFP], желтый флуоресцентный белок [YFP], синий флуоресцентный белок [BFP]) и домены, обладающие одной или несколькими из следующих активностей: метилазная активность, деметилазная активность, активность активации транскрипции (например, VP16 или

VP64), активность подавления транскрипции, активность фактора терминации транскрипции, активность модификации гистонов, активность расщепления РНК и активность связывания нуклеиновых кислот. В других вариантах осуществления белок Cas может быть слит с белком, который связывает молекулы ДНК или другие молекулы, таким как белок связывания мальтозы (MBP), S-tag, ДНК-связывающий домен (DBD) Lex A, ДНК-связывающий домен GAL4A и VP16 вируса простого герпеса (HSV). Дополнительные домены, которые могут быть частью белка слияния, содержащего белок Cas согласно данному документу, раскрыты в публикации заявки на патент США № 2011/0059502, которая включена в данный документ посредством ссылки. В определенных вариантах осуществления, в которых белок Cas слит с гетерологичным белком (например, фактором транскрипции), белок Cas характеризуется активностью распознавания и связывания ДНК (находясь в комплексе с подходящим РНК-компонентом согласно данному документу), но не активностью внесения одностороннего разрыва или расщепления ДНК. Белок Cas, раскрываемый в данном документе, например, может быть слит с CPP (пример белка Cas, ковалентно связанного с CPP). Будет понятно, что в случае необходимости такой белок слияния Cas-CPP также может быть слит с одним или несколькими гетерологичными доменами, описываемыми выше.

Другие примеры гетерологичных доменов, которые могут быть связаны с белком Cas согласно данному документу, включают в себя аминокислотные последовательности, нацеливающие белок в конкретную органеллу (т. е. сигнал локализации). Примеры органелл, на которые происходит нацеливание, включают в себя митохондрии и хлоропласты. Как правило, такие домены нацеливания применяют вместо NLS при нацеливании на сайты внеядерной ДНК. Последовательность для нацеливания в митохондрию (MTS) может быть расположена, например, на N-конце белка Cas или вблизи него. Примеры MTS раскрыты в публикациях заявок на патент США №№ 2007/0011759 и 2014/0135275, которые включены в данный документ посредством ссылки. Последовательность для нацеливания в хлоропласт может быть такой, например, как раскрыто в публикациях заявок на патент США №№ 2010/0192262 или 2012/0042412, которые включены в данный документ посредством ссылки.

Белковый компонент RGEN может быть ассоциирован, например, по меньшей мере с одним РНК-компонентом (тем самым, составляя полную RGEN), который содержит последовательность, комплементарную последовательности целевого сайта на хромосоме или эписоме в клетке. В таких вариантах осуществления RGEN может связываться с последовательностью целевого сайта и необязательно расщеплять одну или обе нити ДНК в последовательности целевого сайта. Например, RGEN может расщеплять одну или обе нити целевой последовательности ДНК. В другом примере RGEN может расщеплять обе нити целевой последовательности ДНК. Будет понятно, что во всех этих вариантах осуществления белковый компонент RGEN может быть ковалентно или нековалентно связан по меньшей мере с одним CPP в комплексе белок RGEN-CPP. Ассоциация комплекса белок RGEN-CPP с РНК-компонентом согласно данному документу может быть охарактеризована как образование комплекса RGEN-CPP. Любое раскрытие в данном документе, относящееся к RGEN, аналогичным образом приложимо к компоненту RGEN комплекса RGEN-CPP, если не указано иное.

RGEN согласно данному документу, которая может расщеплять обе нити целевой последовательности ДНК, как правило, содержит белок Cas, у которого все его эндонуклеазные домены находятся в функциональном состоянии (например, представляют собой эндонуклеазные домены дикого типа или их варианты, сохраняющие некоторую степень активности или полную активность каждого эндонуклеазного домена). Таким образом, белок Cas дикого типа (например, белок Cas9, раскрытый в данном документе) или его вариант, сохраняющий некоторую степень активности или полную активность каждого эндонуклеазного домена белка Cas, представляют собой подходящий пример RGEN, которая может расщеплять обе нити целевой последовательности ДНК. Белок Cas9, содержащий функциональные нуклеазные домены RuvC и HNH, представляет собой пример белка Cas, который может расщеплять обе нити целевой последовательности ДНК. RGEN согласно данному документу, которая может расщеплять обе нити целевой последовательности ДНК, как правило, разрезает обе нити в одном и том же положении, так что в сайте разреза образуются тупые концы (т. е. отсутствуют выступающие концы из нуклеотидов).

RGEN согласно данному документу, которая может расщеплять одну нить целевой последовательности ДНК, может быть охарактеризована в данном документе как обладающая никазой активностью (например, способностью частичного расщепления). Никаза Cas (например, никаза Cas9) согласно данному документу, как правило, содержит один функциональный эндонуклеазный домен, который позволяет Cas расщепить только одну нить (т.е. сделать односторонний разрыв) целевой последовательности ДНК. Например, никаза Cas9 может содержать (i) мутантный, дисфункциональный домен RuvC и (ii) функциональный домен HNH (например, домен HNH дикого типа). В качестве другого примера никаза Cas9 может содержать (i) функциональный домен RuvC (например, домен RuvC дикого типа) и (ii) мутантный, дисфункциональный домен HNH.

Неограничивающие примеры никаз Cas9, подходящих для применения в данном документе, раскрыты в Gasiunas et al. (Proc. Natl. Acad. Sci. U.S.A. 109:E2579-E2586), Jinek et al. (Science 337:816-821), Sapranaukas et al. (Nucleic Acids Res. 39:9275-9282) и в публикации заявки на патент США № 2014/0189896, которые включены в данный документ посредством ссылки. Например, никаза Cas9 согласно данному документу может предусматривать Cas9 из *S. thermophilus* с заменой Asp-31 (например,

Asp-31-Ala) (пример мутантного домена RuvC) или заменой His-865 (например, His-865-Ala), заменой Asn-882 (например, Asn-882-Ala) или заменой Asn-891 (например, Asn-891-Ala) (примеры мутантных доменов HNH). Также, например, никаза Cas9 согласно данному документу может предусматривать Cas9 из *S. pyogenes* с заменой Asp-10 (например, Asp-10-Ala), заменой Glu-762 (например, Glu-762-Ala) или заменой Asp-986 (например, Asp-986-Ala) (примеры мутантного домена RuvC) или заменой His-840 (например, His-840-Ala), заменой Asn-854 (например, Asn-854-Ala) или заменой Asn-863 (например, Asn-863-Ala) (примеры мутантных доменов HNH). Что касается Cas9 из *S. pyogenes*, три субдомена RuvC обычно расположены в положении аминокислотных остатков 1-59, 718-769 и 909-1098 соответственно, а домен HNH расположен в положении аминокислотных остатков 775-908 (Nishimasu et al., Cell 156:935-949).

При необходимости никаза Cas9 согласно данному документу может применяться для различных целей в клетке. Например, никазу Cas9 можно применять для стимуляции HR в последовательности целевого сайта ДНК или вблизи него с подходящим донорным полинуклеотидом. Поскольку подвергнутая одонитевому разрыву ДНК не является субстратом для процессов NHEJ, но распознается процессами HR, внесение одонитевых разрывов в ДНК в специфическом целевом сайте должно сделать сайт более восприимчивым к HR с подходящим донорным полинуклеотидом.

В качестве другого примера можно применять пару никаз Cas9 для повышения специфичности нацеливания на ДНК. В целом, это можно осуществлять путем обеспечения двух никаз Cas9, которые за счет того, что они ассоциированы с РНК-компонентами с разными направляющими последовательностями, нацеливаются и образуют одонитевой разрыв в близкорасположенных последовательностях ДНК на противоположных нитях в участке требуемого нацеливания. Такое близкорасположенное расщепление каждой нити ДНК образует DSB (т. е. DSB с одонитевыми выступающими концами), который затем распознается как субстрат для NHEJ (приводит к образованию вставки/делеции) или HR (приводит к рекомбинации с подходящим донорным полинуклеотидом, если он предусмотрен). В данных вариантах осуществления каждый одонитевой разрыв может располагаться, например, на расстоянии по меньшей мере приблизительно 5, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90 или 100 (или любое целое число от 5 до 100) оснований друг от друга. Один или два белка никазы Cas9 согласно данному документу можно применять в паре никаз Cas9, как описано выше. Например, можно применять никазу Cas9 с мутантным доменом RuvC, но функционирующим доменом HNH (т. е. Cas9 HNH⁺/RuvC⁻) (например, Cas9 HNH⁺/RuvC⁻ из *S. pyogenes*). Каждая никаза Cas9 (например, Cas9 HNH⁺/RuvC⁻) будет нацелена на специфические сайты ДНК, расположенные близко друг от друга (на расстоянии до 100 пар оснований) с применением подходящих РНК-компонентов согласно данному документу с направляющими последовательностями РНК, нацеливающими каждую никазу на каждый специфический сайт ДНК.

В определенных вариантах осуществления RGEN может связываться с последовательностью целевого сайта ДНК, но не расщепляет ни одну нить в последовательности целевого сайта. Такая RGEN может содержать белок Cas, в котором все его нуклеазные домены являются мутантными, дисфункциональными. Например, в данном документе белок Cas9, который может связываться с последовательностью целевого сайта ДНК, но не расщепляет ни одну нить в последовательности целевого сайта, может содержать как мутантный, дисфункциональный домен RuvC, так и мутантный, дисфункциональный домен HNH. Неограничивающие примеры такого белка Cas9 содержат любую из мутаций нуклеазных доменов RuvC и HNH, раскрытые выше (например, Cas9 из *S. pyogenes* с заменой Asp-10, такой как Asp-10-Ala, и заменой His-840, такой как His-840-Ala). Белок Cas согласно данному документу, который связывает, но не расщепляет целевую последовательность ДНК, можно применять для модуляции экспрессии гена, например, в этом случае белок Cas можно сливать с фактором транскрипции (или его частью) (например, репрессором или активатором, таким как любой из раскрытых в данном документе). Например, Cas9, предусматривающий Cas9 из *S. pyogenes* с заменой Asp-10 (например, Asp-10-Ala) и заменой His-840 (например, His-840-Ala), может быть слит с доменами активатора транскрипции VP16 или VP64. Направляющая последовательность, используемая в РНК-компоненте такой RGEN, будет комплементарна последовательности ДНК, например, в промоторе или другом регуляторном элементе (например, интроне) гена.

RGEN согласно данному документу может связываться с последовательностью целевого сайта и необязательно расщеплять одну или обе нити последовательности целевого сайта в хромосоме, эписоме или любой другой молекуле ДНК в геноме клетки. Данное распознавание и связывание целевой последовательности является специфическим при условии, что РНК-компонент RGEN содержит последовательность (направляющую последовательность), которая комплементарна нити целевой последовательности. В определенных вариантах осуществления целевой сайт может быть уникальным (т.е. в данном геноме последовательность целевого сайта встречается один раз).

Длина целевой последовательности согласно данному документу может составлять, например, по меньшей мере 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, или 30 нуклеотидов; находится в диапазоне 13-30 нуклеотидов; в диапазоне 17-25 нуклеотидов или в диапазоне 17-20 нуклеотидов. Эта длина может быть с учетом или без учета последовательности РАМ. Также, нить целевой последовательности согласно данному документу характеризуется достаточной степенью комплементарности с на-

правляющей последовательностью (crRNA или gRNA), чтобы гибридизироваться с направляющей последовательностью и обеспечивать специфическое в отношении последовательности связывание белка Cas или комплекса белка Cas с целевой последовательностью (если подходящий PAM прилегает к целевой последовательности, см. ниже). Степень комплементарности между направляющей последовательностью и нитью соответствующей ей целевой последовательности ДНК составляет, например, по меньшей мере приблизительно 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, или 100%. Целевой сайт согласно данному документу может быть расположен, например, в последовательности, кодирующей продукт гена (например, белок или РНК), или некодирующей последовательности (например, регуляторной последовательности или "избыточной" последовательности).

Последовательность PAM (прилегающего к протоспейсеру мотива) может прилегать к последовательности целевого сайта. Последовательность PAM представляет собой короткую последовательность ДНК, распознаваемую RGEN согласно данному документу. Ассоциированный PAM и первые 11 нуклеотидов целевой последовательности ДНК, по-видимому, важны для нацеливания Cas9/gRNA и расщепления под их действием (Jiang et al., Nat. Biotech. 31:233-239). Длина последовательности PAM согласно данному документу может варьировать в зависимости от применяемого белка Cas или комплекса белка Cas, но, как правило, составляет, например, 2, 3, 4, 5, 6, 7 или 8 нуклеотидов в длину. Последовательность PAM расположена, например, непосредственно ниже или на 2 или 3 нуклеотида ниже последовательности целевого сайта, которая комплементарна нити в целевом сайте, которая в свою очередь комплементарна направляющей последовательности РНК-компонента. В вариантах осуществления данного документа, в которых RGEN представляет собой белок Cas9 с эндонуклеолитической активностью в комплексе с РНК-компонентом, Cas9 связывается с целевой последовательностью под управлением РНК-компонента и расщепляет обе нити непосредственно в направлении 5' от третьего нуклеотидного положения выше последовательности PAM. Рассмотрим следующий пример целевого сайта : последовательности PAM : 5'-NNNNNNNNNNNNNNNNNNNNNNNNXGG-3' (SEQ ID NO: 43) .

N может представлять собой A, C, T или G, а X может представлять собой A, C, T и G в данной последовательности, приведенной в качестве примера (X также может обозначаться N_{PAM}). Последовательность PAM в данном примере представляет собой XGG (подчеркнута). Подходящий комплекс Cas9/РНК-компонент расщепляет данную целевую последовательность непосредственно в направлении 5' от подчеркнутого двойной линией N. Отрезок из N в SEQ ID NO: 43 представляет собой целевую последовательность, которая по меньшей мере на приблизительно 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или 100% идентична, например, направляющей последовательности в РНК-компоненте согласно данному документу (при этом любые T в целевой последовательности ДНК следует выравнивать с любыми U из направляющей последовательности РНК). Направляющая последовательность РНК-компонента из комплекса Cas9 при распознавании и связывании на данной целевой последовательности (которая представляет целевые сайты согласно данному документу) будет подвергаться отжигу с комплементарной

последовательностью отрезка из N; при этом процент комплементарности между направляющей последовательностью и комплементарной нитью целевого сайта составляет, например, по меньшей мере приблизительно 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или 100%. Если для целенаправленного воздействия на SEQ ID NO: 43 в геноме применяют Cas9, то Cas9 будет вносить однонитевой разрыв непосредственно в направлении 5' от подчеркнутого двойной линией N или в таком же положении комплементарной нити, в зависимости от того, какой эндонуклеазный домен в Cas9 является дисфункциональным. Если для целенаправленного воздействия на SEQ ID NO: 43 в геноме применяют Cas9, не обладающую нуклеолитической активностью (оба домена RuvC и HNH являются дисфункциональными), она будет распознавать и связывать целевую последовательность, но не будет разрезать последовательность.

PAM согласно данному документу, как правило, выбирают исходя из типа используемой RGEN. Последовательность PAM согласно данному документу может представлять собой последовательность, распознаваемую RGEN, содержащей Cas, такую как Cas9, полученную, например, из любого из раскрытых в данном документе видов, из которых Cas может быть получена. В определенных вариантах осуществления последовательность PAM может представлять собой последовательность, распознаваемую RGEN, содержащей Cas9, полученную из *S. pyogenes*, *S. thermophilus*, *S. agalactiae*, *N. meningitidis*, *T. denticola*, или *F. novicida*. Например, подходящую Cas9, полученную из *S. pyogenes*, можно было бы применять для нацеливания на геномные последовательности с последовательностью PAM NGG (SEQ ID NO: 44; N может представлять собой A, C, T или G). В качестве других примеров подходящая Cas9 может быть получена от любого из следующих видов, когда подвергаемые нацеливанию последовательности ДНК имеют следующие последовательности PAM: *S. thermophilus* (NNAGAA [SEQ ID NO: 45]), *S. agalactiae* (NGG [SEQ ID NO: 44]), NNAGAAW [SEQ ID NO: 46, W представляет собой A или T], NGGNG [SEQ ID NO: 47]), *N. meningitidis* (NNNGATT [SEQ ID NO: 48]), *T. denticola* (NAAAAC [SEQ ID NO: 49]) или *F. novicida* (NG [SEQ ID NO: 50]) (где N во всех этих конкретных последовательностях PAM представляют собой A, C, T или G). Другие примеры Cas9/PAM, применимого в данном документе,

включают в себя раскрытые в Shah et al. (RNA Biology 10:891-899) и Esvelt et al. (Nature Methods 10:1116-1121), которые включены в данный документ посредством ссылки. Примеры целевых последовательностей согласно данному документу соответствуют SEQ ID NO: 43, но при этом PAM 'XGG' замещается любым из вышеописанных PAM.

РНК-компонент согласно данному документу может содержать последовательность, комплементарную последовательности целевого сайта в хромосоме или эписоме в клетке. RGEN может специфически связываться с последовательностью целевого сайта и необязательно расщеплять одну или обе нити последовательности целевого сайта, на основании данной комплементарности последовательностей. Таким образом, в определенных вариантах осуществления раскрытого изобретения комплементарная последовательность РНК-компонента также может называться направляющей последовательностью или вариабельным нацеливающим доменом.

Длина направляющей последовательности РНК-компонента (например, crRNA или gRNA) согласно данному документу может составлять по меньшей мере 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 или 30 рибонуклеотидов; например, длина находится в диапазоне 13-30 рибонуклеотидов; длина находится в диапазоне 17-25 рибонуклеотидов или длина находится в диапазоне 17-20 рибонуклеотидов. В целом, направляющая последовательность согласно данному документу обладает достаточной степенью комплементарности с нитью целевой последовательности ДНК, чтобы гибридизироваться с целевой последовательностью и обеспечивать специфическое в отношении последовательности связывание белка Cas или комплекса белка Cas с целевой последовательностью (если подходящий PAM прилегает к целевой последовательности). Степень комплементарности между направляющей последовательностью и соответствующей ей целевой последовательности ДНК составляет, например, по меньшей мере приблизительно 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или 100%. Направляющая последовательность может быть сконструирована соответствующим образом, чтобы нацеливать RGEN на целевую последовательность ДНК в клетке.

РНК-компонент согласно данному документу может предусматривать crRNA, которая, например, содержит направляющую последовательность и повторяющуюся (парную для tracrRNA) последовательность. Направляющая последовательность, как

правило, расположена на 5'-конце crRNA или вблизи него (в пределах 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 или более оснований). Ниже направляющей последовательности crRNA находится "повторяющаяся" или "парная для tracrRNA" последовательность, которая комплементарна последовательности на 5'-конце tracrRNA и может гибридизироваться с ней. Направляющая последовательность и парная последовательность для tracrRNA могут непосредственно прилегать или быть разделены, например, 1, 2, 3, 4 или более основаниями. Парная последовательность для tracrRNA, например, по меньшей мере на 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% или 100% комплементарна 5'-концу tracrRNA. В целом, степень комплементарности может указываться с учетом оптимального выравнивания парной последовательности для tracrRNA и 5'-конца последовательности tracrRNA по длине более короткой из двух последовательностей. Длина парной последовательности для tracrRNA согласно данному документу может составлять, например, по меньшей мере 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, или 18 рибонуклеотидов в длину, и она гибридизуется с последовательностью такой же или аналогичной длины (например, плюс-минус 1, 2, 3, 4 или 5 оснований) на 5'-конце tracrRNA. Подходящие примеры парных последовательностей для tracrRNA согласно данному документу включают в себя SEQ ID NO: 51 (guuuuuguasucusaaguuua), SEQ ID NO: 52 (guuuuuguasucusa), SEQ ID NO: 53 (guuuuagagcua) или SEQ ID NO: 54 (guuuuagagcuag) или их варианты, которые (i) имеют последовательность, по меньшей мере на приблизительно 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% или 99% идентичную последовательности 5'-конца tracrRNA и (ii) могут подвергаться отжигу с ней. Длина crRNA согласно данному документу может составлять, например, по меньшей мере приблизительно 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, или 48 рибонуклеотидов; или приблизительно 18-48 рибонуклеотидов; или приблизительно 25-50 рибонуклеотидов.

В вариантах осуществления, в которых в состав RGEN входит белок Cas9 из системы CRISPR типа II, tracrRNA может включаться вместе с crRNA. tracrRNA согласно данному документу содержит, в направлении от 5' к 3', (i) последовательность, которая подвергается отжигу с повторяющимся участком (парная последовательность для tracrRNA) crRNA и (ii) часть, содержащую структуру "стебель-петля". Длина последовательности (i) может быть, например, такой же или аналогичной (например, плюс-минус 1, 2, 3, 4 или 5 оснований) длине любой из парных последовательностей для tracrRNA, раскрытых выше. Общая длина tracrRNA согласно данному документу (т. е. компонентов последовательности [i] и [ii]) может составлять, например, по меньшей мере приблизительно 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, или 90 (или любое целое число от 30 до 90) рибонуклеотидов. tracrRNA может дополнительно включать 1, 2, 3, 4, 5 или более остатков урацила на 3'-конце, которые могут присутствовать вследствие экспрессии tracrRNA с последовательностью терминатора транскрипции.

tracrRNA согласно данному документу может быть получена, например, из любого из перечисленных выше видов бактерий, из которого может быть получена последовательность Cas9. Примеры подходящих последовательностей tracrRNA включают в себя раскрытые патенте США № 8697359 и Chylinski et al. (RNA Biology 10:726-737),, которые включены в данный документ посредством ссылки. Предпочти-

тельная tracrRNA согласно данному документу может быть получена из tracrRNA видов *Streptococcus* (например, *S. pyogenes*, *S. thermophilus*). Другие подходящие примеры tracrRNA согласно данному документу могут включать:

SEQ ID NO:55:
uaagsaaguuuuuuuaaggcuagucscguuaucsaacuugaaaaaguggcaccgagucggu

с,

SEQ ID NO:56:
uaagsaaguuuuuuuaaggcuagucscguuaucsaacuugaaaaagug, или

SEQ ID NO:57:
uaagsaaguuuuuuuaaggcuagucscguuaucsa,

которые получены из tracrRNA *S. pyogenes*. Другие подходящие примеры tracrRNA согласно данному документу могут включать:

SEQ ID NO:58:
uaaaucsuugcagaagcuasacaagaauaaggcuucaugcscgaaaucaacacccugucauuu
uauggcaggguguuuuucguuuuuuaa,

SEQ ID NO:59:
ucsaagaagcuasacaagaauaaggcuucaugcscgaaaucaacacccugucauuuuuauggsa
ggguguuuuucguuuuuua, или

SEQ ID NO:60:
ucsaagaagcuasacaagaauaaggcuucaugcscgaaaucaacacccugucauuuuuauggsa
gggugu,

которые получены из tracrRNA *S. thermophilus*.

Еще одними примерами tracrRNA согласно данному документу являются варианты tracrRNA под вышеуказанными SEQ ID NO, которые (i) имеют последовательность, по меньшей мере на приблизительно 80%, 85%, 90, 91, 92, 93, 94, 95, 96, 97, 98 или 99% идентичную им, и (ii) могут функционировать в качестве tracrRNA (например, последовательность 5'-конца может подвергаться отжигу с парной последовательностью для tracrRNA из crRNA, последовательность ниже последовательности 5'-конца может образовывать одну или несколько шпилек, вариант tracrRNA может образовывать комплекс с белком Cas9).

РНК-компонент RGEN, раскрытой в данном документе (или, иными словами, РНК-компонент, который может ассоциировать с белковым компонентом RGEN) может содержать, например, направляющую РНК (gRNA), содержащую crRNA, функционально связанную или слитую с tracrRNA. В некоторых предпочтительных вариантах осуществления crRNA-компонент gRNA расположен выше tracrRNA-компонента (т. е. такая gRNA содержит, в направлении от 5' к 3', crRNA, функционально связанную с tracrRNA). Например, в состав gRNA может входить любая crRNA и/или tracrRNA (и/или их часть, такая как повторяющаяся последовательность crRNA, парная последовательность для tracrRNA или последовательность 5'-конца tracrRNA), раскрытые в данном документе (например, в вышеуказанных вариантах осуществления).

Парная последовательность для tracrRNA crRNA-компонента gRNA согласно данному документу должна быть способна подвергаться отжигу с 5'-концом tracrRNA-компонента, образуя тем самым шпилечную структуру. Например, любое из приведенных выше раскрытий, касающееся длины и процента комплементарности между парными последовательностями для tracrRNA (из crRNA-компонента) и последовательностями 5'-конца (из tracrRNA-компонента) может характеризовать crRNA- и tracrRNA-компоненты gRNA. Для облегчения данного отжига функциональная связь или слияние crRNA- и tracrRNA-компонентов предпочтительно содержит подходящую рибонуклеотидную последовательность, образующую петли (т.е. последовательность, образующая петли, может связывать вместе crRNA- и tracrRNA-компоненты с образованием gRNA). Подходящие примеры последовательностей РНК, образующих петли, включают в себя GAAA (SEQ ID NO: 36), CAAA (SEQ ID NO: 37) и AAAG (SEQ ID NO: 38). Однако можно использовать более длинные или более короткие последовательности петли, а также можно использовать альтернативные последовательности петли.

Последовательность петли предпочтительно содержит рибонуклеотидный триплет (например, AAA) и дополнительный рибонуклеотид (например, C или G) на каждом конце триплета.

gRNA согласно данному документу образует шпильку ("первую шпильку") при отжиге ее частей, парной последовательности для tracrRNA (из crRNA-компонента) с последовательностью 5'-конца tracrRNA. Одна или несколько (например, 1, 2, 3 или 4) дополнительных шпилечных структур могут образовываться ниже данной первой шпилеки, в зависимости от последовательности tracrRNA-компонента gRNA. Следовательно, gRNA может содержать, например, до пяти шпилечных структур. gRNA может дополнительно содержать 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25,

РНК-компонент RGEN необязательно не имеет 5'-кэпа (7-метилгуанилатного [m^7G] кэпа) (т. е. такой РНК-компонент не имеет m^7G -кэпа на своем 5'-конце). РНК-компонент согласно данному документу может иметь, например, 5'-гидроксильную группу вместо 5'-кэпа. В качестве альтернативы РНК-компонент согласно данному документу может иметь, например, 5'-фосфат вместо 5'-кэпа. Полагают, что РНК-компонент в таких вариантах осуществления может лучше накапливаться в ядре (а именно, после его транскрипции в ядре или после его опосредуемого RGEN импорта в ядро, в зависимости от того, какой РНК-компонент обеспечен в данном документе), поскольку 5'-кэпированная РНК (т. е. РНК с 5' m^7G -кэпом) подвергается экспорту из ядра. Предпочтительные примеры некэпированных РНК-компонентов согласно данному документу включают в себя подходящие gRNA, crRNA и/или tracrRNA. В определенных вариантах осуществления у РНК-компонента согласно данному документу отсутствует 5'-кэп, и необязательно вместо него имеется 5'-гидроксильная группа вследствие автопроцессинга РНК под действием рибозимной последовательности на 5'-конце предшественника РНК-компонента (т.е. РНК-предшественника, содержащего рибозимную последовательность выше РНК-компонента, так что gRNA подвергается опосредуемому рибозимом автопроцессингу с удалением рибозимной последовательности, оставляя тем самым нижерасположенный РНК-компонент без 5'-кэпа). В некоторых других вариантах осуществления РНК-компонент согласно данному документу не образуется путем транскрипции с промотора РНК-полимеразы III (Pol III).

Длина проникающего в клетку пептида (CPP) согласно данному документу может составлять, например, приблизительно 5-30, 5-25, 5-20, 10-30, 10-25 или 10-20 аминокислотных остатков. В качестве других примеров длина CPP может составлять приблизительно 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, или 30 аминокислотных остатков. При этом в дополнительных аспектах данного документа длина CPP может составлять до приблизительно 35, 40, 45, 50, 55, или 60 аминокислотных остатков.

CPP, раскрытый в данном документе, может быть, например, катионным или амфипатическим. Катионный CPP согласно данному документу, как правило, содержит по меньшей мере приблизительно 60% положительно заряженных аминокислот, таких как лизин (K), аргинин (R) и/или гистидин (H). В качестве альтернативы катионный CPP может содержать, например, по меньшей мере 60%, 65%, 70%, 75%, 80%, 85%, 90% или 95% положительно заряженных аминокислот (например, остатков R; остатков K; остатков K и R; остатков K, R и H). В определенных вариантах осуществления катионный CPP может быть охарактеризован как богатый аргинином (например, содержит по меньшей мере 70% или 80% остатков R) или богатый лизином (например, содержит по меньшей мере 70% или 80% остатков L). Примеры катионных CPP, применимых в данном документе, раскрыты в Schmidt et al. (FEBS Lett. 584:1806-1813) и Wender et al. (полилизиновые; Proc. Natl. Acad. Sci. USA 97:13003-13008), которые включены в данный документ посредством ссылки. Другие примеры катионных CPP включают в себя GRKKRRQRRR (SEQ ID NO: 68), RKKRRQRRR (SEQ ID NO: 69) или RKKRRQRR (SEQ ID NO: 70), которые были изначально получены из Tat-белка HIV, и пенетратин (RQIKIWFQNRRMKWKK, SEQ ID NO: 71), который был изначально получен из белка гомеодомена Antennapedia Drosophila.

Другой пример катионного CPP содержит полиаргининовую последовательность, имеющую определенное количество смежных аргининов, достаточное для прямого проникновения CPP и его груза (например, белкового компонента RGEN или RGEN) в клетку. Определенное количество смежных аргининовых остатков в такой полиаргининовой последовательности может составлять, например, по меньшей мере 3, 4, 5, 6, 7, 8, 9, 10 или 10-50 аргининов. В определенных аспектах данного документа CPP может иметь 6 или более смежных аргининовых остатков (например, 6-7, 6-8, 6-9 или 6-10 аргининовых остатков). При необходимости в состав полиаргининового CPP может входить "PolyR" (GGG-GRRRRRRRRLLLLL, SEQ ID NO: 15). Другие примеры полиаргининовых CPP содержат THRLPRRRRRR (SEQ ID NO: 72) или GGRRARRRRRR (SEQ ID NO: 73). В некоторых вариантах осуществления CPP представляет собой активируемый CPP ("ACPP") (Aguilera et al., Integr Biol. (Camb) 1:371-381; включенный в данный документ посредством ссылки). ACPP, как правило, содержат поликатионный CPP (например, из девяти смежных аргининов), соединенный посредством расщепляемого линкера с сопоставимым полианионом (например, из девяти смежных глутаминов), что снижает суммарный заряд до практически нуля и тем самым подавляет адгезию и захват CPP в клетки. После расщепления линкера полианион высвобождается, приводя к демаскировке поликатионной части и ее изначальной адгезивности на местном уровне, тем самым обеспечивая возможность проникновения CPP в клетку. Другим примером в данном документе является полилизиновый CPP; при этом примерами полилизиновых CPP согласно данному документу являются любые из вышеизложенных вариантов осуществления полиаргининовых пептидов, в которых R заменен на K.

Амфипатический CPP согласно данному документу содержит аминокислотную последовательность, содержащую перемежающийся паттерн из полярных/заряженных остатков и неполярных, гидрофобных остатков. Считается, что следующие CPP являются амфипатическими и применимы в определенных аспектах (независимо от того, применим ли к ним в полной мере термин "амфипатические"): CPP, содержащий пептид транспоран-10 (TP10) (например, AGYLLGKINLKACAACAKKIL, SEQ ID NO: 14); CPP из белка кадгерина сосудистого эндотелия, такой как CPP, содержащий pVEC-пептид (на-

пример, LILRRRIRKQАНАНСК, SEQ ID NO: 74; LILRRRIRKQАНАНСК, SEQ ID NO: 13); CPP из трансактиваторного белка Zebra вируса Эпштейна-Барр, такой как CPP, содержащий пептид Zebra (например, ECDSELEIKRYKRVVAVSRKCRACKFKQLLQHYREVAААКСSENDRLRLLKQMC, SEQ ID NO: 12); CPP, содержащий (KFF)₃K-пептид (например, KFFKFFKFFK, SEQ ID NO: 75); CPP, содержащий MAP-пептид (KLALKLALKALKALKLA, SEQ ID NO: 76); CPP, содержащий RRQRRTSKLMKR (SEQ ID NO: 77); CPP, содержащий KALAWEAKLAKALAKALAKHLAKALAKALKCEA (SEQ ID NO: 78). Другие амфипатические CPP, подходящие для данного документа, включают в себя богатые пролином CPP, такие как содержащие по меньшей мере 3, 4, 5, 6, 7 или 8 повторов VHLPPP (SEQ ID NO: 79) или VRLPPP (SEQ ID NO: 80).

В качестве других примеров CPP согласно данному документу может содержать MPG-пептид (например, GALFLGFLGAAGSTMGAWSQPKSKRKV, SEQ ID NO: 81); Рер-1-пептид (например, KETWWETWWTEWSQPKKRRKV, SEQ ID NO: 82) или CPP из белка кальцитонина человека, такой как hCT-пептид (например, LGTYTQDFNKFHTFPQTAIGVGAP, SEQ ID NO: 83; CGNLSTCMLGTY-TQDFNK, SEQ ID NO: 84). Еще одни примеры CPP согласно данному документу включают в себя раскрытые в Regberg et al. (Int. J. Pharm. 464:111-116), который включен в данный документ посредством ссылки.

В качестве альтернативы CPP, подходящий для данного документа, может содержать аминокислотную последовательность, которая, например, по меньшей мере на 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% или 99% идентична любой из аминокислотных последовательностей CPP, раскрытых в данном документе. Такой вариант белка CPP должен обладать активностью CPP, такой как способность опосредовать захват клеткой молекулярного груза (например, аминокислотной последовательности, содержащей один или несколько белковых компонентов RGEN [например, Cas9], или аминокислотной последовательности, содержащей один или несколько белковых компонентов RGEN [например, Cas9], ассоциированных с РНК-компонентом). Тестирование активности варианта CPP можно проводить с помощью целого ряда способов, как например, путем ковалентного связывания его с флуоресцентным белком (например, GFP) и измерения уровня флуоресценции, испускаемой клеткой, которую привели в контакт с комплексом CPP-флуоресцентный белок.

При необходимости CPP согласно данному документу можно модифицировать, чтобы сделать его даже более эффективным в перенесении белкового груза RGEN снаружи клетки внутрь клетки. Например, можно модифицировать CPP так, чтобы он имел липидную группу на одном из его N- или C-конца. Подходящие липидные группы согласно данному документу включают в себя ацильные группы, такие как стеарильная и миристильная группы. Другие примеры липидных групп представляют собой ацильные группы с 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, или 18 атомами углерода. Условия модификации пептидов с помощью липидных групп, применимых в данном документе, раскрыты, например, в Regberg et al. (Int. J. Pharm. 464:111-116) и Anko et al. (Biochim. Biophys. Acta - Biomembranes 1818:915-924), которые включены в данный документ посредством ссылки.

В определенных аспектах данного документа белковый компонент RGEN и по меньшей мере один CPP в данном документе могут быть ковалентно связаны друг с другом в комплекс белок RGEN-CPP. Например, белковый компонент RGEN и по меньшей мере один CPP могут быть слиты вместе в одну аминокислотную последовательность (т. е. белковый компонент RGEN и по меньшей мере один CPP могут входить в состав белка слияния). Таким образом, примером ковалентной связи согласно данному документу может быть пептидная связь, с помощью которой аминокислотная последовательность белкового компонента RGEN слита с аминокислотной последовательностью CPP, так что обе эти аминокислотные последовательности содержатся в одной аминокислотной последовательности. Такой белок слияния (или "химерный белок") может быть охарактеризован как слияние белок RGEN-CPP согласно данному документу. В тех вариантах осуществления, в которых РНК-компонент ассоциирован с белковым компонентом RGEN, такой белок слияния может быть охарактеризован как слияние RGEN-CPP.

Один или несколько CPP могут располагаться, например, на N-конце или C-конце слияния белок RGEN-CPP. В качестве альтернативы один или несколько CPP могут располагаться на обоих N- и C-концах слияния белок RGEN-CPP. В качестве альтернативы, однако, один или несколько CPP могут располагаться в пределах аминокислотной последовательности слияния белок RGEN-CPP. Варианты осуществления данного документа, предусматривающие более одного CPP, могут предусматривать по меньшей мере 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 или 20 CPP, или 5-10, 5-20, или 10-20 CPP. CPP, слитые с белковым компонентом RGEN, могут быть одинаковыми или разными (например, 2, 3, 4 или более разных типов CPP). Один или несколько CPP могут быть слиты непосредственно с аминокислотной последовательностью белка RGEN и/или могут быть слиты с гетерологичным(гетерологичными) доменом(доменами) (например, NLS или последовательностями для нацеливания в другие органеллы, такими как MTS), который слит с белком RGEN.

Слияние между CPP и белковым компонентом RGEN согласно данному документу может быть непосредственным (т. е. аминокислотная последовательность CPP непосредственно связана с аминокислотной последовательностью RGEN с помощью пептидной связи). В качестве альтернативы слияние между CPP и белковым компонентом RGEN может осуществляться с помощью промежуточной аминокис-

лотной последовательности (это пример опосредованной связи между CPP и белковым компонентом RGEN). Примеры промежуточной аминокислотной последовательности включают в себя подходящие линкерные последовательности, содержащие по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90 или 100 аминокислотных остатков, таких как глицин, серин, аланин и/или пролин. Подходящие аминокислотные линкеры раскрыты, например, в патентах США №№ 8828690, 8580922 и 5990275, например, которые включены в данный документ посредством ссылки. Другие примеры промежуточных аминокислотных последовательностей могут содержать один или несколько других типов белков и/или доменов. Например, в состав промежуточной аминокислотной последовательности может входить маркерный белок (например, флуоресцентный белок, такой как раскрытые в данном документе).

Композицию, содержащую комплекс из связанных ковалентной связью белкового компонента RGEN и по меньшей мере одного CPP, как например, в виде белка слияния, можно применять в отношении любого типа клеток, раскрытого в данном документе. Необязательно, однако, данную композицию можно применять в отношении клеток организма, не относящегося к млекопитающим, такого как дрожжи, грибы и растения, но это исключает применение в отношении клеток млекопитающих.

Примеры белков слияния белок RGEN-CPP согласно данному документу могут включать SEQ ID NO: 39 (белок слияния Zebra CPP-Cas9-NLS), 40 (белок слияния PolyR CPP-Cas9-NLS), 41 (белок слияния TP10 CPP-Cas9-NLS) или 42 (белок слияния pVEC CPP-Cas9-NLS). SEQ ID NO: 39-42 представляют собой примеры белков слияния Cas9-CPP. Другие примеры белков слияния белок RGEN-CPP содержат аминокислотную последовательность, которая по меньшей мере на приблизительно 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% или 99% идентична любой из SEQ ID NO: 39-42. Такой вариант белка слияния должен иметь (i) CPP-домен, который может опосредовать захват клеткой белка слияния, и (ii) белок Cas9 со специфической активностью связывания и необязательно активностью расщепления или внесения однонитевого разрыва в ДНК, если он ассоциирован с РНК-компонентом. SEQ ID NO: 39, 40, 41 и 42 содержат Zebra CPP (SEQ ID NO: 12), PolyR CPP (SEQ ID NO: 15), TP10 CPP (SEQ ID NO: 14) и pVEC CPP (SEQ ID NO: 13) соответственно, функционально связанные с белком Cas9 (*S. pyogenes*)-NLS (остатки 2-1379 из SEQ ID NO: 3).

В определенных вариантах осуществления белковый компонент системы направляющий полинуклеотид/эндонуклеаза Cas может быть слит с CPP, где CPP включает в себя:

- (i) CPP из трансактиваторного белка Zebra вируса Эпштейна-Барр,
- (ii) CPP с 6 или более смежными аргининовыми остатками,
- (iii) CPP транспортан-10 (TP10),
- (iv) CPP из белка кадгерина сосудистого эндотелия или
- (v) CPP, выбранный из группы, состоящей из синтетического CPP с девятью аргининовыми остатками, богатого гистидином CPP с девятью аргининовыми остатками и Pas CPP с девятью аргининовыми остатками.

Примеры синтетического CPP с девятью аргининовыми остатками, богатого гистидином CPP с девятью аргининовыми остатками и Pas CPP с девятью аргининовыми остатками раскрыты, например, в Liu et al. (*Advanced Studies in Biology* 5(2):71-88, HIKARI Ltd), который включен в данный документ посредством ссылки.

Другим примером того, как могут быть ковалентно связаны белковый компонент RGEN и по меньшей мере один CPP, является перекрестная сшивка (химическая перекрестная сшивка). Таким образом, пример комплекса белок RGEN-CPP согласно данному документу может содержать белок RGEN, перекрестно сшитый по меньшей мере с одним CPP. Перекрестной сшивкой согласно данному документу называют процесс химического соединения двух или более молекул (в данном случае, белкового компонента RGEN и по меньшей мере одного CPP) с помощью ковалентной(ковалентных) связи(связей). Перекрестную сшивку можно осуществлять с применением целого ряда способов, известных из уровня техники, таких как раскрытые в публикации заявки на патент США № 2011/0190813, патенте США № 8642744 и *Bioconjugate Techniques*, 2nd Edition (G.T. Hermanson, Academic Press, 2008), все из которых включены в данный документ посредством ссылки.

Как правило, CPP можно модифицировать и/или синтезировать так, чтобы он содержал подходящую группу для связывания белка на своем N-конце, C-конце и/или боковой группе аминокислоты для целей перекрестной сшивки CPP с белковым компонентом RGEN. "Группой для связывания белка" называют группу, которая непосредственно способна вступать в реакцию, либо самопроизвольно, либо после активации (например, светом), с доступной функциональной группой боковой цепи на белковом компоненте RGEN при подходящих условиях (например, в водной среде) с образованием ковалентной связи между CPP и белком RGEN. Группа для связывания белка может реагировать, например, с функциональными группами боковой цепи аминокислотного остатка Lys, Cys, Ser, Thr, Tyr, His или Arg в белке RGEN с образованием ковалентной связи с белком. Можно применять, например, либо гомобифункциональную (например, способную связывать амин с амином), либо гетеробифункциональную (например, способную связывать амин с тиолом) группу для связывания белка. В определенных вариантах осуществления группа для связывания белка на CPP может также вступать в реакцию с терминальной

группой (например, N-концом) белка RGEN. Подходящие группы для связывания белка согласно данному документу включают в себя группы, реагирующие с аминогруппой (например, сложный NHS-эфир или сложный имидоэфир), реагирующие с тиольной (сульфгидрильной) группой (например, малеимид, такой как BMOE, BMB или BMH), реагирующие с гидроксильной группой, реагирующие с имидазолилом или реагирующие с гуанидинилом. Иллюстративные группы для связывания белка включают в себя активные сложные эфиры (например, реагирующие с аминогруппой сложные NHS-эфиры) и реагирующие с тиолом малеимидные или йодацетамидные группы. Дополнительные иллюстративные группы для связывания белка, применимые в данном документе, и способы их применения описаны, например, в *Bioconjugate Techniques*, 2nd Edition (G.T. Hermanson, Academic Press, 2008).

Группа для связывания белка согласно данному документу, как правило, может приводить к образованию соединения между CPP и белком RGEN, имеющего длину остова 20 атомов или менее. Например, длина такого соединения может составлять от 1 до 20 атомов или приблизительно 1, 2, 3, 4, 5, 6, 8, 10, 12, 14, 16, 18 или 20 атомов углерода. В определенных вариантах осуществления соединения может быть линейным, разветвленным, циклическим или представлять собой один атом. В определенных случаях один, два, три, четыре или пять или более атомов углерода в остове линкера могут быть замещены гетероатомом серы, азота или кислорода. Связи между атомами остова могут быть насыщенными или ненасыщенными (обычно не более одной, двух или трех ненасыщенных связей в остове линкера). Линкер может включать без ограничений олиго(этиленгликольную); эфирную; тиоэфирную группу; группу третичного амина или алкильную группу, которые могут быть прямыми или разветвленными (например, метильную, этильную, n-пропильную, изо-пропильную, n-бутильную, n-пентильную, трет-бутильную). В качестве других примеров остов линкера может включать циклическую группу, такую как арильная, гетероциклическая или циклоалкильная группа, в которой 2 или более атомов (например, 2, 3 или 4 атома) циклической группы включены в остов.

В определенных вариантах осуществления с белковым компонентом RGEN может быть перекрестно сшито более одного типа CPP (например, 2, 3, 4 или более разных типов CPP). Соотношение (молярное соотношение) CPP и белка RGEN, которое может применяться при перекрестной сшивке, может составлять, например, по меньшей мере приблизительно 1:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1, 15:1, 20:1, 30:1, 40:1, или 50:1. В других аспектах среднее число CPP, перекрестно сшитых с белком RGEN, может составлять по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, или 25, или по меньшей мере 5-10, 5-15, 5-20, или 5-25.

При необходимости белковый компонент RGEN и по меньшей мере один CPP можно перекрестно сшивать в комплекс, дополнительно содержащий один или несколько других белков/пептидов/доменов. Такие другие элементы необязательно можно применять для создания мостиковой связи между белковым компонентом RGEN и CPP, и они могут включать любую из промежуточных аминокислотных последовательностей, описанных выше.

В некоторых аспектах данного документа белковый компонент RGEN и по меньшей мере один CPP в данном документе могут быть нековалентно связаны друг с другом в комплекс белок RGEN-CPP. Без желаниа придерживаться какой-либо конкретной теории или механизма подразумевается, что нековалентная связь между белковым компонентом RGEN и по меньшей мере одним CPP может быть обусловлена электростатическими, Ван-дер-Ваальсовыми и/или гидрофобными силами. В тех вариантах осуществления вариант осуществления, в которых РНК-компонент ассоциирован с белковым компонентом RGEN, такие варианты осуществления могут быть охарактеризованы как содержащие RGEN, которая нековалентно связана по меньшей мере с одним CPP в комплекс RGEN-CPP. Композиция, содержащая белковый компонент RGEN и CPP, которые связаны нековалентно, необязательно может быть охарактеризована как смесь этих компонентов.

В определенных вариантах осуществления белковый компонент RGEN нековалентно связан по меньшей мере с одним CPP с аминокислотной последовательностью, состоящей только из аминокислотной последовательности CPP. Такой CPP, хотя не содержит какую-либо аминокислотную последовательность, "не принадлежащую CPP", необязательно может содержать модификацию, такую как липидная группа, раскрытую в данном документе.

В качестве альтернативы CPP, который нековалентно связан с белковым компонентом RGEN, может входить в состав белка слияния, имеющего как аминокислотную последовательность CPP, так и одну или несколько гетерологичных аминокислотных последовательностей (последовательностей, не принадлежащих белку RGEN).

Гетерологичная последовательность в таких вариантах осуществления может представлять собой последовательность домена или белка (например, флуоресцентного белка, такого как любой из раскрытых в данном документе, или любой домен/белок, перечисленный в раскрытии выше, касающемся слияний Cas). Другим примером является слияние домена димеризации с CPP, при этом данный домен димеризации может связываться с доменом димеризации, связанным или слитым с белковым компонентом RGEN.

Домены лейциновой застежки представляют собой примеры доменов димеризации согласно данному документу. Домены лейциновой застежки могут представлять собой домены из природных белков,

как известно, содержащих такие домены (например, факторы транскрипции), или могут быть разработаны синтетическим образом. Домен лейциновой застежки, связанный с CPP, может вступать в ассоциацию ("застегиваться") с доменом лейциновой застежки белкового компонента RGEN, связывая тем самым CPP и белковый компонент RGEN в комплекс, связанный нековалентной связью. Домены в составе пары доменов лейциновой застежки для нековалентного связывания CPP и белкового компонента RGEN могут быть одинаковыми (такая пара доменов образует гомодимерную лейциновую застежку) или разными (такая пара доменов образует гетеродимерную лейциновую застежку). Примеры доменов лейциновой застежки включают в себя раскрытые в публикациях заявок на патент США №№ 2003/0108869 и 2004/0147721. В определенных аспектах гомодимерная лейциновая застежка может быть образована с помощью домена лейциновой застежки из фактора транскрипции GCN4, при этом в других аспектах гетеродимерная лейциновая застежка может быть образована с помощью доменов лейциновой застежки из факторов транскрипции fos и jun соответственно.

При необходимости комплекс из связанных нековалентной связью белкового компонента RGEN и по меньшей мере одного CPP может дополнительно содержать один или несколько других белков/пептидов/доменов. Такие другие элементы необязательно можно применять для создания мостиковой связи между белковым компонентом RGEN и CPP, и они могут включать любую из промежуточных аминокислотных последовательностей, описанных выше.

В определенных вариантах осуществления с белковым компонентом RGEN может быть нековалентно связано более одного типа CPP (например, 2, 3, 4 или более разных типов CPP). Соотношение (молярное соотношение) CPP и белка RGEN, которое может применяться для получения такого комплекса, может составлять, например, по меньшей мере приблизительно 1:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1, 15:1, 20:1, 30:1, 40:1, или 50:1. В других аспектах среднее число CPP, связанных с помощью нековалентной связи с белком RGEN, может составлять по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, или 25, или по меньшей мере 5-10, 5-15, 5-20, или 5-25.

В определенных вариантах осуществления комплекс из связанных нековалентной связью белкового компонента RGEN и по меньшей мере одного CPP можно получить путем смешивания соответствующего количества каждого компонента (например, чтобы получить соотношение CPP и белка RGEN, раскрытое выше) в водной среде. Подходящая водная среда может содержать, например, буферный раствор, такой как PBS, или бессывороточную среду, такую как DMEM. Смесь можно инкубировать, например, в течение приблизительно 30, 60, 90 или 120 мин при температуре приблизительно 4-45°C, чтобы обеспечить возможность образования комплекса белок RGEN-CPP, связанных нековалентной связью. Подходящий объем (например, минимальный объем, который приемлемо покрывает/в котором в достаточной степени погружены клетки, подлежащие обработке) этого раствора, содержащего комплекс, можно применять в отношении клетки подходящим для типа клетки образом. В вариантах осуществления, в которых РНК-компонент ассоциирован с белковым компонентом RGEN, такое образование RGEN может предусматривать добавление РНК-компонента до, одновременно с или после инкубации CPP с белковым компонентом RGEN.

Композицию, содержащую комплекс из связанных нековалентной связью белкового компонента RGEN и по меньшей мере одного CPP можно применять в отношении любого типа клеток, раскрытого в данном документе. Необязательно, однако, данную композицию можно применять в отношении клеток организма, не относящегося к млекопитающим, такого как дрожжи, грибы и растения, но это исключает применение в отношении клеток млекопитающих.

Комплекс белок RGEN-CPP, в том виде, в котором он может присутствовать в композиции перед применением по отношению к клеткам, может иметь чистоту, например, по меньшей мере приблизительно 30%, 40%, 50%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% или 99%. В определенных вариантах осуществления такая чистота может приводиться на основе белка. В качестве примера, если чистота комплекса составляет по меньшей мере 80%, это будет означать, что по меньшей мере 80% всего белка в композиции составляет данный комплекс. В качестве альтернативы при обозначении чистоты комплекса учитывается не только чистота на основе белка, но также учитываются другие биомолекулы (например, липиды, сахараиды и/или нуклеиновые кислоты). В качестве примера, если чистота комплекса составляет по меньшей мере 80%, это может означать, что по меньшей мере 80% всех биомолекул в композиции согласно данному документу составляет данный комплекс. В определенных вариантах осуществления соединения такие как углеводы, соли и/или липиды и т. п. не влияют на определение процента чистоты согласно данному документу. Все эти раскрытия, касающиеся чистоты, можно также применить к комплексу RGEN-CPP (т.е. белковому компоненту RGEN, ассоциированному с РНК-компонентом).

Композиция согласно данному документу предпочтительно является водной, где растворитель, в котором растворен комплекс белок RGEN-CPP или комплекс RGEN-CPP, содержит по меньшей мере приблизительно 70, 75, 80, 85, 90, 95, 98, или 99 вес.% воды.

Концентрация комплекса в композиции может составлять, например, по меньшей мере приблизительно 0,5, 1,0, 1,5, 2,0, 2,5, 3,0, 3,5, 4,0, 4,5, 5,0, 6,0, 7,0, 8,0, 9,0 или 10,0 μM , или приблизительно 0,5-5,0 μM , 0,5-2,5 μM , 1,0-5,0 μM , 1,0-2,5 μM или 2,5-5,0 μM . Будет понятно, что такие композиции могут на-

ходиться в жидком состоянии.

В определенных вариантах осуществления рН композиции может составлять от приблизительно 4,0 до приблизительно 10,0. В качестве альтернативы рН может составлять приблизительно 4,0, 4,5, 5,0, 5,5, 6,0, 6,5, 7,0, 7,5, 8,0, 8,5, 9,0, 9,5 или 10,0. рН можно регулировать или контролировать путем добавления или включения подходящего буфера, в том числе без ограничений: HEPES, фосфатного (например, PBS), Tris, Tris-HCl, цитратного или их комбинации. Концентрация буфера в композиции согласно данному документу может составлять от 0 мМ до приблизительно 100 мМ или, например, приблизительно 10, 20 или 50 мМ. В определенных аспектах можно применять буфер HEPES (например, ~25 мМ HEPES, такой как 25 мМ HEPES/КОН, рН 7,5, 200 мМ KCl, 20% глицерина, 1 мМ DTT).

Композиция согласно данному документу может необязательно содержать другие компоненты в дополнение к комплексу белок RGEN-CPP или комплексу RGEN-CPP. Например, композиция может содержать одну или несколько солей, таких как соль натрия (например, NaCl, Na₂SO₄). Другие неограничивающие примеры солей включают в себя соли, имеющие (i) катион алюминия, аммония, бария, кальция, хрома (II или III), меди (I или II), железа (II или III), водорода, свинца (II), лития, магния, марганца (II или III), ртути (I или II), калия, серебра, натрия, стронция, олова (II или IV) или цинка и (ii) ацетат-, борат-, бромат-, бромид-, карбонат-, хлорат-, хлорид-, хлорит-, хромат-, дихромат-, дигидрофосфат-, феррицианид-, ферроцианид-, фторид-, гидрокарбонат-, гидрофосфат-, гидросульфат-, гидросульфид-, гидросульфит-, гидрид-, гидроксид-, гипохлорит-, йодат-, йодид-, нитрат-, нитрид-, нитрит-, оксалат-, оксид-, прехлорат-, перманганат-, пероксид-, фосфат-, фосфид-, фосфит-, силикат-, станнат-, станнит-, сульфат-, сульфид-, сульфит-, тартрат- или тиоцианат-анион. Таким образом, любая соль, имеющая, например, катион из (i) выше и анион из (ii) выше, может находиться в композиции согласно данному документу. Соль может присутствовать, например, при вес. % от приблизительно 0,01 до приблизительно 10,00 (или любом значении с шагом одна сотая от 0,01 до 10,00).

Комплекс белок RGEN-CPP согласно данному документу может пересекать (i) клеточную мембрану или (ii) клеточную стенку и клеточную мембрану клетки. В тех вариантах осуществления, в которых белковый компонент RGEN ассоциирован с РНК-компонентом (тем самым составляя полную RGEN), комплекс RGEN-CPP аналогичным образом обладает такой способностью пересекать клеточную мембрану/клеточную стенку. В определенных аспектах данного документа либо комплекс белок RGEN-CPP, либо комплекс RGEN-CPP может пересекать клеточную стенку и клеточную мембрану.

Комплекс белок RGEN-CPP или комплекс RGEN-CPP согласно данному документу необязательно пересекает клеточную стенку, которая содержит гликокаликс (капсулу). Эти варианты осуществления, как правило, относятся к прокариотическим клеткам (например, бактериям), некоторые из которых могут иметь гликокаликс, в зависимости от вида и условий роста.

Без желания придерживаться какой-либо конкретной теории или механизма, подразумевается, что CPP согласно данному документу может доставлять белковый компонент RGEN в клетку посредством процесса эндоцитоза. Примеры такого процесса могут включать макропиноцитоз, опосредуемый клатрином эндоцитоз, опосредуемый кавеолами/липидным рафтом эндоцитоз и/или механизмы опосредуемого рецепторами эндоцитоза (например, опосредуемого фагоцитарным рецептором поглощения, опосредуемого протеогликанами поглощения).

Как только комплекс белок RGEN-CPP или RGEN-CPP оказывается внутри клетки, он может пересекать мембрану органеллы, такую как, например, ядерная мембрана или мембрана митохондрии. В определенных вариантах осуществления эта способность зависит от присутствия по меньшей мере одной последовательности для нацеливания в органеллу (например, NLS, MTS), которая включена в белок RGEN. В других вариантах осуществления, однако, способность пересекать мембрану органеллы, такую как ядерная мембрана или мембрана митохондрии, не зависит от присутствия последовательности для нацеливания в органеллу (т. е. CPP в таких вариантах осуществления может отвечать за обеспечение возможности прохождения RGEN в органеллу, такую как ядро или митохондрии).

Клетка согласно данному документу может представлять собой клетку млекопитающего или клетку организма, не относящегося к млекопитающим, причем последнюю применяют в некоторых предпочтительных вариантах осуществления. В некоторых других аспектах клетка согласно данному документу может представлять собой клетку, существующую (i) в организме/ткани *in vivo*, (ii) в ткани или группе клеток *ex vivo* или (iii) в состоянии *in vitro*.

Микробная клетка согласно данному документу может представлять собой клетку, существующую в выделенном состоянии (например, *in vitro* клетки, культивируемые клетки) или невыделенном состоянии.

В определенных вариантах осуществления микробная клетка представляет собой грибную клетку, такую как дрожжевая клетка. В определенных аспектах данного документа дрожжи могут представлять собой дрожжи, размножающиеся бесполом способом (анаморфные) или половым способом (телеоморфные). Несмотря на то, что дрожжи согласно данному документу, как правило, существуют в одноклеточной форме, некоторые типы таких дрожжей необязательно могут быть способны образовывать псевдогифы (ряды из соединенных почкующихся клеток). В еще других аспектах дрожжи могут быть гаплоидными или диплоидными и/или они могут обладать способностью существовать в виде одной из этих

форм плоидности.

Примеры дрожжей согласно данному документу включают в себя традиционные виды дрожжей и нетрадиционные виды дрожжей. В определенных вариантах осуществления традиционные виды дрожжей представляют собой дрожжи, в которых происходят процессы репарации ДНК путем гомологичной рекомбинации (HR), а не процессы репарации, опосредуемые негомологичным соединением концов (NHEJ). Примеры традиционных видов дрожжей согласно данному документу включают в себя виды из рода *Saccharomyces* (например, *S. cerevisiae*, которые также известны как почкующиеся дрожжи, пекарские дрожжи и/или пивоваренные дрожжи; *S. bayanus*; *S. boulardii*; *S. bulderi*; *S. cariocanus*; *S. cariocus*; *S. chevalieri*; *S. dairenensis*; *S. ellipsoideus*; *S. eubayanus*; *S. exiguus*; *S. florentinus*; *S. kluyveri*; *S. martiniae*; *S. monacensis*; *S. norbensis*; *S. paradoxus*; *S. pastorianus*; *S. spencerorum*; *S. turicensis*; *S. unisporus*; *S. uvarum*; *S. zonatus*) и *Schizosaccharomyces* (например, *S. pombe*, которые также известны как делящиеся дрожжи; *S. cryophilus*; *S. japonicus*; *S. octosporus*).

Нетрадиционные виды дрожжей согласно данному документу не являются традиционными видами дрожжей, такими как виды *Saccharomyces* (например, *S. cerevisiae*) или *Schizosaccharomyces* (например, *S. pombe*). В определенных вариантах осуществления нетрадиционные виды дрожжей могут представлять собой дрожжи, в которых происходят процессы репарации ДНК путем NHEJ, а не процессы репарации, опосредуемые HR. Традиционные виды дрожжей, такие как *S. cerevisiae* и *S. pombe*, как правило, проявляют специфическую интеграцию донорной ДНК с короткими фланкирующими "плечами гомологии" (30-50 п. о.) при неизменных значениях эффективности более 70%, тогда как нетрадиционные виды дрожжей, такие как *Pichia pastoris*, *Pichia stipitis*, *Hansenula polymorpha*, *Yarrowia lipolytica* и *Kluyveromyces lactis* обычно демонстрируют специфическую интеграцию донорной ДНК аналогичной структуры при значениях эффективности менее 1% (Chen et al., PLoS ONE 8:e57 952). Таким образом, предпочтение в отношении процессов HR можно измерить, например, путем трансформации дрожжей с помощью подходящей донорной ДНК и определения степени, с которой она специфически рекомбинирует с геномным сайтом, на который предположительно нацеливается донорная ДНК. Предпочтение в отношении NHEJ (или низкое предпочтение в отношении HR), например, будет выражено, если такой анализ приведет к высокой степени случайной интеграции донорной ДНК в геном дрожжей. Анализы для определения показателя специфической (опосредуемой HR) и/или случайной (опосредуемой NHEJ) интеграции ДНК в дрожжах известны из уровня техники (например, Ferreira and Cooper, Genes Dev. 18:2249-2254; Corrigan et al., PLoS ONE 8:e69628; Weaver et al., Proc. Natl. Acad. Sci. U.S.A. 78:6354-6358; Keeney and Boeke, Genetics 136:849-856).

Принимая во внимание их низкий уровень активности HR, у нетрадиционных видов дрожжей согласно данному документу может

(i) наблюдаться показатель специфического нацеливания подходящей донорной ДНК, имеющей фланкирующие "плечи гомологии" из 30-50 п. о., составляющий, например, менее приблизительно 1%, 2%, 3%, 4%, 5%, 6%, 7% или 8%, и/или (ii) наблюдаться показатель случайной интеграции вышеуказанной донорной ДНК, составляющий, например, более приблизительно 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74% или 75%. Эти показатели (i) специфического нацеливания и/или (ii) случайной интеграции подходящей донорной ДНК могут характеризовать нетрадиционные виды дрожжей, как существующие до обеспечения RGEN, раскрытой в данном документе. Целью обеспечения RGEN в нетрадиционных видах дрожжей в определенных вариантах осуществления является создание сайт-специфических однонитевых разрывов (SSB) или двухнитевых разрывов (DSB) ДНК, чтобы привести дрожжи к HR в специфическом сайте. Таким образом, обеспечение подходящей RGEN в нетрадиционных видах дрожжей, как правило, должно приводить к тому, что дрожжи проявляют повышенный показатель HR с конкретной донорной ДНК. Такой повышенный показатель может быть по меньшей мере в приблизительно 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9- или 10-раз превышать показатель HR в подходящем контроле (например, таких же нетрадиционных видах дрожжей, трансформированных такой же донорной ДНК, но без подходящей RGEN).

Нетрадиционные виды дрожжей согласно данному документу можно культивировать в соответствии с любым способом, известным из уровня техники, таким как описанные в *Non-Conventional Yeasts in Genetics, Biochemistry and Biotechnology: Practical Protocols* (K. Wolf, K.D. Breunig, G. Barth, Eds., Springer-Verlag, Berlin, Germany, 2003), *Yeasts in Natural and Artificial Habitats* (J.F.T. Spencer, D.M. Spencer, Eds., Springer-Verlag, Berlin, Germany, 1997), и/или *Yeast Biotechnology: Diversity and Applications* (T. Satyanarayana, G. Kunze, Eds., Springer, 2009), все из которых включены в данный документ посредством ссылки.

Неограничивающие примеры нетрадиционных видов дрожжей согласно данному документу включают в себя дрожжи следующих родов: *Yarrowia*, *Pichia*, *Schwanniomyces*, *Kluyveromyces*, *Arxula*, *Trichosporon*, *Candida*, *Ustilago*, *Torulopsis*, *Zygosaccharomyces*, *Trigonopsis*, *Cryptococcus*, *Rhodotorula*, *Phaffia*, *Sporobolomyces*, *Pachysolen*, и *Moniliella*. Подходящим примером вида *Yarrowia* является *Y. lipolytica*. Подходящие примеры видов *Pichia* включают в себя *P. pastoris*, *P. methanolica*, *P. stipitis*, *P. anomala* и *P. angusta*. Подходящие примеры видов *Schwanniomyces* включают в себя *S. castellii*, *S. alluvius*, *S. hominis*, *S. occidentalis*, *S. capriottii*, *S. etchellsii*, *S. polymorphus*, *S. pseudopolymorphus*, *S. vanrijiae* и *S. yamadae*.

Подходящие примеры видов *Kluyveromyces* включают в себя *K. lactis*, *K. marxianus*, *K. fragilis*, *K. drosophilae*, *K. thermotolerans*, *K. phaseolusporus*, *K. vanudenii*, *K. waltii*, *K. africanus* и *K. polysporus*. Подходящие примеры видов *Axhula* включают в себя *A. adenivorans* и *A. terrestre*. Подходящие примеры видов *Trichosporon* включают в себя *T. cutaneum*, *T. capitatum*, *T. inkin* и *T. beemeri*. Подходящие примеры видов *Candida* включают в себя *C. albicans*, *C. ascalaphidarum*, *C. amphixiae*, *C. antarctica*, *C. apicola*, *C. argentea*, *C. atlantica*, *C. atmosphaerica*, *C. blattae*, *C. bromeliacearum*, *C. carpophila*, *C. carvajalis*, *C. cerambycidarum*, *C. chauliodes*, *C. corydali*, *C. dosseyi*, *C. dubliniensis*, *C. ergatensis*, *C. fructus*, *C. glabrata*, *C. fermentati*, *C. guilliermondii*, *C. haemulonii*, *C. insectamens*, *C. insectorum*, *C. intermedia*, *C. jeffresii*, *C. kefyr*, *C. keroseneae*, *C. krusei*, *C. lusitaniae*, *C. lyxosophila*, *C. maltosa*, *C. marina*, *C. membranifaciens*, *C. milleri*, *C. mogii*, *C. oleophila*, *C. oregonensis*, *C. parapsilosis*, *C. quercitrusa*, *C. rugosa*, *C. sake*, *C. shehatae*, *C. temnochilae*, *C. tenuis*, *C. theae*, *C. tolerans*, *C. tropicalis*, *C. tsuchiyaе*, *C. sinolaborantium*, *C. sojae*, *C. subhashii*, *C. viswanathii*, *C. utilis*, *C. ubatubensis* и *C. zemplinina*. Подходящие примеры видов *Ustilago* включают в себя *U. avenae*, *U. esculenta*, *U. hordei*, *U. maydis*, *U. nuda* и *U. tritici*. Подходящие примеры видов *Torulopsis* включают в себя *T. geochares*, *T. azyma*, *T. glabrata* и *T. Candida*. Подходящие примеры видов *Zygosaccharomyces* включают в себя *Z. bailii*, *Z. bisporus*, *Z. cidri*, *Z. fermentati*, *Z. florentinus*, *Z. kombuchaensis*, *Z. lentus*, *Z. mellis*, *Z. microellipsoides*, *Z. mrakii*, *Z. pseudorouxii* и *Z. rouxii*. Подходящие примеры видов *Trigonopsis* включают в себя *T. variabilis*. Подходящие примеры видов *Cryptococcus* включают в себя *C. laurentii*, *C. albidus*, *C. neoformans*, *C. gattii*, *C. uniguttulatus*, *C. adeliensis*, *C. aerius*, *C. albidosimilis*, *C. antarcticus*, *C. aquaticus*, *C. ater*, *C. bhutanensis*, *C. consortionis*, *C. curvatus*, *C. phenolicus*, *C. skinneri*, *C. terreus* и *C. vishniacii*. Подходящие примеры видов *Rhodotorula* включают в себя *R. acheniorum*, *R. tula*, *R. acuta*, *R. americana*, *R. araucariae*, *R. arctica*, *R. armeniaca*, *R. aurantiaca*, *R. auriculariae*, *R. bacarum*, *R. benthica*, *R. biourgei*, *R. bogoriensis*, *R. bronchialis*, *R. buffonii*, *R. calyptogenaе*, *R. chungnamensis*, *R. cladiensis*, *R. coralina*, *R. cresolica*, *R. crocea*, *R. cycloclastica*, *R. dairenensis*, *R. diffluens*, *R. evergladiensis*, *R. ferulica*, *R. foliorum*, *R. fragaria*, *R. fujisanensis*, *R. futronensis*, *R. gelatinosa*, *R. glacialis*, *R. glutinis*, *R. gracilis*, *R. graminis*, *R. grinbergii*, *R. himalayensis*, *R. hinnulea*, *R. histolytica*, *R. hylophila*, *R. incarnata*, *R. ingeniosa*, *R. javanica*, *R. koishikawensis*, *R. lactosa*, *R. lamellibrachiae*, *R. laryngis*, *R. lignophila*, *R. lini*, *R. longissima*, *R. ludwigii*, *R. lysinophila*, *R. marina*, *R. martyniae-fragantis*, *R. matritensis*, *R. meli*, *R. minuta*, *R. mucilaginoso*, *R. nitens*, *R. nothofagi*, *R. oryzae*, *R. pacifica*, *R. pallida*, *R. peneaus*, *R. philyla*, *R. phylloplana*, *R. pilatii*, *R. pilimanae*, *R. pinicola*, *R. plicata*, *R. polymorpha*, *R. psychrophenolica*, *R. psychrophila*, *R. pustula*, *R. retinophila*, *R. rosacea*, *R. rosulata*, *R. rubefaciens*, *R. rubella*, *R. rubescens*, *R. rubra*, *R. rubrorugosa*, *R. rufula*, *R. rutila*, *R. sanguinea*, *R. sanniei*, *R. sartoryi*, *R. silvestris*, *R. simplex*, *R. sinensis*, *R. slooffiae*, *R. sonckii*, *R. straminea*, *R. subericola*, *R. suganii*, *R. taiwanensis*, *R. taiwaniana*, *R. terpenoidalis*, *R. terrea*, *R. texensis*, *R. tokyoensis*, *R. ulzamae*, *R. vanillica*, *R. vuilleminii*, *R. yarrowii*, *R. yunnanensis* и *R. zsoitii*. Подходящие примеры видов *Phaffia* включают в себя *P. rhodozyma*. Подходящие примеры видов *Sporobolomyces* включают в себя *S. alborubescens*, *S. bannaensis*, *S. beijingensis*, *S. bischoffiae*, *S. clavatus*, *S. coprosmae*, *S. coprosmicola*, *S. corallinus*, *S. dimmenae*, *S. dracophylli*, *S. elongatus*, *S. gracilis*, *S. inositophilus*, *S. johnsonii*, *S. koalaе*, *S. magnisporus*, *S. novozealandicus*, *S. odoratus*, *S. patagonicus*, *S. productus*, *S. roseus*, *S. sasicola*, *S. shibatanius*, *S. singularis*, *S. subbrunneus*, *S. symmetricus*, *S. syzygii*, *S. taupoensis*, *S. tsugae*, *S. xanthus* и *S. yunnanensis*. Подходящие примеры видов *Pachysolen* и *Moniliella* включают в себя *P. tannophilus* и *M. Pollinis* соответственно. Еще одни примеры нетрадиционных видов дрожжей согласно данному документу включают в себя виды *Pseudozyma* (например, *S. antarctica*), виды *Thodotorula* (например, *T. bogoriensis*), виды *Wickerhamiella* (например, *W. domercqiae*) и виды *Starmerella* (например, *S. bombicola*).

Yarrowia lipolytica является предпочтительным в определенных вариантах осуществления, раскрытых в данном документе.

Примеры подходящих *Y. lipolytica* включают в себя следующие изоляты, доступные из Американской коллекции типовых культур (ATCC, Manassas, VA): обозначение штаммов ATCC #20362, #8862, #8661, #8662, #9773, #15586, #16617, #16618, #18942, #18943, #18944, #18945, #20114, #20177, #20182, #20225, #20226, #20228, #20327, #20255, #20287, #20297, #20315, #20320, #20324, #20336, #20341, #20346, #20348, #20363, #20364, #20372, #20373, #20383, #20390, #20400, #20460, #20461, #20462, #20496, #20510, #20628, #20688, #20774, #20775, #20776, #20777, #20778, #20779, #20780, #20781, #20794, #20795, #20875, #20241, #20422, #20423, #32338, #32339, #32340, #32341, #34342, #32343, #32935, #34017, #34018, #34088, #34922, #34922, #38295, #42281, #44601, #46025, #46026, #46027, #46028, #46067, #46068, #46069, #46070, #46330, #46482, #46483, #46484, #46436, #60594, #62385, #64042, #74234, #76598, #76861, #76862, #76982, #90716, #90811, #90812, #90813, #90814, #90903, #90904, #90905, #96028, #201241, #201242, #201243, #201244, #201245, #201246, #201247, #201249, и/или #201847.

Грибная клетка согласно данному документу может представлять собой дрожжевую клетку (например, как описано выше) или клетку любого другого типа грибов, как например, нитчатого гриба. Например, гриб согласно данному документу может представлять собой гриб из отдела Basidiomycetes, Zygomycetes, Chytridiomycetes или Ascomycetes. Примеры нитчатых грибов согласно данному документу включают в себя грибы из родов *Trichoderma*, *Chrysosporium*, *Thielavia*, *Neurospora* (например, *N. crassa*, *N. sitophila*), *Cryphonectria* (например, *C. parasitica*), *Aureobasidium* (например, *A. pullulans*), *Filibasidium*, *Piromyces*, *Cryptococcus*, *Acremonium*, *Tolypocladium*, *Scytalidium*, *Schizophyllum*, *Sporotrichum*, *Penicil-*

lium (например, *P. bilaiae*, *P. camemberti*, *P. candidum*, *P. chrysogenum*, *P. expansum*, *P. funiculosum*, *P. glaucum*, *P. marneffeii*, *P. roqueforti*, *P. verrucosum*, *P. viridicatum*), *Gibberella* (например, *G. acuminata*, *G. avenacea*, *G. baccata*, *G. circinata*, *G. cyanogena*, *G. fujikuroi*, *G. intricans*, *G. pulcaris*, *G. stilboides*, *G. tricincta*, *G. zaeae*), *Myceliophthora*, *Mucor* (например, *M. rouxii*, *M. circinelloides*), *Aspergillus* (например, *A. niger*, *A. oryzae*, *A. nidulans*, *A. flavus*, *A. lentulus*, *A. terreus*, *A. clavatus*, *A. fumigatus*), *Fusarium* (например, *F. graminearum*, *F. oxysporum*, *F. bubigenum*, *F. solani*, *F. oxysporum*, *F. verticillioides*, *F. proliferatum*, *F. venenatum*), и *Humicola*, и их анаморфы и телеоморфы. При необходимости род и виды грибов согласно данному документу можно определить по морфологии, как раскрыто в Barnett and Hunter (Illustrated Genera of Imperfect Fungi, 3rd Edition, Burgess Publishing Company, 1972). В определенных вариантах осуществления грибок необязательно может быть охарактеризован как вредитель/патоген растения или животного (например, человека).

В определенных аспектах данного документа виды *Trichoderma* включают в себя *T. aggressivum*, *T. amazonicum*, *T. asperellum*, *T. atroviride*, *T. aureoviride*, *T. austrokingii*, *T. brevicompactum*, *T. candidum*, *T. caribbaeum*, *T. catoptron*, *T. cremeum*, *T. ceramicum*, *T. cerinum*, *T. chlorosporum*, *T. chromospermum*, *T. cinnamomeum*, *T. citrinoviride*, *T. crassum*, *T. cremeum*, *T. dingleyeae*, *T. dorotheae*, *T. effusum*, *T. erinaceum*, *T. estonicum*, *T. fertile*, *T. gelatinosus*, *T. ghanense*, *T. hamatum*, *T. harzianum*, *T. helicum*, *T. intricatum*, *T. konilangbra*, *T. koningii*, *T. koningiopsis*, *T. longibrachiatum*, *T. longipile*, *T. minutisporum*, *T. oblongisporum*, *T. ovalisporum*, *T. petersenii*, *T. phyllostahydis*, *T. piluliferum*, *T. pleuroticola*, *T. pleurotum*, *T. polysporum*, *T. pseudokoningii*, *T. pubescens*, *T. reesei*, *T. rogersonii*, *T. rossicum*, *T. saturnisporum*, *T. sinensis*, *T. sinuosum*, *T. spirale*, *T. stramineum*, *T. strigosum*, *T. stromaticum*, *T. surrotundum*, *T. taiwanense*, *T. thailandicum*, *T. thelephoricolum*, *T. theobromicola*, *T. tomentosum*, *T. velutinum*, *T. virens*, *T. viride* и *T. viridescens*. Вид *Trichoderma* согласно данному документу можно культивировать и/или подвергать манипуляциям, как описано, например, в *Trichoderma: Biology and Applications* (P.K. Mukherjee et al., Eds., CABI, Oxfordshire, UK, 2013), который включен в данный документ посредством ссылки.

В определенных вариантах осуществления микробная клетка представляет собой клетку водорослей. Например, клетка водорослей может быть из любого из следующих отделов: *Chlorophyta* (зеленые водоросли), *Rhodophyta* (красные водоросли), *Phaeophyceae* (бурые водоросли), *Bacillariophyceae* (диатомовые) и *Dinoflagellata* (финофлагелляты). В других аспектах клетка водорослей может быть представителем микроскопических водорослей (например, фитопланктон, микрофиты или планктонные водоросли) или макроскопических водорослей (морские водоросли, морская капуста). В качестве дополнительных примеров клетка водорослей согласно данному документу может представлять собой клетку из видов *Porphyta* (красные водоросли), видов *Palmaria*, таких как *P. palmata* (дульсе), видов *Arthrospira*, таких как *A. platensis* (спирулина), *Chlorella* (например, *C. protothecoides*), вида *Chondrus*, такого как *C. crispus* (ирландский мох), *Aphanizomenon*, *Sargassum*, *Cochayuyo*, *Botryococcus* (например, *B. braunii*), *Dunaliella* (например, *D. tertiolecta*), *Gracilaria*, *Pleurochrysis* (например, *P. carterae*), *Ankistrodesmus*, *Cyclotella*, *Hantzschia*, *Nannochloris*, *Nannochloropsis*, *Nitzschia*, *Phaeodactylum* (например, *P. tricorutum*), *Scenedesmus*, *Stichococcus*, *Tetraselmis* (например, *T. suecica*), *Thalassiosira* (например, *T. pseudonana*), *Cryptocodinium* (например, *C. cohnii*), *Neochloris* (например, *N. oleoabundans*) или *Schiochytrium*. Вид водорослей согласно данному документу можно культивировать и/или подвергать манипуляциям, как описано, например, в Thompson (Algal Cell Culture. Encyclopedia of Life Support System (EOLSS), Biotechnology Vol 1, доступном на интернет-сайте eolss.net/sample-chapters), который включен в данный документ посредством ссылки.

В одном варианте осуществления способ включает способ доставки белкового компонента направляемой РНК эндонуклеазы (RGEN) в микробную клетку, причем указанный способ включает: приведение микробной клетки в контакт с композицией, содержащей белковый компонент направляемой РНК эндонуклеазы (RGEN) и по меньшей мере один проникающий в клетку пептид (CPP), где указанный белковый компонент и CPP ковалентно или нековалентно связаны друг с другом в комплекс белок RGEN-CPP, где указанный комплекс белок RGEN-CPP пересекает (i) клеточную мембрану или (ii) клеточную стенку и клеточную мембрану клеткам с попаданием, таким образом, в микробную клетку. Микробные клетки, применимые в способах и композиции, описанных в данном документе, включают в себя клетки, выбранные из клеток видов *Phytophthora*, таких как *Phytophthora capsici* (Lamour et al. 2012. The oomycete broad-host-range pathogen *Phytophthora capsici*. Mol. Plant Pathol. May 13(4): 329-337), видов *Zymoseptoria*, таких как *Septoria tritici* (Testa et al. 2015. Overview of genomic and bioinformatics resources for *Zymoseptoria tritici*. Fungal Genet. Biol. Jun. 79:13-16) и видов *Botrytis*, таких как *Botrytis cinerea* (Hahn M. 2014. Повышение угрозы устойчивости к фунгицидам у патогенных грибов растений: *Botrytis* as a case study. J. Chem. Biol 7:133-141).

Клетка протистов согласно данному документу может быть выбрана, например, из класса *Ciliata* (например, родов *Tetrahymena*, *Paramecium*, *Colpidium*, *Colpoda*, *Glaucocoma*, *Platyophrya*, *Vorticella*, *Potomacus*, *Pseudocohnilembus*, *Euplotes*, *Engelmanniella*, и *Stylonichia*), подтипа *Mastigophora* (жгутиковые), класса *Phytomastigophorea* (например, родов *Euglena*, *Astasia*, *Naematococcus* и *Cryptocodinium*), класса *Zoomastigophorea*, надкласса *Rhizopoda*, класса *Lobosea* (например, рода *Amoeba*) и класса *Eumycetozoa* (например, родов *Dictyostelium* и *Physarum*). Определенные виды протистов согласно данному документу

можно культивировать и/или подвергать манипуляциям, как описано, например, в ATCC® Protistology Culture Guide: tips and techniques for propagating protozoa and algae (2013, доступен на интернет-сайте Американской коллекции типовых культур), который включен в данный документ посредством ссылки. В определенных вариантах осуществления бактерии необязательно могут быть охарактеризованы как вредитель/патоген растения или животного (например, человека).

Бактериальная клетка в определенных вариантах осуществления может представлять собой клетку в форме кокков, бацилл, спиросет, сферопластов, протопластов и т. д. Другие неограничивающие примеры бактерий включают в себя такие, которые являются грамотрицательными и грамположительными. Еще дополнительные неограничивающие примеры бактерий включают в себя бактерии родов *Salmonella* (например, *S. typhi*, *S. enteritidis*), *Shigella* (например, *S. dysenteriae*), *Escherichia* (например, *E. coli*), *Enterobacter*, *Serratia*, *Proteus*, *Yersinia*, *Citrobacter*, *Edwardsiella*, *Providencia*, *Klebsiella*, *Hafnia*, *Ewingella*, *Kluuyvera*, *Morganella*, *Planococcus*, *Stomatococcus*, *Micrococcus*, *Staphylococcus* (например, *S. aureus*, *S. epidermidis*), *Vibrio* (например, *V. cholerae*), *Aeromonas*, *Plesiomonas*, *Haemophilus* (например, *H. influenzae*), *Actinobacillus*, *Pasteurella*, *Mycoplasma* (например, *M. pneumoniae*), *Ureaplasma*, *Rickettsia*, *Coxiella*, *Rochalimaea*, *Ehrlichia*, *Streptococcus* (например, *S. pyogenes*, *S. mutans*, *S. pneumoniae*), *Enterococcus* (например, *E. faecalis*), *Aerococcus*, *Gemella*, *Lactococcus* (например, *L. lactis*), *Leuconostoc* (например, *L. mesenteroides*), *Pedococcus*, *Bacillus* (например, *B. cereus*, *B. subtilis*, *B. thuringiensis*), *Corynebacterium* (например, *C. diphtheriae*), *Arcanobacterium*, *Actinomyces*, *Rhodococcus*, *Listeria* (например, *L. monocytogenes*), *Erysipelothrix*, *Gardnerella*, *Neisseria* (например, *N. meningitidis*, *N. gonorrhoeae*), *Campylobacter*, *Arcobacter*, *Wolinella*, *Helicobacter* (например, *H. pylori*), *Achromobacter*, *Acinetobacter*, *Agrobacterium* (например, *A. tumefaciens*), *Alcaligenes*, *Chryseomonas*, *Comamonas*, *Eikenella*, *Flavimonas*, *Flavobacterium*, *Moraxella*, *Oligella*, *Pseudomonas* (например, *P. aeruginosa*), *Shewanella*, *Weeksella*, *Xanthomonas*, *Bordetella*, *Francisella*, *Brucella*, *Legionella*, *Afipia*, *Bartonella*, *Calymatobacterium*, *Cardiobacterium*, *Streptobacillus*, *Spirillum*, *Peptostreptococcus*, *Peptococcus*, *Sardinia*, *Coprococcus*, *Ruminococcus*, *Propionibacterium*, *Mobiluncus*, *Bifidobacterium*, *Eubacterium*, *Lactobacillus* (например, *L. lactis*, *L. acidophilus*), *Rothia*, *Clostridium* (например, *C. botulinum*, *C. perfringens*), *Bacteroides*, *Porphyromonas*, *Prevotella*, *Fusobacterium*, *Bilophila*, *Leptotrichia*, *Wolinella*, *Acidaminococcus*, *Megasphaera*, *Veilonella*, *Nocardia*, *Actinomadura*, *Nocardiosis*, *Streptomyces*, *Micropolysporas*, *Thermoactinomyces*, *Mycobacterium* (например, *M. tuberculosis*, *M. bovis*, *M. leprae*), *Treponema*, *Borrelia* (например, *B. burgdorferi*), *Leptospira*, и *Chlamydiae*. В определенных вариантах осуществления бактерии необязательно могут быть охарактеризованы как вредитель/патоген растения или животного (например, человека). В определенных вариантах осуществления бактерии могут находиться в смешанной микробной популяции (например, содержащей другие бактерии или содержащей дрожжи и/или другие бактерии).

В определенных вариантах осуществления клетка архей может происходить из любого типа архей, такого как *Euryarchaeota*, *Crenarchaeota*, *Nanoarchaeota*, *Korarchaeota*, *Aigarchaeota* или *Thaumarchaeota*. Клетки архей согласно данному документу, например, могут быть экстремофильными (например, способными расти и/или размножаться в физически или геохимически экстремальных условиях, которые являются губительными для большинства форм жизни). Некоторые примеры экстремофильных архей включают таких, которые являются термофильными (например, могут расти при температурах 45-122°C), гипертермофильными (например, могут расти при температурах 80-122°C), ацидофильными (например, могут расти при уровнях pH 3 или ниже), алкалофильными (например, могут расти при уровнях pH 9 или выше) и/или галофильными (например, могут расти при высоких концентрациях солей [например, 20-30% NaCl]). Примеры видов архей включают таковые из родов *Halobacterium* (например, *H. volcanii*), *Sulfolobus* (например, *S. solfataricus*, *S. acidocaldarius*), *Thermococcus* (например, *T. alkaliphilus*, *T. celer*, *T. chitonophagus*, *T. gammatolerans*, *T. hydrothermalis*, *T. kodakarensis*, *T. litoralis*, *T. peptonophilus*, *T. profundus*, *T. stetteri*), *Methanocaldococcus* (например, *M. thermolithotrophicus*, *M. jannaschii*), *Methanococcus* (например, *M. maripaludis*), *Methanothermobacter* (например, *M. marburgensis*, *M. thermoautotrophicus*), *Archaeoglobus* (например, *A. fulgidus*), *Nitrosopumilus* (например, *N. maritimus*), *Metallosphaera* (например, *M. sedula*), *Ferroplasma*, *Thermoplasma*, *Methanobrevibacter* (например, *M. smithii*), и *Methanosphaera* (например, *M. stadtmannae*).

Примеры клеток насекомых согласно данному документу включают клетки *Spodoptera frugiperda*, клетки *Trichoplusia ni*, клетки *Bombyx mori* и т.п. Клетки *S. frugiperda* включают, например, Sf9 и Sf21. Клетки яичников *T. ni* включают, например, клетки HIGH FIVE (другое название BTI-TN-5B1-4, производимые Invitrogen). Клетки *B. mori*, например, включают N4. Определенные клетки насекомых согласно данному документу можно культивировать и/или подвергать манипуляции, например, как описано в *Growth and Maintenance of Insect cell lines* (2010, Invitrogen, Руководство, часть № 25-0127, MAN0000030), который включен в данный документ посредством ссылки. В других аспектах клетка насекомого может представлять собой клетку вредителя/патогена растений, такого как совка, совка-ипсилон, кукурузная совка, земляная кукурузная блошка, тля сортовая, тля кукурузная корневая, кукурузный мотылек, травяная совка, подземная совка, хрущик японский, малая кукурузная огневка, долгоносик маисовый, *Melanotus communis*, муха ростковая, огневки-травянки, галлица сортовая, сортовая огневка, южный кукурузный долгоносик, блошка 11-точечная Говарда, огневка кукурузная стеблевая, южный картофельный про-

волочник, клещик паутиный, совка стеблевая, тростниковый жук, табачный проволочник, личинка хруща, тля, долгоносик хлопковый, совка хлопковая, совка ни, клоп луговой, трипс, клещик паутиный двупятнистый, совка капустная, долгоносик люцерновый, долгоносик точечный, долгоносик клубеньковый желтоногий, травяная совка, кузнечик, пенница слюнявая, тля гороховая, цикадка картофельная, огневка-травянка, совка маргаритковая, малая кукурузная огневка, табачный трипс, проволочник, пьявица красногрудая, клоп-черепашка, большая злаковая тля, обыкновенная злаковая тля, гессенская муха, бобовый листоед, совка малая, шпанская мушка, листоед виноградный, совка клеверная, мексиканская фасолева коровка, соевая совка, соевый усач, клоп-щитник, цикадка-горбатка люцерновая, совка бархатных бобов, листовертка, совка капустная, гусеница озимой совки, хрущ блестящий зеленый, тля персиковая зеленая, бражник, моль картофельная клубневая, южная медведка, клоп-слепняк, жук-блошка шершавая, овощной долгоносик или белокаемчатый жук. В качестве альтернативы клетка насекомого может представлять собой клетку вредителя/патогена животного (например, человека).

Клетка нематоды, например, может происходить из нематоды, принадлежащей к любому из следующих родов: *Meloidogyne* (галловая нематода), *Pratylenchus* (ранящая нематода), *Heterodera* (цистообразующая нематода), *Globodera* (цистообразующая нематода), *Ditylenchus* (нематода, поражающая стебли и луковицы), *Tylenchulus* (цитрусовая нематода), *Xiphinema* (кинжальная нематода), *Radopholus* (сверлящая нематода), *Rotylenchulus* (почковидная нематода), *Helicotylenchus* (нематода спиральная) или *Belonolaimus* (жалящая нематода). В некоторых вариантах осуществления нематода необязательно может быть охарактеризована как вредитель/патоген растения или животного (например, человека). В других аспектах нематода может представлять собой *C. elegans*.

Клетка рыбы согласно данному документу может представлять собой любую из раскрытых, например, в патентах США №№ 7408095 и 7217564 и *Tissue Culture of Fish Cell Lines* (T. Ott, NWFHS Laboratory Procedures Manual - Second Edition, Chapter 10, 2004), которые включены в данный документ посредством ссылки. В данные источники также раскрыта информация, касающаяся культивирования клеток рыб и/или манипуляций с ними. Неограничивающие примеры клеток рыб могут представлять собой клетки костистых рыб, таких как данио-рерио, медака, гигантский данио или рыба фугу.

Растительная клетка согласно данному документу может представлять собой, например, растительную клетку однодольного растения или растительную клетку двудольного растения. Примеры однодольных растений согласно данному документу включают кукурузу (*Zea mays*), рис (*Oryza sativa*), рожь (*Secale cereale*), сорго (*Sorghum bicolor*, *Sorghum vulgare*), просо (например, африканское просо, *Pennisetum glaucum*), просо культурное (*Panicum miliaceum*), щетинник итальянский (*Setaria italica*), просо пальчатое (*Eleusine coracana*), пшеницу (*Triticum aestivum*), сахарный тростник (*Saccharum spp.*), разновидности овса (*Avena*), ячмень (*Hordeum*), просо прутьевидное (*Panicum virgatum*), ананас (*Ananas comosus*), банан (*Musa spp.*), пальму, декоративные растения и разновидности газонной травы. Примеры двудольных растений согласно данному документу включают сою (*Glycine max*), канолу (*Brassica napus* и *B. campestris*), люцерну (*Medicago sativa*), табак (*Nicotiana tabacum*), *Arabidopsis* (*A. thaliana*), подсолнечник (*Helianthus annuus*), хлопчатник (*Gossypium arboreum*), арахис (*Arachis hypogaea*), томат (*Solanum lycopersicum*), и картофель (*Solanum tuberosum*). Растительная клетка может происходить из любой части растения и/или растения на любой стадии развития.

Растительные клетки согласно данному документу могут быть выращены или регенерированы в растения с применением традиционных условий, см., например, McCormick et al., (1986) *Plant Cell Rep* 5:81-4. Затем регенерированные растения можно выращивать, и опылять либо с помощью растений той же линии, либо с помощью растений другой линии, и идентифицировать полученное потомство, имеющее требуемую характеристику (например, изменение) и/или содержащее введенный полинуклеотид или полипептид. Можно вырастить два или более поколений, чтобы удостовериться в том, что изменение стабильно поддерживается и наследуется, и провести сбор семян.

В определенных вариантах осуществления клетки млекопитающих могут представлять собой клетки человека, примата, не являющегося человеком (например, обезьяны, человекообразной обезьяны), грызуна (например, мыши, крысы, хомячка, морской свинки), кролика, собаки, кошки, коровы, свиньи, лошади, козы или овцы. Другие примеры клеток млекопитающих согласно данному документу включают первичные эпителиальные клетки (например, кератиноциты, клетки цервикального эпителия, клетки эпителия бронхов, клетки эпителия трахеи, эпителиальные клетки почки, эпителиальные клетки сетчатки); устойчивые клеточные линии (например, клетки почек эмбриона 293, клетки цервикального эпителия HeLa, клетки сетчатки PER-C6, MDBK, CRFK, MDCK, CHO, BeWo, клетки Chang, клетки Detroit 562, Hep-2, KB, LS 180, LS 174T, NCI-H-548, RPMI 2650, SW-13, T24, WI-28 VA13, 2RA, WISH, BS-C-1, LLC-MK2, Clone M-3, RAG, TCMK-1, LLC-PK1, PK-15, GH1, GH3, L2, LLC-RC 256, MH1C1, XC, MDOK, VSW, TH-1, B1); любую эпителиальную, мезенхимальную (например, фибробласт), нервную или мышечную клетку из любой ткани или органа (например, кожи, сердца; печени; почки; ободочной кишки; кишечника; пищевода; желудка; нервной ткани, такой как головной мозг или спинной мозг; легкого; сосудистой ткани; лимфоидной ткани, такой как лимфатический узел, аденоид, миндалина, костный мозг или кровь; селезенки) и линии фибробластных или фибробластоподобных клеток (например, TRG-2, IMR-33, клетки Don, GHK-21, клетки цитруллинемии, клетки Dempsey, Detroit 551, Detroit 510, Detroit

525, Detroit 529, Detroit 532, Detroit 539, Detroit 548, Detroit 573, HEL 299, IMR-90, MRC-5, WI-38, WI-26, MiC11, CV-1, COS-1, COS-3, COS-7, Vero, DBS-FrhL-2, BALB/3T3, F9, SV-T2, M-MSV-BALB/3T3, K-BALB, BLO-11, NOR-10, C3H/10T1/2, HSDM1C3, KLN205, клетки McCoy, L-клетки мыши, SCC-PSA1, клетки Swiss/3T3, клетки индийского мунджака, SIRC, клетки Jensen). Способы культивирования линий клеток млекопитающих и манипуляции с ними известны из уровня техники.

В определенных вариантах осуществления микробная клетка может происходить из любого патогена и/или вредителя животного или растения. Примеры таких патогенов/вредителей включают различные типы бактерий, грибов, дрожжей, протистов, нематод и насекомых. Специалистам в данной области техники распознают примеры таких патогенов/вредителей, раскрытые выше.

Как описано в данном документе (см. пример 10), проникающие в клетку пептиды были способны доставлять груз в различные виды эукариот, в том числе в *Phytophthora capsici*, *Septoria tritici* и *Botrytis cinerea*.

В одном варианте осуществления способ, описанный в данном документе, представляет собой способ доставки белкового компонента направляемой РНК эндонуклеазы (RGEN) в микробную клетку, выбранную из группы, состоящей из *Phytophthora capsici*, *Septoria tritici* и *Botrytis cinerea*, при этом указанный способ включает приведение микробной клетки в контакт с композицией, содержащей белковый компонент направляемой РНК эндонуклеазы (RGEN) и по меньшей мере один проникающий в клетку пептид (CPP), где указанный белковый компонент и CPP ковалентно или нековалентно связаны друг с другом в комплекс белок RGEN-CPP, где указанный комплекс белок RGEN-CPP пересекает (i) клеточную мембрану или (ii) клеточную стенку и клеточную мембрану клетки, попадая, таким образом, в микробную клетку.

В определенных вариантах осуществления данного документа композиция может содержать по меньшей мере один белковый компонент комплекса направляющий полинуклеотид/эндонуклеаза Cas и по меньшей мере один проникающий в клетку пептид (CPP), где белковый компонент и CPP ковалентно или нековалентно связаны друг с другом в комплекс полинуклеотид/белок эндонуклеазы-CPP, и где комплекс полинуклеотид/белок эндонуклеазы-CPP может пересекать (i) клеточную мембрану или (ii) клеточную стенку и клеточную мембрану клетки (такой как микробная клетка). Направляющий полинуклеотид и эндонуклеаза Cas способны образовывать комплекс, называемый "комплекс направляющий полинуклеотид/эндонуклеаза Cas", который обеспечивает способность эндонуклеазы Cas вводить двухнитевой разрыв в целевой сайт ДНК.

Раскрытое изобретение также относится к способу доставки белкового компонента направляемой РНК эндонуклеазы (RGEN) в клетку (такую как микробная клетка). Данный способ включает приведение клетки в контакт с композицией, содержащей белковый компонент RGEN и по меньшей мере один проникающий в клетку пептид (CPP), где белковый компонент RGEN и CPP ковалентно или нековалентно связаны друг с другом в комплекс белок RGEN-CPP. В результате данной стадии приведения в контакт комплекс белок RGEN-CPP может пересекать (i) клеточную мембрану или (ii) клеточную стенку и клеточную мембрану клетки, и тем самым проникать в клетку. В определенных вариантах осуществления, в которых белковый компонент RGEN ассоциирован с РНК-компонентом (образуя тем самым RGEN), раскрытый способ направлен на доставку комплекса RGEN-CPP в клетку. Дополнительно, поскольку в определенных вариантах осуществления RGEN можно применять в опосредуемом RGEN нацеливании на ДНК, данный способ необязательно может быть охарактеризован как способ нацеливания на ДНК в клетке.

Данный способ можно осуществлять на практике с применением, например, любого из раскрытых выше вариантов осуществления или представленных ниже примеров, с учетом каждого из признаков способа (например, типа клеток, белкового компонента RGEN, CPP, последовательности для нацеливания на органеллу и т. д.). Таким образом, любой из признаков, раскрытых выше или в примерах, или любую комбинацию данных признаков, можно применять соответствующим образом для характеристики вариантов осуществления способа доставки согласно данному документу. Следующие признаки способа доставки приведены в качестве примеров.

Варианты осуществления способа доставки согласно данному документу включают приведение клетки (такой как микробная клетка) в контакт с композицией, содержащей комплекс белок RGEN-CPP. Предполагается, такое приведение в контакт приводит к взаимодействию комплекса со внешней поверхностью клетки (например, клеточной мембраной, клеточной стенкой), позволяя тем самым компоненту CPP комплекса инициировать прохождение комплекса через (i) клеточную мембрану или (ii) клеточную стенку и клеточную мембрану.

Приведение композиции, содержащей комплекс белок RGEN-CPP, в контакт с клеткой (такой как микробная клетка) можно осуществлять при температуре, которая обеспечивает возможность проникновения комплекса в клетку. Такое приведение в контакт можно осуществлять, например, при любой температуре от приблизительно 4 до 45°C. В неограничивающих вариантах осуществления температура для приведения в контакт может составлять приблизительно 4, 15, 20, 30, 37 или 42°C. Такую же температуру или температурный диапазон можно поддерживать при осуществлении стадии приведения в контакт или модифицировать ее соответствующим образом (например, использовать две или более разные тем-

пературы).

Приведение композиции, содержащей комплекс белок RGEN-CPP, в контакт с клеткой можно осуществлять в течение периода времени, который является достаточным, чтобы обеспечить возможность проникновения комплекса в клетку. Например, клетки можно инкубировать с комплексом белок RGEN-CPP в течение по меньшей мере приблизительно 15, 30, 45, 60, 75, 90, 105, 120, 135, 150, 165, 180, 240, 300, 360, 420, 480, 540, 600, 660, или 720 мин.

Условия среды (например, буфер, вода и концентрация солей, pH, чистота комплекса белок RGEN-CPP), при которых осуществляют приведение в контакт, могут соответствовать любому из условий, раскрытых выше в отношении композиции, содержащей комплекс белок RGEN-CPP. Например, клетки можно инкубировать с комплексом в буфере HEPES (например, ~25 mM HEPES, таком как 25 mM HEPES/КОН, pH 7,5, 200 mM KCl, 20% глицерина, 1 mM DTT) или PBS (например, 1X PBS, pH 7).

Одну или несколько клеток (таких как микробные клетки) можно приводить в контакт с композицией, содержащей комплекс белок RGEN-CPP. Клетка согласно данному документу может представлять собой клетку, существующую (i) в организме/ткани *in vivo*, (ii) в ткани или группе *ex vivo* или (iii) в состоянии *in vitro* (например, культивируемые клетки).

Проникновением комплекса белок RGEN-CPP в клетку согласно данному документу, как правило, называют состояние, когда комплекс полностью пересек (i) клеточную мембрану или (ii) клеточную стенку и клеточную мембрану и содержится в пределах по меньшей мере цитоплазмы клетки. Без желания придерживаться какой-либо конкретной теории или механизма, подразумевается, что комплекс белок RGEN-CPP, удерживаемый вместе с помощью нековалентной связи, либо сохраняется в виде полного или частичного комплекса, либо белковый компонент RGEN отделяется от компонента(ов) CPP комплекса, после того как комплекс белок RGEN-CPP добивается проникновения в клетку. В любом случае белковый компонент RGEN способен ассоциироваться с подходящим РНК-компонентом согласно данному документу; причем такая ассоциация может происходить, например, в цитоплазме, ядре или митохондриях. Данная способность аналогичным образом применима к комплексу белок RGEN-CPP, удерживаемому вместе ковалентной связью.

В определенных вариантах осуществления способа доставки белка RGEN композиция согласно данному документу дополнительно содержит по меньшей мере один РНК-компонент, который ассоциирован с белковым компонентом RGEN комплекса белок RGEN-CPP (т. е. композиция содержит комплекс RGEN-CPP). РНК-компонент в данном варианте осуществления может быть таким, как раскрыт в данном документе, содержащим последовательность, комплементарную последовательности целевого сайта на хромосоме или эписоме в микробной клетке. RGEN может связываться с последовательностью целевого сайта и необязательно расщеплять одну или обе нити ДНК в последовательности целевого сайта. Такой вариант осуществления также можно охарактеризовать как способ доставки комплекса RGEN-CPP в микробную клетку или альтернативно как способ доставки РНК в микробную клетку.

РНК-компонент (например, gRNA) для применения в данном варианте осуществления можно получить с применением любого числа способов, известных из уровня техники. Например, для получения РНК-компонента согласно данному документу можно применять процесс *in vitro* транскрипции. В определенных неограничивающих вариантах осуществления для транскрибирования РНК-компонента с подходящей ДНК-конструкции, кодирующей РНК-компонент, можно применять бактериальные РНК-полимеразы (например, T7, T3, SP6). При необходимости, РНК-компонент можно обработать по меньшей мере до приблизительно 70%, 80%, 90% или 95% чистоты в отношении других биомолекул (например, белка, сахаридов, липидов).

Для получения композиции, содержащей РНК-компонент и комплекс белок RGEN-CPP, РНК-компонент можно растворять в композиции, в которой уже растворен комплекс белок RGEN-CPP, или наоборот (или данные компоненты можно растворять одновременно). При смешивании данных элементов вместе можно применять молярное соотношение РНК-компонента и комплекса белок RGEN-CPP, составляющее, например, по меньшей мере приблизительно 0,5:1, 1,0:1, 1,5:1, 2,0:1, 2,5:1, 3,0:1, 3,5:1 или 4,0:1. В определенных аспектах молярное соотношение РНК-компонента и комплекса белок RGEN-CPP может составлять приблизительно 3,0:1 или может находиться в диапазоне от приблизительно 2,5:1 до 3,5:1, от 2,75:1 до 3,25:1 или от 2,9:1 до 3,1:1. В этом и других аспектах концентрация комплекса белок RGEN-CPP, с которым смешивают РНК-компонент, может составлять по меньшей мере приблизительно 0,5, 1,0, 1,5, 2,0, 2,5, 3,0, 3,5, 4,0, 4,5, 5,0, 6,0, 7,0, 8,0, 9,0 или 10,0 μM или от приблизительно 0,5 до 5,0 μM , от 0,5 до 2,5 μM , от 1,0 до 5,0 μM , от 1,0 до 2,5 μM или от 2,5 до 5,0 μM . Количество времени, требуемое для ассоциации РНК с комплексом белок RGEN-CPP с образованием комплекса RGEN-CPP, может составлять, например, по меньшей мере приблизительно 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 45 или 60 мин.

Другие условия (например, температура, буфер, вода и концентрация солей, pH, чистота комплекса белок RGEN-CPP), при которых РНК-компонент может ассоциировать с комплексом белок RGEN-CPP, могут представлять собой любое из условий, раскрытых выше в отношении (i) композиции, содержащей комплекс белок RGEN-CPP, или (ii) приведения комплекса белок RGEN-CPP в контакт с клеткой. На-

пример, РНК-компонент, такой как gRNA, можно приводить в контакт с комплексом белок RGEN-CPP в буфере HEPES (например, ~25 mM HEPES, таком как 25 mM HEPES/КОН, pH 7,5, 200 mM KCl, 20% глицерина, 1 mM DTT) или PBS (например, 1X PBS, pH 7) при комнатной температуре (например, приблизительно 20-25°C) в течение приблизительно 15 мин. В тех вариантах осуществления, в которых комплекс белок RGEN-CPP удерживается вместе нековалентной связью, ассоциация РНК-компонента с белком RGEN может предусматривать добавление РНК-компонента до, одновременно или после инкубирования CPP с белковым компонентом RGEN.

После ассоциации РНК-компонента с комплексом белок RGEN-CPP полученную композицию, содержащую комплекс RGEN-CPP (например, CPP-Cas9/gRNA), можно, например, сразу же приводить в контакт с клетками. Контакт может осуществляться в условиях среды, при которых, например, происходила ассоциация РНК-компонента и комплекса белок RGEN-CPP (например, см. выше). При необходимости, композицию, содержащую комплекс RGEN-CPP, можно хранить при приблизительно комнатной температуре, при 4°C, или замораживать (например, при -20 или -80°C) для дальнейшего использования. Стабильность комплекса RGEN-CPP и/или его способность проникать в клетки и осуществлять нацеливание на ДНК могут оставаться неизменными или могут характеризоваться по меньшей мере приблизительно 50%, 60%, 70%, 80%, 90% или 95% каждой соответствующей активности, даже если комплекс находится в композиции, которую подвергали одному, двум или более циклам замораживания-оттаивания.

Для приведения в контакт с клеткой композиция, содержащая комплекс белок RGEN-CPP или комплекс RGEN-CPP, необязательно может содержать один или несколько средств для вытеснения объема, которые, предположительно, увеличивают число точек соприкосновения между клеткой и комплексами. Примеры подходящих средств для вытеснения объема согласно данному документу включают глицерин и полиэтиленгликоль (PEG). Другие примеры включают анионный полимер, такой как полиакрилат, полиметилакрилат или анионные полисахаридные полимеры (например, сульфат декстрана). Еще одни примеры средств для вытеснения объема раскрыты в патенте США № 4886741, который включен в данный документ посредством ссылки.

В определенных вариантах осуществления способа доставки белка RGEN клетка (такая как микробная клетка) содержит РНК-компонент, который ассоциирует с белковым компонентом RGEN комплекса белок RGEN-CPP после того, как комплекс белок RGEN-CPP попадает в клетку (т. е. с образованием тем самым комплекса RGEN-CPP в клетке). В данном варианте осуществления РНК-компонент может быть таким, как раскрыт в данном документе, содержащим последовательность, комплементарную последовательности целевого сайта на хромосоме или эписоме в клетке. RGEN может связываться с последовательностью целевого сайта и необязательно расщеплять одну или обе нити ДНК в последовательности целевого сайта.

Один или несколько РНК-компонентов согласно данному документу могут стабильно или временно экспрессироваться в клетке (такой как микробная клетка), например, в которую вводят комплекс белок RGEN-CPP. В качестве примеров временной экспрессии комплекс белок RGEN-CPP может быть (i) доставлен в клетку, которая ранее была модифицирована для временной экспрессии РНК-компонента, (ii) доставлен в клетку совместно с РНК-компонентом или (iii) доставлен в клетку, после чего клетку модифицируют для временной экспрессии РНК-компонента.

Для стабильной и/или временной экспрессии РНК-компонента согласно данному документу, как правило, можно применять полинуклеотидную последовательность ДНК, содержащую (i) промотор, функционально связанный с (ii) нуклеотидной последовательностью, кодирующей РНК-компонент. Такая полинуклеотидная последовательность, например, может входить в состав плазмиды, искусственной хромосомы дрожжей (YAC), космиды, фагмиды, бактериальной искусственной хромосомы (BAC), вируса или линейной ДНК (например, линейный продукт ПЦР), или любого другого типа вектора или конструкции, пригодных для переноса полинуклеотидной последовательности в клетку. Данная полинуклеотидная последовательность может быть способна существовать в клетке временно (т. е. не интегрироваться в геном) или стабильно (т. е. интегрироваться в геном). Также, данная полинуклеотидная последовательность может содержать или не содержать одну или несколько подходящих маркерных последовательностей (например, селективный или фенотипический маркер).

Подходящий промотор, входящий в состав полинуклеотидной последовательности для экспрессии РНК-компонента согласно данному документу, например, может быть конститутивным или индуцируемым. В определенных аспектах промотор может предусматривать сильный промотор, представляющий собой промотор, который может приводить к относительно большому количеству продуктивных инициации транскрипции за единицу времени, и/или он представляет собой промотор, приводящий к более высокому уровню транскрипции гена, чем средний уровень транскрипции генов в клетке, содержащей сильный промотор.

Примеры сильных промоторов, применимых в определенных аспектах данного документа (например, в случае грибных и/или дрожжевых клеток), включают раскрытые в публикациях заявки на патент США №№ 2012/0252079 (DGAT2), 2012/0252093 (EL1), 2013/0089910 (ALK2), 2013/0089911 (SPS19), 2006/0019297 (GPD и GPM), 2011/0059496 (GPD и GPM), 2005/0130280 (FBA, FBAIN, FBAINm),

2006/0057690 (GPAT) и 2010/0068789 (YAT1), которые включены в данный документ посредством ссылки. Другие примеры сильных промоторов включают перечисленные в табл. 2, которые также могут быть применимыми, например, в случае грибных и/или дрожжевых клеток.

Таблица 2. Сильные промоторы

Название промотора	Нативный ген	Ссылка ^a
XPR2	внеклеточная щелочная протеаза	Патент США № 4937189; EP220864
TEF	фактор элонгации трансляции EF1- α (<i>tef</i>)	Патент США № 6265185
GPD, GPM	глицеральдегид-3-фосфат-дегидрогеназа (<i>gpd</i>), фосфоглицератмутаза (<i>gpm</i>)	Патенты США №№ 7259255 и 7459546
GPDIN	глицеральдегид-3-фосфат-дегидрогеназа (<i>gpd</i>)	Патент США № 7459546
GPM/FBAIN	химерная фосфоглицератмутаза (<i>gpm</i>)/фруктозо-бисфосфат-альдолаза (<i>fbal</i>)	Патент США № 7202356
FBA, FBAIN, FBAINm	фруктозо-бисфосфат-альдолаза (<i>fbal</i>)	Патент США № 7202356
GPAT	глицерин-3-фосфат O-ацилтрансфераза (<i>gpac</i>)	Патент США № 7264949
YAT1	фермент-переносчик аммония (<i>yat1</i>)	Публикация заявки на патент США № 2006/0094102
EXP1	экспортный белок	Патент США № 7932077

^a Каждая ссылка в данной таблице включена в данный документ посредством ссылки.

Другие примеры сильных промоторов, применимых в определенных вариантах осуществления данного документа, включают промоторы PGK1, ADH1, TDH3, TEF1, PHO5, LEU2 и GAL1, а также сильные промоторы дрожжей, раскрытые в Velculescu et al. (Cell 88:243-251), который включен в данный документ посредством ссылки.

Промотор для стабильной и/или временной экспрессии РНК-компонента согласно данному документу может представлять собой, например, промотор РНК-полимеразы II (Pol II). Полагают, что все перечисленные выше сильные промоторы являются примерами подходящих промоторов Pol II. Транскрипция за счет промотора Pol II может предусматривать, например, образование комплекса РНК-полимеразы II по меньшей мере из приблизительно 12 белков (например, белков RPB1-RPN12). РНК, транскрибируемая за счет промотора Pol II согласно данному документу, как правило, является 5'-кэпированной (например, содержит группу m⁷G на 5'-конце) и/или содержит, например, полиаденилатный (поли(A)) хвост. При необходимости при экспрессии РНК-компонента за счет промотора Pol II, можно принять способы для удаления 5'-кэпа и/или поли(A)-хвоста из РНК-компонента. Подходящие способы эффективного удаления 5'-кэпа и/или поли(A)-хвоста из транскрибированного за счет Pol II РНК-компонента согласно данному документу включают, например, соответственное применение одного или нескольких рибозимов (см. ниже), самосплайсирующихся интронов группы 1 и самосплайсирующихся интронов группы 2.

В качестве альтернативы промотор для стабильной и/или временной экспрессии РНК-компонента согласно данному документу может представлять собой, например, промотор РНК-полимеразы III (Pol III). Такой промотор, как правило, обеспечивает экспрессию РНК-компонента с определенными 5'- и 3'-концами, поскольку инициацию и терминацию транскрипции с помощью РНК-полимеразы III можно контролировать. Примеры промоторов Pol III, применимых в данном документе, включают промоторы U6 и H1. Другие подходящие промоторы Pol III раскрыты, например, в публикации заявки на патент США № 2010/0160416, которая включена в данный документ посредством ссылки.

Одну или несколько последовательностей рибозимов можно применять для создания определенных 5'- и/или 3'-концов транскрипта, таких как в вариантах осуществления, в которых для экспрессии РНК-компонента в клетке применяют промотор Pol II. Например, нуклеотидная последовательность согласно данному документу, кодирующая РНК-компонент, может дополнительно кодировать рибозим, который расположен выше последовательности, кодирующей РНК-компонент. Таким образом, в определенных вариантах осуществления клетка дополнительно содержит полинуклеотидную последовательность ДНК, содержащую (i) промотор, функционально связанный с (ii) нуклеотидной последовательностью, кодирующей, в направлении от 5' к 3', рибозим и РНК-компонент. Транскрипт, экспрессируемый с такой полинуклеотидной последовательности, автокаталитически удаляет последовательность рибозима с получением РНК с определенными 5'-концом (без 5'-кэпа), но которая содержит последовательность РНК-компонента. Эта "подвергаемая автоматическому процессингу" РНК может предусматривать, например,

сгRNA или gRNA и может образовывать комплекс с белковым компонентом RGEN, таким как Cas9, с образованием тем самым RGEN.

Рибозим согласно данному документу может представлять собой, например, рибозим типа hammerhead (HH), рибозим дельта-вируса гепатита (HDV), рибозим интрона группы I, рибозим РНКазы Р или содержащий шпильку рибозим. Другие неограничивающие примеры рибозимов согласно данному документу включают рибозимы сателлитной РНК Варкуда (VS), глюкозамин-6-фосфат-активируемые рибозимы (glmS) и рибозимы СРЕВ3. В Lilley (Biochem. Soc. Trans. 39:641-646) раскрыта информация, относящаяся к структуре и активности рибозимов. Примеры рибозимов, которые должны подойти для применения в данном документе, включают рибозимы, раскрытые в EP0707638 и патентах США №№ 6063566, 5580967, 5616459, и 5688670,, которые включены в данный документ посредством ссылки. Дополнительная информация, касающаяся применения рибозимов для экспрессии РНК-компонентов с определенными 5'- и/или 3'- концами, раскрыта в публикации заявки на патент США № 62/036652 (поданной 13 августа 2014 года).

В определенных вариантах осуществления полинуклеотид ДНК, содержащий кассету для экспрессии РНК-компонента, содержит подходящую последовательность терминации транскрипции ниже последовательности РНК-компонента. Примеры последовательностей терминации транскрипции, применимых в данном документе, раскрыты в публикации заявки на патент США № 2014/0186906, которая включена в данный документ посредством ссылки. Такие варианты осуществления, как правило, содержат 1, 2, 3, 4, 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 или более остатков, расположенных после конца последовательности РНК-компонента, в зависимости от выбора последовательности терминатора. Все эти дополнительные остатки могут представлять собой остатки U или, например, по меньшей мере 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% или 99% могут представлять собой остатки U, в зависимости от выбора последовательности терминатора. В качестве альтернативы последовательность рибозима (например, рибозима типа hammerhead или рибозима HDV) может быть расположена в направлении 3' (например, на., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 или более нуклеотидов ниже) от последовательности РНК-компонента. 3'-последовательность рибозима может быть расположена соответствующим образом, так что она отщепляет себя от последовательности РНК-компонента; при этом такое расщепление будет приводить к концу транскрипта, например, точно на конце последовательности РНК-компонента или имеющего 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 или более остатков, расположенных после конца последовательности РНК-компонента.

В других примерах РНК-компонент может обеспечиваться в ядре и/или цитоплазме клетки, в которую доставляется комплекс белок RGEN-CPP. Например, можно предполагать, что РНК-компонент, экспрессируемый за счет промотора Pol II без применения 5'-расположенной последовательности рибозима, присутствует как в ядре, так и в цитоплазме. В других вариантах осуществления можно предполагать, что РНК-компонент, экспрессируемый за счет любого типа промотора (например, промотора Pol II или III) и с применением 5'-расположенной последовательности рибозима, присутствует преимущественно в ядре. В определенных аспектах можно предполагать, что РНК-компонент, экспрессируемый за счет промотора Pol III, присутствует преимущественно в ядре. В определенных аспектах РНК-компонент является неэкспортируемым (например, вследствие экспрессии за счет промотора Pol III и/или в результате автопроцессинга рибозима) и, как правило, расположен в ядре, в то время как в других аспектах он является экспортируемым и расположен в ядре и цитоплазме. В целом, после доставки в клетку белковый компонент RGEN комплекса белок RGEN-CPP может ассоциировать с РНК-компонентом (с образованием тем самым RGEN) в цитоплазме и/или ядре (в зависимости от расположения РНК-компонента). Такая ассоциация в ядре обычно обусловлена способностью белкового компонента RGEN согласно данному документу локализоваться в ядре под управлением NLS.

RGEN согласно данному документу применима для опосредуемого RGEN нацеливания на ДНК. Любой из приведенных выше вариантов осуществления, касающихся доставки белкового компонента RGEN в клетку, можно применять для способа нацеливания на ДНК. Например, комплекс белок RGEN-CPP можно приводить в контакт по меньшей мере с одним РНК-компонентом за пределами микробной клетки с образованием комплекса RGEN-CPP, чтобы доставить его в клетку для нацеливания на ДНК в ней. В качестве другого примера комплекс белок RGEN-CPP, после его доставки в микробную клетку, может контактировать по меньшей мере с одним РНК-компонентом внутри микробной клетки с образованием там комплекса RGEN-CPP, который затем может опосредовать нацеливание на ДНК. В следующем раскрытии, касающемся способов нацеливания, упоминается "RGEN", и не упоминается "комплекс RGEN-CPP". Должно быть понятно, что в зависимости от того, применяется ли в способе доставки RGEN согласно данному документу комплекс RGEN-CPP с ковалентной или нековалентной связью (и в зависимости от того, насколько сильная нековалентная связь присутствует в вариантах осуществления с применением комплекса RGEN-CPP с нековалентной связью), ссылка на RGEN ниже ковалентный или нековалентный относится и к такому комплексу RGEN-CPP.

RGEN согласно данному документу, которая может расщеплять одну или обе нити ДНК целевой последовательности ДНК, можно применять, например, в способе нацеливания на ДНК. Такие способы нацеливания на ДНК могут охватывать HR-опосредованное нацеливание на ДНК, если подходящая до-

норная ДНК предусмотрена в способе. Таким образом, в определенных вариантах осуществления микробная клетка в способе нацеливания согласно данному документу может содержать донорный полинуклеотид, содержащий по меньшей мере одну последовательность, гомологичную последовательности в последовательности целевого сайта (последовательности, на которую специфически нацеливается RGEN согласно данному документу) или вблизи нее. Такие варианты осуществления необязательно могут характеризоваться тем, что способ нацеливания дополнительно включает стадию обеспечения подходящего донорного полинуклеотида для микробной клетки.

Донорный полинуклеотид согласно данному документу может подвергаться HR с последовательностью в целевом сайте ДНК или вблизи него, если целевой сайт содержит SSB или DSB (какие можно вводить с применением RGEN согласно данному документу). "Гомологичная последовательность" в пределах донорного полинуклеотида согласно данному документу, например, может содержать или состоять из последовательности по меньшей мере из приблизительно 25, 50, 75, 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000 или 10000 нуклеотидов или приблизительно 50-500, 50-550, 50-600, 50-650 или 50-700 нуклеотидов, которая на 100% идентична последовательности в целевом сайте или вблизи него или, например, по меньшей мере приблизительно 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% или 99% идентична последовательности в целевом сайте или вблизи него.

Донорный полинуклеотид согласно данному документу может иметь две гомологичные последовательности ("плечи гомологии"), например, разделенные последовательностью, которая является гетерологичной относительно последовательности в целевом сайте или вблизи него. HR между таким донорным полинуклеотидом и целевым сайтом, как правило, приводит к замещению последовательности в целевом сайте на гетерологичную последовательность донорного полинуклеотида (т.е. последовательность целевого сайта, расположенная между последовательностями целевого сайта, гомологичными "плечам гомологии" донорного полинуклеотида, замещается гетерологичной последовательностью донорного полинуклеотида). В донорном полинуклеотиде с двумя "плечами гомологии" данные "плечи" могут быть разделены, например, по меньшей мере приблизительно 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 75, 100, 250, 500, 1000, 2500, 5000, 10000, 15000, 20000, 25000 или 30000 нуклеотидами (т.е. длина гетерологичной последовательности в донорном полинуклеотиде может составлять по меньшей мере приблизительно 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 75, 100, 250, 500, 1000, 2500, 5000, 10000, 15000, 20000, 25000 или 30000 нуклеотидов). Длина (например, любая из длин, раскрытых выше для гомологичной последовательности) каждого "плеча гомологии" может быть одинаковой или разной. Процент идентичности (например, любой из % идентичности, раскрытых выше для гомологичной последовательности) каждого плеча с соответствующими гомологичными последовательностями в целевом сайте или вблизи него может быть одинаковым или разным.

Последовательность ДНК, расположенная в последовательности целевого сайта или вблизи нее (в качестве альтернативы поблизости или в непосредственной близости от), которая гомологична соответствующей гомологичной последовательности в донорном полинуклеотиде, может находиться, например, в пределах приблизительно 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 450, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 20000, 30000, 40000, 50000 или 60000 (или любое целое число от 1 до 60000) нуклеотидов (например, приблизительно 1-1000, 100-1000, 500-1000, 1-500 или 100-500 нуклеотидов) от предсказанного сайта разреза RGEN (DSB или одностороннего разрыва) в целевой последовательности. Данные расстояния в нуклеотидах могут быть отсчитаны от сайта разреза до первого нуклеотида гомологичной последовательности в направлении либо выше, либо ниже сайта разреза. Например, последовательность вблизи целевой последовательности, которая гомологична соответствующей последовательности в донорном полинуклеотиде, может начинаться на 500 пар нуклеотидных оснований ниже предсказанного сайта разреза RGEN в целевой последовательности. В вариантах осуществления данного документа, в которых используется донорный полинуклеотид с двумя "плечами гомологии" (например, первое и второе "плечи гомологии" разделены гетерологичной последовательностью), например, гомологичная последовательность (соответствующая в отношении гомологии первому "плечу гомологии" донора) может находиться выше предсказанного сайта разреза RGEN, а гомологичная последовательность (соответствующая в отношении гомологии второму "плечу гомологии" донора) может находиться ниже предсказанного сайта разреза RGEN. Расстояния в нуклеотидах для каждой из данных гомологичных последовательностей, расположенных выше и ниже предсказанного сайта разреза, могут быть одинаковыми или разными и, например, могут представлять собой любое из расстояний в нуклеотидах, раскрытых выше. Например, 3'-конец гомологичной последовательности (соответствующий в отношении гомологии первому "плечу гомологии" донора) может быть расположен на 600 пар нуклеотидных оснований выше предсказанного сайта разреза RGEN, а 5'-конец гомологичной последовательности (соответствующий в отношении гомологии второму "плечу гомологии" донора) может быть расположен на 400 пар нуклеотидных оснований ниже предсказанного сайта разреза RGEN.

В различных аспектах донорный полинуклеотид может доставляться в клетку (такую как микробная клетка) в тот самый момент времени или приблизительно тогда же (например, в пределах 1, 2, 3 или

более часов), когда в клетку доставляется комплекс белок RGEN-CPP. Такая доставка может осуществляться посредством любых известных из уровня техники способов, подходящих для конкретного типа применяемых клеток. Данные методики включают, например, трансформацию (например, трансформацию, опосредованную ацетатом лития [Methods in Enzymology, 194:186-187]), трансфекцию, биобаллистическое воздействие, электропорацию и микроинъекцию. В качестве примеров в патентах США №№ 4880741 и 5071764 и Chen et al. (Appl. Microbiol. Biotechnol. 48:232-235), которые включены в данный документ посредством ссылки, описаны методики переноса ДНК в случае *Y. lipolytica*. Примеры способов доставки, применимых у растений, включают *Agrobacterium*-опосредованную трансформацию и бомбардировку частицами в ходе биобаллистической трансформации.

RGEN, которая расщепляет одну или обе нити ДНК целевой последовательности ДНК, можно применять для создания вставки/делеции в других неограничивающих вариантах осуществления нацеливания на ДНК согласно данному документу. Способ образования вставки/делеции в клетке можно осуществлять, как раскрыто выше для HR-опосредуемого нацеливания, но без дополнительного обеспечения полинуклеотида донорной ДНК, который мог бы подвергаться HR в целевом сайте ДНК или вблизи него (т. е. при данном способе индуцируется NHEJ). Примеры вставок/делеций, которые можно создавать, раскрыты в данном документе. Размер вставки/делеции может составлять, например, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 или более оснований. В определенных вариантах осуществления вставка/делеция может быть даже больше, такой как по меньшей мере приблизительно 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, или 150 оснований. В еще одних вариантах осуществления вставка или делеция могут составлять по меньшей мере приблизительно 500, 750, 1000, или 1500 оснований. В определенных вариантах осуществления при попытке создать вставку/делецию вместо этого в последовательности целевого сайта может быть создана однонуклеотидная замена. Таким образом, способ нацеливания согласно данному документу можно выполнять, например, с целью создания однонуклеотидной замены.

В определенных вариантах осуществления целью способа нацеливания согласно данному документу является образование вставки/делеции, при этом частота образования вставки/делеции у нетрадиционных видов дрожжей (например, *Y. lipolytica*) значительно выше, чем частота, которая наблюдалась бы при применении такой же или подобной стратегии нацеливания у традиционных видов дрожжей, таких как *S. cerevisiae*. Например, в то время как частота образования вставки/делеции у традиционных дрожжей может составлять от приблизительно 0,0001 до 0.001 (DiCarlo et al., Nucleic Acids Res. 41:4336-4343), частота у нетрадиционных видов дрожжей согласно данному документу может составлять по меньшей мере приблизительно 0,05, 0,10, 0,15, 0,20, 0,25, 0,30, 0,35, 0,40, 0,45, 0,50, 0,55, 0,60, 0,65, 0,70, 0,75 или 0,80. Таким образом, частота образования вставки/делеции у нетрадиционных видов дрожжей согласно данному документу может, например, по меньшей мере в приблизительно 50, 100, 250, 500, 750, 1000, 2000, 4000 или 8000 раз превышать частоту, которая наблюдалась бы при применении такой же или подобной RGEN-опосредуемой стратегии нацеливания у традиционных видов дрожжей.

В определенных вариантах осуществления способ нацеливания можно проводить для разрушения одной или нескольких полинуклеотидных последовательностей ДНК, кодирующих белок или некодирующую РНК. Примером такой последовательности, которая может подвергаться нацеливанию для разрушения, является последовательность, кодирующая маркер (т. е. маркерный ген). Неограничивающие примеры маркеров согласно данному документу включают поддающиеся скринингу маркеры и селективируемые маркеры. Поддающийся скринингу маркер согласно данному документу может представлять собой маркер, который делает клетку визуально отличимой при соответствующих условиях. Примеры поддающихся скринингу маркеров включают полинуклеотиды, кодирующие бета-глюкуронидазу (GUS), бета-галактозидазу (lacZ) и флуоресцентные белки (например, GFP, RFP, YFP, BFP). Селектируемый маркер согласно данному документу может представлять собой маркер, который делает клетку устойчивой к селективному средству или селективной среде. Примерами селективируемых маркеров являются маркеры ауксотрофности, такие как HIS3, LEU2, TRP1, MET15, или URA3, которые обеспечивают выживание клеток, таких как дрожжевые клетки, в отсутствие экзогенно обеспечиваемых гистидина, лейцина, триптофана, метионина или урацила соответственно. Другими примерами селективируемых маркеров являются маркеры устойчивости к антибиотикам или противогрибковым средствам, такие как маркеры, которые делают клетку устойчивой к ампициллину, хлорамфениколу, гигромицину В, ноурсетрицину, флеомицину, пуромицину или неомицину (например, G418). Примеры данных способов можно необязательно характеризовать как способы повторного использования маркеров.

По меньшей мере одной целью разрушения маркера в определенных вариантах осуществления может быть повторное использование маркера. Повторное использование маркера представляет собой способ, например, включающий (i) трансформацию клетки с помощью маркера и гетерологичной последовательности ДНК, (ii) отбор трансформированной клетки, содержащей маркер и гетерологичной последовательности ДНК (где клетка, которая селективируется по маркеру, как правило, с более высокой вероятностью содержит гетерологичную последовательность ДНК), (iii) разрушение маркера и затем повторение стадий (i)-(iii) столько раз, сколько необходимо (с применением такого же [или другого] маркера, но с применением в каждом цикле другой гетерологичной последовательности ДНК), для трансформации клеток с помощью нескольких гетерологичных последовательностей ДНК. Одна или несколько гете-

рологических последовательностей в данном способе могут включать маркер сам по себе в форме донорного полинуклеотида (например, маркер, фланкированный "плечами гомологии", для нацеливания на конкретный локус). Примеры способов повторного использования маркера согласно данному документу включают способы с применением URA3 в качестве маркера, такие как определенные способы с использованием дрожжей (например, нетрадиционных видов дрожжей, таких как *Y. lipolytica*).

RGEN согласно данному документу, которая может связываться с последовательностью целевого сайта ДНК, но не расщепляет ни одну нить в последовательности целевого сайта, можно применять в способе нацеливания на ДНК в других вариантах осуществления. Любая RGEN, раскрытая в данном документе, которая содержит только дисфункциональные нуклеазные домены, но сохраняет специфическую активность связывания ДНК, можно применять в данном типе способа нацеливания.

В определенных вариантах осуществления нацеливания на ДНК с применением RGEN, не имеющей функциональных нуклеазных доменов, RGEN может связываться с целевым сайтом и модулировать транскрипцию полинуклеотидной последовательности (т. е. транскрипцию гена). Как правило, RGEN нацеливается на регуляторную последовательность, такую как промотор (например, в пределах 1-1000, 1-500, 1-250, 1-125 или 1-50 оснований выше сайта инициации транскрипции), последовательность, кодирующую 5'-нетранслируемую последовательность РНК, или интрон (например, первый интрон), для осуществления модуляции транскрипции полинуклеотидной последовательности.

В качестве неограничивающего примера RGEN, связанную или слитую с фактором-репрессором транскрипции или его репрессорным доменом, можно применять для репрессии или сайленсинга экспрессии одной или нескольких полинуклеотидных последовательностей. В определенных альтернативных вариантах осуществления RGEN может самостоятельно (без репрессора или его домена) подавлять экспрессию гена; при этом такая RGEN может нацеливаться таким образом, что она подавляет связывание и/или перемещение транскрипции аппарата РНК, необходимого для транскрипции. Способ, предусматривающий любую репрессирующую RGEN, необязательно может характеризоваться как сайленсинг гена или способ сайленсинга транскрипции. Уровень понижающей регуляции транскрипции в способе сайленсинга может составлять приблизительно 100% (полный сайленсинг гена) или по меньшей мере приблизительно 30% (умеренный сайленсинг гена), 40%, 50%, 60%, 70%, 80%, 90% или 95% (значительный сайленсинг гена), например, по сравнению с уровнем транскрипции до применения репрессирующей RGEN.

RGEN, связанную или слитую с фактором-активатором транскрипции или его активаторным доменом, можно применять для повышающей регуляции экспрессии одной или нескольких полинуклеотидных последовательностей. Способ, предусматривающий такую активирующую RGEN, можно необязательно охарактеризовать как способ повышающей регуляции или активации транскрипции. Уровень повышающей регуляции транскрипции в таком способе может составлять, например, по меньшей мере приблизительно 25%, 50%, 75%, 100%, 250%, 500% или 1000%, по сравнению с уровнем транскрипции до применения активирующей RGEN.

В определенных вариантах осуществления RGEN, которая может связываться с последовательностью ДНК целевого сайта, но предпочтительно не расщепляет ни одну нить последовательности целевого сайта, можно применять как инструмент для диагностики (например, зонд для обнаружения последовательности ДНК). Белковый компонент RGEN в ДНК-зонде может быть связан со средством-репортером, таким как, например, репортерный белок (например, флуоресцентный белок, такой как GFP). Специфическое связывание ДНК с помощью RGEN-репортерного белка, определяемое РНК-компонентом RGEN, может быть предусмотрено в системе детекции, при этом с успехом используется активность средства-репортера. Проточная цитометрия (например, сортировка клеток с активированной флуоресценцией [FACS]) и флуоресцентная гибридизация *in situ* (FISH) представляют собой примеры подходящих систем детекции согласно данному документу, в которых применяют флуоресцентный репортер.

Способ нацеливания согласно данному документу можно осуществлять таким образом, что в способе подвергаются нацеливанию, например, два или более целевых сайтов ДНК. Такой способ необязательно может быть охарактеризован как мультиплексный способ. В определенных вариантах осуществления нацеливанию в одно и то же время могут подвергаться два, три, четыре, пять, шесть, семь, восемь, девять, десять или более целевых сайтов. Мультиплексный способ, как правило, выполняют с помощью способа нацеливания согласно данному документу, в котором обеспечено множество разных РНК-компонентов, при этом каждый сконструирован для направления RGEN на уникальный целевой сайт ДНК. Например, два или более разных РНК-компонентов можно применять для получения смеси комплексов RGEN-CPP *in vitro* (например, после проведения раскрытой в данном документе процедуры ассоциации РНК-компонента с комплексом белок RGEN-CPP), затем данную смесь приводят в контакт с клеткой.

Другой аспект мультиплексного нацеливания согласно данному документу может предусматривать обеспечение двух или более разных РНК-компонентов в клетке, которые ассоциируются с белковыми компонентами RGEN комплексов белок RGEN-CPP, которые проникли в клетку. Такой способ может включать, например, обеспечение клетки (i) отдельными полинуклеотидами ДНК, каждый из которых экспрессирует в ней конкретный РНК-компонент, и/или (ii) по меньшей мере одним полинуклеотидом

ДНК, кодирующим два или более РНК-компонентов (например, см. ниже раскрытие, касающееся тандемных кассет рибозим-РНК-компонент).

При помощи мультиплексного способа необязательно можно производить нацеливание на сайты ДНК, расположенные очень близко в пределах одной последовательности (например, промотор или открытая рамка считывания), и/или сайты, которые удалены друг от друга (например, в разных генах и/или хромосомах). Мультиплексный способ в других вариантах осуществления можно осуществлять с подходящими полинуклеотидами донорной ДНК (в случае HR) или без них (в случае NHEJ, приводящему к вставке/делеции и/или замене основания), в зависимости от требуемого результата нацеливания (если применяют RGEN, обладающую эндонуклеазной или никазой активностью). В еще одних вариантах осуществления мультиплексный способ можно осуществлять с помощью репрессирующей или активирующей RGEN, раскрытых в данном документе. Например, могут обеспечиваться множество репрессирующих RGEN, которые оказывают понижающую регуляцию на набор генов, таких как гены, вовлеченные в конкретный метаболический путь.

В определенных вариантах осуществления мультиплексный способ может включать обеспечение клетки полинуклеотидом ДНК, содержащим (i) промотор, функционально связанный с (ii) последовательностью, содержащей более одной кассеты рибозим-РНК-компонент (т. е. тандемные кассеты). Транскрипт, экспрессирующийся с такого полинуклеотида ДНК, может содержать, например, 2, 3, 4, 5, 6, 7, 8, 9, 10, или более кассет. После всех или некоторых последовательностей РНК-компонентов необязательно может быть включена 3'-последовательность рибозима, чтобы обеспечить возможность расщепления и отделения РНК-компонента от расположенной ниже последовательности транскрипта (т.е. тандемные кассеты могут содержать одну или несколько кассет рибозим-РНК-компонент-рибозим). Полинуклеотид ДНК согласно данному документу для экспрессии тандемных кассет рибозим-РНК-компонент-рибозим может быть сконструирован таким образом, что между каждой кассетой расположено приблизительно 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 или более нуклеотидов (например, некодирующая спейсерная последовательность).

Расстояния между каждой кассетой могут быть одинаковыми или разными.

Любую конструкцию или вектор, содержащие полинуклеотид ДНК, кодирующий РНК-компонент, описанный в данном документе, можно вводить в клетку с помощью любых способов, известных из уровня техники, подходящих для конкретного применяемого типа клеток. Например, для доставки донорной ДНК в клетку можно использовать любой из способов, раскрытых выше.

Определенные варианты осуществления данного документа относятся к способу модификации или изменения целевого сайта в геноме микробной клетки, где способ включает приведение микробной клетки в контакт с направляющим полинуклеотидом и эндонуклеазой Cas, ковалентно или нековалентно связанными с CPP, где направляющий полинуклеотид и CPP-эндонуклеаза Cas способны образовывать комплекс, который позволяет эндонуклеазе Cas вводить двухнитевой разрыв в целевой сайт в геноме микробной клетки. Модификация или изменение целевого сайта может включать (i) замещение по меньшей мере одного нуклеотида, (ii) делецию по меньшей мере одного нуклеотида, (iii) вставку по меньшей мере одного нуклеотида или (iv) любую комбинацию (i)-(iii).

Определенные варианты осуществления данного документа относятся к полинуклеотидной последовательности, содержащей нуклеотидную последовательность, кодирующую белок слияния, который содержит белковый компонент направляемой РНК эндонуклеазы (RGEN) и по меньшей мере один проникающий в клетку пептид (CPP). Любой белок слияния, раскрытый в данном документе, например, может кодироваться такой нуклеотидной последовательностью. Нуклеотидная последовательность необязательно может находиться в функциональной связи с промоторной последовательностью. Определенные варианты осуществления предусматривают, например, полинуклеотид (например, вектор или конструкцию), содержащий по меньшей мере одну открытую рамку считывания, кодирующую любое слияние белок RGEN-CPP, раскрытое в данном документе. Такой кодирующий участок необязательно может быть функционально связанным с промоторной последовательностью, подходящей для экспрессии слияния белок RGEN-CPP в клетке (например, клетке бактерий; эукариотической клетке, такой как клетка дрожжей, насекомых или млекопитающих) или, например, в системе *in vitro* экспрессии белка. Примеры вектора или конструкции включают кольцевую (например, плаزمиду) и некольцевую (например, линейную ДНК, такую как амплифицированная последовательность ДНК) полинуклеотидные молекулы.

Определенные варианты осуществления данного документа относятся к способу получения белка слияния белок RGEN-CPP, включающему стадии: обеспечения полинуклеотидной последовательности с нуклеотидной последовательностью, кодирующей белок слияния белок RGEN-CPP, и экспрессию белка слияния белок RGEN-CPP с полинуклеотидной последовательности, в результате чего получают белок слияния белок RGEN-CPP. Стадию экспрессии в таком способе необязательно можно осуществлять в клетке (например, клетке бактерии, такой как *E. coli*; эукариотической клетке, такой как клетка дрожжей [например, *S. cerevisiae*] насекомых или млекопитающих). В качестве альтернативы экспрессию белка слияния белок RGEN-CPP можно осуществлять в системе *in vitro* экспрессии белка (например, в бесклеточных системах экспрессии белка, таких как системы с использованием лизата ретикулоцитов кролика или экстракта из пшеничных зародышей). Кроме того, белок слияния белок RGEN-CPP, полученный на

стадии экспрессии, необязательно можно выделять. Такое выделение можно осуществлять, например, путем получения композиции, характеризующейся любым из раскрытых выше признаков (например, чистотой, рН, буфером и/или содержанием солей).

К неограничивающим примерам композиций и способов, раскрытых в данном документе, относятся следующие.

1. Композиция, содержащая по меньшей мере один белковый компонент направляемой РНК эндонуклеазы (RGEN) и по меньшей мере один проникающий в клетку пептид (CPP), где белковый компонент и CPP ковалентно или нековалентно связаны друг с другом в комплекс белок RGEN-CPP, и где комплекс белок RGEN-CPP может пересекать (i) клеточную мембрану или (ii) клеточную стенку и клеточную мембрану клетки.

2. Композиция согласно варианту осуществления 1, где белковый компонент RGEN ассоциирован по меньшей мере с одним РНК-компонентом, который содержит последовательность, комплементарную последовательности целевого сайта на хромосоме или эписоме в клетке, где RGEN может связываться с последовательностью целевого сайта и необязательно расщеплять одну или обе нити ДНК в последовательности целевого сайта.

3. Композиция согласно варианту осуществления 2, где РНК-компонент содержит направляющую РНК (gRNA), содержащую РНК CRISPR (crRNA), функционально связанную с транскрибирующей РНК CRISPR (tracrRNA).

4. Композиция согласно варианту осуществления 2, где RGEN может расщеплять одну или обе нити ДНК в последовательности целевого сайта.

5. Композиция согласно варианту осуществления 1, где RGEN содержит аминокислотную последовательность CRISPR-ассоциированного (Cas) белка-9 (Cas9).

6. Композиция согласно варианту осуществления 1, где белковый компонент RGEN и CPP связаны ковалентно.

7. Композиция согласно варианту осуществления 1, где белковый компонент RGEN и CPP связаны нековалентно.

8. Композиция согласно варианту осуществления 1, где CPP является катионным или амфипатическим.

9. Композиция согласно варианту осуществления 1, где CPP включает в себя (i) CPP из транскрипционного белка Zebra вируса Эпштейна-Барр, (ii) CPP с 6 или более смежными аргининовыми остатками, (iii) CPP транспортан-10 (TP10) или (iv) CPP из белка кадгерина сосудистого эндотелия.

10. Композиция согласно варианту осуществления 1, где комплекс белок RGEN-CPP может пересекать клеточную стенку и клеточную мембрану клетки.

11. Клетка, содержащая композицию по варианту осуществления 1.

12. Способ доставки белкового компонента направляемой РНК эндонуклеазы (RGEN) в клетку, причем способ включает приведение клетки в контакт с композицией, содержащей белковый компонент направляемой РНК эндонуклеазы (RGEN) и по меньшей мере один проникающий в клетку пептид (CPP), где белковый компонент и CPP ковалентно или нековалентно связаны друг с другом в комплекс белок RGEN-CPP, где комплекс белок RGEN-CPP пересекает (i) клеточную мембрану или (ii) клеточную стенку и клеточную мембрану клетки, вследствие чего он попадает в клетку.

13. Способ согласно варианту осуществления 12, где (i) композиция дополнительно содержит по меньшей мере один РНК-компонент, который ассоциирован с белковым компонентом RGEN; или (ii) клетка содержит РНК-компонент, где РНК-компонент ассоциирует с белковым компонентом RGEN после того, как комплекс белок RGEN-CPP попадает в клетку; где РНК-компонент содержит последовательность, комплементарную последовательности целевого сайта на хромосоме или эписоме в клетке, где RGEN может связываться с последовательностью целевого сайта и необязательно расщеплять одну или обе нити ДНК в последовательности целевого сайта.

14. Способ согласно варианту осуществления 13, где RGEN может расщеплять одну или обе нити ДНК в последовательности целевого сайта.

15. Способ согласно варианту осуществления 14, где клетка дополнительно содержит донорный полинуклеотид, содержащий по меньшей мере одну последовательность, гомологичную последовательности в последовательности целевого сайта или вблизи нее.

16. Способ согласно варианту осуществления 12, где клетка является клеткой организма, не относящегося к млекопитающим.

17. Композиция, содержащая по меньшей мере один белковый компонент из комплекса направляющий полинуклеотид/эндонуклеаза Cas и по меньшей мере один проникающий в клетку пептид (CPP), где белковый компонент и CPP ковалентно или нековалентно связаны друг с другом в комплексе направляющий полинуклеотид/эндонуклеаза Cas-CPP, и где комплекс направляющий полинуклеотид/эндонуклеаза Cas-CPP может пересекать (i) клеточную мембрану или (ii) клеточную стенку и клеточную мембрану клетки, где клетка необязательно представляет собой растительную клетку.

18. Композиция согласно варианту осуществления 17, где эндонуклеаза Cas представляет собой эндонуклеазу Cas9, оптимизированную для растений.

19. Композиция согласно варианту осуществления 17, где направляющий полинуклеотид содержит (i) первый домен нуклеотидной последовательности, который комплементарен нуклеотидной последовательности в целевой ДНК, и (ii) второй домен нуклеотидной последовательности, который взаимодействует с эндонуклеазой Cas, где первый домен нуклеотидной последовательности и второй домен нуклеотидной последовательности составлены дезоксирибонуклеиновыми кислотами (DNA), рибонуклеиновыми кислотами (RNA) или их комбинацией.

20. Композиция согласно варианту осуществления 17, где комплекс направляющий полинуклеотид/эндонуклеаза Cas-CPP может пересекать клеточную стенку растительной клетки.

21. Композиция согласно варианту осуществления 17, где CPP включает в себя (i) CPP из трансактиваторного белка Zebra вируса Эпштейна-Барр, (ii) CPP с 6 или более смежными аргининовыми остатками, (iii) CPP транспортан-10 (TP10), (iv) CPP из белка кадгерина сосудистого эндотелия или (v) CPP, выбранный из группы, состоящей из синтетического CPP с девятью аргининовыми остатками, богатого гистидином CPP с девятью аргининовыми остатками и Ras CPP с девятью аргининовыми остатками.

22. Композиция согласно варианту осуществления 20, где растительная клетка представляет собой клетку однодольного или двудольного растения.

23. Композиция согласно варианту осуществления 22, где однодольное растение выбрано из группы, состоящей из маиса, риса, сорго, ржи, ячменя, пшеницы, проса, разновидностей овса, сахарного тростника, газонной травы и проса прутьевидного.

24. Композиция согласно варианту осуществления 22, где двудольное растение выбрано из группы, состоящей из сои, канолы, люцерны, подсолнечника, хлопчатника, табака, арахиса, картофеля, табака, Arabidopsis и сафлора.

25. Способ модификации целевого сайта в геноме клетки, причем способ включает обеспечение направляющего полинуклеотида, проникающего в клетку пептида (CPP) и эндонуклеазы Cas для клетки, где направляющий полинуклеотид, эндонуклеаза Cas и CPP ковалентно или нековалентно связаны друг с другом в комплекс направляющий полинуклеотид/эндонуклеаза Cas-CPP, и где комплекс направляющий полинуклеотид/эндонуклеаза Cas-CPP может пересекать (i) клеточную мембрану или (ii) клеточную стенку и клеточную мембрану клетки, где клетка необязательно представляет собой растительную клетку.

26. Способ согласно варианту осуществления 25, дополнительно включающий обнаружение по меньшей мере одной растительной клетки, которая имеет модификацию в целевом сайте, где модификация в целевом сайте выбрана из группы, состоящей из (i) замещения по меньшей мере одного нуклеотида, (ii) делеции по меньшей мере одного нуклеотида, (iii) вставки по меньшей мере одного нуклеотида и (iv) любой комбинации (i)-(iii).

27. Способ согласно варианту осуществления 25, где растительная клетка представляет собой клетку однодольного или двудольного растения.

28. Композиция, содержащая по меньшей мере один белковый компонент направляемой РНК эндонуклеазы (RGEN) и по меньшей мере один проникающий в клетку пептид (CPP), где белковый компонент и CPP ковалентно или нековалентно связаны друг с другом в комплекс белок RGEN-CPP, и где комплекс белок RGEN-CPP может пересекать (i) клеточную мембрану или (ii) клеточную стенку и клеточную мембрану микробной клетки.

29. Композиция согласно варианту осуществления 28, где белковый компонент RGEN ассоциирован по меньшей мере с одним РНК-компонентом, который содержит последовательность, комплементарную последовательности целевого сайта на хромосоме или эписоме в микробной клетке, где RGEN может связываться с последовательностью целевого сайта и необязательно расщеплять одну или обе нити ДНК в последовательности целевого сайта.

30. Композиция согласно варианту осуществления 28, где комплекс белок RGEN-CPP может пересекать клеточную стенку и клеточную мембрану микробной клетки.

31. Микробная клетка, содержащая композицию по варианту осуществления 28.

32. Способ доставки белкового компонента направляемой РНК эндонуклеазы (RGEN) в микробную клетку, причем способ включает приведение микробной клетки в контакт с композицией, содержащей белковый компонент направляемой РНК эндонуклеазы (RGEN) и по меньшей мере один проникающий в клетку пептид (CPP), где белковый компонент и CPP ковалентно или нековалентно связаны друг с другом в комплекс белок RGEN-CPP, где комплекс белок RGEN-CPP пересекает (i) клеточную мембрану или (ii) клеточную стенку и клеточную мембрану микробной клетки с попаданием, таким образом, в микробную клетку.

33. Способ согласно варианту осуществления 32, где (i) композиция дополнительно содержит по меньшей мере один РНК-компонент, который ассоциирован с белковым компонентом RGEN; или (ii) микробная клетка содержит РНК-компонент, где РНК-компонент ассоциирует с белковым компонентом RGEN после того, как комплекс белок RGEN-CPP попадает в микробную клетку; где РНК-компонент содержит последовательность, комплементарную последовательности целевого сайта на хромосоме или эписоме в микробной клетке, где RGEN может связываться с последовательностью целевого сайта и необязательно расщеплять одну или обе нити ДНК в последовательности целевого сайта.

34. Способ согласно варианту осуществления 33, где RGEN может расщеплять одну или обе нити ДНК в последовательности целевого сайта.

35. Способ по варианту осуществления 34, где микробная клетка дополнительно содержит донорный полинуклеотид, содержащий по меньшей мере одну последовательность, гомологичную последовательности в последовательности целевого сайта или вблизи нее.

36. Способ согласно варианту осуществления 32, где микробная клетка представляет собой дрожжевую клетку.

37. Композиция, содержащая по меньшей мере один белковый компонент комплекса направляющий полинуклеотид/эндонуклеаза Cas и по меньшей мере один проникающий в клетку пептид (CPP), где белковый компонент и CPP ковалентно или нековалентно связаны друг с другом в комплекс направляющий полинуклеотид/эндонуклеаза Cas-CPP, и где комплекс направляющий полинуклеотид/эндонуклеаза Cas-CPP может пересекать (i) клеточную мембрану или (ii) клеточную стенку и клеточную мембрану микробной клетки.

38. Композиция согласно варианту осуществления 37, где комплекс направляющий полинуклеотид/эндонуклеаза Cas-CPP может пересекать клеточную стенку микробной клетки.

39. Способ модификации целевого сайта в геноме микробной клетки, причем способ включает обеспечение направляющего полинуклеотида, проникающего в клетку пептида (CPP) и эндонуклеазы Cas для микробной клетки, где направляющий полинуклеотид, эндонуклеаза Cas и CPP ковалентно или нековалентно связаны друг с другом в комплекс направляющий полинуклеотид/эндонуклеаза Cas-CPP, и где комплекс направляющий полинуклеотид/эндонуклеаза Cas-CPP может пересекать (i) клеточную мембрану или (ii) клеточную стенку и клеточную мембрану микробной клетки.

40. Способ согласно варианту осуществления 39, дополнительно включающий обнаружение по меньшей мере одной микробной клетки, которая имеет модификацию в целевом сайте, где модификация в целевом сайте выбрана из группы, состоящей из (i) замещения по меньшей мере одного нуклеотида, (ii) делеции по меньшей мере одного нуклеотида, (iii) вставки по меньшей мере одного нуклеотида и (iv) любой комбинации (i)-(iii).

Примеры

Раскрытое изобретение дополнительно определено в следующих примерах. Следует понимать, что эти примеры, хотя и указывают на определенные предпочтительные аспекты настоящего изобретения, приведены только для иллюстрации. Из приведенного выше обсуждения и данных примеров специалист в данной области техники сможет установить существенные характеристики настоящего изобретения, и не отклоняясь от его идеи и объема, сможет осуществить различные изменения и модификации настоящего изобретения для адаптации его к различным областям применения и условиям.

Пример 1. Векторы для экспрессии белка слияния Cas9-CPP (проникающий в клетку пептид) в *E. coli*.

В данном примере векторы, сконструированные для индуцируемой экспрессии белков трансляционного слияния, содержащих белок Cas9 и проникающий в клетку пептид (CPP), получали и тестировали в отношении экспрессии в *E. coli*. Показано, что белки слияния Cas9-CPP, как ожидалось, экспрессировались в *E. coli*, и впоследствии их очищали.

Открытую рамку считывания гена Cas9 из *Streptococcus pyogenes* M1 GAS (SF370) подвергали кодон-оптимизации для экспрессии в *Yagowia* с помощью стандартных методик с получением SEQ ID NO: 1. Последовательность ДНК, кодирующую одинарный сигнал внутриядерной локализации (NLS) вируса обезьян 40 (SV40) плюс короткий линкер (4 аминокислоты), вставляли после последнего смыслового кодона SEQ ID NO: 1 с получением SEQ ID NO: 2. SEQ ID NO: 2 кодирует аминокислотную последовательность, показанную под SEQ ID NO: 3. Последние семь аминокислот SEQ ID NO: 3 кодируют добавленный NLS, в то время как остатки в положениях 1369-1372 SEQ ID NO: 3 кодируют добавленный линкер. Последовательность Cas9-NLS (SEQ ID NO: 2), кодон-оптимизированную для *Yagowia*, связывали с конститутивным промотором *Yagowia*, FBA1 (SEQ ID NO: 4), с помощью стандартных методик молекулярной биологии. Кассета экспрессии Cas9, кодон-оптимизированная для *Yagowia*, содержащая конститутивный промотор FBA1, Cas9, кодон-оптимизированную для *Yagowia*, и NLS SV40, изложена под SEQ ID NO: 5. Данную кассету экспрессии Cas9 (SEQ ID NO: 5) клонировали в плазмиду pZUF с получением конструкции pZUFCas9 (фиг. 1, SEQ ID NO: 6).

Последовательность Cas9-NLS, кодон-оптимизированную для *Yagowia*, подвергали ПЦР амплификации на основании pZUFCas9 (SEQ ID NO: 6) с применением стандартных методик молекулярной биологии. Праймеры для ПЦР-реакции представляли собой SEQ ID NO: 7 (прямой) и SEQ ID NO: 8 (обратный), которые добавляли к амплифицированному ДНК-продукту 5'-сайт для EcoRI и 3'-сайт для HindIII соответственно. Добавленный 5'-сайт для EcoRI замещал стартовый кодон ATG открытой рамки считывания Cas9-NLS (ORF) в амплифицированном продукте. Амплифицированный продукт (SEQ ID NO: 9) расщепляли с помощью EcoRI и HindIII, а затем очищали с применением ZymoClean™ и колонок для концентрации (Zymo Research, Ирвин, Калифорния). Очищенный фрагмент ДНК клонировали в сайты для EcoRI и HindIII в плазмиде pBAD/HisB от Life Technologies (Карлсбрад, Калифорния) (фиг. 2A, SEQ ID NO: 10) для создания плазмидной конструкции pRF48 (фиг. 2B, SEQ ID NO: 11). Плазмида pRF48

способна экспрессировать в *E. coli* Cas9-NLS, содержащую гексагистидиновую (6xHis) метку на ее N-конце.

Для слияния последовательности проникающего в клетку пептида (CPP) с Cas9-NLS получали отдельные полинуклеотидные последовательности ДНК, при этом каждая была кодон-оптимизированной для экспрессии в *E. coli* и содержала последовательность, кодирующую 6xHis метку, связанную с аминокислотной последовательностью конкретного CPP: пептида Zebra (ECDSELEIKRYKRVRVARSKCRACKFKQLLQHYREVAATAKSSSENDRLRLLKQMC, SEQ ID NO: 12) из трансактиваторного белка Zebra вируса Эпштейна-Барр; пептида pVEC (LLILRRRIRKQAHANSK, SEQ ID NO: 13) из белка кадгерина эндотелия мыши; пептида TP10 (AGYLLGKINLKACAACAKKIL, SEQ ID NO: 14) из белка нейропептида галанина и синтетического богатого аргинином пептида "PolyR" (GGGGRRRRRRRRRLLLL, SEQ ID NO: 15). Каждая полинуклеотидная последовательность ДНК содержала на 5'-конце сайт рестрикции для NcoI и на 3'-конце сайт для EcoRI, чтобы получить последовательности для клонирования, имеющие следующую структуру: NcoI-6xHis-CPP-EcoRI (SEQ ID NO: 16-19). Каждую из SEQ ID NO: 16-19 отдельно клонировали в сайты для NcoI и EcoRI pRF48, в результате чего получали плазмидные конструкции, обеспечивающие экспрессию определенных белков слияния 6xHis-CPP-Cas9-NLS в *E. coli*. В частности, плазмидную конструкцию pRF144 (фиг. 3A, SEQ ID NO: 20) получали для экспрессии слияния 6xHis-Zebra CPP-Cas9-NLS; плазмидную конструкцию pRF145 (фиг. 3B, SEQ ID NO: 21) получали для экспрессии слияния 6xHis-PolyR CPP-Cas9-NLS; плазмидную конструкцию pRF146 (фиг. 3C, SEQ ID NO: 22) получали для экспрессии слияния 6xHis-TP10 CPP-Cas9-NLS и плазмидную конструкцию pRF162 (фиг. 3D, SEQ ID NO: 23) получали для экспрессии слияния 6xHis-pVEC CPP-Cas9-NLS.

Каждую из плазмид pRF48, pRF144, pRF145, pRF146 и pRF162 по отдельности трансформировали в компетентные клетки TOP10 (Life Technologies). Клетки выращивали в течение ночи при 37°C со встряхиванием (220 об/мин) в L-бульоне (Miller), содержащем 0,4% (вес/объем) глюкозы и 100 мкг/мл ампициллина. Каждую предварительную культуру разводили 1:100 в 2X среде YT, содержащей 100 мкг/мл ампициллина и далее выращивали при 37°C со встряхиванием (220 об/мин). Когда культуры достигали OD₆₀₀ приблизительно 0,5, экспрессию белка на основании каждой плазмиды индуцировали путем добавления L-арабинозы в конечной концентрации 0,2% (вес/объем). Культуры выращивали в течение дополнительных 18 ч при 18°C со встряхиванием (200 об/мин). Клетки осаждали при 5000×g в течение 15 мин при 4°C. Среду удаляли и клеточные осадки замораживали при -80°C в течение по меньшей мере 4 часов. Клеточные осадки размораживали в течение 15 минут на льду и ресуспендировали в 15 мл буфера для лизиса (20 mM Tris, pH 7,5, 500 mM NaCl, 1 mM MgCl₂, 10 mM имидазола, 120 единиц/мл ДНКазы I, 1 mM PMSF, 1 mM DTT) на литр исходной культуры. Клетки лизировали путем пропускания через большой пресс Френча для клеток при 16000 psi (фунтов на квадратный дюйм). Клеточный дебрис осаждали при 20000×g в течение 30 мин при 4°C. Надосадочные жидкости переносили в 50 мл конические пробирки, в которые добавляли 2 мл 50% суспензии смолы Ni-NTA (Qiagen) для связывания 6xHis метки каждого экспрессированного белка слияния. Содержимое каждой пробирки перемешивали путем медленного вращения при 4°C в течение 1 ч, а затем вносили в пустую колонку для гравитационного элюирования, через которую обеспечивали протекание надосадочной жидкости. Отбирали образец проточной фракции (75 мкл), добавляли к 25 мкл 4x-восстановленного буфера Лэммли и хранили на льду. Смолу промывали четыре раза в каждой колонке с помощью 5 мл промывочного буфера (20 mM Tris, pH 7,5, 500 mM NaCl, 10 mM имидазола, 1 mM PMSF, 1 mM DTT). Из каждого промывочного раствора отбирали образец (75 мкл), добавляли к 25 мкл 4x-восстановленного буфера Лэммли и хранили на льду. 1 мл аликвоты элюирующего буфера (20 mM Tris, pH 7,5, 500 mM NaCl, 1 mM MgCl₂, 500 mM имидазола, 1 mM PMSF, 1 mM DTT) вносили на смолу в каждой колонке и проводили инкубацию в течение 10 мин. Элюирование белка отслеживали по поглощению при 280 нм. Из каждого элюирования отбирали образец (75 мкл), добавляли к 25 мкл 4x-восстановленного буфера Лэммли и хранили на льду. Для каждого эксперимента по экспрессии плазмиды фракции, содержащие элюированный белок из колонки, объединяли, загружали на диализную мембрану 10000 MWCO и диализовали против буфера для диализа (25 mM HEPES/KOH, pH 7,5, 200 mM KCl, 20% глицерина, 1 mM DTT) при 4°C в течение по меньшей мере 14 часов. Концентрацию белка для каждого продукта диализа определяли с применением метода Бредфорда и поглощения при 565 нм. Очищенный белок разделяли на две аликвоты, одну из которых замораживали при -80°C, а другую хранили на льду при 4°C. Образцы, отобранные в ходе процесса очистки на колонке для каждого эксперимента по экспрессии плазмиды, нагревали при 95°C в течение 5 мин и загружали на 8% (вес/объем) Tris-глициновый полиакриламидный разделяющий гель с 4% (вес/объем) концентрирующего геля. Белки разделяли с помощью электрофореза при 200 вольт в течение 30 минут и окрашивали с помощью Кумасси голубого. Гель, полученный в ходе процесса очистки 6xHis-Zebra-Cas9-NLS, показан на фиг. 4 в качестве примера.

Таким образом, экспрессировали и выделяли четыре разных белка слияния CPP-Cas9. Эти белки слияния являются примерами комплексов белок RGEN-CPP согласно данному документу.

Пример 2. Экспрессия короткой направляющей РНК (sgRNA) с применением *in vitro* транскрипции.

В данном примере конструировали последовательность ДНК, которая кодирует sgRNA, слитую с

рибозимами на ее 5'- и 3'-концах (называемую "RGR") соответственно. Последовательность RGR обеспечивает возможность *in vitro* транскрипции sgRNA с точно определенными концами с помощью РНК-полимеразы T7.

На фиг. 5 изображена молекула sgRNA, которая представляет собой одиночную молекулу РНК, содержащую два участка, вариабельный нацеливающий домен (VT) (направляющая последовательность) и домен распознавания эндонуклеазы Cas (CER) (SEQ ID NO: 24 представляет собой пример CER). Участок VT может представлять собой, например, 20-мерный полинуклеотид РНК, который идентичен подвещающейся нацеливанию молекуле нуклеиновой кислоты. VT-домен определяет целевой сайт для расщепления в целевом сайте, который лежит в направлении 5' от мотива PAM. CER-домен взаимодействует с белком Cas9 и обеспечивает возможность VT-домену взаимодействовать с белком Cas9 и направлять опосредованное им расщепление (Jinek et al., Science 337:816-821). Как VT-, так и CER-домен необходимы для функционирования sgRNA.

Добавление рибозима типа HammerHead (HH) на 5'-конце и рибозима дельта-вируса гепатита (HDV) на 3'-конце последовательности sgRNA обеспечивает возможность экспрессии sgRNA с помощью любого промотора без учета определенных требований к транскрипции некоторых РНК-полимераз (например, для РНК-полимеразы T7 требуется один транскрибируемый G-остаток непосредственно после сайта инициации транскрипции, но она лучше функционирует в случае трех транскрибируемых G-остатков). Когда такая sgRNA экспрессируется, присутствующие в транскрипте pre-sgRNA рибозимы самоотщепляются, тем самым отделяясь от транскрипта и оставляя немодифицированную sgRNA.

Получали последовательность ДНК, кодирующую sgRNA, которая нацеливается на locus Can1-1 (SEQ ID NO: 25) в *Yarrowia lipolytica*; при этом данная sgRNA содержит SEQ ID NO: 24 в качестве ее CER-домена. Последовательность, кодирующую sgRNA, связывали на ее 5'-конце с последовательностью, кодирующей рибозим типа HH (SEQ ID NO: 26), и на ее 3'-конце с последовательностью, кодирующей рибозим HDV (SEQ ID NO: 27), так что первые 6 оснований рибозима типа HH были обратными комплементарными первым 6 основаниям участка VT sgRNA. Эта конкретная RGR sgRNA кодируется SEQ ID NO: 28. Затем RGR sgRNA под SEQ ID NO: 28 связывали с промотором РНК-полимеразы T7 (SEQ ID NO: 29) с помощью стандартных методик молекулярной биологии с получением плазмиды pRF46 (SEQ ID NO: 30).

Последовательность, кодирующую T7-RGR sgRNA, подвергали амплификации с помощью ПЦР с плазмиды pRF46 (SEQ ID NO: 30) с применением стандартных методик. Праймерами для ПЦР-реакции были SEQ ID NO: 31 (прямой праймер T7) и SEQ ID NO: 32 (обратный праймер gRNArev1). Продукт ПЦР очищали с помощью осаждения этанолом и ресуспендировали в ddH₂O; данную ДНК применяли в качестве матрицы для реакции *in vitro* транскрипции. ДНК-матрицу добавляли в конечной концентрации 150 нМ в 20 мкл реакционной смеси для *in vitro* транскрипции (набор MEGAShortscript™ T7, Life Technologies). Обеспечивали протекание реакции в течение различных промежутков времени (2 ч, 4 ч, 6 ч и в течение ночи) для определения подходящих условий для *in vitro* транскрипции (фиг. 6). Затем реакционные смеси обрабатывали 10 единицами ДНКазы I в течение 15 мин при 37°C для удаления ДНК-матрицы. РНК осаждали с применением этанола и стандартных протоколов. Для каждых 20 мкл реакционной смеси в ходе *in vitro* транскрипции получили от 60 до 100 мкг РНК.

Так проводили *in vitro* синтез sgRNA с определенными 5'- и 3'-концами. Как продемонстрировано в примере 3 ниже, sgRNA, полученную с помощью *in vitro* транскрипции, можно ассоциировать с белком слияния Cas9-CPP с образованием комплекса RGEN-CPP.

Пример 3. Специфическое *in vitro* расщепление целевой последовательности ДНК с применением белка слияния Cas9-CPP, находящегося в комплексе с sgRNA.

В данном примере тестировали нацеливающую эндонуклеазную функцию белка слияния Zebra CPP-Cas9 (содержащего SEQ ID NO: 39) в комплексе с sgRNA, чтобы подтвердить, что слияние с CPP не препятствует эндонуклеазной активности Cas9.

В анализе *in vitro* расщепления Can1 полинуклеотид ДНК (SEQ ID NO: 35), содержащий целевую последовательность Can1-1 под SEQ ID NO: 25, подвергали амплификации с помощью ПЦР из клеток *Y. lipolytica* (ATCC 20362) и очищали с применением стандартных методик. Праймерами для ПЦР-реакции были SEQ ID NO: 33 (прямой праймер IV-up) и SEQ ID NO: 34 (обратный праймер IV-down).

Очищенный белок слияния Zebra CPP-Cas9 (600 нг, получен в примере 1), sgRNA, нацеливающуюся на целевой сайт Can1-1 (250 нг, получена в примере 2), NEBuffer 3.1 (New England BioLabs, Ипсвич, Массачусетс), и ДНК для анализа расщепления Can1 (150 нг, SEQ ID NO: 35) смешивали в 10 мкл реакционной смеси (объем доводили до конечного объема с помощью ddH₂O). В качестве отрицательных контролей также готовили реакционные смеси с отсутствием либо белка слияния Zebra CPP-Cas9, либо sgRNA. В качестве положительного контроля вместо Zebra CPP-Cas9 в реакционной смеси применяли белок Cas9 дикого типа (PNA Bio, Таузенд-Оукс, Калифорния). Реакционные смеси инкубировали при 37°C в течение 60 мин. Затем в каждую реакционную смесь добавляли РНКазу I (4 мкг) и инкубировали при 37°C в течение 15 мин для разрушения sgRNA. Для остановки реакции добавляли стоп-раствор (1 мкл; 30% [вес/объем] глицерина, 1,2% [вес/объем] SDS, 250 мМ EDTA, pH 8,0), затем реакционные смеси дополнительно инкубировали в течение 15 мин при 37°C. Каждую реакционную смесь загрузили в

1,2% FlashGel™ (Lonza, Базель, Швейцария) и подвергали электрофорезу в течение 10 мин при 200 вольт (фиг. 7). Характер расщепления целевой ДНК, обеспечиваемый Zebra CPP-Cas9, согласовывался с характером расщепления, обеспечиваемым Cas9 дикого типа (фиг. 7), указывая тем самым на нормальное функционирование Zebra CPP-Cas9 *in vitro*. Более того, данная активность не подавлялась при применении Zebra CPP-Cas9/sgRNA, который подвергали двум циклам замораживания-оттаивания.

Таким образом, белок слияния CPP-Cas9, находящийся в комплексе с подходящей sgRNA (т. е. пример комплекса RGEN-CPP) обладал специфической активностью расщепления ДНК. Было показано, что данная активность аналогична активности комплекса Cas9 дикого типа-sgRNA, указывая тем самым на то, что слияние с CPP не подавляет эндонуклеотическую функцию Cas9-sgRNA. Хотя белок слияния CPP-Cas9 в данном примере содержал SEQ ID NO: 39 (Zebra CPP-Cas9), предполагается, что белок слияния CPP-Cas9, содержащий, например, SEQ ID NO: 40, 41 или 42, также характеризуется активностью расщепления, находясь в ассоциации с подходящей sgRNA в качестве РНК-компонента.

Пример 4. Доставка комплекса CPP-Cas9/sgRNA в дрожжевые клетки и расщепление целевой ДНК в них.

В данном примере белок слияния Zebra CPP-Cas9 (содержащий SEQ ID NO: 39) в комплексе с sgRNA (Zebra CPP-Cas9/sgRNA) тестировали в отношении способности проникать в дрожжевые клетки после простого контакта с клетками. Zebra CPP-Cas9/sgRNA, специфический в отношении Can1-1, обладать способностью проникать в клетки и расщеплять ген Can1, придавая тем самым клеткам устойчивость к канаванину.

Дрожжевые клетки *Y. lipolytica* (ATCC 20362) выращивали в жидкой среде YPD (2% глюкозы, 2% пептона, 1% дрожжевого экстракта) при 30°C со встряхиванием (220 об./мин.) до OD₆₀₀=0,5 (примерно 5×10⁶ клеток на мл культуры). Очищенный белок слияния Zebra CPP-Cas9 (получен в примере 1) и sgRNA, нацеливающуюся на целевой сайт Can1-1 (получена в примере 2), смешивали в молярном соотношении 1:3 соответственно в буфере для диализа, применяемом в примере 1, и предварительно инкубировали при комнатной температуре в течение 15 мин для обеспечения возможности ассоциации sgRNA с Zebra CPP-Cas9. 5×10⁵ клеток *Y. lipolytica* смешивали с раствором Zebra CPP-Cas9/sgRNA, так что конечная концентрация Zebra CPP-Cas9 составляла 1 мкМ, 2,5 мкМ или 5 мкМ. В качестве отрицательного контроля клетки также смешивали с 5 мкМ конечной концентрацией Zebra CPP-Cas9 отдельно (без sgRNA в качестве РНК-компонента). Все растворы клетки-Cas9 инкубировали при 30°C со встряхиванием (220 об./мин.) в течение 2 ч. Затем клетки серийно разводили в 1000 и 10000 раз. Каждое серийное разведение (100 мкл) высевали на полную среду, не содержащую аргинина (CM-Arg), и оставляли для восстановления в течение 48 ч при 30°C.

Колонии на планшетах с 10⁻³-разведением подсчитывали для определения общего количества высевных клеток. Колонии переносили на планшеты, содержащие CM-Arg с канаванином (60 мкг/мл), с помощью пересева методом реплик. Колонии оставляли расти при 30°C в течение 48 ч. Подсчитывали количество канавагин-устойчивых колоний и делили на общее число колоний (из планшетов без канавагина) для определения частоты мутаций в каждом случае. Приведение клеток в контакт с комплексами Zebra CPP-Cas9/sgRNA приводило к получению колоний, которые были устойчивы к канаванину, с частотой от приблизительно 2 до 10% от общего количества колоний (фиг. 8). Предполагалось, что данная устойчивость к канаванину обусловлена потерей функции гена Can1 в результате образования вставки/делеции в/вблизи предсказанного сайта расщепления для Cas9 в кодирующей последовательности гена Can1. Однако приведение клеток в контакт с Zebra CPP-Cas9 отдельно (без sgRNA) не приводило к получению устойчивых к канаванину колоний (фиг. 8), указывая на то, что устойчивость к канаванину у экспериментальных клеток зависела специфичности белка CPP-Cas9, которую придавала ему sgRNA. Учитывая природу дрожжевых клеток, комплексы CPP-Cas9/sgRNA, по-видимому, должны были пересечь как клеточную стенку, так и структуры клеточной мембраны, чтобы опосредовать специфическое нацеливание на ДНК.

Таким образом, белок слияния CPP-Cas9, находящийся в комплексе с подходящей sgRNA (т. е. пример комплекса RGEN-CPP), способен проникать в дрожжевые клетки (пересекать клеточную стенку и клеточную мембрану) и нацеливаться на специфическую последовательность ДНК в них. Хотя белок слияния CPP-Cas9 в данном примере содержал SEQ ID NO: 39 (Zebra CPP-Cas9), предполагается, что белок слияния CPP-Cas9, содержащий, например, SEQ ID NO: 40, 41 или 42, также обладает активностью проникновения в клетки и активностью специфического нацеливания на ДНК в клетках, находясь в ассоциации с подходящей sgRNA в качестве РНК-компонента.

Пример 5. CPP-облегченная доставка комплекса Cas9/sgRNA в растительные клетки и расщепление целевой ДНК в них.

CPP-облегченную доставку белка в клетки сои можно тестировать путем инкубирования каллюсных клеток сои с флуоресцентными белками DS-RED, слитыми с CPP. Ожидалось, что флуоресцентные сигналы будут регистрироваться в случае обработки CPP-DS-RED, но не в контролях, которые инкубировали только с белками DS-RED. Данным образом можно тестировать различные CPP, что помогает идентифицировать наиболее эффективные CPP для проникновения в растительную клетку и доставки

белкового груза. Некоторые примеры CPP, которые можно тестировать, включают: (i) CPP из трансактиваторного белка Zebra вируса Эпштейна-Барр, (ii) CPP с 6 или более смежными аргининовыми остатками, (iii) CPP транспортан-10 (TP10), (iv) CPP из белка кадгерина сосудистого эндотелия или (vi) CPP, выбранный из группы, состоящей из синтетического CPP с девятью аргининовыми остатками, богатого гистидином CPP с девятью аргининовыми остатками и Pas CPP с девятью аргининовыми остатками. Примеры синтетических CPP с девятью аргининовыми остатками, богатого гистидином CPP с девятью аргининовыми остатками и Pas CPP с девятью аргининовыми остатками раскрыты, например, в Liu et al. (*Advanced Studies in Biology* 5(2):71-88, HIKARI Ltd).

In vitro транслированные белки Cas9 и синтетическую sgRNA можно смешивать с CPP, самими по себе или в виде слияния (например, CPP-DS-RED выше), и инкубировать с каллусом сои для тестирования того, может ли Cas9/sgRNA переноситься в клетки. После проникновения в клетки комплекс Cas9/sgRNA может распознавать геномную мишень, определяемую нацеливающей последовательностью sgRNA, для создания двухнитевых разрывов ДНК (DSB). Спонтанная репарация DSB с помощью клеточного аппарата может приводить к мутациям вследствие нехомологичного соединения концов (NHEJ) или интеграции гена вследствие гомологичной рекомбинации, если донорная ДНК присутствует. CPP также можно ковалентно связывать с белками Cas9 для потенциально большей эффективности. Успех доставки CPP-Cas9/sgRNA в клетки сои и, таким образом, перенос комплекса CPP-эндонуклеаза Cas через клеточную стенку растительной клетки и мембрану растительной клетки можно проверить, например, путем выявления мутаций или интеграции генов в специфический целевой сайт с помощью ПЦР-анализа.

Пример 6. Экспрессия и очистка белков CPP-dsREDexpress из клеток *E. coli*.

Для быстрой оценки способности данного проникающего в клетку пептида проникать в специфический тип клеток создавали слияния CPP с белком dsREDexpress (SEQ ID NO: 85), обеспечивали их экспрессию в клетках *E. coli* и очищали. Слияния белков CPP-dsREDexpress представляют собой инструмент, обеспечивающий быструю оценку доставки груза в данный тип клеток с помощью данного CPP. Это обеспечивает возможность выбора молекулы CPP, специфической в отношении вида, типа клеток или штамма, чтобы максимизировать доставку груза быстрым и высокопроизводительным образом, путем оценки флуоресценции клеток с помощью микроскопического анализа или анализа проточной цитометрии.

Ген dsREDexpress, кодон-оптимизированный для *E. coli* (SEQ ID NO: 86) синтезировали (IDT DNA) и клонировали в сайты NcoI/HinDIII в pBAD/HisB (SEQ ID NO: 87), получая pRF161 (SEQ ID NO: 88). dsREDexpress, кодон-оптимизированный *E. coli*, содержал внутренний сайт для EcoRI, так что расщепление плазмиды с помощью NcoI/EcoRI обеспечило бы возможность замещения his-метки различными последовательностями his-метка-CPP для получения плазмид для экспрессии слияния his метка-CPP-dsREDexpress. Различные слияния his-метка-CPP; TAT (SEQ ID NO: 89), TLM (SEQ ID NO: 90), MPG1 (SEQ ID NO: 91), pep1 (SEQ ID NO: 92) и CFFKDEL (SEQ ID NO: 93) подвергали кодон-оптимизации для *E. coli* и фланкировали в рамке с помощью сайта для NcoI на 5'-конце и сайта для EcoRI на 3'-конце (SEQ ID NO: 94-98 соответственно), а также клонировали с применением стандартных методик в сайты для NcoI/EcoRI в pRF161 (SEQ ID NO: 88), замещая последовательность his-метки соответствующим слиянием his-метка-CPP и получая плазмиды pRF224 (his-TAT-dsREDexpress SEQ ID NO: 99), pRF214 (his-TLM-dsREDexpress SEQ ID NO: 100), pRF213 (his-MPG1-dsREDexpress SEQ ID NO: 101), pRF217 (his-pep1-dsREDexpress SEQ ID NO: 102), pRF216 (his-CFFKDEL-dsREDexpress SEQ ID NO: 103). Последовательности вставленных фрагментов проверяли с применением стандартных методик секвенирования и олигонуклеотида 36 (SEQ ID NO: 104).

His-Zebra (SEQ ID NO: 105), His-tp10 (SEQ ID NO: 106) и His-pVEC (SEQ ID NO: 107), кодон-оптимизированные для *E. coli*, подвергали амплификации с помощью ПЦР из pRF144 (SEQ ID NO: 108), pRF162 (SEQ ID NO: 109) и pRF146 (SEQ ID NO: 110) соответственно с применением олигонуклеотида 36 (SEQ ID NO: 104) и олигонуклеотида 153 (SEQ ID NO: 111) с помощью стандартных методик ПЦР. ПЦР-фрагменты клонировали в сайты для NcoI/EcoRI в pRF161 (SEQ ID NO: 88) с получением плазмид pRF186 (his-Zebra-dsREDexpress SEQ ID NO: 112), pRF192 (his-tp10-dsREDexpress SEQ ID NO: 113) и pRF190 (his-pVEC-dsREDexpress SEQ ID NO: 114). Последовательности проверяли с применением олигонуклеотида 36 (SEQ ID NO: 104).

Белки слияния CPP-dsREDexpress с his-меткой экспрессировали с применением стандартных методик. Вкратце, клетки предварительно культивировали либо в 10 мл ZYM-505 (1% N-Z амина, 0,5% дрожжевого экстракта, 5% глицерина, 1,0% декстрозы, 25 мМ Na₂HPO₄, 25 мМ KH₂PO₄, 50 мМ NH₄Cl, 5 мМ Na₂SO₄, 1x раствор следовых количеств металлов (Teknova), 5×10⁻⁵% тиамина, 2 мМ MgCl₂, 100 мкг/мл ампициллина), либо в лизогенном бульоне (1% триптона, 0,5% дрожжевого экстракта, 1% хлорида натрия, 100 мкг/мл ампициллина, 0,4% декстрозы) в 125 мл колбах в течение 12-16 ч при 37°C и 220 об./мин. Предварительные культуры разводили 1:1000 (ZYM-505) в 500 мл ZYM-5052 (1% N-Z амина, 0,5% дрожжевого экстракта, 5% глицерина, 0,5% декстрозы, 2% L-арабинозы, 25 мМ Na₂HPO₄, 25 мМ KH₂PO₄, 50 мМ NH₄Cl, 5 мМ Na₂SO₄, 1x раствор следовых количеств металлов (Teknova), 5×10⁻⁵% тиа-

мина, 2 мМ MgCl₂, 100 мкг/мл ампициллина) или 1:100 (бульон для лизиса) в 500 мл 2хYT (1,6% триптона, 1% дрожжевого экстракта, 0,5% NaCl, 100 мкг/мл ампициллина) и выращивали при 37°C, 220 об./мин. в 2,9 л колбах Фернбаха до OD₆₀₀ ~0,5. L-арабинозу добавляли до конечной концентрации 0,1% в культуре с 2х YT и температуру всех культур меняли на 18°C, 220 об./мин., в течение 20-30 ч для экспрессии белка. Клетки собирали при 5000 об./мин. в течение 10 мин, обработанную среду отбрасывали, а клеточные осадки замораживали при -80°C.

Клеточные осадки размораживали и ресуспендировали в денатурирующем буфере для лизиса (50 мМ Tris, pH 8,0, 150 мМ NaCl, 8 М мочевины, 20 мМ имидазола) и лизировали путем пропускания через пресс Френча для клеток при 16000 psi дважды. Твердые осадки отделяли от надосадочной жидкости путем центрифугирования при 10000 д, 4°C в течение 15 мин. 20 мкл очищенного от примесей экстракта смешивали с 20 мкл 2х буфера Лэммли (4% SDS, 20% глицерина, 100 мМ DTT, 0,004% бромфенолового синего, 125 мМ Tris, pH 6,8), нагревали до 95°C в течение 5 мин и замораживали при -20°C, чтобы сохранить для анализа. Очищенный от примесей экстракт смешивали с 6 мл 50% (объем/объем) суспензии никель-NTA-агароза в течение 1 ч при комнатной температуре. Гранулы осаждали из смеси при 2000 об./мин. в течение 5 мин. Надосадочную жидкость удаляли и отбирали 20 мкл для получения образца очищенного от примесей экстракта.

Осажденные гранулы ресуспендировали в 10 мл денатурирующего буфера для лизиса и вносили на хроматографическую колонку с гравитационным потоком. Жидкости давали возможность вытечь, при этом оставался слой из уплотненных гранул. Слой промывали с помощью серии промывок с применением различных соотношений промывочного буфера 1 (50 мМ Tris, pH 8,0, 150 мМ NaCl, 8 М мочевины, 20 мМ имидазола) и промывочного буфера 2 (50 мМ Tris, pH 8,0, 500 мМ NaCl, 20 мМ имидазола) для снижения концентрации денатурирующего средства (мочевины) и повышения концентрации NaCl, а также обеспечения возможности рефолдинга белка в колонке. Вкратце, колонку промывали с помощью (буфер 1: буфер 2): 10 мл 1:0 (8 М мочевины, 150 мМ NaCl), 10 мл 7:1 (7 М мочевины, 194 мМ NaCl), 10 мл 3:1 (6 М мочевины, 238 мМ NaCl) 10 мл 5:3 (5 М мочевины, 281 мМ NaCl), 10 мл 1:1 (4 М мочевины, 325 мМ NaCl), 20 мл 3:5 (3 М мочевины, 369 мМ NaCl), 20 мл 1:3 (2 М мочевины, 413 мМ NaCl), 20 мл 3:13 (1,5 М мочевины, 434 мМ NaCl), 20 мл 1:5 (1 М мочевины, 456 мМ NaCl), 20 мл 1:15 (0,5 М мочевины, 478 мМ NaCl) и 30 мл 0:1 (0 М мочевины, 500 мМ NaCl). Белок элюировали в нативных условиях с помощью буфера для элюирования (50 мМ Tris, pH 8,0, 500 мМ NaCl, 10% глицерина, 500 мМ имидазола) в 10х 1 мл фракциях. Фракции, содержащие элюированный белок dsREDexpress или CPP-dsREDexpress, были окрашены в красный цвет. Красные фракции объединяли и диализировали на 10000 MWCO регенерированной целлюлозной диализной мембране против 1000 объемов буфера для диализа (50 мМ Tris, pH 8,0, 10% глицерина) в течение ночи при комнатной температуре. Раствор белка удаляли из диализной мембраны и подвергали стерилизации фильтрованием с применением 0,22 мкм мембраны Tuffryn®. 20 мкл раствора белка обрабатывали как в случае очищенного от примесей экстракта клеток.

Образцы, отобранные в ходе очистки в буфере Лэммли, нагревали до 95°C в течение 5 мин и загружали в 12,5% PAGE-гель. Гель прогоняли при постоянном напряжении 200 вольт в течение 1 ч и окрашивали с применением красителя Simply blue. Пример типичного PAGE-геля, который использовали для очистки меченых белков CPP-dsREDexpress, показан на фиг. 9. Концентрацию общего белка для каждого очищенного белка определяли с применением анализа Coomassie Plus от Pierce™ с альбумином бычьей сыворотки в качестве стандарта. Концентрация каждого очищенного слияния CPP-dsREDexpress приведена в табл. 3.

Таблица 3. Концентрация очищенных слияний белка dsREDexpress

Белок	мг/мл	мкМ
dsREDexpress (SEQ ID NO: 700)	3,8	137
MPG1-dsREDexpress (SEQ ID NO: 751)	0,5	17
pVEC-dsREDexpress (SEQ ID NO: 752)	2,0	68
CFFKDEL-dsREDexpress (SEQ ID NO: 753)	1,5	54
TLM-dsREDexpress (SEQ ID NO: 754)	2,5	86
Zebra-dsREDexpress (SEQ ID NO: 755)	0,5	18
pep1-dsREDexpress (SEQ ID NO: 756)	0,3	10
tp10-dsREDexpress (SEQ ID NO: 757)	0,9	33

Пример 7. Экспрессия и очистка дополнительных белков CPP-Cas9 из клеток E. coli.

Для доставки Cas9 в различные типы клеток может потребоваться прикрепление Cas9 к различным молекулам CPP. С целью выделения различных белков слияния CPP-Cas9 разные CPP сливали с Cas9 в векторе экспрессии для E. coli. Эти белки экспрессировали и очищали из клеток E. coli для применения в CPP-опосредуемой доставке рибонуклеопротеинового комплекса Cas9/sgRNA в клетки.

Для получения кассет экспрессии со слиянием His-CFFKDEL-Cas9 (SEQ ID NO: 115) и His-MPG1-Cas9 (SEQ ID NO: 116) фрагменты NcoI/EcoRI pRF216 (CFFKDEL SEQ ID NO: 103) или pRF213 (MPG1 SEQ ID NO: 101) клонировали в те же сайты плазмиды pRF48 для экспрессии белка Cas9 (SEQ ID NO:

117) с применением стандартных методик с получением плазмиды pRF243 (his-CFFKDEL-Cas9 SEQ ID NO: 118) и pRF238 (his-MPG1-Cas9, SEQ ID NO: 119) соответственно. Правильность структуры кассет со слияниями MPG1-Cas9 или CFFKDEL-Cas9 подтверждали с помощью секвенирования по Сэнгеру с олигонуклеотидом 36 (SEQ ID NO: 104).

Белки слияния CPP-Cas9 с his-меткой экспрессировали с применением стандартных методик. Вкратце, клетки предварительно культивировали либо в 10 мл ZYM-505 (1% N-Z амина, 0,5% дрожжевого экстракта, 5% глицерина, 1,0% декстрозы, 25 мМ Na₂HPO₄, 25 мМ KH₂PO₄, 50 мМ NH₄Cl, 5 мМ Na₂SO₄, 1x раствор следовых количеств металлов (Teknova), 5×10⁻⁵% тиамина, 2 мМ MgCl₂, 100 мкг/мл ампициллина), либо в лизогенном бульоне (1% триптона, 0,5% дрожжевого экстракта, 1% хлорида натрия, 100 мкг/мл ампициллина, 0,4% декстрозы) в 125 мл колбах в течение 12-16 ч при 37°C и 220 об./мин. Предварительные культуры разводили 1:1000 (ZYM-505) в 500 мл ZYM-5052 (1% N-Z амина, 0,5% дрожжевого экстракта, 5% глицерина, 0,5% декстрозы, 2% L-арабинозы, 25 мМ Na₂HPO₄, 25 мМ KH₂PO₄, 50 мМ NH₄Cl, 5 мМ Na₂SO₄, 1x раствор следовых количеств металлов (Teknova), 5×10⁻⁵% тиамина, 2 мМ MgCl₂, 100 мкг/мл ампициллина) или 1:100 (бульон для лизиса) в 500 мл 2xYT (1,6% триптона, 1% дрожжевого экстракта, 0,5% NaCl, 100 мкг/мл ампициллина) и выращивали при 37°C, 220 об./мин. в 2,9 л колбах Фернбаха до OD₆₀₀ ~0,5. L-арабинозу добавляли до конечной концентрации 0,1% в культуры с 2x YT и температуру всех культур меняли на 18°C, 220 об/мин, в течение 20-30 ч для экспрессии белка. Клетки собирали при 5000 об/мин в течение 10 мин, отработанную среду отбрасывали и клеточные осадки замораживали при -80°C. Белки очищали, как описано в примере 1. Конечные концентрации очищенных белков CPP-Cas9, определенные с помощью анализа Coomassie Plus (Pierce™), перечислены в табл. 4.

Таблица 4. Концентрация очищенных белков CPP-Cas9

Белок	mg/ml	µM
Zebra-Cas9 (SEQ ID NO: 758)	1,5	9
CFFKDEL-Cas9 (SEQ ID NO: 730)	4,6	28
MPG1-Cas9 (SEQ ID NO: 731)	3,8	23
pVEC-Cas9 (SEQ ID NO: 759)	2,5	15

Пример 8. Опосредуемое CPP-Cas9/qRNA нацеливание на ген в клетках *E. coli*.

Данный пример демонстрирует обработку клеток *Escherichia coli* с помощью рибонуклеопротеиновых комплексов CPP-Cas9/sgRNA, которые обеспечивают нацеливание sgRNA на ген galK *E. coli*. Проникновение CPP-Cas9/sgRNA в клетку обеспечивает осуществление нацеливания и расщепления в гене galK, что приводит к инактивации гена вследствие подверженных ошибкам механизмов репарации ДНК, которые можно фенотипически отслеживать как устойчивость к галактозе. Этот способ обусловлен доставкой груза Cas9/sgRNA в клетки посредством CPP-опосредуемой доставки.

Ген galK *E. coli* (SEQ ID NO: 120) отвечает за фенотип чувствительности к галактозе, наблюдаемый у мутантов galE в присутствии сахара галактозы. Когда галактоза попадает в клетку, она фосфорилируется галактокиназой, продуктом гена galK (SEQ ID NO: 120). Галактозофосфат является токсичным для клеток. В клетках дикого типа галактозофосфат дополнительно метаболизируется продуктами генов gale (SEQ ID NO: 121) и galT (SEQ ID NO: 122) и применяется в качестве источника углерода. У мутантов с потерей функции, galE или galT, галактозофосфат накапливается, что приводит к гибели клеток. Следовательно, мутации с потерей функции в гене galK могут быть выбраны при фоновом генотипе мутанта galE, как обеспечивающие возможность образования колонии в присутствии галактозы.

Для получения sgRNA (SEQ ID NO: 135), нацеливающейся на ген galK (SEQ ID NO: 120) в целевом сайте galK2-1 (SEQ ID NO: 134), получали матрицу для *in vitro* транскрипции (SEQ ID NO: 131). Вначале продукт ПЦР из ДНК, кодирующей CER-домен (SEQ ID NO: 123), амплифицировали из pRF291 (SEQ ID NO: 125) с применением прямого праймера CER (SEQ ID NO: 126) и универсального обратного праймера (SEQ ID NO: 127) в стандартной ПЦР-реакции (SEQ ID NO: 124). Продукт ПЦР, кодирующий CER (SEQ ID NO: 124), очищали с применением колонок Zymo™ clean & concentrator 25 и элюировали с помощью 35 мкл ddH₂O. При амплификации матрицы для *in vitro* транскрипции sgRNA применяли мультиплексную ПЦР, предусматривающую 4 праймера, универсальный прямой праймер, содержащий промотор T7 (SEQ ID NO: 128), специфический в отношении мишени прямой праймер, содержащий некоторую часть промотора T7 и некоторую часть целевого сайта (SEQ ID NO: 129), специфический в отношении мишени обратный праймер, содержащий некоторую часть целевого сайта и перекрывание с CER-доменом (SEQ ID NO: 130), и универсальный обратный праймер (SEQ ID NO: 127). ПЦР-реакцию проводили с применением мастер-микса Phusion flash, содержащего 15 нМ продукта ПЦР CER-домена (SEQ ID NO: 124), 1 мкМ каждого универсального прямого (SEQ ID NO: 128) и обратного праймеров (SEQ ID NO: 127) и 300 нМ каждого из специфического в отношении мишени прямого (SEQ ID NO: 129) и специфического в отношении мишени обратного (SEQ ID NO: 130) праймеров. Циклы ПЦР-реакции проводили как в случае стандартной реакции. Матрицу для *in vitro* транскрипции sgRNA (SEQ ID NO: 131), очищали с применением колонок Zymo clean & concentrator 25 и элюировали с помощью 35 мкл ddH₂O. Матрица для *in vitro* транскрипции sgRNA (SEQ ID NO: 131) содержала промотор T7 (SEQ ID NO: 132), ДНК, кодирую-

щую вариабельный нацеливающий домен для galK2-1 (SEQ ID NO: 133), и ДНК, кодирующую CER-домен (SEQ ID NO: 125). Реакцию *in vitro* транскрипции для получения sgRNA galK2-1 (SEQ ID NO: 135) выполняли, как описано в примере 2.

СРР-опосредуемую доставку нуклеопротеиновых комплексов Cas9/sgRNA выполняли путем выращивания штамма *E. coli* с делецией по galE в лизогенном бульоне (1% триптона, 0,5% дрожжевого экстракта, 1% NaCl) в течение ночи при 37°C, 220 об./мин. Культуру разводили 1:100 в свежем лизогенном бульоне и выращивали при 37°C, 220 об./мин. в течение 2 ч для получения клеток на фазе экспоненциального роста. СРР-Cas9 (pVEC-Cas9 (SEQ ID NO: 144), Zebra-Cas9 (SEQ ID NO: 143), MPG1-Cas9 (SEQ ID NO: 116), CFFKDEL-Cas9 (SEQ ID NO: 115)) инкубировали при конечной концентрации 10 мкМ либо в присутствии, либо в отсутствие 10 мкМ sgRNA galK2-1 (SEQ ID NO: 135) в объеме 50 мкл в течение 30 мин при комнатной температуре. В случае обработки 1,2 мл клеток осаждали при 3000 об./мин. в течение 3 мин, надосадочную жидкость отбрасывали, а клетки ресуспендировали в 600 мкл LB, содержащего буфер с 2х нуклеазой (200 мМ NaCl, 100 мМ Tris-HCl, 20 мМ MgCl₂, 200 мкг/мл BSA, pH 7,9). 50 мкл суспензии клеток смешивали с каждой реакционной смесью, а также только с контролем, содержащим только gRNA, и клетками без обработки. Образцы инкубировали при 37°C, 220 об./мин. в течение 4 ч. По 100 мкл разведений образцов из расчета 10⁻³, 10⁻⁴ и 10⁻⁵ высевали на планшеты с лизогенным бульоном для подсчета количества жизнеспособных клеток в конце обработки, остальную часть реакционной смеси высевали на планшеты с лизогенным бульоном и инкубировали в течение ночи при 37°C. Жизнеспособные клетки из разведения 10⁻⁵ подсчитывали для определения количества жизнеспособных колониеобразующих единиц (CFU), посеянных на планшет с образцами на лизогенном бульоне. Планшеты с образцами подвергали пересеву методом реплик с помощью стандартных методик на минимальную среду А (1 г/л (NH₄)₂SO₄, 4,5 г/л KH₂PO₄, 10,5 г/л K₂HPO₄, 0,5 г/л цитрата натрия·2H₂O, 1 мМ MgSO₄·7H₂O, 5×10⁻⁵% тиамин), загущенную с помощью 1,5% (вес/объем) бактоагара, содержащего 0,2% (вес/объем) глицерина и 0,2% (вес/объем) галактозы в качестве источников углерода. Планшеты инкубировали при 37°C в течение 24 ч и затем оценивали в отношении образования колоний. Каждая CFU из штамма galE на планшете, содержащем галактозу, представляет собой событие инактивации гена для гена galK. Результаты посева методом реплик показаны в табл. 5.

Таблица 5. Частота инактивации гена galK в мутантных по galE клетках *E. coli*, обработанных с помощью СРР-Cas9/sgRNA

Белок Cas9	sgRNA	CFU на галактозе	CFU, посеянные на галактозу	Частота Gal ^R CFU	Кратная частота Gal ^R /частота необработанных Gal ^R
Отсутствует	Отсутствует	21	1,65×10 ⁸	1,27×10 ⁻⁷	1,00
pVEC-Cas9	Отсутствует	21	1,18×10 ⁸	1,78×10 ⁻⁷	1,39
pVEC-Cas9	galK2-1	15	1,23×10 ⁸	1,22×10 ⁻⁷	0,96
MPG1-Cas9	Отсутствует	22	1,34×10 ⁸	1,65×10 ⁻⁷	1,29
MPG1-Cas9	galK2-1	16	1,11×10 ⁸	1,44×10 ⁻⁷	1,13
Zebra-Cas9	Отсутствует	29	1,89×10 ⁸	1,53×10 ⁻⁷	1,20
Zebra-Cas9	galK2-1	25	8,88×10 ⁷	2,82×10 ⁻⁷	2,21
CFFKDEL-Cas9	Отсутствует	29	1,24×10 ⁸	2,34×10 ⁻⁷	1,84
CFFKDEL-Cas9	galK2-1	63	1,24×10 ⁸	5,10×10 ⁻⁷	4,00
Отсутствует	galK2-1	31	1,42×10 ⁸	2,19×10 ⁻⁷	1,72

Обработка клеток *E. coli* с помощью рибонуклеопротеиновых комплексов СРР-Cas9/sgRNA в некоторых случаях увеличивала частоту инактивации galK приблизительно в 4 раза на фоне необработанных клеток. Данное увеличение не наблюдалось у клеток, обработанных только с помощью СРР-Cas9 или sgRNA, что дает основание полагать, что повышенная инактивация гена galK была связана с попаданием рибонуклеопротеина СРР-Cas9/sgRNA в клетку и образованием двухнитевых разрывов ДНК в целевом сайте galK2-1 в пределах гена galK.

Пример 9. Доставка СРР-белка dsREDexpress в клетки архей.

Для тестирования доставки груза в клетки архей с применением проникающих в клетку пептидов и определения кандидатных СРР, которые преодолевают клеточную стенку архей, включающую элементы, которые аналогичны таковым в клеточных стенках бактериальной и эукариотической клетки (например, фосфолипиды), а также мембраны и элементы, которые характерны только для архей (например, S-слой), клетки архей обрабатывали с помощью слияний СРР-белок dsREDexpress. СРР, идентифицированные в

данном скрининге, можно будет применять для доставки другого груза (например, рибонуклеопротеинового комплекса Cas9/sgRNA) в клетки архей.

Архею *Halobacterium salinarum* ATCC19700 выращивали на среде 213 (250 г/л NaCl, 10 г/л MgSO₄·7H₂O, 5 г/л KCl, 0,2 г/л CaCl₂·6H₂O, 10 г/л дрожжевого экстракта, 2,5 г/л триптона), загущенной с помощью 1,5% бактоагара, при 37°C до образования колоний (4 дня). Одиночную колонию использовали для инокуляции 50 мл среды 213 в 250 мл колбе. Культуру выращивали при 37°C, 220 об./мин. до достижения OD₆₀₀ примерно 0,5, что указывает на фазу экспоненциального роста. Готовили смеси из 100 мкл клеток без белков или с 5 мкМ dsREDexpress (SEQ ID NO: 85), или 5 мкМ MPG1-dsREDexpress (SEQ ID NO: 136), или 5 мкМ pVEC-dsREDexpress (SEQ ID NO: 137), или 5 мкМ CFFKDEL-dsREDexpress (SEQ ID NO: 138), или 5 мкМ TLM-dsREDexpress (SEQ ID NO: 139), или 5 мкМ pep1-dsREDexpress (SEQ ID NO: 141) или 5 мкМ tp10 dsRED-express (SEQ ID NO: 142) в 24-луночном блоке. Смеси инкубировали в течение 4 часов при 37°C, 220 об./мин. Клетки промывали дважды средой 213, не содержащей триптон и дрожжевой экстракт, и ресуспендировали в 100 мкл среды 213, не содержащей триптон и дрожжевой экстракт. Клетки анализировали в отношении флуоресценции в красном канале проточного цитометра Accuri C5 для определения, какие CPP-метки доставляли груз dsREDexpress в клетки *H. salinarum*. Необработанные клетки использовали для создания гейта для анализа данных проточной цитометрии, разграничивающего неокрасные и красные клетки, так что гейт приводил к частоте ложноположительных событий 0,2%, необработанных клеток, которые попадали в красный гейт (табл. 6).

Таблица 6. CPP-доставка dsREDexpress в *H. salinarum*

Обработка	Процент популяции в гейте красных клеток ± стандартное отклонение ¹	Кратное увеличение популяции красных клеток относительно обработки dsREDexpress отдельно
Без dsREDexpress	0,21±0,06	0,73
dsREDexpress	0,29±0,21	1,00
MPG1-dsREDexpress	0,37±0,08	1,27
pVEC-dsREDexpress	16,87±9,90	57,50
CFFKDEL-dsREDexpress	0,33±0,14	1,14
TLM-dsREDexpress	2,03±1,02	6,93
pep1-dsREDexpress	0,36±0,18	1,23
tp10-dsREDexpress	0,91±0,27	3,09

¹Данные соответствуют трем повторностям ± стандартное отклонение.

Доставка груза dsREDexpress в клетки архей демонстрирует, что по меньшей мере три проникающие в клетку пептиды (pVEC, TLM, tp10) способны доставлять белковый груз в клетки архей с эффективностью в 50 раз превышающую доставку белка dsREDexpress отдельно, что дает основание полагать, что эти три CPP-мотива можно применять для доставки другого груза в клетки архей (например, рибонуклеопротеинового комплекса Cas9). Кроме того, поскольку CPP-мотивы доставляют груз в клетки, составляющие вплоть до 16% всей клеточной популяции, это дает основание полагать, что доставка груза с помощью CPP в клетки архей является эффективным процессом.

Пример 10. Доставка CPP-белка dsREDexpress в эукариотические клетки.

Для тестирования способности проникающих в клетку пептидов доставлять груз в разные виды эукариотических организмов группу из трех видов, *Phytophthora capsici* (Oomycete), *Septoria tritici* (настоящие грибы) и *Botrytis cinerea* (настоящие грибы), обрабатывали с помощью различных слияний CPP-dsREDexpress. Доставку груза dsREDexpress отслеживали для различных фрагментов CPP с помощью анализа FACS для определения процентной доли клеток, в которые груз был доставлен. Если CPP будут способны доставлять груз dsREDexpress в данные клетки, то это дает основание полагать, что данные CPP будут способны доставлять другие виды груза в данные классы эукариотических клеток (например, рибонуклеопротеинового комплекса Cas9/sgRNA).

P. capsici выращивали на среде V8 (20% сок V8, 4,5 г/л CaCO₃), загущенной с помощью 1,8% бактоагара, при 23°C в темноте в течение 3 дней. Затем планшет помещали на свет при 23°C в течение дополнительных 7 дней. Планшеты охлаждали при 4°C в течение 30 мин. На планшет наливали воду, чтобы только покрыть поверхность, и оставляли для инкубации в течение 30 мин при комнатной температуре. Жидкость удаляли для сбора зооспор. Наличие зооспор подтверждали с помощью микроскопического анализа. Равный объем 2х среды для инцистирования (40 г/л триптона, 10 г/л дрожжевого экстракта, 200 мл/л 10х SOC-солей [5,84 г/л NaCl, 1,86 г/л KCl, 20,3 г/л MgCl₂·6H₂O, 24,6 г/л MgSO₄·7H₂O, 36 г/л декстрозы], 36,4 г/л сорбитола, 1,47 г/л CaCl₂·2H₂O) добавляли к зооспорам и аккуратно перемешивали. Зооспоры в среде для инцистирования инкубировали в течение 20 мин при комнатной температуре. Инцистирование подтверждали микроскопически. Споры осаждали и ресуспендировали в равном объеме среды YMA (2 г/л дрожжевого экстракта, 4 г/л солодового экстракта) и подсчитывали с применением гемо-

цитометра. Зооспоры разводили до концентрации 3×10^7 спор/мл в YMA. 100 мкл зооспор в YMA смешивали с различными белками слияния dsREDexpress (новый пример 5, табл. N1) до конечной концентрации белка 5 мкМ. Смеси инкубировали при 25°C, 400 об./мин. в течение 2 ч. Клетки промывали дважды фосфатно-солевым буферным раствором (PBS) (8г/л NaCl, 0,2г/л KCl, 1,44г/л Na₂HPO₄·2H₂O, 0,24г/л KH₂PO₄ pH 6,8) и ресуспендировали в конечном объеме 200 мкл PBS. Захват белков слияния dsREDexpress отслеживали с применением проточной цитометрии, как в случае с *Halobacterium salinarium* (пример 9). Процент клеток, в которые груз был успешно доставлен, определяли путем проведения произвольного гейта у клеток, обработанных dsREDexpress, так что 0,1% популяции оценивалось как ложноположительный красный объект (1:1000 клеток). Результаты данной обработки можно видеть в табл. 7. pVEC, pep1 и tp10 обеспечивали получение в 5,8, 5,5 и 1,8 раза больше красных клеток, чем при обработке клеток dsREDexpress отдельно, что дает основание полагать, что данные фрагменты CPP могут быть кандидатами для доставки другого вида груза в Oomycetes (например, рибонуклеопротеинового комплекса Cas9/sgRNA)

Таблица 7. CPP-доставка dsREDexpress в *Phytophthora capsici*

Обработка	Процент популяции в гейте красных клеток ± стандартное отклонение ¹	Кратное увеличение популяции красных клеток относительно обработки dsREDexpress отдельно
dsREDexpress	0,10±0,03	1,00
pVEC-dsREDexpress	0,56±0,16	5,79
CFFKDEL-dsREDexpress	0,01±0,01	0,07
TLM-dsREDexpress	0,00±0,00	0,00
pep1-dsREDexpress	0,53±0,29	5,52
tp10-dsREDexpress	0,17±0,14	1,76
MPG-dsREDexpress	0,00±0,00	0,00
Zebra-dsREDexpress	0,03±0,05	0,34

¹Данные соответствуют трем биологическим повторностям ± стандартное отклонение.

B. cinerea выращивали на среде PDA (24 г/л картофельно-декстрозного бульона), загущенной с помощью 1,8% бактоагара, в темноте в течение 5-10 дней. Конидии собирали в воде с помощью стерильного пластикового шпателя и фильтровали через 2 слоя марли. Конидии подсчитывали на гемоцитометре и разводили до 3×10^7 конидий на мл в среде YMA. 100 мкл конидий в YMA смешивали с различными белками слияния dsREDexpress (новый пример 5, табл. N1) до конечной концентрации белка 5 мкМ. Смеси инкубировали при 25°C, 400 об./мин. в течение 2 ч. Клетки промывали дважды фосфатно-солевым буферным раствором (PBS) (8г/л NaCl, 0,2г/л KCl, 1,44г/л Na₂HPO₄·2H₂O, 0,24г/л KH₂PO₄ pH 6,8) и ресуспендировали в конечном объеме 200 мкл PBS. Захват белков слияния dsREDexpress отслеживали с применением проточной цитометрии, как в случае с *Halobacterium salinarium* (пример 8). Процент клеток, в которые груз был успешно доставлен, определяли путем проведения произвольного гейта у клеток, обработанных dsREDexpress, так что 0,1% популяции оценивалось как ложноположительный красный объект (1:1000 клеток). Результаты данной обработки можно видеть в табл. 8.

Таблица 8. CPP-доставка dsREDexpress в *Botrytis cinerea*

Обработка	Процент популяции в гейте красных клеток ± стандартное отклонение ¹	Кратное увеличение популяции красных клеток относительно обработки dsREDexpress отдельно
dsREDexpress	0,12±0,04	1,00
pVEC-dsREDexpress	0,08±0,10	0,68
CFFKDEL-dsREDexpress	0,03±0,01	0,22
TLM-dsREDexpress	0,01±0,01	0,05
pep1-dsREDexpress	0,01±0,01	0,05
tp10-dsREDexpress	0,03±0,02	0,24
MPG-dsREDexpress	0,01±0,02	0,11
Zebra-dsREDexpress	0,01±0,02	0,11

¹Данные соответствуют трем биологическим повторностям ± стандартное отклонение.

S. tritici выращивали на среде YMA, загущенной с помощью 1,8% бактоагара, при 23°C на свету. Конидии собирали через 5-10 дней с помощью стерильного пластикового шпателя и воды. Конидии подсчитывали на гемоцитометре и разводили до 3×10^7 конидий в среде YMA. 100 мкл конидий в YMA смешивали с различными белками слияния dsREDexpress (новый пример 5, табл. N1) до конечной концен-

трации белка 5 мкМ. Смеси инкубировали при 25°C, 400 об./мин. в течение 2 ч. Клетки промывали дважды фосфатно-солевым буферным раствором (PBS) (8г/л NaCl, 0,2г/л KCl, 1,44г/л Na₂HPO₄·2H₂O, 0,24г/л KH₂PO₄ pH 6,8) и ресуспендировали в конечном объеме 200 мкл PBS. Захват белков слияния dsREDexpress отслеживали с применением проточной цитометрии, как в случае с *Halobacterium salinarium* (пример 9). Процент клеток, в которые груз был успешно доставлен, определяли путем проведения произвольного гейта у клеток, обработанных dsREDexpress, так что 0,1% популяции оценивалось как ложноположительный красный объект (1:1000 клеток). Результаты данной обработки можно видеть в табл. 9. pVEC, TLM, pep1 и tp10 повышали доставку dsREDexpress в 25, 4, 3 и 5 раз соответственно по сравнению с dsREDexpress отдельно. Это дает основание полагать, что данные CPP будут хорошими кандидатами для доставки другого груза в настоящие грибы (например, рибонуклеопротеинового комплекса Cas9/sgRNA).

Таблица 9. CPP-доставка dsREDexpress в *Septoria tritici*

Обработка	Процент популяции в гейте красных клеток \pm стандартное отклонение ¹	Кратное увеличение популяции красных клеток относительно обработки dsREDexpress отдельно
dsREDexpress	0,12 \pm 0,03	1,00
pVEC-dsREDexpress	3,02 \pm 0,91	25,2
CFPKDEL-dsREDexpress	0,00 \pm 0,01	0,03
TLM-dsREDexpress	0,48 \pm 0,14	4,03
pep1-dsREDexpress	0,37 \pm 0,21	3,06
tp10-dsREDexpress	0,71 \pm 0,69	5,94
MPG-dsREDexpress	0,14 \pm 0,05	1,17
Zebra-dsREDexpress	0,00 \pm 0,00	0,00

¹ Данные соответствуют трем биологическим повторностям \pm стандартное отклонение.

Пример 11. Доставка семи CPP-dsRED и двух CPP-метка RFP в семь кишечных бактерий.

В данном примере тестировали эффективность CPP в доставке двух белков-грузов, dsRED и метка RFP, в 7 видов кишечных бактерий (для которых были продемонстрированы положительное влияние на физиологию организма хозяина).

Бактериальные клетки выращивали на соответствующих средах (см. таблицу 10) в течение ночи при 37° С в ротационном шейкере при 150 об./мин. в анаэробной палатке (80% N₂, 15% CO₂ и 5% H₂). Для данного анализа 1 \times 10⁸ бактериальных клеток смешивали с 5 мкМ конечной концентрацией белков CPP-dsRED и CPP-метка RFP в 96-луночном планшете, за которым следовало два часа разрастания при 37°C. Для измерения сигналов флуоресценции dsRED и RFP в клетках бактериальные клетки собирали путем центрифугирования (3500 \times g, 4°C, 20 мин.) и дважды промывали фосфатно-солевым буферным раствором (100 мкл на лунку). Интенсивности флуоресценции определяли количественно с помощью планшетридера Tecan Spark 10M (Tecan, Меннедорф, Швейцария), оснащенного фильтрами для длины волны возбуждения 554 нм и длины волны испускания 586 нм с шириной полосы 10 нм. Исходные значения флуоресценции вычитали из значений необработанных клеток (фон). В качестве минимального порога доставки CPP внутрь клеток считали значения интенсивности флуоресценции, составляющие 7000.

Таблица 10. Культуральная среда для 7 видов бактерий

Бактерии	Тип	Культуральная среда
<i>Bacteroides thetaiotaomicron</i>	Bacteroidetes	Сердечно-мозговая вытяжка, дополненная 10% крови крупного рогатого скота (ВНИ-кровь)
<i>Eubacterium hallii</i>	Firmicutes	ВНИ-кровь
<i>Faecalibacterium prausnitzii</i>	Firmicutes	ВНИ-кровь
<i>Blautia hydrogenotrophica</i>	Firmicutes	YCFA
<i>Bacteroides fragilis</i>	Bacteroidetes	ВНИ-кровь
<i>Prevotella histicola</i>	Bacteroidetes	ВНИ-кровь
<i>Clostridium scindens</i>	Firmicutes	YCFA

Как показано в табл. 11, данные результаты указывают на то, что пять CPP, включая MPG, pVEC,

TLM, ZEBRA и pep1, эффективно доставлялись в анаэробные кишечные бактерии, принадлежащие к типам Firmicutes и Bacteroidetes, что указывает на то, что CPP могут пересекать клеточную мембрану данных бактерий (табл. 9).

Таблица 11. Дифференциальные эффективности доставки CPP в различные бактериальные штаммы, как показано согласно интенсивности флуоресценции выше порогового значения 7000

	MPG-1- dsRED	pVEC- dsRED	TLM- dsRED	ZEBRA- dsRED	pep1- dsRED
<i>Bacteroides thetaiotaomicron</i>	-	-	-	10230	16657
<i>Eubacterium hallii</i>	10015	17156	-	16894	7004
<i>Faecalibacterium prausnitzii</i>	-	40525	14998	17014	12696
<i>Blautia hydrogenotrophica</i>	-	11770	14612	9623	-
<i>Bacteroides fragilis</i>	-	14783	-	15026	-
<i>Prevotella histicola</i>	-	-	-	22416	-
<i>Clostridium scindens</i>	-	17677	32492	-	-

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Способ доставки белкового компонента направляемой РНК-эндонуклеазы (RGEN) в микробную клетку, причем указанный способ включает

приведение микробной клетки в контакт с композицией, содержащей белковый компонент направляемой РНК-эндонуклеазы (RGEN) и по меньшей мере один проникающий в клетку пептид (CPP),

где указанный белковый компонент и CPP ковалентно или нековалентно связаны друг с другом в комплекс белок RGEN-CPP, и

где композиция дополнительно содержит по меньшей мере один РНК-компонент, который ассоциирован с белковым компонентом RGEN,

где РНК-компонент содержит последовательность, комплементарную последовательности целевого сайта на хромосоме или эписоме в клетке, где RGEN может связываться с последовательностью целевого сайта и, необязательно, расщеплять одну или обе нити ДНК в последовательности целевого сайта,

где указанный комплекс белок RGEN-CPP пересекает (i) клеточную мембрану или (ii) клеточную стенку и клеточную мембрану клетки, вследствие чего он попадает в микробную клетку.

2. Способ по п.1, где RGEN может расщеплять одну или обе нити ДНК в последовательности целевого сайта.

3. Способ по п.2, где микробная клетка содержит донорный полинуклеотид, содержащий по меньшей мере одну последовательность, гомологичную последовательности в последовательности целевого сайта или вблизи нее, и где донорный полинуклеотид интегрируется в последовательность целевого сайта или вблизи нее с помощью гомологичной рекомбинации.

4. Способ модификации целевого сайта в геноме микробной клетки, причем способ включает обеспечение направляющего полинуклеотида, проникающего в клетку пептида (CPP) и эндонуклеазы Cas для клетки, где указанный направляющий полинуклеотид, эндонуклеаза Cas и CPP ковалентно или нековалентно связаны друг с другом в комплекс направляющий полинуклеотид/эндонуклеаза Cas-CPP, и где указанный комплекс направляющий полинуклеотид/эндонуклеаза Cas-CPP может пересекать (i) клеточную мембрану или (ii) клеточную стенку и клеточную мембрану микробной клетки.

5. Способ по п.1, где РНК-компонент содержит направляющую РНК (gRNA), содержащую РНК CRISPR (crRNA), функционально связанную с транскриптивирующей РНК CRISPR (tracrRNA).

6. Способ по п.1, где RGEN содержит аминокислотную последовательность CRISPR-ассоциированного (Cas) белка-9 (Cas9).

7. Способ по п.1, где белковый компонент RGEN и CPP связаны ковалентно.

8. Способ по п.1, где белковый компонент RGEN и CPP связаны нековалентно.

9. Способ по п.1 или 4, где CPP является катионным или амфипатическим.

10. Способ по п.1 или 4, где CPP включает

(i) CPP из трансактиваторного белка Zebra вируса Эпштейна-Барр,

(ii) CPP с 6 или более смежными аргининовыми остатками,

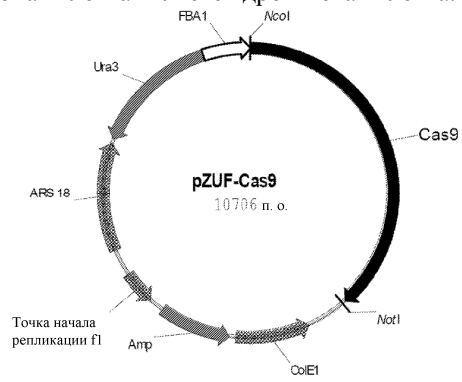
(iii) CPP транспортан-10 (TP10) или

(iv) CPP из белка кадгерина сосудистого эндотелия.

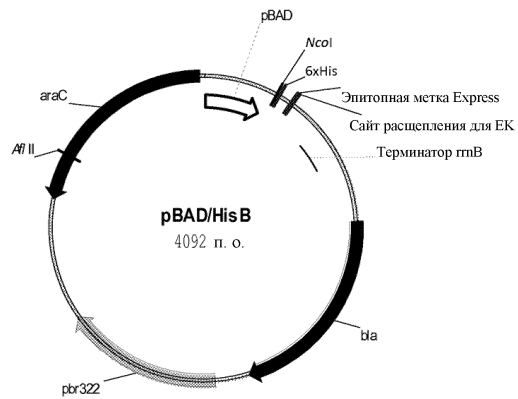
11. Способ по п.1, где указанный комплекс белок RGEN-CPP может пересекать клеточную стенку и клеточную мембрану микробной клетки.

12. Способ по п.4, указанный комплекс направляющий полинуклеотид/эндонуклеаза Cas-CPP может пересекать клеточную стенку и клеточную мембрану микробной клетки.

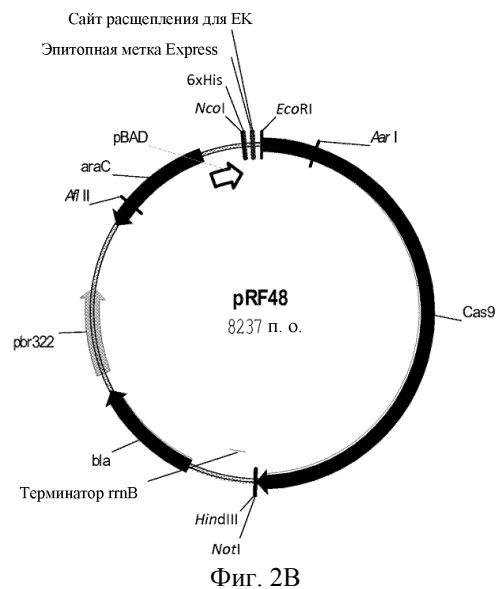
13. Способ по любому из предыдущих пунктов, где микробной клеткой является грибная клетка.
 14. Способ по п.13, где грибная клетка является дрожжевая клетка.



Фиг. 1

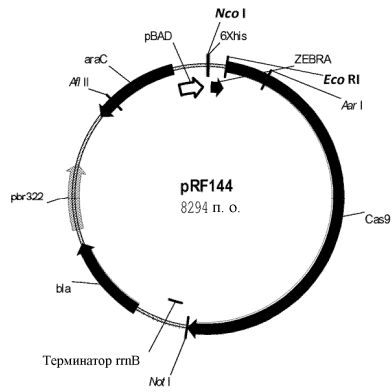


Фиг. 2А

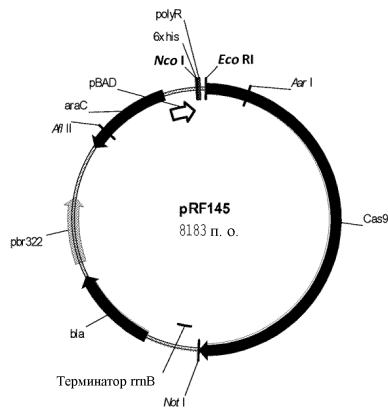


Фиг. 2В

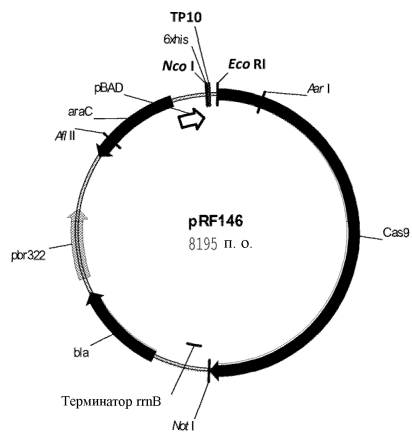
038321



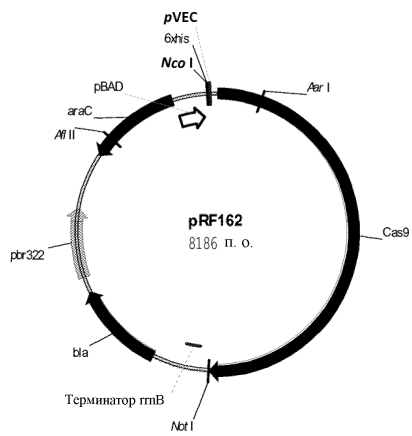
Фиг. 3А



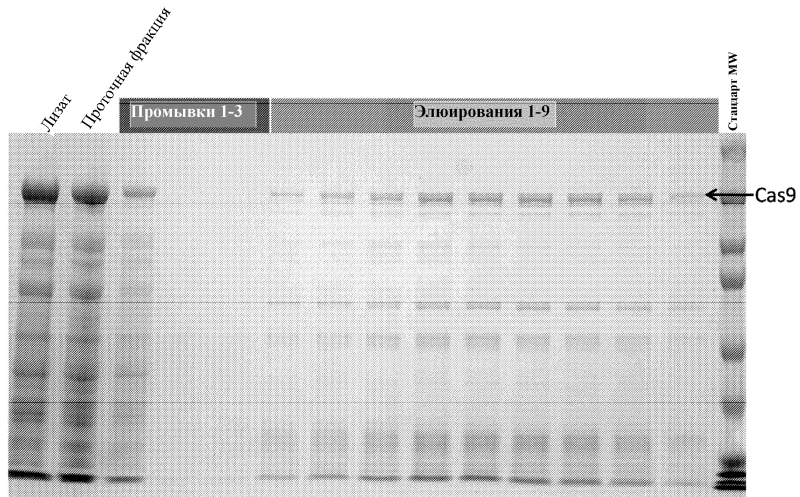
Фиг. 3В



Фиг. 3С

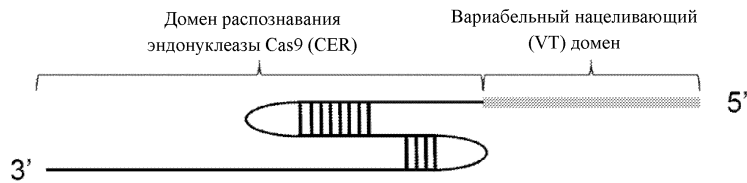


Фиг. 3D

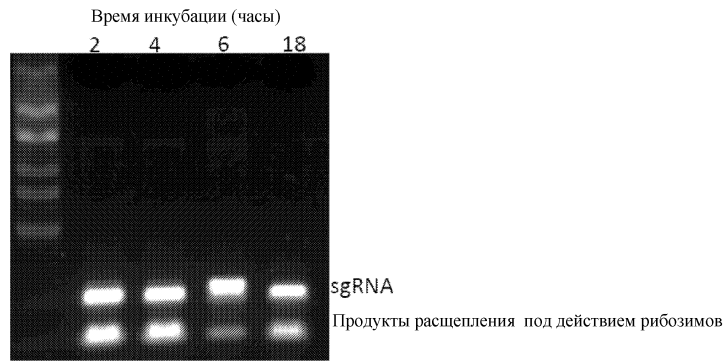


Фиг. 4

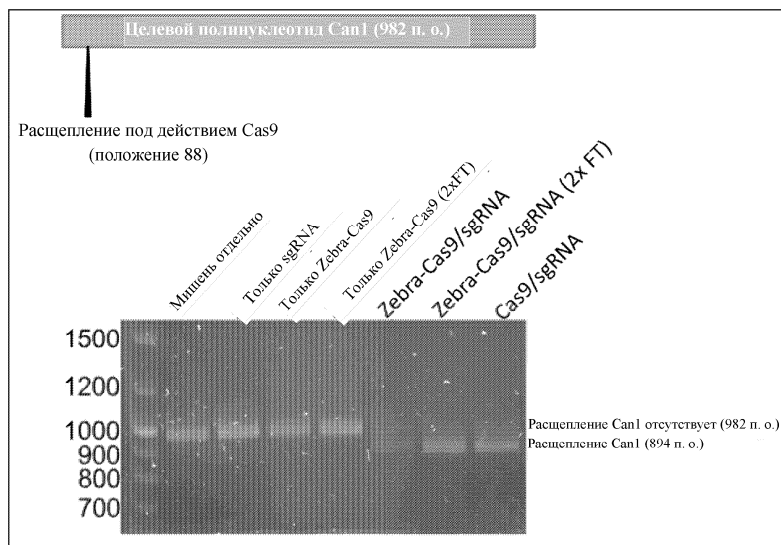
Одиночный направляющий полинуклеотид (например, sgRNA)



Фиг. 5

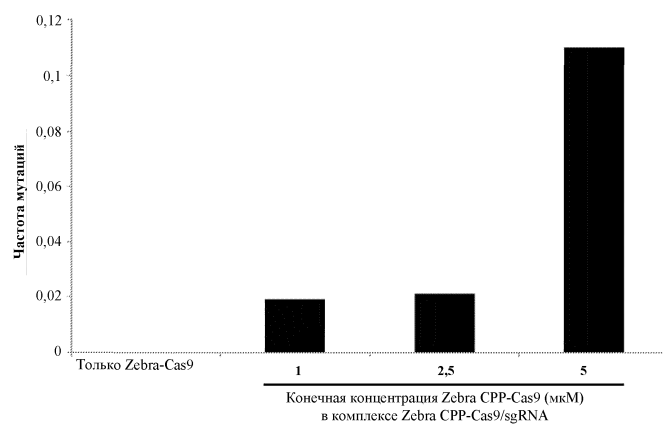


Фиг. 6

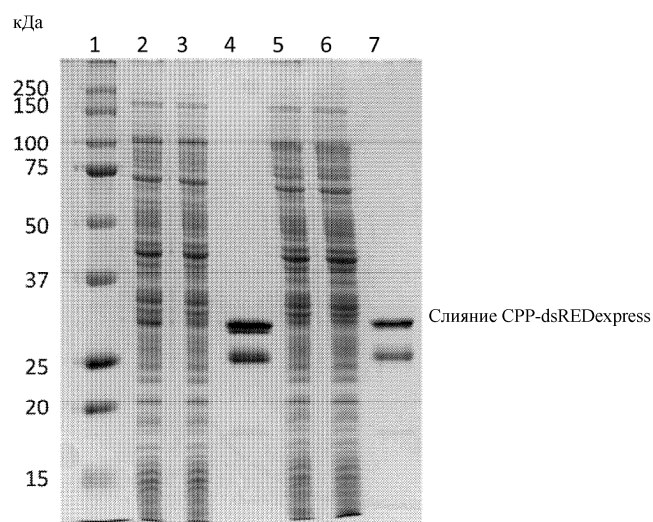


Фиг. 7

038321



Фиг. 8



Фиг. 9