

(19)



**Евразийское  
патентное  
ведомство**

(11) **045278**

(13) **B1**

**(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ**

**(45)** Дата публикации и выдачи патента  
**2023.11.10**

**(21)** Номер заявки  
**201990860**

**(22)** Дата подачи заявки  
**2017.09.28**

**(51)** Int. Cl. *A61K 38/46* (2006.01)  
*A61K 48/00* (2006.01)  
*C12N 9/22* (2006.01)  
*C12N 15/11* (2006.01)  
*C12N 15/55* (2006.01)

---

**(54) РНК-НАПРАВЛЯЕМЫЕ МОДИФИЦИРУЮЩИЕ НУКЛЕИНОВЫЕ КИСЛОТЫ  
ФЕРМЕНТЫ И СПОСОБЫ ИХ ПРИМЕНЕНИЯ**

---

**(31)** 62/402,849

**(32)** 2016.09.30

**(33)** US

**(43)** 2019.10.31

**(86)** PCT/US2017/054047

**(87)** WO 2018/064352 2018.04.05

**(71)(73)** Заявитель и патентовладелец:  
**ТЕ РИДЖЕНТС ОФ ТЕ  
ЮНИВЕРСИТИ ОФ КАЛИФОРНИЯ  
(US)**

**(72)** Изобретатель:  
**Дудна Дженифер А., Бэнфилд  
Джилиан Ф., Бёрнстайн Дэвид,  
Харрингтон Лукас Бенджамин (US)**

**(74)** Представитель:  
**Фелицына С.Б. (RU)**

**(56)** BURSTEIN et al. "New CRISPR-Cas systems from uncultivated microbes," Nature, 22 December 2016 (22.12.2016), Vol. 542, Pgs. 237-241, entire document

US-A1-20160208243

DELTCHEVA et al. "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III," Nature, 31 March 2011 (31.03.2011), Vol. 471, Pgs. 1-19 [Original Pgs. 602-7] entire document

US-A1-20160138008

YANG et al. "New CRISPR-Cas systems discovered," Cell Res. 21 February 2017 (21.02.2017), Vol. 27, Pgs. 313-314, entire document

---

**(57)** В изобретении предложены белки CasY, нуклеиновые кислоты, кодирующие белки CasY, и модифицированные клетки-хозяева, содержащие белки CasY и/или кодирующие их нуклеиновые кислоты. Белки CasY можно использовать во множестве предложенных применений. В изобретении предложены CasY-направляющие РНК, которые связываются с белками CasY и обеспечивают белкам CasX специфичность последовательностей, нуклеиновые кислоты, кодирующие CasY-направляющие РНК; и модифицированные клетки-хозяева, содержащие CasY-направляющие РНК и/или кодирующие их нуклеиновые кислоты. CasY-направляющие РНК можно использовать во множестве применений, которые предложены. В изобретении предложены способы идентификации РНК-направляемой эндонуклеазы CRISPR.

---

**B1**

**045278**

**045278**

**B1**

### Перекрестная ссылка

Данная заявка заявляет приоритет по предварительной заявке на патент США № 62/402849, поданной 30 сентября 2016 г., которая в полном объеме включена в данный документ посредством ссылки.

### Включение посредством ссылки перечня последовательностей, предоставленного в виде текстового файла

Перечень последовательностей предоставлен вместе с данным документом в виде текстового файла "BERK-343WO\_SeqList\_ST25.txt", созданного 28 сентября 2017 г., и имеющего размер 244 кбайт. Сохранение этого текстового файла в полном объеме включено в данный документ посредством ссылки.

### Введение

Система CRISPR-Cas, являющаяся примером пути, который был неизвестен в науке до наступления эры секвенирования ДНК, на сегодняшний день известна как обеспечивающая приобретенный иммунитет против фагов и вирусов бактериям и археям. Интенсивные исследования, проводимые в последнее десятилетие, позволили раскрыть биохимию этой системы. Система CRISPR-Cas состоит из белков Cas, которые вовлечены в обнаружение, нацеливание и расщепление чужеродной ДНК или РНК, и матрицы CRISPR, которая содержит прямые повторы, фланкирующие короткие спейсерные последовательности, которые направляют белки Cas к их мишеням. CRISPR-Cas класса 2 являются упрощенными версиями, в которых один белок Cas, связанный с РНК, отвечает за связывание и расщепление целевой последовательности. Программируемая природа этих минимальных систем сделала возможным их применение в качестве гибкой технологии, революционной в области редактирования генома.

Современные технологии CRISPR-Cas основаны на системах, полученных из культивируемых бактерий, оставляя неохваченными большинство организмов, для которых не проводили выделение. На сегодняшний день было обнаружено только несколько систем CRISPR/Cas класса 2. В данной области техники существует необходимость в дополнительных системах CRISPR/Cas класса 2 (например, комбинациях белков Cas плюс направляющая РНК).

### Сущность изобретения

В настоящем изобретении предложены направляемые РНК эндонуклеазные полипептиды, называемые в данном документе полипептидами "CasY" (также называемые "белками CasY"); нуклеиновые кислоты, кодирующие полипептиды CasY; и модифицированные клетки-хозяева, содержащие полипептиды CasY и/или кодирующие их нуклеиновые кислоты. Полипептиды CasY можно использовать во множестве применений, которые предложены.

В настоящем изобретении предложены направляющие РНК (называемые в данном документе "CasY-направляющими РНК"), которые связываются с белками CasY и обеспечивают белкам CasX специфичность последовательностей; нуклеиновые кислоты, кодирующие CasY-направляющие РНК; и модифицированные клетки-хозяева, содержащие CasY-направляющие РНК и/или кодирующие их нуклеиновые кислоты. CasY-направляющие РНК можно использовать во множестве применений, которые предложены.

В настоящем изобретении предложены способы идентификации РНК-направляемой эндонуклеазы CRISPR.

### Краткое описание графических материалов

На фиг. 1 приведены примеры последовательностей белка CasY природного происхождения.

На фиг. 2 приведено выравнивание последовательностей белков CasY природного происхождения.

На фиг. 3 (панели a-b) приведено схематическое представление доменов CasY. Также приведены результаты различных поисков в целях идентификации гомологов CasY. Также изображены части CasY-содержащих локусов CRISPR, которые были идентифицированы.

На фиг. 4 приведено схематическое изображение локусов CasY и C2c3. Белки интерференции показаны в зеленом цвете, комплекующие белки - в красном. Повторы, собранные посредством РНК-структуры, показаны справа, демонстрируя выраженную шпильку на 5' конце, что позволяет предположить наличие самопроцессинга матрицы CRISPR, обусловленного CasY.

На фиг. 5 (панели a-d) проиллюстрированы эксперименты, проводимые (PAM-зависимая плазмидная интерференция CasY), чтобы определить последовательность PAM для CasY.

На фиг. 6 (панели a-b) представлены "повторные" последовательности CasY-направляющих РНК природного происхождения и пример гибридизации CasY-направляющей РНК с целевой ДНК. (Сверху вниз, SEQ ID NO: 11-15 и 20).

На фиг. 7 (панели a-b) представлены новые идентифицированные системы CRISPR-Cas из некультивируемых организмов, а, соотношение основных линий дифференцировки с выделенными представителями и без них во всех бактериях и археях на основании данных Hug et al.<sup>32</sup>. Результаты отображают огромный масштаб все еще малоизученной биологии в этих доменах. Cas9 архей и новый CRISPR-CasY были обнаружены исключительно в линиях дифференцировки с отсутствием выделенных представителей. b, организация локусов недавно открытых систем CRISPR-Cas.

На фиг. 8 (панели a-b) представлено разнообразие матрицы CRISPR ARMAN-1 и определение PAM-последовательности Cas9 ARMAN-1. a, матрицы CRISPR, реконструированные из 15 разных образцов КШВ. Белые прямоугольники указывают на повторы, а цветные ромбы указывают на спейсеры

(идентичные спейсеры имеют одинаковый цвет; уникальные спейсеры окрашены в черный цвет). Консервативная область матрицы выделена (справа). Разнообразие недавно полученных спейсеров (слева) указывает на активность системы. Анализ, который также включает в себя фрагменты CRISPR из считанных данных, представлен на фиг. 14. b, один предполагаемый вирусный контиг, реконструированный из метагеномных данных КШВ, содержит 56 протоспейсеров (красные вертикальные столбики) из матриц ARMAN-1 CRISPR. c, анализ последовательности позволил выявить консервативный PAM-мотив "NGG" ниже протоспейсеров в нецелевой цепи.

На фиг. 9 (панели a-d) представлены данные, демонстрирующие, что CasX опосредует программируемую ДНК-интерференцию в *E. coli*. a, схема анализа плазмидной интерференции CasX. *E. coli*, экспрессирующую минимальный локус CasX, трансформируют плазмидой, содержащей спейсер, совпадающий с последовательностью в матрице CRISPR (мишень), или плазмидой, содержащей несовпадающий спейсер (немишень). После трансформации культуры высевают и проводят количественную оценку колониеобразующих единиц (КОЕ). b, серийные разведения *E. coli*, экспрессирующей нацеленный на спейсер 1 локус CasX планктомицетов (sX. 1) и трансформированной указанной мишенью (sX1, спейсер 1 CasX; sX2, спейсер 2 CasX; НМ, не-мишень). c, плазмидная интерференция, обусловленная CasX дельта-протеобактерий. Эксперименты проводили в трех повторностях, а результаты приведены как среднее  $\pm$  с.о. d, анализ истощения PAM для локуса CasX планктомицетов, экспрессируемого в *E. coli*. PAM-последовательности, истощенные более чем в 30 раз по сравнению с контрольной библиотекой, использовали для создания WebLogo.

На фиг. 10 (панели a-c) представлены данные, демонстрирующие, что CasX является комплексом CRISPR с двойной направляющей, a, картирование природных РНК-последовательностей (метатранскриптомные данные) относительно локуса CasX CRISPR, схематически изображенного ниже (красная стрелка, предполагаемая *trans*-РНК; белые прямоугольники, повторяющиеся последовательности; зеленые ромбы, спейсерные последовательности). На вставке приведен детализированный вид первого повтора и спейсера. b, схема интерференции двухцепочечной ДНК CasX. Сайт процессинга РНК указан черными стрелками, c, результаты анализа плазмидной интерференции с нокаутом предполагаемой *trans*-РНК в локусе CasX (М, мишень; НМ, не-мишень). Эксперименты проводили в трех повторностях, а результаты приведены как среднее  $\pm$  с.о.

На фиг. 11 (панели a-c) представлены данные, демонстрирующие, что экспрессии локуса CasY в *E. coli* достаточно для ДНК-интерференции, a, схемы локусов CasY и соседних белков, b, WebLogo 5' PAM-последовательностей, истощенных в более чем 3 раза CasY по сравнению с контрольной библиотекой, c, плазмидная интерференция *E. coli*, экспрессирующих CasY.1 и трансформированных мишенями, содержащими указанные PAM. Эксперименты проводили в трех повторностях, а результаты приведены как среднее  $\pm$  с.о.

На фиг. 12 (панели a-b) приведены недавно идентифицированные CRISPR-Cas в контексте известных систем, a, упрощенное филогенетическое дерево универсального белка Cas1. Типы CRISPR известных систем указаны на клиньях и ответвлениях; недавно описанные системы выделены жирным. Подробная филогенетика Cas1 представлена в дополнительных данных 2. b, предполагаемый эволюционный сценарий, давший начало системе типа II архей в результате рекомбинации между локусами типа II-B и типа II-C.

На фиг. 13 показано, что Cas9 архей из ARMAN-4 можно обнаружить в многочисленных контигах с вырожденной матрицей CRISPR. Cas9 из ARMAN-4 выделен темно-красным в 16 разных контигах. Белки с предполагаемыми доменами или функциями помечены, тогда как гипотетические белки оставлены непомеченными. Пятнадцать контигов содержат два вырожденных прямых повтора (несовпадение в одну п.о.) и один консервативный спейсер. Оставшийся контиг содержит только один прямой повтор. В отличие от ARMAN-1 не было обнаружено дополнительных белков Cas, смежных с Cas9 в ARMAN-4.

На фиг. 14 представлена полная реконструкция матриц CRISPR ARMAN-1. Реконструкция матриц CRISPR, которые содержат референсные собранные последовательности, а также сегменты матрицы, реконструированные из коротких ридов ДНК. Зеленые стрелки указывают на повторы, а цветные стрелки указывают на спейсеры CRISPR (идентичные спейсеры имеют одинаковый цвет, тогда как уникальные спейсеры окрашены в черный цвет). В системах CRISPR спейсеры обычно добавляются в одном направлении, поэтому большое разнообразие спейсеров с левой стороны связано с недавним комплектованием.

На фиг. 15 (панели a-b) показано, что спейсеры ARMAN-1 картируются на геномы представителей семейства архей. a, протоспейсеры (красные стрелки) из ARMAN-1 картируются на геном ARMAN-2, наноархей из той же среды. Шесть протоспейсеров однозначно картируются на часть генома, фланкируемую двумя длинными концевыми повторами (ДКП), а два дополнительных протоспейсера идеально совпадают с ДКП (синий и зеленый). Эта область, вероятно, представляет собой транспозон, что позволяет предположить, что система CRISPR-Cas из ARMAN-1 играет роль в подавлении мобилизации этого элемента. b, протоспейсеры также картируются на архею *Thermoplasmatales* (I-плазма), другого представителя экосистемы Richmond Mine, которого можно обнаружить в тех же образцах, что и организмы ARMAN. Протоспейсеры объединяются в кластеры в пределах области генома, кодирующей короткие гипотетические белки, что позволяет предположить, что они также представляют мобильный элемент.

На фиг. 16 (панели а-е) представлена прогнозируемая вторичная структура cr-РНК и tracr-РНК ARMAN-1. а, повтор CRISPR и анти-повтор tracr-РНК показаны в черном цвете, тогда как полученная из спейсера последовательность показана в виде серии зеленых N. Из этого локуса невозможно спрогнозировать четкий сигнал терминации, поэтому исследовали три разные длины tracr-РНК на основании их вторичной структуры - 69, 104 и 179 в красном, синем и розовом цвете, соответственно. b, сконструированная одиночная направляющая РНК, соответствующая двойной направляющей в а. с, двойная направляющая для Cas9 ARMAN-4 с двумя разными шпильками на 3' конце tracr-РНК (75 и 122). d, сконструированная одиночная направляющая РНК, соответствующая двойной направляющей в с.е, условия, исследуемые в *in vivo* анализе нацеливания с *E. coli*.

На фиг. 17 (панели а-б) представлена схема очистки для *in vitro* биохимических исследований, а, Cas9 ARMAN-1 (AR1) и ARMAN-4 (AR4) экспрессировали и очищали в разных условиях, приведенных в дополнительных материалах. Белки, выделенные синими прямоугольниками, исследовали в отношении активности расщепления *in vitro*, b, очищенные фракции AR1-Cas9 и AR4-Cas9 разделяли в 10% ДСН-ПААГ геле.

На фиг. 18 представлены недавно идентифицированные системы CRISPR-Cas в сравнении с известными белками. Сходство CasX и CasY с известными белками на основании следующих поисковых запросов: (1) поиск Blast по неизбыточной (non-redundant - NR) белковой базе данных NCB (2) поиск с помощью скрытой Марковской модели (НММ (от англ. "Hidden markov model")) по базе данных НММ всех известных белков и (3) поиск дальней гомологии с применением HHpred<sup>30</sup>.

На фиг. 19 (панели а-д) представлены данные, относящиеся к программируемой ДНК-интерференции, обусловленной CasX. а, анализ плазмидной интерференции для CasX2 (планктомицеты) и CasX1 (дельта-протеобактерии), продолжение фиг. 9, панель с (sX1, спейсер 1 CasX; sX2, спейсер 2 CasX; НМ, не-мишень). Эксперименты проводили в трех повторностях, а результаты приведены как среднее  $\pm$  с.о. b, серийное разведение *E. coli*, экспрессирующих локус CasX и трансформированных указанной мишенью, продолжение фиг. 9, панель b. с, анализ истощения PAM для CasX дельта-протеобактерий и d, CasX планктомицетов, экспрессируемых в *E. coli*. PAM-последовательности, истощенные более указанного порогового значения истощения PAM (PAM depletion value threshold - PDVT) по сравнению с контрольной библиотекой, использовали для создания WebLogo.

На фиг. 20 представлено эволюционное дерево гомологов Cas9. Максимально правдоподобное филогенетическое дерево белков Cas9, на котором ранее описанные системы окрашены на основании их типа: II-A в синий, II-B в зеленый и II-C в пурпурный цвет. Cas9 архей объединяется в кластер с системами CRISPR-Cas типа II-C вместе с двумя недавно описанными бактериальными Cas9 из некультивируемых бактерий.

На фиг. 21 представлена таблица условий расщепления, проанализированных для Cas9 из ARMAN-1 и ARMAN-4.

#### Определения.

В контексте данного документа "гетерологичный" означает нуклеотидную или полипептидную последовательность, которую нельзя обнаружить в природной нуклеиновой кислоте или природном белке, соответственно. Например, в отношении полипептида CasY гетерологичный полипептид содержит аминокислотную последовательность из белка, отличного от полипептида CasY. В некоторых случаях часть белка CasY от одного вида сливаются с частью белка CasY от другого вида. Следовательно, последовательность CasY от каждого вида может считаться гетерологичной по отношению к другому. В качестве другого примера белок CasY (например, белок dCasY) можно сливать с активным доменом отличного от CasY белка (например, гистон-деацетилазы), а последовательность активного домена может считаться гетерологичным полипептидом (она гетерологична относительно белка CasY).

Термины "полинуклеотид" и "нуклеиновая кислота", взаимозаменяемо употребляемые в данном документе, относятся к полимерной форме нуклеотидов любой длины, как рибонуклеотидов, так и дезоксирибонуклеотидов. Таким образом, этот термин включает в себя, но не ограничивается этим, одно-, двух- или многоцепочечную ДНК или РНК, геномную ДНК, кДНК гибриды ДНК-РНК или полимер, содержащий пуриновые и пиримидиновые основания или другие природные, химически или биохимически модифицированные, неприродные или дериватизированные нуклеотидные основания. Термины "полинуклеотид" и "нуклеиновая кислота" следует понимать как включающие в себя, применительно к описанному варианту реализации, одноцепочечные (такие как смысловые или антисмысловые) и двухцепочечные полинуклеотиды.

Термины "полипептид", "пептид" и "белок" взаимозаменяемо употребляются в данном документе и относятся к полимерной форме аминокислот любой длины, которая может включать в себя генетически кодируемые и негенетически кодируемые аминокислоты, химически или биохимически модифицированные или дериватизированные аминокислоты и полипептиды с модифицированными пептидными остовами. Этот термин включает в себя слитые белки, включая, но не ограничиваясь этим, слитые белки с гетерологичной аминокислотной последовательностью, слияния с гетерологичной и гомологичной лидерными последовательностями, с N-концевыми остатками метионина или без них; иммунологически меченные белки и т.п.

В контексте данного документа термин "природного происхождения" применительно к нуклеиновой кислоте, белку, клетке или организму относится к нуклеиновой кислоте, клетке, белку или организму, которые можно обнаружить в природе.

В контексте данного документа подразумевается, что термин "выделенный" описывает полинуклеотид, полипептид или клетку, которые находятся в окружении, отличном от того, в котором полинуклеотид, полипептид или клетка находятся в природе. Выделенная генетически модифицированная клетка-хозяин может находиться в смешанной популяции генетически модифицированных клеток-хозяев.

В контексте данного документа термин "экзогенная нуклеиновая кислота" относится к нуклеиновой кислоте, которая обычно или в естественном состоянии не присутствует и/или не вырабатывается данными бактерией, организмом или клеткой в природе. В контексте данного документа термин "эндогенная нуклеиновая кислота" относится к нуклеиновой кислоте, которая обычно присутствует и/или вырабатывается данными бактерией, организмом или клеткой в природе. "Эндогенная нуклеиновая кислота" также называется "нативной нуклеиновой кислотой" или нуклеиновой кислотой, которая является "нативной" для данных бактерии, организма или клетки.

В контексте данного документа термин "рекомбинантный" означает, что конкретная нуклеиновая кислота (ДНК или РНК) является продуктом различных комбинаций этапов клонирования, рестрикции и/или лигирования, приводящих к получению конструкции, имеющей структурную кодирующую или некодирующую последовательность, отличную от эндогенных нуклеиновых кислот, которые можно обнаружить в природных системах. В общем случае последовательности ДНК, кодирующие структурную кодирующую последовательность, можно собирать из фрагментов кДНК и коротких олигонуклеотидных линкеров или из ряда синтетических олигонуклеотидов, чтобы получить синтетическую нуклеиновую кислоту, которая может экспрессироваться из рекомбинантной транскрипционной единицы, содержащейся в клетке или в бесклеточной транскрипционной и трансляционной системе. Такие последовательности могут быть предоставлены в форме открытой рамки считывания, не прерываемой внутренними нетранслируемыми последовательностями, или нитронами, которые, как правило, присутствуют в эукариотических генах. Геномную ДНК, содержащую релевантные последовательности, также можно использовать для образования рекомбинантного гена или транскрипционной единицы. Последовательности нетранслируемой ДНК могут находиться 5' или 3' от открытой рамки считывания, где такие последовательности не препятствуют манипуляции или экспрессии кодирующих областей и могут фактически действовать, модулируя выработку необходимого продукта различными механизмами (смотрите "регуляторные последовательности ДНК" ниже).

Таким образом, например, термины "рекомбинантный" полинуклеотид или "рекомбинантная" нуклеиновая кислота относятся к молекулам, которые не существуют в природе, например, получены с помощью искусственной комбинации двух в ином случае разделенных сегментов последовательности посредством вмешательства человека. Искусственную комбинацию часто получают посредством химического синтеза или посредством искусственной манипуляции с выделенными сегментами нуклеиновых кислот, например, с помощью методов генной инженерии. Это обычно делают для замещения кодона избыточным кодоном, кодирующим такую же или консервативную аминокислоту, как правило, внося или удаляя при этом участок распознавания последовательности. В альтернативном варианте это делают для соединения сегментов нуклеиновых кислот с необходимыми функциями для создания необходимой комбинации функций. Искусственную комбинацию часто получают посредством химического синтеза или посредством искусственной манипуляции с выделенными сегментами нуклеиновых кислот, например, с помощью методов генной инженерии.

Аналогично, термин "рекомбинантный" полипептид относится к полипептиду, который не существует в природе, например, получен с помощью искусственной комбинации двух в ином случае разделенных сегментов аминокислотной последовательности посредством вмешательства человека. Таким образом, например, полипептид, который содержит гетерологичную аминокислотную последовательность, является рекомбинантным.

Под "конструкцией" или "вектором" подразумевается рекомбинантная нуклеиновая кислота, в общем случае рекомбинантная ДНК, которая была создана в целях экспрессии и/или размножения конкретной(ых) нуклеотидной(ых) последовательности(ей) или предназначена для использования при конструировании других рекомбинантных нуклеотидных последовательностей.

Термины "регуляторные последовательности ДНК", "контрольные элементы" и "регуляторные элементы", взаимозаменяемо употребляемые в данном документе, относятся к транскрипционным и трансляционным регуляторным последовательностям, таким как промоторы, энхансеры, сигналы полиаденилирования, терминаторы, сигналы деградации белков и т.п., которые обеспечивают и/или регулируют экспрессию кодирующей последовательности и/или выработку кодируемого полипептида в клетке-хозяине.

Термин "трансформация" взаимозаменяемо употребляется в данном документе с термином "генетическая модификация" и относится к перманентному или временному генетическому изменению, индуцированному в клетке после внесения в клетку новой нуклеиновой кислоты (например, ДНК, экзогенной для клетки). Генетическое изменение ("модификацию") можно осуществлять путем включения новой

нуклеиновой кислоты в геном клетки-хозяина или путем временного или стабильного поддержания новой нуклеиновой кислоты в виде эписомального элемента. Если клетка является эукариотической клеткой, перманентное генетическое изменение в общем случае обеспечивается путем внесения новой ДНК в геном клетки. В прокариотических клетках перманентные изменения можно вносить в хромосому или посредством применения внехромосомных элементов, таких как плазмиды и экспрессионные векторы, которые могут содержать один или более селективных маркеров, способствующих их поддержанию в рекомбинантной клетке-хозяине. Подходящие способы генетической модификации включают в себя вирусное инфицирование, трансфекцию, конъюгацию, слияние протопластов, электропорацию, технологию генной пушки, осаждение фосфатом кальция, прямую микроинъекцию и т.п. Выбор способа в общем случае зависит от типа трансформируемой клетки и условий, в которых проходит трансформация (т.е. *in vitro*, *ex vivo* или *in vivo*). Общее обсуждение этих способов можно найти в Ausubel, et al., *Short Protocols in Molecular Biology*, 3<sup>rd</sup> ed., Wiley & Sons, 1995.

Термин "функционально связанный" относится к взаимному расположению, в котором описываемые таким образом компоненты находятся во взаимосвязи, обеспечивающей их функционирование предполагаемым для них образом. Например, промотор функционально связан с кодирующей последовательностью, если промотор влияет на ее транскрипцию или экспрессию. В контексте данного документа термины "гетерологичный промотор" и "гетерологичные регуляторные области" относятся к промоторам и другим регуляторным областям, которые обычно не связаны с данной нуклеиновой кислотой в природе. Например, "транскрипционная регуляторная область, гетерологичная для кодирующей области" представляет собой транскрипционную регуляторную область, которая обычно не связана с кодирующей областью в природе.

В контексте данного документа "клетка-хозяин" означает *in vivo* или *in vitro* эукариотическую клетку, прокариотическую клетку или клетку из многоклеточного организма (например, линии клеток), культивируемую в виде одноклеточного компонента, причем эти эукариотические или прокариотические клетки могут использоваться или использовались в качестве реципиентов для нуклеиновой кислоты (например, экспрессионного вектора) и включают в себя потомство исходной клетки, которая была генетически модифицирована нуклеиновой кислотой. Понятно, что потомство одной клетки не обязательно может быть полностью идентичным по морфологии или по комплементарной последовательности геномной или общей ДНК с исходной родительской клеткой вследствие природных, случайных или умышленных мутаций. "Рекомбинантная клетка-хозяин" (также называемая "генетически модифицированной клеткой-хозяином") представляет собой клетку-хозяина, в которую была внесена гетерологичная нуклеиновая кислота, например, экспрессионный вектор. Например, исследуемая прокариотическая клетка-хозяин представляет собой генетически модифицированную прокариотическую клетку-хозяина (например, бактерию) посредством внесения в подходящую прокариотическую клетку-хозяина гетерологичной нуклеиновой кислоты, например, экзогенной нуклеиновой кислоты, чужеродной (обычно не присутствующей в природе) для прокариотической клетки-хозяина, или рекомбинантной нуклеиновой кислоты, обычно не присутствующей в прокариотической клетке-хозяине; а исследуемая эукариотическая клетка-хозяин представляет собой генетически модифицированную эукариотическую клетку-хозяина посредством внесения в подходящую эукариотическую клетку-хозяина гетерологичной нуклеиновой кислоты, например, экзогенной нуклеиновой кислоты, чужеродной для эукариотической клетки-хозяина, или рекомбинантной нуклеиновой кислоты, обычно не присутствующей в эукариотической клетке-хозяине.

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

Полинуклеотид или полипептид обладает определенным процентом "идентичности последовательности" с другим полинуклеотидом или полипептидом, что означает, что при выравнивании этот процент оснований или аминокислот является одинаковым, и они находятся в одинаковых относительных позициях при сравнении двух последовательностей. Сходство последовательностей можно определять рядом разных способов. Чтобы определить идентичность последовательностей, можно проводить выравнивание последовательностей, используя способы и компьютерные программы, включая BLAST, доступную в сети по адресу [ncbi.nlm.nih.gov/BLAST](http://ncbi.nlm.nih.gov/BLAST). Смотрите, например, Altschul et al. (1990), *J. Mol. Biol.* 215:403-10. Другим алгоритмом выравнивания является FASTA, доступный в пакете компании Genetics Computing Group (GCG) из г. Мэдисон, штат Висконсин, США, стопроцентной дочерней компании Oxford Molecular Group, Inc. Другие методики выравнивания описаны в *Methods in Enzymology*, vol. 266: Computer

Methods for Macromolecular Sequence Analysis (1996), ed. Doolittle, Academic Press, Inc., подразделение Harcourt Brace & Co., Сан Диего, штат Калифорния, США. Особый интерес представляют программы, которые допускают наличие гэпов (пропусков) в последовательности. Одним из типов алгоритмов, который допускает наличие гэпов в выравниваниях последовательностей, является алгоритм Смита-Уотермана. Смотрите Meth. Mol. Biol. 70: 173-187 (1997). Также для выравнивания последовательностей можно использовать программу GAP на основании метода выравнивания Нидлмана и Вунша. Смотрите J. Mol. Biol. 48: 443-453 (1970).

В контексте данного документа термины "лечение", "лечить" и т.п. относятся к получению желаемого фармакологического и/или физиологического эффекта. Эффект может быть профилактическим с точки зрения полного или частичного предотвращения заболевания или его симптомов и/или может быть терапевтическим с точки зрения частичного или полного излечения заболевания и/или неблагоприятного явления, связанного с этим заболеванием. В контексте данного документа термин "лечение" охватывает любое лечение заболевания у млекопитающего, например, у человека, и включает в себя: (а) предотвращение появления заболевания у субъекта, который может быть предрасположен к заболеванию, но у которого это заболевание еще не диагностировано; (б) ингибирование заболевания, т.е. прекращение его развития; и (с) облегчение заболевания, т.е. регрессию заболевания.

Термины "индивид", "субъект", "хозяин" и "пациент", взаимозаменяемо употребляемые в данном документе, относятся к отдельному организму, например, млекопитающему, включая, но не ограничиваясь этим, мышей, обезьян, людей, млекопитающих сельскохозяйственных животных, млекопитающих спортивных животных и млекопитающих домашних животных.

Перед тем как продолжить описание настоящего изобретения, необходимо принять во внимание, что это изобретение не ограничено конкретными описанными вариантами реализации, поскольку они, конечно, могут варьироваться. Следует также понимать, что употребляемая в данном документе терминология применяется только для описания конкретных вариантов реализации и не имеет ограничительного характера, поскольку объем настоящего изобретения ограничен только прилагаемой формулой изобретения.

Когда приводится диапазон значений, следует понимать, что данное изобретение охватывает каждое промежуточное значение с точностью до десятого знака нижнего предела, если в контексте явно не указано иное, между верхним и нижним пределом этого диапазона, а также любое другое указанное или промежуточное значение в таком указанном диапазоне. Верхний и нижний пределы этих меньших диапазонов могут быть независимо включены в меньшие диапазоны и также включены в данное изобретение с учетом любого специальным образом исключенного предела в указанном диапазоне. Если указанный диапазон включает в себя один или оба предела, диапазоны, исключаящие один или оба этих включенных предела, также включены в данное изобретение.

Если не указано иное, все употребляемые в данном документе технические и научные термины имеют общепринятое значение, понятное любому специалисту в данной области техники, к которой имеет отношение настоящее изобретение. Хотя способы и материалы, аналогичные или эквивалентные описанным в данном документе, также можно использовать при практической реализации или испытании настоящего изобретения, далее будут описаны предпочтительные способы и материалы. Все публикации, упомянутые в данном документе, включены в данный документ посредством ссылки с целью раскрытия и описания способов и/или материалов, в связи с которыми цитируются эти публикации.

Следует отметить, что употребление в данном документе и в прилагаемой формуле изобретения форм единственного числа включает в себя отсылку на множественное число, если в контексте явно не указано иное. Таким образом, например, отсылка к "полипептиду CasY" включает в себя множество таких полипептидов, а отсылка к "направляющей РНК" включает в себя отсылку к одной или более направляющим РНК и их эквивалентам, известным специалистам в данной области техники, и т.д. Следует также отметить, что в формулу изобретения могут вноситься правки с целью исключения любых необязательных элементов. Следовательно, данное утверждение должно служить в качестве предварительного основания для использования такой исчерпывающей терминологии, как "исключительно", "только" и т.п. в связи с указанием элементов формулы изобретения, или использования "отрицательного" признака.

Следует понимать, что некоторые особенности данного изобретения, которые для ясности описаны в контексте отдельных вариантов реализации, также могут быть представлены в комбинации в виде одного варианта реализации. И наоборот, различные особенности данного изобретения, которые для краткости описаны в контексте одного варианта реализации, также могут быть представлены отдельно или в любой подходящей подкомбинации. Все комбинации вариантов реализации, которые относятся к данному изобретению, явным образом включены в данное изобретение и описаны в данном документе точно так же, как если бы все без исключения комбинации были отдельно и явным образом описаны. Кроме того, все подкомбинации различных вариантов реализации и их элементов также явным образом включены в данное изобретение и описаны в данном документе точно так же, как если бы все без исключения подкомбинации были отдельно и явным образом описаны.

Обсуждаемые в данном документе публикации представлены исключительно в отношении их содержания до даты подачи настоящей заявки. Ничто в настоящем документе не следует интерпретировать

как признание того, что данное изобретение не имеет права датировать такую публикацию более ранним числом в силу более раннего изобретения. Более того, даты представленных публикаций могут отличаться от фактических дат публикаций, что требует независимого подтверждения.

#### **Подробное описание сущности изобретения**

В настоящем изобретении предложены направляемые РНК эндонуклеазные полипептиды, называемые в данном документе полипептидами "CasY" (также называемые "белками CasY"); нуклеиновые кислоты, кодирующие полипептиды CasY; и модифицированные клетки-хозяева, содержащие полипептиды CasY и/или кодирующие их нуклеиновые кислоты. Полипептиды CasY можно использовать во множестве применений, которые предложены.

В настоящем изобретении предложены направляющие РНК (называемые в данном документе "CasY-направляющими РНК"), которые связываются с белками CasY и обеспечивают белкам CasX специфичность последовательностей; нуклеиновые кислоты, кодирующие CasY-направляющие РНК; и модифицированные клетки-хозяева, содержащие CasY-направляющие РНК и/или кодирующие их нуклеиновые кислоты. CasY-направляющие РНК можно использовать во множестве применений, которые предложены.

В настоящем изобретении предложены способы идентификации РНК-направляемой эндонуклеазы CRISPR.

Композиции.

Белки CRISPR/CasY и направляющие РНК.

CRISPR/Cas-эндонуклеаза (например, белок CasY) взаимодействует с (связывается с) соответствующей направляющей РНК (например, CasY-направляющей РНК) с образованием рибонуклеопротеинового (РНП) комплекса, который нацелен на конкретный сайт в целевой нуклеиновой кислоте посредством спаривания оснований между направляющей РНК и целевой последовательностью в целевой молекуле нуклеиновой кислоты. Направляющая РНК содержит нуклеотидную последовательность (направляющую последовательность), комплементарную последовательности (целевому сайту) целевой нуклеиновой кислоты. Таким образом, белок CasY образует комплекс с CasY-направляющей РНК, а направляющая РНК обеспечивает специфичность последовательности РНП-комплекса за счет направляющей последовательности. Белок CasY комплекса обеспечивает сайт-специфическую активность. Другими словами, белок CasY направляется к целевому сайту (например, стабилизируется в целевом сайте) в последовательности целевой нуклеиновой кислоты (например, хромосомной последовательности или внехромосомной последовательности, например, эписомальной последовательности, миникольцевой последовательности, митохондриальной последовательности, хлоропластной последовательности и т.д.) за счет его ассоциации с направляющей РНК.

В настоящем изобретении предложены композиции, содержащие полипептид CasY (и/или нуклеиновую кислоту, кодирующую полипептид CasY) (например, полипептид CasY может представлять собой существующий в природе белок, уникальный белок CasY, белок dCasY, химерный белок CasY и т.д.). В настоящем изобретении предложены композиции, содержащие CasY-направляющую РНК (и/или нуклеиновую кислоту, кодирующую CasY-направляющую РНК). В настоящем изобретении предложены композиции, содержащие (а) полипептид CasY (и/или нуклеиновую кислоту, кодирующую полипептид CasY) (например, полипептид CasY может представлять собой существующий в природе белок, уникальный белок CasY, белок dCasY, химерный белок CasY и т.д.) и (b) CasY-направляющую РНК (и/или нуклеиновую кислоту, кодирующую CasY-направляющую РНК). В настоящем изобретении предложен комплекс нуклеиновая кислота/белок (РНП-комплекс), содержащий: (а) полипептид CasY согласно настоящему изобретению (например, полипептид CasY может представлять собой существующий в природе белок, уникальный белок CasY, белок dCasY, химерный белок CasY и т.д.) и (b) CasY-направляющую РНК.

Белок CasY.

Полипептид CasY (этот термин взаимозаменяемо употребляется с термином "белок CasY") может связывать и/или модифицировать (например, расщеплять, создавать одноцепочечный разрыв ("ник"), метилировать, деметилировать и т.д.) целевую нуклеиновую кислоту и/или полипептид, связанный с целевой нуклеиновой кислотой (например, метилирование или ацетилирование гистоновых хвостов) (например, в некоторых случаях белок CasY имеет партнера по слиянию с определенной активностью, а в некоторых случаях сам белок CasY обеспечивает нуклеазную активность). В некоторых случаях белок CasY является белком природного происхождения (например, встречается в природе в прокариотических клетках). В других случаях белок CasY не является белком природного происхождения (например, белок CasY представляет собой вариантный белок CasY, химерный белок и т.п.).

Методы анализа для определения того, взаимодействует ли данный белок с CasY-направляющей РНК, могут представлять собой любые удобные методы анализа связывания, в которых исследуется связывание между белком и нуклеиновой кислотой. Подходящие методы анализа связывания (например, анализ сдвига в геле) известны специалистам в данной области техники (например, методы анализа, которые включают в себя добавление CasY-направляющей РНК и белка к целевой нуклеиновой кислоте). Методы анализа для определения того, имеет ли белок определенную активность (например, для определения того, имеет ли белок нуклеазную активность, которая состоит в расщеплении целевой нуклеино-

вой кислоты, и/или некоторую гетерологичную активность), могут представлять собой любой удобный метод анализа (например, любой удобный метод анализа расщепления нуклеиновых кислот, в котором исследуется расщепление нуклеиновых кислот). Подходящие методы анализа (например, анализ расщепления) известны специалисту в данной области техники.

Белок CasY природного происхождения функционирует как эндонуклеаза, которая катализирует двухцепочечный разрыв в конкретной последовательности в целевой двухцепочечной ДНК (дцДНК). Специфичность последовательности обеспечивается ассоциированной направляющей РНК, которая гибридизируется с целевой последовательностью в целевой ДНК. CasY-направляющая РНК природного происхождения представляет собой сг-РНК, при этом сг-РНК содержит (i) направляющую последовательность, которая гибридизируется с целевой последовательностью в целевой ДНК, и (ii) белок-связывающий сегмент, который содержит структуру типа "стебель-петля" (шпилька - двойная спираль дцРНК), которая связывается с белком CasY.

В некоторых вариантах реализации изобретения белок CasY согласно предложенным способам и/или композициям представляет собой белок природного происхождения (дикого типа) (или получен из него). Примеры белков CasY природного происхождения проиллюстрированы на фиг. 1 и приведены как SEQ ID NO: 1-7. Примеры белков CasY природного происхождения проиллюстрированы на фиг. 1 и приведены как SEQ ID NO: 1-8. Выравнивание типовых белков CasY природного происхождения представлено на фиг. 2 (белки обозначены как "Y1.", "Y2.", "Y3." и т.д.). Частичные ДНК-каркасы 7 локусов CRISPR CasY природного происхождения (собранные по данным секвенирования) приведены как SEQ ID NO: 21-27. Важно отметить, что этот недавно открытый белок (CasY) является коротким по сравнению с ранее идентифицированными эндонуклеазами CRISPR-Cas, и, следовательно, применение этого белка в качестве альтернативы обеспечивает преимущество, состоящее в том, что нуклеотидная последовательность, кодирующая белок, является относительно короткой. Это целесообразно, например, в случаях, когда необходима нуклеиновая кислота, кодирующая белок CasY, например, в ситуациях, в которых применяется вирусный вектор (например, ААВ-вектор) для доставки в клетку, такую как эукариотическая клетка (например, клетка млекопитающего, клетка человека, клетка мыши, *in vitro*, *ex vivo*, *in vivo*), для исследовательских и/или клинических применений. Также в данном документе отмечается, что бактерии, имеющие локусы CasY CRISPR, присутствовали в образцах окружающей среды, которые были собраны при низкой температуре (например, 10-17°C). Таким образом, ожидается, что CasY сможет функционировать при низких температурах (например, 10-14°C, 10-17°C, 10-20°C) (например, лучше, чем другие эндонуклеазы Cas, известные на данный момент).

В некоторых случаях белок CasY (согласно предложенным композициям и/или способам) содержит аминокислотную последовательность, имеющую 20% или более идентичности последовательности (например, 30% или более, 40% или более, 50% или более, 60% или более, 70% или более, 80% или более, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с последовательностью белка CasY, приведенной как SEQ ID NO: 1. Например, в некоторых случаях белок CasY содержит аминокислотную последовательность, имеющую 50% или более идентичности последовательности (например, 60% или более, 70% или более, 80% или более, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с последовательностью белка CasY, приведенной как SEQ ID NO: 1. В некоторых случаях белок CasY содержит аминокислотную последовательность, имеющую 80% или более идентичности последовательности (например, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с последовательностью белка CasY, приведенной как SEQ ID NO: 1. В некоторых случаях белок CasY содержит аминокислотную последовательность, имеющую 90% или более идентичности последовательности (например, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с последовательностью белка CasY, приведенной как SEQ ID NO: 1. В некоторых случаях белок CasY содержит аминокислотную последовательность, имеющую последовательность белка CasY, приведенную как SEQ ID NO: 1. В некоторых случаях белок CasY содержит аминокислотную последовательность, имеющую последовательность белка CasY, приведенную как SEQ ID NO: 1, за исключением того, что последовательность содержит аминокислотную замену (например, 1, 2 или 3 аминокислотные замены), которая снижает природную каталитическую активность белка (например, в аминокислотных позициях, описанных ниже).

В некоторых случаях белок CasY содержит аминокислотную последовательность, имеющую 20% или более идентичности последовательности (например, 30% или более, 40% или более, 50% или более, 60% или более, 70% или более, 80% или более, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с последовательностью белка CasY, приведенной как SEQ ID NO: 2. В некоторых случаях белок CasY содержит аминокислотную последовательность, имеющую 50% или более идентичности последовательности (например, 60% или более, 70% или более, 80% или более, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с последовательностью белка CasY, приведенной как SEQ ID NO: 2. В некоторых случаях белок CasY содержит аминокис-









### Белковые домены CasY.

Домены белка CasY проиллюстрированы на фиг. 3. Как видно на схематическом представлении на фиг. 3 (аминокислоты пронумерованы на основании белка CasY1 (SEQ ID NO: 1)), белок CasY содержит N-концевой домен длиной приблизительно 800-1000 аминокислот (например, около 815 для CasY1 и около 980 для CasY5) и C-концевой домен, который содержит 3 частичных домена RuvC (RuvC-I, RuvC-II и RuvC-III также называемые в данном документе субдоменами), которые не являются непрерывными по отношению к первичной аминокислотной последовательности белка CasY, но образуют домен RuvC после выработки и сворачивания белка. Таким образом, в некоторых случаях белок CasY (согласно предложенным композициям и/или способам) содержит аминокислотную последовательность с N-концевым доменом (например, не включая какую-либо слитую гетерологичную последовательность, такую как СЯЛ, и/или домен с каталитической активностью), имеющую длину в диапазоне от 750 до 1050 аминокислот (например, от 750 до 1025, от 750 до 1000, от 750 до 950, от 775 до 1050, от 775 до 1025, от 775 до 1000, от 775 до 950, от 800 до 1050, от 800 до 1025, от 800 до 1000 или от 800 до 950 аминокислот). В некоторых случаях белок CasY (согласно предложенным композициям и/или способам) содержит аминокислотную последовательность, имеющую длину (например, не включая какую-либо слитую гетерологичную последовательность, такую как СЯЛ, и/или домен с каталитической активностью) в диапазоне от 750 до 1050 аминокислот (например, от 750 до 1025, от 750 до 1000, от 750 до 950, от 775 до 1050, от 775 до 1025, от 775 до 1000, от 775 до 950, от 800 до 1050, от 800 до 1025, от 800 до 1000 или от 800 до 950 аминокислот), которая является N-концевой относительно расщепленного домена RuvC (например, 3 частичных доменов RuvC-I, RuvC-II и RuvC-III).

В некоторых случаях белок CasY (согласно предложенным композициям и/или способам) содержит аминокислотную последовательность, имеющую 20% или более идентичности последовательности (например, 30% или более, 40% или более, 50% или более, 60% или более, 70% или более, 80% или более, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с N-концевым доменом (например, доменом, проиллюстрированным в виде аминокислот 1-812 для CasY1 на фиг. 3, панель а) последовательности белка CasY, приведенной как SEQ ID NO: 1. Например, в некоторых случаях белок CasY содержит аминокислотную последовательность, имеющую 50% или более идентичности последовательности (например, 60% или более, 70% или более, 80% или более, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с N-концевым доменом (например, доменом, проиллюстрированным в виде аминокислот 1-812 для CasY1 на фиг. 3, панель а) последовательности белка CasY, приведенной как SEQ ID NO: 1. В некоторых случаях белок CasY содержит аминокислотную последовательность, имеющую 80% или более идентичности последовательности (например, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с N-концевым доменом (например, доменом, проиллюстрированным в виде аминокислот 1-812 для CasY1 на фиг. 3, панель а) последовательности белка CasY, приведенной как SEQ ID NO: 1. В некоторых случаях белок CasY содержит аминокислотную последовательность, имеющую 90% или более идентичности последовательности (например, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с N-концевым доменом (например, доменом, проиллюстрированным в виде аминокислот 1-812 для CasY1 на фиг. 3, панель а) последовательности белка CasY, приведенной как SEQ ID NO: 1. В некоторых случаях белок CasY содержит аминокислотную последовательность, содержащую аминокислоты 1-812 последовательности белка CasY, приведенной как SEQ ID NO: 1.

В некоторых случаях белок CasY (согласно предложенным композициям и/или способам) содержит аминокислотную последовательность, имеющую 20% или более идентичности последовательности (например, 30% или более, 40% или более, 50% или более, 60% или более, 70% или более, 80% или более, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с N-концевым доменом (например, доменом, проиллюстрированным в виде аминокислот 1-812 для CasY1 на фиг. 3, панель а) последовательности белка CasY, приведенной в любой из SEQ ID NO: 1-4. Например, в некоторых случаях белок CasY содержит аминокислотную последовательность, имеющую 50% или более идентичности последовательности (например, 60% или более, 70% или более, 80% или более, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с N-концевым доменом (например, доменом, проиллюстрированным в виде аминокислот 1-812 для CasY1 на фиг. 3, панель а) последовательности белка CasY, приведенной в любой из SEQ ID NO: 1-4. В некоторых случаях белок CasY содержит аминокислотную последовательность, имеющую 80% или более идентичности последовательности (например, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с N-концевым доменом (например, доменом, проиллюстрированным в виде аминокислот 1-812 для CasY1 на фиг. 3, панель а) последовательности белка CasY, приведенной в любой из SEQ ID NO: 1-4. В некоторых случаях белок CasY содержит аминокислотную последовательность, имеющую 90% или более идентичности последовательности (например, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с N-концевым доменом (например, доменом, проиллюстрированным в виде аминокислот 1-812 для CasY1 на фиг. 3, панель а) последовательности белка CasY, приведенной в любой из SEQ ID NO: 1-4. В некоторых случаях белок CasY содержит аминокислотную последовательность, имеющую 90% или более идентичности последовательности (например, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с N-концевым доменом (например, доменом, проиллюстрированным в виде аминокислот 1-812 для CasY1 на фиг. 3, панель а) последовательности белка CasY, приведенной в любой из SEQ ID NO: 1-4.







100% идентичности последовательности) с N-концевым доменом (например, доменом, проиллюстрированным в виде аминокислот 1-812 для CasY1 на фиг. 3, панель а) любой из последовательностей белка CasY, приведенных как SEQ ID NO: 1-8; и вторую аминокислотную последовательность, С-концевую относительно первой аминокислотной последовательности, которая содержит расщепленный домен RuvC (например, 3 частичных домена RuvC-RuvC-RuvC-II и RuvC-III). Например, в некоторых случаях белок CasY содержит первую аминокислотную последовательность, имеющую 50% или более идентичности последовательности (например, 60% или более, 70% или более, 80% или более, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с N-концевым доменом (например, доменом, проиллюстрированным в виде аминокислот 1-812 для CasY1 на фиг. 3, панель а) любой из последовательностей белка CasY, приведенных как SEQ ID NO: 1-8; и вторую аминокислотную последовательность, С-концевую относительно первой аминокислотной последовательности, которая содержит расщепленный домен RuvC (например, 3 частичных домена RuvC-RuvC-RuvC-II и RuvC-III). В некоторых случаях белок CasY содержит первую аминокислотную последовательность, имеющую 80% или более идентичности последовательности (например, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с N-концевым доменом (например, доменом, проиллюстрированным в виде аминокислот 1-812 для CasY1 на фиг. 3, панель а) любой из последовательностей белка CasY, приведенных как SEQ ID NO: 1-8; и вторую аминокислотную последовательность, С-концевую относительно первой аминокислотной последовательности, которая содержит расщепленный домен RuvC (например, 3 частичных домена RuvC-RuvC-RuvC-II и RuvC-III). В некоторых случаях белок CasY содержит первую аминокислотную последовательность, имеющую 90% или более идентичности последовательности (например, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с N-концевым доменом (например, доменом, проиллюстрированным в виде аминокислот 1-812 для CasY1 на фиг. 3, панель а) любой из последовательностей белка CasY, приведенных как SEQ ID NO: 1-8; и вторую аминокислотную последовательность, С-концевую относительно первой аминокислотной последовательности, которая содержит расщепленный домен RuvC (например, 3 частичных домена RuvC-RuvC-RuvC-II и RuvC-III). В некоторых случаях белок CasY содержит аминокислотную последовательность, соответствующую аминокислотам 1-812 последовательности белка CasY, приведенной как SEQ ID NO: 1; и вторую аминокислотную последовательность, С-концевую относительно первой аминокислотной последовательности, которая содержит расщепленный домен RuvC (например, 3 частичных домена RuvC-RuvC-RuvC-II и RuvC-III).

В некоторых вариантах реализации расщепленный домен RuvC белка CasY (согласно предложенным композициям и/или способам) содержит область между субдоменами RuvC-II и RuvC-III, которая больше субдомена RuvC-III. Например, в некоторых случаях соотношение длины области между субдоменами RuvC-II и RuvC-III и длины субдомена RuvC-III составляет 1,1 или более (например, 1,2). В некоторых случаях соотношение длины области между субдоменами RuvC-II и RuvC-III и длины субдомена RuvC-III превышает 1. В некоторых случаях соотношение длины области между субдоменами RuvC-II и RuvC-III и длины субдомена RuvC-III превышает 1 и составляет от 1 до 1,5 (например, от 1 до 1,4, от 1 до 1,3 или от 1 до 1,2).

В некоторых вариантах реализации (для белка CasY согласно предложенным композициям и/или способам) соотношение длины субдомена RuvC-II и длины субдомена RuvC-III составляет 2 или менее (например, 1,8 или менее, 1,7 или менее, 1,6 или менее, 1,5 или менее или 1,4 или менее). Например, в некоторых случаях соотношение длины субдомена RuvC-II и длины субдомена RuvC-III составляет 1,5 или менее (например, 1,4 или менее). В некоторых вариантах реализации соотношение длины субдомена RuvC-II и длины субдомена RuvC-III находится в диапазоне от 1 до 2 (например, от 1,1 до 2, от 1,2 до 2, от 1 до 1,8, от 1,1 до 1,8, от 1,2 до 1,8, от 1 до 1,6, от 1,1 до 1,6, от 1,2 до 1,6, от 1 до 1,4, от 1,1 до 1,4 или от 1,2 до 1,4).

В некоторых случаях (для белка CasY согласно предложенным композициям и/или способам) соотношение длины области между субдоменами RuvC-II и RuvC-III и длины субдомена RuvC-III превышает 1. В некоторых случаях соотношение длины области между субдоменами RuvC-II и RuvC-III и длины субдомена RuvC-III превышает 1 и составляет от 1 до 1,3 (например, от 1 до 1,2).

В некоторых случаях (для белка CasY согласно предложенным композициям и/или способам) область между субдоменами RuvC-II и RuvC-III имеет длину по меньшей мере 60 аминокислот (например, длину по меньшей мере 65, 68 или 70 аминокислот). В некоторых случаях область между субдоменами RuvC-II и RuvC-III имеет длину в диапазоне 60-110 аминокислот (например, в диапазоне 60-105, 60-100, 60-95, 60-90, 65-110, 65-105, 65-100, 65-95 или 65-90 аминокислот).

В некоторых случаях белок CasY (согласно предложенным композициям и/или способам) содержит первую аминокислотную последовательность, имеющую 50% или более идентичности последовательности (например, 60% или более, 70% или более, 80% или более, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с N-концевым доменом (например, доменом, проиллюстрированным в виде аминокислот 1-812 для CasY1 на фиг. 3, панель а) любой из последовательностей белка CasY, приведенных как SEQ ID NO: 1-4; и вторую









от 1 до 1,5 (например, от 1 до 1,2); (ix) область между субдоменами RuvC-II и RuvC-III имеет длину по меньшей мере 60 аминокислот (например, длину по меньшей мере 65 или по меньшей мере 70 аминокислот); (x) область между субдоменами RuvC-II и RuvC-III имеет длину по меньшей мере 65 аминокислот; (xi) область между субдоменами RuvC-II и RuvC-III имеет длину в диапазоне 60-110 аминокислот (например, в диапазоне 60-105, 60-100, 60-95, 60-90, 65-110, 65-105, 65-100, 65-95 или 65-90 аминокислот) или (xii) область между субдоменами RuvC-II и RuvC-III имеет длину в диапазоне 65-95 аминокислот.

В некоторых случаях белок CasY (согласно предложенным композициям и/или способам) содержит первую аминокислотную последовательность, имеющую 75% или более идентичности последовательности (например, 80% или более, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с N-концевым доменом (например, доменом, проиллюстрированным в виде аминокислот 1-812 для CasY1 на фиг. 3, панель а) любой из последовательностей белка CasY, приведенных как SEQ ID NO: 1-8; и вторую аминокислотную последовательность, С-концевую относительно первой аминокислотной последовательности, которая содержит 3 частичных домена RuvC-RuvC-RuvC-II и RuvC-III - где: (i) соотношение длины области между субдоменами RuvC-II и RuvC-III и длины субдомена RuvC-III составляет 1,1 или более (например, 1,2); (ii) соотношение длины области между субдоменами RuvC-II и RuvC-III и длины субдомена RuvC-III превышает 1; (iii) соотношение длины области между субдоменами RuvC-II и RuvC-III и длины субдомена RuvC-III превышает 1 и составляет от 1 до 1,5 (например, от 1 до 1,4, от 1 до 1,3, от 1 до 1,2); (iv) соотношение длины субдомена RuvC-II и длины субдомена RuvC-III составляет 2 или менее (например, 1,8 или менее, 1,7 или менее, 1,6 или менее, 1,5 или менее или 1,4 или менее); (v) соотношение длины субдомена RuvC-II и длины субдомена RuvC-III составляет 1,5 или менее (например, 1,4 или менее); (vi) соотношение длины субдомена RuvC-II и длины субдомена RuvC-III находится в диапазоне от 1 до 2 (например, от 1,1 до 2, от 1,2 до 2, от 1 до 1,8, от 1,1 до 1,8, от 1,2 до 1,8, от 1 до 1,6, от 1,1 до 1,6, от 1,2 до 1,6, от 1 до 1,4, от 1,1 до 1,4 или от 1,2 до 1,4); (vii) соотношение длины области между субдоменами RuvC-II и RuvC-III и длины субдомена RuvC-III превышает 1; (viii) соотношение длины области между субдоменами RuvC-II и RuvC-III и длины субдомена RuvC-III превышает 1 и составляет от 1 до 1,5 (например, от 1 до 1,2); (ix) область между субдоменами RuvC-II и RuvC-III имеет длину по меньшей мере 60 аминокислот (например, длину по меньшей мере 65 или по меньшей мере 70 аминокислот); (x) область между субдоменами RuvC-II и RuvC-III имеет длину по меньшей мере 65 аминокислот; (xi) область между субдоменами RuvC-II и RuvC-III имеет длину в диапазоне 60-110 аминокислот (например, в диапазоне 60-105, 60-100, 60-95, 60-90, 65-110, 65-105, 65-100, 65-95 или 65-90 аминокислот) или (xii) область между субдоменами RuvC-II и RuvC-III имеет длину в диапазоне 65-95 аминокислот.

В некоторых случаях белок CasY (согласно предложенным композициям и/или способам) содержит первую аминокислотную последовательность, имеющую 85% или более идентичности последовательности (например, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с N-концевым доменом (например, доменом, проиллюстрированным в виде аминокислот 1-812 для CasY1 на фиг. 3, панель а) любой из последовательностей белка CasY, приведенных как SEQ ID NO: 1-8; и вторую аминокислотную последовательность, С-концевую относительно первой аминокислотной последовательности, которая содержит 3 частичных домена RuvC-RuvC-RuvC-II и RuvC-III - где: (i) соотношение длины области между субдоменами RuvC-II и RuvC-III и длины субдомена RuvC-III составляет 1,1 или более (например, 1,2); (ii) соотношение длины области между субдоменами RuvC-II и RuvC-III и длины субдомена RuvC-III превышает 1; (iii) соотношение длины области между субдоменами RuvC-II и RuvC-III и длины субдомена RuvC-III превышает 1 и составляет от 1 до 1,5 (например, от 1 до 1,4, от 1 до 1,3, от 1 до 1,2); (iv) соотношение длины субдомена RuvC-II и длины субдомена RuvC-III составляет 2 или менее (например, 1,8 или менее, 1,7 или менее, 1,6 или менее, 1,5 или менее или 1,4 или менее); (v) соотношение длины субдомена RuvC-II и длины субдомена RuvC-III составляет 1,5 или менее (например, 1,4 или менее); (vi) соотношение длины субдомена RuvC-II и длины субдомена RuvC-III находится в диапазоне от 1 до 2 (например, от 1,1 до 2, от 1,2 до 2, от 1 до 1,8, от 1,1 до 1,8, от 1,2 до 1,8, от 1 до 1,6, от 1,1 до 1,6, от 1,2 до 1,6, от 1 до 1,4, от 1,1 до 1,4 или от 1,2 до 1,4); (vii) соотношение длины области между субдоменами RuvC-II и RuvC-III и длины субдомена RuvC-III превышает 1; (viii) соотношение длины области между субдоменами RuvC-II и RuvC-III и длины субдомена RuvC-III превышает 1 и составляет от 1 до 1,5 (например, от 1 до 1,2); (ix) область между субдоменами RuvC-II и RuvC-III имеет длину по меньшей мере 60 аминокислот (например, длину по меньшей мере 65 или по меньшей мере 70 аминокислот); (x) область между субдоменами RuvC-II и RuvC-III имеет длину по меньшей мере 65 аминокислот; (xi) область между субдоменами RuvC-II и RuvC-III имеет длину в диапазоне 60-110 аминокислот (например, в диапазоне 60-105, 60-100, 60-95, 60-90, 65-110, 65-105, 65-100, 65-95 или 65-90 аминокислот) или (xii) область между субдоменами RuvC-II и RuvC-III имеет длину в диапазоне 65-95 аминокислот.

В некоторых случаях белок CasY (согласно предложенным композициям и/или способам) содержит первую аминокислотную последовательность с N-концевым доменом (например, не включая какую-либо слитую гетерологичную последовательность, такую как СЯЛ, и/или домен с каталитической активностью), имеющую длину в диапазоне от 750 до 1050 аминокислот (например, от 750 до 1025, от 750 до

1000, от 750 до 950, от 775 до 1050, от 775 до 1025, от 775 до 1000, от 775 до 950, от 800 до 1050, от 800 до 1025, от 800 до 1000 или от 800 до 950 аминокислот); и вторую аминокислотную последовательность (С-концевую относительно первой), содержащую расщепленный домен RuvC с 3 частичными доменами RuvC-RuvC-RuvC-II и RuvC-III где: (i) соотношение длины области между субдоменами RuvC-II и RuvC-III и длины субдомена RuvC-III составляет 1,1 или более (например, 1,2); (ii) соотношение длины области между субдоменами RuvC-II и RuvC-III и длины субдомена RuvC-III превышает 1; (iii) соотношение длины области между субдоменами RuvC-II и RuvC-III и длины субдомена RuvC-III превышает 1 и составляет от 1 до 1,5 (например, от 1 до 1,4, от 1 до 1,3, от 1 до 1,2); (iv) соотношение длины субдомена RuvC-II и длины субдомена RuvC-III составляет 2 или менее (например, 1,8 или менее, 1,7 или менее, 1,6 или менее, 1,5 или менее или 1,4 или менее); (v) соотношение длины субдомена RuvC-II и длины субдомена RuvC-III составляет 1,5 или менее (например, 1,4 или менее); (vi) соотношение длины субдомена RuvC-II и длины субдомена RuvC-III находится в диапазоне от 1 до 2 (например, от 1,1 до 2, от 1,2 до 2, от 1 до 1,8, от 1,1 до 1,8, от 1,2 до 1,8, от 1 до 1,6, от 1,1 до 1,6, от 1,2 до 1,6, от 1 до 1,4, от 1,1 до 1,4 или от 1,2 до 1,4); (vii) соотношение длины области между субдоменами RuvC-II и RuvC-III и длины субдомена RuvC-III превышает 1; (viii) соотношение длины области между субдоменами RuvC-II и RuvC-III и длины субдомена RuvC-III превышает 1 и составляет от 1 до 1,5 (например, от 1 до 1,2); (ix) область между субдоменами RuvC-II и RuvC-III имеет длину по меньшей мере 60 аминокислот (например, длину по меньшей мере 65 или по меньшей мере 70 аминокислот); (x) область между субдоменами RuvC-II и RuvC-III имеет длину по меньшей мере 65 аминокислот; (xi) область между субдоменами RuvC-II и RuvC-III имеет длину в диапазоне 60-110 аминокислот (например, в диапазоне 60-105, 60-100, 60-95, 60-90, 65-110, 65-105, 65-100, 65-95 или 65-90 аминокислот) или (xii) область между субдоменами RuvC-II и RuvC-III имеет длину в диапазоне 65-95 аминокислот.

В некоторых случаях белок CasY (согласно предложенным композициям и/или способам) содержит аминокислотную последовательность, имеющую 20% или более идентичности последовательности (например, 30% или более, 40% или более, 50% или более, 60% или более, 70% или более, 80% или более, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с С-концевым доменом (например, доменом, проиллюстрированным в виде аминокислот 812-1125 для CasY1 на фиг. 3, панель а) последовательности белка CasY, приведенной как SEQ ID NO: 1. Например, в некоторых случаях белок CasY содержит аминокислотную последовательность, имеющую 50% или более идентичности последовательности (например, 60% или более, 70% или более, 80% или более, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с С-концевым доменом (например, доменом, проиллюстрированным в виде аминокислот 812-1125 для CasY1 на фиг. 3, панель а) последовательности белка CasY, приведенной как SEQ ID NO: 1. В некоторых случаях белок CasY содержит аминокислотную последовательность, имеющую 80% или более идентичности последовательности (например, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с С-концевым доменом (например, доменом, проиллюстрированным в виде аминокислот 812-1125 для CasY1 на фиг. 3, панель а) последовательности белка CasY, приведенной как SEQ ID NO: 1. В некоторых случаях белок CasY содержит аминокислотную последовательность, имеющую 90% или более идентичности последовательности (например, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с С-концевым доменом (например, доменом, проиллюстрированным в виде аминокислот 812-1125 для CasY1 на фиг. 3, панель а) последовательности белка CasY, приведенной как SEQ ID NO: 1. В некоторых случаях белок CasY содержит аминокислотную последовательность, содержащую аминокислоты 812-1125 последовательности белка CasY, приведенной как SEQ ID NO: 1.

В некоторых случаях белок CasY (согласно предложенным композициям и/или способам) содержит аминокислотную последовательность, имеющую 20% или более идентичности последовательности (например, 30% или более, 40% или более, 50% или более, 60% или более, 70% или более, 80% или более, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с С-концевым доменом (например, доменом, проиллюстрированным в виде аминокислот 812-1125 для CasY1 на фиг. 3, панель а) любой из последовательностей белка CasY, приведенных как SEQ ID NO: 1-4. Например, в некоторых случаях белок CasY содержит аминокислотную последовательность, имеющую 50% или более идентичности последовательности (например, 60% или более, 70% или более, 80% или более, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с С-концевым доменом (например, доменом, проиллюстрированным в виде аминокислот 812-1125 для CasY1 на фиг. 3, панель а) любой из последовательностей белка CasY, приведенных как SEQ ID NO: 1-4. В некоторых случаях белок CasY содержит аминокислотную последовательность, имеющую 80% или более идентичности последовательности (например, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с С-концевым доменом (например, доменом, проиллюстрированным в виде аминокислот 812-1125 для CasY1 на фиг. 3, панель а) любой из последовательностей белка CasY, приведенных как SEQ ID NO: 1-4. В некоторых случаях белок CasY









как СЯЛ, и/или домен с каталитической активностью), имеющую длину в диапазоне от 750 до 1050 аминокислот (например, от 750 до 1025, от 750 до 1000, от 750 до 950, от 775 до 1050, от 775 до 1025, от 775 до 1000, от 775 до 950, от 800 до 1050, от 800 до 1025, от 800 до 1000 или от 800 до 950 аминокислот); и вторую аминокислотную последовательность, расположенную С-терминально относительно первой аминокислотной последовательности, имеющую 50% или более идентичности последовательности (например, 60% или более, 70% или более, 80% или более, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с С-концевым доменом (например, доменом, проиллюстрированным в виде аминокислот 812-1125 для CasY1 на фиг. 3, панель а) любой из последовательностей белка CasY, приведенных как SEQ ID NO: 1-8. В некоторых случаях белок CasY содержит первую аминокислотную последовательность (N-концевой домен) (например, не включая какую-либо слитую гетерологичную последовательность, такую как СЯЛ, и/или домен с каталитической активностью), имеющую длину в диапазоне от 750 до 1050 аминокислот (например, от 750 до 1025, от 750 до 1000, от 750 до 950, от 775 до 1050, от 775 до 1025, от 775 до 1000, от 775 до 950, от 800 до 1050, от 800 до 1025, от 800 до 1000 или от 800 до 950 аминокислот); и вторую аминокислотную последовательность, расположенную С-терминально относительно первой аминокислотной последовательности, имеющую 80% или более идентичности последовательности (например, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с С-концевым доменом (например, доменом, проиллюстрированным в виде аминокислот 812-1125 для CasY1 на фиг. 3, панель а) любой из последовательностей белка CasY, приведенных как SEQ ID NO: 1-8. В некоторых случаях белок CasY содержит первую аминокислотную последовательность (N-концевой домен) (например, не включая какую-либо слитую гетерологичную последовательность, такую как СЯЛ, и/или домен с каталитической активностью), имеющую длину в диапазоне от 750 до 1050 аминокислот (например, от 750 до 1025, от 750 до 1000, от 750 до 950, от 775 до 1050, от 775 до 1025, от 775 до 1000, от 775 до 950, от 800 до 1050, от 800 до 1025, от 800 до 1000 или от 800 до 950 аминокислот); и вторую аминокислотную последовательность, расположенную С-терминально относительно первой аминокислотной последовательности, имеющую 90% или более идентичности последовательности (например, 95% или более, 97% или более, 98% или более, 99% или более или 100% идентичности последовательности) с С-концевым доменом (например, доменом, проиллюстрированным в виде аминокислот 812-1125 для CasY1 на фиг. 3, панель а) любой из последовательностей белка CasY, приведенных как SEQ ID NO: 1-8. В некоторых случаях белок CasY содержит первую аминокислотную последовательность (N-концевой домен) (например, не включая какую-либо слитую гетерологичную последовательность, такую как СЯЛ, и/или домен с каталитической активностью), имеющую длину в диапазоне от 750 до 1050 аминокислот (например, от 750 до 1025, от 750 до 1000, от 750 до 950, от 775 до 1050, от 775 до 1025, от 775 до 1000, от 775 до 950, от 800 до 1050, от 800 до 1025, от 800 до 1000 или от 800 до 950 аминокислот); и вторую аминокислотную последовательность, расположенную С-терминально относительно первой аминокислотной последовательности, содержащую фрагмент аминокислотной последовательности любой из последовательностей белка CasY, приведенных как SEQ ID NO: 1-8, которая соответствует аминокислотам 812-1125 последовательности белка CasY, приведенной как SEQ ID NO: 1.

Варианты CasY.

Вариантный белок CasY имеет аминокислотную последовательность, которая отличается по меньшей мере одной аминокислотой (например, имеет делецию, вставку, замену, слияние) по сравнению с аминокислотной последовательностью соответствующего белка CasY дикого типа. Белок CasY, который расщепляет только одну цепь двухцепочечной целевой нуклеиновой кислоты, называется в данном документе "никазой" (например, "никазой CasY"). Белок CasY, который практически не обладает нуклеазной активностью, называется в данном документе мертвым белком CasY ("dCasY" от англ. "dead CasY") (с оговоркой, что нуклеазная активность может обеспечиваться гетерологичным полипептидом - партнером по слиянию - в случае химерного белка CasY, который более подробно описан ниже). В случае любых описанных в данном документе вариантных белков CasY (например, никазы CasY, dCasY, химерного CasY), вариант CasY может содержать последовательность белка CasY с такими же параметрами, что описаны выше (например, присутствующие домены, процент идентичности и т.п.).

Варианты - каталитическая активность.

В некоторых случаях белок CasY представляет собой вариантный белок CasY, например, мутировавший относительно каталитически активной последовательности природного происхождения, и проявляет сниженную активность расщепления (например, проявляет 90% или менее, 80% или менее, 70% или менее, 60% или менее, 50% или менее, 40% или менее или 30% или менее активности расщепления) по сравнению с соответствующей последовательностью природного происхождения. В некоторых случаях такой вариантный белок CasY является каталитически "мертвым" белком (практически не обладает активностью расщепления) и может называться "dCasY". В некоторых случаях вариантный белок CasY представляет собой никазу (расщепляет только одну цепь двухцепочечной целевой нуклеиновой кислоты, например, двухцепочечной целевой ДНК). Как более подробно описано в данном документе, в некоторых случаях белок CasY (в некоторых случаях белок CasY с активностью расщепления дикого типа и в

некоторых случаях вариантный CasY со сниженной активностью расщепления, например, dCasY или никаза CasY) слит (конъюгирован) с гетерологичным полипептидом, обладающим представляющей интерес активностью (например, представляющей интерес каталитической активностью), с образованием слитого белка (химерного белка CasY).

Каталитические остатки CasY включают в себя D828, E914, D1074 при нумерации в соответствии с CasY1 (SEQ ID NO: 1) (эти остатки подчеркнуты на фиг. 1 для SEQ ID NO: 1), (смотрите, например, выравнивание на фиг. 2, панели a и b).

Таким образом, в некоторых случаях белок CasY имеет сниженную активность, а одна или более из вышеописанных аминокислот (или одна или более соответствующих аминокислот любого белка CasY) являются мутировавшими (например, с замещением аланином). В некоторых случаях вариантный белок CasY является каталитически "мертвым" белком (является каталитически неактивным) и называется "dCasY". Белок dCasY может быть слит с партнером по слиянию, который обеспечивает активность, и, в некоторых случаях, dCasY (например, один, без партнера по слиянию, который обеспечивает каталитическую активность, но который может содержать СЯЛ при экспрессии в эукариотической клетке) может связываться с целевой ДНК и может блокировать трансляцию РНК-полимеразы из целевой ДНК. В некоторых случаях вариантный белок CasY представляет собой никазу (расщепляет только одну цепь двухцепочечной целевой нуклеиновой кислоты, например, двухцепочечной целевой ДНК).

Варианты - химерный CasY (т.е. слитые белки).

Как отмечено выше, в некоторых случаях белок CasY (в некоторых случаях белок CasY с активностью расщепления дикого типа и в некоторых случаях вариантный CasY со сниженной активностью расщепления, например, dCasY или никаза CasY) слит (конъюгирован) с гетерологичным полипептидом, обладающим представляющей интерес активностью (например, представляющей интерес каталитической активностью), с образованием слитого белка (химерного белка CasY). Гетерологичный полипептид, с которым может быть слит белок CasY, называется в данном документе "партнером по слиянию".

В некоторых случаях партнер по слиянию может модулировать транскрипцию (например, ингибировать транскрипцию, повышать транскрипцию) целевой ДНК. Например, в некоторых случаях партнер по слиянию представляет собой белок (или домен из белка), который ингибирует транскрипцию (например, репрессор транскрипции, белок, который функционирует путем рекрутирования белков-ингибиторов транскрипции, модификации целевой ДНК, такой как метилирование, рекрутирования модификаторов ДНК, модуляции гистонов, связанных с целевой ДНК, рекрутирования модификаторов гистонов, таких как те, которые модифицируют ацетилирование и/или метилирование гистонов, и т.п.). В некоторых случаях партнер по слиянию представляет собой белок (или домен из белка), который повышает транскрипцию (например, активатор транскрипции, белок, который действует путем рекрутирования белков-активаторов транскрипции, модификации целевой ДНК, такой как деметилирование, рекрутирования модификаторов ДНК, модуляции гистонов, связанных с целевой ДНК, рекрутирования модификаторов гистонов, таких как те, которые модифицируют ацетилирование и/или метилирование гистонов, и т.п.).

В некоторых случаях химерный белок CasY содержит гетерологичный полипептид, который обладает ферментативной активностью, которая модифицирует целевую нуклеиновую кислоту (например, нуклеазной активностью, метилтрансферазной активностью, деметилазной активностью, активностью ДНК-репарации, ДНК-повреждающей активностью, деаминирующей активностью, дисмутазной активностью, алкилирующей активностью, депурицирующей активностью, окислительной активностью, активностью образования пиримидиновых димеров, интегразной активностью, транспозазной активностью, рекомбиназной активностью, полимеразной активностью, лигазной активностью, геликазной активностью, фотолиазной активностью или гликозилазной активностью).

В некоторых случаях химерный белок CasY содержит гетерологичный полипептид, который обладает ферментативной активностью, которая модифицирует полипептид (например, гистон), связанный с целевой нуклеиновой кислотой (например, метилтрансферазной активностью, деметилазной активностью, ацетилтрансферазной активностью, деацетилазной активностью, киназной активностью, фосфатазной активностью, убиквитинлигазной активностью, деубиквитинизирующей активностью, аденилирующей активностью, деаденилирующей активностью, сумоилирующей активностью, десумоилирующей активностью, рибозилирующей активностью, дерибозилирующей активностью, миристоилирующей активностью или демиристоилирующей активностью).

Примеры белков (или их фрагментов), которые можно использовать для повышения транскрипции, включают в себя, но не ограничиваются этим: активаторы транскрипции, такие как VP16, VP64, VP48, VP160, субдомен р65 (например, из NFκB) и домен активации EDLL и/или активации TAL (например, для активности в растениях); гистоновые лизин-метилтрансферазы, такие как SET1A, SET1B, MLL1-5, ASH1, SYMD2, NSD1 и т.п.; гистоновые лизин-деметилазы, такие как JHDM2a/b, UTX, JMJD3 и т.п.; гистоновые ацетилтрансферазы, такие как GCN5, PCAF, CBP, p300, TAF1, TIP60/PLIP, MOZ/MYST3, MORF/MYST4, SRC1, ACTR, P160, CLOCK и т.п.; и ДНК-деметилазы, такие как TET (Ten-Eleven Translocation)-диоксигеназа 1 (TET1CD), TET1, DME, DML1, DML2, ROS1 и т.п.

Примеры белков (или их фрагментов), которые можно использовать для снижения транскрипции, включают в себя, но не ограничиваются этим: репрессоры транскрипции, такие как Kriiprel-

ассоциированный бокс (KRAB или SKD); домен репрессии KOX1; домен взаимодействия Mad mSIN3 (SID); домен-репрессор ERF (ERD), домен репрессии SRDX (например, для репрессии в растениях) и т.п.; гистоновые лизин-метилтрансферазы, такие как Pr-SET7/8, SUV4-20H1, RIZ1 и т.п.; гистоновые лизин-деметилазы, такие как JMJD2A/JHDM3A, JMJD2B, JMJD2C/GASC1, JMJD2D, JARID1A/RBP2, JARID1B/PLU-1, JARID1C/SMCX, JARID1D/SMCY и т.п.; гистоновые лизин-деацетилазы, такие как HDAC1, HDAC2, HDAC3, HDAC8, HDAC4, HDAC5, HDAC7, HDAC9, SIRT1, SIRT2, HDAC11 и т.п.; ДНК-метилазы, такие как HhaI ДНК-m5c-метилтрансфераза (M.HhaI), ДНК-метилтрансфераза 1 (DNMT1), ДНК-метилтрансфераза 3a (DNMT3a), ДНК-метилтрансфераза 3b (DNMT3b), MET DRM3 (растения), ZMET2, CMT1, CMT2 (растения) и т.п.; и периферические элементы рекрутирования, такие как ламин А, ламин В и т.п.

В некоторых случаях партнер по слиянию обладает ферментативной активностью, которая модифицирует целевую нуклеиновую кислоту (например, оцРНК, дцРНК, оцДНК, дцДНК). Примеры ферментативной активности, которую может обеспечивать партнер по слиянию, включают в себя, но не ограничиваются этим: нуклеазную активность, такую, как обеспечивается рестрикционным ферментом (например, нуклеазой FokI), метилтрансферазную активность, такую, как обеспечивается метилтрансферазой (например, HhaI ДНК-m5c-метилтрансферазой (M.HhaI), ДНК-метилтрансферазой 1 (DNMT1), ДНК-метилтрансферазой 3a (DNMT3a), ДНК-метилтрансферазой 3b (DNMT3b), MET DRM3 (растения), ZMET2, CMT1, CMT2 (растения) и т.п.); деметилазную активность, такую, как обеспечивается деметилазой (например, TET (Ten-Eleven Translocation)-диоксигеназой 1 (TET1CD), TET1, DME, DML1, DML2, ROS1 и т.п.), активность ДНК-репарации, ДНК-повреждающую активность, деаминирующую активность, такую, как обеспечивается деаминазой (например, ферментом цитозиндеаминазой, таким как крысиный APOBEC1), дисмутазную активность, алкилирующую активность, депурицирующую активность, окислительную активность, активность образования пиримидиновых димеров, интегразную активность, такую, как обеспечивается интегразой и/или резольвазой (например, Gin-инвертазой, такой как гиперактивный мутант Gin-инвертазы, GinH106Y; интегразой вируса иммунодефицита человека типа 1 (IN); Tn3-резольвазой и т.п.), транспозазную активность, рекомбиназную активность, такую, как обеспечивается рекомбиназой (например, рекомбиназой каталитического домена Gin), полимеразную активность, лигазную активность, геликазную активность, фотолиазную активность и гликозилазную активность).

В некоторых случаях партнер по слиянию обладает ферментативной активностью, которая модифицирует белок, связанный с целевой нуклеиновой кислотой (например, оцРНК, дцРНК, оцДНК, дцДНК) (например, гистон, РНК-связывающий белок, ДНК-связывающий белок и т.п.). Примеры ферментативной активности (которая модифицирует белок, связанный с целевой нуклеиновой кислотой), которую может обеспечивать партнер по слиянию, включают в себя, но не ограничиваются этим: метилтрансферазную активность, такую, как обеспечивается гистоновой метилтрансферазой (HMT) (например, гомологом 1 супрессора мозаицизма 3-9 (SUV39H1, также известным как KMT1A), эухроматической гистоновой лизин-метилтрансферазой 2 (G9A, также известной как KMT1C и EHMT2), SUV39H2, ESET/SETDB1 и т.п., SET1A, SET1B, MLL1-5, ASH1, SYMD2, NSD1, DOT1L, Pr-SET7/8, SUV4-20H1, EZH2, RIZ1), деметилазную активность, такую, как обеспечивается гистоновой деметилазой (например, лизин-деметилазой 1A (KDM1A, также известной как LSD1), JHDM2a/b, JMJD2A/JHDM3A, JMJD2B, JMJD2C/GASC1, JMJD2D, JARID1A/RBP2, JARID1B/PLU-1, JARID1C/SMCX, JARID1D/SMCY, UTX, JMJD3 и т.п.), ацетилтрансферазную активность, такую, как обеспечивается гистоновой ацетилтрансферазой (например, каталитическим ядром/фрагментом человеческой ацетилтрансферазы p300, GCN5, PCAF, CBP, TAF1, TIP60/PLIP, MOZ/MYST3, MORF/MYST4, HBO1/MYST2, HMOF/MYST1, SRC1, ACTR, P160, CLOCK и т.п.), деацетилазную активность, такую, как обеспечивается а гистоновой деацетилазой (например, HDAC1, HDAC2, HDAC3, HDAC8, HDAC4, HDAC5, HDAC7, HDAC9, SIRT1, SIRT2, HDAC11 и т.п.), киназную активность, фосфатазную активность, убиквитинлигазную активность, деубиквитинизирующую активность, аденилирующую активность, деаденилирующую активность, сумоилирующую активность, десумоилирующую активность, рибозилирующую активность, дерибозилирующую активность, миристоилирующую активность и демиростоилирующую активность.

Дополнительными примерами подходящих партнеров по слиянию являются домен дестабилизации дигидрофолатредуктазы (DHFR) (например, для создания химически регулируемого белка CasY) и транзитный пептид хлоропластов. Подходящие транзитные пептиды хлоропластов включают в себя, но не ограничиваются этим:

MASMISSAVTTVSRASRGQSAAMAPFGGLKSMTGFPVRKVNTDITSITSNGGRVKCMQ  
VWPPIGKKKFETLSYLPPLTRDSRA (SEQ ID NO:83);

MASMISSAVTTVSRASRGQSAAMAPFGGLKSMTGFPVRKVNTDITSITSNGGRVKS  
(SEQ ID NO:84);

MASSMLSSATMVASPAQATMVAPFNGLKSSAAFPATRKANNDITSITSNGGRVNCMQV  
WPPIEKKKFETLSYLPDLTDSGGRVNC (SEQ ID NO:85);

MAQVSRICNGVQNPISLISNLSKSSQRKSPLSVSLKTQQHPRAYPISSSWGLKKSGMTLIGS  
ELRPLKVMSSVSTAC (SEQ ID NO:86);

MAQVSRICNGVWNPISLISNLSKSSQRKSPLSVSLKTQQHPRAYPISSSWGLKKSGMTLIG  
SELRPLKVMSSVSTAC (SEQ ID NO:87);

MAQINNMAQGIQTLNPNNSNFHKPQVPKSSSFLVFGSKKLKNSANSMLVLKKSIFMQLF  
CSFRISASVATAC (SEQ ID NO:88);

MAALVTSQLATSGTVLSVTDRFRPGFQGLRPRNPADAALGMRTVGASAAPKQSRKPH  
RFDRRCLSMVV (SEQ ID NO:89);

MAALTSQLATSATGFGIADRSAPSSLLRHGFQGLKPRSPAGGDATSLSVTTSARATPKQ  
QRSVQRGSRRFPSVVVC (SEQ ID NO:90);

MASSVLSSAAVATRSNVAQANMVAPFTGLKSAASFPVSRKQNLDTIASNGGRVQC  
(SEQ ID NO:91);

MESLAATSVFAPSRVAVPAARALVRAGTVVPTRRTSSTSGTSGVKCSAAVTPQASPVISR  
SAAAA (SEQ ID NO:92) и

MGAAATSMQSLKFSNRLVPPSRRLSPVNNVTCNNLPKSAAPVRTVKCCASSWNSTING  
AAATTNGASAASS (SEQ ID NO:93).

В некоторых случаях слитый полипептид CasY согласно настоящему изобретению содержит: а) полипептид CasY согласно настоящему изобретению и б) транзитный пептид хлоропластов. Таким образом, например, комплекс CRISPR-CasY может быть нацелен на хлоропласты. В некоторых случаях это нацеливание можно обеспечивать за счет наличия N-концевого удлинения, называемого транзитным пептидом хлоропластов (ТПХ) или пластидным транзитным пептидом. Хромосомные трансгены из бактериальных источников должны иметь последовательность, кодирующую последовательность ТПХ, слитую с последовательностью, кодирующей экспрессируемый полипептид, в случае, если экспрессируемый полипептид необходимо локализовать в растительной пластиде (например, хлоропласте). Соответственно, локализацию экзогенного полипептида в хлоропласте часто осуществляют посредством функционального связывания полинуклеотидной последовательности, кодирующей последовательность ТПХ, с 5' областью полинуклеотида, кодирующего экзогенный полипептид. ТПХ удаляется на этапе процессинга во время транслокации в пластиду. На эффективность процессинга при этом может влиять аминокислотная последовательность ТПХ и близлежащие последовательности в NH<sub>2</sub>-конце пептида. Другими вариантами для нацеливания на хлоропласта, которые были описаны, являются сигнальная последовательность *cab-m7* кукурузы (патент США № 7022896, WO 97/41228), сигнальная последовательность глутатионредуктазы гороха (WO 97/41228) и ТПХ, описанные в US2009029861.

В некоторых случаях слитый полипептид CasY согласно настоящему изобретению может содержать: а) полипептид CasY согласно настоящему изобретению и б) пептид эндосомального высвобождения. В некоторых случаях полипептид эндосомального высвобождения содержит аминокислотную последовательность GLFXALLXLLXSLWXLXLLXA (SEQ ID NO:94), где каждый X независимо выбран из лизина, гистидина и аргинина. В некоторых случаях полипептид эндосомального высвобождения содержит аминокислотную последовательность GLFHALLHLLHSLWHLLLHA (SEQ ID NO:95).

Примеры некоторых вышеперечисленных партнеров по слиянию (и не только), используемых в контексте слияния с белками Cas9, цинковыми пальцами и/или TALE (для сайт-специфической модификации целевой нуклеиновой кислоты, модуляции транскрипции и/или модификации целевого белка, например, модификации гистонов), смотрите, например, в: Nomura et al., J Am Chem Soc. 2007 Jul 18; 129(28):8676-7; Rivenbark et al., Epigenetics. 2012 Apr; 7(4):350-60; Nucleic Acids Res. 2016 Jul 8;

44(12):5615-28; Gilbert et al., *Cell*. 2013 Jul 18; 154(2):442-51; Kearns et al., *Nat Methods*. 2015 May; 12(5):401-3; Mendenhall et al., *Nat Biotechnol*. 2013 Dec; 31(12): 1133-6; Hilton et al., *Nat Biotechnol*. 2015 May; 33(5):510-7; Gordley et al., *Proc Natl Acad Sci USA*. 2009 Mar 31; 106(13):5053-8; Akopian et al., *Proc Natl Acad Sci USA*. 2003 Jul 22; 100(15):8688-91; Tan et al., *J Virol*. 2006 Feb; 80(4): 1939-48; Tan et al., *Proc Natl Acad Sci USA*. 2003 Oct 14; 100(21):11997-2002; Papworth et al., *Proc Natl Acad Sci USA*. 2003 Feb 18; 100(4):1621-6; Sanjana et al., *Nat Protoc*. 2012 Jan 5; 7(1): 171-92; Beerli et al., *Proc Natl Acad Sci U S A*. 1998 Dec 8; 95(25):14628-33; Snowden et al., *Curr Biol*. 2002 Dec 23; 12(24):2159-66; Xu et al., *Cell Discov*. 2016 May 3; 2:16009; Komor et al., *Nature*. 2016 Apr 20; 533(7603):420-4; Chaikind et al., *Nucleic Acids Res*. 2016 Aug 11; Choudhury et al., *Oncotarget*. 2016 Jun 23; Du et al., *Cold Spring Harb Protoc*. 2016 Jan 4; Pham et al., *Methods Mol Biol*. 2016; 1358:43-57; Balboa et al., *Stem Cell Reports*. 2015 Sep 8; 5(3):448-59; Hara et al., *Sci Rep*. 2015 Jun 9; 5:11221; Piatek et al., *Plant Biotechnol J*. 2015 May; 13(4):578-89; Hu et al., *Nucleic Acids Res*. 2014 Apr; 42(7):4375-90; Cheng et al., *Cell Res*. 2013 Oct; 23(10):1163-71 и Maeder et al., *Nat Methods*. 2013 Oct; 10(10):977-9.

Дополнительные подходящие гетерологические полипептиды включают в себя, но не ограничиваются этим, полипептиды, которые прямо и/или опосредованно обеспечивают повышение транскрипции и/или трансляции целевой нуклеиновой кислоты (например, активаторы транскрипции или их фрагменты, белки или их фрагменты, которые рекрутируют активаторы транскрипции, низкомолекулярные/восприимчивые к лекарственным препаратам регуляторы транскрипции и/или трансляции, белки, регулирующие трансляцию и т.д.). Неограничивающие примеры гетерологических полипептидов для осуществления повышения или снижения транскрипции включают в себя домены активаторов транскрипции и репрессоров транскрипции. В некоторых таких случаях химерный полипептид CasY нацеливается направляющей нуклеиновой кислотой (направляющей РНК) на конкретную локацию (т.е. последовательность) в целевой нуклеиновой кислоте и осуществляет locus-специфическую регуляцию, такую как блокирование связывания РНК-полимеразы с промотором (что избирательно ингибирует функцию активации транскрипции), и/или модификация локального хроматинового статуса (например, когда используется слитая последовательность, которая модифицирует целевую нуклеиновую кислоту или модифицирует полипептид, связанный с целевой нуклеиновой кислотой). В некоторых случаях изменения являются временными (например, репрессия или активация транскрипции). В некоторых случаях изменения являются наследуемыми (например, когда осуществляют эпигенетические модификации в целевой нуклеиновой кислоте или в белках, связанных с целевой нуклеиновой кислотой, например, нуклеосомных гистонах).

Неограничивающие примеры гетерологических полипептидов для применения при нацеливании на целевую нуклеиновую кислоту, такую как оцРНК, включают в себя (но не ограничиваются этим): факторы сплайсинга (например, домены RS); компоненты белковой трансляции (например, факторы инициации трансляции, элонгации и/или высвобождения; например, eIF4G); РНК-метилазы; РНК-редактирующие ферменты (например, РНК-деаминазы, например, аденозин-деаминазу, действующую на РНК (ADAR), включая редактирующие ферменты A-I и/или C-U); геликазы; РНК-связывающие белки и т.п. Следует понимать, что гетерологический полипептид может включать в себя целый белок или в некоторых случаях может включать в себя фрагмент белка (например, функциональный домен).

Гетерологический полипептид предложенного химерного полипептида CasY может представлять собой любой домен, способный взаимодействовать с оцРНК (которая в контексте этого изобретения включает в себя внутримолекулярные и/или межмолекулярные вторичные структуры, например, двойные спирали двухцепочечной РНК, такие как шпильки, структуры типа "стебель-петля" и т.д.), временно или необратимо, прямо или непрямо, включая, но не ограничиваясь этим, эффекторный домен, выбранный из группы, содержащей: эндонуклеазы (например, РНКазу II домен CRR22 DYW, дайсер и домены PIN (N-конец PiIT) из белков, таких как SMG5 и SMG6); белки и белковые домены, ответственные за стимуляцию расщепления РНК (например, CPSF, CstF, CFIm и CFIIIm); экзонуклеазы (например, XRN-1 или экзонуклеаза T); deadенилазы (например, HNT3); белки и белковые домены, ответственные за нонсенс-опосредованное разрушение РНК (например, UPF1, UPF2, UPF3, UPF3b, RNP S1, Y14, DEK, REF2 и SRm160); белки и белковые домены, ответственные за стабилизацию РНК (например, PABP); белки и белковые домены, ответственные за репрессию трансляции (например, Ago2 и Ago4); белки и белковые домены, ответственные за стимуляцию трансляции (например, Staufen); белки и белковые домены, ответственные за (например, способные осуществлять) модуляцию трансляции (например, факторы трансляции, такие как факторы инициации, факторы элонгации, факторы высвобождения и т.д., например, eIF4G); белки и белковые домены, ответственные за полиаденилирование РНК (например, PAP1, GLD-2 и Star-PAP); белки и белковые домены, ответственные за полиуридилрование РНК (например, CI D1 и терминальная уридилат-трансфераза); белки и белковые домены, ответственные за локализацию РНК (например, из IMP1, ZBP1, She2p, She3p и Bicaudal-D); белки и белковые домены, ответственные за ядерное удержание РНК (например, Rgr6); белки и белковые домены, ответственные за ядерный экспорт РНК (например, TAP, NXF1, THO, TREX, REF и Aly); белки и белковые домены, ответственные за репрессию сплайсинга РНК (например, PTB, Sam68 и hnRNP A1); белки и белковые домены, ответственные за стимуляцию сплайсинга РНК (например, богатые серином/аргином домены (SR)); белки и бел-

ковые домены, ответственные за снижение эффективности транскрипции (например, FUS (TLS)); и белки и белковые домены, ответственные за стимуляцию транскрипции (например, CDK7 и Tat ВИЧ). В альтернативном варианте эффекторный домен может быть выбран из группы, содержащей эндонуклеазы; белки и белковые домены, способные стимулировать расщепление РНК; экзонуклеазы; деаденилазы; белки и белковые домены, имеющие активность нонсенс-опосредованного разрушения РНК; белки и белковые домены, способные стабилизировать РНК; белки и белковые домены, способные репрессировать трансляцию; белки и белковые домены, способные стимулировать трансляцию; белки и белковые домены, способные модулировать трансляцию (например, факторы трансляции, такие как факторы инициации, факторы элонгации, факторы высвобождения и т.д., например, eIF4G); белки и белковые домены, способные осуществлять полиаденилирование РНК; белки и белковые домены, способные осуществлять полиуридилрование РНК; белки и белковые домены, имеющие активность локализации РНК; белки и белковые домены, способные осуществлять ядерное удержание РНК; белки и белковые домены, имеющие активность ядерного экспорта РНК; белки и белковые домены, способные репрессировать сплайсинг РНК; белки и белковые домены, способные стимулировать сплайсинг РНК; белки и белковые домены, способные снижать эффективность транскрипции; и белки и белковые домены, способные стимулировать транскрипцию. Другим подходящим гетерологичным полипептидом является РНК-связывающий домен PUF, который более подробно описан в WO 2012068627, которая в полном объеме включена в данный документ посредством ссылки.

Некоторые факторы сплайсинга РНК, которые можно использовать (целыми или в виде фрагментов) в качестве гетерологичных полипептидов для химерного полипептида CasY, имеют модульную организацию с отдельными специфическими к последовательности РНК-связывающими модулями и сплайсинг-эффекторными доменами. Например, представители семейства богатых серином/аргинином (SR) белков содержат N-концевые мотивы распознавания РНК (MPP), которые связываются с экзонными энхансерами сплайсинга (ЭЭС) в пре-мРНК, и C-концевые RS-домены, которые стимулируют включение экзона. В качестве другого примера hnRNP-белок hnRNP A1 связывается с экзонными энхансерами сплайсинга (ЭЭС) посредством MPP-доменов и ингибирует включение экзона посредством C-концевого богатого глицином домена. Некоторые факторы сплайсинга могут регулировать альтернативное использование сайта сплайсинга (сс) посредством связывания с регуляторными последовательностями между двумя альтернативными сайтами. Например, ASF/SF2 может распознавать ЭЭС и стимулировать использование проксимальных интронных сайтов, тогда как hnRNP A1 может связываться с ЭЭС и сдвигать сплайсинг в направлении использования дистальных интронных сайтов. Одним из применений для таких факторов является создание ЭЭС, которые модулируют альтернативный сплайсинг эндогенных генов, в частности, генов, ассоциированных с заболеванием. Например, пре-мРНК Bcl-x дает две изоформы сплайсинга с двумя альтернативными 5' сайтами сплайсинга для кодирования белков с противоположными функциями. Длинная изоформа сплайсинга Bcl-xL является эффективным ингибитором апоптоза, экспрессируемым в долгоживущих постмитотических клетках, а его регуляция повышена во многих раковых клетках, защищая клетки от апоптотических сигналов. Короткая изоформа Bcl-xS является проапоптотической изоформой и экспрессируется на высоких уровнях в клетках с интенсивным метаболизмом (например, в развивающихся лимфоцитах). Соотношение между двумя изоформами сплайсинга Bcl-x регулируется некоторым количеством  $\omega$ -элементов, расположенных в коровой области экзона или области удлинения экзона (т.е. между двумя альтернативными 5' сайтами сплайсинга). Больше примеров можно найти в WO 2010075303, которая в полном объеме включена в данный документ посредством ссылки.

Дополнительные подходящие партнеры по слиянию включают в себя, но не ограничиваются этим, белки (или их фрагменты), являющиеся граничными элементами (например, CTCF), белки и их фрагменты, которые обеспечивают периферическое рекрутирование (например, ламин А, ламин В и т.д.), белковые стыковочные элементы (например, FKBP/FRB, Pll1/Aby1 и т.д.).

Примеры различных дополнительных подходящих гетерологичных полипептидов (или их фрагментов) для предложенного химерного полипептида CasY включают в себя, но не ограничиваются этим, описанные в следующих заявках (эти публикации относятся к другим эндонуклеазам CRISPR, таким как Cas9, но описанные партнеры по слиянию также можно использовать с CasY): патентные заявки согласно РСТ: WO2010075303, WO2012068627 и WO2013155555, и могут быть найдены, например, в патентах и патентных заявках США:

8906616; 8895308; 8889418; 8889356; 8871445; 8865406;  
 8795965; 8771945; 8697359; 20140068797; 20140170753; 20140179006; 20140179770;  
 20140186843; 20140186919; 20140186958; 20140189896; 20140227787; 20140234972;  
 20140242664; 20140242699; 20140242700; 20140242702; 20140248702; 20140256046;  
 20140273037; 20140273226; 20140273230; 20140273231; 20140273232; 20140273233;  
 20140273234; 20140273235; 20140287938; 20140295556; 20140295557; 20140298547;  
 20140304853; 20140309487; 20140310828; 20140310830; 20140315985; 20140335063;  
 20140335620; 20140342456; 20140342457; 20140342458; 20140349400; 20140349405;  
 20140356867; 20140356956; 20140356958; 20140356959; 20140357523; 20140357530;  
 20140364333 и 20140377868;

которые все в полном объеме включены в данный документ посредством ссылки.

В некоторых случаях гетерологичный полипептид (партнер по слиянию) обеспечивает субклеточную локализацию, т.е. гетерологичный полипептид содержит последовательность субклеточной локализации (например, сигнал ядерной локализации (СЯЛ) для нацеливания на ядро, последовательность для удержания слитого белка за пределами ядра, например, последовательность ядерного экспорта (ПЯЭ), последовательность для удержания слитого белка в цитоплазме, сигнал митохондриальной локализации для нацеливания на митохондрии, сигнал локализации в хлоропластах для нацеливания на хлоропласты, сигнал удержания в ЭР и т.п.). В некоторых вариантах реализации слитый полипептид CasY не содержит СЯЛ, поэтому белок не нацелен на ядро (что может быть преимуществом, например, когда целевой нуклеиновой кислотой является РНК, присутствующая в цитозоле). В некоторых вариантах реализации гетерологичный полипептид может обеспечивать наличие метки (т.е. гетерологичный полипептид является выявляемой меткой) для облегчения отслеживания и/или очистки (например, флуоресцентный белок, например, зеленый флуоресцентный белок (ЗФБ), ЖФБ, КФБ, ГФБ, mCherry, tdTomato и т.п.; гистидиновая метка, например, а метка 6XHis; гемагглютининовая (НА) метка; метка FLAG; метка Мус и т.п.).

В некоторых случаях белок CasY (например, белок CasY дикого типа, вариантный белок CasY, химерный белок CasY, белок dCasY, химерный белок CasY, в котором часть CasY обладает сниженной нуклеазной активностью, такой как белок dCasY, слитый с партнером по слиянию, и т.п.) содержит (слит) сигнал ядерной локализации (СЯЛ) (например, в некоторых случаях, 2 или более, 3 или более, 4 или более или 5 или более СЯЛ). Таким образом, в некоторых случаях, полипептид CasY содержит один или более СЯЛ (например, 2 или более, 3 или более, 4 или более или 5 или более СЯЛ). В некоторых случаях один или более СЯЛ (2 или более, 3 или более, 4 или более или 5 или более СЯЛ) размещены на или вблизи (например, в пределах 50 аминокислот) N-конца и/или С-конца. В некоторых случаях один или более СЯЛ (2 или более, 3 или более, 4 или более или 5 или более СЯЛ) размещены на или вблизи (например, в пределах 50 аминокислот) N-конца. В некоторых случаях один или более СЯЛ (2 или более, 3 или более, 4 или более или 5 или более СЯЛ) размещены на или вблизи (например, в пределах 50 аминокислот) С-конца. В некоторых случаях один или более СЯЛ (3 или более, 4 или более или 5 или более СЯЛ) размещены на или вблизи (например, в пределах 50 аминокислот) как N-конца, так и С-конца. В некоторых случаях один СЯЛ размещен на N-конце, а другой СЯЛ размещен на С-конце.

В некоторых случаях белок CasY (например, белок CasY дикого типа, вариантный белок CasY, химерный белок CasY, белок dCasY, химерный белок CasY, в котором часть CasY обладает сниженной нуклеазной активностью, такой как белок dCasY, слитый с партнером по слиянию, и т.п.) содержит (слит) от 1 до 10 СЯЛ (например, 1-9, 1-8, 1-7, 1-6, 1-5, 2-10, 2-9, 2-8, 2-7, 2-6 или 2-5 СЯЛ). В некоторых случаях белок CasY (например, белок CasY дикого типа, вариантный белок CasY, химерный белок CasY, белок dCasY, химерный белок CasY, в котором часть CasY обладает сниженной нуклеазной активностью, такой как белок dCasY, слитый с партнером по слиянию, и т.п.) содержит (слит) от 2 до 5 СЯЛ (например, 2-4 или 2-3 СЯЛ).

Неограничивающие примеры СЯЛ включают в себя последовательность СЯЛ, полученную из: СЯЛ большого Т-антигена вируса SV40, имеющего аминокислотную последовательность PKKKRKV (SEQ ID NO: 96); СЯЛ из нуклеоплазмина (например, двухсторонний СЯЛ нуклеоплазмينا с последовательностью KRPAATKKAGQAKKKK (SEQ ID NO: 97)); СЯЛ с-мус, имеющего аминокислотную последовательность PAAKRVKLD (SEQ ID NO: 98) или RQRRNELKRSP (SEQ ID NO: 99); СЯЛ hRNPA1 M9, имеющего последовательность NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY (SEQ ID NO: 100); последовательности RMRIZFKNKGKDTAELRRRRVEVSVELRKAkkDEQILKRRNV

(SEQ ID NO: 101) IBV-домена из импортина-альфа; последовательностей VSRKRPRP (SEQ ID NO: 102) и PPKKARED (SEQ ID NO: 103) Т-протеина миомы; последовательности PPKKKKPL (SEQ ID NO: 104) человеческого р53; последовательности SALIKKKKKMAP (SEQ ID NO: 105) мышинового c-abl IV; последовательностей DRLRR (SEQ ID NO: 106) и PKQKKRK (SEQ ID NO: 107) вируса гриппа NS1; последовательности RKLKKKIKKL (SEQ ID NO: 108) антигена вируса гепатита дельта; последовательности REKKKFLKRR (SEQ ID NO: 109) мышинового белка Мх1; последовательности KRKGDEV DGVDEVAKKKSKK (SEQ ID NO: 110) человеческой поли(АДФ-рибоза)полимеразы и последовательности RKCLQAGMNLEARKTKK (SEQ ID NO: 111) рецепторов стероидного гормона (человеческого) глюкокортикоида. В общем случае СЯЛ (или некоторое количество СЯЛ) имеют достаточную силу, чтобы стимулировать накопление белка CasY в выявляемом количестве в ядре эукариотической клетки. Выявление накопления в ядре можно проводить любым удобным способом. Например, с белком CasY можно сливать выявляемый маркер, чтобы можно было визуализировать его локацию в клетке. Также можно выделять из клеток клеточные ядра, содержание которых затем анализировать любым подходящим способом выявления белка, таким как иммуногистохимия, вестерн-блоттинг или анализ ферментативной активности. Накопление в ядрах также можно определять косвенным способом.

В некоторых случаях слитый белок CasY содержит "домен белковой трансдукции" или ДБТ (также известный как КПП - клеточно-проникающий пептид), который относится к полипептиду, полинуклеотиду, углеводу или органическому или неорганическому соединению, которое облегчает перемещение через липидный бислой, мицеллы, клеточную мембрану, мембрану органелл или везикулярную мембрану. ДБТ, присоединенный к другой молекуле, которая может представлять собой от малой полярной молекулы до большой макромолекулы и/или наночастицы, облегчает перемещение молекулы через мембрану, например, при переходе из внеклеточного пространства во внутриклеточное пространство или из цитозоля в органеллу. В некоторых вариантах реализации ДБТ ковалентно связан с аминоконцом полипептида (например, связан с CasY дикого типа с образованием слитого белка или связан с вариантным белком CasY, таким как dCasY, никазный CasY или химерный белок CasY, с образованием слитого белка). В некоторых вариантах реализации ДБТ ковалентно связан с карбокси-концом полипептида (например, связан с CasY дикого типа с образованием слитого белка или связан с вариантным белком CasY, таким как dCasY, никазный CasY или химерный белок CasY, с образованием слитого белка). В некоторых случаях ДБТ вставлен внутри слитого полипептида CasY (т.е. не на N- или C-конце слитого полипептида CasY) в подходящем для вставки сайте. В некоторых случаях предложенный слитый полипептид CasY содержит (конъюгирован с, слит с) один или более ДБТ (например, два или более, три или более, четыре или более ДБТ). В некоторых случаях ДБТ содержит сигнал ядерной локализации (СЯЛ) (например, в некоторых случаях 2 или более, 3 или более, 4 или более или 5 или более СЯЛ). Таким образом, в некоторых случаях, слитый полипептид CasY содержит один или более СЯЛ (например, 2 или более, 3 или более, 4 или более или 5 или более СЯЛ). В некоторых вариантах реализации ДБТ ковалентно связан с нуклеиновой кислотой (например, CasY-направляющей нуклеиновой кислотой, полинуклеотидом, кодирующим CasY-направляющую нуклеиновую кислоту, полинуклеотидом, кодирующим слитый полипептид CasY, донорным полинуклеотидом и т.д.). Примеры ДБТ включают в себя, но не ограничиваются этим, минимальный ундекапептидный домен белковой трансдукции (соответствующий остаткам 47-57 ТАТ ВИЧ-1, составляющим YGRKKRRQRRR; SEQ ID NO:112); полиаргининовую последовательность, содержащую достаточное число аргининов для непосредственного попадания в клетку (например, 3, 4, 5, 6, 7, 8, 9, 10 или 10-50 аргининов); домен VP22 (Zender et al. (2002) Cancer Gene Ther. 9(6):489-96); домен белковой трансдукции Antennapedia дрозофилы (Noguchi et al. (2003) Diabetes 52(7): 1732-1737); усеченный пептид человеческого кальцитонина (Trehin et al. (2004) Pharm. Research 21:1248-1256); полилизин (Wender et al. (2000) Proc. Natl. Acad. Sci. USA 97:13003-13008);

RRQRRTSKLMKR (SEQ ID NO:113);

транспортан

GWTLNSAGYLLGKINLKALAALAKKIL (SEQ ID NO:114);

KALAWEAKLAKALAKALAKHLAKALAKALKCEA (SEQ ID NO:115)

и RQIKIWFQNRRMKWKK (SEQ ID NO:116).

Типовые ДБТ включают в себя, но не ограничиваются этим, YGRKKRRQRRR (SEQ ID NO:117), RKKRRQRRR (SEQ ID NO:118); аргининовый гомополимер от 3 остатков аргинина до 50 остатков аргинина; типовые аминокислотные последовательности домена ДБТ включают в себя, но не ограничиваются этим, любые из нижеприведенных:

YGRKKRRQRRR (SEQ ID NO:119);

RKKRRQRR (SEQ ID NO:120);

YARAAARQARA (SEQ ID NO:121);

THRLPRRRRRR (SEQ ID NO:122) и GGRRARRRRR (SEQ ID NO:123).

В некоторых вариантах реализации ДБТ представляет собой активируемый КПП (АКПП) (Aguilera et al. (2009) Integr Biol (Camb) June; 1(5-6): 371-381). АКПП содержат поликатионный КПП (например,

Arg9 или "R9"), связанный посредством расщепляемого линкера с соответствующим полианионом (например, Glu9 или "E9"), который снижает суммарный заряд практически до нуля и тем самым ингибирует адгезию и поглощение клетками. После расщепления линкера происходит высвобождение полианиона, что приводит к локальной демаскировке полиаргинина и его природной адгезивности, "активируя", таким образом, перемещение АКПП через мембрану.

Линкеры (например, для партнеров по слиянию).

В некоторых вариантах реализации предложенный белок CasY может быть слит с партнером по слиянию посредством линкерного полипептида (например, одного или более линкерных полипептидов). Линкерный полипептид может иметь любую из множества аминокислотных последовательностей. Белки могут быть соединены спейсерным пептидом, в общем случае гибким, хотя не исключаются и другие химические связи. Подходящие линкеры включают в себя полипептиды длиной от 4 аминокислот до 40 аминокислот или длиной от 4 аминокислот до 25 аминокислот. Эти линкеры можно получать, используя синтетические, кодирующие линкеры олигонуклеотиды для сопряжения белков, или же они могут кодировать последовательностью нуклеиновой кислоты, кодирующей слитый белок. Можно использовать пептидные линкеры с некоторой степенью гибкости. Связующие пептиды могут иметь практически любую аминокислотную последовательность, принимая во внимание, что предпочтительные линкеры должны иметь последовательность, которая приводит к получению в общем случае гибкого пептида. При создании гибкого пептида имеет смысл применение небольших аминокислот, таких как глицин и аланин. Создание таких последовательностей является рутинным для специалистов в данной области техники. Множество разных линкеров являются коммерчески доступными и считаются подходящими для применения.

Примеры линкерных полипептидов включают в себя глициновые полимеры (G)<sub>n</sub>, глицин-сериновые полимеры (включая, например, (GS)<sub>n</sub>, GSGGS<sub>n</sub>(SEQ ID NO: 124), GGSGGS<sub>n</sub>(SEQ ID NO: 125) и GGGS<sub>n</sub>(SEQ ID NO: 126), где n - целое число, равное по меньшей мере одному), глицин-аланиновые полимеры, аланин-сериновые полимеры. Типовые линкеры могут содержать аминокислотные последовательности, включая, но не ограничиваясь этим, GSGG (SEQ ID NO: 127), GSGGG (SEQ ID NO: 128), GSGSG (SEQ ID NO: 129), GSGGG (SEQ ID NO: 130), GGGSG (SEQ ID NO: 131), GSSSG (SEQ ID NO: 132) и т.п. Обычному специалисту в данной области техники понятно, что конструкция пептида, конъюгированного с любым необходимым элементом, может содержать линкеры, которые являются полностью или частично гибкими, так что линкер может включать в себя гибкий линкер, а также одну или более частей, которые обеспечивают менее гибкую структуру.

Выявляемые метки.

В некоторых случаях полипептид CasY согласно настоящему изобретению содержит выявляемую метку. Подходящие выявляемые метки и/или фрагменты, которые могут обеспечивать выявляемый сигнал, включают в себя, но не ограничиваются этим, фермент, радиоактивный изотоп, представителя конкретной связывающей пары; флуорофор; флуоресцентный белок; квантовую точку и т.п.

Подходящие флуоресцентные белки включают в себя, но не ограничиваются этим, зеленый флуоресцентный белок (ЗФБ) или его варианты, синий флуоресцентный вариант ЗФБ (СФБ), голубой флуоресцентный вариант ЗФБ (ГФБ), желтый флуоресцентный вариант ЗФБ (ЖФБ), усиленный ЗФБ (УЗФБ), усиленный ГФБ (УГФБ), усиленный ЖФБ (УЖФБ), GFPs65T, Emerald, Topaz (TYFP), Venus, Citrine, mCitrine, GFPuv, дестабилизированный УЗФБ (дУЗФБ), дестабилизированный УГФБ (дУГФБ), дестабилизированный УЖФБ (дУЖФБ), mCFPm, Cerulean, T-Sapphire, CyPet, YPet, mKO, HcRed, t-HcRed, DsRed, DsRed2, DsRed-мономер, J-Red, димер2, t-димер2(12), mRFP1, поциллопорин, ЗФБ Renilla, ЗФБ Monster, раGFP, белок Kaede и "разжигаемый" белок, фикобилипротеины и конъюгаты фикобилипротеинов, включая В-фикоэритрин, R-фикоэритрин и аллофикоцианин. Другие примеры флуоресцентных белков включают в себя mHoneydew, mBanana, mOrange, dTomato, tdTomato, mTangerine, mStrawberry, mCherry, mGrape1, mRaspberry, mGrape2, mPlum (Shaner et al. (2005) Nat. Methods 2:905-909) и т.п. Любой из множества флуоресцентных и цветных белков вида Anthozoan, описанных, например, в Matz et al. (1999) Nature Biotechnol. 17:969-973, подходит для применения.

Подходящие ферменты включают в себя, но не ограничиваются этим, пероксидазу хрена (HRP), щелочную фосфатазу (AP), бета-галактозидазу (GAL), глюкозо-6-фосфатдегидрогеназу, бета-N-ацетилглюкозаминидазу, Р-глюкуронидазу, инвертазу, ксантинооксидазу, люциферазу светлячков, глюкозооксидазу (GO) и т.п.

Мотив, прилегающий к протоспейсеру (PAM).

Белок CasY связывается с целевой ДНК в целевой последовательности, определяемой областью комплементарности между ДНК-нацеливающей РНК и целевой ДНК. Как и в случае многих эндонуклеаз CRISPR, сайт-специфическое связывание (и/или расщепление) двухцепочечной целевой ДНК происходит в локациях, определяемых как (i) комплементарностью, допускающей спаривание оснований между направляющей РНК и целевой ДНК; так и (ii) коротким мотивом [называемым мотивом, прилегающим к протоспейсеру (PAM (от англ. protospacer adjacent motif))] в целевой ДНК.

В некоторых вариантах реализации РАМ для белка CasY расположен непосредственно 5' к целевой последовательности некомплементарной цепи целевой ДНК (комплементарная цепь гибридизируется с направляющей последовательностью направляющей РНК, тогда как некомплементарная цепь напрямую не гибридизируется с направляющей РНК и является обратно комплементарной некомплементарной цепи). В некоторых вариантах реализации (например, когда используется описанный в данном документе CasY1) последовательность РАМ некомплементарной цепи расположена 5'-ТА-3' (и в некоторых случаях ХТА, где Х представляет собой С, А или Т). В качестве примера смотрите фиг. 5 и 7 (на которых РАМ представляет собой ТА или СТА, если рассматривать РАМ как ХТА, где Х представляет собой С, А или Т). В некоторых вариантах реализации (например, когда используется описанный в данном документе CasY1) последовательность РАМ некомплементарной цепи расположена 5'-ТА-3' (и в некоторых случаях НТА, где Н представляет собой С, А или Т). В качестве примера смотрите фиг. 5 и 7 (на которых РАМ представляет собой ТА или СТА, если рассматривать РАМ как НТА, где Н представляет собой С, А или Т). В некоторых случаях (например, когда используется описанный в данном документе CasY2) последовательность РАМ некомплементарной цепи представляет собой 5'-YR-3' фланкирующую последовательность 5' от мишени (где Y представляет собой Т или С, а R представляет собой А или G). В некоторых случаях (например, когда используется описанный в данном документе CasY2) последовательность РАМ некомплементарной цепи представляет собой 5'-TR-3' (например, 5'-DTR-3') (где R представляет собой А или G, а D представляет собой А, G или Т). В качестве примера смотрите фиг. 5d.

В некоторых случаях разные белки CasY (т.е. белки CasY различных видов) может быть выгодно использовать в различных предложенных способах, чтобы обеспечить преимущество различных ферментативных характеристик разных белков CasY (например, для разных предпочтений относительно последовательностей РАМ; для повышения или снижения ферментативной активности; для повышения или снижения уровня клеточной токсичности; для изменения баланса между НГСК, гомологичной репарацией, однопочечными разрывами, двухцепочечными разрывами и т.д.; для обеспечения преимущества короткой общей последовательности и т.п.). Для белков CasY разных видов могут требоваться разные последовательности РАМ в целевой ДНК. Таким образом, для конкретного выбранного белка CasY требования к последовательности РАМ могут отличаться от последовательности 5'-ТА-3' (или ХТА, НТА), описанной выше. В данной области техники известны и являются рутинными различные способы (включая *in silico* и/или практические способы) определения соответствующей последовательности РАМ, и можно использовать любой подходящий способ. Описанная в данном документе последовательность ТА (ХТА, НТА) РАМ была определена с помощью анализа истощения РАМ (например, смотрите фиг. 5 в рабочих примерах ниже).

#### CasY-направляющая РНК.

Молекула нуклеиновой кислоты, которая связывается с белком CasY, образуя рибонуклеопротеиновый комплекс (РНП), и нацеливает комплекс на конкретную локацию в целевой нуклеиновой кислоте (например, целевой ДНК), называется в данном документе "CasY-направляющей РНК" или просто "направляющей РНК". Следует понимать, что в некоторых случаях можно создавать гибрид ДНК/РНК так, что CasY-направляющая РНК содержит основания ДНК помимо оснований РНК, но в данном документе для описания такой молекулы все равно используется термин "CasY-направляющая РНК".

Можно сказать, что CasY-направляющая РНК содержит два сегмента, нацеливающий сегмент и белок-связывающий сегмент. Нацеливающий сегмент CasY-направляющей РНК содержит нуклеотидную последовательность (направляющую последовательность), которая является комплементарной (и, следовательно, гибридизируется с ней) конкретной последовательности (целевому сайту) в целевой нуклеиновой кислоте (например, целевой оцРНК, целевой оцДНК, комплементарной цепи двухцепочечной целевой ДНК и т.д.). Белок-связывающий сегмент (или "белок-связывающая последовательность") взаимодействует с (связывается с) полипептидом CasY. Белок-связывающий сегмент предложенной CasY-направляющей РНК содержит два комплементарных участка нуклеотидов, которые гибридизируются друг с другом с образованием двойной спирали двухцепочечной РНК (двойной спирали дцРНК). Сайт-специфическое связывание и/или расщепление целевой нуклеиновой кислоты (например, геномной ДНК) может происходить в локациях (например, целевой последовательности целевого локуса), определяемых комплементарностью, допускающей спаривание оснований между CasY-направляющей РНК (направляющей последовательностью CasY-направляющей РНК) и целевой нуклеиновой кислотой.

CasY-направляющая РНК и белок CasY, например, слитый полипептид CasY, образуют комплекс (например, связываются посредством нековалентных взаимодействий). CasY-направляющая РНК обеспечивает комплексу целевую специфичность посредством включения нацеливающего сегмента, который содержит направляющую последовательность (нуклеотидную последовательность, комплементарную последовательности целевой нуклеиновой кислоты). Белок CasY комплекса обеспечивает сайт-специфическую активность (например, активность расщепления, обеспечиваемую белком CasY, и/или активность, обеспечиваемую партнером по слиянию в случае химерного белка CasY). Другими словами, белок CasY направляется к последовательности целевой нуклеиновой кислоты (например, целевой последовательности) за счет его ассоциации с CasY-направляющей РНК.

"Направляющую последовательность", также называемую "нацеливающей последовательностью", CasY-направляющей РНК можно модифицировать так, чтобы CasY-направляющая РНК могла нацели-

вать белок CasY (например, белок CasY природного происхождения, слитый полипептид CasY (химерный CasY) и т.п.) на любую необходимую последовательность любой необходимой целевой нуклеиновой кислоты, за исключением (например, как описано в данном документе) того, что можно принимать во внимание последовательность РАМ. Таким образом, например, CasY-направляющая РНК может иметь направляющую последовательность с комплементарностью с (например, может гибридизироваться с) последовательностью в нуклеиновой кислоте в эукариотической клетке, например, вирусной нуклеиновой кислоте, эукариотической нуклеиновой кислоте (например, эукариотической хромосоме, хромосомной последовательности, эукариотической РНК и т.д.) и т.п.

Направляющая последовательность CasY-направляющей РНК Предложенная CasY-направляющая РНК содержит направляющую последовательность (т.е. нацеливающую последовательность), которая представляет собой нуклеотидную последовательность, комплементарную последовательности (целевому сайту) в целевой нуклеиновой кислоте. Другими словами, направляющая последовательность CasY-направляющей РНК может взаимодействовать с целевой нуклеиновой кислотой (например, двухцепочечной ДНК (дцДНК), одноцепочечной ДНК (оцДНК), одноцепочечной РНК (оцРНК) или двухцепочечной РНК (дцРНК)) специфическим к последовательности образом посредством гибридизации (т.е. спаривания оснований). Направляющую последовательность CasY-направляющей РНК можно модифицировать (например, методами генетической инженерии)/конструировать для гибридизации с любой необходимой целевой последовательностью (например, принимая во внимание РАМ, например, при нацеливании на мишень дцДНК) в целевой нуклеиновой кислоте (например, эукариотической целевой нуклеиновой кислоте, такой как геномная ДНК).

В некоторых вариантах реализации процент комплементарности между направляющей последовательностью и целевым сайтом целевой нуклеиновой кислоты составляет 60% или более (например, 65% или более, 70% или более, 75% или более, 80% или более, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100%). В некоторых случаях процент комплементарности между направляющей последовательностью и целевым сайтом целевой нуклеиновой кислоты составляет 80% или более (например, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100%). В некоторых случаях процент комплементарности между направляющей последовательностью и целевым сайтом целевой нуклеиновой кислоты составляет 90% или более (например, 95% или более, 97% или более, 98% или более, 99% или более или 100%). В некоторых случаях процент комплементарности между направляющей последовательностью и целевым сайтом целевой нуклеиновой кислоты составляет 100%.

В некоторых случаях процент комплементарности между направляющей последовательностью и целевым сайтом целевой нуклеиновой кислоты составляет 100% на протяжении семи смежных крайних 3' нуклеотидов целевого сайта целевой нуклеиновой кислоты.

В некоторых случаях процент комплементарности между направляющей последовательностью и целевым сайтом целевой нуклеиновой кислоты составляет 60% или более (например, 70% или более, 75% или более, 80% или более, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100%) на протяжении 17 или более (например, 18 или более, 19 или более, 20 или более, 21 или более, 22 или более) смежных нуклеотидов. В некоторых случаях процент комплементарности между направляющей последовательностью и целевым сайтом целевой нуклеиновой кислоты составляет 80% или более (например, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100%) на протяжении 17 или более (например, 18 или более, 19 или более, 20 или более, 21 или более, 22 или более) смежных нуклеотидов. В некоторых случаях процент комплементарности между направляющей последовательностью и целевым сайтом целевой нуклеиновой кислоты составляет 90% или более (например, 95% или более, 97% или более, 98% или более, 99% или более или 100%) на протяжении 17 или более (например, 18 или более, 19 или более, 20 или более, 21 или более, 22 или более) смежных нуклеотидов. В некоторых случаях процент комплементарности между направляющей последовательностью и целевым сайтом целевой нуклеиновой кислоты составляет 100% на протяжении 17 или более (например, 18 или более, 19 или более, 20 или более, 21 или более, 22 или более) смежных нуклеотидов.

В некоторых случаях процент комплементарности между направляющей последовательностью и целевым сайтом целевой нуклеиновой кислоты составляет 60% или более (например, 70% или более, 75% или более, 80% или более, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100%) на протяжении 19 или более (например, 20 или более, 21 или более, 22 или более) смежных нуклеотидов. В некоторых случаях процент комплементарности между направляющей последовательностью и целевым сайтом целевой нуклеиновой кислоты составляет 80% или более (например, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100%) на протяжении 19 или более (например, 20 или более, 21 или более, 22 или более) смежных нуклеотидов. В некоторых случаях процент комплементарности между направляющей последовательностью и целевым сайтом целевой нуклеиновой кислоты составляет 90% или более (например, 95% или более, 97% или более, 98% или более, 99% или более или 100%) на протяжении 19 или более (например, 20 или более, 21 или более, 22 или более) смежных нуклеотидов. В некоторых случаях процент компле-

ментарности между направляющей последовательностью и целевым сайтом целевой нуклеиновой кислоты составляет 100% на протяжении 19 или более (например, 20 или более, 21 или более, 22 или более) смежных нуклеотидов.

В некоторых случаях процент комплементарности между направляющей последовательностью и целевым сайтом целевой нуклеиновой кислоты составляет 60% или более (например, 70% или более, 75% или более, 80% или более, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100%) на протяжении 17-25 смежных нуклеотидов. В некоторых случаях процент комплементарности между направляющей последовательностью и целевым сайтом целевой нуклеиновой кислоты составляет 80% или более (например, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100%) на протяжении 17-25 смежных нуклеотидов. В некоторых случаях процент комплементарности между направляющей последовательностью и целевым сайтом целевой нуклеиновой кислоты составляет 90% или более (например, 95% или более, 97% или более, 98% или более, 99% или более или 100%) на протяжении 17-25 смежных нуклеотидов. В некоторых случаях процент комплементарности между направляющей последовательностью и целевым сайтом целевой нуклеиновой кислоты составляет 100% на протяжении 17-25 смежных нуклеотидов.

В некоторых случаях процент комплементарности между направляющей последовательностью и целевым сайтом целевой нуклеиновой кислоты составляет 60% или более (например, 70% или более, 75% или более, 80% или более, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100%) на протяжении 19-25 смежных нуклеотидов. В некоторых случаях процент комплементарности между направляющей последовательностью и целевым сайтом целевой нуклеиновой кислоты составляет 80% или более (например, 85% или более, 90% или более, 95% или более, 97% или более, 98% или более, 99% или более или 100%) на протяжении 19-25 смежных нуклеотидов. В некоторых случаях процент комплементарности между направляющей последовательностью и целевым сайтом целевой нуклеиновой кислоты составляет 90% или более (например, 95% или более, 97% или более, 98% или более, 99% или более или 100%) на протяжении 19-25 смежных нуклеотидов. В некоторых случаях процент комплементарности между направляющей последовательностью и целевым сайтом целевой нуклеиновой кислоты составляет 100% на протяжении 19-25 смежных нуклеотидов.

В некоторых случаях направляющая последовательность имеет длину в диапазоне 17-30 нуклеотидов (нт) (например, 17-25, 17-22, 17-20, 19-30, 19-25, 19-22, 19-20, 20-30, 20-25 или 20-22 нт). В некоторых случаях направляющая последовательность имеет длину в диапазоне 17-25 нуклеотидов (нт) (например, 17-22, 17-20, 19-25, 19-22, 19-20, 20-25 или 20-22 нт). В некоторых случаях направляющая последовательность имеет длину 17 или более нт (например, 18 или более, 19 или более, 20 или более, 21 или более или 22 или более нт; 19 нт, 20 нт, 21 нт, 22 нт, 23 нт, 24 нт, 25 нт и т.д.). В некоторых случаях направляющая последовательность имеет длину 19 или более нт (например, 20 или более, 21 или более или 22 или более нт; 19 нт, 20 нт, 21 нт, 22 нт, 23 нт, 24 нт, 25 нт и т.д.). В некоторых случаях направляющая последовательность имеет длину 17 нт. В некоторых случаях направляющая последовательность имеет длину 18 нт. В некоторых случаях направляющая последовательность имеет длину 19 нт. В некоторых случаях направляющая последовательность имеет длину 20 нт. В некоторых случаях направляющая последовательность имеет длину 21 нт. В некоторых случаях направляющая последовательность имеет длину 22 нт. В некоторых случаях направляющая последовательность имеет длину 23 нт.

Белок-связывающий сегмент CasY-направляющей РНК.

Белок-связывающий сегмент предложенной CasY-направляющей РНК взаимодействует с белком CasY. CasY-направляющая РНК направляет связанный белок CasY к конкретной нуклеотидной последовательности в целевой нуклеиновой кислоте посредством вышеописанной направляющей последовательности. Белок-связывающий сегмент CasY-направляющей РНК содержит два комплементарные друг другу участка нуклеотидов, которые гибридизируются с образованием двойной спирали двухцепочечной РНК (двойной спирали дцРНК). Таким образом, белок-связывающий сегмент содержит двойную спираль дцРНК.

В некоторых случаях область двойной спирали дцРНК включает в себя диапазон 5-25 пар оснований (п.о.) (например, 5-22, 5-20, 5-18, 5-15, 5-12, 5-10, 5-8, 8-25, 8-22, 8-18, 8-15, 8-12, 12-25, 12-22, 12-18, 12-15, 13-25, 13-22, 13-18, 13-15, 14-25, 14-22, 14-18, 14-15, 15-25, 15-22, 15-18, 17-25, 17-22 или 17-18 п.о., например, 5 п.о., 6 п.о., 7 п.о., 8 п.о., 9 п.о., 10 п.о. и т.д.). В некоторых случаях область двойной спирали дцРНК включает в себя диапазон 6-15 пар оснований (п.о.) (например, 6-12, 6-10 или 6-8 п.о., например, 6 п.о., 7 п.о., 8 п.о., 9 п.о., 10 п.о. и т.д.). В некоторых случаях область двойной спирали содержит 5 или более п.о. (например, 6 или более, 7 или более или 8 или более п.о.). В некоторых случаях область двойной спирали содержит 6 или более п.о. (например, 7 или более или 8 или более п.о.). В некоторых случаях не все нуклеотиды области двойной спирали являются спаренными, и, следовательно, образующая двойную спираль область может содержать выступ. В контексте данного документа термин "выступ" употребляется для обозначения участка нуклеотидов (который может состоять из одного нуклеотида), которые не являются частью двухцепочечной двойной спирали, но которые окружены 5' и 3' нуклеотидами, которые являются частью спирали, и, следовательно, выступ считается частью области двойной спирали. В некоторых случаях дцРНК содержит 1 или более выступов (например, 2 или более, 3

или более или 4 или более выступов). В некоторых случаях двойная спираль дцРНК содержит 2 или более выступов (например, 3 или более или 4 или более выступов). В некоторых случаях двойная спираль дцРНК содержит 1-5 выступов (например, 1-4, 1-3, 2-5, 2-4 или 2-3 выступов).

Таким образом, в некоторых случаях участки нуклеотидов, которые гибридизируются друг с другом с образованием двойной спирали дцРНК, имеют 70%-100% комплементарности (например, 75%-100%, 80%-100%, 85%-100%, 90%-100%, 95%-100% комплементарности) друг с другом. В некоторых случаях участки нуклеотидов, которые гибридизируются друг с другом с образованием двойной спирали дцРНК, имеют 70%-100% комплементарности (например, 75%-100%, 80%-100%, 85%-100%, 90%-100%, 95%-100% комплементарности) друг с другом. В некоторых случаях участки нуклеотидов, которые гибридизируются друг с другом с образованием двойной спирали дцРНК, имеют 85%-100% комплементарности (например, 90%-100%, 95%-100% комплементарности) друг с другом. В некоторых случаях участки нуклеотидов, которые гибридизируются друг с другом с образованием двойной спирали дцРНК, имеют 70%-95% комплементарности (например, 75%-95%, 80%-95%, 85%-95%, 90%-95% комплементарности) друг с другом.

Другими словами, в некоторых вариантах реализации двойная спираль дцРНК содержит два участка нуклеотидов, которые имеют 70%-100% комплементарности (например, 75%-100%, 80%-100%, 85%-100%, 90%-100%, 95%-100% комплементарности) друг с другом. В некоторых случаях двойная спираль дцРНК содержит два участка нуклеотидов, которые имеют 85%-100% комплементарности (например, 90%-100%, 95%-100% комплементарности) друг с другом. В некоторых случаях двойная спираль дцРНК содержит два участка нуклеотидов, которые имеют 70%-95% комплементарности (например, 75%-95%, 80%-95%, 85%-95%, 90%-95% комплементарности) друг с другом.

Область двойной спирали предложенной CasY-направляющей РНК может содержать одну или более (1, 2, 3, 4, 5, и т.д.) мутаций относительно области двойной спирали природного происхождения. Например, в некоторых случаях спаривание оснований может сохраняться, хотя нуклеотиды из каждого сегмента, участвующие в спаривании оснований, могут быть разными. В некоторых случаях область двойной спирали предложенной CasY-направляющей РНК содержит больше спаренных оснований, меньше спаренных оснований, меньший выступ, больший выступ, меньше выступов, больше выступов или любую удобную их комбинацию по сравнению с областью двойной спирали природного происхождения (CasY-направляющей РНК природного происхождения).

Примеры различных Cas9-направляющих РНК можно найти в литературе, и в некоторых случаях вариации, аналогичные вносимым в Cas9-направляющие РНК, также можно вносить в CasY-направляющие РНК согласно настоящему изобретению (например, мутации в двойной спирали дцРНК, продление 5' или 3' конца для прибавления стабильности для обеспечения взаимодействия с другим белком и т.п.). Например, смотрите Jinek et al., *Science*. 2012 Aug 17; 337(6096):816-21; Chylinski et al., *RNA Biol*. 2013 May; 10(5):726-37; Ma et al., *Biomed Res Int*. 2013; 2013:270805; Hou et al., *Proc Natl Acad Sci USA*. 2013 Sep 24; 110(39):15644-9; Jinek et al., *Elife*. 2013; 2:e00471; Pattanayak et al., *Nat Biotechnol*. 2013 Sep; 31(9):839-43; Qi et al., *Cell*. 2013 Feb 28; 152(5):1173-83; Wang et al., *Cell*. 2013 May 9; 153(4):910-8; Auer et al., *Genome Res*. 2013 Oct 31; Chen et al., *Nucleic Acids Res*. 2013 Nov 1; 41(20):e19; Cheng et al., *Cell Res*. 2013 Oct; 23(10):1163-71; Cho et al., *Genetics*. 2013 Nov; 195(3):1177-80; DiCarlo et al., *Nucleic Acids Res*. 2013 Apr; 41(7):4336-43; Dickinson et al., *Nat Methods*. 2013 Oct; 10(10): 1028-34; Ebin et al., *Sci Rep*. 2013; 3:2510; Fujii et al., *Nucleic Acids Res*. 2013 Nov 1; 41(20):e187; Hu et al., *Cell Res*. 2013 Nov; 23(11):1322-5; Jiang et al., *Nucleic Acids Res*. 2013 Nov 1; 41(20):e188; Larson et al., *Nat Protoc*. 2013 Nov; 8(11):2180-96; Mali et al., *Nat Methods*. 2013 Oct; 10(10):957-63; Nakayama et al., *Genesis*. 2013 Dec; 51(12):835-43; Ran et al., *Nat Protoc*. 2013 Nov; 8(11):2281-308; Ran et al., *Cell*. 2013 Sep 12; 154(6):1380-9; Upadhyay et al., *G3 (Bethesda)*. 2013 Dec 9; 3(12):2233-8; Walsh et al., *Proc Natl Acad Sci USA*. 2013 Sep 24; 110(39):15514-5; Xie et al., *Mol Plant*. 2013 Oct 9; Yang et al., *Cell*. 2013 Sep 12; 154(6):1370-9; Briner et al., *Mol Cell*. 2014 Oct 23; 56(2):333-9 и патенты и заявки на патенты США:

8906616; 8895308; 8889418; 8889356;

8871445; 8865406; 8795965; 8771945; 8697359; 20140068797; 20140170753; 20140179006;

20140179770; 20140186843; 20140186919; 20140186958; 20140189896; 20140227787;

20140234972; 20140242664; 20140242699; 20140242700; 20140242702; 20140248702;

20140256046; 20140273037; 20140273226; 20140273230; 20140273231; 20140273232;  
 20140273233; 20140273234; 20140273235; 20140287938; 20140295556; 20140295557;  
 20140298547; 20140304853; 20140309487; 20140310828; 20140310830; 20140315985;  
 20140335063; 20140335620; 20140342456; 20140342457; 20140342458; 20140349400;  
 20140349405; 20140356867; 20140356956; 20140356958; 20140356959; 20140357523;  
 20140357530; 20140364333 и 20140377868;

которые все в полном объеме включены в данный документ посредством ссылки.

CasY-направляющая РНК содержит как направляющую последовательность, так и два участка ("образующие двойную спираль сегменты") нуклеотидов, которые гибридизируются с образованием двойной спирали дцРНК белок-связывающего сегмента. Конкретная последовательность заданной CasY-направляющей РНК может быть характерной для вида, которому принадлежит cr-РНК. Примеры подходящих CasY-направляющих РНК предложены в данном документе.

Примеры направляющих последовательностей РНК.

Последовательности повторов (ненаправляющая часть последовательности типовых CasY-направляющих РНК), проиллюстрированные на фиг. 6 (панели а и b), получены из природного локуса CasY1-Y5. В некоторых случаях предложенная CasY-направляющая РНК содержит (например, помимо направляющей последовательности) последовательность cr-РНК  
 CTCCGAAAGTATCGGGGATAAAGGC (SEQ ID NO: 31) [РНК представляет собой  
 CUCCGAAAGUAUCGGGGAUAAAGGC (SEQ ID NO: 11)] (например, смотрите фиг. 6). В некоторых случаях предложенная CasY-направляющая РНК содержит нуклеотидную последовательность, имеющую 80% или более идентичности (например, 85% или более, 90% или более, 93% или более, 95% или более, 97% или более, 98% или более или 100% идентичности) с последовательностью cr-РНК  
 CTCCGAAAGTATCGGGGATAAAGGC (SEQ ID NO: 31) [РНК представляет собой  
 CUCCGAAAGUAUCGGGGAUAAAGGC (SEQ ID NO: 11)]. В некоторых случаях предложенная CasY-направляющая РНК содержит нуклеотидную последовательность, имеющую 90% или более идентичности (например, 93% или более, 95% или более, 97% или более, 98% или более или 100% идентичности) с последовательностью cr-РНК  
 CTCCGAAAGTATCGGGGATAAAGGC (SEQ ID NO: 31) [РНК представляет собой  
 CUCCGAAAGUAUCGGGGAUAAAGGC (SEQ ID NO: 11)].

В некоторых случаях предложенная CasY-направляющая РНК содержит (например, помимо направляющей последовательности) последовательность cr-РНК  
 CACCGAAATTTGGAGAGGATAAAGGC (SEQ ID NO: 32) [РНК представляет собой  
 CACCGAAAUUUGGAGAGGAUAAGGC (SEQ ID NO: 12)] (например, смотрите фиг. 6).

В некоторых случаях предложенная CasY-направляющая РНК содержит нуклеотидную последовательность, имеющую 80% или более идентичности (например, 85% или более, 90% или более, 93% или более, 95% или более, 97% или более, 98% или более или 100% идентичности) с последовательностью cr-РНК  
 CACCGAAATTTGGAGAGGATAAAGGC (SEQ ID NO: 32) [РНК представляет собой  
 CACCGAAAUUUGGAGAGGAUAAGGC (SEQ ID NO: 12)]. В некоторых случаях предложенная CasY-направляющая РНК содержит нуклеотидную последовательность, имеющую 90% или более идентичности (например, 93% или более, 95% или более, 97% или более, 98% или более или 100% идентичности) с последовательностью cr-РНК  
 CACCGAAATTTGGAGAGGATAAAGGC (SEQ ID NO: 32) [РНК представляет собой  
 CACCGAAAUUUGGAGAGGAUAAGGC (SEQ ID NO: 12)].

В некоторых случаях предложенная CasY-направляющая РНК содержит (например, помимо направляющей последовательности) последовательность cr-РНК  
 CTCCGAATTATCGGGAGGATAAAGGC (SEQ ID NO: 33) [РНК представляет собой  
 CUCCGAAUUAUCGGGAGGAUAAGGC (SEQ ID NO: 13)] (например, смотрите фиг. 6).

В некоторых случаях предложенная CasY-направляющая РНК содержит нуклеотидную последовательность, имеющую 80% или более идентичности (например, 85% или более, 90% или более, 93% или более, 95% или более, 97% или более, 98% или более или 100% идентичности) с последовательностью cr-РНК  
 CTCCGAATTATCGGGAGGATAAAGGC (SEQ ID NO: 33) [РНК представляет собой  
 CUCCGAAUUAUCGGGAGGAUAAGGC (SEQ ID NO: 13)]. В некоторых случаях предложенная CasY-направляющая РНК содержит нуклеотидную последовательность, имеющую 90% или более идентичности (например, 93% или более, 95% или более, 97% или более, 98% или более или 100%





нуклеотидную последовательность, кодирующую первую CasY-направляющую РНК, и нуклеотидную последовательность, кодирующую вторую CasY-направляющую РНК; или г) рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую слитый полипептид CasY согласно настоящему изобретению, нуклеотидную последовательность, кодирующую первую CasY-направляющую РНК, и нуклеотидную последовательность, кодирующую вторую CasY-направляющую РНК; или некоторую вариацию (а)-(г).

Нуклеиновые кислоты.

В настоящем изобретении предложены одна или более нуклеиновых кислот, содержащих одно или более из: донорной полинуклеотидной последовательности, нуклеотидной последовательности, кодирующей полипептид CasY (например, полипептид CasY дикого типа, никазный белок CasY, белок dCasY, химерный белок CasY и т.п.), CasY-направляющей РНК и нуклеотидной последовательности, кодирующей CasY-направляющую РНК. В настоящем изобретении предложена нуклеиновая кислота, содержащая нуклеотидную последовательность, кодирующую слитый полипептид CasY. В настоящем изобретении предложен рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую слитый полипептид CasY. В настоящем изобретении предложен рекомбинантный экспрессионный вектор, содержащий: а) нуклеотидную последовательность, кодирующую полипептид CasY; и б) нуклеотидную последовательность, кодирующую CasY-направляющую(ие) РНК. В настоящем изобретении предложен рекомбинантный экспрессионный вектор, содержащий: а) нуклеотидную последовательность, кодирующую слитый полипептид CasY; и б) нуклеотидную последовательность, кодирующую CasY-направляющую(ие) РНК. В некоторых случаях нуклеотидная последовательность, кодирующая белок CasY, и/или нуклеотидная последовательность, кодирующая CasY-направляющую РНК, функционально связаны с промотором, функциональным в выбранном клеточном типе (например, в прокариотической клетке, эукариотической клетке, клетке растения, клетке животного, клетке млекопитающего, клетке примата, клетке грызуна, клетке человека и т.д.).

В некоторых случаях нуклеотидная последовательность, кодирующая полипептид CasY согласно настоящему изобретению, является кодон-оптимизированной. Этот тип оптимизации может включать в себя мутацию кодирующей CasY нуклеотидной последовательности, направленную на имитацию предпочтительного использования кодонов предполагаемого организма- или клетки-хозяина с кодированием того же самого белка. Таким образом, кодоны можно изменять, но кодируемый белок остается неизменным. Например, если предполагаемой целевой клеткой является человеческая клетка, можно использовать человеческую кодон-оптимизированную кодирующую CasY нуклеотидную последовательность. В качестве другого неограничивающего примера, если предполагаемой целевой клеткой является мышьяная клетка, то можно создавать мышьяную кодон-оптимизированную кодирующую CasY нуклеотидную последовательность. В качестве другого неограничивающего примера, если предполагаемой целевой клеткой является растительная клетка, то можно создавать растительную кодон-оптимизированную кодирующую CasY нуклеотидную последовательность. В качестве другого неограничивающего примера, если предполагаемой целевой клеткой является клетка насекомого, то можно создавать кодон-оптимизированную кодирующую CasY нуклеотидную последовательность насекомого.

В настоящем изобретении предложены один или более рекомбинантных экспрессионных векторов, которые содержат (в некоторых случаях - в разных рекомбинантных экспрессионных векторах, а в некоторых случаях - в одном рекомбинантном экспрессионном векторе): (i) нуклеотидную последовательность донорной матричной нуклеиновой кислоты (при этом донорная матрица содержит нуклеотидную последовательность, имеющую гомологию с целевой последовательностью целевой нуклеиновой кислоты (например, целевым геномом)); (ii) нуклеотидную последовательность, которая кодирует CasY-направляющую РНК, которая гибридизируется с целевой последовательностью целевого локуса целевого генома (например, функционально связанную с промотором, функциональным в целевой клетке, такой как эукариотическая клетка); и (iii) нуклеотидную последовательность, кодирующую белок CasY (например, функционально связанную с промотором, функциональным в целевой клетке, такой как эукариотическая клетка). В настоящем изобретении предложены один или более рекомбинантных экспрессионных векторов, которые содержат (в некоторых случаях - в разных рекомбинантных экспрессионных векторах, а в некоторых случаях - в одном рекомбинантном экспрессионном векторе): (i) нуклеотидную последовательность донорной матричной нуклеиновой кислоты (при этом донорная матрица содержит нуклеотидную последовательность, имеющую гомологию с целевой последовательностью целевой нуклеиновой кислоты (например, целевым геномом)); и (ii) нуклеотидную последовательность, которая кодирует CasY-направляющую РНК, которая гибридизируется с целевой последовательностью целевого локуса целевого генома (например, функционально связанную с промотором, функциональным в целевой клетке, такой как эукариотическая клетка). В настоящем изобретении предложены один или более рекомбинантных экспрессионных векторов, которые содержат (в некоторых случаях - в разных рекомбинантных экспрессионных векторах, а в некоторых случаях - в одном рекомбинантном экспрессионном векторе): (i) нуклеотидную последовательность, которая кодирует CasY-направляющую РНК, которая

гибридируется с целевой последовательностью целевого локуса целевого генома (например, функционально связанную с промотором, функциональным в целевой клетке, такой как эукариотическая клетка); и (ii) нуклеотидную последовательность, кодирующую белок CasY (например, функционально связанную с промотором, функциональным в целевой клетке, такой как эукариотическая клетка).

Подходящие экспрессионные векторы включают в себя вирусные экспрессионные векторы (например, вирусные векторы на основе вируса осповакцины; полиовируса; аденовируса (смотрите, например, Li et al., *Invest Ophthalmol Vis Sci* 35:2543-2549, 1994; Borrás et al., *Gene Ther* 6:515-524, 1999; Li and Davidson, *PNAS* 92:7700-7704, 1995; Sakamoto et al., *Hum Gene Ther* 5:1088-1097, 1999; WO 94/12649, WO 93/03769; WO 93/19191; WO 94/28938; WO 95/11984 и WO 95/00655); аденоассоциированного вируса (AAB) (смотрите, например, Ali et al., *Hum Gene Ther* 9:81-86, 1998; Flannery et al., *PNAS* 94:6916-6921, 1997; Bennett et al., *Invest Ophthalmol Vis Sci* 38:2857-2863, 1997; Jomary et al., *Gene Ther* 4:683-690, 1997; Rolling et al., *Hum Gene Ther* 10:641-648, 1999; Ali et al., *Hum Mol Genet* 5:591-594, 1996; Srivastava in *WO* 93/09239, Samulski et al., *J. Vir.* (1989) 63:3822-3828; Mendelson et al., *Viol.* (1988) 166:154-165; и Flotte et al., *PNAS* (1993) 90:10613-10617); SV40; вируса простого герпеса; вируса иммунодефицита человека (смотрите, например, Miyoshi et al., *PNAS* 94:10319-23, 1997; Takahashi et al., *J Virol* 73:7812-7816, 1999); ретровирусный вектор (например, вируса лейкемии мышей, вируса некроза селезенки, и векторы, полученные из ретровирусов, таких как вирус саркомы Рауса, вирус саркомы Харви, вирус лейкемии птиц, лентивирус, вирус иммунодефицита человека, вирус миелопротоплазматической саркомы и вирус опухоли молочной железы) и т.п. В некоторых случаях рекомбинантный экспрессионный вектор согласно настоящему изобретению представляет собой рекомбинантный аденоассоциированный вирусный (AAB) вектор. В некоторых случаях рекомбинантный экспрессионный вектор согласно настоящему изобретению представляет собой рекомбинантный лентивирусный вектор. В некоторых случаях рекомбинантный экспрессионный вектор согласно настоящему изобретению представляет собой рекомбинантный ретровирусный вектор.

В зависимости от используемой системы хозяин/вектор в экспрессионном векторе можно использовать любое число подходящих транскрипционных и трансляционных регуляторных элементов, включая конститутивные и индуцибельные промоторы, энхансеры транскрипции, терминаторы транскрипции и т.д.

В некоторых вариантах реализации нуклеотидная последовательность, кодирующая CasY-направляющую РНК, функционально связана с регуляторным элементом, например, транскрипционным регуляторным элементом, таким как промотор. В некоторых вариантах реализации нуклеотидная последовательность, кодирующая белок CasY или слитый полипептид CasY, функционально связана с регуляторным элементом, например, транскрипционным регуляторным элементом, таким как промотор.

Транскрипционный регуляторный элемент может представлять собой промотор. В некоторых случаях промотор является конститутивно активным промотором. В некоторых случаях промотор является регулируемым промотором. В некоторых случаях промотор является индуцибельным промотором. В некоторых случаях промотор является тканеспецифическим промотором. В некоторых случаях промотор является специфическим в отношении клеточного типа промотором. В некоторых случаях транскрипционный регуляторный элемент (например, промотор) является функциональным в целевом клеточном типе или в целевой клеточной популяции. Например, в некоторых случаях транскрипционный регуляторный элемент может быть функциональным в эукариотических клетках, например, гемопоэтических стволовых клетках (например, мобилизованных CD34(+) клетках периферической крови (мПК), CD34(+) клетках костного мозга (КМ) и т.д.).

Неограничивающие примеры эукариотических промоторов (промоторов, функциональных в эукариотических клетках) включают в себя промоторы EF1 $\alpha$ , немедленно-ранние промоторы из цитомегаловируса (ЦМВ), промоторы тимидинкиназы вируса простого герпеса (ВПГ), ранние и поздние промоторы SV40, длинные концевые повторы (ДКП) из ретровируса и мышинный металлотионеин-I. Выбор подходящего вектора и промотора находится в пределах компетенции специалиста в данной области техники. Экспрессионный вектор также может содержать сайт связывания рибосомы для инициации трансляции и терминатор транскрипции. Экспрессионный вектор также может содержать соответствующие последовательности для амплификации экспрессии. Экспрессионный вектор также может содержать нуклеотидные последовательности, кодирующие белковые метки (например, метку 6xHis, гемагглютининовую метку, флуоресцентный белок и т.д.), которые могут быть слиты с белком CasY, приводя, таким образом, к созданию химерного полипептида CasY.

В некоторых вариантах реализации нуклеотидная последовательность, кодирующая CasY-направляющую РНК и/или слитый полипептид CasY, функционально связана с индуцибельным промотором. В некоторых вариантах реализации нуклеотидная последовательность, кодирующая CasY-направляющую РНК и/или слитый полипептид CasY, функционально связана с конститутивным промотором.

Промотор может быть конститутивно активным промотором (т.е. промотором, который конститутивно находится в активном/"включенном" состоянии), он может быть индуцибельным промотором (т.е. промотором, чье состояние, активное/"включенное" или неактивное/"выключенное", регулируется внеш-

ним стимулом, например, конкретной температурой, наличием конкретного соединения или белка), он может быть пространственно ограниченным промотором (т.е. транскрипционным регуляторным элементом, энхансером и т.д.) (например, тканеспецифическим промотором, специфическим в отношении клеточного типа промотором и т.д.) и он может быть временно ограниченным промотором (т.е. промотор находится во "включенном" состоянии или "выключенном" состоянии во время конкретных стадий эмбрионального развития или во время конкретных стадий биологического процесса, например, цикл волосяного фолликула у мышей).

Подходящие промоторы могут быть получены из вирусов и, следовательно, называются вирусными промоторами, или же они могут быть получены из любого организма, включая прокариотические или эукариотические организмы. Подходящие промоторы можно использовать для стимуляции экспрессии любой РНК-полимеразой (например, pol pol I pol III). Типовые промоторы включают в себя, но не ограничиваются этим, ранний промотор SV40, промотор длинных концевых повторов (ДКП) опухоли молочной железы мышей; аденовирусный основной поздний промотор (Ad MLP); промотор вируса простого герпеса (ВПГ), промотор цитомегаловируса (ЦМВ (англ., CMV)), такой как область немедленно-раннего промотора ЦМВ (CMVIE), промотор вируса саркомы Пауса (BCP), малый ядерный промотор U6 человека (U6) (Miyagishi et al., Nature Biotechnology 20, 497-500 (2002)), усиленный промотор U6 (например, Xia et al., Nucleic Acids Res. 2003 Sep 1; 31(17)), промотор H1 человека (H1) и т.п.

В некоторых случаях нуклеотидная последовательность, кодирующая CasY-направляющую РНК, функционально связана с промотором (находится под управлением промотора), функциональным в эукариотической клетке (например, промотором U6, усиленным промотором U6, промотором H1 и т.п.). Как понятно специалисту в данной области техники, при экспрессии РНК (например, направляющей РНК) из нуклеиновой кислоты (например, экспрессионного вектора) с применением промотора U6 (например, в эукариотической клетке) или другого промотора PolIII в РНК может быть необходимо внести мутации в случае наличия нескольких Т в ряд (кодирующих U в РНК). Это связано с тем, что ряд Т (например, 5 Т) в ДНК может действовать как терминатор для полимеразы III (PolIII). Таким образом, чтобы гарантировать транскрипцию направляющей РНК в эукариотической клетке, иногда может быть необходимо модифицировать последовательность, кодирующую направляющую РНК, чтобы устранить сплошные участки Т. В некоторых случаях нуклеотидная последовательность, кодирующая белок CasY (например, белок CasY дикого типа, никазный белок CasY, белок dCasY, химерный белок CasY и т.п.), функционально связана с промотором, функциональным в эукариотической клетке (например, промотором ЦМВ, промотором EF1 $\alpha$ , регулируемым эстрогеновым рецептором промотором и т.п.).

Примеры индуцибельных промоторов включают в себя, но не ограничиваются этим, промотор РНК-полимеразы T7, промотор РНК-полимеразы T3, изопропил-бета-D-тиогалактопиранозид (ИПТГ)-регулируемый промотор, индуцируемый лактозой промотор, промотор теплового шока, тетрациклин-регулируемый промотор, стероид-регулируемый промотор, металл-регулируемый промотор, регулируемый эстрогеновым рецептором промотор и т.д. Следовательно, индуцибельные промоторы могут регулироваться молекулами, включая, но не ограничиваясь этим, доксициклин; эстроген и/или аналог эстрогена; ИПТГ и т.д.

Подходящие для применения индуцибельные промоторы включают в себя любые описанные в данном документе или известные специалисту в данной области техники индуцибельные промоторы. Примеры индуцибельных промоторов включают в себя, без ограничений, химически/биохимически-регулируемые и физически-регулируемые промоторы, такие как спирторегулируемые промоторы, тетрациклин-регулируемые промоторы (например, ангидротетрациклин (aTc)-чувствительные промоторы и другие тетрациклин-чувствительные промоторные системы, которые включают в себя тетрациклиновый репрессорный белок (tetR), тетрациклиновую операторную последовательность (tetO) и тетрациклиновый трансактивационный слитый белок (tTA)), стероид-регулируемые промоторы (например, промоторы на основании крысиного рецептора глюкокортикоидов, человеческого рецептора эстрогена, рецепторов экдизонов мотыльков и промоторы из суперсемейства стероидных/ретиноидных/тироидных рецепторов), металл-регулируемые промоторы (например, промоторы, полученные из генов металлотионеина (белков, которые связывают и секвенируют ионы металлов) дрожжей, мышей и человека), патогенез-регулируемые промоторы (например, индуцируемые салициловой кислотой, этиленом или бензотиадиазолом (BTH)), индуцируемые температурой/теплом промоторы (например, промоторы теплового шока) и светорегулируемые промоторы (например, светочувствительные промоторы из клеток растений).

В некоторых случаях промотор является пространственно ограниченным промотором (т.е. специфическим в отношении клеточного типа промотором, тканеспецифическим промотором и т.д.) так, что в многоклеточном организме промотор является активным (т.е. "включенным") в подгруппе конкретных клеток. Пространственно ограниченные промоторы также могут называться энхансерами, транскрипционными регуляторными элементами, регуляторными последовательностями и т.д. Можно использовать любой подходящий пространственно ограниченный промотор при условии, что промотор является функциональным в целевой клетке-хозяине (например, эукариотической клетке; прокариотической клетке).

В некоторых случаях промотор является обратимым промотором. Подходящие обратимые промоторы, включая обратимые индуцибельные промоторы, известны в данной области техники. Такие обра-

тимые промоторы могут быть выделены и получены из многих организмов, например, эукариот и прокариот. В данной области техники хорошо известна модификация обратимых промоторов, полученных из первого организма, для использования во втором организме, например, когда первый организм является прокариотом, а второй эукариотом; первый эукариотом, а второй прокариотом и т.д. Такие обратимые промоторы и системы на основе таких обратимых промоторов, но также содержащие дополнительные регуляторные белки, включают в себя, но не ограничиваются этим, спирторегулируемые промоторы (например, промотор гена алкогольдегидрогеназы I (alcA), промоторы, чувствительные к алкоголь-трансактивирующим белкам (AlcR) и т.д.), тетрациклин-регулируемые промоторы (например, промоторные системы, включая Tet-активаторы, TetON, TetOFF и т.д.), стероид-регулируемые промоторы (например, промоторные системы крысиных рецепторов глюкокортикоидов, промоторные системы человеческих рецепторов эстрогена, ретиноидные промоторные системы, тироидные промоторные системы, экдизоновые промоторные системы, мифепристоновые промоторные системы и т.д.), металл-регулируемые промоторы (например, металлотионеиновые промоторные системы и т.д.), патогенез-регулируемые промоторы (например, регулируемые салициловой кислотой промоторы, этилен-регулируемые промоторы, бензотиадиазол-регулируемые промоторы и т.д.), температурно-регулируемые промоторы (например, индуцибельные промоторы теплового шока (например, HSP-70, HSP-90, промотор теплового шока сои и т.д.), светорегулируемые промоторы, синтетические индуцибельные промоторы и т.п.

Способы внесения нуклеиновой кислоты (например, нуклеиновой кислоты, содержащей донорную полинуклеотидную последовательность, одну или более нуклеиновых кислот, кодирующих белок CasY и/или CasY-направляющую РНК, и т.п.) в клетку-хозяина известны в данной области техники, а для внесения нуклеиновой кислоты (например, экспрессионной конструкции) в клетку можно использовать любой удобный способ. Подходящие способы включают в себя, например, вирусную инфекцию, трансфекцию, липофекцию, электропорацию, осаждение фосфатом кальция, полиэтиленмин (ПЭИ)-опосредованную трансфекцию, ДЭАЭ-декстран-опосредованную трансфекцию, липосомно-опосредованную трансфекцию, технологию генной пушки, осаждение фосфатом кальция, прямую микроинъекцию, опосредованную наночастицами доставку нуклеиновой кислоты и т.п.

Внесение рекомбинантного экспрессионного вектора в клетки может происходить в любой культуральной среде и в любых культуральных условиях, которые способствуют выживаемости клеток. Внесение рекомбинантного экспрессионного вектора в целевую клетку можно проводить *in vivo* или *ex vivo*. Внесение рекомбинантного экспрессионного вектора в целевую клетку можно проводить *in vitro*.

В некоторых вариантах реализации белок CasY может быть получен в виде РНК. РНК может быть получена путем прямого химического синтеза или может быть транскрибирована *in vitro* из ДНК (например, кодирующей белок CasY). После синтеза РНК можно вносить в клетку любым из хорошо известных способов внесения в клетки нуклеиновых кислот (например, путем микроинъекции, электропорации, трансфекции и т.д.).

Нуклеиновые кислоты можно внедрять в клетки, используя хорошо разработанные способы трансфекции; смотрите, например, Angel and Yanik (2010) PLoS ONE 5(7): e11756, и коммерчески доступные реагенты TransMessenger® от Qiagen, набор для трансфекции РНК Stemfect™ от Stemgent и набор для трансфекции мРНК TransIT® от Minis Bio LLC. Также смотрите Beumer et al. (2008) PNAS 105(50):19821-19826.

Векторы можно напрямую вносить в целевую клетку-хозяина. Другими словами, клетки приводят в контакт с векторами, содержащими предложенные нуклеиновые кислоты (например, рекомбинантными экспрессионными векторами, имеющими донорную матричную последовательность и кодирующими CasY-направляющую РНК; рекомбинантными экспрессионными векторами, кодирующими белок CasY; и т.д.), так, чтобы векторы поглощались клетками. Способы приведения клеток в контакт с содержащими нуклеиновые кислоты векторами, которые являются плазмидами, включают в себя электропорацию, трансфекцию с помощью хлорида кальция, микроинъекцию и липофекцию и хорошо известны в данной области техники. Для доставки вирусных векторов можно приводить клетки в контакт с вирусными частицами, содержащими предложенные вирусные экспрессионные векторы.

Ретровирусы, например, лентивирусы, подходят для применения в способах согласно настоящему изобретению. Обычно используемые ретровирусные векторы являются "дефективными", т.е. неспособными вырабатывать вирусные белки для продуктивного инфицирования. Правильнее сказать, что для репликации вектора требуется рост в пакующей линии клеток. Для создания вирусных частиц, содержащих представляющие интерес нуклеиновые кислоты, ретровирусные нуклеиновые кислоты, содержащие нуклеиновую кислоту, пакуют в вирусные капсиды с помощью пакующей линии клеток. Разные пакующие линии клеток обеспечивают разные оболочечные белки (экотропные, амфотропные или ксенотропные) для включения в капсид, эти оболочечные белки определяют специфичность вирусной частицы в отношении клеток (экотропные для мышей и крыс; амфотропные для большинства типов клеток млекопитающих, включая человека, собак и мышей; и ксенотропные для большинства типов клеток млекопитающих, за исключением мышинных клеток). Можно использовать соответствующую пакующую линию клеток, чтобы гарантировать, что упакованные вирусные частицы нацелены на клетки. Способы внесе-

ния предложенных экспрессионных векторов в пакующие линии клеток и сбора вирусных частиц, созданных пакующими линиями, хорошо известны в данной области техники. Нуклеиновые кислоты также можно вносить путем прямой микроинъекции (например, инъекции РНК).

Векторы, используемые для внесения нуклеиновых кислот, кодирующих CasY-направляющую РНК и/или полипептид CasY, в целевую клетку-хозяина, могут содержать подходящие промоторы для стимуляции экспрессии, что означает активацию транскрипции представляющей интерес нуклеиновой кислоты. Другими словами, в некоторых случаях представляющая интерес нуклеиновая кислота функционально связана с промотором. Они могут включать в себя универсальные промоторы, например, промотор ЦМВ- $\beta$ -актина, или индуцибельные промоторы, такие как промоторы, активные в конкретных клеточных популяциях или которые восприимчивы к присутствию лекарственных препаратов, таких как тетрациклин. Под активацией транскрипции подразумевается, что транскрипция в целевой клетке повышается выше исходных уровней в 10 раз, в 100 раз, чаще - в 1000 раз. Кроме того, векторы, используемые для внесения нуклеиновой кислоты, кодирующей CasY-направляющую РНК и/или белок CasY, в клетку, могут содержать последовательности нуклеиновых кислот, которые кодируют селективные маркеры в целевых клетках, для определения клеток, поглотивших CasY-направляющую РНК и/или белок CasY.

Нуклеиновая кислота, содержащая нуклеотидную последовательность, кодирующую полипептид CasY или слитый полипептид CasY, в некоторых случаях представляет собой РНК. Таким образом, слитый белок CasY можно вносить в клетки в виде РНК. Способы внесения РНК в клетки известны в данной области техники и могут включать в себя, например, прямую инъекцию, трансфекцию или любой другой способ для внесения ДНК. В ином случае белок CasY можно вносить в клетки в виде полипептида. Такой полипептид может необязательно быть слит с полипептидным доменом, который повышает растворимость продукта. Домен может быть связан с полипептидом посредством сайта расщепления определенной протеазой, например, последовательности TEV, которая расщепляется TEV-протеазой. Линкер также может содержать одну или более гибких последовательностей, например, от 1 до 10 остатков глицина. В некоторых вариантах реализации расщепление слитого белка проводят в буфере, который поддерживает растворимость продукта, например, в присутствии от 0,5 до 2 М мочевины, в присутствии полипептидов и/или полинуклеотидов для повышения растворимости и т.п. Представляющие интерес домены включают в себя эндосомолитические домены, например, домен HA гриппа; и другие полипептиды, которые способствуют выработке, например, домен IF2, домен GST, домен GRPE и т.п. Полипептид может быть приготовлен с улучшением его стабильности. Например, пептиды могут быть ПЭГилированными, в случае чего полиэтиленокси-группа обеспечивает повышенное время жизни в кровотоке.

В дополнительном или альтернативном варианте полипептид CasY согласно настоящему изобретению может быть слит с доменом, обеспечивающим проникновение полипептида, что способствует поглощению клеткой. В данной области техники известен ряд обеспечивающих проникновение доменов, которые можно использовать в неинтегрирующихся полипептидах согласно настоящему изобретению, включая пептиды, пептидомиметики и непептидные носители. Например, обеспечивающий проникновение пептид может быть получен из третьей альфа-спирали гена *Antennapedia* транскрипционного фактора *Drosophila melanogaster*, называемого пенетратинном, который содержит аминокислотную последовательность RQIKIWFQNRRMKWKK (SEQ ID NO: 133). В качестве другого примера, обеспечивающий проникновение пептид содержит аминокислотную последовательность основной области tat ВИЧ-1, которая может включать в себя, например, аминокислоты 49-57 белка tat природного происхождения. Другие обеспечивающие проникновение домены включают в себя полиаргининовые мотивы, например, область аминокислот 34-56 белка rev ВИЧ-1, нонааргинин, октааргинин и т.п. (Смотрите, например, Futaki et al. (2003) *Curr Protein Pept Sci.* 2003 Apr; 4(2):87-9 и 446 и Wender et al. (2000) *Proc. Natl. Acad. Sci. U.S.A.* 2000 Nov. 21; 97(24):13003-8; опубликованные патентные заявки США 20030220334; 20030083256; 20030032593 и 20030022831, специально включенные в данный документ посредством ссылки в отношении описания транслокационных пептидов и пептоидов). Нонааргининовая (R9) последовательность является одним из наиболее эффективных ДБТ, которые были изучены (Wender et al., 2000; Uemura et al., 2002). Сайт, в котором проводится слияние, может быть выбран так, чтобы оптимизировать биологическую активность, характеристики секреции или связывания полипептида. Оптимальный сайт определяют с помощью рутинных экспериментов.

Полипептид CasY согласно настоящему изобретению может быть получен *in vitro* или с помощью эукариотических или прокариотических клеток и может подвергаться дополнительной обработке путем разворачивания, например, тепловой денатурации, восстановления дитиотреитолом и т.д., и может подвергаться дополнительному повторному сворачиванию с помощью известных в данной области техники способов.

Представляющие интерес модификации, которые не меняют первичную последовательность, включают в себя химическую дериватизацию полипептидов, например, ацилирование, ацетилирование, карбоксилирование, амидирование и т.д. Также включены модификации гликозилирования, например, осуществляемые путем модификации профиля гликозилирования полипептида во время его синтеза и обработки или на этапах дополнительной обработки; например, посредством воздействия на полипептид

ферментов, которые влияют на гликозилирование, таких как гликозилирующие или дегликозилирующие ферменты млекопитающих. Также включены последовательности, имеющие фосфорилированные аминокислотные остатки, например, фосфотирозин, фосфосерин или фосфотреонин.

Также для включения в варианты реализации настоящего изобретения подходят аминокислоты (например, кодирующие CasY-направляющую РНК, кодирующие слитый белок CasY и т.д.) и белки (например, слитый белок CasY, полученный из белка дикого типа или вариантного белка), которые были модифицированы с помощью стандартных способов молекулярной биологии и синтетической химии для улучшения их устойчивости к протеолитической деградации, для изменения специфичности целевой последовательности, для оптимизации характеристик растворимости, для изменения активности белка (например, активности модуляции транскрипции, ферментативной активности и т.д.) или для того, чтобы сделать их более подходящими. Аналоги таких полипептидов включают в себя содержащие остатки, отличные от L-аминокислот природного происхождения, например, D-аминокислоты или синтетические аминокислоты не природного происхождения. D-аминокислоты могут быть замещены на некоторые или все аминокислотные остатки.

Полипептид CasY согласно настоящему изобретению можно готовить посредством *in vitro* синтеза, используя известные в данной области техники традиционные способы. Доступны различные коммерческие аппараты для синтеза, например, автоматизированные синтезаторы от Applied Biosystems, Inc., Beckman и т.д. Используя синтезаторы, можно замещать аминокислоты природного происхождения неприродными аминокислотами. Конкретную последовательность и способ приготовления определяют из соображений удобства, экономической целесообразности, необходимой чистоты и т.п.

При необходимости, во время синтеза или во время экспрессии в пептид можно вносить различные группы, которые обеспечивают связывание с другими молекулами или с поверхностью. Таким образом, цистеины можно использовать для получения тиоэфиров, гистидины - для связывания с металлоионным комплексом, карбоксильные группы - для образования амидов или сложных эфиров, аминокислотные группы - для образования амидов и т.п.

Полипептид CasY согласно настоящему изобретению также можно выделять и очищать в соответствии с традиционными способами рекомбинантного синтеза. Можно готовить лизат от экспрессионного хозяина и очищать этот лизат, используя высокоэффективную жидкостную хроматографию (ВЭЖХ), эксклюзионную хроматографию, гель-электрофорез, аффинную хроматографию или другой способ очистки. В большинстве случаев используемые композиции содержат 20% или более по массе необходимого продукта, чаще 75% или более по массе, предпочтительно 95% или более по массе и в терапевтических целях обычно 99,5% или более по массе в связи с примесными веществами, связанными со способом получения продукта и его очистки.

Обычно процентное значение приведено на основании общего содержания белка. Таким образом, в некоторых случаях полипептид CasY или слитый полипептид CasY согласно настоящему изобретению является по меньшей мере на 80% чистым, по меньшей мере на 85% чистым, по меньшей мере на 90% чистым, по меньшей мере на 95% чистым, по меньшей мере на 98% чистым или по меньшей мере на 99% чистым (например, свободным от примесных веществ, отличных от CasY белков или других макромолекул и т.д.).

Чтобы индуцировать расщепление или любую необходимую модификацию в целевой нуклеиновой кислоте (например, геномной ДНК) или любую необходимую модификацию в полипептиде, связанном с целевой нуклеиновой кислотой, CasY-направляющую РНК и/или полипептид CasY согласно настоящему изобретению и/или донорную матричную последовательность, подлежащие внесению в виде нуклеиновых кислот или полипептидов, вносят в клетки в течение от около 30 мин до около 24 ч, например, 1 ч, 1,5 ч, 2 ч, 2,5 ч, 3 ч, 3,5 ч, 4 ч, 5 ч, 6 ч, 7 часов, 8 ч, 12 ч, 16 ч, 18 ч, 20 ч или в течение любого другого периода времени от около 30 мин до около 24 ч, что можно повторять с частотой от около одного раза в сутки до около одного раза каждые 4 суток, например, каждые 1,5 суток, каждые 2 суток, каждые 3 суток или с любой другой частотой от около одного раза в сутки до около одного раза каждые четверо суток. Агент(ы) можно вносить в предложенные клетки один или более раз, например, один раз, два раза, три раза или более трех раз, и оставлять клетки для инкубации с агентом(ами) в течение некоторого количества времени после каждого контакта, например, на 16-24 ч, после чего среду заменяют свежей средой, а клетки культивируют дальше.

В случаях, в которых в клетку вносят два или более разных нацеливающих комплекса (например, две разные CasY-направляющие РНК, комплементарные разным последовательностям в одной или в разных целевых нуклеиновых кислотах), комплексы можно вносить одновременно (например, в виде двух полипептидов и/или нуклеиновых кислот) или доставлять одновременно. В альтернативном варианте их можно вносить последовательно, например, первым внося нацеливающий комплекс, а за ним - второй нацеливающий комплекс и т.д. или наоборот.

Для улучшения доставки ДНК-вектора в целевую клетку ДНК можно защищать от повреждения и облегчать ее попадание в клетку, например, используя липоплексы и полиплексы. Таким образом, в некоторых случаях нуклеиновая кислота согласно настоящему изобретению (например, рекомбинантный экспрессионный вектор согласно настоящему изобретению) может быть покрыта липидами с упорядо-

ченной структурой, такой как мицеллы или липосомы. Упорядоченную структуру в комплексе с ДНК называют липоплексом. Существует три типа липидов: анионные (отрицательно заряженные), нейтральные или катионные (положительно заряженные). Пригодность липоплексов, в которых используются катионные липиды, для генного переноса была подтверждена. Катионные липиды благодаря своему положительному заряду образуют естественный комплекс с отрицательно заряженной ДНК. Также из-за своего заряда они взаимодействуют с клеточной мембраной. Затем происходит эндоцитоз липоплекса и высвобождение ДНК в цитоплазму. Катионные липиды также защищают от деградации ДНК клеткой.

Комплексы полимеров с ДНК называются полиплексами. Большинство полиплексов состоит из катионных полимеров, а их выработка регулируется ионными взаимодействиями. Одним из основных отличий между способами действия полиплексов и липоплексов является то, что полиплексы не могут высвобождать свою ДНК-нагрузку в цитоплазму, поэтому в данном случае необходима котрансфекция с эндосомолитическими агентами (для лизиса эндосомы, образуемой во время эндоцитоза), такими как инактивированный аденовирус. Однако это не всегда так; полимеры, такие как полиэтиленмин, имеют свой собственный способ разрушения эндосомы, как хитозан и триметилхитозан.

Для генетической модификации стволовых клеток можно также использовать дендримеры, сильно разветвленные макромолекулы сферической формы. Поверхность дендримерной частицы можно функционализировать для изменения ее свойств. В частности, можно сконструировать катионный дендример (т.е. с положительным поверхностным зарядом). В случае присутствия генетического материала, такого как ДНК-плазмида, комплементарность заряда приводит к временной ассоциации нуклеиновой кислоты с катионным дендримером. После достижения своего пункта назначения комплекс дендример-нуклеиновая кислота может поглощаться клеткой посредством эндоцитоза.

В некоторых случаях нуклеиновая кислота согласно изобретению (например, экспрессионный вектор) содержит сайт вставки представляющей интерес направляющей последовательности. Например, нуклеиновая кислота может содержать сайт вставки для представляющей интерес направляющей последовательности, причем сайт вставки расположен непосредственно рядом с нуклеотидной последовательностью, кодирующей часть CasY-направляющей РНК, которая не меняется, когда меняют направляющую последовательность для гибридизации с необходимой целевой последовательностью (например, последовательности, которые обеспечивают CasY-связывающий аспект направляющей РНК, например, последовательности, которые обеспечивают двойные спирали дцРНК CasY-направляющей РНК - эта часть направляющей РНК также может называться "каркасом" или "константной областью" направляющей РНК). Таким образом, в некоторых случаях предложенная нуклеиновая кислота (например, экспрессионный вектор) содержит нуклеотидную последовательность, кодирующую CasY-направляющую РНК, за исключением того, что часть, кодирующая часть направляющей последовательности направляющей РНК, является последовательностью вставки (сайтом вставки). Сайт вставки может представлять собой любую нуклеотидную последовательность, используемую для вставки необходимой последовательности. "Сайты вставки" для применения в различных способах известны специалистам в данной области техники, и можно использовать любой удобный сайт вставки. Сайт вставки может подходить для любого способа манипуляции с последовательностями нуклеиновых кислот. Например, в некоторых случаях сайт вставки представляет собой сайт множественного клонирования (СМК) (например, сайт, содержащий одну или более последовательностей распознавания рестрикционных ферментов), сайт для безлигазного клонирования, сайт для рекомбинационного клонирования (например, рекомбинации на основании сайтов att), нуклеотидную последовательность, распознаваемую технологией CRISPR/Cas (например, Cas9), и т.п.

Сайт вставки может иметь любую необходимую длину, которая может зависеть от типа сайта вставки (например, может зависеть от того, содержит ли (и как много) сайт одну или более последовательностей распознавания рестрикционных ферментов, содержит ли сайт целевой сайт для белка CRISPR/Cas и т.д.). В некоторых случаях сайт вставки предложенной нуклеиновой кислоты имеет длину 3 или более нуклеотидов (нт) (например, длину 5 или более, 8 или более, 10 или более, 15 или более, 17 или более, 18 или более, 19 или более, 20 или более, 25 или более или 30 или более нт). В некоторых случаях длина сайта вставки предложенной нуклеиновой кислоты находится в диапазоне от 2 до 50 нуклеотидов (нт) (например, от 2 до 40 нт, от 2 до 30 нт, от 2 до 25 нт, от 2 до 20 нт, от 5 до 50 нт, от 5 до 40 нт, от 5 до 30 нт, от 5 до 25 нт, от 5 до 20 нт, от 10 до 50 нт, от 10 до 40 нт, от 10 до 30 нт, от 10 до 25 нт, от 10 до 20 нт, от 17 до 50 нт, от 17 до 40 нт, от 17 до 30 нт, от 17 до 25 нт). В некоторых случаях длина сайта вставки предложенной нуклеиновой кислоты находится в диапазоне от 5 до 40 нт.

Модификации нуклеиновых кислот.

В некоторых вариантах реализации предложенная нуклеиновая кислота (например, CasY-направляющая РНК) имеет одну или более модификаций, например, модификацию оснований, модификацию остова и т.д. для придания нуклеиновой кислоте нового или усиленного свойства (например, улучшенной стабильности). Нуклеозид представляет собой комбинацию основания и сахара. Часть основания в нуклеозиде обычно представляет собой гетероциклическое основание. Двумя наиболее распространенными классами таких гетероциклических оснований являются пурины и пиримидины. Нуклеотиды представляют собой нуклеозиды, которые дополнительно содержат фосфатную группу, ковалентно

связанную с сахарной частью нуклеозида. В случае нуклеозидов, которые содержат пентофуранозильный сахар, фосфатная группа может быть связана с 2', 3' или 5' гидроксильным фрагментом сахара. При образовании олигонуклеотидов фосфатные группы ковалентно связывают соседние нуклеозиды друг с другом с образованием линейного полимерного соединения. В свою очередь, соответствующие концы этого линейного полимерного соединения могут быть дополнительно соединены с образованием кольцевого соединения, однако подходящими являются линейные соединения. Кроме того, линейные соединения могут иметь внутреннюю комплементарность нуклеотидных оснований и, следовательно, могут сворачиваться с образованием полностью или частично двухцепочечного соединения. В олигонуклеотидах фосфатные группы обычно относятся к образующим межнуклеозидный остов олигонуклеотида. Обычная связь или остов РНК или ДНК представляет собой 3'-5' фосфодиэфирную связь.

Подходящие модификации нуклеиновых кислот включают в себя, но не ограничиваются этим: 2'-О-метил модифицированные нуклеотиды, 2' фтор-модифицированные нуклеотиды, модифицированные в виде закрытой нуклеиновой кислоты (ЗНК) нуклеотиды, модифицированные в виде пептидой нуклеиновой кислоты (ПНК) нуклеотиды, нуклеотиды с тиофосфатными связями и 5'-кэп (например, 7-метилгуанилатный кэп (m7G)). Дополнительные подробности и дополнительные модификации описаны ниже.

2'-О-Метил-модифицированный нуклеотид (также называемый 2'-О-метил РНК) представляет собой модификацию РНК природного происхождения, которую можно обнаружить в тРНК и других малых РНК и которая возникает на основе посттранскрипционных модификаций. Можно непосредственно синтезировать олигонуклеотиды, которые содержат 2'-О-метил РНК. Эта модификация повышает  $T_m$  двойных спиралей РНК:РНК, но приводит только к небольшим изменениям стабильности РНК:ДНК. Она является стабильной в отношении действия одноцепочечных рибонуклеаз и, как правило, в 5-10 раз менее восприимчива к ДНКазам, чем ДНК. Ее обычно используют в антисмысловых олигонуклеотидах как средство повышения стабильности и аффинности связывания с целевой мРНК.

2' Фтор-модифицированные нуклеотиды (например, 2' фтор-основания) содержат фтор-модифицированную рибозу, которая повышает аффинность связывания ( $T_m$ ) и также придает относительную устойчивость к нуклеазам по сравнению с нативной РНК. Эти модификации обычно применяют в рибозимах и миРНК для улучшения стабильности в сыворотке или других биологических жидкостях.

ЗНК-основания имеют модификацию в рибозном остове, которая закрывает основание в С3'-концевой позиции, что поддерживает геометрию двойной спирали РНК А-типа. Эта модификация существенно повышает  $T_m$  и также является очень устойчивой к нуклеазам. Некоторое количество вставок ЗНК можно размещать в олигонуклеотиде в любой позиции за исключением 3'-конца. Были описаны применения, от антисмысловых олигонуклеотидов до гибридизационных зондов, для выявления ОНП и аллель-специфической ПНР. Вследствие сильного повышения  $T_m$ , вызываемого ЗНК, они также могут приводить к повышению образования димеров праймеров, а также образованию само-шпилек. В некоторых случаях число ЗНК, включенных в один олигонуклеотид, составляет 10 оснований или менее.

Тиофосфатная (ТФ) связь (т.е. тиофосфатное соединение) замещает атом серы немостиковым кислородом в фосфатном остове нуклеиновой кислоты (например, олигонуклеотида). Эта модификация делает межнуклеозидную связь устойчивой к деградации нуклеазами. Тиофосфатные связи можно вносить между последними 3-5 нуклеотидами на 5'- или 3'-конце олигонуклеотида для ингибирования деградации экзонуклеазами. Включение тиофосфатных связей в олигонуклеотид (например, по всему олигонуклеотиду) также может снизить разрушение эндонуклеазами.

В некоторых вариантах реализации предложенная нуклеиновая кислота содержит один или более нуклеотидов, которые являются 2'-О-метил-модифицированными нуклеотидами. В некоторых вариантах реализации предложенная нуклеиновая кислота (например, дцРНК, миРНК и т.д.) содержит один или более 2' фтор-модифицированных нуклеотидов. В некоторых вариантах реализации предложенная нуклеиновая кислота (например, дцРНК, миРНК и т.д.) содержит одно или более ЗНК-оснований. В некоторых вариантах реализации предложенная нуклеиновая кислота (например, дцРНК, миРНК и т.д.) содержит один или более нуклеотидов, связанных тиофосфатной связью (т.е. предложенная нуклеиновая кислота содержит одну или более тиофосфатных связей). В некоторых вариантах реализации предложенная нуклеиновая кислота (например, дцРНК, миРНК и т.д.) содержит 5'-кэп (например, 7-метилгуанилатный кэп (m7G)). В некоторых вариантах реализации предложенная нуклеиновая кислота (например, дцРНК, миРНК и т.д.) содержит комбинацию модифицированных нуклеотидов. Например, предложенная нуклеиновая кислота (например, дцРНК, миРНК и т.д.) может содержать 5'-кэп (например, 7-метилгуанилатный кэп (m7G)) в дополнение к одному или более нуклеотидам с другими модификациями (например, 2'-О-метил-нуклеотиду, и/или 2' фтор-модифицированному нуклеотиду, и/или ЗНК-основанию, и/или тиофосфатной связью).

Модифицированные остовы и модифицированные межнуклеозидные связи.

Примеры подходящих нуклеиновых кислот (например, CasY-направляющих РНК), содержащих модификации, включающих в себя нуклеиновые кислоты, содержащие модифицированные остовы или неприродные межнуклеозидные связи. Нуклеиновые кислоты, имеющие модифицированные остовы, включают в себя те, которые содержат атом фосфора в остове, и те, которые не содержат атом фосфора в

остове.

Подходящие модифицированные олигонуклеотидные остовы, содержащие атом фосфора, включают в себя, например, тиофосфаты, хиральные тиофосфаты, дитиофосфаты, фосфотриэфиры, аминокилфосфотриэфиры, металльные и другие алкильные фосфонаты, включая 3'-алкилен фосфонаты, 5'-алкилен фосфонаты и хиральные фосфонаты, фосфинаты, фосфорамидаты, включая 3'-аминофосфорамидат и аминокилфосфорамидаты, фосфородиамидаты, тионофосфорамидаты, тионоалкилфосфонаты, тионоалкилфосфотриэфиры, селенофосфаты и боранофосфаты, имеющие нормальные 3'-5' связи, их 2'-5' связанные аналоги и имеющие обратную полярность, причем одна или более межнуклеотидных связей являются 3'-3', 5'-5' или 2'-2' связями. Подходящие олигонуклеотиды, имеющие обратную полярность, содержат одинарную 3'-3' связь в крайней 3' межнуклеотидной связи, т.е. один инвертированный нуклеозидный остаток, который может быть основным (отсутствует нуклеобазовое основание или на его месте присутствует гидроксильная группа). Также включены различные соли (такие как, например, калиевые или натриевые), смешанные соли и свободные кислоты.

В некоторых вариантах реализации предложенная нуклеиновая кислота содержит одну или более тиофосфатных и/или содержащих гетероатом межнуклеозидных связей, в частности  $-\text{CH}_2\text{-NH-O-CH}_2-$ ,  $-\text{CH}_2\text{-N(CH}_2\text{)-O-CH}_2-$  (известную как метиленовый (метилимино) или ММІ-остов),  $-\text{CH}_2\text{-O-N(CH}_3\text{)-CH}_2-$ ,  $-\text{CH}_2\text{-N(CH}_3\text{)-N(CH}_3\text{)-CH}_2-$  и  $-\text{O-N(CH}_3\text{)-CH}_2\text{-CH}_2-$  (где нативная фосфодиэфирная межнуклеотидная связь представлена как  $-\text{O-P(=O)(OH)-O-CH}_2-$ ). Межнуклеозидные связи ММІ-типа описаны в вышеупомянутом патенте США № 5489677, содержание которого в полном объеме включено в данный документ посредством ссылки. Подходящие амидные межнуклеозидные связи описаны в патенте США № 5602240, содержание которого в полном объеме включено в данный документ посредством ссылки.

Также подходят нуклеиновые кислоты, имеющие морфолиновые остовные структуры, описанные, например, в патенте США № 5034506. Например, в некоторых вариантах реализации предложенная нуклеиновая кислота содержит 6-членное морфолиновое кольцо вместо рибозного кольца. В некоторых из этих вариантов реализации фосфородиамидатная или другая отличная от фосфодиэфирной межнуклеозидная связь замещает тиофосфатную связь.

Подходящие модифицированные полинуклеотидные остовы, которые не содержат атом фосфора, имеют остовы, образованные короткоцепочечными алкильными или циклоалкильными межнуклеозидными связями, смешанными гетероатомными и алкильными или циклоалкильными межнуклеозидными связями или одной или более короткоцепочечными гетероатомными или гетероциклическими межнуклеозидными связями. Они включают в себя имеющие морфолиновые связи (частично образованные из сахарной части нуклеозида); силоксановые остовы; сульфидные, сульфоксидные и сульфоновые остовы; формацетильные и тиоформацетильные остовы; метилен-формацетильные и тиоформацетильные остовы; рибоацетильные остовы; алкен-содержащие остовы; сульфаматные остовы; метиленимино- и метиленигидразино-остовы; сульфонатные и сульфонамидные остовы; амидные остовы и другие имеющие смешанные N, O, S и  $\text{CH}_2$  составляющие.

Миметики.

Предложенная нуклеиновая кислота может представлять собой миметик нуклеиновой кислоты. Подразумевается, что термин "миметик", применяемый к полинуклеотидам, включает в себя полинуклеотиды, в которых только фуранозное кольцо или фуранозное кольцо и межнуклеотидная связь замещены нефуранозными группами, причем замещение только фуранозного кольца также называется в данной области техники сахарным суррогатом. Гетероциклический фрагмент основания или модифицированный гетероциклический фрагмент основания сохраняют для гибридизации с соответствующей целевой нуклеиновой кислотой. Одна такая нуклеиновая кислота, полинуклеотидный миметик, который, как было показано, обладает превосходными гибридизационными свойствами, называется пептидной нуклеиновой кислотой (ПНК). В ПНК сахарный остов полинуклеотида замещен амид-содержащим остовом, в частности, аминоэтилглициновым остовом. Нуклеотиды удерживаются и связаны прямо или непрямо с аза-атомами азота амидной части остова.

Один полинуклеотидный миметик, который, как сообщалось, обладает превосходными гибридизационными свойствами, представляет собой пептидную нуклеиновую кислоту (ПНК). Остов ПНК-соединений состоит из двух или более связанных аминоэтилглициновых единиц, которые обеспечивают амид-содержащий остов ПНК. Гетероциклические фрагменты основания связаны прямо или непрямо с аза-атомами азота амидной части остова. Типовые патенты США, в которых описано приготовление ПНК-соединений, включают в себя, но не ограничиваются этим: патенты США №№ 5539082; 5714331 и 5719262, содержание которых в полном объеме включено в данный документ посредством ссылки.

Другой изучаемый класс полинуклеотидных миметиков основан на связанных морфолиновых единицах (морфолиновая нуклеиновая кислота), содержащих гетероциклические основания, присоединенные к морфолиновому кольцу. Сообщалось о ряде связующих групп, которые связывают мономерные морфолиновые единицы в морфолиновую нуклеиновую кислоту. Один класс связующих групп был выбран для получения неионных олигомерных соединений. Неионные морфолиновые олигомерные соединения менее вероятно будут вступать в нежелательные взаимодействия с клеточными белками. Морфолиновые полинуклеотиды представляют собой неионные миметики олигонуклеотидов, которые менее

вероятно будут вступать в нежелательные взаимодействия с клеточными белками (Dwayne A. Braasch and David R. Corey, *Biochemistry*, 2002, 41(14), 4503-4510). Морфолиновые полинуклеотиды описаны в патенте США № 5034506, содержание которого в полном объеме включено в данный документ посредством ссылки. Был получен ряд соединений в пределах класса морфолиновых полинуклеотидов, имеющих ряд разных связующих групп, соединяющих мономерные субъединицы.

Дополнительный класс полинуклеотидных миметиков называется циклогексенильными нуклеиновыми кислотами (ЦНК). Фуранозное кольцо, обычно присутствующее в молекуле ДНК/РНК, замещено циклогексенильным кольцом. Получали ДМТ ЦНК-защищенные фосфорамидитные мономеры и использовали для синтеза олигомерного соединения, используя принципы классической фосфорамидитной химии. Получали и исследовали полностью модифицированные ЦНК-олигомерные соединения и олигонуклеотиды, содержащие конкретные модифицированные ЦНК позиции (смотрите Wang et al., *J. Am. Chem. Soc.*, 2000, 122, 8595-8602, содержание которой в полном объеме включено в данный документ посредством ссылки). В общем случае включение мономеров ЦНК в цепь ДНК повышает стабильность гибрида ДНК/РНК. Олигоаденилаты ЦНК образуют комплексы с комплементарными последовательностями РНК и ДНК с такой же стабильностью, что и у нативных комплексов. С помощью исследований ЯМР и кругового дихроизма было показано, что включение структур ЦНК в природные структуры нуклеиновых кислот осуществляется с легкой конформационной адаптацией.

Дополнительные модификации включают в себя закрытые нуклеиновые кислоты (ЗНК), в которых 2'-гидроксильная группа связана с 4' атомом углерода сахарного кольца, формируя, таким образом, 2'-С,4'-С-оксиметиленовую связь, формируя, таким образом, бициклический сахарный фрагмент. Связь может представлять собой метилен (-CH<sub>2</sub>-), группу, соединяющую 2' атом кислорода с 4' атомом углерода, где n равен 1 или 2 (Singh et al., *Chem. Commun.*, 1998, 4, 455-456, содержание которой в полном объеме включено в данный документ посредством ссылки). ЗНК и аналоги ЗНК демонстрируют очень высокую температурную стабильность двойной спирали с комплементарными ДНК и РНК (T<sub>m</sub>=+3 - +10°C), стабильность в отношении 3'-эксонуклеолитической дегградации и хорошие свойства растворимости. Были описаны эффективные и нетоксичные антисмысловые олигонуклеотиды, содержащие ЗНК (например, Wahlestedt et al., *Proc. Natl. Acad. Sci. U.S.A.*, 2000, 97, 5633-5638, содержание которой в полном объеме включено в данный документ посредством ссылки).

Были описаны синтез и приготовление мономеров ЗНК аденина, цитозина, гуанина, 5-метилцитозина, тимина и урацила, наряду с их олигомеризацией и свойствами распознавания нуклеиновых кислот (например, Koshkin et al., *Tetrahedron*, 1998, 54, 3607-3630, содержание которой в полном объеме включено в данный документ посредством ссылки). ЗНК и их приготовление также описаны в WO 98/39352 и WO 99/14226, а также в заявках США 20120165514, 20100216983, 20090041809, 20060117410, 20040014959, 20020094555 и 20020086998, содержание которых в полном объеме включено в данный документ посредством ссылки.

Модифицированные сахарные компоненты.

Предложенная нуклеиновая кислота также может содержать один или более замещенных сахарных фрагментов. Подходящие полинуклеотиды содержат сахарозамещающую группу, выбранную из: OH; F; O-, S- или N-алкила; O-, S- или N-алкенила; O-, S- или N-алкинила или O-алкил-O-алкила, где алкил, алкенил и алкинил могут быть замещенными или незамещенными C<sub>1</sub>-C<sub>10</sub> алкилом или C<sub>2</sub>-C<sub>10</sub> алкенилом и алкинилом. В частности, подходят O((CH<sub>2</sub>)<sub>n</sub>O)<sub>m</sub>CH<sub>2</sub>, O(CH<sub>2</sub>)<sub>n</sub>OCH<sub>3</sub>, O(CH<sub>2</sub>)<sub>n</sub>NH<sub>2</sub>, O(CH<sub>2</sub>)<sub>n</sub>CH<sub>3</sub>, O(CH<sub>2</sub>)<sub>n</sub>ONH<sub>2</sub> и O(CH<sub>2</sub>)<sub>n</sub>ON((CH<sub>2</sub>)<sub>n</sub>CH<sub>3</sub>)<sub>2</sub>, где n и m равны от 1 до 10. Другие подходящие полинуклеотиды содержат сахарозамещающую группу, выбранную из: C<sub>1</sub>-C<sub>10</sub> низшего алкила, замещенного низшего алкила, алкенила, алкинила, алкарила, аралкила, O-алкарила или O-аралкила, SH, SCH<sub>3</sub>, OCN, Cl, Br, CN, CF<sub>3</sub>, OCF<sub>3</sub>, SOCH<sub>3</sub>, SO<sub>2</sub>CH<sub>3</sub>, ONO<sub>2</sub>, NO<sub>2</sub>, N<sub>3</sub>, NH<sub>2</sub>, гетероциклоалкила, гетероциклоалкарила, аминоклиламино, полиалкиламино, замещенного силила, РНК-расщепляющей группы, репортерной группы, интеркалятора, группы для улучшения фармакокинетических свойств олигонуклеотида или группы для улучшения фармакодинамических свойств олигонуклеотида и других заместителей, имеющих аналогичные свойства. Подходящие модификации включают в себя 2'-метоксиэтокси (2'-O-CH<sub>2</sub>CH<sub>2</sub>OCH<sub>3</sub>, также известный как 2'-O-(2-метоксиэтил) или 2'-МОЕ) (Martin et al., *Helv. Chim. Acta*, 1995, 78, 486-504, содержание которой в полном объеме включено в данный документ посредством ссылки), т.е. алкилалкокси-группу. Дополнительные подходящие модификации включают в себя 2'-диметиламинооксиэтокси, т.е. O(CH<sub>2</sub>)<sub>2</sub>ON(CH<sub>3</sub>)<sub>2</sub>-группу, также известную как 2'-DMAOE, описанную в нижеприведенных примерах, и 2'-диметиламиноэтоксиоксиэтокси (также известную в данной области техники как 2'-O-диметиламиноэтоксиэтил или 2'-DMAEOE), т.е. 2'-O-CH<sub>2</sub>-O-CH<sub>2</sub>-N(CH<sub>3</sub>)<sub>2</sub>.

Другие подходящие сахарозамещающие группы включают в себя метокси(-O-CH<sub>2</sub>), аминпропокси (-OCH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>NH<sub>2</sub>), аллил (-CH<sub>2</sub>-CH=CH<sub>2</sub>), -O-аллил(-O-CH<sub>2</sub>-CH=CH<sub>2</sub>) и фтор (F). 2'-сахарозамещающие группы могут находиться в арабино-положении (вверх) или рибо-положении (вниз). Подходящей 2'-арабино-модификацией является 2'-F. Аналогичные модификации также можно осуществлять в других положениях олигомерного соединения, в частности, в 3'-положении сахара на 3'-концевом нуклеозиде или в 2'-5'-связанных олигонуклеотидах и в 5'-положении 5'-концевого нуклеотида. Олигомерные соединения также могут содержать сахарные миметики, такие как циклобутильные фрагменты,

на месте пентофуранозильного сахара.

Модификации и замены оснований.

Предложенная нуклеиновая кислота также может содержать модификации или замены в нуклеосооснованиях (часто называемых в данной области техники просто "основаниями"). В контексте данного документа "немодифицированные" или "природные" нуклеосооснования включают в себя пуриновые основания аденин (А) и гуанин (G) и пиримидиновые основания тимин (Т), цитозин (С) и урацил (U). Модифицированные нуклеосооснования включают в себя другие синтетические и природные нуклеосооснования, такие как 5-метилцитозин (5-me-C), 5-гидроксиметилцитозин, ксантин, гипоксантин, 2-аминоаденин, 6-метил и другие алкильные производные аденина и гуанина, 2-пропил и другие алкильные производные аденина и гуанина, 2-тиоурацил, 2-тиотимин и 2-тиоцитозин, 5-галогенурацил и цитозин, 5-пропинил (-C=C-CH<sub>3</sub>) урацил и цитозин и другие алкильные производные пиримидиновых оснований, 6-азоурацил, цитозин и тимин, 5-урацил (псевдоурацил), 4-тиоурацил, 8-галоген, 8-амино, 8-тиол, 8-тиоалкил, 8-гидроксил и другие 8-замещенные аденины и гуанины, 5-галоген, в частности, 5-бром, 5-трифторметил и другие 5-замещенные урацилы и цитозины, 7-метилгуанин и 7-метиладенин, 2-F-аденин, 2-амино-аденин, 8-азагуанин и 8-азааденин, 7-деазагуанин и 7-деазааденин и 3-деазагуанин и 3-деазааденин. Дополнительные модифицированные нуклеосооснования включают в себя трициклические пиримидины, такие как феноксазинцитидин(1Н-пиримидо(5,4-b)(1,4)бензоксазин-2(3Н)-он), фенотиазинцитидин (1Н-пиримидо(5,4-b)(1,4)бензотиазин-2(3Н)-он), G-зажимы, такие как замещенный феноксазинцитидин (например, 9-(2-аминоэтокси)-Н-пиримидо(5,4-(b)(1,4)бензоксазин-2(3Н)-он), карбазолцитидин (2Н-пиримидо(4,5-b)индол-2-он), пиридоиндолцитидин (Н-пиридо(3',2':4,5)пирроло(2,3-d)пиримидин-2-он).

Гетероциклические фрагменты основания также могут включать в себя те, в которых пуриновое или пиримидиновое основание замещено другими гетероциклами, например, 7-деазааденином, 7-деазагуанозином, 2-аминопиридином и 2-пиридоном. Дополнительные нуклеосооснования включают в себя описанные в патенте США № 3687808, описанные в *The Concise Encyclopedia Of Polymer Science And Engineering*, pages 858-859, Kroschwitz, J. ed. John Wiley & Sons, 1990, описанные в *Englisch et al., Angewandte Chemie, International Edition*, 1991, 30, 613 и описанные *Sanghvi, Y. S., Chapter 15, Antisense Research and Applications*, pages 289-302, Crooke, S. T. and Lebleu, B., ed., CRC Press, 1993; содержание которых в полном объеме включено в данный документ посредством ссылки. Некоторые из этих нуклеосооснований применимы для повышения аффинности связывания олигомерного соединения. Они включают в себя 5-замещенные пиримидины, 6-азапиримидины и N-2, N-6 и O-6-замещенные пурины, включая 2-аминопропиладенин, 5-пропинилурацил и 5-пропинилцитозин. Было показано, что 5-метилцитозинового замены повышают стабильность двойной спирали нуклеиновой кислоты на 0,6-1,2°C (*Sanghvi et al., eds., Antisense Research and Applications*, CRC Press, Boca Raton, 1993, pp. 276-278; содержание которой в полном объеме включено в данный документ посредством ссылки) и являются подходящими заменами оснований, например, в комбинации с 2'-О-метоксиэтильными сахарными модификациями.

Конъюгаты.

Другая возможная модификация предложенной нуклеиновой кислоты включает в себя химическое связывание с полинуклеотидом одного или более фрагментов или конъюгатов, которые повышают активность, клеточное распределение или клеточное поглощение олигонуклеотида. Эти фрагменты или конъюгаты могут включать в себя конъюгатные группы, ковалентно связанные с функциональными группами, такими как первичные или вторичные гидроксильные группы. Конъюгатные группы включают в себя, но не ограничиваются этим, интеркаляторы, репортерные молекулы, полиамины, полиамиды, полиэтиленгликоли, полиэфиры, группы, которые повышают фармакодинамические свойства олигомеров, группы, которые повышают фармакокинетические свойства олигомеров. Подходящие конъюгатные группы включают в себя, но не ограничиваются этим, холестерин, липиды, фосфолипиды, биотин, феназин, фолат, фенантридин, антрахинон, акридин, флуоресцеины, родамины, кумарины и красители. Группы, которые повышают фармакодинамические свойства, включают в себя группы, которые улучшают поглощение, повышают устойчивость к деградации и/или усиливают специфическую в отношении последовательности гибридизацию с целевой нуклеиновой кислотой. Группы, которые повышают фармакокинетические свойства, включают в себя группы, которые улучшают поглощение, распределение, метаболизм или выведение предложенной нуклеиновой кислоты.

Конъюгатные фрагменты включают в себя, но не ограничиваются этим, липидные фрагменты, такие как холестеринный фрагмент (*Letsinger et al., Proc. Natl. Acad. Sci. USA*, 1989, 86, 6553-6556), желтую кислоту (*Manoharan et al., Bioorg. Med. Chem. Lett.*, 1994, 4, 1053-1060), тиоэфир, например, гексил-S-третилтиол (*Manoharan et al., Ann. N.Y. Acad. Sci.*, 1992, 660, 306-309; *Manoharan et al., Bioorg. Med. Chem. Lett.*, 1993, 3, 2765-2770), тиохолестерин (*Oberhauser et al., Nucl. Acids Res.*, 1992, 20, 533-538), алифатическую цепь, например, остаток додекандиола или ундецила (*Saison-Behmoaras et al., EMBO J.*, 1991, 10, 1111-1118; *Kabanov et al., FEBS Lett.*, 1990, 259, 327-330; *Svinarchuk et al., Biochimie*, 1993, 75, 49-54), фосфолипид, например, дигексадецил-рац-глицерин или триэтиламмоний 1,2-ди-О-гексадецил-рац-глицеро-3-Н-фосфонат (*Manoharan et al., Tetrahedron Lett.*, 1995, 36, 3651-3654; *Shea et al., Nucl. Acids*

Res., 1990, 18, 3777-3783), цепь полиамина или полиэтиленгликоля (Manoharan et al., *Nucleosides & Nucleotides*, 1995, 14, 969-973) или адамантан-уксусную кислоту (Manoharan et al., *Tetrahedron Lett.*, 1995, 36, 3651-3654), пальмитиловый фрагмент (Mishra et al., *Biochim. Biophys. Acta*, 1995, 1264, 229-237) или октадециламиноновый или гексиламинокарбонил-оксихолестеринный фрагмент (Crooke et al., *J. Pharmacol. Exp. Ther.*, 1996, 277, 923-937).

Конъюгат может содержать "домен белковой трансдукции" или ДБТ (также известный как КПП - клеточно-проникающий пептид), который относится к полипептиду, полинуклеотиду, углеводу или органическому или неорганическому соединению, которое облегчает перемещение через липидный бислой, мицеллы, клеточную мембрану, мембрану органелл или везикулярную мембрану. ДБТ, присоединенный к другой молекуле, которая может представлять собой от малой полярной молекулы до большой макромолекулы и/или наночастицы, облегчает перемещение молекулы через мембрану, например, при переходе из внеклеточного пространства во внутриклеточное пространство или из цитозоля в органеллу (например, ядро). В некоторых вариантах реализации ДБТ ковалентно связан с 3' концом экзогенного полинуклеотида. В некоторых вариантах реализации ДБТ ковалентно связан с 5' концом экзогенного полинуклеотида. Примеры ДБТ включают в себя, но не ограничиваются этим, минимальный ундекапептидный домен белковой трансдукции (соответствующий остаткам 47-57 ТАТ ВИЧ-1, составляющим YGRKKRRQRRR; SEQ ID NO:112); полиаргининовую последовательность, содержащую достаточное число аргининов для непосредственного попадания в клетку (например, 3, 4, 5, 6, 7, 8, 9, 10 или 10-50 аргининов); домен VP22 (Zender et al. (2002) *Cancer Gene Ther.* 9(6):489-96); домен белковой трансдукции *Antennapedia* дрозофилы (Noguchi et al. (2003) *Diabetes* 52(7):1732-1737); усеченный пептид человеческого кальцитонина (Trehin et al. (2004) *Pharm. Research* 21:1248-1256); полилизин (Wender et al. (2000) *Proc. Natl. Acad. Sci. USA* 97:13003-13008); RRQRRTSKLMKR SEQ ID NO:113); транспортан GWTLNSAGYLLGKINLKALAALAKKIL SEQ ID NO:114); KALAWKAKLAKALAKALAKHLAKALAKALKCEA SEQ ID NO:115) и RQIKIWFQNRRMKWKK SEQ ID NO:116).

Типовые ДБТ включают в себя, но не ограничиваются этим, YGRKKRRQRRR (SEQ ID NO:117), RKKRRQRRR (SEQ ID NO:118); аргининовый гомополимер от 3 остатков аргинина до 50 остатков аргинина; типовые аминокислотные последовательности домена ДБТ включают в себя, но не ограничиваются этим, любые из нижеприведенных:

YGRKKRRQRRR SEQ ID NO:119);  
RKKRRQRR SEQ ID NO:120); YARAAARQARA SEQ ID NO:121); THRLPRRRRRR SEQ ID NO:122) и GRRARRRRRR SEQ ID NO:123).

В некоторых вариантах реализации ДБТ представляет собой активируемый КПП (АКПП) (Aguilera et al. (2009) *Integr Biol (Camb)* June; 1(5-6):371-381). АКПП содержат поликатионный КПП (например, Arg9 или "R9"), связанный посредством расщепляемого линкера с соответствующим полианионом (например, Glu9 или "E9"), который снижает суммарный заряд практически до нуля и тем самым ингибирует адгезию и поглощение клетками. После расщепления линкера происходит высвобождение полианиона, что приводит к локальной демаскировке полиаргинина и его природной адгезивности, "активируя", таким образом, перемещение АКПП через мембрану.

Внесение компонентов в целевую клетку.

CasY-направляющую РНК (или нуклеиновую кислоту, содержащую кодирующую ее нуклеотидную последовательность), и/или полипептид CasY согласно настоящему изобретению (или нуклеиновую кислоту, содержащую кодирующую его нуклеотидную последовательность), и/или слитый полипептид CasY согласно настоящему изобретению (или нуклеиновую кислоту, которая содержит нуклеотидную последовательность, кодирующую слитый полипептид CasY согласно настоящему изобретению), и/или донорный полинуклеотид (донорную матрицу) можно вносить в клетку-хозяина любым из ряда хорошо известных способов.

Любые из множества соединений и способов можно использовать для доставки в целевую клетку системы CasY согласно настоящему изобретению, например, при этом система CasY содержит: а) полипептид CasY согласно настоящему изобретению и CasY-направляющую РНК; б) полипептид CasY согласно настоящему изобретению, CasY-направляющую РНК и донорную матричную нуклеиновую кислоту; в) слитый полипептид CasY согласно настоящему изобретению и CasY-направляющую РНК; г) слитый полипептид CasY согласно настоящему изобретению, CasY-направляющую РНК и донорную матричную нуклеиновую кислоту; е) мРНК, кодирующую полипептид CasY согласно настоящему изобретению, и CasY-направляющую РНК; ф) мРНК, кодирующую полипептид CasY согласно настоящему изобретению, CasY-направляющую РНК и донорную матричную нуклеиновую кислоту; г) мРНК, кодирующую слитый полипептид CasY согласно настоящему изобретению, и CasY-направляющую РНК; д) мРНК, кодирующую слитый полипептид CasY согласно настоящему изобретению, CasY-направляющую РНК и донорную матричную нуклеиновую кислоту; и) рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобре-

тению, и нуклеотидную последовательность, кодирующую CasY-направляющую РНК; j) рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобретению, нуклеотидную последовательность, кодирующую CasY-направляющую РНК, и нуклеотидную последовательность, кодирующую донорную матричную нуклеиновую кислоту; k) рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую слитый полипептид CasY согласно настоящему изобретению, и нуклеотидную последовательность, кодирующую CasY-направляющую РНК; l) рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую слитый полипептид CasY согласно настоящему изобретению, нуклеотидную последовательность, кодирующую CasY-направляющую РНК, и нуклеотидную последовательность, кодирующую донорную матричную нуклеиновую кислоту; m) первый рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобретению, и второй рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую CasY-направляющую РНК; n) первый рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобретению, и второй рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую CasY-направляющую РНК, и донорную матричную нуклеиновую кислоту; o) первый рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую слитый полипептид CasY согласно настоящему изобретению, и второй рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую CasY-направляющую РНК; p) первый рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую слитый полипептид CasY согласно настоящему изобретению, и второй рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую CasY-направляющую РНК, и донорную матричную нуклеиновую кислоту; q) рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобретению, нуклеотидную последовательность, кодирующую первую CasY-направляющую РНК, и нуклеотидную последовательность, кодирующую вторую CasY-направляющую РНК; или r) рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую слитый полипептид CasY согласно настоящему изобретению, нуклеотидную последовательность, кодирующую первую CasY-направляющую РНК, и нуклеотидную последовательность, кодирующую вторую CasY-направляющую РНК; или некоторую вариацию (a)-(r). В качестве неограничивающего примера систему CasY согласно настоящему изобретению можно комбинировать с липидом. В качестве другого неограничивающего примера систему CasY согласно настоящему изобретению можно комбинировать с частицей или инкапсулировать в частице.

Способы внесения нуклеиновой кислоты в клетку-хозяина известны в данной области техники, а для внесения предложенной нуклеиновой кислоты (например, экспрессионной конструкции/экспрессионного вектора) в целевую клетку (например, прокариотическую клетку, эукариотическую клетку, клетку растения, клетку животного, клетку млекопитающего, клетку человека и т.п.) можно использовать любой удобный способ. Подходящие способы включают в себя, например, вирусную инфекцию, трансфекцию, конъюгацию, слияние протопластов, липофекцию, электропорацию, осаждение фосфатом кальция, полиэтиленимин (ПЭИ)-опосредованную трансфекцию, ДЭАЭ-декстран-опосредованную трансфекцию, липосомно-опосредованную трансфекцию, технологию генной пушки, осаждение фосфатом кальция, прямую микроинъекцию, опосредованную наночастицами доставку нуклеиновой кислоты (смотрите, например, Panyam et al., *Adv Drug Deliv Rev.* 2012 Sep 13. pii: S0169-409X(12)00283-9. doi: 10.1016/j.addr.2012.09.023) и т.п.

В некоторых случаях полипептид CasY согласно настоящему изобретению получен в виде нуклеиновой кислоты (например, мРНК, ДНК, плазмиды, экспрессионного вектора, вирусного вектора и т.д.), которая кодирует полипептид CasY. В некоторых случаях полипептид CasY согласно настоящему изобретению получен непосредственно в виде белка (например, без связанной с ним направляющей РНК или со связанной с ним направляющей РНК, т.е. в виде рибонуклеопротеинового комплекса). Полипептид CasY согласно настоящему изобретению можно вносить в клетку (вводить в клетку) любым удобным способом; а такие способы известны специалистам в данной области техники. В качестве иллюстративного примера полипептид CasY согласно настоящему изобретению можно вводить непосредственно в клетку (например, с или без CasY-направляющей РНК или нуклеиновой кислоты, кодирующей CasY-направляющую РНК, и с или без донорного полинуклеотида). В качестве другого примера заранее сформированный комплекс полипептида CasY согласно настоящему изобретению и CasY-направляющей РНК (РНП) можно вносить в клетку (например, эукариотическую клетку) (например, посредством инъекции, посредством нуклеофекции; посредством домена белковой трансдукции (ДБТ), конъюгированного с одним или более компонентами, например, конъюгированного с белком CasY, конъюгированного с направляющей РНК, конъюгированного с полипептидом CasY согласно настоящему изобретению и направляющей РНК; и т.д.).

В некоторых случаях слитый полипептид CasY (например, dCasY, слитый с партнером по слиянию, никазный CasY, слитый партнером по слиянию, и т.д.) согласно настоящему изобретению получен в виде нуклеиновой кислоты (например, мРНК, ДНК, плазмиды, экспрессионного вектора, вирусного вектора и

т.д.), которая кодирует слитый полипептид CasY. В некоторых случаях слитый полипептид CasY согласно настоящему изобретению получен непосредственно в виде белка (например, без связанной с ним направляющей РНК или со связанной с ним направляющей РНК, т.е. в виде рибонуклеопротеинового комплекса). Слитый полипептид CasY согласно настоящему изобретению можно вносить в клетку (вводить в клетку) любым удобным способом; а такие способы известны специалистам в данной области техники. В качестве иллюстративного примера слитый полипептид CasY согласно настоящему изобретению можно вводить непосредственно в клетку (например, с или без нуклеиновой кислоты, кодирующей CasY-направляющую РНК, и с или без донорного полинуклеотида). В качестве другого примера заранее сформированный комплекс слитого полипептида CasY согласно настоящему изобретению и CasY-направляющей РНК (РНП) можно вносить в клетку (например, посредством инъекции, посредством нуклеофекции; посредством домена белковой трансдукции (ДБТ), конъюгированного с одним или более компонентами, например, конъюгированного со слитым белком CasY, конъюгированного с направляющей РНК, конъюгированного со слитым полипептидом CasY согласно настоящему изобретению и направляющей РНК; и т.д.).

В некоторых случаях нуклеиновую кислоту (например, CasY-направляющую РНК; нуклеиновую кислоту, содержащую нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобретению; и т.д.) доставляют в клетку (например, целевую клетку-хозяина) и/или полипептид (например, полипептид CasY; слитый полипептид CasY) в частице или связанной с частицей. В некоторых случаях систему CasY согласно настоящему изобретению доставляют в клетку в частице или связанной с частицей. Термины "частица" и "наночастица" можно употреблять взаимозаменяемо, в зависимости от ситуации. Рекombинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобретению и/или CasY-направляющую РНК, мРНК, содержащую нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобретению, и направляющую РНК можно вводить одновременно, используя частицы или липидные оболочки; например, полипептид CasY и CasY-направляющую РНК, например, в виде комплекса (например, рибонуклеопротеинового (РНП) комплекса), можно доставлять посредством частицы, например, частицы для доставки, содержащей липид или липидоид и гидрофильный полимер, например, катионный липид и гидрофильный полимер, например, при этом катионный липид содержит 1,2-диолеил-3-триметиламмонийпропан (DOTAP) или 1,2-дигетрадеканол-сн-глицеро-3-фосфохолин (DMPC), и/или при этом гидрофильный полимер содержит этиленгликоль или полиэтиленгликоль (ПЭГ); и/или при этом частица дополнительно содержит холестерин (например, частица из состава 1=DOTAP 100, DMPC 0, ПЭГ 0, холестерин 0; состава номер 2=DOTAP 90, DMPC 0, ПЭГ 10, холестерин 0; состава номер 3=DOTAP 90, DMPC 0, ПЭГ 5, холестерин 5). Например, частицу можно получать, используя многоэтапный процесс, в котором смешивают вместе полипептид CasY и CasY-направляющую РНК, например, в молярном соотношении 1:1, например, при комнатной температуре, например, в течение 30 мин, например, в стерильном, не содержащем нуклеазы 1 x фосфатном буферном солевом растворе (ФСБ); а DOTAP, DMPC, ПЭГ и холестерин, в зависимости от состава, отдельно растворяют в спирте, например, 100% этаноле; и смешивают два раствора вместе для образования частиц, содержащих комплексы.

Полипептид CasY согласно настоящему изобретению (или мРНК, содержащую нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобретению; или рекombинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобретению) и/или CasY-направляющую РНК (или нуклеиновую кислоту, такую как один или более экспрессионных векторов, кодирующих CasY-направляющую РНК) можно доставлять одновременно, используя частицы или липидные оболочки. Например, можно использовать биоразлагаемые структурированные наночастицы типа ядро-оболочка с ядром из поли( $\beta$ -аминоэфира) (РВАЕ) в оболочке из фосфолипидного бислоя. В некоторых случаях используют частицы/наночастицы на основе самособираемых биоадгезивных полимеров; такие частицы/наночастицы можно применять для пероральной доставки пептидов, внутривенной доставки пептидов и назальной доставки пептидов, например, в головной мозг. Также предусматриваются другие варианты реализации, такие как пероральное всасывание или глазная доставка гидрофобных лекарственных препаратов. Можно использовать технологию молекулярной оболочки, которая включает в себя применение сконструированной защищенной полимерной оболочки, которая доставляется в болезненный участок. Можно использовать дозы около 5 мг/кг, однократные или многократные, в зависимости от различных факторов, например, целевой ткани.

Липидоидные соединения (например, описанные в патентной заявке США 20110293703) также применимы при введении полинуклеотидов и могут использоваться для доставки полипептида CasY согласно настоящему изобретению, слитого полипептида CasY согласно настоящему изобретению, РНП согласно настоящему изобретению, нуклеиновой кислоты согласно настоящему изобретению или системы CasY согласно настоящему изобретению, например, при этом система CasY содержит: а) полипептид CasY согласно настоящему изобретению и CasY-направляющую РНК; б) полипептид CasY согласно настоящему изобретению, CasY-направляющую РНК и донорную матричную нуклеиновую кислоту; в) слитый полипептид CasY согласно настоящему изобретению и CasY-направляющую РНК; д) слитый

полипептид CasY согласно настоящему изобретению, CasY-направляющую РНК и донорную матричную нуклеиновую кислоту; е) мРНК, кодирующую полипептид CasY согласно настоящему изобретению, и CasY-направляющую РНК; ф) мРНК, кодирующую полипептид CasY согласно настоящему изобретению, CasY-направляющую РНК и донорную матричную нуклеиновую кислоту; г) мРНК, кодирующую слитый полипептид CasY согласно настоящему изобретению, и CasY-направляющую РНК; h) мРНК, кодирующую слитый полипептид CasY согласно настоящему изобретению, CasY-направляющую РНК и донорную матричную нуклеиновую кислоту; i) рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобретению, и нуклеотидную последовательность, кодирующую CasY-направляющую РНК; j) рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобретению, нуклеотидную последовательность, кодирующую CasY-направляющую РНК, и нуклеотидную последовательность, кодирующую донорную матричную нуклеиновую кислоту; k) рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую слитый полипептид CasY согласно настоящему изобретению, и нуклеотидную последовательность, кодирующую CasY-направляющую РНК; l) рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую слитый полипептид CasY согласно настоящему изобретению, нуклеотидную последовательность, кодирующую CasY-направляющую РНК, и нуклеотидную последовательность, кодирующую донорную матричную нуклеиновую кислоту; m) первый рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобретению, и второй рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую CasY-направляющую РНК; n) первый рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобретению, и второй рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую CasY-направляющую РНК, и донорную матричную нуклеиновую кислоту; o) первый рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую слитый полипептид CasY согласно настоящему изобретению, и второй рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую CasY-направляющую РНК; p) первый рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую слитый полипептид CasY согласно настоящему изобретению, и второй рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую CasY-направляющую РНК, и донорную матричную нуклеиновую кислоту; q) рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобретению, нуклеотидную последовательность, кодирующую первую CasY-направляющую РНК, и нуклеотидную последовательность, кодирующую вторую CasY-направляющую РНК; или r) рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую слитый полипептид CasY согласно настоящему изобретению, нуклеотидную последовательность, кодирующую первую CasY-направляющую РНК, и нуклеотидную последовательность, кодирующую вторую CasY-направляющую РНК; или некоторую вариацию (a)-(r). В одном аспекте аминокислотные липидоидные соединения комбинируют с агентом, который необходимо доставить в клетку или организм субъекта, с образованием микрочастиц, наночастиц, липосом или мицелл. Аминокислотные липидоидные соединения можно комбинировать с другими аминокислотными липидоидными соединениями, полимерами (синтетическими или природными), поверхностно-активными веществами, холестерином, углеводами, белками, липидами и т.д. для образования частиц. Затем эти частицы можно необязательно комбинировать с фармацевтическим вспомогательным веществом для образования фармацевтической композиции.

Поли(бета-аминоспирт) (РВАА) можно использовать для доставки полипептида CasY согласно настоящему изобретению, слитого полипептида CasY согласно настоящему изобретению, РНП согласно настоящему изобретению, нуклеиновой кислоты согласно настоящему изобретению или системы CasY согласно настоящему изобретению в целевую клетку. В патентной публикации США № 20130302401 описан класс поли(бета-аминоспиртов) (РВАА), которые готовили с помощью комбинаторной полимеризации.

Можно использовать частицы на основе сахара, например, GalNAc, описанные по ссылке на WO2014118272 (включенную в данный документ посредством ссылки) и Nair, J K et al., 2014, Journal of the American Chemical Society 136 (49), 16958-16961), можно использовать для доставки полипептида CasY согласно настоящему изобретению, слитого полипептида CasY согласно настоящему изобретению, РНП согласно настоящему изобретению, нуклеиновой кислоты согласно настоящему изобретению или системы CasY согласно настоящему изобретению в целевую клетку.

В некоторых случаях используют липидные наночастицы (ЛНЧ) для доставки полипептида CasY согласно настоящему изобретению, слитого полипептида CasY согласно настоящему изобретению, РНП согласно настоящему изобретению, нуклеиновой кислоты согласно настоящему изобретению или системы CasY согласно настоящему изобретению в целевую клетку. Отрицательно заряженные полимеры, такие как РНК, можно загружать в ЛНЧ при низких значениях pH (например, pH 4), когда ионизируемые

липиды проявляют положительный заряд. Однако при физиологических значениях pH ЛНЧ демонстрируют низкий поверхностный заряд, совместимый с большим временем нахождения в циркуляции. В фокусе находятся четыре вида ионизируемых катионных липидов, а именно 1,2-дидолеил-3-диметиламмонийпропан (DLinDAP), 1,2-дидолеилокси-3-N,N-диметиламинопропан (DLinDMA), 1,2-дидолеилокси-кето-N,N-диметил-3-аминопропан (DLinKDMA) и 1,2-дидолеил-4-(2-диметиламиноэтил)-[1,3]-диоксолан (DLinKC2-DMA). Получение ЛНЧ описано, например, в Rosin et al. (2011) *Molecular Therapy* 19:1286-2200). Можно использовать катионные липиды 1,2-дидолеил-3-диметиламмонийпропан (DLinDAP), 1,2-дидолеилокси-3-N,N-диметиламинопропан (DLinDMA), 1,2-дидолеилокси-кето-N,N-диметил-3-аминопропан (DLinK-DMA), 1,2-дидолеил-4-(2-диметиламиноэтил)-[1,3]-диоксолан (DLinKC2-DMA), (3-о-[2"-(метоксиполиэтиленгликоль 2000)сукциноил]-1,2-димиристоил-сн-гликоль (PEG-S-DMG) и R-3-[(ω-метокси-поли(этиленгликоль)2000) карбамоил]-1,2-димиристоилоксипропил-3-амин (PEG-C-DOMG). Нуклеиновая кислота (например, CasY-направляющая РНК; нуклеиновая кислота согласно настоящему изобретению и т.д.) может быть инкапсулирована в ЛНЧ, содержащих DLinDAP, DLinDMA, DLinK-DMA и DLinKC2-DMA (катионный липид: DSPC:CHOL: PEGS-DMG или PEG-C-DOMG в молярном соотношении 40:10:40:10). В некоторых случаях включают 0,2% SP-DiOC18.

Конструкции сферических нуклеиновых кислот (СНК<sup>TM</sup>) и другие наночастицы (в частности, наночастицы золота) можно использовать для доставки полипептида CasY согласно настоящему изобретению, слитого полипептида CasY согласно настоящему изобретению, РНП согласно настоящему изобретению, нуклеиновой кислоты согласно настоящему изобретению или системы CasY согласно настоящему изобретению в целевую клетку. Смотрите, например, Cutler et al., *J. Am. Chem. Soc.* 2011 133:9254-9257, Hao et al., *Small*. 2011:3158-3162, Zhang et al., *ACS Nano*. 2011 5:6962-6970, Cutler et al., *J. Am. Chem. Soc.* 2012 134:1376-1391, Young et al., *Nano Lett.* 2012 12:3867-71, Zheng et al., *Proc. Natl. Acad. Sci. USA*. 2012 109:11975-80, Mirkin, *Nanomedicine* 2012 7:635-638 Zhang et al., *J. Am. Chem. Soc.* 2012 134:16488-1691, Weintraub, *Nature* 2013 495:S14-S16, Choi et al., *Proc. Natl. Acad. Sci. USA*. 2013 110(19): 7625-7630, Jensen et al., *Sci. Transl. Med.* 5, 209ra152 (2013) и Mirkin, et al., *Small*, 10:186-192.

Самособираемые наночастицы с РНК можно конструировать с полиэтиленимином (ПЭИ), ПЭГилированным пептидным лигандом Arg-Gly-Asp (RGD), присоединенным на дистальном конце полиэтиленгликоля (ПЭГ).

В общем случае "наночастица" относится к любой частице, имеющей диаметр менее 1000 нм. В некоторых случаях наночастицы, подходящие для применения для доставки полипептида CasY согласно настоящему изобретению, слитого полипептида CasY согласно настоящему изобретению, РНП согласно настоящему изобретению, нуклеиновой кислоты согласно настоящему изобретению или системы CasY согласно настоящему изобретению в целевую клетку, имеют диаметр 500 нм или менее, например, от 25 нм до 35 нм, от 35 нм до 50 нм, от 50 нм до 75 нм, от 75 нм до 100 нм, от 100 нм до 150 нм, от 150 нм до 200 нм, от 200 нм до 300 нм, от 300 нм до 400 нм или от 400 нм до 500 нм. В некоторых случаях наночастицы, подходящие для применения для доставки полипептида CasY согласно настоящему изобретению, слитого полипептида CasY согласно настоящему изобретению, РНП согласно настоящему изобретению, нуклеиновой кислоты согласно настоящему изобретению или системы CasY согласно настоящему изобретению в целевую клетку, имеют диаметр от 25 нм до 200 нм. В некоторых случаях наночастицы, подходящие для применения для доставки полипептида CasY согласно настоящему изобретению, слитого полипептида CasY согласно настоящему изобретению, РНП согласно настоящему изобретению, нуклеиновой кислоты согласно настоящему изобретению или системы CasY согласно настоящему изобретению в целевую клетку, имеют диаметр 100 нм или менее. В некоторых случаях наночастицы, подходящие для применения для доставки полипептида CasY согласно настоящему изобретению, слитого полипептида CasY согласно настоящему изобретению, РНП согласно настоящему изобретению, нуклеиновой кислоты согласно настоящему изобретению или системы CasY согласно настоящему изобретению в целевую клетку, имеют диаметр от 35 нм до 60 нм.

Наночастицы, подходящие для применения для доставки полипептида CasY согласно настоящему изобретению, слитого полипептида CasY согласно настоящему изобретению, РНП согласно настоящему изобретению, нуклеиновой кислоты согласно настоящему изобретению или системы CasY согласно настоящему изобретению в целевую клетку, могут быть получены в разных формах, например, в виде твердых наночастиц (например, металла, такого как серебро, золото, железо, титан), не-металла, липидных твердых веществ, полимеров), суспензий наночастиц или их комбинаций. Можно готовить металлические, диэлектрические или полупроводниковые наночастицы, а также гибридные структуры (например, наночастицы типа ядро-оболочка). Наночастицы, полученные из полупроводникового материала, также могут представлять собой меченные квантовые точки, если они достаточно невелики (как правило меньше 10 нм) для того, чтобы происходило квантование энергетических электронных уровней. Такие наномасштабные частицы используют в биомедицинских применениях в качестве носителей лекарственных препаратов или визуализирующих агентов, и их можно адаптировать в тех же целях в настоящем изобретении.

Полутвердые и мягкие наночастицы также подходят для применения для доставки полипептида CasY согласно настоящему изобретению, слитого полипептида CasY согласно настоящему изобретению, РНП согласно настоящему изобретению, нуклеиновой кислоты согласно настоящему изобретению или системы CasY согласно настоящему изобретению в целевую клетку. Прототипом полутвердой наночастицы является липосома.

В некоторых случаях используют экзосомы для доставки полипептида CasY согласно настоящему изобретению, слитого полипептида CasY согласно настоящему изобретению, РНП согласно настоящему изобретению, нуклеиновой кислоты согласно настоящему изобретению или системы CasY согласно настоящему изобретению в целевую клетку. Экзосомы представляют собой эндогенные нановезикулы, которые переносят РНК и белки, и которые могут доставлять РНК в головной мозг и другие целевые органы.

В некоторых случаях используют липосомы для доставки полипептида CasY согласно настоящему изобретению, слитого полипептида CasY согласно настоящему изобретению, РНП согласно настоящему изобретению, нуклеиновой кислоты согласно настоящему изобретению или системы CasY согласно настоящему изобретению в целевую клетку. Липосомы представляют собой сферические везикулярные структуры, состоящие из одно- или многослойного липидного бислоя, окружающего внутренние водные компартменты, и относительно непроницаемого внешнего липофильного фосфолипидного бислоя. Липосомы можно получать из нескольких разных типов липидов; однако для создания липосом наиболее часто используют фосфолипиды. Хотя образование липосом является спонтанным, когда липидную пленку смешивают с водным раствором, его можно ускорить, применяя силу в форме встряхивания с помощью гомогенизатора, соникатора или экструзионного аппарата. К липосомам можно добавлять несколько других добавок, чтобы модифицировать их структуру и свойства. Например, в смесь липосом можно добавлять холестерин или сфингомиелин, чтобы способствовать стабилизации липосомной структуры и чтобы предотвратить протекание внутреннего содержимого липосомы. Липосомный состав может состоять, главным образом, из природных фосфолипидов и липидов, таких как 1,2-дистеароил-сн-глицеро-3-фосфатидилхолин (DSPC), сфингомиелин, фосфатидилхолины яиц и моносиалоганглиозид.

Стабильные частицы нуклеиновая кислота-липид (SNALP) можно использовать для доставки полипептида CasY согласно настоящему изобретению, слитого полипептида CasY согласно настоящему изобретению, РНП согласно настоящему изобретению, нуклеиновой кислоты согласно настоящему изобретению или системы CasY согласно настоящему изобретению в целевую клетку. Состав SNALP может содержать липиды 3-N-[(метоксиполи(этиленгликоль)2000)карбамоил]-1,2-димиристилокси-пропиламин (PEG-C-DMA), 1,2-дидолеилокси-N,N-диметил-3-аминопропан (DLinDMA), 1,2-дистеароил-сн-глицеро-3-фосфохолин (DSPC) и холестерин в молярном процентном соотношении 2:40:10:48. Липосомы SNALP можно готовить, смешивая D-Lin-DMA и PEG-C-DMA с дистеароилфосфатидилхолином (DSPC), холестерином и миРНК, используя соотношение 25:1 липид/миРНК и молярное соотношение 48/40/10/2 холестерин/D-Lin-DMA/DSPC/PEG-C-DMA. Получаемые в результате липосомы SNALP могут иметь размер около 80-100 нм. SNALP может содержать синтетический холестерин (Sigma-Aldrich, Сент-Луис, штат Миссури, США), дипальмитоилфосфатидилхолин (Avanti Polar Lipids, Алабама, штат Алабама, США), 3-N-[(w-метоксиполи(этиленгликоль)2000)карбамоил]-1,2-димиристоилоксипропиламин и катионный 1,2-дидолеилокси-3-N,N-диметиламинопропан. SNALP может содержать синтетический холестерин (Sigma-Aldrich), 1,2-дистеароил-сн-глицеро-3-фосфохолин (DSPC; Avanti Polar Lipids Inc.), PEG-cDMA и 1,2-дидолеилокси-3-(N,N-диметил)аминопропан (DLinDMA).

Другие катионные липиды, такие как аминоклипид 2,2-дидолеил-4-диметиламиноэтил-[1,3]-диоксолан (DLin-KC2-DMA) можно использовать для доставки полипептида CasY согласно настоящему изобретению, слитого полипептида CasY согласно настоящему изобретению, РНП согласно настоящему изобретению, нуклеиновой кислоты согласно настоящему изобретению или системы CasY согласно настоящему изобретению в целевую клетку. Может предусматриваться заранее сформированная везикула со следующей липидной композицией: аминоклипид, дистеароилфосфатидилхолин (DSPC), холестерин и (R)-2,3-бис(октадецилокси)пропил-1-(метоксиполи(этиленгликоль)2000)пропилкарбамат (ПЭГ-липид) в молярном соотношении 40/10/40/10, соответственно, и с соотношением FVII миРНК/общий липид приблизительно 0,05 мас./мас. Чтобы гарантировать узкое распределение частиц по размеру в диапазоне 70-90 нм и низкий индекс полидисперсности  $0,11 \pm 0,04$  ( $n=56$ ), частицы можно до трех раз экструдировать через 80 нм мембраны перед добавлением направляющей РНК. Можно использовать частицы, содержащие высокоэффективный аминоклипид 16, в котором молярное соотношение четырех липидных компонентов 16, DSPC, холестерина и ПЭГ-липид (50/10/38,5/1,5), которое можно дополнительно оптимизировать, чтобы усилить *in vivo* активность.

Липиды можно смешивать с системой CasY согласно настоящему изобретению, или ее компонентом(ами), или кодирующими их нуклеиновыми кислотами для образования липидных наночастиц (ЛНЧ). Подходящие липиды включают в себя, но не ограничиваются этим, DLin-KC2-DMA4, C12-200 и колипиды дистеароилфосфатидилхолина, холестерин и PEG-DMG можно смешивать с системой CasY или ее компонентом согласно настоящему изобретению, используя процедуру спонтанного образования вези-

кул. Молярное соотношение компонентов может составлять около 50/10/38,5/1,5 (DLin-KC2-DMA или C12-200/ дистеароилфосфатидилхолин/ холестерин/PEG-DMG).

Систему CasY согласно настоящему изобретению или ее компонент можно доставлять инкапсулированными в микросферах PLGA, таких как дополнительно описаны в опубликованных заявках США 20130252281, 20130245107 и 20130244279.

Суперзаряженные белки можно использовать для доставки полипептида CasY согласно настоящему изобретению, слитого полипептида CasY согласно настоящему изобретению, РНП согласно настоящему изобретению, нуклеиновой кислоты согласно настоящему изобретению или системы CasY согласно настоящему изобретению в целевую клетку. Суперзаряженные белки представляют собой класс сконструированных или природных белков с необычно высоким положительным или отрицательным суммарным теоретическим зарядом. Как суперотрицательно, так и суперположительно заряженные белки демонстрируют способность противостоять термически или химически индуцируемой агрегации. Суперположительно заряженные белки также способны проникать в клетки млекопитающих. Связывание с этими белками нагрузки, такой как плазмидная ДНК, РНК или другие белки, может обеспечить функциональную доставку этих макромолекул в клетки млекопитающих как *in vitro*, так и *in vivo*.

Клеточно-проникающие пептиды (КПП) можно использовать для доставки полипептида CasY согласно настоящему изобретению, слитого полипептида CasY согласно настоящему изобретению, РНП согласно настоящему изобретению, нуклеиновой кислоты согласно настоящему изобретению или системы CasY согласно настоящему изобретению в целевую клетку. КПП, как правило, имеют аминокислотный состав, который содержит относительно большое количество положительно заряженных аминокислот, таких как лизин или аргинин, или имеет последовательности, которые содержат чередующийся паттерн из полярных/заряженных аминокислот и неполярных гидрофобных аминокислот.

Имплантируемое устройство можно использовать для доставки полипептида CasY согласно настоящему изобретению, слитого полипептида CasY согласно настоящему изобретению, РНП согласно настоящему изобретению, нуклеиновой кислоты согласно настоящему изобретению (например, CasY-направляющей РНК, нуклеиновой кислоты, кодирующей CasY-направляющую РНК, нуклеиновой кислоты, кодирующей полипептид CasY, донорной матрицы и т.п.) или системы CasY согласно настоящему изобретению в целевую клетку (например, целевую клетку *in vivo*, когда целевая клетка представляет собой целевую клетку в циркуляции, целевую клетку в ткани, целевую клетку в органе и т.д.). Имплантируемое устройство, подходящее для применения для доставки полипептида CasY согласно настоящему изобретению, слитого полипептида CasY согласно настоящему изобретению, РНП согласно настоящему изобретению, нуклеиновой кислоты согласно настоящему изобретению или системы CasY согласно настоящему изобретению в целевую клетку (например, целевую клетку *in vivo*, когда целевая клетка представляет собой целевую клетку в циркуляции, целевую клетку в ткани, целевую клетку в органе и т.д.), может содержать контейнер (например, резервуар, матрицу и т.д.), который содержит полипептид CasY, слитый полипептид CasY, РНП или систему CasY (или ее компонент, например, нуклеиновую кислоту согласно настоящему изобретению).

Подходящее имплантируемое устройство может содержать полимерный субстрат, такой как матрица, например, который используется как тело устройства, и в некоторых случаях дополнительные формирующие материалы, такие как металлы или дополнительные полимеры, и материалы для повышения видимости и визуализации. Имплантируемое устройство для доставки может обеспечивать преимущество локального высвобождения в течение длительного периода, когда высвобождение подлежащих доставке полипептида и/или нуклеиновой кислоты происходит непосредственно в целевом участке, например, внеклеточном матриксе (ВКМ), сосудистой системе, окружающей опухоль, болезненной ткани и т.д. Подходящие имплантируемые устройства для доставки включают в себя устройства, пригодные для применения при полостной доставке, например, в брюшную полость, и/или любого другого типа введения, при котором система доставки лекарственного препарата не прикреплена или не присоединена, содержащие биостабильный, и/или разлагаемый, и/или биоабсорбируемый полимерный субстрат, который необязательно может быть, например, матрицей. В некоторых случаях подходящее имплантируемое устройство для доставки лекарственного препарата содержит разлагаемые полимеры, а основным механизмом высвобождения является объемная эрозия. В некоторых случаях подходящее имплантируемое устройство для доставки лекарственного препарата содержит неразлагаемые или медленно разлагаемые полимеры, а основным механизмом высвобождения является диффузия, а не объемная эрозия, при этом внешняя часть функционирует как мембрана, а внутренняя часть функционирует как резервуар для лекарственного препарата, на который практически не влияет окружение в течение длительного периода (например, от около недели до около нескольких месяцев). Также необязательно можно использовать комбинации разных полимеров с разными механизмами высвобождения. Градиент концентрации можно поддерживать эффективно постоянным в течение значительного периода общего периода высвобождения, и, следовательно, скорость диффузии является эффективно постоянной (что называется "нулевым режимом" диффузии). Под термином "постоянный" подразумевается скорость диффузии, поддерживаемая выше нижнего предела терапевтической эффективности, но которая может необязательно демонстрировать начальное резкое повышение и/или может флуктуировать, например, повышаясь или снижаясь

до определенной степени. Скорость диффузии можно поддерживать таким образом в течение длительного периода и можно считать постоянной до определенного уровня для оптимизации терапевтически эффективного периода, например, эффективного периода подавления.

В некоторых случаях имплантируемую систему для доставки конструируют для защиты нуклеотидного терапевтического агента от деградации, химической по природе или вследствие действия ферментов и других факторов в организме субъекта.

Участок для имплантации устройства или целевой участок может быть выбран для обеспечения максимальной терапевтической эффективности. Например, устройство для доставки можно имплантировать в или вблизи опухолевого окружения или кровеносных сосудов, связанных с опухолью. Целевой локацией может быть, например: 1) головной мозг в участках дегенерации, как при болезни Паркинсона или Альцгеймера в базальных ганглиях, белом и сером веществе; 2) позвоночник, как в случае амиотрофического бокового склероза (АБС); 3) шейка матки; 4) суставы с активным и хроническим воспалением; 5) дерма в случае псориаза; 6) симпатические и сенсорные нервные участки для анальгезирующего эффекта; 7) кости; 8) участок острой или хронической инфекции; 9) внутривагинальная; 10) внутренне ухо - слуховая система, лабиринт внутреннего уха, вестибулярная система; 11) эндотрахеальная; 12) внутрисердечная; коронарная, эпикардальная; 13) мочевыводящие пути или мочевого пузыря; 14) желчевыводительная система; 15) паренхиматозная ткань, включая и не ограничиваясь этим почки, печень, селезенку; 16) лимфатические узлы; 17) слюнные железы; 18) десны; 19) интраартикулярная (внутри суставов); 20) внутриглазная; 21) ткань головного мозга; 22) желудочки головного мозга; 23) полости, включая брюшную полость (например, но без ограничения, в случае рака яичников); 24) внутри пищевода; и 25) внутри прямой кишки; и 26) внутри сосудистой системы.

Способ введения, такой как имплантация, может необязательно уже использоваться для других типов тканевой имплантации, и/или тканевого введения, и/или для получения тканевых образцов необязательно без модификаций, или в альтернативном варианте необязательно только с небольшими модификациями в таких способах. Такие способы необязательно включают в себя, но не ограничиваются этим, метод брахитерапии, биопсию, эндоскопию с или без применения ультразвука, например, стереотактические методы, применяемые в случае тканей головного мозга, лапароскопию, включая имплантацию с помощью лапароскопа в суставы, органы брюшной полости, стенку мочевого пузыря и полости тела.

Модифицированные клетки-хозяева.

В настоящем изобретении предложена модифицированная клетка, содержащая полипептид CasY согласно настоящему изобретению и/или нуклеиновую кислоту, содержащую нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобретению. В настоящем изобретении предложена модифицированная клетка, содержащая полипептид CasY согласно настоящему изобретению, при этом модифицированная клетка является клеткой, которая обычно не содержит полипептид CasY согласно настоящему изобретению. В настоящем изобретении предложена модифицированная клетка (например, генетически модифицированная клетка), содержащая нуклеиновую кислоту, содержащую нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобретению. В настоящем изобретении предложена генетически модифицированная клетка, которая была генетически модифицирована мРНК, содержащей нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобретению. В настоящем изобретении предложена генетически модифицированная клетка, которая была генетически модифицирована рекомбинантным экспрессионным вектором, содержащим нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобретению. В настоящем изобретении предложена генетически модифицированная клетка, которая была генетически модифицирована рекомбинантным экспрессионным вектором, содержащим: а) нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобретению; и б) нуклеотидную последовательность, кодирующую CasY-направляющую РНК согласно настоящему изобретению. В настоящем изобретении предложена генетически модифицированная клетка, которая была генетически модифицирована рекомбинантным экспрессионным вектором, содержащим: а) нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобретению; б) нуклеотидную последовательность, кодирующую CasY-направляющую РНК согласно настоящему изобретению; и с) нуклеотидную последовательность, кодирующую донорную матрицу.

Клетка, которая служит в качестве реципиента для полипептида CasY согласно настоящему изобретению и/или нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобретению и/или CasY-направляющую РНК согласно настоящему изобретению, может быть любой из множества клеток, включая, например, *in vitro* клетки; *in vivo* клетки; *ex vivo* клетки; первичные клетки; раковые клетки; клетки животных; клетки растений; клетки водорослей; клетки грибов и т.д. Клетка, которая служит в качестве реципиента для полипептида CasY согласно настоящему изобретению и/или нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобретению и/или CasY-направляющую РНК согласно настоящему изобретению, называется "клеткой-хозяином" или "целевой клеткой". Клетка-хозяин или целевая клетка может быть реципиентом системы CasY согласно настоящему изобретению. Клетка-хозяин или целевая клетка может быть реципиентом РНП CasY согласно настоящему изобретению.

Клетка-хозяин или целевая клетка может быть реципиентом одного компонента системы CasY согласно настоящему изобретению.

Неограничивающие примеры клеток (целевых клеток) включают в себя: прокариотическую клетку, эукариотическую клетку, бактериальную клетку, клетку архей, клетку одноклеточного эукариотического организма, клетку простейших, клетку из растения (например, клетки из растительных культур, фруктов, овощей, зерновых, сои, кукурузы, маиса, пшеницы, семян, томатов, риса, маниоки, сахарного тростника, тыквы, сена, картофеля, хлопка, конопли, табака, цветковых растений, хвойных, голосеменных, покрытосеменных, папоротников, плаунов, антоцеровых, печеночников, мхов, двудольных, однодольных и т.д.), клетку водорослей (например, *Botryococcus braunii*, *Chlamydomonas reinhardtii*, *Nannochloropsis gaditana*, *Chlorella pyrenoidosa*, *Sargassum patens*, *C. agardh* и т.п.), клетку морских водорослей (например, бурых водорослей), клетку грибов (например, дрожжевую клетку, клетку из грибов), клетку животного, клетку беспозвоночного животного (например, плодовой мушки, стрекающих, иглокожих, нематоды и т.д.), клетку позвоночного животного (например, рыбы, амфибии, рептилии, птицы, млекопитающего), клетку млекопитающего (например, копытных (например, свиньи, коровы, козы, овцы); грызуна (например, крысы, мыши); отличного от человека примата; человека; кошачьих (например, кошки); собачьих (например, собаки) и т.д.) и т.п. В некоторых случаях клетка является клеткой, которая не происходит из природного организма (например, клетка может быть синтетически созданной клеткой; также называемой искусственной клеткой).

Клетка может быть *in vitro* клеткой (например, стабильной культивируемой линией клеток). Клетка может быть *ex vivo* клеткой (культивируемой клеткой из отдельного организма). Клетка может быть *in vivo* клеткой (например, клеткой в отдельном организме). Клетка может быть выделенной клеткой. Клетка может быть клеткой внутри организма. Клетка может быть организмом. Клетка может быть клеткой в культуре клеток (например, *in vitro* культуре клеток). Клетка может быть клеткой из коллекции клеток. Клетка может быть прокариотической клеткой или быть полученной из прокариотической клетки. Клетка может быть бактериальной клеткой или быть полученной из бактериальной клетки. Клетка может быть клеткой архей или быть полученной из клетки архей. Клетка может быть эукариотической клеткой или быть полученной из эукариотической клетки. Клетка может быть растительной клеткой или быть полученной из растительной клетки. Клетка может быть клеткой животного или быть полученной из клетки животного. Клетка может быть клеткой беспозвоночного или быть полученной из клетки беспозвоночного. Клетка может быть клеткой позвоночного или быть полученной из клетки позвоночного. Клетка может быть клеткой млекопитающего или быть полученной из клетки млекопитающего. Клетка может быть клеткой грызуна или быть полученной из клетки грызуна. Клетка может быть клеткой человека или быть полученной из клетки человека. Клетка может быть микробной клеткой или быть полученной из микробной клетки. Клетка может быть клеткой грибов или быть полученной из клетки грибов. Клетка может быть клеткой насекомого. Клетка может быть клеткой членистоногого. Клетка может быть клеткой простейшего. Клетка может быть клеткой гельминта.

Подходящие клетки включают в себя стволовые клетки (например, эмбриональные стволовые (ЭС) клетки, индуцированные плюрипотентные стволовые (иПС) клетки; половые клетки (например, ооциты, сперматозоиды, оогонии, сперматогонии и т.д.); соматические клетки, например, фибробласты, олигодендроциты, глиальные клетки, гемопоэтические клетки, нейроны, мышечные клетки, костные клетки, гепатоциты, панкреатические клетки и т.д.

Подходящие клетки включают в себя человеческие эмбриональные стволовые клетки, фетальные кардиомиоциты, миофибробласты, мезенхимальные стволовые клетки, аутоотрансплантированные размноженные кардиомиоциты, адипоциты, тотипотентные клетки, плюрипотентные клетки, кровяные стволовые клетки, миобласты, взрослые стволовые клетки, клетки костного мозга, мезенхимальные клетки, эмбриональные стволовые клетки, паренхимальные клетки, эпителиальные клетки, эндотелиальные клетки, мезотелиальные клетки, фибробласты, остеобласты, хондроциты, экзогенные клетки, эндогенные клетки, стволовые клетки, гемопоэтические стволовые клетки, полученные из костного мозга клетки-предшественники, миокардиальные клетки, скелетные клетки, фетальные клетки, недифференцированные клетки, мультипотентные клетки-предшественники, унипотентные клетки-предшественники, моноциты, сердечные миобласты, скелетные миобласты, макрофаги, капиллярные эндотелиальные клетки, ксеногенные клетки, аллогенные клетки и постнатальные стволовые клетки.

В некоторых случаях клетка представляет собой иммунную клетку, нейрон, эпителиальную клетку, эндотелиальную клетку или стволовую клетку. В некоторых случаях иммунная клетка представляет собой Т-клетку, В-клетку, моноцит, естественную клетку-киллера, дендритную клетку или макрофаг. В некоторых случаях иммунная клетка представляет собой цитотоксическую Т-клетку. В некоторых случаях иммунная клетка представляет собой хелперную Т-клетку. В некоторых случаях иммунная клетка представляет собой регуляторную Т-клетку (Treg).

В некоторых случаях клетка представляет собой стволовую клетку. Стволовые клетки включают в себя взрослые стволовые клетки. Взрослые стволовые клетки также называются соматическими стволовыми клетками.

Взрослые стволовые клетки присутствуют в дифференцированной ткани, но сохраняют свойства самообновления и способность давать начало множеству типов клеток, обычно типов клеток, типичных

для ткани, в которой находятся стволовые клетки. Специалистам в данной области техники известны многочисленные примеры соматических стволовых клеток, включая мышечные стволовые клетки; гемопоэтические стволовые клетки; эпителиальные стволовые клетки; нейральные стволовые клетки; мезенхимальные стволовые клетки; стволовые клетки молочной железы; кишечные стволовые клетки; мезодермальные стволовые клетки; эндотелиальные стволовые клетки; обонятельные стволовые клетки; стволовые клетки нервного гребня и т.п.

Представляющие интерес стволовые клетки включают в себя стволовые клетки млекопитающих, при этом термин "млекопитающее" относится к любому животному, классифицированному как млекопитающее, включая людей; отличных от человека приматов; домашних и сельскохозяйственных животных; а также содержащихся в зоопарках, лабораторных, спортивных или домашних животных, таких как собаки, лошади, кошки, коровы, мыши, крысы, кролики и т.д. В некоторых случаях стволовая клетка является человеческой стволовой клеткой. В некоторых случаях стволовая клетка является стволовой клеткой грызуна (например, мыши; крысы). В некоторых случаях стволовая клетка является стволовой клеткой отличного от человека примата.

Стволовые клетки могут экспрессировать один или более маркеров стволовых клеток, например, SOX9, KRT19, KRT7, LGR5, CA9, FXYD2, CDH6, CLDN18, TSPAN8, BPIFB1, OLFM4, CDH17 и PPARGC1A.

В некоторых вариантах реализации стволовая клетка является гемопоэтической стволовой клеткой (ГСК). ГСК являются мезодермальными клетками, которые можно выделять из костного мозга, крови, пуповинной крови, фетальной печени и желточного мешка. ГСК делятся на  $CD34^+$  и  $CD3^-$ . ГСК могут восстанавливать популяции линий дифференцировки эритроидных клеток, нейтрофилов-макрофагов, мегакариоцитов и лимфоидных гемопоэтических клеток *in vivo*. ГСК можно индуцировать *in vitro* для прохождения по меньшей мере нескольких ведущих к самообновлению клеточных делений и можно индуцировать для дифференцировки в те же линии, которые наблюдаются *in vivo*. Следовательно, ГСК можно индуцировать для дифференцировки в одну или более из эритроидных клеток, мегакариоцитов, нейтрофилов, макрофагов и лимфоидных клеток.

В других вариантах реализации стволовая клетка является нейральной стволовой клеткой (НСК). Нейральные стволовые клетки (НСК) способны дифференцироваться в нейроны и глию (включая олигодендроциты и астроциты). Нейральная стволовая клетка является мультипотентной стволовой клеткой, которая способна проходить множество делений и в определенных условиях может давать дочерние клетки, которые представляют собой нейральные стволовые клетки или нейральные клетки-предшественники, которые могут быть нейробластами или глиобластами, например, клетки, которые должны стать одним или более типами нейронов и глиальных клеток, соответственно. Способы получения НСК известны в данной области техники.

В других вариантах реализации стволовая клетка является мезенхимальной стволовой клеткой (МСК). МСК, изначально полученные из эмбриональной мезодермы и выделенные из взрослого костного мозга, могут дифференцироваться с образованием мышц, костей, хрящей, жира, костномозговой стромы и сухожилий. Способы выделения МСК известны в данной области техники; а для получения МСК можно использовать любой известный способ. Смотрите, например, патент США № 5736396, в котором описано выделение человеческих МСК.

В некоторых случаях клетка является растительной клеткой. Растительная клетка может быть клеткой однодольного растения. Растительная клетка может быть клеткой двудольного растения.

В некоторых случаях клетка является растительной клеткой. Например, клетка может быть клеткой одного из основных сельскохозяйственных растений, например, ячменя, бобов (сухих съедобных), канолы, кукурузы, хлопка (Пима), хлопка ("Апланд"), льняного семени, сена (люцерны), сена (отличного от люцерны), овса, арахиса, риса, сорго, сои, сахарной свеклы, сахарного тростника, подсолнечника (масла), подсолнечника (не масла), сладкого картофеля, табака (Берли), табака (дымовой сушки), томатов, пшеницы (дурум), пшеницы (весенней), пшеницы (озимой) и т.п. В качестве другого примера клетка представляет собой клетку овощных культурных растений, которые включают в себя, но не ограничиваются этим, ростки люцерны, листья алоэ, корень триостренника, стрелолист, артишоки, спаржу, ростки бамбука, цветы банана, ростки бобов, бобы, свекольную ботву, свеклу, горький огурец, китайскую капусту, брокколи, ботву брокколи (рапини), брюссельскую капусту, кочанную капусту, ростки капусты, листья кактуса (нопалес), бутылочную тыкву, кардон, морковь, цветную капусту, сельдерей, чайот, китайский артишок (хорогу), китайскую капусту, китайский сельдерей, лук пахучий, бок-чой, листья хризантем (тунг хо), листовую капусту, стебли кукурузы, сладкую кукурузу, огурцы, дайкон, молодые листья одуванчика, колоказию, доу миао (верхушки гороха), донква (зимнюю восковую тыкву), баклажан, цикорий салатный, эскариоль, рахисы папоротника, полевой кресс-салат, фризе, гай чой (китайскую горчицу), гай ян, галангал (сиамский, тайский имбирь), чеснок, корень имбиря, лопушник, зелень, ганновверскую салатную зелень, киноа, иерусалимские артишоки, хикама, кормовую капусту, кольраби, марь белую (триллиум), латук (биб), латук (бостонский), латук (бостонский красный), латук (зеленолистный), латук (айсберг), латук (лollo rossa), латук (дуболистный - зеленый), латук (дуболистный - красный), латук (обработанный), латук (краснолистный), латук (ромен), латук (красный ромен), латук (крас-

ную горчицу), линкок, ло бок, спаржевые бобы, корень лотоса, полевой салат, листья агавы (мэги), ксантозому, смесь месклан, курчаволистную горчицу, люффу (тыкву мочальную), му, мокву (волосатую тыкву), грибы, горчицу, нагаимо, бамию, водяной шпинат, зеленый лук, опо (длинную тыкву), декоративную кукурузу, горянку обыкновенную, петрушку, пастернак, горох, перец (конусный), перец, тыквы, радикио, ростки редьки, редьку, зелень рапса, зелень рапса, ревень, ромен (бэби рэд), брюкву, солерос (энтаду ползучую), синкву (гранистую/ребристую люффу), шпинат, тыкву, соломенные блоки, сахарный тростник, сладкий картофель, мангольд, тамаринд, таро, листья таро, побеги таро, китайскую плоскую капусту, леуцену, тиндору, физалис, томаты, томаты (черри), томаты (виноградные), томаты (сливки), куркуму, ботву молодой репы, репу, водяной орех, диоскорею, ямс (разные виды), юй чой, юку (кассаву) и т.п.

В некоторых случаях клетка может быть клеткой членистоногого. Например, клетка может быть клеткой подотряда, семейства, подсемейства, группы, подгруппы или вида, например,

*Chelicerata, Myriapodia, Hexipodia, Arachnida, Insecta, Archaeognatha, Thysanura, Palaeoptera, Ephemeroptera, Odonata, Anisoptera, Zygoptera, Neoptera, Exopterygota, Plecoptera, Embioptera, Orthoptera, Zoraptera, Dermaptera, Dictyoptera, Notoptera, Grylloblattidae, Mantophasmatidae, Phasmatodea, Blattaria, Isoptera, Mantodea, Parapneuroptera, Psocoptera, Thysanoptera, Phthiraptera, Hemiptera, Endopterygota или Holometabola, Hymenoptera, Coleoptera, Strepsiptera, Raphidioptera, Megaloptera, Neuroptera, Mecoptera, Siphonaptera, Diptera, Trichoptera или Lepidoptera.*

В некоторых случаях клетка может быть клеткой насекомого. Например, в некоторых случаях клетка является клеткой москита, кузнечика, клопа, мухи, блохи, пчелы, осы, муравья, вши, мотылька или жука.

Наборы.

В настоящем изобретении предложен набор, содержащий систему CasY согласно настоящему изобретению или компонент системы CasY согласно настоящему изобретению.

Набор согласно настоящему изобретению может содержать: а) полипептид CasY согласно настоящему изобретению и CasY-направляющую РНК; б) полипептид CasY согласно настоящему изобретению, CasY-направляющую РНК и донорную матричную нуклеиновую кислоту; в) слитый полипептид CasY согласно настоящему изобретению и CasY-направляющую РНК; г) слитый полипептид CasY согласно настоящему изобретению, CasY-направляющую РНК и донорную матричную нуклеиновую кислоту; д) мРНК, кодирующую полипептид CasY согласно настоящему изобретению, и CasY-направляющую РНК; е) мРНК, кодирующую полипептид CasY согласно настоящему изобретению, CasY-направляющую РНК и донорную матричную нуклеиновую кислоту; ж) мРНК, кодирующую слитый полипептид CasY согласно настоящему изобретению, и CasY-направляющую РНК; з) мРНК, кодирующую слитый полипептид CasY согласно настоящему изобретению, CasY-направляющую РНК и донорную матричную нуклеиновую кислоту; и) рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобретению, и нуклеотидную последовательность, кодирующую CasY-направляющую РНК; л) рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобретению, CasY-направляющую РНК и донорную матричную нуклеиновую кислоту; м) рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую слитый полипептид CasY согласно настоящему изобретению, нуклеотидную последовательность, кодирующую CasY-направляющую РНК, и нуклеотидную последовательность, кодирующую донорную матричную нуклеиновую кислоту; н) первый рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобретению, и второй рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую CasY-направляющую РНК; о) первый рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую слитый полипептид CasY согласно настоящему изобретению, и второй рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую CasY-направляющую РНК, и донорную матричную нуклеиновую кислоту; п) первый рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую слитый полипептид CasY согласно настоящему изобретению, и второй рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую CasY-направляющую РНК, и донорную матричную нуклеиновую кислоту; р) первый рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую слитый полипептид CasY согласно настоящему изобретению, и

второй рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую CasY-направляющую РНК, и донорную матричную нуклеиновую кислоту; q) рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобретению, нуклеотидную последовательность, кодирующую первую CasY-направляющую РНК, и нуклеотидную последовательность, кодирующую вторую CasY-направляющую РНК; или r) рекомбинантный экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую слитый полипептид CasY согласно настоящему изобретению, нуклеотидную последовательность, кодирующую первую CasY-направляющую РНК, и нуклеотидную последовательность, кодирующую вторую CasY-направляющую РНК; или некоторую вариацию (a)-(r).

Набор согласно настоящему изобретению может содержать: а) вышеописанный компонент системы CasY согласно настоящему изобретению или может содержать систему CasY согласно настоящему изобретению и б) один или более дополнительных реагентов, например, i) буфер; ii) ингибитор протеаз; iii) ингибитор нуклеаз; iv) реагент, необходимый для проявления или визуализации выявляемой метки; v) положительную и/или отрицательную контрольную целевую ДНК; vi) положительную и/или отрицательную контрольную CasY-направляющую РНК и т.п. Набор согласно настоящему изобретению может содержать: а) вышеописанный компонент системы CasY согласно настоящему изобретению или может содержать систему CasY согласно настоящему изобретению и б) терапевтический агент.

Набор согласно настоящему изобретению может содержать рекомбинантный экспрессионный вектор, содержащий: а) сайт вставки для вставки нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую часть CasY-направляющей РНК, которая гибридизируется с целевой нуклеотидной последовательностью в целевой нуклеиновой кислоте; и б) нуклеотидную последовательность, кодирующую CasY-связывающую часть CasY-направляющей РНК. Набор согласно настоящему изобретению может содержать рекомбинантный экспрессионный вектор, содержащий: а) сайт вставки для вставки нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую часть CasY-направляющей РНК, которая гибридизируется с целевой нуклеотидной последовательностью в целевой нуклеиновой кислоте; б) нуклеотидную последовательность, кодирующую CasY-связывающую часть CasY-направляющей РНК; и с) нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобретению.

#### Применение.

Полипептид CasY согласно настоящему изобретению или слитый полипептид CasY согласно настоящему изобретению можно применять в ряде способов (например, в комбинации с CasY-направляющей РНК и в некоторых случаях дополнительно в комбинации с донорной матрицей). Например, полипептид CasY согласно настоящему изобретению можно применять для (i) модификации (например, расщепления, например, создания одноцепочечного разрыва; метилирования; и т.д.) целевой нуклеиновой кислоты (ДНК или РНК; одноцепочечной или двухцепочечной); (ii) модуляции транскрипции целевой нуклеиновой кислоты; (iii) мечения целевой нуклеиновой кислоты; (iv) связывания целевой нуклеиновой кислоты (например, в целях выделения, мечения, визуализации, отслеживания и т.д.); (v) модификации полипептида (например, гистона), связанного с целевой нуклеиновой кислотой; и т.п. Таким образом, в настоящем изобретении предложен способ модификации целевой нуклеиновой кислоты. В некоторых случаях способ согласно настоящему изобретению для модификации целевой нуклеиновой кислоты включает в себя приведение целевой нуклеиновой кислоты в контакт с: а) полипептидом CasY согласно настоящему изобретению и б) одной или более (например, двумя) CasY-направляющими РНК. В некоторых случаях способ согласно настоящему изобретению для модификации целевой нуклеиновой кислоты включает в себя приведение целевой нуклеиновой кислоты в контакт с: а) полипептидом CasY согласно настоящему изобретению; б) CasY-направляющей РНК и с) донорной нуклеиновой кислотой (например, донорной матрицей). В некоторых случаях этап приведения в контакт проводят в клетке *in vitro*. В некоторых случаях этап приведения в контакт проводят в клетке *in vivo*. В некоторых случаях этап приведения в контакт проводят в клетке *ex vivo*.

Так как способ, в котором используется полипептид CasY, включает в себя связывание полипептида CasY с конкретной областью в целевой нуклеиновой кислоте (посредством нацеливания на нее с помощью связанной CasY-направляющей РНК), эти способы в общем случае называются в данном документе способами связывания (например, способ связывания целевой нуклеиновой кислоты). Однако следует понимать, что в некоторых случаях, хотя способ связывания может приводить не более чем к связыванию целевой нуклеиновой кислоты, в других случаях способ может иметь разные конечные результаты (например, способ может приводить к модификации целевой нуклеиновой кислоты, например, расщеплению/метилованию и т.д., модуляции транскрипции из целевой нуклеиновой кислоты; модуляции трансляции целевой нуклеиновой кислоты; редактированию генома; модуляции белка, связанного с целевой нуклеиновой кислотой; выделению целевой нуклеиновой кислоты и т.д.).

Примеры подходящих способов смотрите, например, в Jinek et al., *Science*. 2012 Aug 17; 337(6096):816-21; Chylinski et al., *RNA Biol*. 2013 May; 10(5):726-37; Ma et al., *Biomed Res Int*. 2013; 2013:270805; Hou et al., *Proc Natl Acad Sci USA*. 2013 Sep 24; 110(39):15644-9; Jinek et al., *Elife*. 2013; 2:e00471; Pattanayak et al., *Nat Biotechnol*. 2013 Sep; 31(9):839-43; Qi et al., *Cell*. 2013 Feb 28; 152(5):1173-83; Wang et al., *Cell*. 2013 May 9; 153(4):910-8; Auer et al., *Genome Res*. 2013 Oct 31; Chen et al., *Nucleic*

Acids Res. 2013 Nov 1; 41(20):e19; Cheng et al., Cell Res. 2013 Oct; 23(10):1163-71; Cho et al., Genetics. 2013 Nov; 195(3):1177-80; DiCarlo et al., Nucleic Acids Res. 2013 Apr; 41(7):4336-43; Dickinson et al., Nat Methods. 2013 Oct; 10(10):1028-34; Ebina et al., Sci Rep. 2013; 3:2510; Fujii et al., Nucleic Acids Res. 2013 Nov 1; 41(20):e187; Hu et al., Cell Res. 2013 Nov; 23(11):1322-5; Jiang et al., Nucleic Acids Res. 2013 Nov 1; 41(20):e188; Larson et al., Nat Protoc. 2013 Nov; 8(11):2180-96; Mali et al., Nat Methods. 2013 Oct; 10(10):957-63; Nakayama et al., Genesis. 2013 Dec; 51(12):835-43; Ran et al., Nat Protoc. 2013 Nov; 8(11):2281-308; Ran et al., Cell. 2013 Sep 12; 154(6):1380-9; Upadhyay et al., G3 (Bethesda). 2013 Dec 9; 3(12):2233-8; Walsh et al., Proc Natl Acad Sci USA. 2013 Sep 24; 110(39): 15514-5; Xie et al., Mol Plant. 2013 Oct 9; Yang et al., Cell. 2013 Sep 12; 154(6):1370-9; а также патенты и патентные заявки США:

8906616; 8895308; 8889418; 8889356; 8871445; 8865406; 8795965; 8771945;  
8697359; 20140068797; 20140170753; 20140179006; 20140179770; 20140186843;  
20140186919; 20140186958; 20140189896; 20140227787; 20140234972; 20140242664;  
20140242699; 20140242700; 20140242702; 20140248702; 20140256046; 20140273037;  
20140273226; 20140273230; 20140273231; 20140273232; 20140273233; 20140273234;  
20140273235; 20140287938; 20140295556; 20140295557; 20140298547; 20140304853;  
20140309487; 20140310828; 20140310830; 20140315985; 20140335063; 20140335620;  
20140342456; 20140342457; 20140342458; 20140349400; 20140349405; 20140356867;  
20140356956; 20140356958; 20140356959; 20140357523; 20140357530; 20140364333 и  
20140377868;

которые все в полном объеме включены в данный документ посредством ссылки.

Например, в настоящем изобретении предложены (но оно не ограничивается ими) способы расщепления целевой нуклеиновой кислоты; способы редактирования целевой нуклеиновой кислоты; способы модуляции транскрипции из целевой нуклеиновой кислоты; способы выделения целевой нуклеиновой кислоты, способы связывания целевой нуклеиновой кислоты, способы визуализации целевой нуклеиновой кислоты, способы модификации целевой нуклеиновой кислоты и т.п.

В контексте данного документа термины/выражения "приводить в контакт целевую нуклеиновую кислоту" и "приведение в контакт целевой нуклеиновой кислоты", например, с полипептидом CasY или со слитым полипептидом CasY и т.д., включают в себя все способы приведения в контакт целевой нуклеиновой кислоты. Например, полипептид CasY можно вводить в клетку в виде белка, РНК (кодирующей полипептид CasY) или ДНК (кодирующей полипептид CasY); тогда как CasY-направляющую РНК можно вводить в виде направляющей РНК или в виде нуклеиновой кислоты, кодирующей направляющую РНК. Следовательно, например, при осуществлении способа в клетке (например, внутри клетки *in vitro*, внутри клетки *in vivo*, внутри клетки *ex vivo*) способ, который включает в себя приведение в контакт целевой нуклеиновой кислоты, охватывает внесение в клетку любых или всех компонентов в их активном/конечном состоянии (например, в форме белка(ов) для полипептида CasY; в форме белка для слитого полипептида CasY; в форме РНК в некоторых случаях для направляющей РНК) и также охватывает внесение в клетку одной или более нуклеиновых кислот, кодирующих один или более компонентов (например, нуклеиновых кислот, содержащих нуклеотидные последовательности, кодирующие полипептид CasY или слитый полипептид CasY, нуклеиновых кислот, содержащих нуклеотидные последовательности, кодирующие направляющие РНК, нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую донорную матрицу, и т.п.). Так как способы также можно осуществлять *in vitro* вне клетки, способ, который включает в себя приведение в контакт целевой нуклеиновой кислоты (если не указано иное), охватывает приведение в контакт вне клетки *in vitro*, внутри клетки *in vitro*, внутри клетки *in vivo*, внутри клетки *ex vivo* и т.д.

В некоторых случаях способ согласно настоящему изобретению для модификации целевой нуклеиновой кислоты включает в себя внесение в целевую клетку локуса CasY, например, нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую полипептид CasY, а также нуклеотидные последовательности длиной от около 1 тысячи оснований (т.о.) до 5 т.о., окружающие CasY-кодирующую нуклеотидную последовательность, из клетки (например, в некоторых случаях клетки, которая в своем естественном состоянии (состоянии, в котором она находится в природе) содержит локус CasY), содержащей локус CasY, при этом целевая клетка обычно (в своем естественном состоянии) не содержит локус CasY. Однако одну или более спейсерных последовательностей, кодирующих направляющие последовательности для кодируемых cr-РНК, можно модифицировать так, чтобы осуществлять нацеливание на одну или более целевых последовательностей. Таким образом, например, в некоторых случаях способ согласно настоящему изобретению для модификации целевой нуклеиновой кислоты

включает в себя внесение в целевую клетку локуса CasY, например, нуклеиновой кислоты, полученной из клеточного источника (например, в некоторых случаях клетки, которая в своем естественном состоянии (состоянии, в котором она находится в природе) содержит локус CasY), при этом нуклеиновая кислота имеет длину от 100 нуклеотидов (нт) до 5 т.о. (например, длину от 100 нт до 500 нт, от 500 нт до 1 т.о., от 1 т.о. до 1,5 т.о., от 1,5 т.о. до 2 т.о., от 2 т.о. до 2,5 т.о., от 2,5 т.о. до 3 т.о., от 3 т.о. до 3,5 т.о., от 3,5 т.о. до 4 т.о. или от 4 т.о. до 5 т.о.) и содержит нуклеотидную последовательность, кодирующую полипептид CasY. Как отмечено выше, в некоторых таких случаях одну или более спейсерных последовательностей, кодирующих направляющие последовательности для кодируемых cr-РНК, можно модифицировать так, чтобы осуществлять нацеливание на одну или более целевых последовательностей. В некоторых случаях способ включает в себя внесение в целевую клетку: i) локуса CasY и ii) донорной ДНК-матрицы. В некоторых случаях целевая нуклеиновая кислота находится в бесклеточной композиции *in vitro*. В некоторых случаях целевая нуклеиновая кислота находится в целевой клетке. В некоторых случаях целевая нуклеиновая кислота находится в целевой клетке, при этом целевая клетка является прокариотической клеткой. В некоторых случаях целевая нуклеиновая кислота находится в целевой клетке, при этом целевая клетка является эукариотической клеткой. В некоторых случаях целевая нуклеиновая кислота находится в целевой клетке, при этом целевая клетка является клеткой млекопитающего. В некоторых случаях целевая нуклеиновая кислота находится в целевой клетке, при этом целевая клетка является клеткой растения.

В некоторых случаях способ согласно настоящему изобретению для модификации целевой нуклеиновой кислоты включает в себя приведение целевой нуклеиновой кислоты в контакт с полипептидом CasY согласно настоящему изобретению или со слитым полипептидом CasY согласно настоящему изобретению. В некоторых случаях способ согласно настоящему изобретению для модификации целевой нуклеиновой кислоты включает в себя приведение целевой нуклеиновой кислоты в контакт с полипептидом CasY и CasY-направляющей РНК. В некоторых случаях способ согласно настоящему изобретению для модификации целевой нуклеиновой кислоты включает в себя приведение целевой нуклеиновой кислоты в контакт с полипептидом CasY, первой CasY-направляющей РНК и второй CasY-направляющей РНК. В некоторых случаях способ согласно настоящему изобретению для модификации целевой нуклеиновой кислоты включает в себя приведение целевой нуклеиновой кислоты в контакт с полипептидом CasY согласно настоящему изобретению и CasY-направляющей РНК и донорной ДНК-матрицей.

Представляющие интерес целевые нуклеиновые кислоты и целевые клетки.

Полипептид CasY согласно настоящему изобретению или слитый полипептид CasY согласно настоящему изобретению, связанный с CasY-направляющей РНК, может связываться с целевой нуклеиновой кислотой и в некоторых случаях может связывать и модифицировать целевую нуклеиновую кислоту. Целевая нуклеиновая кислота может представлять собой любую нуклеиновую кислоту (например, ДНК, РНК), может быть двухцепочечной или одноцепочечной, может представлять собой любой тип нуклеиновой кислоты (например, хромосому (геномную ДНК), полученную из хромосомы, хромосомную ДНК, плазмидную, вирусную, внеклеточную, внутриклеточную, митохондриальную, хлоропластную, линейную, кольцевую и т.д.) и может происходить из любого организма (например, при условии, что CasY-направляющая РНК содержит нуклеотидную последовательность, которая гибридизируется с целевой последовательностью в целевой нуклеиновой кислоте, так, чтобы можно было осуществлять нацеливание на целевую нуклеиновую кислоту).

Целевая нуклеиновая кислота может представлять собой ДНК или РНК. Целевая нуклеиновая кислота может быть двухцепочечной (например, дцДНК, дцРНК) или одноцепочечной (например, оцРНК, оцДНК). В некоторых случаях целевая нуклеиновая кислота является одноцепочечной. В некоторых случаях целевая нуклеиновая кислота является одноцепочечной РНК (оцРНК). В некоторых случаях целевая оцРНК (например, целевая клеточная оцРНК, вирусная оцРНК и т.д.) выбрана из: мРНК, рРНК, тРНК, некодирующей РНК (нкРНК), длинной некодирующей РНК (днкРНК) и микроРНК (миРНК). В некоторых случаях целевая нуклеиновая кислота является одноцепочечной ДНК (оцДНК) (например, вирусной ДНК). Как отмечено выше, в некоторых случаях целевая нуклеиновая кислота является одноцепочечной.

Целевая нуклеиновая кислота может находиться где угодно, например, вне клетки *in vitro*, внутри клетки *in vitro*, внутри клетки *in vivo*, внутри клетки *ex vivo*. Подходящие целевые клетки (которые содержат целевые нуклеиновые кислоты, такие как геномная ДНК) включают в себя, но не ограничиваются этим: бактериальную клетку; клетку архей; клетку одноклеточного эукариотического организма; клетку растения; клетку водоросли, например, *Botryococcus braunii*, *Chlamydomonas reinhardtii*, *Nannochloropsis gaditana*, *Chlorella pyrenoidosa*, *Sargassum patens*, *C. agardh* и т.п.; клетку гриба (например, дрожжевую клетку); клетку животного; клетку беспозвоночного животного (например, плодовой мушки, стрекающих, иглокожих, нематоды и т.д.); клетку насекомого (например, москита; пчелы; сельскохозяйственного вредителя и т.д.); клетку паукообразных (например, паука; клеща и т.д.); клетку позвоночного животного (например, рыбы, амфибии, рептилии, птицы, млекопитающего); клетку млекопитающего (например, клетку грызуна, клетку человека; клетку отличного от человека млекопитающего; клетку грызуна (например, мыши, крысы); клетку зайцеобразных (например, кролика); клетку копытных (например, коровы, лошади, верблюда, ламы, викуньи, овцы, козы и т.д.); клетку морского млекопитающего (например,

кита, тюленя, морского слона, дельфина, морского льва и т.д.) и т.п. Интерес может представлять любой тип клеток (например, стволовые клетки, например, эмбриональные стволовые (ЭС) клетки, индуцированные плюрипотентные стволовые (иПС) клетки, половые клетки (например, ооциты, сперматозоиды, оогонии, сперматогонии и т.д.), взрослые стволовые клетки, соматические клетки, например, фибробласты, гемопоэтические клетки, нейроны, мышечные клетки, костные клетки, гепатоциты, панкреатические клетки; *in vitro* или *in vivo* эмбриональные клетки эмбриона на любой стадии, например, 1-клеточной, 2-клеточной, 4-клеточной, 8-клеточной и т.д. стадии эмбриона данио; и т.д.).

Клетки могут быть из стабильных линий клеток или же они могут быть первичными клетками, причем выражения "первичные клетки", "линии первичных клеток" и "первичные культуры" взаимозаменяемо употребляются в данном документе для обозначения клеток и культур клеток, которые были получены от субъекта и выращивались *in vitro* в течение ограниченного числа пассажей, т.е. делений культуры. Например, первичные культуры представляют собой культуры, которые могли пассировать 0 раз, 1 раз, 2 раза, 4 раза, 5 раз, 10 раз или 15 раз, но недостаточно времени, чтобы пройти через стадию кризиса. Как правило, линии первичных клеток поддерживают в течение менее 10 пассажей *in vitro*. Целевые клетки могут представлять собой одноклеточные организмы и/или могут быть выращены в культуре. Если клетки являются первичными клетками, их можно получать от индивида любым удобным способом. Например, лейкоциты удобно получать с помощью афереза, лейкоцитафереза, разделения в градиенте плотности и т.д., тогда как клетки из тканей, таких как кожа, мышцы, костный мозг, селезенка, печень, поджелудочная железа, легкие, кишечник, желудок и т.д., удобно получать с помощью биопсии.

В некоторых из вышеуказанных применений предложенные способы можно использовать, чтобы индуцировать расщепление целевой нуклеиновой кислоты, модификацию целевой нуклеиновой кислоты и/или связывание целевых нуклеиновых кислот (например, для визуализации, для сбора и/или анализа и т.д.) в митотических или постмитотических клетках *in vivo* и/или *ex vivo* и/или *in vitro* (например, для нарушения выработки белка, кодируемого целевой мРНК, для расщепления или какой-либо иной модификации целевой ДНК, для генетической модификации целевой клетки и т.п.). Так как направляющая РНК обеспечивает специфичность посредством гибридизации с целевой нуклеиновой кислотой, представляющая интерес в описанных способах митотическая или постмитотическая клетка может включать в себя любую клетку из любого организма (например, бактериальную клетку, клетку архей, клетку одноклеточного эукариотического организма, клетку растения, клетку водоросли, например, *Botryococcus braunii*, *Chlamydomonas reinhardtii*, *Nannochloropsis gaditana*, *Chlorella pyrenoidosa*, *Sargassum patens*, *C. agardh* и т.п., клетку гриба (например, дрожжевую клетку), клетку животного, клетку беспозвоночного животного (например, плодовой мушки, стрекающих, иглокожих, нематоды и т.д.), клетку позвоночного животного (например, рыбы, амфибии, рептилии, птицы, млекопитающего), клетку млекопитающего, клетку грызуна, клетку человека и т.д.). В некоторых случаях предложенный белок CasY (и/или нуклеиновую кислоту, кодирующую белок, такую как ДНК и/или РНК), и/или CasY-направляющую РНК (и/или ДНК, кодирующую направляющую РНК), и/или донорную матрицу, и/или РНП можно вносить в организм индивида (т.е. целевая клетка может существовать *in vivo*) (например, млекопитающего, крысы, мыши, свиньи, примата, отличного от человека примата, человека и т.д.). В некоторых случаях такое введение можно осуществлять в целях лечения и/или предотвращения заболевания, например, путем редактирования генома целевых клеток.

Клетки растений включают в себя клетки однодольных и клетки двудольных. Клетки могут представлять собой клетки корней, клетки листьев, клетки ксилемы, клетки флоэмы, клетки камбия, клетки апикальной меристемы, клетки паренхимы, клетки колленхимы, клетки склеренхимы и т.п. Клетки растений включают в себя клетки сельскохозяйственных растений, таких как пшеница, кукуруза, рис, сорго, просо, соя и т.д. Клетки растений включают в себя клетки сельскохозяйственных фруктовых и орехоплодных растений, например, растений, которые дают абрикосы, апельсины, лимоны, яблоки, сливы, персики, миндаль и т.д.

Дополнительные примеры целевых клеток перечислены выше в разделе, озаглавленном "Модифицированные клетки". Неограничивающие примеры клеток (целевых клеток) включают в себя: прокариотическую клетку, эукариотическую клетку, бактериальную клетку, клетку архей, клетку одноклеточного эукариотического организма, клетку простейших, клетку из растения (например, клетки из растительных культур, фруктов, овощей, зерновых, сои, кукурузы, маиса, пшеницы, семян, томатов, риса, маниоки, сахарного тростника, тыквы, сена, картофеля, хлопка, конопли, табака, цветковых растений, хвойных, голосеменных, покрытосеменных, папоротников, плаунов, антоцеровых, печеночников, мхов, двудольных, однодольных и т.д.), клетку водорослей (например, *Botryococcus braunii*, *Chlamydomonas reinhardtii*, *Nannochloropsis gaditana*, *Chlorella pyrenoidosa*, *Sargassum patens*, *C. agardh* и т.п.), клетку морских водорослей (например, бурых водорослей), клетку грибов (например, дрожжевую клетку, клетку из грибов), клетку животного, клетку беспозвоночного животного (например, плодовой мушки, стрекающих, иглокожих, нематоды и т.д.), клетку позвоночного животного (например, рыбы, амфибии, рептилии, птицы, млекопитающего), клетку млекопитающего (например, копытных, свиньи, коровы, козы, овцы); грызуна (например, крысы, мыши); отличного от человека примата; человека; кошачьих (например, кошки); собачьих (например, собаки) и т.д.) и т.п. В некоторых случаях клетка является клеткой, которая

не происходит из природного организма (например, клетка может быть синтетически созданной клеткой; также называемой искусственной клеткой).

Клетка может быть *in vitro* клеткой (например, стабильной культивируемой линией клеток). Клетка может быть *ex vivo* клеткой (культивируемой клеткой из отдельного организма). Клетка может быть *in vivo* клеткой (например, клеткой в отдельном организме). Клетка может быть выделенной клеткой. Клетка может быть клеткой внутри организма. Клетка может быть организмом. Клетка может быть клеткой в культуре клеток (например, *in vitro* культуре клеток). Клетка может быть клеткой из коллекции клеток. Клетка может быть прокариотической клеткой или быть полученной из прокариотической клетки. Клетка может быть бактериальной клеткой или быть полученной из бактериальной клетки. Клетка может быть клеткой архей или быть полученной из клетки архей. Клетка может быть эукариотической клеткой или быть полученной из эукариотической клетки. Клетка может быть растительной клеткой или быть полученной из растительной клетки. Клетка может быть клеткой животного или быть полученной из клетки животного. Клетка может быть клеткой позвоночного или быть полученной из клетки позвоночного. Клетка может быть клеткой млекопитающего или быть полученной из клетки млекопитающего. Клетка может быть клеткой грызуна или быть полученной из клетки грызуна. Клетка может быть клеткой человека или быть полученной из клетки человека. Клетка может быть микробной клеткой или быть полученной из микробной клетки. Клетка может быть клеткой грибов или быть полученной из клетки грибов. Клетка может быть клеткой насекомого. Клетка может быть клеткой членистоногого. Клетка может быть клеткой простейшего. Клетка может быть клеткой гельминта.

Подходящие клетки включают в себя стволовые клетки (например, эмбриональные стволовые (ЭС) клетки, индуцированные плюрипотентные стволовые (иПС) клетки; половые клетки (например, ооциты, сперматозоиды, оогонии, сперматогонии и т.д.); соматические клетки, например, фибробласты, олигодендроциты, глиальные клетки, гемопоэтические клетки, нейроны, мышечные клетки, костные клетки, гепатоциты, панкреатические клетки и т.д.

Подходящие клетки включают в себя человеческие эмбриональные стволовые клетки, фетальные кардиомиоциты, миофибробласты, мезенхимальные стволовые клетки, аутотрансплантированные размноженные кардиомиоциты, адипоциты, тотипотентные клетки, плюрипотентные клетки, кровяные стволовые клетки, миобласты, взрослые стволовые клетки, клетки костного мозга, мезенхимальные клетки, эмбриональные стволовые клетки, паренхимальные клетки, эпителиальные клетки, эндотелиальные клетки, мезотелиальные клетки, фибробласты, остеобласты, хондроциты, экзогенные клетки, эндогенные клетки, стволовые клетки, гемопоэтические стволовые клетки, полученные из костного мозга клетки-предшественники, миокардиальные клетки, скелетные клетки, фетальные клетки, недифференцированные клетки, мультипотентные клетки-предшественники, унипотентные клетки-предшественники, моноциты, сердечные миобласты, скелетные миобласты, макрофаги, капиллярные эндотелиальные клетки, ксеногенные клетки, аллогенные клетки и постнатальные стволовые клетки.

В некоторых случаях клетка представляет собой иммунную клетку, нейрон, эпителиальную клетку, эндотелиальную клетку или стволовую клетку. В некоторых случаях иммунная клетка представляет собой Т-клетку, В-клетку, моноцит, естественную клетку-киллера, дендритную клетку или макрофаг. В некоторых случаях иммунная клетка представляет собой цитотоксическую Т-клетку. В некоторых случаях иммунная клетка представляет собой хелперную Т-клетку. В некоторых случаях иммунная клетка представляет собой регуляторную Т-клетку (Treg).

В некоторых случаях клетка представляет собой стволовую клетку. Стволовые клетки включают в себя взрослые стволовые клетки. Взрослые стволовые клетки также называются соматическими стволовыми клетками.

Взрослые стволовые клетки присутствуют в дифференцированной ткани, но сохраняют свойства самообновления и способность давать начало множеству типов клеток, обычно типов клеток, типичных для ткани, в которой находятся стволовые клетки. Специалистам в данной области техники известны многочисленные примеры соматических стволовых клеток, включая мышечные стволовые клетки; гемопоэтические стволовые клетки; эпителиальные стволовые клетки; нейральные стволовые клетки; мезенхимальные стволовые клетки; стволовые клетки молочной железы; кишечные стволовые клетки; мезодермальные стволовые клетки; эндотелиальные стволовые клетки; обонятельные стволовые клетки; стволовые клетки нервного гребня и т.п.

Представляющие интерес стволовые клетки включают в себя стволовые клетки млекопитающих, при этом термин "млекопитающее" относится к любому животному, классифицированному как млекопитающее, включая людей; отличных от человека приматов; домашних и сельскохозяйственных животных; а также содержащихся в зоопарках, лабораторных, спортивных или домашних животных, таких как собаки, лошади, кошки, коровы, мыши, крысы, кролики и т.д. В некоторых случаях стволовая клетка является человеческой стволовой клеткой. В некоторых случаях стволовая клетка является стволовой клеткой грызуна (например, мыши; крысы). В некоторых случаях стволовая клетка является стволовой клеткой отличного от человека примата.

Стволовые клетки могут экспрессировать один или более маркеров стволовых клеток, например,

SOX9, KRT19, KRT7, LGR5, CA9, FXD2, CDH6, CLDN18, TSPAN8, BPIFB1, OLFM4, CDH17 и PPARGC1A.

В некоторых вариантах реализации стволовая клетка является гемопоэтической стволовой клеткой (ГСК). ГСК являются мезодермальными клетками, которые можно выделять из костного мозга, крови, пуповинной крови, фетальной печени и желточного мешка. ГСК делятся на CD34<sup>+</sup> и CD3<sup>-</sup>. ГСК могут восстанавливать популяции линий дифференцировки эритроидных клеток, нейтрофилов-макрофагов, мегакариоцитов и лимфоидных гемопоэтических клеток *in vivo*. ГСК можно индуцировать *in vitro* для прохождения по меньшей мере нескольких ведущих к самообновлению клеточных делений и можно индуцировать для дифференцировки в те же линии, которые наблюдаются *in vivo*. Следовательно, ГСК можно индуцировать для дифференцировки в одну или более из эритроидных клеток, мегакариоцитов, нейтрофилов, макрофагов и лимфоидных клеток.

В других вариантах реализации стволовая клетка является нейральной стволовой клеткой (НСК). Нейральные стволовые клетки (НСК) способны дифференцироваться в нейроны и глию (включая олигодендроциты и астроциты). Нейральная стволовая клетка является мультипотентной стволовой клеткой, которая способна проходить множество делений и в определенных условиях может давать дочерние клетки, которые представляют собой нейральные стволовые клетки или нейральные клетки-предшественники, которые могут быть нейробластами или глиобластами, например, клетки, которые должны стать одним или более типами нейронов и глиальных клеток, соответственно. Способы получения НСК известны в данной области техники.

В других вариантах реализации стволовая клетка является мезенхимальной стволовой клеткой (МСК). МСК, изначально полученные из эмбриональной мезодермы и выделенные из взрослого костного мозга, могут дифференцироваться с образованием мышц, костей, хрящей, жира, костномозговой стромы и сухожилий. Способы выделения МСК известны в данной области техники; а для получения МСК можно использовать любой известный способ. Смотрите, например, патент США № 5736396, в котором описано выделение человеческих МСК.

В некоторых случаях клетка является растительной клеткой. Растительная клетка может быть клеткой однодольного растения. Растительная клетка может быть клеткой двудольного растения.

В некоторых случаях клетка является растительной клеткой. Например, клетка может быть клеткой одного из основных сельскохозяйственных растений, например, ячменя, бобов (сухих съедобных), канолы, кукурузы, хлопка (Пима), хлопка ("Апланд"), льняного семени, сена (люцерны), сена (отличного от люцерны), овса, арахиса, риса, сорго, сои, сахарной свеклы, сахарного тростника, подсолнечника (масла), подсолнечника (не масла), сладкого картофеля, табака (Берли), табака (дымовой сушки), томатов, пшеницы (дурум), пшеницы (весенней), пшеницы (озимой) и т.п. В качестве другого примера клетка представляет собой клетку овощных культурных растений, которые включают в себя, но не ограничиваются этим, например, ростки люцерны, листья алоэ, корень триостренника, стрелолист, артишоки, спаржу, ростки бамбука, цветы банана, ростки бобов, бобы, свекловичную ботву, свеклу, горький огурец, китайскую капусту, брокколи, ботву брокколи (рапини), брюссельскую капусту, кочанную капусту, ростки капусты, листья кактуса (нопалес), бутылочную тыкву, кардон, морковь, цветную капусту, сельдерей, чайот, китайский артишок (хорогу), китайскую капусту, китайский сельдерей, лук пахучий, бок-чой, листья хризантем (тунг хо), листовую капусту, стебли кукурузы, сладкую кукурузу, огурцы, дайкон, молодые листья одуванчика, колоказию, доу миао (верхушки гороха), донква (зимнюю восковую тыкву), баклажан, цикорий салатный, эскариоль, рахисы папоротника, полевой кресс-салат, фризье, гай чой (китайскую горчицу), гай ян, галангал (сиамский, тайский имбирь), чеснок, корень имбиря, лопушник, зелень, ганноверскую салатную зелень, киноа, иерусалимские артишоки, хикама, кормовую капусту, кольраби, марь белую (триллиум), латук (биб), латук (бостонский), латук (бостонский красный), латук (зеленолиственный), латук (айсберг), латук (лолло росса), латук (дуболистный - зеленый), латук (дуболистный - красный), латук (обработанный), латук (краснолистный), латук (ромен), латук (красный ромен), латук (красную горчицу), линкок, ло бок, спаржевые бобы, корень лотоса, полевой салат, листья агавы (мэги), ксантозому, смесь месклан, курчаволистную горчицу, люффу (тыкву мочальную), му, мокву (волосатую тыкву), грибы, горчицу, нагаимо, бамию, водяной шпинат, зеленый лук, опо (длинную тыкву), декоративную кукурузу, горлянку обыкновенную, петрушку, пастернак, горох, перец (конусный), перец, тыквы, радикио, ростки редьки, редьку, зелень рапса, зелень рапса, ревень, ромен (бэби рэд), брюкву, солерос (энтаду ползучую), синкву (границкую/ребристую люффу), шпинат, тыкву, соломенные блоки, сахарный тростник, сладкий картофель, мангольд, тамаринд, таро, листья таро, побеги таро, китайскую плоскую капусту, леуцену, тиндору, физалис, томаты, томаты (черри), томаты (виноградные), томаты (сливки), куркуму, ботву молодой репы, репу, водяной орех, диоскорею, ямс (разные виды), юй чой, юку (кассаву) и т.п.

В некоторых случаях клетка может быть клеткой членистоногого. Например, клетка может быть клеткой подотряда, семейства, подсемейства, группы, подгруппы или вида, например,

*Chelicerata, Myriapodia, Hexipodia, Arachnida, Insecta, Archaeognatha, Thysamura, Palaeoptera, Ephemeroptera, Odonata, Anisoptera, Zygoptera, Neoptera, Exopterygota, Plecoptera, Embioptera, Orthoptera, Zoraptera, Dermaptera, Dictyoptera, Notoptera, Grylloblattidae, Mantophasmatidae, Phasmatodea, Blattaria, Isoptera, Mantodea, Parapneuroptera, Psocoptera, Thysanoptera, Phthiraptera, Hemiptera, Endopterygota или Holometabola, Hymenoptera, Coleoptera, Strepsiptera, Raphidioptera, Megaloptera, Neuroptera, Mecoptera, Siphonaptera, Diptera, Trichoptera или Lepidoptera.*

В некоторых случаях клетка может быть клеткой насекомого. Например, в некоторых случаях клетка является клеткой москита, кузнечика, клопа, мухи, блохи, пчелы, осы, муравья, вши, мотылька или жука.

Внесение компонентов в целевую клетку.

Cas9-направляющую РНК (или нуклеиновую кислоту, содержащую кодирующую ее нуклеотидную последовательность), и/или слитый полипептид Cas9 (или нуклеиновую кислоту, содержащую кодирующую его нуклеотидную последовательность), и/или донорный полинуклеотид можно вносить в клетку-хозяина любым из ряда хорошо известных способов.

Способы внесения нуклеиновой кислоты в клетку-хозяина известны в данной области техники, а для внесения нуклеиновой кислоты (например, экспрессионной конструкции) в целевую клетку (например, эукариотическую клетку, клетку человека, стволовую клетку, клетку-предшественника и т.п.) можно использовать любой удобный способ. Подходящие способы более подробно описаны в другом месте данного документа и включают в себя, например, инфекцию вирусом или бактериофагом, трансфекцию, конъюгацию, слияние протопластов, липофекцию, электропорацию, осаждение фосфатом кальция, полиэтиленимин (ПЭИ)-опосредованную трансфекцию, ДЭАЭ-декстран-опосредованную трансфекцию, липосомно-опосредованную трансфекцию, технологию генной пушки, осаждение фосфатом кальция, прямую микроинъекцию, опосредованную наночастицами доставку нуклеиновой кислоты (смотри, например, Panyam et al., Adv Drug Deliv Rev. 2012 Sep 13. pii: S0169-409X(12)00283-9. doi:10.1016/j.addr.2012.09.023) и т.п. Любой или все компоненты можно вносить в клетку в виде композиции (например, включая любую удобную комбинацию: полипептида CasY, CasY-направляющей РНК, донорного полинуклеотида и т.д.), используя известные способы, например, такие как нуклеофекция.

Донорный полинуклеотид (донорная матрица).

Направляемый CasY-направляющей РНК белок CasY в некоторых случаях создает сайт-специфические двухцепочечные разрывы (ДЦР) или одноцепочечные разрывы (ОЦР) (например, когда белок CasY является вариантом никазы) в целевых нуклеиновых кислотах в виде двухцепочечной ДНК (дцДНК), репарация которых осуществляется посредством негомологичного соединения концов (НГСК) или гомологичной рекомбинации (ГР).

В некоторых случаях приведение в контакт целевой ДНК (с белком CasY и CasY-направляющей РНК) происходит в условиях, которые допускают негомологичное соединение концов или гомологичную репарацию. Таким образом, в некоторых случаях предложенный способ включает в себя приведение целевой ДНК в контакт с донорным полинуклеотидом (например, посредством внесения донорного полинуклеотида в клетку), при этом донорный полинуклеотид, часть донорного полинуклеотида, копия донорного полинуклеотида или часть копии донорного полинуклеотида интегрируется в целевую ДНК. В некоторых случаях способ не включает в себя приведение клетки в контакт с донорным полинуклеотидом, а целевую ДНК модифицируют так, чтобы удалить нуклеотиды в целевой ДНК.

В некоторых случаях CasY-направляющую РНК (или кодирующую ее ДНК) и белок CasY (или кодирующую его нуклеиновую кислоту, такую как РНК или ДНК, например, один или более экспрессионных векторов) вводят совместно (например, приводят в контакт с целевой нуклеиновой кислотой, вводимой в клетки, и т.д.) с последовательностью донорного полинуклеотида, который содержит по меньшей мере один сегмент с гомологией с целевой последовательностью ДНК, предложенные способы можно использовать для добавления, т.е. вставки или замены, материала нуклеиновой кислоты в целевую последовательность ДНК (например, для "нокина" нуклеиновой кислоты, например, кодирующей белок, миРНК, микроРНК и т.д.), для добавления метки (например, 6×His, флуоресцентного белка (например, зеленого флуоресцентного белка; желтого флуоресцентного белка и т.д.), гемагглютинина (HA), FLAG и т.д.), для добавления регуляторной последовательности к гену (например, промотора, сигнала полиаденилирования, последовательности внутренней посадки рибосомы (IRES), 2A-пептида, стартового кодона, стоп-кодона, сигнала сплайсинга, сигнала локализации и т.д.), для модификации последовательности нуклеиновой кислоты (например, внесения мутации, удаления вызывающей заболевание мутации путем внесения правильной последовательности) и т.п. Следовательно, комплекс, содержащий CasY-направляющую РНК и белок CasY, можно использовать в любом *in vitro* или *in vivo* применении, в кото-

ром необходимо модифицировать ДНК сайт-специфическим, т.е. "прицельным", образом, например, с помощью генного нокаута, генного нокина, генного редактирования, генного мечения и т.д., что применяется, например, в генной терапии, например, для лечения заболевания, или в качестве противовирусного, антипатогенного или противоракового терапевтического средства, для получения генетически модифицированных организмов в сельском хозяйстве, для крупномасштабной выработки белков клетками в терапевтических, диагностических или исследовательских целях, для индукции ИПС клеток, биологических исследований, нацеливания на гены патогенов для их удаления или замещения и т.д.

В применениях, в которых необходима вставка полинуклеотидной последовательности в геном в месте расщепления целевой последовательности, в клетку также можно вносить донорный полинуклеотид (нуклеиновую кислоту, содержащую донорную последовательность). Под "донорной последовательностью", или "донорным полинуклеотидом", или "донорной матрицей" подразумевается последовательность нуклеиновой кислоты, подлежащая вставке в сайте, расщепляемом белком CasY (например, после расщепления дцДНК, после создания одноцепочечного разрыва в целевой ДНК, после двойного создания одноцепочечного разрыва в целевой ДНК и т.п.). Донорный полинуклеотид может обладать достаточной гомологией с геномной последовательностью в целевом сайте, например, 70%, 80%, 85%, 90%, 95% или 100% гомологией с нуклеотидными последовательностями, фланкирующими целевой сайт, например, в пределах 50 оснований или менее от целевого сайта, например, в пределах около 30 оснований, в пределах около 15 оснований, в пределах около 10 оснований, в пределах около 5 оснований или непосредственно фланкирующими целевой сайт, чтобы поддерживать гомологичную репарацию между ним и геномной последовательностью, с которой он гомологичен. Приблизительно 25, 50, 100 или 200 нуклеотидов или более 200 нуклеотидов гомологии последовательности между донорной и геномной последовательностью (или любое целочисленное значение между 10 и 200 нуклеотидами или более) могут поддерживать гомологичную репарацию. Донорный полинуклеотид может иметь любую длину, например, 10 нуклеотидов или более, 50 нуклеотидов или более, 100 нуклеотидов или более, 250 нуклеотидов или более, 500 нуклеотидов или более, 1000 нуклеотидов или более, 5000 нуклеотидов или более и т.д.

Донорная последовательность, как правило, не идентична с геномной последовательностью, которую она замещает. Правильнее сказать, что донорная последовательность может содержать по меньшей мере одну или более одноосновных замен, вставок, делеций, инверсий или перестроек относительно геномной последовательности при условии наличия достаточной гомологии для поддержания гомологичной репарации (например, для генной коррекции, например, для преобразования вызывающей заболевание пары оснований или не вызывающей заболевание пары оснований). В некоторых вариантах реализации донорная последовательность содержит негомологичную последовательность, фланкируемую двумя областями гомологии, так, что гомологичная репарация между целевой областью ДНК и двумя фланкирующими последовательностями приводит к вставке негомологичной последовательности в целевой области. Донорные последовательности также могут содержать векторный остов, содержащий последовательности, негомологичные представляющей интерес области ДНК, которые не предназначены для вставки в представляющую интерес область ДНК. В общем случае гомологичные области донорной последовательности имеют по меньшей мере 50% идентичности последовательности с геномной последовательностью, с которой необходима рекомбинация. В определенных вариантах реализации наблюдается 60%, 70%, 80%, 90%, 95%, 98%, 99% или 99,9% идентичности последовательности. Может наблюдаться любое значение идентичности последовательности между 1% и 100% в зависимости от длины донорного полипептида.

Донорная последовательность может содержать определенные различия в последовательности по сравнению с геномной последовательностью, например, рестрикционные сайты, нуклеотидный полиморфизм, селективные маркеры (например, гены резистентности к лекарственным препаратам, флуоресцентные белки, ферменты и т.д.) и т.д., которые можно использовать для оценки успешности вставки донорной последовательности в сайте расщепления или в некоторых случаях можно использовать в других целях (например, для того, чтобы показать экспрессию в целевом геномном локусе). В некоторых случаях при расположении в кодирующей области такие различия в нуклеотидной последовательности не меняют аминокислотную последовательность или создают молчащие аминокислотные изменения (т.е. изменения, которые не влияют на структуру или функцию белка). В альтернативном варианте эти различия в последовательности могут включать в себя фланкирующие рекомбинационные последовательности, такие как последовательности FLP, loxP и т.п., которые могут активироваться позднее для удаления маркерной последовательности.

В некоторых случаях донорную последовательность вводят в клетку в виде одноцепочечной ДНК. В некоторых случаях донорную последовательность вводят в клетку в виде двухцепочечной ДНК. Ее можно вносить в клетку в линейной или кольцевой форме. При внесении в линейной форме концы донорной последовательности могут быть защищены (например, от экзонуклеолитической деградации) любым удобным способом, и такие способы известны специалистам в данной области техники. Например, можно добавлять один или более дидезоксирибонуклеотидных остатков к 3' концу линейной молекулы, и/или самокомплементарные олигонуклеотиды можно лигировать с одним или обоими концами. Смотрите, например, Chang et al. (1987) Proc. Natl. Acad. Sci. USA 84:4959-4963; Nehls et al. (1996) Science

272:886-889. Дополнительные способы защиты экзогенных полинуклеотидов от деградации включают в себя, но не ограничиваются этим, добавление концевых аминокислот и применение модифицированных межнуклеотидных связей, таких как, например, тиофосфаты, фосфорамидаты и остатки О-метилрибозы или дезоксирибозы. В качестве альтернативного варианта защиты концов линейной донорной последовательности за пределами области гомологии можно добавлять дополнительную длину последовательностей, которые могут подвергаться деградации, не влияя на рекомбинацию. Донорную последовательность можно вносить в клетку в виде части векторной молекулы, имеющей дополнительные последовательности, такие как, например, точки начала репликации, промоторы и гены, кодирующие резистентность к антибиотикам. Кроме того, донорные последовательности можно вносить в виде оголенной нуклеиновой кислоты, когда нуклеиновая кислота находится в комплексе с таким агентом, как липосома или полоксамер, или можно доставлять с помощью вирусов (например, аденовируса, ААВ), как описано в другом месте данного документа для нуклеиновых кислот, кодирующих CasY-направляющую РНК, и/или слитый полипептид CasY, и/или донорный полинуклеотид.

Трансгенные, отличные от человека организмы.

Как описано выше, в некоторых случаях нуклеиновую кислоту (например, рекомбинантный экспрессионный вектор) согласно настоящему изобретению (например, нуклеиновую кислоту, содержащую нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобретению; нуклеиновую кислоту, содержащую нуклеотидную последовательность, кодирующую слитый полипептид CasY согласно настоящему изобретению; и т.д.) используют как трансген для создания трансгенного отличного от человека организма, который вырабатывает полипептид CasY или слитый полипептид CasY согласно настоящему изобретению. В настоящем изобретении предложен трансгенный отличный от человека организм, содержащий нуклеотидную последовательность, кодирующую полипептид CasY или слитый полипептид CasY согласно настоящему изобретению.

Трансгенные, отличные от человека животные.

В настоящем изобретении предложено трансгенное отличное от человека животное, которое содержит трансген, содержащий нуклеиновую кислоту, содержащую нуклеотидную последовательность, кодирующую полипептид CasY или слитый полипептид CasY. В некоторых вариантах реализации геном трансгенного отличного от человека животного содержит нуклеотидную последовательность, кодирующую полипептид CasY или слитый полипептид CasY согласно настоящему изобретению. В некоторых случаях трансгенное отличное от человека животное является гомозиготным в отношении генетической модификации. В некоторых случаях трансгенное отличное от человека животное является гетерозиготным в отношении генетической модификации. В некоторых вариантах реализации трансгенное отличное от человека животное является позвоночным, например, рыбой (например, лососем, форелью, данио, золотой рыбкой, иглобрюхом, рыбой-ангелом и т.д.), амфибией (лягушкой, тритоном, саламандрой и т.д.), птицей (например, курицей, индюшкой и т.д.), рептилией (например, змеей, ящерицей и т.д.), отличным от человека млекопитающим (например, копытным, например, свиньей, коровой, козой, овцой и т.д.; зайцеобразным (например, кроликом); грызуном (например, крысой, мышью); отличным от человека приматом и т.д.) и т.д. В некоторых случаях трансгенное отличное от человека животное является беспозвоночным. В некоторых случаях трансгенное отличное от человека животное является насекомым (например, комаром; сельскохозяйственным вредителем и т.д.). В некоторых случаях трансгенное отличное от человека животное является паукообразным.

Нуклеотидные последовательности, кодирующие полипептид CasY или слитый полипептид CasY согласно настоящему изобретению, могут находиться под управлением неизвестного промотора (т.е. быть функционально связанными с ним) (например, когда нуклеиновая кислота случайным образом интегрируется в геном клетки-хозяина) или могут находиться под управлением известного промотора (т.е. быть функционально связанными с ним). Подходящие известные промоторы могут быть любыми известными промоторами и включают в себя конститутивно активные промоторы (например, промотор ЦМВ), индуцибельные промоторы (например, промотор теплового шока, тетрациклин-регулируемый промотор, стероид-регулируемый промотор, металл-регулируемый промотор, промотор, регулируемый рецептором эстрогена, и т.д.), пространственно ограниченные и/или временно ограниченные промоторы (например, тканеспецифический промотор, специфический в отношении клеточного типа промотор и т.д.) и т.д.

Трансгенные растения.

Как описано выше, в некоторых случаях нуклеиновую кислоту (например, рекомбинантный экспрессионный вектор) согласно настоящему изобретению (например, нуклеиновую кислоту, содержащую нуклеотидную последовательность, кодирующую полипептид CasY согласно настоящему изобретению; нуклеиновую кислоту, содержащую нуклеотидную последовательность, кодирующую слитый полипептид CasY согласно настоящему изобретению; и т.д.) используют как трансген для создания трансгенного растения, которое вырабатывает полипептид CasY или слитый полипептид CasY согласно настоящему изобретению. В настоящем изобретении предложено трансгенное растение, содержащее нуклеотидную последовательность, кодирующую полипептид CasY или слитый полипептид CasY согласно настоящему изобретению. В некоторых вариантах реализации геном трансгенного растения содержит предложенную

нуклеиновую кислоту. В некоторых вариантах реализации трансгенное растение является гомозиготным в отношении генетической модификации. В некоторых вариантах реализации трансгенное растение является гетерозиготным в отношении генетической модификации.

Способы внесения экзогенных нуклеиновых кислот в клетки растений хорошо известны в данной области техники. Такие клетки растений считаются "трансформированными" согласно определению выше. Подходящие способы включают в себя вирусную инфекцию (например, вирусами, имеющими двухцепочечную ДНК), трансфекцию, конъюгацию, слияние протопластов, электропорацию, технологию генной пушки, осаждение фосфатом кальция, прямую микроинъекцию, технологию с применением усов из карбида кремния, *Agrobacterium*-опосредованную трансформацию и т.п. Выбор способа в общем случае зависит от типа трансформируемой клетки и условий, в которых проходит трансформация (т.е. *in vitro*, *ex vivo* или *in vivo*).

Способы трансформации на основе почвенной бактерии *Agrobacterium tumefaciens* являются исключительно удобными для внесения молекулы экзогенной нуклеиновой кислоты в сосудистое растение. Форма дикого типа *Agrobacterium* содержит T<sub>i</sub> (опухоль-индуцирующую) плазмиду, которая управляет выработкой туморогенных корончатых галлов на растениях-хозяевах. Перенос опухоль-индуцирующей области T-ДНК T<sub>i</sub>-плазмиды в геном растения требует наличия кодируемых T<sub>i</sub>-плазмидой генов вирулентности, а также границ T-ДНК, которые представляют собой набор прямых повторов ДНК, которые обозначают область, предназначенную для переноса. Вектор на основе *Agrobacterium* является модифицированной формой T<sub>i</sub>-плазмиды, в которой опухоль-индуцирующие функции заменены представляющей интерес последовательностью нуклеиновой кислоты, предназначенной для внесения в растение-хозяина.

В *Agrobacterium*-опосредованной трансформации в общем случае применяются совместно интегрирующиеся векторы или бинарные векторные системы, в которых компоненты T<sub>i</sub>-плазмиды разделены между вектором-помощником, который постоянно находится в хозяине *Agrobacterium* и несет гены вирулентности, и вектором-челноком, который содержит представляющий интерес ген, связанный последовательностями T-ДНК. Ряд бинарных векторов хорошо известны в данной области техники и являются коммерчески доступными, например, от Clontech (Пало-Альто, штат Калифорния). Способы совместного культивирования *Agrobacterium* с культивируемыми растительными клетками или поврежденной тканью, такой как, например, листовая ткань, корневые эксплантаты, подсемядоли, части стеблей или клубней, также хорошо известны в данной области техники. Смотрите, например, Glick and Thompson, (eds.), *Methods in Plant Molecular Biology and Biotechnology*, Boca Raton, Fla.: CRC Press (1993).

Для получения предложенного трансгенного растения также можно использовать опосредованную микрочастицами трансформацию. Этот способ, впервые описанный Klein et al. (*Nature* 327:70-73 (1987)), основан на применении микрочастиц, таких как золотые или вольфрамовые, покрытых молекулами необходимой нуклеиновой кислоты посредством осаждения хлоридом кальция, спермидином или полиэтиленгликолем. Микрочастицы, ускоренные до высоких скоростей, направляют в ткани покрытосеменных растений, используя устройство, такое как BIOLISTIC PD-1000 (Biorad; Геркулес, штат Калифорния).

Нуклеиновую кислоту согласно настоящему изобретению (например, нуклеиновую кислоту (например, рекомбинантный экспрессионный вектор), содержащую нуклеотидную последовательность, кодирующую полипептид CasY или слитый полипептид CasY согласно настоящему изобретению) можно вносить в растение таким способом, чтобы нуклеиновая кислота могла попадать в клетку(и) растения, например, следуя *in vivo* или *ex vivo* протоколу. Под "*in vivo*" подразумевается, что нуклеиновую кислоту вводят в живое тело растения, например, посредством инфильтрации. Под "*ex vivo*" подразумевается, что клетки или эксплантаты модифицированы вне растения, а затем такие клетки или органы регенерируют в растении. Был описан ряд векторов, подходящих для стабильной трансформации клеток растений или для создания трансгенных растений, включая описанные в Weissbach and Weissbach (1989) *Methods for Plant Molecular Biology* Academic Press, и Gelvin et al. (1990) *Plant Molecular Biology Manual*, Kluwer Academic Publishers. Конкретные примеры включают в себя полученные с помощью T<sub>i</sub>-плазмиды *Agrobacterium tumefaciens*, а также описанные Herrera-Estrella et al. (1983) *Nature* 303: 209, Bevan (1984) *Nucl Acid Res.* 12: 8711-8721, Klee (1985) *Bio/Technolo* 3:637-642. В альтернативном варианте можно использовать не содержащие T<sub>i</sub> векторы для переноса ДНК в растения и клетки, используя способы доставки свободной ДНК. Используя эти способы можно получать трансгенные растения, такие как пшеница, рис (Christou (1991) *Bio/Technology* 9:957-9 and 4462) и кукуруза (Gordon-Kamm (1990) *Plant Cell* 2:603-618). Незрелый эмбрион также может служить хорошей целевой тканью в случае однодольных для способов прямой доставки ДНК с применением генной пушки (Weeks et al. (1993) *Plant Physiol* 102:1077-1084; Vasil (1993) *Bio/Technolo* 10:667-674; Wan and Lemeaux (1994) *Plant Physiol* 104:37-48 и для *Agrobacterium*-опосредованного переноса ДНК (Ishida et al. (1996) *Nature Biotech* 14:745-750). Типовыми способами внесения ДНК в хлоропласты являются биолистическая бомбардировка, трансформация протопластов полиэтиленгликолем и микроинъекция (Danieli et al., *Nat. Biotechnol* 16:345-348, 1998; Staub et al., *Nat. Biotechnol* 18: 333-338, 2000; O'Neill et al., *Plant J.* 3:729-738, 1993; Knoblauch et al., *Nat. Biotechnol* 17: 906-909; патенты США №№ 5451513, 5545817, 5545818 и 5576198; междун. заявка № WO 95/16783 и Boynton et al., *Methods in Enzymology* 217: 510-536 (1993), Svab et al., *Proc. Natl. Acad. Sci. USA* 90: 913-917 (1993),

и McBride et al., Proc. Natl. Acad. Sci. USA 91:7301-7305 (1994)). Любой вектор, подходящий для таких способов, как биолистическая бомбардировка, трансформация протопластов полиэтиленгликолем и микроинъекция, подойдет в качестве нацеливающего вектора для трансформации хлоропластов. Любой вектор на основе двухцепочечной ДНК можно использовать как трансформационный вектор, в особенности, когда в способе внесения не используется *Agrobacterium*.

Растения, которые могут быть генетически модифицированы, включают в себя зерновые, кормовые культуры, фрукты, овощи, масляные растения, пальмы, лесохозяйственные растения и виноград. Конкретными примерами растений, которые могут быть модифицированы, являются следующие: маис, бананы, арахис, полевой горох, подсолнечник, томаты, канола, табак, пшеница, ячмень, овес, картофель, соя, хлопок, гвоздика, сорго, люпин и рис.

В настоящем изобретении предложены трансформированные растительные клетки, ткани, растения и продукты, которые содержат трансформированные растительные клетки. Признаком предложенных трансформированных клеток, тканей и продуктов, которые их содержат, является наличие предложенной нуклеиновой кислоты, интегрированной в геном, и выработка растительными клетками полипептида CasY или слитого полипептида CasY согласно настоящему изобретению. Рекомбинантные растительные клетки согласно настоящему изобретению применимы как популяции рекомбинантных клеток или как ткани, семена, цельные растения, стебли, фрукты, листья, корни, цветы, стебли, клубни, зерна, корм для животных, поле растений и т.п.

Нуклеотидные последовательности, кодирующие полипептид CasY или слитый полипептид CasY согласно настоящему изобретению, могут находиться под управлением неизвестного промотора (т.е. быть функционально связанными с ним) (например, когда нуклеиновая кислота случайным образом интегрируется в геном клетки-хозяина) или могут находиться под управлением известного промотора (т.е. быть функционально связанными с ним). Подходящие известные промоторы могут быть любыми известными промоторами и включают в себя конститутивно активные промоторы, индуцибельные промоторы, пространственно ограниченные и/или временно ограниченные промоторы и т.д.

Способы идентификации РНК-направляемой эндонуклеазы CRISPR.

Предложены способы идентификации РНК-направляемой эндонуклеазы CRISPR. Например, в некоторых вариантах реализации такой способ включает в себя этап выявления во множестве метагеномных нуклеотидных последовательностей нуклеотидной последовательности, кодирующей полипептид Cas1. Белки Cas1 известны в данной области техники и находятся вблизи локусов CRISPR систем CRISPR класса 2, тех систем CRISPR, которые содержат один эффекторный белок, который функционирует как эндонуклеаза, а для его надлежащего функционирования не требуется взаимодействие с комплексом белков. Хотя сам белок Cas1 вовлечен в комплектацию новых целевых последовательностей в локусе CRISPR и, следовательно, не является необходимым эффекторным белком для идентификации этим способом, присутствие белка Cas1 вблизи локуса CRISPR является показателем того, что по меньшей мере один из других белков Cas, присутствующих рядом с локусом, может быть эффекторным белком (РНК-направляемой эндонуклеазой).

В контексте данного документа термин "метагеномика" означает параллельный анализ нуклеиновых кислот, выделенных из множества микроорганизмов (например, бактерий, архей и т.д.) в образце, например, образце окружающей среды, таком как образец, который содержит неизвестное количество прокариот (бактерий/архей) и может содержать прокариоты, еще не обнаруженные и/или не изученные до этого момента. Нуклеиновые кислоты можно выделять из такого образца любым удобным способом, и в общем случае нуклеиновые кислоты выделяют вместе из всего образца так, что перед анализом неизвестно, из какого организма происходит каждая заданная молекула нуклеиновой кислоты. В некоторых вариантах реализации образец содержит неизвестную смесь и/или неизвестное количество микроорганизмов. После этого можно проводить секвенирование нуклеиновых кислот для создания множества метагеномных последовательностей. В некоторых случаях предложенный способ идентификации РНК-направляемой эндонуклеазы CRISPR включает в себя этап выделения образца (например, образца окружающей среды). В некоторых случаях предложенный способ идентификации РНК-направляемой эндонуклеазы CRISPR включает в себя этап выделения нуклеиновых кислот из образца и/или анализа образца для создания множества метагеномных нуклеотидных последовательностей из образца.

После идентификации белка Cas1 предложенный способ идентификации РНК-направляемой эндонуклеазы CRISPR может включать в себя этап обнаружения матрицы CRISPR (матрицы повтор-спейс-повтор) вблизи кодирующей Cas1 нуклеотидной последовательности. Затем способ может включать в себя этап клонирования (например, из образца нуклеиновой кислоты, из которого было получено множество метагеномных нуклеотидных последовательностей) локуса CRISPR, содержащего обнаруженную матрицу CRISPR, в экспрессионный вектор для создания рекомбинантного экспрессионного вектора с локусом CRISPR. Затем локус CRISPR можно исследовать в отношении функции, анализируя способность рекомбинантного экспрессионного вектора с локусом CRISPR расщеплять целевую нуклеиновую кислоту. Можно использовать любой удобный способ. В некоторых вариантах реализации этап анализа включает в себя внесение рекомбинантного экспрессионного вектора с локусом CRISPR и целевой нуклеиновой кислоты в клетку, например, гетерологичную клетку-хозяина, такую как клетка *E. coli*. Напри-

мер, смотрите анализы истощения РАМ в рабочих примерах ниже (фиг. 5). В некоторых случаях этап анализа включает в себя внесение в популяцию клеток-хозяев (например, клеток *E. coli*) плазмидной библиотеки, причем каждая плазида из библиотеки содержит от 4 до 10 (например, от 5 до 10, от 5 до 8, от 6 до 10, от 6 до 8, 5, 6, 7, 8) нуклеотидов, рандомизированных 5' и/или 3' относительно целевой последовательности. Клетки-хозяева могут уже содержать подлежащий исследованию рекомбинантный экспрессионный вектор с локусом CRISPR, или же рекомбинантный экспрессионный вектор с локусом CRISPR можно вносить после библиотеки. Только те из исследуемых локусов CRISPR, которые являются функциональными и, следовательно, содержат функциональную РНК-направляемую эндонуклеазу CRISPR, будут обеспечивать способность расщепления плазмид, содержащих целевую последовательность. Причина включения рандомизированных последовательностей 5' и 3' относительно целевой последовательности состоит в том, что последовательность РАМ, необходимая для конкретной эндонуклеазы, может быть неизвестной в начале эксперимента.

Если экспрессионный вектор может расщеплять целевую нуклеиновую кислоту (например, с соответствующей целевой последовательностью и РАМ, такой как целевая последовательность, которая совпадает по меньшей мере с одним спейсером матрицы CRISPR), то локус CRISPR содержит нуклеотидную последовательность, кодирующую предполагаемую РНК-направляемую эндонуклеазу CRISPR. Таким образом, затем можно определить открытую рамку считывания из локуса CRISPR, которая кодирует РНК-направляемую эндонуклеазу CRISPR. В некоторых случаях необходимо идентифицировать ранее неизвестную РНК-направляемую эндонуклеазу CRISPR, и, следовательно, в некоторых случаях идентифицированный полипептид имеет менее чем 20% идентичности аминокислотной последовательности (например, менее чем 15%, менее чем 10%, менее чем 5% идентичности аминокислотной последовательности) с аминокислотной последовательностью полипептида известной РНК-направляемой эндонуклеазы CRISPR.

#### Примеры неограничивающих аспектов изобретения

Аспекты, включая варианты реализации, предложенного предмета изобретения, описанные выше, могут обеспечивать преимущество отдельно или в комбинации с одним или более другими аспектами или вариантами реализации. Без ограничения вышеприведенного описания ниже приведены определенные неограничивающие аспекты изобретения, пронумерованные 1-123. После прочтения этого описания специалистам в данной области техники станет понятно, что каждый из отдельно пронумерованных аспектов можно использовать или комбинировать с любым из предыдущих или следующих отдельно пронумерованных аспектов. Подразумевается, что это обеспечит основу для всех таких комбинаций аспектов и не ограничено комбинациями аспектов, явным образом приведенными ниже:

Аспекты.

1. Композиция, содержащая:

a) полипептид CasY или молекулу нуклеиновой кислоты, кодирующую полипептид CasY; и

b) CasY-направляющую РНК или одну или более молекул ДНК, кодирующих CasY-направляющую РНК.

2. Композиция по п.1, отличающаяся тем, что полипептид CasY содержит аминокислотную последовательность, имеющую 50% или более идентичности с аминокислотной последовательностью, приведенной в SEQ ID NO: 1 или SEQ ID NO: 2 (или аминокислотной последовательностью, приведенной в любой из SEQ ID NO: 1-8).

3. Композиция по п.1 или 2, отличающаяся тем, что CasY-направляющая РНК содержит нуклеотидную последовательность, имеющую 80% или более идентичности с последовательностью cr-РНК, приведенной в любой из SEQ ID NO: 11-15.

4. Композиция по п.1 или 2, отличающаяся тем, что полипептид CasY слит с последовательностью СЯЛ.

5. Композиция по любому из пп.1-4, отличающаяся тем, что композиция содержит липид.

6. Композиция по любому из пп.1-4, отличающаяся тем, что a) и b) находятся внутри липосомы.

7. Композиция по любому из пп.1-4, отличающаяся тем, что a) и b) находятся внутри частицы.

8. Композиция по любому из пп.1-7, содержащая одно или более из: буфера, ингибитора нуклеаз и ингибитора протеаз.

9. Композиция по любому из пп.1-8, отличающаяся тем, что полипептид CasY содержит аминокислотную последовательность, имеющую 85% или более идентичности с аминокислотной последовательностью, приведенной в SEQ ID NO: 1 или SEQ ID NO: 2 (или аминокислотной последовательностью, приведенной в любой из SEQ ID NO: 1-8).

10. Композиция по любому из пп.1-9, отличающаяся тем, что полипептид CasY представляет собой нуклеазу, которая может расщеплять только одну цепь двухцепочечной молекулы целевой нуклеиновой кислоты.

11. Композиция по любому из пп.1-9, отличающаяся тем, что полипептид CasY представляет собой каталитически неактивный полипептид CasY (dCasY).

12. Композиция по п.10 или 11, отличающаяся тем, что полипептид CasY содержит одну или более мутаций в положении, соответствующем выбранным из: D672, E769 и D935 из SEQ ID NO: 1.

13. Композиция по любому из пп.1-12, дополнительно содержащая донорную матричную ДНК.
14. Слитый полипептид CasY, содержащий полипептид CasY, слитый с гетерологичным полипептидом.
15. Слитый полипептид CasY по п.14, отличающийся тем, что полипептид CasY содержит аминокислотную последовательность, имеющую 50% или более идентичности с аминокислотной последовательностью, приведенной в SEQ ID NO: 1 или SEQ ID NO: 2 (или аминокислотной последовательностью, приведенной в любой из SEQ ID NO: 1-8).
16. Слитый полипептид CasY по п.14, отличающийся тем, что полипептид CasY содержит аминокислотную последовательность, имеющую 85% или более идентичности с аминокислотной последовательностью, приведенной в SEQ ID NO: 1 или SEQ ID NO: 2 (или аминокислотной последовательностью, приведенной в любой из SEQ ID NO: 1-8).
17. Слитый полипептид CasY по любому из пп.14-16, отличающийся тем, что полипептид CasY представляет собой нуклеазу, которая может расщеплять только одну цепь двухцепочечной молекулы целевой нуклеиновой кислоты.
18. Слитый полипептид CasY по любому из пп.14-17, отличающийся тем, что полипептид CasY представляет собой каталитически неактивный полипептид CasY (dCasY).
19. Слитый полипептид CasY по п.17 или 18, отличающийся тем, что полипептид CasY содержит одну или более мутаций в положении, соответствующем выбранным из: D672, E769 и D935 из SEQ ID NO: 1.
20. Слитый полипептид CasY по любому из пп.14-19, отличающийся тем, что гетерологичный полипептид слит с N-концом и/или C-концом полипептида CasY.
21. Слитый полипептид CasY по любому из пп.14-20, содержащий СЯЛ.
22. Слитый полипептид CasY по любому из пп.14-21, отличающийся тем, что гетерологичный полипептид представляет собой нацеливающий полипептид, который обеспечивает связывание с компонентом клеточной поверхности на целевой клетке или целевом типе клеток.
23. Слитый полипептид CasY по любому из пп.14-21, отличающийся тем, что гетерологичный полипептид проявляет ферментативную активность, которая модифицирует целевую ДНК.
24. Слитый полипептид CasY по п.23, отличающийся тем, что гетерологичный полипептид проявляет один или более видов ферментативной активности, выбранных из: нуклеазной активности, метилтрансферазной активности, деметилазной активности, активности ДНК-репарации, ДНК-повреждающей активности, деаминирующей активности, дисмутазной активности, алкилирующей активности, депурицирующей активности, окислительной активности, активности образования пиримидиновых димеров, интегразной активности, транспозазной активности, рекомбиназной активности, полимеразной активности, лигазной активности, геликазной активности, фотолиазной активности и гликозилазной активности.
25. Слитый полипептид CasY по п.24, отличающийся тем, что гетерологичный полипептид проявляет один или более видов ферментативной активности, выбранных из: нуклеазной активности, метилтрансферазной активности, деметилазной активности, деаминирующей активности, депурицирующей активности, интегразной активности, транспозазной активности и рекомбиназной активности.
26. Слитый полипептид CasY по любому из пп.14-21, отличающийся тем, что гетерологичный полипептид проявляет ферментативную активность, которая модифицирует целевой полипептид, связанный с целевой нуклеиновой кислотой.
27. Слитый полипептид CasY по п.26, отличающийся тем, что гетерологичный полипептид проявляет гистон-модифицирующую активность.
28. Слитый полипептид CasY по п.26 или 27, отличающийся тем, что гетерологичный полипептид проявляет один или более видов ферментативной активности, выбранных из: метилтрансферазной активности, деметилазной активности, ацетилтрансферазной активности, деацетилазной активности, киназной активности, фосфатазной активности, убиквитинлигазной активности, деубиквитинизирующей активности, аденилирующей активности, деаденилирующей активности, сумоилирующей активности, десумоилирующей активности, рибозилирующей активности, дерибозилирующей активности, миристоилирующей активности, демиристоилирующей активности, гликозилирующей активности (например, от O-GlcNAc-трансферазы) и дегликозилирующей активности.
29. Слитый полипептид CasY по п.28, отличающийся тем, что гетерологичный полипептид проявляет один или более видов ферментативной активности, выбранных из: метилтрансферазной активности, деметилазной активности, ацетилтрансферазной активности и деацетилазной активности.
30. Слитый полипептид CasY по любому из пп.14-21, отличающийся тем, что гетерологичный полипептид представляет собой полипептид эндосомального высвобождения.
31. Слитый полипептид CasY по п.30, отличающийся тем, что полипептид эндосомального высвобождения содержит аминокислотную последовательность, выбранную из  
GLFXALLXLLXSLWXLLLXA (SEQ ID NO:94) и  
GLFHALLHLLHSLWHLLLHA (SEQ ID NO:95), где каждый X независимо выбран из лизина, гистидина и аргинина.

32. Слитый полипептид CasY по любому из пп.14-21, отличающийся тем, что гетерологичный полипептид представляет собой транзитный пептид хлоропластов.

33. Слитый полипептид CasY по п.32, отличающийся тем, что транзитный пептид хлоропластов содержит аминокислотную последовательность, выбранную из

MASMISSAVTTVSRASRGQSAAMAPFGGLKSMTGFPVRKVNTDITSITSNGGRVKCMQ

VWPPIGKKKFETLSYLPPLTRDSRA (SEQ ID NO:83);

MASMISSAVTTVSRASRGQSAAMAPFGGLKSMTGFPVRKVNTDITSITSNGGRVKS

(SEQ ID NO:84);

MASSMLSSATMVASPAQATMVAPFNGLKSSAAFPATRKANNDITSITSNGGRVNCMQV

WPPIEKKKFETLSYLPDLTDSGGRVNC (SEQ ID NO:85);

MAQVSRICNGVQNPSLISNLSKSSQRKSPLSVSLKTQQHPRAYPISSSWGLKKSGMTLIGS

ELRPLKVMSSVSTAC (SEQ ID NO:86);

MAQVSRICNGVWNPSLISNLSKSSQRKSPLSVSLKTQQHPRAYPISSSWGLKKSGMTLIG

SELRPLKVMSSVSTAC (SEQ ID NO:87);

MAQINNMAQGIQTLNPNNSNFHKPQVPKSSSFLVFGSKKLKNSANSMLVLKKDSIFMQLF

CSFRISASVATAC (SEQ ID NO:88);

MAALVTSQLATSGTVLSVTDRFRPQGLRPRNPADAALGMRTVGASAAPKQSRKPH

RFDRRCLSMVV (SEQ ID NO:89);

MAALTTSQLATSATGFGIADRSAPSSLLRHGFQGLKPRSPAGGDATSLSVTTSARATPKQ

QRSVQRGSRFPVVC (SEQ ID NO:90);

MASSVLSSAAVATRSNVAQANMVAPFTGLKSAASFPVSRKQNLDTIASNGGRVQC

(SEQ ID NO:91);

MESLAATSVFAPSRVAVPAARALVRAGTVVPTRRTSSTSGTSGVKCSAAVTPQASPVISR

SAAAA (SEQ ID NO:92) и

MGAAATSMQSLKFSNRLVPPSRRLSPVPNNVTCNNLPKSAAPVRTVKCCASSWNSTING

AAATTNGASAASS (SEQ ID NO:93).

34. Слитый полипептид CasY по любому из пп.14-21, отличающийся тем, что гетерологичный полипептид представляет собой белок, который повышает или снижает транскрипцию.

35. Слитый полипептид CasY по п.34, отличающийся тем, что гетерологичный полипептид представляет собой домен репрессии транскрипции.

36. Слитый полипептид CasY по п.34, отличающийся тем, что гетерологичный полипептид представляет собой домен активации транскрипции.

37. Слитый полипептид CasY по любому из пп.14-21, отличающийся тем, что гетерологичный полипептид представляет собой белок-связывающий домен.

38. Молекула нуклеиновой кислоты, кодирующая слитый полипептид CasY по любому из пп.14-37.

39. Молекула нуклеиновой кислоты по п.38, отличающаяся тем, что нуклеотидная последовательность, кодирующая слитый полипептид CasY, функционально связана с промотором.

40. Молекула нуклеиновой кислоты по п.39, отличающаяся тем, что промотор является функциональным в эукариотической клетке.

41. Молекула нуклеиновой кислоты по п.40, отличающаяся тем, что промотор является функциональным в одной или более из: клетки растения, клетки гриба, клетки животного, клетки беспозвоночного, клетки мухи, клетки позвоночного, клетки млекопитающего, клетки примата, клетки отличного от человека примата и клетки человека.

42. Молекула нуклеиновой кислоты по любому из пп.39-41, отличающаяся тем, что промотор представляет собой один или более из: конститутивного промотора, индуцибельного промотора, специфического в отношении клеточного типа промотора и тканеспецифического промотора.

43. Молекула нуклеиновой кислоты по любому из пп.38-42, отличающаяся тем, что молекула ДНК представляет собой рекомбинантный экспрессионный вектор.

44. Молекула нуклеиновой кислоты по п.43, отличающаяся тем, что рекомбинантный экспрессионный вектор представляет собой рекомбинантный аденоассоциированный вирусный вектор, рекомбинантный ретровирусный вектор или рекомбинантный лентивирусный вектор.

45. Молекула нуклеиновой кислоты по п.39, отличающаяся тем, что промотор является функциональным в прокариотической клетке.
46. Молекула нуклеиновой кислоты по п.38, отличающаяся тем, что молекула нуклеиновой кислоты представляет собой мРНК.
47. Одна или более молекул нуклеиновой кислоты, кодирующие:
- (a) CasY-направляющую РНК и
  - (b) полипептид CasY.
48. Одна или более молекул нуклеиновой кислоты по п.47, отличающиеся тем, что полипептид CasY содержит аминокислотную последовательность, имеющую 50% или более идентичности с аминокислотной последовательностью, приведенной в SEQ ID NO: 1 или SEQ ID NO: 2 (или аминокислотной последовательностью, приведенной в любой из SEQ ID NO: 1-8).
49. Одна или более молекул нуклеиновой кислоты по п.47, отличающиеся тем, что полипептид CasY содержит аминокислотную последовательность, имеющую 85% или более идентичности с аминокислотной последовательностью, приведенной в SEQ ID NO: 1 или SEQ ID NO: 2 (или аминокислотной последовательностью, приведенной в любой из SEQ ID NO: 1-8).
50. Одна или более молекул нуклеиновой кислоты по любому из пп.47-49, отличающиеся тем, что CasY-направляющая РНК содержит нуклеотидную последовательность, имеющую 80% или более идентичности с последовательностью cr-РНК, приведенной в любой из SEQ ID NO: 11-15.
51. Одна или более молекул нуклеиновой кислоты по любому из пп.47-50, отличающиеся тем, что полипептид CasY слит с последовательностью СЯЛ.
52. Одна или более молекул нуклеиновой кислоты по любому из пп.47-51, отличающиеся тем, что указанные одна или более молекул нуклеиновой кислоты содержат нуклеотидную последовательность, кодирующую CasY-направляющую РНК, которая функционально связана с промотором.
53. Одна или более молекул нуклеиновой кислоты по любому из пп.47-52, отличающиеся тем, что указанные одна или более молекул нуклеиновой кислоты содержат нуклеотидную последовательность, кодирующую полипептид CasY, которая функционально связана с промотором.
54. Одна или более молекул нуклеиновой кислоты по п.52 или 53, отличающиеся тем, что промотор, функционально связанный с нуклеотидной последовательностью, кодирующей CasY-направляющую РНК, и/или промотор, функционально связанный с нуклеотидной последовательностью, кодирующей полипептид CasY, является функциональным в эукариотической клетке.
55. Одна или более молекул нуклеиновой кислоты по п.54, отличающиеся тем, что промотор является функциональным в одной или более из: клетки растения, клетки гриба, клетки животного, клетки беспозвоночного, клетки мухи, клетки позвоночного, клетки млекопитающего, клетки примата, клетки отличного от человека примата и клетки человека.
56. Одна или более молекул нуклеиновой кислоты по любому из пп.53-55, отличающиеся тем, что промотор представляет собой один или более из: конститутивного промотора, индуцибельного промотора, специфического в отношении клеточного типа промотора и тканеспецифического промотора.
57. Одна или более молекул нуклеиновой кислоты по любому из пп.47-56, отличающаяся тем, что одна или более молекул нуклеиновой кислоты представляют собой один или более рекомбинантных экспрессионных векторов.
58. Одна или более молекул нуклеиновой кислоты по п.57, отличающиеся тем, что один или более из рекомбинантных экспрессионных векторов выбраны из: одного или более аденоассоциированных вирусных векторов, одного или более рекомбинантных ретровирусных векторов или одного или более рекомбинантных лентивирусных векторов.
59. Одна или более молекул нуклеиновой кислоты по п.53, отличающиеся тем, что промотор является функциональным в прокариотической клетке.
60. Эукариотическая клетка, содержащая одно или более из:
- a) полипептида CasY или молекулы нуклеиновой кислоты, кодирующей полипептид CasY,
  - b) слитого полипептида CasY или молекулы нуклеиновой кислоты, кодирующей слитый полипептид CasY, и
  - c) CasY-направляющей РНК или молекулы нуклеиновой кислоты, кодирующей CasY-направляющую РНК.
61. Эукариотическая клетка по п.60, содержащая молекулу нуклеиновой кислоты, кодирующую полипептид CasY, при этом указанная молекула нуклеиновой кислоты интегрирована в геномную ДНК клетки.
62. Эукариотическая клетка по п.60 или 61, отличающаяся тем, что эукариотическая клетка представляет собой клетку растения, клетку млекопитающего, клетку насекомого, клетку паукообразного, клетку гриба, клетку птицы, клетку рептилии, клетку амфибии, клетку беспозвоночного, клетку мыши, клетку крысы, клетку примата, клетку отличного от человека примата или клетку человека.
63. Клетка, содержащая слитый полипептид CasY или молекулу нуклеиновой кислоты, кодирующую слитый полипептид CasY.
64. Клетка по п.63, отличающаяся тем, что клетка представляет собой прокариотическую клетку.

65. Клетка по п.63 или 64, содержащая молекулу нуклеиновой кислоты, кодирующую слитый полипептид CasY, при этом указанная молекула нуклеиновой кислоты интегрирована в геномную ДНК клетки.

66. Способ модификации целевой нуклеиновой кислоты, включающий в себя приведение целевой нуклеиновой кислоты в контакт с:

а) полипептидом CasY и

б) CasY-направляющей РНК, содержащей направляющую последовательность, которая гибридизируется с целевой последовательностью целевой нуклеиновой кислоты,

при этом указанное приведение в контакт приводит к модификации целевой нуклеиновой кислоты полипептидом CasY.

67. Способ по п.66, отличающийся тем, что указанная модификация представляет собой расщепление целевой нуклеиновой кислоты.

68. Способ по п.66 или 67, отличающийся тем, что целевая нуклеиновая кислота выбрана из: двухцепочечной ДНК, одноцепочечной ДНК, РНК, геномной ДНК и внехромосомной ДНК.

69. Способ по любому из пп.66-68, отличающийся тем, что указанное приведение в контакт происходит *in vitro* вне клетки.

70. Способ по любому из пп.66-68, отличающийся тем, что указанное приведение в контакт происходит внутри клетки в культуре.

71. Способ по любому из пп.66-68, отличающийся тем, что указанное приведение в контакт происходит внутри клетки *in vivo*.

72. Способ по п.70 или 71, отличающийся тем, что клетка представляет собой эукариотическую клетку.

73. Способ по п.72, отличающийся тем, что клетка выбрана из: клетки растения, клетки гриба, клетки млекопитающего, клетки рептилии, клетки насекомого, клетки птицы, клетки рыбы, клетки паразита, клетки членистоногого, клетки беспозвоночного, клетки позвоночного, клетки грызуна, клетки мыши, клетки крысы, клетки примата, клетки отличного от человека примата и клетки человека.

74. Способ по п.70 или 71, отличающийся тем, что клетка представляет собой прокариотическую клетку.

75. Способ по любому из пп.66-74, отличающийся тем, что указанное приведение в контакт приводит к редактированию генома.

76. Способ по любому из пп.66-75, отличающийся тем, что указанное приведение в контакт включает в себя внесение в клетку: (а) полипептида CasY или молекулы нуклеиновой кислоты, кодирующей полипептид CasY, и (б) CasY-направляющей РНК или молекулы нуклеиновой кислоты, кодирующей CasY-направляющую РНК.

77. Способ по п.76, отличающийся тем, что указанное приведение в контакт дополнительно включает в себя: внесение в клетку донорной матричной ДНК.

78. Способ по любому из пп.66-77, отличающийся тем, что CasY-направляющая РНК содержит нуклеотидную последовательность, имеющую 80% или более идентичности с последовательностью sg-РНК, приведенной в любой из SEQ ID NO: 11-15.

79. Способ по любому из пп.66-78, отличающийся тем, что полипептид CasY слит с последовательностью СЯЛ.

80. Способ модуляции транскрипции из целевой ДНК, модификации целевой нуклеиновой кислоты или модификации белка, связанного с целевой нуклеиновой кислотой, включающий в себя приведение целевой нуклеиновой кислоты в контакт с:

а) слитым полипептидом CasY, содержащим полипептид CasY, слитый с гетерологичным полипептидом;

б) CasY-направляющей РНК, содержащей направляющую последовательность, которая гибридизируется с целевой последовательностью целевой нуклеиновой кислоты,

81. Способ по п.80, отличающийся тем, что CasY-направляющая РНК содержит нуклеотидную последовательность, имеющую 80% или более идентичности с последовательностью sg-РНК, приведенной в любой из SEQ ID NO: 11-15.

82. Способ по п.80 или 81, отличающийся тем, что слитый полипептид CasY содержит последовательность СЯЛ.

83. Способ по любому из пп.80-82, отличающийся тем, что указанная модификация не является расщеплением целевой нуклеиновой кислоты.

84. Способ по любому из пп.80-83, отличающийся тем, что целевая нуклеиновая кислота выбрана из: двухцепочечной ДНК, одноцепочечной ДНК, РНК, геномной ДНК и внехромосомной ДНК.

85. Способ по любому из пп.80-84, отличающийся тем, что указанное приведение в контакт происходит *in vitro* вне клетки.

86. Способ по любому из пп.80-84, отличающийся тем, что указанное приведение в контакт происходит внутри клетки в культуре.

87. Способ по любому из пп.80-84, отличающийся тем, что указанное приведение в контакт происходит внутри клетки *in vivo*.

88. Способ по п.86 или 87, отличающийся тем, что клетка представляет собой эукариотическую клетку.

89. Способ по п.88, отличающийся тем, что клетка выбрана из: клетки растения, клетки гриба, клетки млекопитающего, клетки рептилии, клетки насекомого, клетки птицы, клетки рыбы, клетки паразита, клетки членистоногого, клетки беспозвоночного, клетки позвоночного, клетки грызуна, клетки мыши, клетки крысы, клетки примата, клетки отличного от человека примата и клетки человека.

90. Способ по п.86 или 87, отличающийся тем, что клетка представляет собой прокариотическую клетку.

91. Способ по любому из пп.80-90, отличающийся тем, что указанное приведение в контакт включает в себя внесение в клетку: (а) слитого полипептида CasY или молекулы нуклеиновой кислоты, кодирующей слитый полипептид CasY, и (b) CasY-направляющей РНК или молекулы нуклеиновой кислоты, кодирующей CasY-направляющую РНК.

92. Способ по любому из пп.80-91, отличающийся тем, что полипептид CasY представляет собой каталитически неактивный полипептид CasY (dCasY).

93. Способ по любому из пп.80-92, отличающийся тем, что полипептид CasY содержит одну или более мутаций в положении, соответствующем выбранным из: D672, E769 и D935 из SEQ ID NO: 1.

94. Способ по любому из пп.80-93, отличающийся тем, что гетерологичный полипептид проявляет ферментативную активность, которая модифицирует целевую ДНК.

95. Способ по п.94, отличающийся тем, что гетерологичный полипептид проявляет один или более видов ферментативной активности, выбранных из: нуклеазной активности, метилтрансферазной активности, деметилазной активности, активности ДНК-репарации, ДНК-повреждающей активности, деаминирующей активности, дисмутазной активности, алкилирующей активности, депуринизирующей активности, окислительной активности, активности образования пиримидиновых димеров, интегразной активности, транспозазной активности, рекомбиназной активности, полимеразной активности, лигазной активности, геликазной активности, фотолиазной активности и гликозилазной активности.

96. Способ по п.95, отличающийся тем, что гетерологичный полипептид проявляет один или более видов ферментативной активности, выбранных из: нуклеазной активности, метилтрансферазной активности, деметилазной активности, деаминирующей активности, депуринизирующей активности, интегразной активности, транспозазной активности и рекомбиназной активности.

97. Способ по любому из пп.80-93, отличающийся тем, что гетерологичный полипептид проявляет ферментативную активность, которая модифицирует целевой полипептид, связанный с целевой нуклеиновой кислотой.

98. Способ по п.97, отличающийся тем, что гетерологичный полипептид проявляет гистон-модифицирующую активность.

99. Способ по п.97 или 98, отличающийся тем, что гетерологичный полипептид проявляет один или более видов ферментативной активности, выбранных из: метилтрансферазной активности, деметилазной активности, ацетилтрансферазной активности, деацетилазной активности, киназной активности, фосфатазной активности, убиквитинлигазной активности, деубиквитинизирующей активности, аденилирующей активности, деаденилирующей активности, сумоилирующей активности, десумоилирующей активности, рибозилирующей активности, дерибозилирующей активности, миристоилирующей активности, демиристоилирующей активности, гликозилирующей активности (например, от O-GlcNAc-трансферазы) и дегликозилирующей активности.

100. Способ по п.99, отличающийся тем, что гетерологичный полипептид проявляет один или более видов ферментативной активности, выбранных из: метилтрансферазной активности, деметилазной активности, ацетилтрансферазной активности и деацетилазной активности.

101. Способ по любому из пп.80-93, отличающийся тем, что гетерологичный полипептид представляет собой белок, который повышает или снижает транскрипцию.

102. Способ по п.101, отличающийся тем, что гетерологичный полипептид представляет собой домен репрессии транскрипции.

103. Способ по п.101, отличающийся тем, что гетерологичный полипептид представляет собой домен активации транскрипции.

104. Способ по любому из пп.80-93, отличающийся тем, что гетерологичный полипептид представляет собой белок-связывающий домен.

105. Трансгенный, многоклеточный, отличный от человека организм, геном которого содержит трансген, содержащий нуклеотидную последовательность, кодирующую одно или более из:

- а) полипептида CasY,
- б) слитого полипептида CasY и
- в) CasY-направляющей РНК.

106. Трансгенный, многоклеточный, отличный от человека организм по п.105, отличающийся тем, что полипептид CasY содержит аминокислотную последовательность, имеющую 50% или более идентичности аминокислотной последовательности с аминокислотной последовательностью, приведенной в SEQ ID NO: 1 или SEQ ID NO: 2 (или аминокислотной последовательностью, приведенной в любой из SEQ ID NO: 1-8).

107. Трансгенный, многоклеточный, отличный от человека организм по п.105, отличающийся тем, что полипептид CasY содержит аминокислотную последовательность, имеющую 85% или более идентичности аминокислотной последовательности с аминокислотной последовательностью, приведенной в SEQ ID NO: 1 или SEQ ID NO: 2 (или аминокислотной последовательностью, приведенной в любой из SEQ ID NO: 1-8).

108. Трансгенный, многоклеточный, отличный от человека организм по любому из пп.105-107, отличающийся тем, что организм представляет собой растение, однодольное растение, двудольное растение, беспозвоночное животное, насекомое, членистоногое, паукообразное, паразита, червя, стрекающее, позвоночное животное, рыбу, рептилию, амфибию, копытное, птицу, свинью, лошадь, овцу, грызуна, мышь, крысу или отличного от человека примата.

109. Система, содержащая:

- a) полипептид CasY и CasY-направляющую РНК;
- b) полипептид CasY, CasY-направляющую РНК и донорную матричную ДНК;
- c) слитый полипептид CasY и CasY-направляющую РНК;
- d) слитый полипептид CasY, CasY-направляющую РНК и донорную матричную ДНК;
- e) мРНК, кодирующую полипептид CasY, и CasY-направляющую РНК;
- f) мРНК, кодирующую полипептид CasY; CasY-направляющую РНК и донорную матричную ДНК;
- g) мРНК, кодирующую слитый полипептид CasY, и CasY-направляющую РНК;
- h) мРНК, кодирующую слитый полипептид CasY; CasY-направляющую РНК и донорную матричную ДНК;
- i) один или более рекомбинантных экспрессионных векторов, содержащих: i) нуклеотидную последовательность, кодирующую полипептид CasY; и ii) нуклеотидную последовательность, кодирующую CasY-направляющую РНК;
- j) один или более рекомбинантных экспрессионных векторов, содержащих: i) нуклеотидную последовательность, кодирующую полипептид CasY; ii) нуклеотидную последовательность, кодирующую CasY-направляющую РНК; и iii) донорную матричную ДНК;
- к) один или более рекомбинантных экспрессионных векторов, содержащих: i) нуклеотидную последовательность, кодирующую слитый полипептид CasY; и ii) нуклеотидную последовательность, кодирующую CasY-направляющую РНК; и
- l) один или более рекомбинантных экспрессионных векторов, содержащих: i) а нуклеотидную последовательность, кодирующую слитый полипептид CasY; ii) нуклеотидную последовательность, кодирующую CasY-направляющую РНК; и донорную матричную ДНК.

110. Система CasY по п.109, отличающаяся тем, что полипептид CasY содержит аминокислотную последовательность, имеющую 50% или более идентичности аминокислотной последовательности с аминокислотной последовательностью, приведенной в SEQ ID NO: 1 или SEQ ID NO: 2 (или аминокислотной последовательностью, приведенной в любой из SEQ ID NO: 1-8).

111. Система CasY по п.109, отличающаяся тем, что полипептид CasY содержит аминокислотную последовательность, имеющую 85% или более идентичности аминокислотной последовательности с аминокислотной последовательностью, приведенной в SEQ ID NO: 1 или SEQ ID NO: 2 (или аминокислотной последовательностью, приведенной в любой из SEQ ID NO: 1-8).

112. Система CasY по любому из пп.109-111, отличающаяся тем, что донорная матричная нуклеиновая кислота имеет длину от 8 нуклеотидов до 1000 нуклеотидов.

113. Система CasY по любому из пп.109-111, отличающаяся тем, что донорная матричная нуклеиновая кислота имеет длину от 25 нуклеотидов до 500 нуклеотидов.

114. Набор, содержащий систему CasY по любому из пп.109-113.

115. Набор по п.114, отличающийся тем, что компоненты набора находятся в одном контейнере.

116. Набор по п.114, отличающийся тем, что компоненты набора находятся в разных контейнерах.

117. Стерильный контейнер, содержащий систему CasY по любому из пп.109-116.

118. Стерильный контейнер по п.117, отличающийся тем, что контейнер представляет собой шприц.

119. Имплантируемое устройство, содержащее систему CasY по любому из пп.109-116.

120. Имплантируемое устройство по п.119, отличающееся тем, что система CasY находится в матрице.

121. Имплантируемое устройство по п.119, отличающееся тем, что система CasY находится в резервуаре.

122. Способ идентификации РНК-направляемой эндонуклеазы CRISPR, включающий в себя обнаружение во множестве метагеномных нуклеотидных последовательностей нуклеотидной последовательности, кодирующей полипептид Cas1;

обнаружение матрицы CRISPR вблизи кодирующей Cas1 нуклеотидной последовательности;

клонирование из образца нуклеиновой кислоты, из которого было получено множество метагеномных нуклеотидных последовательностей, локуса CRISPR, содержащего обнаруженную матрицу CRISPR, в экспрессионный вектор для создания рекомбинантного экспрессионного вектора с локусом CRISPR;

анализ способности рекомбинантного экспрессионного вектора с локусом CRISPR расщеплять це-

левую нуклеиновую кислоту, причем локус CRISPR, обладающий способностью расщеплять целевую нуклеиновую кислоту, содержит нуклеотидную последовательность, кодирующую РНК-направляемую эндонуклеазу CRISPR;

идентификацию в локусе CRISPR открытой рамки считывания, кодирующей полипептид, который имеет менее чем 20% идентичности аминокислотной последовательности с аминокислотной последовательностью полипептида известной РНК-направляемой эндонуклеазы CRISPR.

123. Способ по п.122, отличающийся тем, что указанный анализ включает в себя внесение рекомбинантного экспрессионного вектора с локусом CRISPR и целевой нуклеиновой кислоты в клетку.

### Примеры

Следующие примеры приводятся с тем, чтобы предоставить специалистам в данной области техники полное раскрытие и описание того, как осуществлять и применять настоящее изобретение, однако они не претендуют ни на ограничение объема того, что авторы рассматривают как свое изобретение, ни на то, что нижеприведенные эксперименты представляют собой все или единственные проведенные эксперименты. Были приняты меры для того, чтобы сохранить точность в отношении используемых чисел (например, количеств, значений температуры и т.д.), однако следует учитывать некоторые экспериментальные ошибки и отклонения. Если не указано иное, части означают массовые части, молекулярная масса означает среднюю молекулярную массу, температура выражена в градусах Цельсия, а давление представляет атмосферное давление или давление, близкое к атмосферному. Могут употребляться стандартные сокращения, например, п.о., пара(ы) оснований; т. п. н., тысяча(и) пар нуклеотидов; пл, пиколитр(ы); с или с, секунда(ы); мин, минута(ы); ч или час, час(ы); ак, аминокислота(ы); нт, нуклеотид(ы); в/м, внутримышечный(о); в/б, внутрибрюшинный(о); п/к, подкожный(о) и т.п.

#### Пример 1.

Описанная в данном документе работа включает в себя анализ метагеномных образцов микробных сообществ из грунтовых вод, осадков и кислых шахтных вод. Были обнаружены новые системы CRISPR-Cas класса 2, которые не представлены в культивируемых организмах.

Фиг. 3. Домены CasY и поиск сходства, (панель а).

Схематическое представление домена для CasY, определенное по выравниванию дальних гомологов с AcCpf1 с применением HHpred. Консервативные каталитические остатки помечены красными столбиками над белками. CasY содержит расщепленный домен RuvC в С-концевой области (RuvC-RuvC-II и RuvC-III) и большой новый N-концевой домен. Ниже схемы показаны лучшие результаты совпадения на основании следующего поиска: (1) поиск BLAST по всем белкам в NCBI (база данных NR, включая модельные белки и белки внешней среды). (2) профильный поиск на основе скрытой Марковской модели (НММ) на основании моделей, построенных с использованием всех белков Cas, описанных в Makarova et al., Nat Rev Microbiol. 2015 Nov; 13(11):722-36 и Shmakov et al., Mol Cell. 2015 Nov 5; 60(3):385-97). (3) поиск дальних гомологов на основании HHpred. Совпадения отмечены цветом на основании их значимости, и также приведены диапазон результатов и E-значения. Следует отметить, что CasY имеет только локальные совпадения. 812 N-концевая аминокислота CasY имеет только одно очень незначительное частичное совпадение. Вместе эти результаты указывают на то, что CasY является новым белком Cas. (панель b) Разные CasY-содержащие каркасы локусов CRISPR были сконструированы по данным последовательности.

#### Пример 2.

Фиг. 4. Схематическая диаграмма локусов CasY и C2c3.

Белки интерференции показаны в зеленом цвете, комплекующие белки - в красном. Повторы, собранные посредством РНК-структуры, показаны справа, демонстрируя выраженную шпильку на 5' конце, что позволяет предположить наличие самопроцессинга матрицы CRISPR, обусловленного CasY.

Фиг. 5 (панели a-d). РАМ-зависимая плазмидная интерференция CasY (панель а).

Анализ истощения РАМ проводили с CasY. E. coli, содержащие CRISPR-локус CasY, трансформировали плазмидной библиотекой с 7 нуклеотидами, рандомизированными 5' или 3' относительно целевой последовательности. Проводили отбор в отношении целевой плазмиды и объединение трансформантов. Рандомизированную область амплифицировали и готовили к глубокому секвенированию. Определяли истощенные последовательности и использовали для генерации лого РАМ. (Панель b) лого РАМ, созданное для CasY. 1, демонстрировало сильную преференцию в отношении последовательностей, содержащих фланкирующую последовательность 5'-ТА-3', расположенную 5' относительно мишени. 3' РАМ обнаружен не был. (Панель c) напрямую анализировали четыре разных РАМ, чтобы проверить РАМ, определенные в анализе истощения РАМ. (Панель d) лого РАМ, созданные для CasY.2, демонстрировали преференцию в отношении последовательностей, содержащих фланкирующую последовательность 5'-YR-3' и/или 5'-TR-3' (например, 5'-DTR-3') (нижнее пороговое значение и верхнее пороговое значение, соответственно), расположенную 5' относительно мишени (где Y представляет собой T или C; R представляет собой A или G; а D представляет собой A, G или T). 3' РАМ обнаружен не был.

Фиг. 6. (Панель а) последовательности "повторов" из CasY-направляющих РНК природного происхождения (для локусов CasY Y1-Y6). (Панель b) схема направляемого РНК CasY расщепления ДНК. Белок CasY связывается с cr-РНК (CasY-направляющей РНК) в области повтора (черный, повтор; красный,

спейсер). Спаривание оснований направляющей последовательности направляющей РНК с целевой последовательностью (синий), содержащей правильный мотив, прилегающий к протоспейсеру (PAM), приводит к двухцепочечному расщеплению целевой ДНК.

Пример 3. Новые системы CRISPR-Cas, полученные из некультивируемых микробов.

Адаптивные иммунные системы CRISPR-Cas совершили революцию в области редактирования генома, обеспечив программируемые ферменты, способные осуществлять сайт-специфическое расщепление ДНК. Однако современные технологии CRISPR-Cas основаны исключительно на системах, полученных из культивируемых бактерий, оставляя неохваченными большинство ферментов из организмов, для которых не проводили выделение. В данных, приведенных в данном документе, показано применение независимой от культивации метагеномики с геномным разрешением, определение новых систем CRISPR-Cas, включая первый обнаруженный Cas9 в домене архей. Этот дивергентный фермент Cas9 был обнаружен в малоизученных наноархеях как часть активной системы CRISPR-Cas. В бактериях были обнаружены две ранее неизвестные системы, CRISPR-CasX и CRISPR-CasY, которые относятся к одним из самых совершенных систем, определенных на данный момент. Следует отметить, что все необходимые функциональные компоненты были определены с помощью метагеномики, что обеспечивает подтверждение стабильной РНК-направляемой активности ДНК-интерференции в *E. coli*. Данные в этом документе показывают, что поиск по микробным сообществам окружающей среды в комбинации с экспериментами в живых клетках дает доступ к беспрецедентному разнообразию геномов, чье содержание расширит репертуар биотехнологий на основе микроорганизмов.

Результаты.

Анализировали наборы терамасштабных метагеномных данных по микробным сообществам из грунтовых вод, осадков и кислых шахтных вод в поисках систем CRISPR-Cas класса 2, которые не представлены среди культивируемых организмов. Были определены первые белки Cas9 в домене архей и обнаружены две новые системы CRISPR-Cas, CRISPR-CasX и CRISPR-CasY, в некультивируемых бактериях (фиг. 7). Следует отметить, что как Cas9 архей, так и CasY кодировались исключительно в геномах организмов из линий дифференцировки с отсутствием известных выделенных представителей.

Первое определение Cas9 архей.

Одним из отличительных признаков CRISPR-Cas9 было ее предполагаемое наличие только в бактериальном домене. Следовательно, было неожиданно обнаружить белки Cas9, кодируемые в геномах наноархей ARMAN-1 (*Candidatus Micrarchaeum acidiphilum* ARMAN-1) и ARMAN-4 (*Candidatus Parvarchaeum acidiphilum* ARMAN-4) в наборах метагеномных данных из кислых шахтных вод (КШВ). Эти открытия расширили наличие Cas9-содержащих систем CRISPR до другого домена.

Ген cas9 ARMAN-4 был обнаружен в 16 разных образцах в одном геномном контексте, но без других смежных генов cas (несмотря на центральное расположение в нескольких контигах последовательности ДНК >25 т. п.о.) и только с одной смежной повторно-спейсерной единицей CRISPR (фиг. 13). Отсутствие типичной матрицы CRISPR и cas1, который кодирует универсальную интегразу CRISPR, указывает на отсутствие способности системы к приобретению новых спейсеров. Для спейсерной последовательности невозможно было определить мишень, но, учитывая сохранность локуса в образцах, собранных в течение нескольких лет, его функция в "однонацеленной" системе CRISPR-Cas не может быть исключена за такое время.

И, наоборот, локус CRISPR-Cas в ARMAN-1, восстановленный по 15 разным образцам, содержит крупные матрицы CRISPR, смежные с генами cas1, cas2, cas4 и cas9. Были реконструированы многочисленные альтернативные матрицы CRISPR ARMAN-1 с практически консервативным концом (вероятно, состоящим из наиболее старых спейсеров) и вариабельной областью, в которую было включено много различных спейсеров (фиг. 8а и 14). Учитывая эту гипервариабельность в содержании спейсеров, эти данные показывают, что система CRISPR-Cas9 ARMAN-1 является активной в исследуемых популяциях.

Следует отметить, что 56 из предполагаемых спейсерных мишеней (протоспейсеров) системы CRISPR-Cas9 ARMAN-1 были расположены в одном 10 т. п.о. геномном фрагменте, который, вероятно, является вирусом ARMAN-1, учитывая, что он кодирует высокую плотность коротких гипотетических белков (фиг. 8b). Действительно, криоэлектронная томографическая реконструкция часто определяет наличие вирусных частиц, присоединенных к клеткам ARMAN. Протоспейсеры ARMAN-1 также получены из предполагаемого транспозона в геноме ARMAN-2 (другой представитель наноархей) и предполагаемого мобильного элемента в геномах архей Thermoplasmatales, включая элемент I-плазмы из той же экосистемы (фиг. 15). Между клетками ARMAN и Thermoplasmatales наблюдали наличие прямых цитоплазматических "мостиков", что предполагает близкую взаимосвязь между ними. CRISPR-Cas9 ARMAN-1 может, таким образом, защищать от распространения транспозона между этими организмами, роль, аналогичная рiPHK-опосредованной защите против транспозиции в эукариотической зародышевой линии.

В активных нацеленных на ДНК системах CRISPR-Cas используются 2-4 п.о. прилегающие к протоспейсеру мотивы (PAM), расположенные за целевыми последовательностями для разделения на "своих" и "чужих". Исследование последовательностей, смежных с целевыми геномными последовательностями, действительно выявило сильную преференцию в отношении "NGG" PAM в ARMAN-1 (фиг. 8с). В

Cas9 также используются два отдельных транскрипта, РНК CRISPR (сг-РНК) и транс-активирующая РНК CRISPR (транс-РНК), для РНК-направляемого расщепления ДНК. Предполагаемая транс-РНК была идентифицирована вблизи систем CRISPR-Cas9 ARMAN-1 и ARMAN-4 (фиг. 16). Ранее предполагалось, что системы CRISPR типа II отсутствуют у архей вследствие недостатка хозяйского фактора, РНКазы II ответственного за созревание направляющего комплекса сгРНК-транс-РНК. Следует отметить, что в геноме ARMAN-1 (по оценке полно на 95%) не было идентифицировано гомологов РНКазы III и не было предсказано внутренних промоторов для матрицы CRISPR, что позволяет предположить наличие еще неизвестного механизма генерации направляющей РНК. Биохимические эксперименты для исследования активности расщепления белков Cas9 ARMAN-1 и ARMAN-4, очищенных из *E. coli* и дрожжей, и анализ *in vivo* нацеливания на *E. coli* не выявили никакой регистрируемой активности (смотрите фиг. 21 и 17).

CRISPR-CasX является новой направляемой двойной РНК системой CRISPR.

Кроме Cas9, было обнаружено и экспериментально подтверждено только три семейства эффекторных белков Cas класса 2: Cpf1, C2c1 и C2c2. Было сделано предположение, что другой ген, c2c3, который был идентифицирован только в небольших фрагментах ДНК, также кодирует это семейство белков. Новый тип системы CRISPR-Cas класса 2 был обнаружен в геномах двух бактерий, неоднократно выделенных из образцов грунтовых вод и осадков. Высокая консервативность этой системы в двух организмах, принадлежащих разным типам, дельта-протеобактериям и планктомицетам, позволяет предположить наличие недавнего переноса между типами. Эта недавно открытая система включает в себя Cas1, Cas2, Cas4 и неохарактеризованный ~980 ак белок, называемый в данном документе CasX. Матрицы CRISPR, связанные с каждым CasX, имели очень сходные повторы из 37 пар оснований, спейсеры из 33-34 пар оснований и предполагаемую транс-РНК между опероном Cas и матрицей CRISPR (фиг. 7b). Поиск BLAST выявил только слабое сходство ( $e$ -значение  $> 1 \times 10^{-4}$ ) с транспозазами, при этом сходство было ограничено конкретными областями С-конца CasX. Определение дальней гомологии и белковое моделирование позволило идентифицировать домен RuvC вблизи С-конца CasX с организацией, аналогичной обнаруживаемой в системах CRISPR-Cas типа V (фиг. 18). Оставшаяся часть белка CasX (630 N-концевых аминокислот) не демонстрировала выявляемого сходства ни с одним известным белком, что позволяет предположить, что он является новым эффектором класса 2. Комбинация транс-РНК и отдельных белков Cas1, Cas2 и Cas4 является уникальной среди систем класса V. Кроме того, CasX значительно меньше, чем любой из известных белков V: 980 ак по сравнению с типичным размером более 1200 ак для Cpf1, C2c1 и C2c3.

Далее задумались над тем, может ли CasX, несмотря на свой небольшой размер и неканоническое содержание локуса, быть способным к РНК-направляемому нацеливанию на ДНК, аналогично ферментам Cas9 и Cpf1. Чтобы исследовать эту возможность, синтезировали плазмиду, кодирующую минимальный локус CRISPR-CasX, содержащий casX, короткую матрицу из повторов и спейсеров и промежуточные некодирующие области. При экспрессии в *E. coli* этот минимальный локус блокировал трансформацию плазмидой, несущей целевую последовательность, определенную в метагеномном анализе (фиг. 9a-c, фиг. 19). Кроме того, интерференция с трансформацией возникала только тогда, когда спейсерная последовательность в мини-локусе совпадала с последовательностью протоспейсера в мишени плазмиды. Чтобы определить последовательность PAM для CasX, повторяли анализ трансформации в *E. coli*, используя плазмиду, содержащую 5' или 3' рандомизированную последовательность, смежную с целевым сайтом. Этот анализ позволил выявить строгую преференцию в отношении последовательности "TTTCN", расположенной непосредственно 5' от последовательности протоспейсера (фиг. 9d). Преференции в отношении 3' PAM не наблюдали (фиг. 19). В соответствии с этим открытием "TTCA" была последовательностью, обнаруженной выше предполагаемого протоспейсера CRISPR-CasX дельта-протеобактерии, которая была идентифицирована в образцах окружающей среды. Следует отметить, что оба локуса CRISPR-CasX имеют одинаковую последовательность PAM наряду с высокой степенью гомологии белка CasX.

Примеры направляемых одиночной РНК и двойной РНК систем существуют среди локусов CRISPR типа V. Мета-транскриптомные данные по образцам окружающей среды использовали для определения того, необходима ли транс-РНК для активности нацеливания на ДНК CasX. Этот анализ позволил выявить некодирующий РНК-транскрипт с последовательностью, комплементарной повтору CRISPR, кодируемому между открытой рамкой считывания Cas2 и матрицей CRISPR (фиг. 10a). Транскриптомное картирование дополнительно позволяет предположить, что РНК CRISPR (сгРНК) процессируется до включения 22 нт повтора и 20 нт смежного спейсера, аналогично процессингу сг-РНК, который имеет место в системах CRISPR-Cas9 (фиг. 10a). Кроме того, был идентифицирован 2-нт 3' выступ, что согласуется с опосредованным РНКазой III процессингом двойной спирали сгРНК-транс-РНК (фиг. 10b). Чтобы определить зависимость активности CasX от предполагаемой транс-РНК, эту область удаляли из вышеописанного минимального локуса CRISPR-CasX и повторяли анализ плазмидной интерференции. Удаление предполагаемой транс-РНК-кодирующей последовательности из плазмиды CasX устраняло стабильную интерференцию трансформации, наблюдаемую в ее присутствии (фиг. 10c). Вместе эти результаты позволяют утвердить CasX как новый функциональный, нацеленный на ДНК, направляемый двойной РНК фермент CRISPR.

CRISPR-CasY, система, обнаруживаемая исключительно в бактериальных линиях дифференцировки с отсутствием изолятов.

Был определен другой новый белок Cas класса 2, кодируемый в геноме определенных бактерий candidate phyla radiation (CPR). Эти бактерии, как правило, имеют небольшой размер клеток (на основании крио-ТЭМ данных и обогащения путем фильтрации), очень маленькие геномы и ограниченные возможности биосинтеза, что указывает на то, что они, наиболее вероятно, являются симбионтами. Новый ~1200 ак белок Cas, называемый в данном документе CasY, является частью минимальной системы CRISPR-Cas, которая содержит, главным образом, Cas1 и матрицу CRISPR (фиг. 11a). Большинство матриц CRISPR имеют необычайно короткие спейсеры длиной 17-19 нт, но одна система, в которой отсутствует Cas1 (CasY.5), имеет более длинные спейсеры (27-29 нт). Шесть примеров идентифицированных белков CasY не имели существенного сходства последовательности ни с одним белком из общедоступных баз данных. Чувствительный поиск с применением профильных моделей (HMM), построенных на основе опубликованных белков Cas<sup>3,4</sup>, показал, что четыре из шести белков CasY имели локальное сходство (е-значения  $4 \times 10^{-11}$  -  $3 \times 10^{-18}$ ) с C2c3 в С-концевой области, перекрывающейся с доменами RuvC, и небольшой области (~45 ак) N-конца (смотрите фиг. 18). C2c3 являются предположительными эффекторами Cas типа V, которые были идентифицированы в коротких контигах без таксономической принадлежности и не были подтверждены экспериментально. Как и CasY, C2c3 были обнаружены за матрицами с короткими спейсерами и Cas1, но без других белков Cas. Следует отметить, что два из белков CasY, идентифицированных в этом исследовании, не имели существенного сходства с C2c3, несмотря на существенное сходство последовательности (наилучшие совпадения Blast: е-значения  $6 \times 10^{-85}$ ,  $7 \times 10^{-75}$ ) с другими белками CasY.

Учитывая низкую гомологию CRISPR-CasY с любыми экспериментально подтвержденными локусами CRISPR, далее задумались, может ли эта система вызывать РНК-направляемую ДНК-интерференцию, но вследствие короткой длины спейсеров не существует достоверной информации о возможном мотиве PAM, который может требоваться для такой активности. Чтобы провести работу в этом направлении, синтезировали целый локус CRISPR-CasY. 1 с укороченной матрицей CRISPR и вносили в *E. coli* в плазмидном векторе. Затем эти клетки стимулировали в анализе трансформации, используя целевую плазмиду с последовательностью, совпадающей со спейсерной последовательностью в матрице и содержащей смежную рандомизированную 5' или 3' область для определения возможного PAM. Анализ трансформантов выявил истощение последовательностей, содержащих 5' ТА непосредственно вблизи целевой последовательности (фиг. 11b). Используя эту идентифицированную последовательность PAM, исследовали локус CasY. 1 против плазмид, содержащих один PAM. Плазмидная интерференция была продемонстрирована только в присутствии мишени, содержащей идентифицированную последовательность 5' ТА PAM (фиг. 11c). Таким образом, эти данные показывают, что CRISPR-CasY обладает активностью ДНК-интерференции.

#### Обсуждение.

Были идентифицированы новые адаптивные иммунные системы CRISPR-Cas класса 2 в геномах из некультивируемых бактерий и архей и изучены их характеристики. Эволюционный анализ Cas1 (фиг. 12a), который является универсальным для активных локусов CRISPR, позволяет предположить, что описанная в данном документе система Cas9 архей явным образом не попадает ни в один из существующих подтипов типа II. Филогенетика Cas1 (а также существование cas4) объединяет его в один кластер вместе с системами типа II-B, хотя последовательность Cas9 имеет намного больше сходства с белками типа II-C (фиг. 20). Таким образом, система архей типа II могла возникнуть как слияние систем типа II-C и типа II-B (фиг. 12b). Аналогично, филогенетический анализ Cas1 показал, что Cas1 из системы CRISPR-CasX далек от любой другой известной системы типа V. Существует предположение, что системы типа V являются результатом слияния транспозона с адаптационным модулем (Cas1-Cas2) из предковой системы типа I. Следовательно, было сделано предположение, что система CRISPR-CasX возникла после события слияния, отличного от тех, которые дали начало ранее описанным системам типа V. Удивительно, но как в системе CRISPR-CasY, так и в предполагаемой системе C2c3 отсутствует Cas2, белок, который считался важным для интеграции ДНК в локус CRISPR. Учитывая, что все системы CRISPR-Cas считаются потомками предковой системы типа которая содержала Cas1 и Cas2, системы CRISPR-CasY и C2c3 могут иметь отличного от других систем CRISPR-Cas предка, или, в альтернативном варианте, Cas2 мог быть утерян во время эволюционного развития.

Описанное в данном документе открытие Cas9 в археях и двух ранее неизвестных систем CRISPR-Cas в бактериях основано на использовании обширных баз данных по последовательностям ДНК и РНК, полученных из комплексных естественных микробных сообществ. В случае CasX и CasY геномный контекст был важным для прогнозирования функций, которые не были бы очевидными из информации по несобраным последовательностям. Кроме того, идентификация предполагаемой trasg-РНК, а также целевых вирусных последовательностей, не обнаруженных путем анализа метагеномных данных, задала направление функционального тестирования. Интересно, что некоторые из наиболее компактных локусов CRISPR-Cas, идентифицированных на сегодняшний день, были обнаружены в организмах с очень

маленькими геномами. Следствием небольшого размера генома является вероятная зависимость этих организмов от других представителей сообщества в отношении основных метаболических потребностей, и, таким образом, они в большинстве остались за пределами традиционных основанных на культивировании способов. Ограниченное число белков, необходимых для интерференции, делает эти минимальные системы в особенности ценными для разработки новых инструментов редактирования генома. Важным является то, что, как показано в данном документе, метагеномные открытия, связанные с системами CRISPR-Cas, не ограничены *in silico* наблюдениями, но могут быть включены в экспериментальные условия, в которых можно исследовать их функцию. Учитывая, что на сегодняшний день можно проводить зондирование практически всех сред, в которых существует жизнь, с помощью метагеномных способов с геномным разрешением, предполагается, что описанный в данном документе компьютерно-экспериментальный подход сильно расширит разнообразие известных систем CRISPR-Cas, обеспечивая новые технологии для биологических исследований и клинических применений.

### Способы

Метагеномика и метатранскриптомика.

Анализировали метагеномные образцы из трех разных мест: (1) Образцы кислых шахтных вод (КШВ), собранные между 2006 и 2010 гг. в Richmond Mine, Iron Mountain, Калифорния (2) образцы грунтовых вод и осадков, собранные между 2007 и 2013 гг. в центре Rifle Integrated Field Research (IFRC) вблизи реки Колорадо рядом с Рифлом, Колорадо. (3) грунтовые воды, собранные в 2009 и 2014 гг. в Crystal Geysir, холодном CO<sub>2</sub>-гейзере на плато Колорадо в Юте.

Для данных по КШВ (кислых шахтных вод) способы экстракции ДНК и секвенирование коротких ридов были описаны Deneff and Banfield (2012) и Miller et al. (2011). Для данных по Рифлу экстракция ДНК и РНК, а также секвенирование, сборка и реконструкция генома были описаны Anantharaman et al. (2016) и Brown et al. (2015). Для образцов из Crystal Geysir способы соответствовали описанным Probst et al. (2016) и Emerson et al. (2015). Вкратце, ДНК экстрагировали из образцов, используя набор для выделения ДНК PowerSoil (MoBio Laboratories Inc., Карлсбад, штат Калифорния, США). РНК экстрагировали из 0,2 мкм фильтров, собранных из шести образцов грунтовых вод из Рифла за 2011 г., как описано Brown et al. (2015). ДНК секвенировали на платформе Illumina HiSeq2000, а метатранскриптомную кДНК - на платформе 5500XL SOLiD. В случае недавно опубликованных данных по Crystal Geysir и повторно-го анализа данных КШВ (кислых шахтных вод) последовательности собирали, используя IDBA-UD. Картирование ридов ДНК и РНК (кДНК), используемое для определения покрытия секвенирования и геномной экспрессии, соответственно, проводили с помощью Bowtie2. Открытые рамки считывания (ОРС) прогнозировали для собранных каркасов, используя Prodigal. Каркасы из набора данных по Crystal Geysir сортировали на основании характера распространения дифференциального покрытия, используя комбинацию ABAWACA, ABAWACA2 (<https://github.com/CK7>) Maxbin2, и частоты появления тетрауклеотидов, используя получаемые самоорганизующиеся карты (ESOM). Геномы обрабатывали вручную, используя % содержания GC, таксономическую принадлежность и полноту генома. Ошибки в построении каркасов корректировали, используя ga2.py (<https://github.com/christophertbrown>).

Компьютерный анализ CRISPR-Cas.

Проводили сканирование собранных контигов из различных образцов в отношении известных белков Cas, используя профили скрытой Марковской модели (HMM), которые строили, используя пакет HMMer, на основании выравнивания из Makarova et al. и Shmakov et al. Матрицы CRISPR идентифицировали, используя локальную версию программного обеспечения CrisprFinder. Локусы, которые содержали как Cas1, так и матрицу CRISPR, анализировали дополнительно, если одна из десяти ОРС, смежных с геном cas1, кодировала неохарактеризованный белок больше 800 а.к., и в том же контиге не было идентифицировано известных генов cas интерференции. Эти крупные белки дополнительно анализировали как потенциальные эффекторы Cas класса 2. Потенциальные эффекторы объединяли в кластеры в белковые семейства на основании сходства последовательностей, используя MCL. Эти семейства белков расширяли путем построения HMM, представляющих каждое из этих семейств, а затем использовали их для поиска по метагеномным базам данных в отношении сходных белков Cas. Чтобы убедиться, что эти белковые семейства действительно являются новыми, используя BLAST, проводили поиск известных гомологов по избыточным (nr) и метагеномным (env\_nr) белковым базам данных NCBI а также поиск HMM по базе UniProt KnowledgeBase. Новыми белками считались только белки с отсутствием полно-размерных совпадений (>25% длины белка). Поиск дальней гомологии предполагаемых белков Cas проводили, используя HHpred из пакета HH. Имеющие высокую оценку совпадения HHpred использовали для вывода доменной архитектуры на основании сравнения с разрешенными кристаллическими структурами, а вторичную структуру прогнозировали с помощью JPred4. База данных HMM, включая новые открытые белки Cas, доступна в дополнительных данных 1.

Спейсерные последовательности определяли по собранным данным, используя CrisprFinder. CRASS использовали для обнаружения дополнительных спейсеров в коротких ридах ДНК соответствующих образцов. Затем определяли мишени спейсеров (протоспейсеры) с помощью поиска BLAST (используя "-task blastn-short") по соответствующим собранным метагеномным данным в отношении совпадений с ≤1 несовпадением в спейсерах. Совпадения, принадлежащие контигам, которые содержали ассоциирован-

ный повтор, отфильтровывали (чтобы избежать определения матриц CRISPR как протоспейсеров). Мотивы, прилегающие к протоспейсеру (PAM), определяли посредством выравнивания областей, фланкирующих протоспейсеры, и визуализировали с помощью WebLogo. Структуры РНК прогнозировали, используя mFold. Разнообразие матриц CRISPR анализировали, вручную выравнивая спейсеры, повторы и фланкирующие последовательности из собранных данных. Ручное выравнивание и визуализацию контингов проводили с помощью Geneious 9.1.

Для филогенетического анализа белков Cas1 и Cas9 новых идентифицированных систем использовали белки из Makarova et al. и Shmakov et al. Неизбыточный набор составляли путем объединения в кластеры белков с  $\geq 90\%$  идентичности, используя CD-HIT. Выравнивание проводили с помощью MAFFT, а филогенетику с максимальным правдоподобием конструировали, используя RAxML с PROTGAMMALG в качестве замещающей модели и со 100 выборками методом бутстрепа. Дерево Cas1 укореняли, используя ветвь, ведущую к каспозонам. Деревья визуализировали, используя FigTree 1.4.1 (<http://tree.bio.ed.ac.uk/software/figtree/>) и iTOL v3.

Создание гетерологичных плазмид.

Метагеномные контиги собирали в минимальные плазмиды интерференции CRISPR путем удаления белков, связанных с комплектованием для CasX, и уменьшением размера матрицы CRISPR для CasX и CasY. Минимальный локус синтезировали как Gblocks (Integrated DNA Technology) и собирали, используя Gibson Assembly.

Анализ истощения PAM.

Анализ истощения PAM проводили, как было описано ранее, с некоторой модификацией. Плазмидные библиотеки, содержащие рандомизированные последовательности PAM, собирали путем ренатурации ДНК-олигонуклеотида, содержащего мишень с 7 нт рандомизированной областью PAM, с праймером и расширяли с помощью фрагмента Кленова (NEB). Двухцепочечную ДНК расщепляли с помощью EcoRI и NcoI и лигировали в остов pUC19. Лигированную библиотеку трансформировали в DH5 $\alpha$ , собирали  $>10^8$  клеток и проводили выделение и очистку плазмид. 200 нг объединенной библиотеки трансформировали в электрокомпетентные *E. coli*, несущие локус CRISPR, или контрольную плазмиду без этого локуса. Трансформированные клетки высевали в селективную среду, содержащую карбенициллин (100 мг л<sup>-1</sup>) и хлорамфеникол (30 мг л<sup>-1</sup>) в течение 30 ч при 25°C. Экстрагировали плазмидную ДНК и амплифицировали последовательность PAM с адаптерами для секвенирования Illumina. Экстрагировали 7 нт область PAM и рассчитывали частоту появления PAM для каждой 7 нт последовательности. Последовательности PAM, истощенные выше определенного порогового значения, использовали для создания WebLogo.

Плазмидная интерференция.

Предполагаемые мишени, определенные в анализе метагеномных последовательностей или в анализе истощения PAM, клонировали в плазмиду pUC19.10 нг целевой плазмиды трансформировали в электрокомпетентные *E. coli* (NEB-стабильные), содержащие плазмиду с локусами CRISPR. Клетки восстанавливали в течение 2 ч при 25°C и соответствующее разведение высевали в селективную среду. Планшеты инкубировали при 25°C и подсчитывали колониеобразующие единицы. Все эксперименты по плазмидной интерференции проводили в трех повторностях, а электрокомпетентные клетки готовили независимо для каждого эксперимента.

Экспрессия и очистка белка ARMAN-Cas9.

Экспрессионные конструкции для Cas9 из ARMAN-1 (AR1) и ARMAN-4 (AR4) собирали из gBlocks (Integrated DNA Technologies), которые были кодон-оптимизированными в отношении *E. coli*. Собранные гены клонировали в экспрессионный вектор на основе pET в виде N-концевого His<sub>6</sub>-МБР или His<sub>6</sub> слитого белка. Экспрессионные векторы трансформировали в клетки BL21(DE3) *E. coli* и выращивали в бульоне LB при 37°C. Для белковой экспрессии клетки индуцировали во время средней фазы логарифмического роста с 0,4 мМ IPTG (изопропил- $\beta$ -D-1-тиогаллактопиранозид) и инкубировали в течение ночи при 16°C. Все последующие этапы проводили при 4°C. Клеточный осадок ресуспендировали в лизисном буфере (50 мМ Трис-НСl, рН 8, 500 мМ NaCl, 1 мМ ТКЭФ, 10 мМ имидазола) 0,5% Тритон X-100 и дополняли полной смесью ингибиторов протеаз (Roche) перед лизисом путем обработки ультразвуком. Лизат очищали путем центрифугирования при 15000 g в течение 40 мин и вводили в Superflow Ni-NTA-агарозу (Qiagen) партией. Смола интенсивно промывали промывочным буфером А (50 мМ Трис-НСl, рН 8, 500 мМ NaCl, 1 мМ ТКЭФ, 10 мМ имидазола), после этого - 5 объемами колонки промывочного буфера В (50 мМ Трис-НСl, рН 8, 1М NaCl, 1 мМ ТКЭФ, 10 мМ имидазола). Белок элюировали из смолы Ni-NTA с помощью элюирующего буфера (50 мМ Трис-НСl, рН 8, 500 мМ NaCl, 1 мМ ТКЭФ, 300 мМ имидазола). Метку His<sub>6</sub>-МБР удаляли с помощью протеазы TEV во время ночного диализа против промывочного буфера А. Расщепленный Cas9 удаляли из аффинной метки с помощью второй Ni-NTA-агарозной колонки. Проводили диализ белка в IEX-буфер А (50 мМ Трис-НСl, рН 7,5, 300 мМ NaCl, 1 мМ ТКЭФ, 5% глицерина) перед введением в 5 мл гепариновую колонку HiTrap (GE Life Sciences). Cas9 элюировали в линейном градиенте NaCl (0,3-1,5 М). Фракции объединяли и концентрировали с помощью 30 кДа концентратора-центрифуги (Thermo Fisher). При необходимости Cas9 дополнительно очищали с помощью

эксклюзионной хроматографии в 200 пг колонке Superdex (GE Life Sciences) и хранили в IEX-буфере А для последующего анализа расщепления. В случае дрожжевой экспрессии AR1-Cas9 клонировали в экспрессионный вектор Gal1/10 His6-MBP TEV Ura S. cerevisiae (№ в банке плазмид Addgene 48305). Вектор трансформировали в штамм BY4741 URA3 и выращивали культуры в среде при 30°C. При ОП600 ~0,6 индуцировали белковую экспрессию с помощью 2 мас./об.% галактозы и инкубировали в течение ночи при 16°C. Очистку белка проводили, как описано выше.

In vitro транскрипция РНК и очистка олигонуклеотидов.

Реакции in vitro транскрипции проводили, как было описано ранее,<sup>65</sup> используя синтетические ДНК-матрицы, содержащие последовательность промотора T7. Все in vitro транскрибированные направляющие РНК и целевые РНК или ДНК очищали в денатурирующем ПААГ. Двухцепочечные целевые РНК или ДНК гибридизировали в 20 мМ Трис HCl, pH 7,5 и 100 мМ NaCl путем инкубации при 95°C в течение 1 мин с последующим медленным охлаждением до комнатной температуры. Гибриды очищали с помощью нативного ПААГ.

In vitro анализ расщепления.

Очищенные олигонуклеотиды ДНК и РНК снабжали радиоактивными метками, используя киназу полинуклеотида Т4 (NEB) и [ $\gamma$ -<sup>32</sup>P] АТФ (Perkin-Elmer) в 1× РНК-буфере в течение 30 мин при 37°C. РНК термически инактивировали при 65°C в течение 20 мин и удаляли свободный АТФ из реакций меченения, используя колонки illustra Microspin G-25 (GE Life Sciences). CrРНК и tracr-РНК смешивали в эквимолярных количествах в 1× рефолдинг-буфере (50 мМ Трис HCl, pH 7,5, 300 мМ NaCl, 1 мМ ТКЭФ, 5% глицерина) и инкубировали при 70°C в течение 5 мин, а затем медленно охлаждали до комнатной температуры. Реакции дополняли до конечной концентрации металла 1 мМ и после этого нагревали при 50°C в течение 5 мин. После медленного охлаждения до комнатной температуры пересвернутые направляющие помещали на лед. Если концентрация соли не указана для буфера, Cas9 восстанавливали эквимолярным количеством направляющей в 1× расщепляющего буфера (50 мМ Трис HCl, pH 7,5, 300 мМ NaCl, 1 мМ ТКЭФ, 5% глицерина, 5 мМ двухвалентного металла) при 37°C в течение 10 мин. Реакции расщепления проводили в 1× расщепляющего буфера с 10× избытком Cas9-направляющего комплекса относительно радиоактивно меченой мишени при 37°C или при указанной температуре. Реакции гасили в эквивалентном объеме гелевого загрузочного буфера, дополненного 50 мМ ЭДТК. Продукты расщепления разделяли в 10% денатурирующем ПААГ и визуализировали путем визуализации на фосфорном покрытии.

In vivo анализ интерференции E. coli.

Анализ трансформации E. coli для AR1- и AR4-Cas9 проводили согласно ранее опубликованным схемам<sup>66</sup>. Вкратце, E. coli, трансформированным направляющими РНК, придавали электрокомпетентность. Затем клетки трансформировали 9 фмоль плазмиды, кодирующей Cas9 дикого типа или каталитически неактивный Cas9 (dCas9). Серийные разведения восстановленных клеток высевали в LB-планшеты с селективными антибиотиками. Колонии подсчитывали через 16 ч при 37°C.

Таблица 1

Данные по организмам и геномной локации, в которой была идентифицирована система CRISPR-Cas, а также информация по числу и средней длине реконструированных спейсеров и длине повторов (НД, нет данных). Спейсеры ARMAN-1 были реконструированы по 16 образцам

Таксономическая группа	Cas-эффектор	Доступ NCBI	Координаты	Длина повтора	Количество спейсеров	Средняя длина спейсера
ARMAN-1	Cas9	MOEG01000017	1827..7130	36	271	34,5
ARMAN-4	Cas9	KY040241	11779..14900	36	1	36
Дельта-протеобактерии	CasX	MGP01000094	4319..9866	37	5	33,6
Планктомицеты	CasX	MNYZ01000150	1..5586	37	7	32,3
Candidatus Katanobacterium	CasY.1	MOEN01000029	459..5716	26	14	17,1
Candidatus Vogelbacterium	CasY.2	MOEJ01000028	7322..13087	26	18	17,3
Candidatus Vogelbacterium	CasY.3	MOEK01000006	1..4657	26	12	17,3
Candidatus Parcubacterium	CasY.4	KY040242	1..5193	25	13	18,4
Candidatus Komeilbacterium	CasY.5	MOEI01000022	2802..7242	36	8	26
Candidatus Kerfeldbacterium	CasY.6	MHKD01000036	11503..15366	НД	НД	НД

Хотя настоящее изобретение было описано с отсылкой на конкретные варианты его реализации, однако специалистам в данной области техники следует понимать, что могут быть осуществлены различные изменения и выполнена замена эквивалентов без отступления от сути и объема этого изобретения. Кроме того, могут быть осуществлены многочисленные модификации с целью адаптации конкретной ситуации, материала, химической композиции, процесса, этапа или этапов процесса к цели, сути и объему настоящего изобретения. Предполагается, что все такие модификации входят в объем прилагаемой формулы изобретения.

#### ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Композиция для модификации целевой нуклеиновой кислоты, содержащая:

а) полипептид, содержащий аминокислотную последовательность, имеющую 50% или большую идентичность по аминокислотной последовательности с аминокислотной последовательностью, представленной в любой из SEQ ID NO: 1-7, и три частичных домена RuvC, которые образуют единый домен RuvC; и

б) направляющую РНК, содержащую нуклеотидную последовательность, которая комплементарна нуклеотидной последовательности в целевой нуклеиновой кислоте, где указанный полипептид способен образовывать комплекс с указанной направляющей РНК и комплекс способен связывать целевую нуклеиновую кислоту.

2. Композиция по п.1, отличающаяся тем, что полипептид содержит аминокислотную последовательность, имеющую 80% или большую идентичность по аминокислотной последовательности с аминокислотной последовательностью, представленной в любой из SEQ ID NO: 1-7.

3. Композиция по п.1 или 2, отличающаяся тем, что направляющая РНК содержит нуклеотидную последовательность, имеющую 80% или большую идентичность по нуклеотидной последовательности с последовательностью sg-РНК, приведенной в любой из SEQ ID NO: 11-15.

4. Композиция по п.1 или 2, отличающаяся тем, что полипептид слит с последовательностью сигнала ядерной локализации.

5. Композиция по любому из пп.1-4, отличающаяся тем, что композиция содержит липид.

6. Композиция по любому из пп.1-4, отличающаяся тем, что а) и б) находятся внутри липосомы.

7. Композиция по любому из пп.1-4, отличающаяся тем, что а) и б) находятся внутри частицы.
8. Композиция по любому из пп.1-7, содержащая одно или более из: буфера, ингибитора нуклеаз и ингибитора протеаз.
9. Композиция по любому из пп.1-8, отличающаяся тем, что полипептид содержит аминокислотную последовательность, имеющую 90% или большую идентичность с аминокислотной последовательностью, представленной в любой из SEQ ID NO: 1-7.
10. Композиция по любому из пп.1-9, отличающаяся тем, что полипептид представляет собой никазу, которая может расщеплять только одну цепь двухцепочечной молекулы целевой нуклеиновой кислоты.
11. Композиция по любому из пп.1-9, отличающаяся тем, что полипептид представляет собой каталитически неактивный полипептид.
12. Композиция по п.10 или 11, отличающаяся тем, что полипептид содержит одну или более мутаций в положении, соответствующем выбранному из: D672, E769 и D935 из SEQ ID NO: 1.
13. Композиция по любому из пп.1-12, дополнительно содержащая донорную матричную ДНК.
14. Слитый полипептид, способный формировать комплекс с направляющей РНК, содержащейся в композиции по п.1, который содержит:
  - а) полипептид, который содержит аминокислотную последовательность, имеющую 50% или большую идентичность по аминокислотной последовательности с аминокислотной последовательностью, представленной в любой из SEQ ID NO: 1-7, и три частичных домена RuvC, которые образуют единый домен RuvC; и
  - б) гетерологичный полипептид.
15. Слитый полипептид по п.14, отличающийся тем, что полипептид содержит аминокислотную последовательность, имеющую 85% или большую идентичность по аминокислотной последовательности с аминокислотной последовательностью, представленной в любой из SEQ ID NO: 1-7.
16. Слитый полипептид по п.14, отличающийся тем, что полипептид содержит аминокислотную последовательность, имеющую 90% или большую идентичность по аминокислотной последовательности с аминокислотной последовательностью, представленной в любой из SEQ ID NO: 1-7.
17. Слитый полипептид по любому из пп.14-16, отличающийся тем, что полипептид представляет собой никазу, которая может расщеплять только одну цепь двухцепочечной молекулы целевой нуклеиновой кислоты.
18. Слитый полипептид по любому из пп.14-17, отличающийся тем, что полипептид представляет собой каталитически неактивный полипептид.
19. Слитый полипептид по п.17 или 18, отличающийся тем, что полипептид содержит одну или более мутаций в положении, соответствующем выбранному из: D672, E769 и D935 из SEQ ID NO: 1.
20. Слитый полипептид по любому из пп.14-19, отличающийся тем, что гетерологичный полипептид слит с N-концом и/или C-концом полипептида.
21. Слитый полипептид по любому из пп.14-20, содержащий сигнал ядерной локализации.
22. Слитый полипептид по любому из пп.14-21, отличающийся тем, что гетерологичный полипептид представляет собой нацеливающий полипептид, который обеспечивает связывание с компонентом клеточной поверхности на целевой клетке или целевом типе клеток.
23. Слитый полипептид по любому из пп.14-21, отличающийся тем, что гетерологичный полипептид проявляет ферментативную активность, которая модифицирует целевую ДНК.
24. Слитый полипептид по п.23, отличающийся тем, что гетерологичный полипептид проявляет один или более видов ферментативной активности, выбранных из: нуклеазной активности, метилтрансферазной активности, деметилазной активности, активности ДНК-репарации, ДНК-повреждающей активности, деаминирующей активности, дисмутазной активности, алкилирующей активности, депуринизирующей активности, окислительной активности, активности образования пиримидиновых димеров, интегразной активности, транспозазной активности, рекомбиназной активности, полимеразной активности, лигазной активности, геликазной активности, фотолиазной активности и гликозилазной активности.
25. Слитый полипептид по п.24, отличающийся тем, что гетерологичный полипептид проявляет один или более видов ферментативной активности, выбранных из: нуклеазной активности, метилтрансферазной активности, деметилазной активности, деаминирующей активности, депуринизирующей активности, интегразной активности, транспозазной активности и рекомбиназной активности.
26. Слитый полипептид по любому из пп.14-21, отличающийся тем, что гетерологичный полипептид проявляет ферментативную активность, которая модифицирует целевой полипептид, связанный с целевой нуклеиновой кислотой.
27. Слитый полипептид по п.26, отличающийся тем, что гетерологичный полипептид проявляет гистон-модифицирующую активность.
28. Слитый полипептид по п.26 или 27, отличающийся тем, что гетерологичный полипептид проявляет один или более видов ферментативной активности, выбранных из: метилтрансферазной активности, деметилазной активности, ацетилтрансферазной активности, деацетилазной активности, киназной активности, фосфатазной активности, убиквитинлигазной активности, деубиквитинизирующей активности,

аденилирующей активности, деаденилирующей активности, сумоилирующей активности, десумоилирующей активности, рибозилирующей активности, дерибозилирующей активности, миристоилирующей активности, демиростоилирующей активности, гликозилирующей активности (например, O-GlcNAc-трансферазы) и дегликозилирующей активности.

29. Слитый полипептид по п.28, отличающийся тем, что гетерологичный полипептид проявляет один или более видов ферментативной активности, выбранных из: метилтрансферазной активности, деметилазной активности, ацетилтрансферазной активности и деацетилазной активности.

30. Слитый полипептид по любому из пп.14-21, отличающийся тем, что гетерологичный полипептид представляет собой полипептид эндосомального высвобождения.

31. Слитый полипептид по п.30, отличающийся тем, что полипептид эндосомального высвобождения содержит аминокислотную последовательность, выбранную из: GLFXALLXLLXSLWXLLLXA (SEQ ID NO: 94) и GLFHALLHLLHSLWHLLLHA (SEQ ID NO: 95), где каждый X независимо выбран из лизина, гистидина и аргинина.

32. Слитый полипептид по любому из пп.14-21, отличающийся тем, что гетерологичный полипептид представляет собой транзитный пептид хлоропластов.

33. Слитый полипептид по п.32, отличающийся тем, что транзитный пептид хлоропластов содержит аминокислотную последовательность, выбранную из

MASMISSSAVTTVSRASRGQSAAMAPFGGLKSMTGFPVRKVNTDITSITSNGGRVKCMQ  
VWPPIGKKKFETLSYLPPLTRDSRA (SEQ ID NO:83);

MASMISSSAVTTVSRASRGQSAAMAPFGGLKSMTGFPVRKVNTDITSITSNGGRVKS  
(SEQ ID NO:84);

MASSMLSSATMVASPAQATMVAPFNGLKSSAAFPATRKANNDITSITSNGGRVNCMQV  
WPPIEKKKFETLSYLPDLTDSGGRVNC (SEQ ID NO:85);

MAQVSRICNGVQNPSLISNLSKSSQRKSPLSVSLKTQQHPRAYPISSSWGLKKSGMTLIGS  
ELRPLKVMSSVSTAC (SEQ ID NO:86);

MAQVSRICNGVWNPSLISNLSKSSQRKSPLSVSLKTQQHPRAYPISSSWGLKKSGMTLIG  
SELRPLKVMSSVSTAC (SEQ ID NO:87);

MAQINNMAQGIQTLNPNNSNFHKPQVPKSSSFLVFGSKKLKNSANSMLVLKKDSIFMQLF  
CSFRISASVATAC (SEQ ID NO:88);

MAALVTSQLATSGTVLSVTDRFRPGFQGLRPRNPADAALGMRTVGASAAPKQSRKPH  
RFDRRCLSMVV (SEQ ID NO:89);

MAALTTSQLATSATGFGIADRSAPSSLLRHGFQGLKPRSPAGGDATSLSVTTSARATPKQ  
QRSVQRGSRRFPSVVVC (SEQ ID NO:90);

MASSVLSSAAVATRSNVAQANMVAPFTGLKSAASFPVSRKQNLDTIASNGGRVQC  
(SEQ ID NO:91);

MESLAATSVFAPSRVAVPAARALVRAGTVVPTRRTSSTSGTSGVKCSAAVTPQASPVISR  
SAAAA (SEQ ID NO:92) и

MGAAATSMQSLKFSNRLVPPSRRLSPVPNNVTCNNLPKSAAPVRTVKCCASSWNSTING  
AAATTNGASAASS (SEQ ID NO:93).

34. Слитый полипептид по любому из пп.14-21, отличающийся тем, что гетерологичный полипептид представляет собой белок, который повышает или снижает транскрипцию.

35. Слитый полипептид по п.34, отличающийся тем, что гетерологичный полипептид представляет собой домен репрессии транскрипции.

36. Слитый полипептид по п.34, отличающийся тем, что гетерологичный полипептид представляет собой домен активации транскрипции.

37. Слитый полипептид по любому из пп.14-21, отличающийся тем, что гетерологичный полипептид представляет собой белок-связывающий домен.

38. Нуклеиновая кислота, содержащая нуклеотидную последовательность, кодирующую слитый полипептид по любому из пп.14-37.

39. Нуклеиновая кислота по п.38, отличающаяся тем, что указанная нуклеотидная последовательность функционально связана с промотором.

40. Нуклеиновая кислота по п.39, отличающаяся тем, что промотор является функциональным в эукариотической клетке.

41. Нуклеиновая кислота по п.40, отличающаяся тем, что промотор является функциональным в одной или более из: клетки растения, клетки гриба, клетки животного, клетки беспозвоночного, клетки мухи, клетки позвоночного, клетки млекопитающего, клетки примата, клетки отличного от человека примата и клетки человека.

42. Нуклеиновая кислота по любому из пп.39-41, отличающаяся тем, что промотор представляет собой один или более из: конститутивного промотора, индуцибельного промотора, специфического в отношении клеточного типа промотора и тканеспецифического промотора.

43. Нуклеиновая кислота по любому из пп.38-42, отличающаяся тем, что молекула ДНК представляет собой рекомбинантный экспрессионный вектор.

44. Нуклеиновая кислота по п.43, отличающаяся тем, что рекомбинантный экспрессионный вектор представляет собой рекомбинантный аденоассоциированный вирусный вектор, рекомбинантный ретровирусный вектор или рекомбинантный лентивирусный вектор.

45. Нуклеиновая кислота по п.39, отличающаяся тем, что промотор является функциональным в прокариотической клетке.

46. Нуклеиновая кислота по п.38, отличающаяся тем, что представляет собой мРНК.

47. Нуклеиновая кислота, содержащая:

а) нуклеотидную последовательность, кодирующую полипептид, где полипептид содержит аминокислотную последовательность, имеющую 50% или большую идентичность по аминокислотной последовательности с аминокислотной последовательностью, представленной в любой из SEQ ID NO: 1-7, и три частичных домена RuvC, которые образуют единый домен RuvC; и

б) нуклеотидную последовательность, кодирующую направляющую РНК, причем направляющая РНК содержит нуклеотидную последовательность, комплементарную нуклеотидной последовательности в целевой нуклеиновой кислоте, при этом полипептид способен образовывать комплекс с направляющей РНК, и где комплекс способен связывать целевую нуклеиновую кислоту.

48. Нуклеиновая кислота по п.47, отличающаяся тем, что полипептид содержит аминокислотную последовательность, имеющую 85% или большую идентичность по аминокислотной последовательности с аминокислотной последовательностью, представленной в любой из SEQ ID NO: 1-7.

49. Нуклеиновая кислота по п.47, отличающаяся тем, что полипептид содержит аминокислотную последовательность, имеющую 90% или большую идентичность по аминокислотной последовательности с аминокислотной последовательностью, представленной в любой из SEQ ID NO: 1-7.

50. Нуклеиновая кислота по любому из пп.47-49, отличающаяся тем, что направляющая РНК содержит нуклеотидную последовательность, имеющую 80% или большую идентичность по аминокислотной последовательности с последовательностью sg-РНК, приведенной в любой из SEQ ID NO: 11-15.

51. Нуклеиновая кислота по любому из пп.47-50, отличающаяся тем, что полипептид слит с последовательностью сигнала ядерной локализации.

52. Нуклеиновая кислота по любому из пп.47-51, отличающаяся тем, что нуклеотидная последовательность, кодирующая направляющую РНК, функционально связана с промотором.

53. Нуклеиновая кислота по любому из пп.47-52, отличающаяся тем, что нуклеотидная последовательность, кодирующая полипептид, функционально связана с промотором.

54. Нуклеиновая кислота по п.52 или 53, отличающаяся тем, что промотор, функционально связанный с нуклеотидной последовательностью, кодирующей направляющую РНК, и/или промотор, функционально связанный с нуклеотидной последовательностью, кодирующей полипептид, является функциональным в эукариотической клетке.

55. Нуклеиновая кислота по п.54, отличающаяся тем, что промотор является функциональным в одной или более из: клетки растения, клетки гриба, клетки животного, клетки беспозвоночного, клетки мухи, клетки позвоночного, клетки млекопитающего, клетки примата, клетки отличного от человека примата и клетки человека.

56. Нуклеиновая кислота по любому из пп.53-55, отличающаяся тем, что промотор представляет собой один или более из: конститутивного промотора, индуцибельного промотора, специфического в отношении клеточного типа промотора и тканеспецифического промотора.

57. Нуклеиновая кислота по любому из пп.47-56, отличающаяся тем, что представляет собой рекомбинантный экспрессионный вектор.

58. Нуклеиновая кислота по п.57, отличающаяся тем, что рекомбинантный экспрессионный вектор выбран из: одного или более аденоассоциированных вирусных векторов, одного или более рекомбинантных ретровирусных векторов или одного или более рекомбинантных лентивирусных векторов.

59. Нуклеиновая кислота по п.53, отличающаяся тем, что промотор является функциональным в прокариотической клетке.

60. Эукариотическая клетка для модификации целевой нуклеиновой кислоты, содержащая одно или более из:  
 а) полипептида или молекулы нуклеиновой кислоты, кодирующей полипептид, где полипептид содержит аминокислотную последовательность, имеющую 50% или большую идентичность по аминокислотной последовательности с аминокислотной последовательностью, представленной в любой из SEQ ID NO: 1-7, и три частичных домена RuvC, которые образуют единый домен RuvC;

б) слитого полипептида, содержащего: i) полипептид, содержащий аминокислотную последовательность, имеющую 50% или большую идентичность по аминокислотной последовательности с аминокислотной последовательностью, представленной в любой из SEQ ID NO: 1-7, и три частичных домена RuvC, которые образуют единый домен RuvC; и ii) партнера по слиянию, или молекулы нуклеиновой кислоты, кодирующей слитый полипептид, и

с) направляющей РНК или молекулы нуклеиновой кислоты, кодирующей направляющую РНК, где направляющая РНК содержит нуклеотидную последовательность, которая комплементарна нуклеотидной последовательности в целевой нуклеиновой кислоте, где полипептид способен образовывать комплекс с направляющей РНК и комплекс способен связывать целевую нуклеиновую кислоту.

61. Эукариотическая клетка по п.60, содержащая молекулу нуклеиновой кислоты, кодирующую полипептид, при этом указанная молекула нуклеиновой кислоты интегрирована в геномную ДНК клетки.

62. Эукариотическая клетка по п.60 или 61, отличающаяся тем, что эукариотическая клетка представляет собой клетку растения, клетку млекопитающего, клетку насекомого, клетку паукообразного, клетку гриба, клетку птицы, клетку рептилии, клетку амфибии, клетку беспозвоночного, клетку мыши, клетку крысы, клетку примата, клетку отличного от человека примата или клетку человека.

63. Клетка для модификации целевой нуклеиновой кислоты, содержащая слитый полипептид, содержащий: i) полипептид, содержащий аминокислотную последовательность, имеющую 50% или большую идентичность по аминокислотной последовательности с аминокислотной последовательностью, представленной в любой из SEQ ID NO: 1-7, и три частичных домена RuvC, которые образуют единый домен RuvC; и ii) партнера по слиянию, или молекулу нуклеиновой кислоты, кодирующую слитый полипептид.

64. Клетка по п.63, отличающаяся тем, что клетка представляет собой прокариотическую клетку.

65. Клетка по п.63 или 64, содержащая молекулу нуклеиновой кислоты, кодирующую слитый полипептид, при этом указанная молекула нуклеиновой кислоты интегрирована в геномную ДНК клетки.

66. Способ модификации целевой нуклеиновой кислоты, включающий в себя приведение целевой нуклеиновой кислоты в контакт с:

а) полипептидом, содержащим аминокислотную последовательность, имеющую 50% или большую идентичность по аминокислотной последовательности с аминокислотной последовательностью, представленной в любой из SEQ ID NO: 1-7, и три частичных домена RuvC, которые образуют единый домен RuvC; и

б) направляющей РНК, содержащей направляющую последовательность, которая гибридизуется с целевой последовательностью целевой нуклеиновой кислоты,

где полипептид способен образовывать комплекс с направляющей РНК и комплекс способен связывать целевую нуклеиновую кислоту,

при этом указанное приведение в контакт приводит к модификации целевой нуклеиновой кислоты полипептидом.

67. Способ по п.66, отличающийся тем, что указанная модификация представляет собой расщепление целевой нуклеиновой кислоты.

68. Способ по п.66 или 67, отличающийся тем, что целевая нуклеиновая кислота выбрана из: двухцепочечной ДНК, одноцепочечной ДНК, РНК, геномной ДНК и внехромосомной ДНК.

69. Способ по любому из пп.66-68, отличающийся тем, что указанное приведение в контакт происходит *in vitro* вне клетки.

70. Способ по любому из пп.66-68, отличающийся тем, что указанное приведение в контакт происходит внутри клетки в культуре.

71. Способ по любому из пп.66-68, отличающийся тем, что указанное приведение в контакт происходит внутри клетки *in vivo*.

72. Способ по п.70 или 71, отличающийся тем, что клетка представляет собой эукариотическую клетку.

73. Способ по п.72, отличающийся тем, что клетка выбрана из: клетки растения, клетки гриба, клетки млекопитающего, клетки рептилии, клетки насекомого, клетки птицы, клетки рыбы, клетки паразита, клетки членистоногого, клетки беспозвоночного, клетки позвоночного, клетки грызуна, клетки мыши, клетки крысы, клетки примата, клетки отличного от человека примата и клетки человека.

74. Способ по п.70 или 71, отличающийся тем, что клетка представляет собой прокариотическую клетку.

75. Способ по любому из пп.66-74, отличающийся тем, что указанное приведение в контакт приводит к редактированию генома.

76. Способ по любому из пп.66-75, отличающийся тем, что указанное приведение в контакт включает в себя внесение в клетку: (а) полипептида или молекулы нуклеиновой кислоты, кодирующей поли-

пептид, и (b) направляющей РНК или молекулы нуклеиновой кислоты, кодирующей направляющую РНК.

77. Способ по п.76, отличающийся тем, что указанное приведение в контакт дополнительно включает в себя внесение в клетку донорной матричной ДНК.

78. Способ по любому из пп.66-77, отличающийся тем, что направляющая РНК содержит нуклеотидную последовательность, имеющую 80% или большую идентичность с последовательностью cr-РНК, приведенной в любой из SEQ ID NO: 11-15.

79. Способ по любому из пп.66-78, отличающийся тем, что полипептид слит с последовательностью сигнала ядерной локализации.

80. Способ модуляции транскрипции из целевой ДНК, модификации целевой нуклеиновой кислоты или модификации белка, связанного с целевой нуклеиновой кислотой, включающий в себя приведение целевой нуклеиновой кислоты в контакт со:

а) слитым полипептидом, содержащим полипептид, слитый с гетерологичным полипептидом, где полипептид содержит аминокислотную последовательность, имеющую 50% или большую идентичность по аминокислотной последовательности с аминокислотной последовательностью, представленной в любой из SEQ ID NO: 1-7, и три частичных домена RuvC, которые образуют единый домен RuvC; и

б) направляющей РНК, содержащей направляющую последовательность, которая гибридизуется с целевой последовательностью целевой нуклеиновой кислоты, где полипептид способен образовывать комплекс с направляющей РНК и где комплекс способен связывать целевую нуклеиновую кислоту.

81. Способ по п.80, отличающийся тем, что направляющая РНК содержит нуклеотидную последовательность, имеющую 80% или большую идентичность с последовательностью cr-РНК, приведенной в любой из SEQ ID NO: 11-15.

82. Способ по п.80 или 81, отличающийся тем, что слитый полипептид содержит последовательность сигнала ядерной локализации.

83. Способ по любому из пп.80-82, отличающийся тем, что указанная модификация не является расщеплением целевой нуклеиновой кислоты.

84. Способ по любому из пп.80-83, отличающийся тем, что целевая нуклеиновая кислота выбрана из: двухцепочечной ДНК, одноцепочечной ДНК, РНК, геномной ДНК и внехромосомной ДНК.

85. Способ по любому из пп.80-84, отличающийся тем, что указанное приведение в контакт происходит *in vitro* вне клетки.

86. Способ по любому из пп.80-84, отличающийся тем, что указанное приведение в контакт происходит внутри клетки в культуре.

87. Способ по любому из пп.80-84, отличающийся тем, что указанное приведение в контакт происходит внутри клетки *in vivo*.

88. Способ по п.86 или 87, отличающийся тем, что клетка представляет собой эукариотическую клетку.

89. Способ по п.88, отличающийся тем, что клетка выбрана из: клетки растения, клетки гриба, клетки млекопитающего, клетки рептилии, клетки насекомого, клетки птицы, клетки рыбы, клетки паразита, клетки членистоногого, клетки беспозвоночного, клетки позвоночного, клетки грызуна, клетки мыши, клетки крысы, клетки примата, клетки отличного от человека примата и клетки человека.

90. Способ по п.86 или 87, отличающийся тем, что клетка представляет собой прокариотическую клетку.

91. Способ по любому из пп.80-90, отличающийся тем, что указанное приведение в контакт включает в себя внесение в клетку: (а) слитого полипептида или молекулы нуклеиновой кислоты, кодирующей слитый полипептид, и (b) направляющей РНК или молекулы нуклеиновой кислоты, кодирующей направляющую РНК.

92. Способ по любому из пп.80-91, отличающийся тем, что полипептид представляет собой каталитически неактивный полипептид.

93. Способ по любому из пп.80-92, отличающийся тем, что полипептид содержит одну или более мутаций в положении, соответствующем выбранному из: D672, E769 и D935 из SEQ ID NO: 1.

94. Способ по любому из пп.80-93, отличающийся тем, что гетерологичный полипептид проявляет ферментативную активность, которая модифицирует целевую ДНК.

95. Способ по п.94, отличающийся тем, что гетерологичный полипептид проявляет один или более видов ферментативной активности, выбранных из: нуклеазной активности, метилтрансферазной активности, деметилазной активности, активности ДНК-репарации, ДНК-повреждающей активности, деаминирующей активности, дисмутазной активности, алкилирующей активности, депуринизирующей активности, окислительной активности, активности образования пиримидиновых димеров, интегразной активности, транспозазной активности, рекомбиназной активности, полимеразной активности, лигазной активности, геликазной активности, фотолиазной активности и гликозилазной активности.

96. Способ по п.95, отличающийся тем, что гетерологичный полипептид проявляет один или более видов ферментативной активности, выбранных из: нуклеазной активности, метилтрансферазной активности, деметилазной активности, деаминирующей активности, депуринизирующей активности, интегразной активности, транспозазной активности и рекомбиназной активности.

97. Способ по любому из пп.80-93, отличающийся тем, что гетерологичный полипептид проявляет ферментативную активность, которая модифицирует целевой полипептид, связанный с целевой нуклеиновой кислотой.

98. Способ по п.97, отличающийся тем, что гетерологичный полипептид проявляет гистон-модифицирующую активность.

99. Способ по п.97 или 98, отличающийся тем, что гетерологичный полипептид проявляет один или более видов ферментативной активности, выбранных из: метилтрансферазной активности, деметилазной активности, ацетилтрансферазной активности, деацетилазной активности, киназной активности, фосфатазной активности, убиквитинлигазной активности, деубиквитинизирующей активности, аденилирующей активности, деаденилирующей активности, сумоилирующей активности, десумоилирующей активности, рибозилирующей активности, дерибозилирующей активности, миристоилирующей активности, демиристоилирующей активности, гликозилирующей активности (например, O-GlcNAc-трансферазы) и дегликозилирующей активности.

100. Способ по п.99, отличающийся тем, что гетерологичный полипептид проявляет один или более видов ферментативной активности, выбранных из: метилтрансферазной активности, деметилазной активности, ацетилтрансферазной активности и деацетилазной активности.

101. Способ по любому из пп.80-93, отличающийся тем, что гетерологичный полипептид представляет собой белок, который повышает или снижает транскрипцию.

102. Способ по п.101, отличающийся тем, что гетерологичный полипептид представляет собой домен репрессора транскрипции.

103. Способ по п.101, отличающийся тем, что гетерологичный полипептид представляет собой домен активации транскрипции.

104. Способ по любому из пп.80-93, отличающийся тем, что гетерологичный полипептид представляет собой белок-связывающий домен.

105. Трансгенный, многоклеточный, отличный от человека организм для получения полипептида, содержащего аминокислотную последовательность, имеющую 50% или большую идентичность по аминокислотной последовательности с аминокислотной последовательностью, представленной в любой из SEQ ID NO: 1-7, и три частичных домена RuvC, которые образуют единый домен RuvC,

где геном указанного отличного от человека организма содержит трансген, содержащий нуклеотидную последовательность, кодирующую указанный полипептид.

106. Трансгенный, многоклеточный, отличный от человека организм для получения слитого полипептида, содержащего: i) полипептид, содержащий аминокислотную последовательность, имеющую 50% или большую идентичность по аминокислотной последовательности с аминокислотной последовательностью, представленной в любой из SEQ ID NO: 1-7, и три частичных домена RuvC, которые образуют единый домен RuvC; и ii) партнера по слиянию,

где геном указанного отличного от человека организма содержит трансген, содержащий нуклеотидную последовательность, кодирующую указанный слитый полипептид.

107. Трансгенный многоклеточный, отличный от человека организм по п.105 или 106, отличающийся тем, что геном организма содержит нуклеотидную последовательность, кодирующую направляющую РНК, причем направляющая РНК содержит нуклеотидную последовательность, комплементарную нуклеотидной последовательности в целевой нуклеиновой кислоте, при этом полипептид способен образовывать комплекс с направляющей РНК, и при этом комплекс способен связывать целевую нуклеиновую кислоту.

108. Трансгенный, многоклеточный, отличный от человека организм по п.105 или 106, отличающийся тем, что полипептид содержит аминокислотную последовательность, имеющую 85% или большую идентичность по аминокислотной последовательности с аминокислотной последовательностью, представленной в любой из SEQ ID NO: 1-7, и три частичных домена RuvC, которые образуют единый домен RuvC.

109. Трансгенный, многоклеточный, отличный от человека организм по п.105 или 106, отличающийся тем, что полипептид содержит аминокислотную последовательность, имеющую 90% или большую идентичность по аминокислотной последовательности с аминокислотной последовательностью, представленной в любой из SEQ ID NO: 1-7, и три частичных домена RuvC, которые образуют единый домен RuvC.

110. Трансгенный, многоклеточный, отличный от человека организм по любому из пп.105-109, отличающийся тем, что организм представляет собой растение, однодольное растение, двудольное растение, беспозвоночное животное, насекомое, членистоногое, паукообразное, паразита, червя, стрекающее, позвоночное животное, рыбу, рептилию, амфибию, копытное, птицу, свинью, лошадь, овцу, грызуна, мышь, крысу или отличного от человека примата.

111. Система для модулирования транскрипции с целевой ДНК, модификации целевой нуклеиновой кислоты или модификации белка, ассоциированного с целевой нуклеиновой кислотой, содержащая:

- a) полипептид и направляющую РНК;
- b) полипептид, направляющую РНК и донорную матричную ДНК;

- с) слитый полипептид и направляющую РНК;
- д) слитый полипептид, направляющую РНК и донорную матричную ДНК;
- е) мРНК, кодирующую полипептид, и направляющую РНК;
- ф) мРНК, кодирующую полипептид, направляющую РНК и донорную матричную ДНК;
- г) мРНК, кодирующую слитый полипептид, и направляющую РНК;
- h) мРНК, кодирующую слитый полипептид, направляющую РНК и донорную матричную ДНК;
- и) один или более рекомбинантных экспрессионных векторов, содержащих: i) нуклеотидную последовательность, кодирующую полипептид; и ii) нуклеотидную последовательность, кодирующую направляющую РНК;
- j) один или более рекомбинантных экспрессионных векторов, содержащих: i) нуклеотидную последовательность, кодирующую полипептид; ii) нуклеотидную последовательность, кодирующую направляющую РНК; и iii) донорную матричную ДНК;
- к) один или более рекомбинантных экспрессионных векторов, содержащих: i) нуклеотидную последовательность, кодирующую слитый полипептид; и ii) нуклеотидную последовательность, кодирующую направляющую РНК; и
- l) один или более рекомбинантных экспрессионных векторов, содержащих: i) нуклеотидную последовательность, кодирующую слитый полипептид; ii) нуклеотидную последовательность, кодирующую направляющую РНК, и донорную матричную ДНК,

где полипептид содержит аминокислотную последовательность, имеющую 50% или большую идентичность по аминокислотной последовательности с аминокислотной последовательностью, представленной в любой из SEQ ID NO: 1-7, и три частичных домена RuvC, которые образуют единый домен RuvC;

где слитый полипептид включает: i) полипептид, содержащий аминокислотную последовательность, имеющую 50% или большую идентичность по аминокислотной последовательности с аминокислотной последовательностью, представленной в любой из SEQ ID NO: 1-7, и три частичных домена RuvC, которые образуют единый домен RuvC; и ii) партнера по слиянию, и

где направляющая РНК содержит нуклеотидную последовательность, которая комплементарна нуклеотидной последовательности в целевой нуклеиновой кислоте, причем полипептид способен образовывать комплекс с направляющей РНК, и при этом комплекс способен связывать целевую нуклеиновую кислоту.

112. Система по п.111, отличающаяся тем, что полипептид содержит аминокислотную последовательность, имеющую 85% или большую идентичность по аминокислотной последовательности с аминокислотной последовательностью, представленной в любой из SEQ ID NO: 1-7, и три частичных домена RuvC, которые образуют единый домен RuvC.

113. Система по п.111, отличающаяся тем, что полипептид содержит аминокислотную последовательность, имеющую 95% или большую идентичность по аминокислотной последовательности с аминокислотной последовательностью, представленной в любой из SEQ ID NO: 1-7, и три частичных домена RuvC, которые образуют единый домен RuvC.

114. Система по любому из пп.111-113, отличающаяся тем, что донорная матричная нуклеиновая кислота имеет длину от 8 до 1000 нуклеотидов.

115. Система по любому из пп.111-113, отличающаяся тем, что донорная матричная нуклеиновая кислота имеет длину от 25 до 500 нуклеотидов.

116. Набор, содержащий систему по любому из пп.111-115, для модулирования транскрипции с целевой ДНК, модификации целевой нуклеиновой кислоты или модификации белка, ассоциированного с целевой нуклеиновой кислотой.

117. Набор по п.116, отличающийся тем, что компоненты набора находятся в одном контейнере.

118. Набор по п.116, отличающийся тем, что компоненты набора находятся в разных контейнерах.

119. Стерильный контейнер для включения в набор по п.116, содержащий систему по любому из пп.111-115.

120. Стерильный контейнер по п.119, отличающийся тем, что контейнер представляет собой шприц.

121. Имплантируемое устройство для модуляции транскрипции с целевой ДНК, модификации целевой ДНК или модификации белка, ассоциированного с целевой ДНК, содержащее систему по любому из пп.111-115.

122. Имплантируемое устройство по п.121, отличающееся тем, что система находится в матрице.

123. Имплантируемое устройство по п.121, отличающееся тем, что система находится в резервуаре.

124. Композиция для модификации целевой нуклеиновой кислоты, содержащая:

а) нуклеиновую кислоту, содержащую нуклеотидную последовательность, кодирующую полипептид, где полипептид включает аминокислотную последовательность, имеющую 50% или большую идентичность по аминокислотной последовательности с аминокислотной последовательностью, представленной в любой из SEQ ID NO: 1-7, и три частичных домена RuvC, которые образуют единый домен RuvC; и

б) одну или несколько молекул ДНК, содержащих нуклеотидные последовательности, кодирующие направляющую РНК, при этом направляющая РНК содержит нуклеотидную последовательность, ком-

плементарную нуклеотидной последовательности в целевой нуклеиновой кислоте, где полипептид способен образовывать комплекс с направляющей РНК, при этом комплекс способен связывать целевую нуклеиновую кислоту.

125. Композиция по п.124, отличающаяся тем, что содержит липид.

126. Композиция по п.124, отличающаяся тем, что а) и б) находится внутри липосомы.

127. Композиция по п.124, в которой а) и б) находятся внутри частицы.

128. Композиция по любому из пп.125-127, содержащая одно или несколько из следующего: буфер, ингибитор нуклеазы и ингибитор протеазы.

129. Композиция по любому из пп.125-128, отличающаяся тем, что кодируемый полипептид содержит аминокислотную последовательность, имеющую 90% или большую идентичность с любой из аминокислотных последовательностей от SEQ ID NO: 1 до SEQ ID NO: 7, и три частичных домена RuvC, которые образуют единый домен RuvC.

130. Композиция по любому из пп.125-129, отличающаяся тем, что кодируемый полипептид представляет собой нуклеазу, способную расщеплять только одну цепь двухцепочечной молекулы целевой нуклеиновой кислоты.

131. Композиция по любому из пп.125-129, отличающаяся тем, что кодируемый полипептид представляет собой каталитически неактивный полипептид.

132. Композиция по п.130 или 131, отличающаяся тем, что кодируемый полипептид содержит одну или несколько мутаций в положении, соответствующем тем, которые выбраны из: D672, E769 и D935 SEQ ID NO: 1.

133. Композиция по любому из пп.125-132, дополнительно содержащая донорную матричную ДНК.

&gt;CasY1

MRKKLFFKGYILHNKRLVYTGKAAIRS IKYPLVAPNKTALNNLSEKI IYDYEHLFGPLNVA  
 SYARNSNRYSLVDFWIDSLRAGVIWQSKSTSLIDLISKLEGSKSPSEKIFEQIDFELKNK  
 LDKEQFKDI ILLNTGIRSSSNVRSRGRFLKCFKEEFRDTEEVACVDKWSKDLIVEGKS  
 ILVSKQFLYWEEEFGIKIFPHFKDNHDLPKLTTFFVEPSLEFS PHLPLANCLERLKKFDIS  
 RESLLGLDNNFSAFSNYFNELFNLLSRGEIKKIVTAVLAVSKSWENEPELEKRLHFLSEK  
 AKLLGYPKLTSSWADYRMI IGGKIKSWHSNYTEQLIKVREDLKKHQIALDKLQEDLKKVV  
 DSSLREQIEAQREALPLLDTMLKEKDFSDDELYRFILSDFKSLLNGSYQRYIQTEER  
 KEDRDVTKKYKDLYSNLRNI PRFFGESKKEQFNKFKINKSLPTTIDVGLKILEDIRNALETV  
 SVRKPPSITEEYVTKQLEKLSRKYKINAFNSNRFKQITEQVLRKYNNGELPKISEVFYRY  
 FRESHVAIRILPVKISNPRKDISYLLDKYQISPDWKN SNPGEVVDLIEIYKLT LGWLLSC  
 NKDFSMDFSYDLKLFPEAASLIKFNFGSCLSGYYSKMI FNCITSEIKGMITLYTRDKFV  
 VRYVTQMIGSNQKFPLLCLVGEKQTKNFSRNWGVLI EEKGD LGEEKNQEKCLIFKDKTDF  
 AKAKEVEIFKNNIWRIRTSKYQIQFLNRLFKKTKEWDLMNVLVSEPSLVLEEEWGVSWDK  
 DKLLPLLKKEKSCEERLYSLPLNLVPATDYKEQSAEIEQRNTYLGLDVGEGFGVAYAVR  
 IVRDRIELLSWGFLKDPALRKIRERVQDMKKKQVMVAFSSSSTAVARVREMAIHSLRNQI  
 HSLALAYKAKI IYEGEISISNFETGGNRMAKI YRSIKVSDVYRESGADTLVSEMIWGKKNKQ  
 MGNHISSYATSYTCCNCARTPFELVIDNDKEYEKGGDEFIFNVGDEKKVRGFLQKSLLGK  
 TIKGKEVLKSIKEYARPPIREVLLLEGEDVEQLLKRRGNSYIYRCPFCGYKTDADIQAAALN  
 IACRGYISDNAKDAVKEGERKLDYILEVRKLWEKNGAVLRSKFL (SEQ ID NO: 1)

&gt;CasY2

MQKVRKTLSEVHKNPYGKVRNAKTGYSLQIERLSYTGKEGMRSFKI PLENKNKEVFDEF  
 VKKIRNDYISQVGLNLS DWYEHYQEQEHYSLADFWLDSL RAGVIFAHKETEIKNLISK  
 IRGDKSIVDKFNASIKKKHADLYALVDIKALYDFLTSDARRGLKTEEEFFNSKRNTLFPK  
 FRKKDNKAVDLWVKKFIFGLDNKDKLNFTKKFIFGDPNPQIKYDHTFFFHQDINFDLERIT  
 TPKELISTYKFKFLGKNKDLYGSDETTEQDKMVLGFHNNHGAFSKYFNASLEAFRGRDNS  
 LVEQIINNSPYWNSHRKELEKRIIFLQVQSKKIKETELGKPHEYLASFGGKFESWVSNYL  
 RQEEEVKRQLFGYEENKKGQKKEIVGNKQELDKIIRGTDEYEIKALSKETIGLTQKCLKL  
 LEQLKDSVDDYTLISLYRQLLIVELRIRLNVEFQETYPELIGKSEKDKKDAKNKRADKRYP  
 QIFKDIKLI PNFLGETKQMVYKFFIRSADILYEGINFIDQIDKQITQNLPCFKNDKERI  
 EFTKQFETLRRKYILMNSSRFHHVIEGI INNRKLIEMKKRENSLKTFSDSKFVLSKLF  
 LKKGKKYENEVYYTFYINPKARDQRRIKIVLDINGNNSVGILODLVQKLPKWDDI IKNK  
 DMGELIDAIEIEKVRLGILIALYCEHKFKIKKELLSLDLFAAYQYLELEDDPEELSGTN  
 LGRFLOSLVCSEIKGAINKISRTEYIERYTQPMNTEKNYPLLINKEGKATWHIAAKDDL  
 SKKKGGGTVAMNQKIGKNFFGKQDYKTVFMLQDKRFDLLTSKYHLQFLSKTLDTGGGSWW  
 KNKNIDLNLSSYSFIFEQKVKVEWDLTNLDHPIKIKPSENSDDRRLFVSI PFVIKPKQTK  
 RKDLQTRVNYMGIDIGEYGLAWTI INIDLKNNKINKISKQGFIEPLTHKVRDYVATIKD  
 NQVRGTFGMPDTKLARLRENATSLRNQVHDIAMRYDAKPVYEFELSNFETGSNKVKVIY  
 DSVKRADIGRGQNNTEADNTEVNLVWGKTSKQFGSQIGAYATSYICSEFCGYSPYEFENS  
 KSGDEEGARDNLYQMKKLSRPSLEDFLQGNPVYKTRDFDKYKNDQRLQKTGDKDGWKT  
 HRGNTAIYACQKCRHISDADIQASYWIALKQVVRDFYKDKEMDGDLIQGDNKDKRKVNEL  
 NRLIGVHKDVPPIINKNLITSLDINLL (SEQ ID NO: 2)

&gt;CasY3

MKAKKSFYNQKRKFGKRGYRLHDERIAYSGGIGSMRSIKYELKDSYGIAGLRNR  
 IADATISDNKWLYGNINLNDYLEWRSSKTDKQIEDGDRESSLLGFWLEALRLGFVFSKQS  
 HAPNDFNETALQDLFETLDDDLKHVLDKRWKDFIKIGTPKTNDQGRLLKQIKNLLKGNK  
 REEIEKTLNESDDELKEKINRIADVFAKNKSDKYTIFKLDKPNTEKYPRINDVQVAFFCH  
 PDFEEITERDRKTKTLDLIINRFNKRYEITENKKDDKTSNRMALYSLNQGYIPRVLNDFLFL  
 FVKDNEDDFSQFLSDLENFFSFSNEQIKIKERLKKLKKYAEPIPGKPOLADKWDDYASD  
 FGGKLESWYSNRIEKLKKIPEVSDLRNLEKIRNVLKKQNNASKILELSQKIEEYIRDY  
 GVSFEKPEIIEKFSWINKTKDGQKQVYVAKMADREFIEKLDLWMADLRSQLEYNQDNKV  
 SFKKKGKKEIEELGVLDFAFNKAKKNKSTKNENGWQQLSESIQSAPLFFGEGNRVRNEEV  
 YNLKDLLFSEIKNVENILMSSEAEDLNKIKIEYKEDGAKKGNVVLNVLARFYARFNEGDY  
 GGWNKVKTVLENIAREAGTDFSKYGNMNNRNAGRFYLNGRERQVFTLIKFEKSITVEKIL  
 ELVKLPSLLDEAYRDLVNNKNHKLDRDVIQLSKTIMALVLSHSDKEKQIGGNYIHSKLSG  
 YNALISKRDFISRYSVQTTNGTQCKLAIGKGSKKGNEIDRYFYAFQFFKNDDSKINLKV  
 IKNNSHKNIDFNDNENKINALQVYSSNYQIQFLDWF FEKHQGKKSLEVGGSFTIAEKSL  
 TIDWSGSNPRVGFKRSDTEEKRVFVVSQPFPTLIPDDEDKERRKERMIKTKNRFIGIDIGEY  
 GLAWSLIEVDNGDKNNRGIRQLESGFITDNQQQVLKKNVKSQRQNIQRTFTSPDTKIAR  
 LRESLIGSYKNQLESMLVAKKANLSFEYEVSGFEVGGKRVAKIYDSIKRGSVRKKDNNSQ  
 NDQSWGKKGINNEWSFETTAAGTSQFCTHCKRWSSLAIVDIEEYELKDYNDNLFKVKINDG  
 EVRLLGKKGWRSGEKIKGKELFGPVKDAMRPNVDGLGMKIVKRYLKLDLRDWVSRYGNM  
 AIFICPYVDCHHISHADKQAAFNIIV (SEQ ID NO: 3)

&gt;CasY4

MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSSGGRTVPREIVSAINDDYV  
 GLYGLSNFDDLYNAEKRNEEKVYSVLDWFYDCVQYGAVFSYTAPGLLKNVAEVRGGSYEL  
 TKTLKGSPLYDELQIDKVIKFLNKKEISRANGSLDKLKKDIIDCFKAEYRERHKDQCENK  
 ADDIKNAKKDAGASLGERQKFLFRDFFGISEQSENDKPSFTNPLNLTCCLLPFDTVNNNR  
 NRGEVLFNKLKEYAQKLDKNEGSLEMWEYIGIGNSGTAFSNFLGEGFLGRLENKITEK  
 KAMMDITDAWRGQEQEELKRLRILAALTIKLRPKFDNHWGGYRSDINGKLSWLQNY  
 INQTVKIKEDLKGHKDLKAKEMINRFGESDTKEAVVSSLESIEKIVPDDSADDEKP  
 DIPAIAYRRFLSDGRLTLNRFVQREDVQEALIKERLEAEKKKKPKKRKKKSDAEDEKET  
 IDFKELFPHLAKPLKLVNPFYGDSKRELYKYYKNAAIYTDALWKAVEKIYKSAFSSSLKN  
 SFFDTEFDKDFFIKRLQKIFSVYRRENTDKWKPIVKNFAPYCDIVSLAENEVLYKPKQS  
 RSRKSAIDKNRVRPSTENIAKAGIALARELSVAGFDWKDLLKKEEHEEYIDLIELHKT  
 ALALLAVTETQLDISALDFVENGTVKDFMKTDRDGNLVLEGRFLEMFSQSIVFSELRLGLA  
 GLMSRKEFITRSAIQTMNGKQAEELLYIPHEFQSAKITTPKEMSRFLDLAPAEFATSLEP  
 ESLSEKSLKLMRYYPHYFGYELTRTGQIGDGGVAENALRLEKSPVKKREIKCKQYKT  
 LGRGQNKIVLYVRSSYYQTQFLEWFLHRPKNVQTDVAVSGSFLIDEKKVKTRWNYDALTV  
 ALEPVSGSERVFVSQPFITFPEKSAEEEGQRYLIGIDIGEYGIAYTALEITGDSAKILDQN  
 FISDPQLKTLREEVKGLKLDQRRGTFAMPSTKIARIRESLVHSLRNRIHHLALKHKKAKIV  
 YELEVSRFEEGKQKIKKVYATLKKADVSEIDADKNLQTTVWGKLAVASEISASYTSQFC  
 GACKKLWRAEMQVDEITITTQELIGTVRVIKGGTLIDAIKDFMRPPIFDENDTFFPKYRDF  
 CDKHHISKMRGNSCLFICPFCRANADADIQASQTIALLRYVKEEKKVEDYFERFRKLKN  
 IKVLGQMKKI (SEQ ID NO: 4)

&gt;CasY5

MAESKQMOCRKCGASMKYEVIIGLGKKSCRYMCPDCGNHTSARKIQNKKKRD  
 KKYGSASKAQSQRIVAGALYPDKKVQTIKTYKYPADLNGEVHDSGVAEKIAQAIQEDEI  
 GLLGPSSEYACWIASQKQSEPYVVDFFWFDVAVCAGGVFAYSGARLLSTVLQLSGEEESVLR  
 AALASSPFVDDINLAQAQAEKFLAVSRRTGQDKLGKRIGECFAEGRLEALGIKDRMREFVQA  
 IDVAQTAGQRFAAKLKIFGISOMPEAKQWNNDSSGLTVCILPDYVPEENRADQLVVLRRR  
 LREIAYCMGIEDEAGFEHLGIDPGALSNFNSGNPKRGFLGRLLNNDI IALANNMSAMTPY  
 WEGRKGELIERLAWLKHRAEGLYLKEPHFGNSWADHRSRIFSRIAGWLSGCAGKLIKIAKD  
 QISGVRTDLFLLKRLLDVAVPQSAPSPDFIASISALDRFLEAAESSQDPAEQVRALYAFHL  
 NAPAVRSIANKAVQRSQEWLIKELDAVDHLEFNKAFPPFFSDTGKKKKKGANSNGAPSE  
 EYTTETESIQQPEDAEQEVNGQEGNGASKNQKQKQRI PRFFGEGSRSEYRILTEAPQYFD  
 MFCNNMRAIFMQLESQPRKAPRDFKCFLQNRLOKLYKQTFLNARSNKCRALLESVLI SWG  
 EFYTYGANEEKFRLRHEASERSSDPDYVVOQALELARRLFLEGFWEWRDCSAGERVDLVEI  
 HKKAISFLLAITQAEVSVGSYNWLGNSTVSRYL SVAGTDTLYGTQLEEFLNATVLSQMRG  
 LAIRLSSQELKDGFDVQLESSQDNLQHLLVYRASRDLAACKRATCPAELDPKILVLPVG  
 AFIASVMKMIERGDEPLAGAYLRHRPHSFGWQIRVRGVAEVGMDQGTALAFQKPTSESEPF  
 KIKPFSAQYGPVLWLNSSSYSQSQYLDGFLSQPKNWSMRVLPQAGSVRVEQRVALIWNLQ  
 AGKMLRERSGARAFFMPVPPFSFRPSSGSGDEAVLAPNRYLGLFPHSGGIEYAVVDVLD SAG  
 FKILERGTI AVNGFSQKRGERQEEAHREKQRRGISDIGRKKPVQAEVDAANELHRKYTDV  
 ATRLGCRIVVQWAPQPKPGTAPTAQT VYARAVRTEAPRSGNQEDHARMKSSWGTYWGTYW  
 EKRPEDILGISTQVYWTGGIGESCPAVAVALLGHIRATSTQTEWEKEEVVFGRLKKFFP  
 S (SEQ ID NO: 6)

&gt;CasY6

MAESKQMOCRKCGASMKYEVIIGLGKKSCRYMCPDCGNHTSARKIQNKKKRD  
 KKYGSASKAQSQRIVAGALYPDKKVQTIKTYKYPADLNGEVHDRGVAEKIEQAIQEDEI  
 GLLGPSSEYACWIASQKQSEPYVVDFFWFDVAVCAGGVFAYSGARLLSTVLQLSGEEESVLR  
 AALASSPFVDDINLAQAQAEKFLAVSRRTGQDKLGKRIGECFAEGRLEALGIKDRMREFVQA  
 IDVAQTAGQRFAAKLKIFGISOMPEAKQWNNDSSGLTVCILPDYVPEENRADQLVVLRRR  
 LREIAYCMGIEDEAGFEHLGIDPGALSNFNSGNPKRGFLGRLLNNDI IALANNMSAMTPY  
 WEGRKGELIERLAWLKHRAEGLYLKEPHFGNSWADHRSRIFSRIAGWLSGCAGKLIKIAKD  
 QISGVRTDLFLLKRLLDVAVPQSAPSPDFIASISALDRFLEAAESSQDPAEQVRALYAFHL  
 NAPAVRSIANKAVQRSQEWLIKELDAVDHLEFNKAFPPFFSDTGKKKKKGANSNGAPSE  
 EYTTETESIQQPEDAEQEVNGQEGNGASKNQKQKQRI PRFFGEGSRSEYRILTEAPQYFD  
 MFCNNMRAIFMQLESQPRKAPRDFKCFLQNRLOKLYKQTFLNARSNKCRALLESVLI SWG  
 EFYTYGANEEKFRLRHEASERSSDPDYVVOQALELARRLFLEGFWEWRDCSAGERVDLVEI  
 HKKAISFLLAITQAEVSVGSYNWLGNSTVSRYL SVAGTDTLYGTQLEEFLNATVLSQMRG  
 LAIRLSSQELKDGFDVQLESSQDNLQHLLVYRASRDLAACKRATCPAELDPKILVLPAG  
 AFIASVMKMIERGDEPLAGAYLRHRPHSFGWQIRVRGVAEVGMDQGTALAFQKPTSESEPF  
 KIKPFSAQYGPVLWLNSSSYSQSQYLDGFLSQPKNWSMRVLPQAGSVRVEQRVALIWNLQ  
 AGKMLRERSGARAFFMPVPPFSFRPSSGSGDEAVLAPNRYLGLFPHSGGIEYAVVDVLD SAG  
 FKILERGTI AVNGFSQKRGERQEEAHREKQRRGISDIGRKKPVQAEVDAANELHRKYTDV  
 ATRLGCRIVVQWAPQPKPGTAPTAQT VYARAVRTEAPRSGNQEDHARMKSSWGTYWSTYW  
 EKRPEDILGISTQVYWTGGIGESCPAVAVALLGHIRATSTQTEWEKEEVVFGRLKKFFP  
 S (SEQ ID NO: 7)

&gt;CasY7

MKRIILNSLKVAAALRLLFRGKGSSELVKTVMKYPLVSPVQGAVEELAEAIRHDNLHLFGQKEI  
 VDLMEKDEGTQVYSVVDVFWLDTLRLGMFFSPSANALKITLGKFNSDQVSPFRKVLQSPF  
 FLAGRLKVEPAERILSVEIRKIGKRENRVENYAADVETCFIGQLSSDEKQSIQKLANDIW  
 DSKDHEEQRMLKADFFAIPLIKDPKAVTEEDPENETAGKQKPLELCVCLVPELYTRGFGS  
 IADFLVQRLTLRRDKMSTDTAEDCLEYVGIIEEEKGNMNSLLGTFLKNLQGDGFEQIFQF  
 MLGSYVGWQKEDVLRERLDLLAEKVKRLPKPKFAGEWSGHRMFLHGQLKSWSSNFFRLF  
 NETRELLESIKSDIQHATMLISYVEEKGGYHPQLLSQYRKLMEQLPALRTKVLDPETIEMT  
 HMSEAVRSYIMIHKSVAGFLPDLLES LDRDKDREFLLSIFPRIPKIDKKTKEIVAWELPG  
 EPEEGYLFTANNLFRNFLENPKHVPRFMAERI PEDWTRLRSAPVWFDMVKQWQKVVNQL  
 VESPGALYQFNESFLRQRLQAMLTVYKRD LQTEKFLKLLADVCRPLVDFDFGLGGNDII FK  
 SCQDPRKQWQTVIPLSVPADVYTACEGLAIRLRET LGFEWKNLKGHEREDFLRLHQLLGN  
 LLFWIRDAKLVVKLEDWMNNPCVQBYVEARKAIDLPLEIFGFVPIFLNGYLFSELRQLE  
 LLLRRKSVMTSYSVKTTGSPNRLFQLVYLP LNPSDPEKKNSNFFQERLDTPTGLSRRFLD  
 LTLDAFAGKLLTDPVTQELKTMAGFYDHLFGFKLPCKLAAMSNHFGSSSKMVVLAKPKKG  
 VASNIGFEPIPDPAHPVFRVRSWPELKYLEGLLYLPEDTPTTELAETS VSCQSVSSVA  
 FDLKNLTTILGRVGEFRVTADQPFKLTPIIPEKEESFIGKTYLGLDAGERSGVGFIVTV  
 DGDGYEVQRLGVHEDTQLMALQQVASKSLKEPVFQPLRKGTFRQQERTIRKSLRGCYWNFY  
 HALMIKYRAKVVHEESVGSGLVQWLRFAQKDLKKADVLPKKGKNGVDKCKRESSAQD  
 TLWGGAFSKKEEQQIAFEVQAAGSSQFCLKCGWWFQGMREVN RVQESGVVLDWNRSIVT  
 FLIESSGEKVYGFSPQOLEKGFDPDIETFKKMVRDFMRPPMFDKGRPAAAYERFVLGRR  
 HRRYRFDKVFEEFRGSRALFICPRVGGNFDHSSEQSAVV LALIGYIADKEGMSGKKL VY  
 VRLAELMAEWKLLKCLERSRVEEQSSAQ (SEQ ID NO: 5)

&gt;CasY18

MKRIAKFRHDKPVKREAWSKGYRVHKNRI INKVTRSIKYPLVVKDEWKKRLIDDAAHDYRWLVG  
 PINYSDWCRDPNQYSILEFWIDFLCVGGVFQSSHSNICRLAIQLSGGSVFEQEWKDLSPFVRAN  
 LIQGIKPAEFIGFLTAEFRSSSNPKNFISKFFEGSNEDLES LTNEFASIVDFIKAKDISLLRKS  
 LPSCCKIAPNLWEKAVGSHSTNELLKLLTKYTRVMLVAEPSHS DRVFSQTVLQSNQDDPELTG  
 PLPSHKVKGASYLFIPEFTIREVNLDKISKLDLSAKSKLAVEQVKKLSELTSDFKQIENQSEAYF  
 GLSTSFNELSNFLGILIRTLR NAPEAILKDQIALCAPLDKDI LKITLDWLCDRAQALPENPRFE  
 TNWAEYRSLGGKIKSWFSNYENFFEIPQAASSQONNREK KLGNRSAIRALNLKKEAFEKARE  
 TFKGDKGTLEKIDLAYRLGSI SPEVLQCD EGLKLYQQFNDELLVLNETINQKFQDAKRDIAK  
 KEKESFEKLQRNLSSPLPRIPEFFGERAKKGYQKARVSPK LARHLLLECLNDWLARFAKVEESAF  
 SEKEFQRILDWLRTSDFLPVFIRKSKDPPSWLRYIARVATGKY YFWVSEYSRKRQVIIDKPIAQ  
 NPLKELISWFLLNKDAF SRDNELEFKGLSSKMVTLARIMAGILRDRGEG LKELQAMTSKLDNIGL  
 LHPSFSVPVTD SLKDAAFYRAFFSELEGLLNIGRSRLTIERITLQSQQSKNKKTRRPLMPEPFI  
 NEDKEVFLAFPKFETKNKVKGTRVYNSPDEVNWL LSPIRSSKGQLSFMFRCLSEDAKIMTTSG  
 GCSYIVEFKLLEAQEEVLSIHDCDIIPRAFVSIPTFLER ESEETKPDWKPNRFMGVDIGEYAV  
 AYCIVIEKGTDSIEILD CGIVRNGAHRVLKEKVDRLKRRQRSMTFGAMDT SIAAARES LVGNRYRN  
 RLHAIALKHGAKLVY EYEVSAFESGGNRIKKVYETLKKSDCTGETEADKNARKHIWGETNAVGD  
 QIGAGWTSQTC AKCGRSFGADLKAGNFGVAVPVEKVEDSKGHYAYHEFPFEDGLKVRGFLKPN  
 KII SDQKELAKAVHAYMRPPLVALGKRKLPKNARYRRGNSSLFRCPFSDCGFTADADIQAAYNI  
 AVKQLYKPKKGYPKERKQDFVILKPKPEPSKLEDKQFYRPN (SEQ ID NO: 8)

Фиг. 1



	237	244	251	261	265	275
1. CasY1	LKKFDISR	ES	-----L	LGLDN-NFS	SNFYFNE	-----L
2. CasY2	LYGSDETTE	D	QLKML	LCFHN-NHG	AFSKYFNA	-----S
3. CasY3	YEITENKK	DD	KTSNR	MALYSLN	OGYIPRVL	ND
4. CasY4	AOKL	DKNEGS	LEMWEY	IGTGN-SG	TAFSNFL	GE
5. CasY5	AYCMGIE	EA	-----G	FELGL	-----D	FGALSNFS
6. CasY6	AYCMGIE	EA	-----G	FELGL	-----D	FGALSNFS
7. CasY7	RDKMSTDAE	DCLEY	VGL	EEKGN	GVNSLL	GT
	285	292	302	310	320	330
1. CasY1	VISKSW	-----E	NEPE	EKRRIHF	USEKAKL	-----G
2. CasY2	NSIPY	W	-----N	SHRKE	EKRRIHF	QVOSK
3. CasY3	FFSFS	-----N	EIKIKI	EKRRIHF	QVOSK	-----G
4. CasY4	ITDAW	RGQEE	EE	EKRRI	DAALTI	K
5. CasY5	MTIPY	W	-----E	GRKGE	IERLAW	LKHRAE
6. CasY6	MTIPY	W	-----E	GRKGE	IERLAW	LKHRAE
7. CasY7	SYVGT	-----Q	KE	EDV	ERL	DL
	340	343	353	363	373	
1. CasY1	TEQLIK	VRED	-----K	KH	QIA	DKLQ
2. CasY2	LRQEE	EMKRC	Q	FGYEEN	KKGQ	KK
3. CasY3	IEK	LKKI	PE	SV	-----S	DL
4. CasY4	IN	OTV	KIK	ED	-----K	GH
5. CasY5	AG	LKI	AK	Q	-----S	GV
6. CasY6	AG	LKI	AK	Q	-----S	GV
7. CasY7	F	RLF	NE	T	REL	-----E
	383	392	395	396	398	
1. CasY1	L	KEK	D	F	S	D
2. CasY2	L	KL	E	Q	L	K
3. CasY3	L	EY	R	D	Y	G
4. CasY4	N	P	D	S	A	D
5. CasY5	D	R	F	L	E	A
6. CasY6	D	R	F	L	E	A
7. CasY7	M	E	Q	L	P	A
	412	413	419	424		
1. CasY1	L	NGS	Y	Q	R	-----Y
2. CasY2	L	N	V	E	F	Q
3. CasY3	L	N	E	Y	N	Q
4. CasY4	L	N	R	F	V	Q
5. CasY5	A	N	K	A	V	Q
6. CasY6	A	N	K	A	V	Q
7. CasY7	L	D	R	D	K	D
	430	433	443	453	463	
1. CasY1	-----Y	KD	L	S	N	L
2. CasY2	-----Y	P	Q	I	F	K
3. CasY3	-----Q	Q	L	S	E	S
4. CasY4	D	F	K	E	L	-----F
5. CasY5	E	T	E	S	I	Q
6. CasY6	E	T	E	S	I	Q
7. CasY7	T	A	N	N	L	-----F

	473	483	490	497	507	
1. CasY1	GLKIEDIRNALETVSVRK	---	PPSITTEYVTIKO	---	LEKLSRKYKINAFNS	
2. CasY2	GINFTDQIDKQITQNL	LPC-FKNDKERIEETEKO	---	FETLRRKYYL	MNS	
3. CasY3	FIKNVENTILMSSEAEDL	KN--IKIEYKEDGAKK	GNVVLNVLARFYAR	---	FNEDGYGGW	
4. CasY4	LWKAVEKITYKSAFSS	LKNSFFD	DFDFFIKR	---	LQKIFSVYRR	
5. CasY5	FCNNMRAIFMQL	ESQPRKA	---	PRDFKCFLONR	---	
6. CasY6	FCNNMRAIFMQL	ESQPRKA	---	PRDFKCFLONR	---	
7. CasY7	MVKQWQKVVNQ	LIVESPGAL	---	YQFNESFLROR	---	
	512	523	522	525	535	
1. CasY1	NRFKQITTEQV	---	---	LRKYNNGELPKIS	EVFVYR	
2. CasY2	SRFHHTVIEGT	IINNRKLIEMKKRE	ENSELKT	TSIS	SKFVLSKLF	
3. CasY3	NKVKVTLENT	---	---	AREAGTDFSKY	GNNNNR	
4. CasY4	DKWKPIVKN	---	---	FAPYCDIVSLA	ENLVLYK	
5. CasY5	NKCRALLESV	---	---	LISWGEFYTYG	ANKKFR	
6. CasY6	NKCRALLESV	---	---	LISWGEFYTYG	ANKKFR	
7. CasY7	EKFIKLLADV	---	---	CRPLVDF	FFGLGCNDI	
	550	551	561	571	577	587
1. CasY1	I	---	LPVKISNPRKID	TSYLLDKYQIS	PDW	---
2. CasY2	INPKARDQRR	KIVLDDINGNS	SVGILQDLVQ	KLKPKWDDI	IKKNDMG	ELIDALEIEK
3. CasY3	IKF	---	---	SITVEKILEL	LVKLP	PSLLDEAYRDL
4. CasY4	AAI	---	DKNRVRLP	STENIAKAGI	ALARELS	VAGRDW
5. CasY5	SDP	---	---	YVVOALE	ETARRLF	FLGFEM
6. CasY6	SDP	---	---	YVVOALE	ETARRLF	FLGFEM
7. CasY7	P	---	---	LSVPADV	TACGLAT	ELRET
	597	607	615	625	630	640
1. CasY1	TLGWLLS	CNKDFSMDFSSY	---	DLKLPPEASL	TKNF	---
2. CasY2	RTGILLALY	CHFKFKK	ELLSLDL	FASAYQVLE	---	DDPEETS
3. CasY3	IMALMLSHS	DKKQLGG	---	---	---	---
4. CasY4	ATALLLAVT	ETOLD	T	SAL--DFV	ENGTVKDF	FMKTR
5. CasY5	ATS	FLLAIT	QAEMSV	GSY--N	WLGNST	VSRVLSVA
6. CasY6	ATS	FLLAIT	QAEMSV	GSY--N	WLGNST	VSRVLSVA
7. CasY7	LGNL	FWIR	DAKLV	VKLE--	DWMN	PCVQEV
	650	660	678			
1. CasY1	TSEIKG	MITLYTR	DKFV	VRYVT	OMIG	SNO
2. CasY2	CSEIKG	AINKIS	RTI	EYIER	YTV	OPM
3. CasY3	HSKLS	SGYNAL	TSKR	DFIS	RSV	OTT
4. CasY4	FSELR	GLAGL	MSKKE	FIT	RSAT	OTM
5. CasY5	LSOMR	GLAIR	LSSQ	ELK	DGED	VOLES
6. CasY6	LSOMR	GLAIR	LSSQ	ELK	DGED	VOLES
7. CasY7	FSELR	GLLE	LLRR	KIS	VMT	SYSVK
	679	687	697			
1. CasY1	---	---	---	LVGEK	QTK	NFSRN
2. CasY2	---	---	---	NKEG	KATW	---
3. CasY3	---	---	---	KGKS	KGN	EIDRY
4. CasY4	TPKEMS	RAF	LDL	PAEF	ATS	LEPESL
5. CasY5	CPAELD	PKIL	VLP	VGA	FIAS	VMKMI
6. CasY6	CPAELD	PKIL	VLP	VGA	FIAS	VMKMI
7. CasY7	TPTGL	SRR	FLDL	TLDA	FAGKL	---

	704	711	721	729	737	747
1. CasY1	GEEK---NQEKCLIFKDKTDFAK--AKEVEIFKNNT--WRIRTSKYQIQFLNRLFKKT					
2. CasY2	SKKKG--GTVAMNQKIGKNFFGKQDYKTVFMLQDKR--FDLLTSKYHLQFLSKTLDTG					
3. CasY3	SKIN-----LQVIKNNSHK--NIDFNDNENKTNALQVYSSNYQIQFLDWFEEKH					
4. CasY4	GVAANALRLEKSPVKREIKCKQ--YKTLGRGQNKI-VLYVRSYQYQTFLEWFLHRIP					
5. CasY5	GMDQ---GTALAFQKPTSEPFK--IKPFSAQYGPV-LWLNSSYSQSOYLDGFLSOP					
6. CasY6	GMDQ---GTALAFQKPTSEPFK--IKPFSAQYGPV-LWLNSSYSQSOYLDGFLSOP					
7. CasY7	HPGSSSKMVVLAKPKKGVASNIG--FEPIDPAHPV--FRVRSWPELKYLEGGLIYLP					
	754	762	772	782	789	799
1. CasY1	----KQWDLMNQVLSLSEPSLVLEEEWGVSWDKDKLLPLLK--KEKSCCEERLYYSLE					
2. CasY2	GGSWWKNKN--IDNLSYSFIFDKKVKVEMDLTNLDHPTIKKPSENSDDRRTFVSIPE					
3. CasY3	----QGKK-TSLEVGGSFITAEKSLTLDWVSGSNPRVGFK--RSDTEEKRVFVISOPE					
4. CasY4	----KNVQ-TDVAVSGSFLIDKVKTRWNYDALTVALE--PVSGSERVYVISOPE					
5. CasY5	----KNWS-MRVLPOAGSVRVLQVAVLIWNLQAGKMRLE--RSGARAFVMPVPE					
6. CasY6	----KNWS-MRVLPOAGSVRVLQVAVLIWNLQAGKMRLE--RSGARAFVMPVPE					
7. CasY7	----EDTTP-LITTELAETSIVSICQSVSSVAFDLKNLTTITLG--RVGEFRVTDQPE					
	809	818	827	836	842	851
1. CasY1	LNIVPATDYKEQSAE--TEQRNTYLGIDVCE-FGVAVAVVRI----VRDRIEILISWG					
2. CasY2	FVTKPK---QTKRKD--QTRVNYMGIDIGE-YGLAWTITNIDL--KNKKINKTISKQG					
3. CasY3	FTLIPDDEDKERRKERMKTKNRFIGIDIGE-YGLAWSTIEVDNGDKNNRGIROLESQ					
4. CasY4	FTLIFPE---KSAEE---GORYLGDIGE-YGLAYTADEI----TGDSAKITLDQN					
5. CasY5	FSFRPS---GSGDEA--VLAPNRYLGLFPHS-GGIEYAVVDVL----DSAGFKILERG					
6. CasY6	FSFRPS---GSGDEA--VLAPNRYLGLFPHS-GGIEYAVVDVL----DSAGFKILERG					
7. CasY7	EKLITPI---IPEKEE--SFIGKTYLGDAGERSGVGEAVITV----DGDGYEVORLG					
	861	870	876	886	896	
1. CasY1	FKKDPALRKIRE-RVQDMK---KQVMAVFSSSSTAVARVREMAIHSILRNQIHSIALI					
2. CasY2	FVYVPLTHKVRD-YVATIK---DNQVRGTFGMPDTKLARLRENAITSILRNQVHDITAM					
3. CasY3	FTDNQQQVLIK-NVKSWE---QNLROTFTSPDTKIARLRESLIGSYKMOLESILVV					
4. CasY4	FTSDPQLKTIK-EVQGLK---LDQVRGTFAMPSTKIARLRESLVHSILRNRIHHTAL					
5. CasY5	TAVNGFSQKRG-EROEEAH--REKQRGHSIDIGRKKPVOAEVDAANELHRKYTDVAT					
6. CasY6	TAVNGFSQKRG-EROEEAH--REKQRGHSIDIGRKKPVOAEVDAANELHRKYTDVAT					
7. CasY7	VHEDTQLMALIQQVASKSLKEPVFQPLRKGTF---RQQRIRKISLRGCYWNFYHALVIT					
	906	916	926	936	942	947
1. CasY1	AYKAKIIVYIISISNFEETGGNRMAKTYRSIKVSDIYR-----ESGADITLVSEMI					
2. CasY2	RYDAKPVYEFETISNFEETGSNKVKVITYDSVKRAIDIGRGO-----NNTEDANTEVNIV					
3. CasY3	AKKANLSFEYEMSGFEVGGKRVAKIYDSIKRGSVRK-----KDNNSQNDQS					
4. CasY4	KHKAKIVYELMSRFEEGKQIKKVVATLKKADIVYS-----EIDADKNLQTTV					
5. CasY5	RLGCRIVVQWAPQPKPGTAPTACTVYARAVRTEAPRS-----GNOEDHARMKSS					
6. CasY6	RLGCRIVVQWAPQPKPGTAPTACTVYARAVRTEAPRS-----GNOEDHARMKSS					
7. CasY7	KYRAKVVHESVWSSGLVGQWLRFAQDKLKKADVLPKKGKNGVDKKKRESAQTLL					
	956	967	977	986	996	
1. CasY1	WG-----KKNKQMGNHISISYATSYTCNCAR-TPFELVIDNDKEYEKG----G					
2. CasY2	WG-----KTSKQFGSQIGAYATSYICSFQGY-SPYYEFENSKSGDEEG----A					
3. CasY3	WG-----KKGINEWSFETTAAAGTSQFCTHCKR-WSSLAIVDIEEYELKD----Y					
4. CasY4	WG-----KLAVASEISASYTSCFGACKKLWRAEMQVDEITTT-----					
5. CasY5	WGYTWGTYWEKRPEDILGISTQVYWTGGIGESCP-----					
6. CasY6	WGYTWSTYWEKRPEDILGISTQVYWTGGIGESCP-----					
7. CasY7	WG----GAFS-KKEEQQIAFEVQVACSSQFCLKCGW-WFQLGMREVNRVQESGVVLDW					

	1002	1009	1019	1028	1042
1. CasY1	DEFIIFNVGDIIEKKVGRGFLQKSL	-GKTIK	GKEVLKSIK	EYAR	PPPIREV-----
2. CasY2	RDNLYQMK-----	KLSRPSLE--	DFLQGNPVYKTF	RDFDKYKNDQR-----	
3. CasY3	NDNLFKVKIN--	DGEVRLLGKKGWRS	GEEKIKGKELFGPVK	DAMRIPNVDGL	-GMKIVKR
4. CasY4	MELHGTVR-----		VIKGGTLIDA	KDFMR	PIFDENDT
5. CasY5	-----				PFPHY
6. CasY6	-----				
7. CasY7	NRSIVTFLIESSG	EKVYGFSPQ	QLEKGRPDIET	FKKMVR	RDFMRPPMFD

	1043	1046	1055	1062	1070	1080		
1. CasY1	----L	LEGEDVEQLL	K----	RRGNSYIYR	CPF--	CGYKTDADI	QAALNLA	CRGYISD
2. CasY2	----L	QKTGD	KDGEWKT--	HRGNTAIYAC	QK--	CRHISDADI	QASYW	IALKQVVRD
3. CasY3	KYL---	KLDLRDWVS	----	RYGNMAIF	ICPYVD	CHHISHADK	QA	AFNTAV
4. CasY4	RDF-----	CDKHHISK	----	MRCNSCL	ICPF--	CRANADA	DIQAS	OTIALLRYVKE
5. CasY5	-----							AVAVALLGHIRA
6. CasY6	-----							AVAVALLGHIRA
7. CasY7	ERFVLGRR	RRYRF	DKVFEER	FRSAL	FIICPRV	GICGNFDHS	SEQSA	AVVLLALLGYIAD

	1090	1100	1110	1114	1122	1125
1. CasY1	NAKDAVK	GERKLDYILE	VVRK	WE-----	KNGAV	RSATFL
2. CasY2	FYKDKEM	GDLIQGD	NKDKR	KVNELNRL	LIGVH	KDVP
3. CasY3						
4. CasY4	EKKVE--		DYFERFRK			KNIKVI
5. CasY5	TSTQTEW					KEEVV
6. CasY6	TSTQTEW					KEEVV
7. CasY7	KEGMSGKK		LVIYRLAEL	MAE-----	WKLKK	LES

B

	818	827	836	844	850
1. CasY1	IIEQRNT	YLGIDV	GCF--	GVA	AVVRI--
2. CasY2	IQTRV	YMGID	IGEY--	GLAW	TIIN
3. CasY3	IKTNR	FIGID	IGEY--	GLAWS	LIIEVD
4. CasY4		GQRYL	GIDIGEY--	GIAY	TALEI--
5. CasY5	VLAPNR	YLGIF	PHSG--	GIEY	AVVDV--
6. CasY6	VLAPNR	YLGIF	PHSG--	GIEY	AVVDV--
7. CasY7		GKTYL	GIDAGERS	GVG	FAIVTV--
8. AsCpf1_5843	EHPETPI	IGIDR	GER--	NLI	YITV--
7. LbCpf1_SID6	HDDNPYV	IGIDR	GER--	NLL	YIVV--

	859	868	874	884	889
1. CasY1	LRKIRE	-RVQDMK	----	KKQ	MAV
2. CasY2	THKVRD	-YVATIK	----	DNQ	VRGT
3. CasY3	QQVLRK	-NVKSWR	----	QNI	ROTE
4. CasY4	LKTIRE	-EVKGLK	----	LDQ	RGTE
5. CasY5	FSQKRG	-EROEEAH	----	REK	RRGIS
6. CasY6	FSQKRG	-EROEEAH	----	REK	RRGIS
7. CasY7	LMAIQ	QVASK	SLKEPV	FQPLR	KGTER--
8. AsCpf1_5843	LNTIQ	QFDYQ	KKL----	DN	REK
7. LbCpf1_SID6	GIRIKT	-DYHSL	----	DK	KEKERE

	899	909	919	928	937
1. CasY1	NQIHS	LALAKKIIY	EISISN	FETGGNRM	--AKIYRSIKVSDVYK
2. CasY2	NOVHD	LAMRYDAKPIY	EFEISN	FETGSKVM	--KVIYDSVKRADIGRIG
3. CasY3	NQLES	LMVAKKANISF	EYEVSG	FVGGKRV	--AKIYDSIKRGSVRIK
4. CasY4	NRRIHH	LALKKIAKIY	ELEVSR	FEEGKQKI	--KIVYATLKKADVYS
5. CasY5	RKIYTD	VATRLGCRIV	QWAPQ	PKPGTAPTA	--QTVYARAVRTEIAPR
6. CasY6	RKIYTD	VATRLGCRIV	QWAPQ	PKPGTAPTA	--QTVYARAVRTEIAPR
7. CasY7	RFYHA	LMIKYRAKVIH	EESVGS	SGLVGQWL	--RAFQKDLKKADVLPKKG
8. AsCpf1_5843	HEIIVD	LMIHVQIAVV	LENLNF	GFKSKRTGIAE	KIAYVQQFEEKMLIDK

7. LbCpf1\_sid6 HKIICE LVEKVDIAVIAL E DLNSG FKN SRVKM - EKQVYQKFEKMLIDK

	942	955	956	960
1. CasY1	-----ESGAD	TLVSEMI	WIC	-----KKN-KQMGNHISISY
2. CasY2	---QNNTEAD	NTEVNLV	WIC	-----KTS-KQFGSQIGIY
3. CasY3	-----KDNNS	QNDQS	WIC	-----KKGINWSFETTA
4. CasY4	-----EIDAD	KNLQTTV	WIC	-----KLA---VASEISIS
5. CasY5	----SGNQED	HARMKSS	WIC	-YTWG-----TYWEKRPEDILGISTQVY
6. CasY6	----SGNQED	HARMKSS	WIC	-YTWS-----TYWEKRPEDILGISTQVY
7. CasY7	KNGVDKKKRE	SSAQDTL	WIC	-----GAFSKKEE-QQIAFEMQIA
8. AsCpf1_5843	----LNCLVL	KDYPAEKV	G	VLNPHYQLTDQFTSFAKMGTSQSGFLFYVPAIP

7. LbCpf1\_sid6 ----LNYMVD KKSNPCAT G GALKGYQITNKFESFKSMSTQNGFIFYIPAW

	970	980	994	995
1. CasY1	ATSYTCCNC	ARTPFELV	IDNDKEYEK	-----GGDEFLIFN
2. CasY2	ATSYICSF	CGYSPYEF	ENSKSGDEE	-----GARDNLIYQ
3. CasY3	GTSQFCTHC	KRWSSLA	IVDIEEYELK	-----DYNLNLFK
4. CasY4	YTSQFCGAC	KKLWRAEM	QVDETITTO	-----LIGT
5. CasY5	WTGGIGESC	PAVAVAL	LGHIRATSTO	-----
6. CasY6	WTGGIGESC	PAVAVAL	LGHIRATSTO	-----
7. CasY7	GTSQFCLKC	GWWFQLG	MREVNRVQES	SGVVL-----DWNRSIVT
8. AsCpf1_5843	YTSKIDPLT	GFVDPFV	WKTIKNHES	RKKHFLEGDFLHY--DVKTKGDFILH

7. LbCpf1\_sid6 LTSKIDPSTGFVNLLKTKYTSIADS K K - FISSFDRIMYVPEEDLFEFALD

	1004	1005	1012	1016
1. CasY1	VG	-----	DEKKVRGF	-----LQKSLLGKTIK
2. CasY2	MKK	-----	-----	-----LSRPSLEDFLQG
3. CasY3	VKI	-----	NDGEVRLLGK	-----KGWRSGEKIKG
4. CasY4	VRV	-----	-----	-----IKG
5. CasY5	-----	-----	-----	-----TEWEKE
6. CasY6	-----	-----	-----	-----TEWEKE
7. CasY7	FLIE	-----	SSGEKVYGFSP	-----QOLEKGFEPDI
8. AsCpf1_5843	FKMNRNLS	FQRLPG	FMPAWDIV	FEKNETQFDAQGTPFIAGKRIVPIEN

7. LbCpf1\_sid6 YK-----NFSRTDADYIKKWKLYSYGNRIRIFAA-----AKKNNVFAWTE

	1 026	1 036	1 042	1 043	1 052
1. CasY1	KEVILKSIKREYARPIPI		REIV		LEGE DVEQL---
2. CasY2	NPVYKTFRDFDKYKN		DQR		LQKTGDKDGEWK-
3. CasY3	KELEFGPVRDAMRPNV		DGLGMKIVKRKYLK		---
4. CasY4	GTLIDAIKDFMRPIPI		FDENDTFFPKYRDFC		DKHHISK-
5. CasY5	EVVVFGRLKKKFF-PS				
6. CasY6	EVVVFGRLKKKFF-PS				
7. CasY7	ETFKKMVRDFMRPMPM		FDRKGRPAAAYERFV		LGRRHRRYRFDKV
8. AsCpf1_5843	HRFTGRYRDLY-PEANELIALLEEKGI		VERDGSNILPK		LEND DSHAITDM
7. LbCpf1_SID6	VCLTTSAYKELF		NKYGINYQQG		DIRALICEQS DKAFYSSE

	1 053	1 061	1 067
1. CasY1	LKRRGNSYIYRCPF		CGYKT
2. CasY2	THRGNTAIYACQK		CRHIS
3. CasY3	LDLRDWVSRYGNMAIFICPYV		DCHHS
4. CasY4	KMRGNSCLFICPF		CRANA
5. CasY5			
6. CasY6			
7. CasY7	FERFGRSALFICPRV		GIGNFD
8. AsCpf1_5843	VALIRSVLQMRNS-NAATGEDYINSF		VVRDLNGVCFDSRFQNP
7. LbCpf1_SID6	MALMSLMLQMRNSITGRTDVDFLISF		VVKNSDGFYDSRNYEAQENAILPK

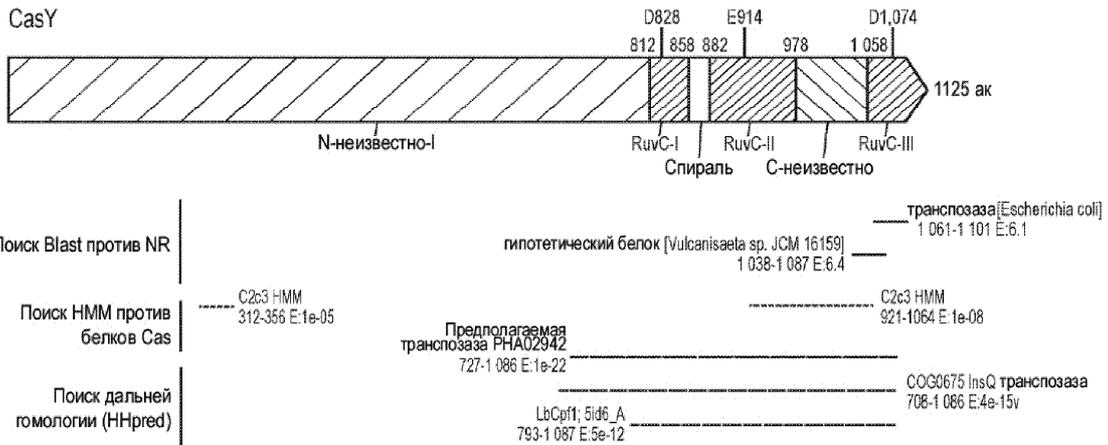
	1 072	1 083	1 093	1 103	1 113
1. CasY1	DADIQAALNLACRGYISDN		AKDAVKEGERKIDYILEV		RKIWEKNGAV---
2. CasY2	DADIQAISYWTALKQVVRD		IFYKDKEMDGDLIQGDNKDKR		KVNELNRLIGVH
3. CasY3	HADKQAFNIAV				
4. CasY4	DADIQAISQTTALLRYVKEEK				VEDVFERFRK
5. CasY5					
6. CasY6					
7. CasY7	HSEQSAVVVLLALIGYIADKEG		MSGK-KLVVVR		LAEELMAEWKLLK---
8. AsCpf1_5843	DADANGAYHIALKGQL				LNLHLKESKDKLKLQN
7. LbCpf1_SID6	NADANGAYNIAKVLW				AITGQFKKAEDEKLDKVKI---

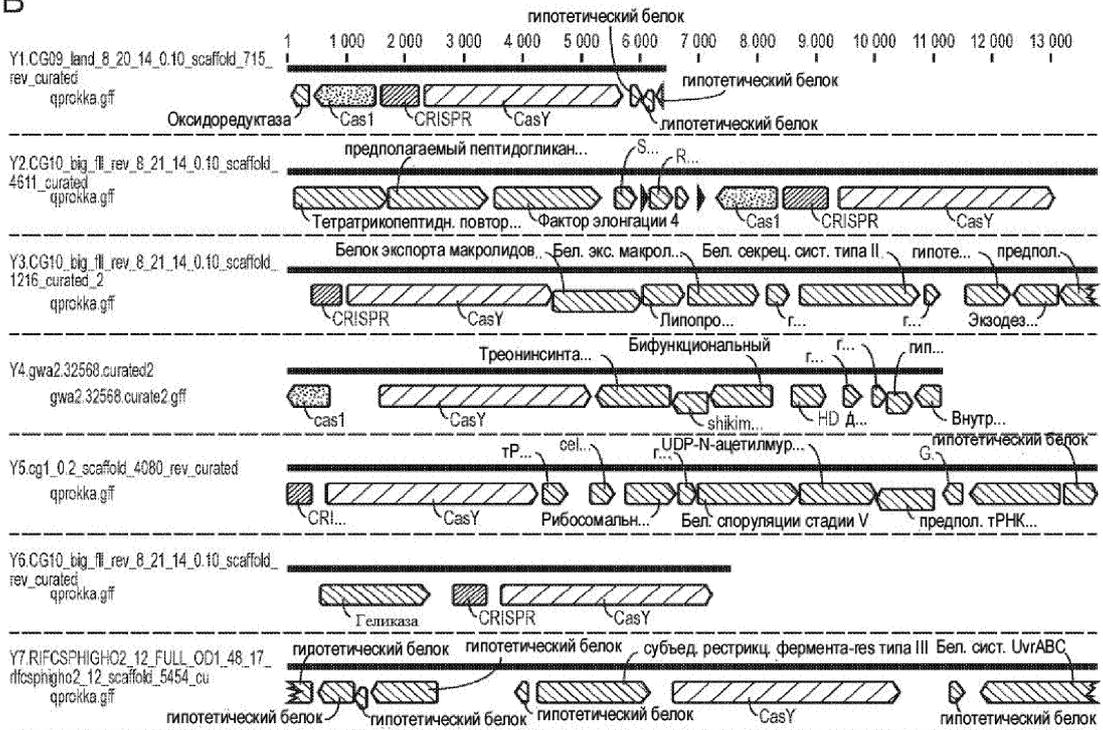
	1 119	1 125
1. CasY1	RSAKF	
2. CasY2	KDVPINKNLITSLDINLL	
3. CasY3		
4. CasY4	KNIKVIG	
5. CasY5		
6. CasY6		
7. CasY7	ERSRVEE	
8. AsCpf1_5843	GISNQDWLAYIQELRN	
7. LbCpf1_SID6	ASNKEWLEYAQTIVK	

Фиг. 2

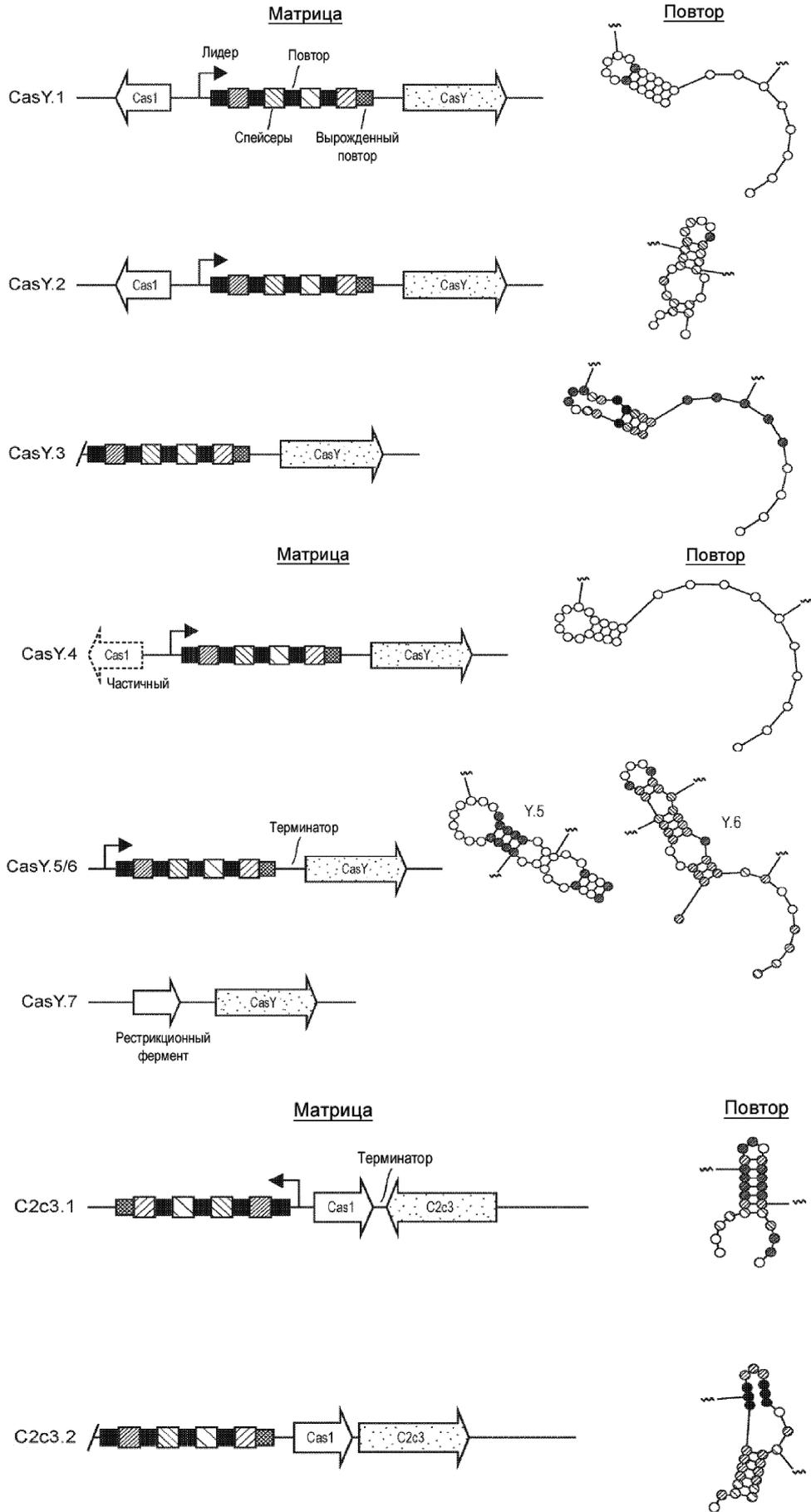
A



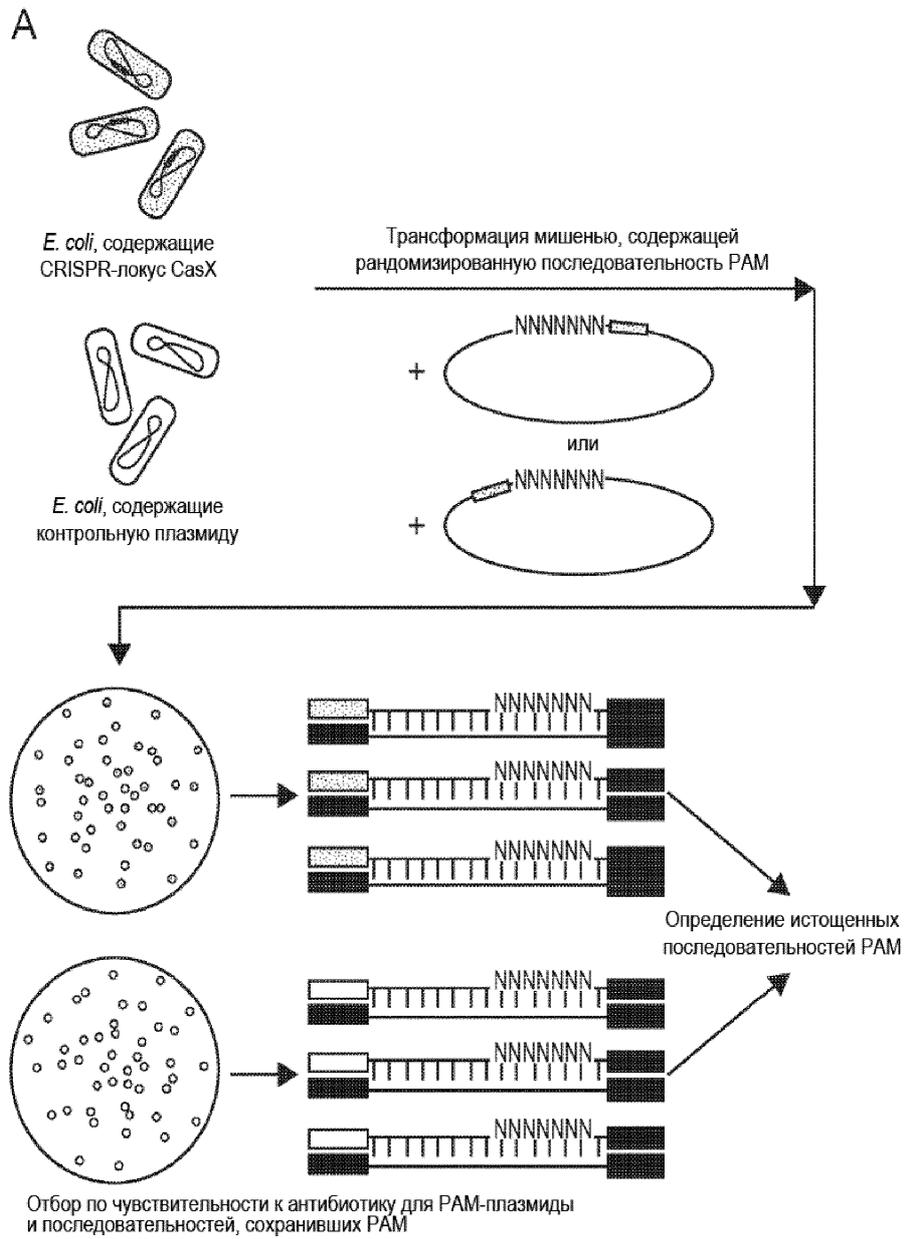
B

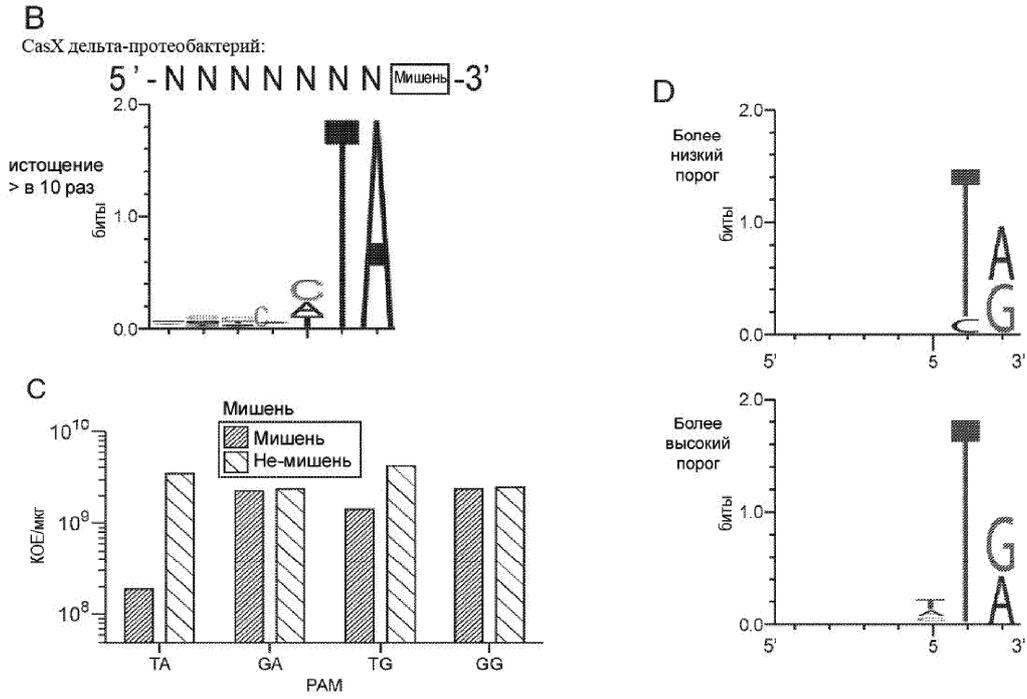


Фиг. 3



Фиг. 4





Фиг. 5

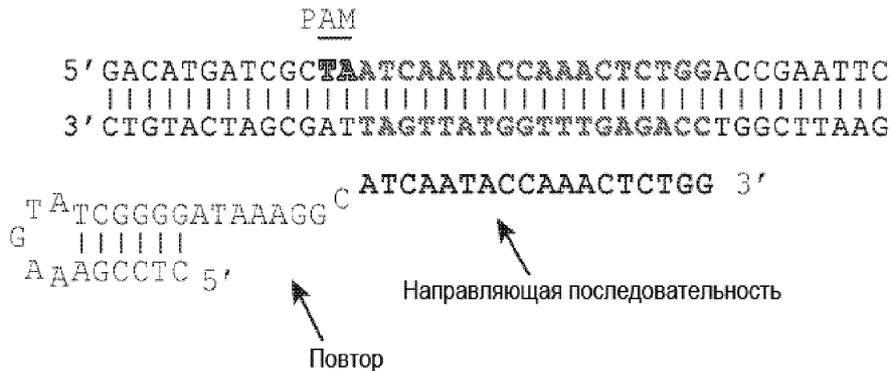
**А**

Повтор

CasY1 CTCCGAAAGTATCGGGGATAAAGGC  
 CasY2 CACCGAAATTTGGAGAGGATAAGGC  
 CasY3 CTCCGAATTATCGGGAGGATAAGGC  
 CasY4 CCCCGAATATAGGGGACAAAAGGC

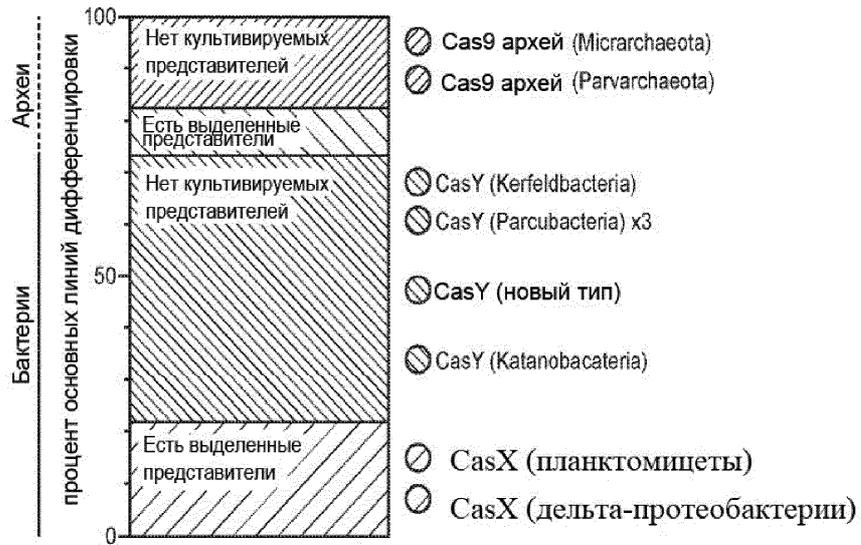
CasY5 GTCTAGACATACAGGTGAAAGGTGAGAGTAAAGAC  
 (и Y6)

**В**

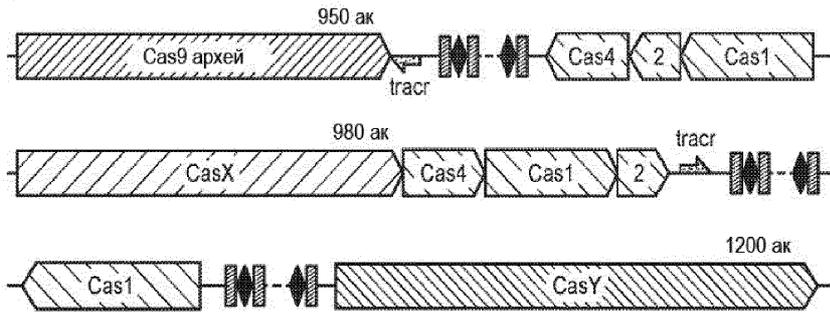


Фиг. 6

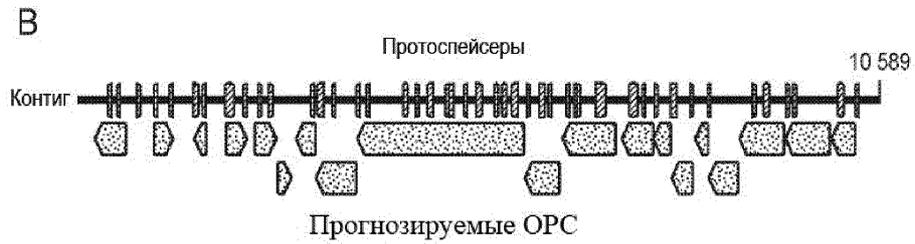
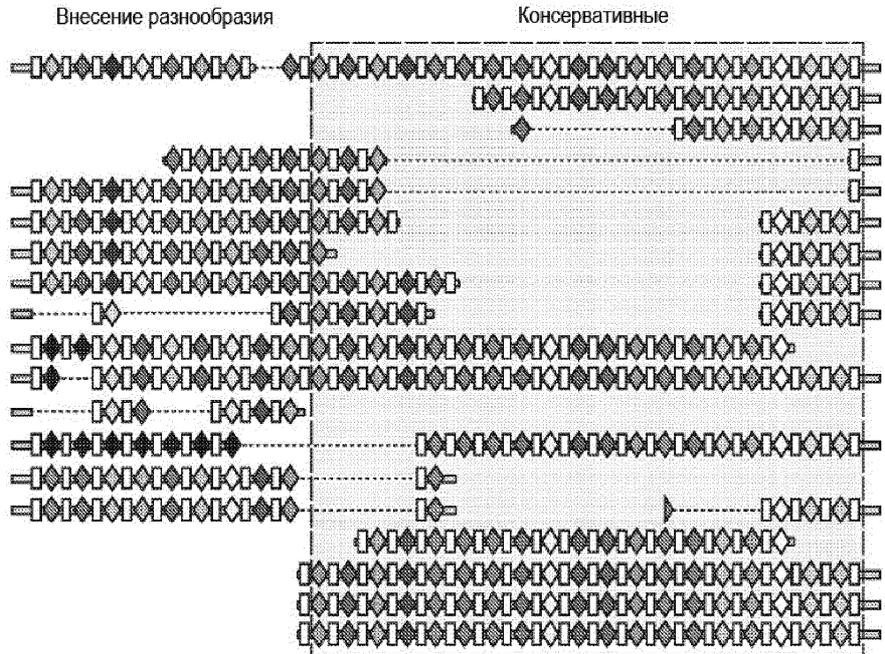
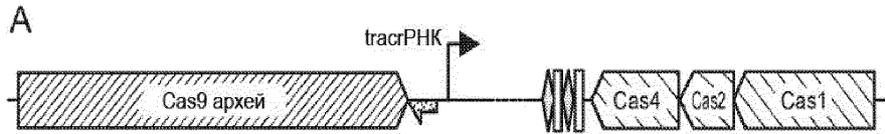
A

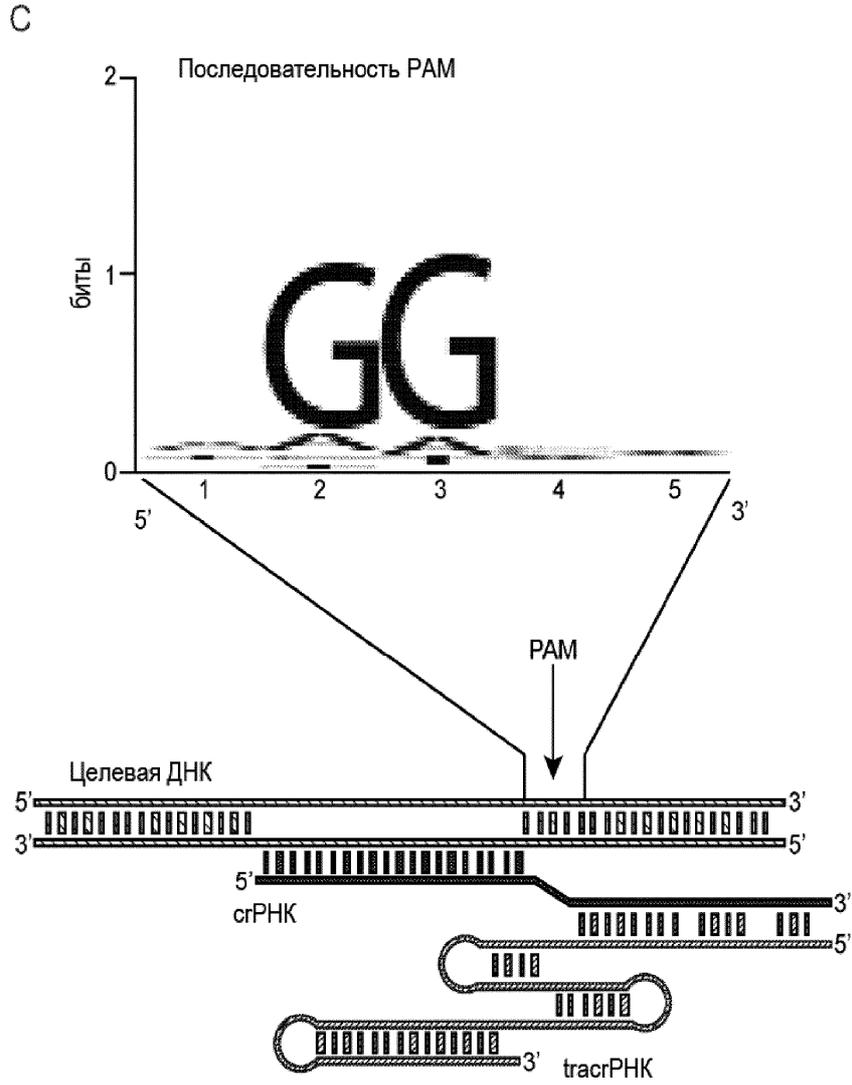


B

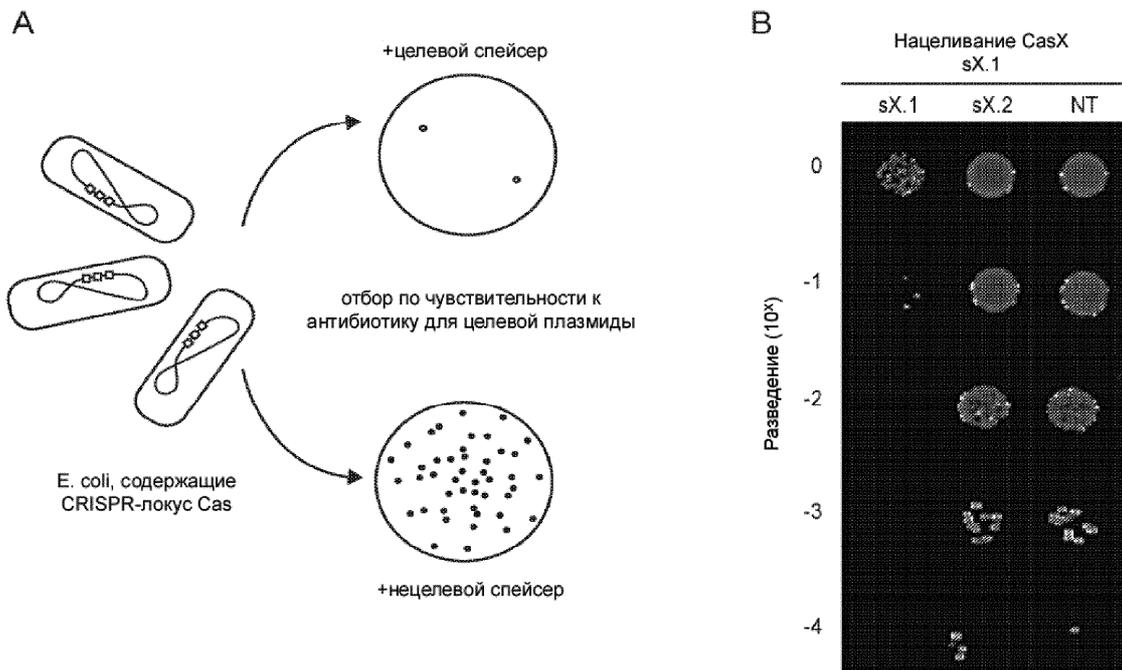


Фиг. 7

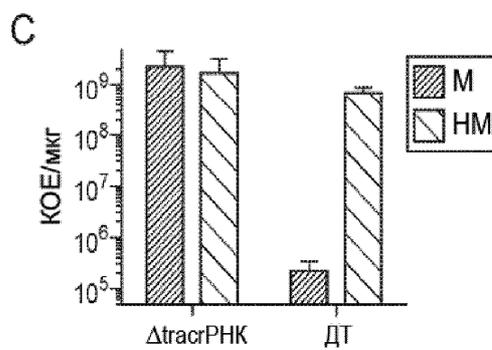
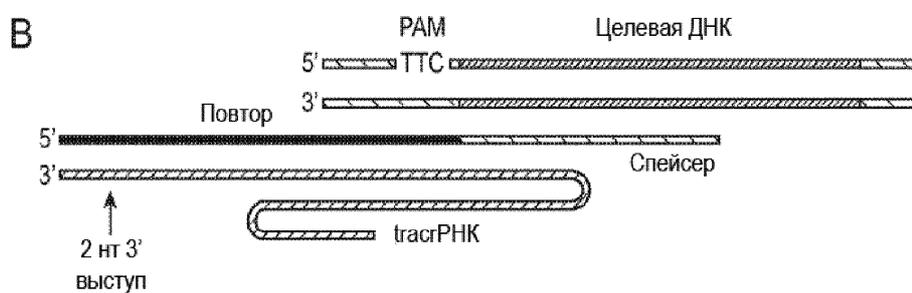
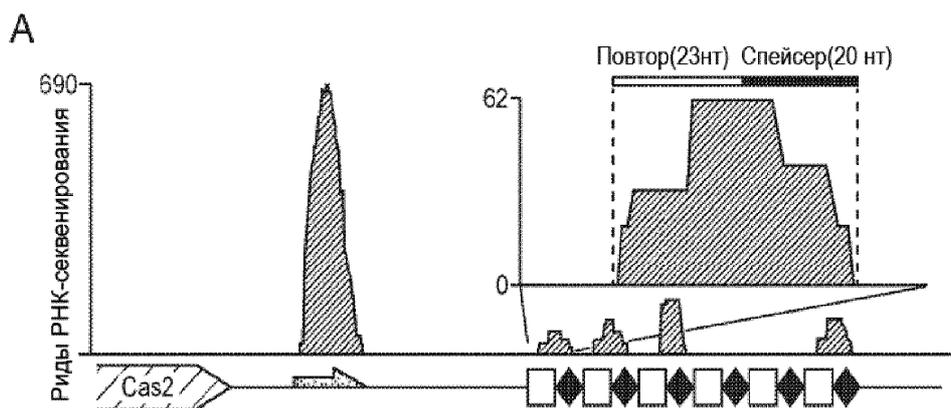




Фиг. 8

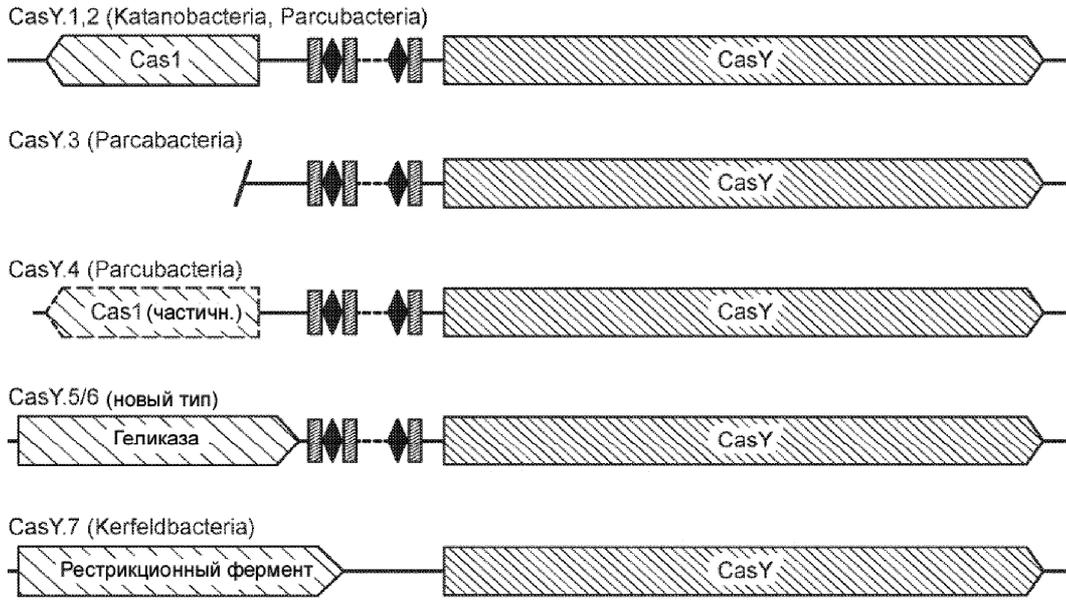




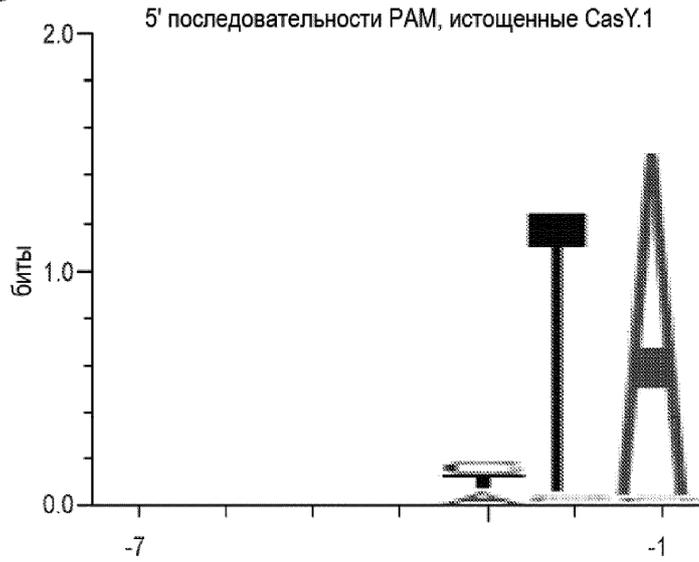


Фиг. 10

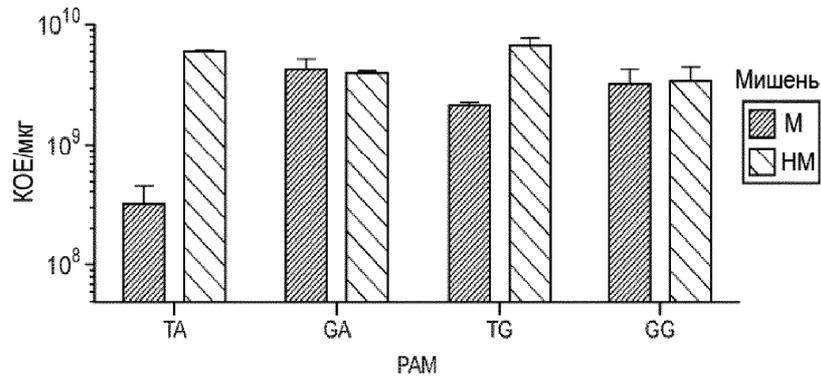
A



B

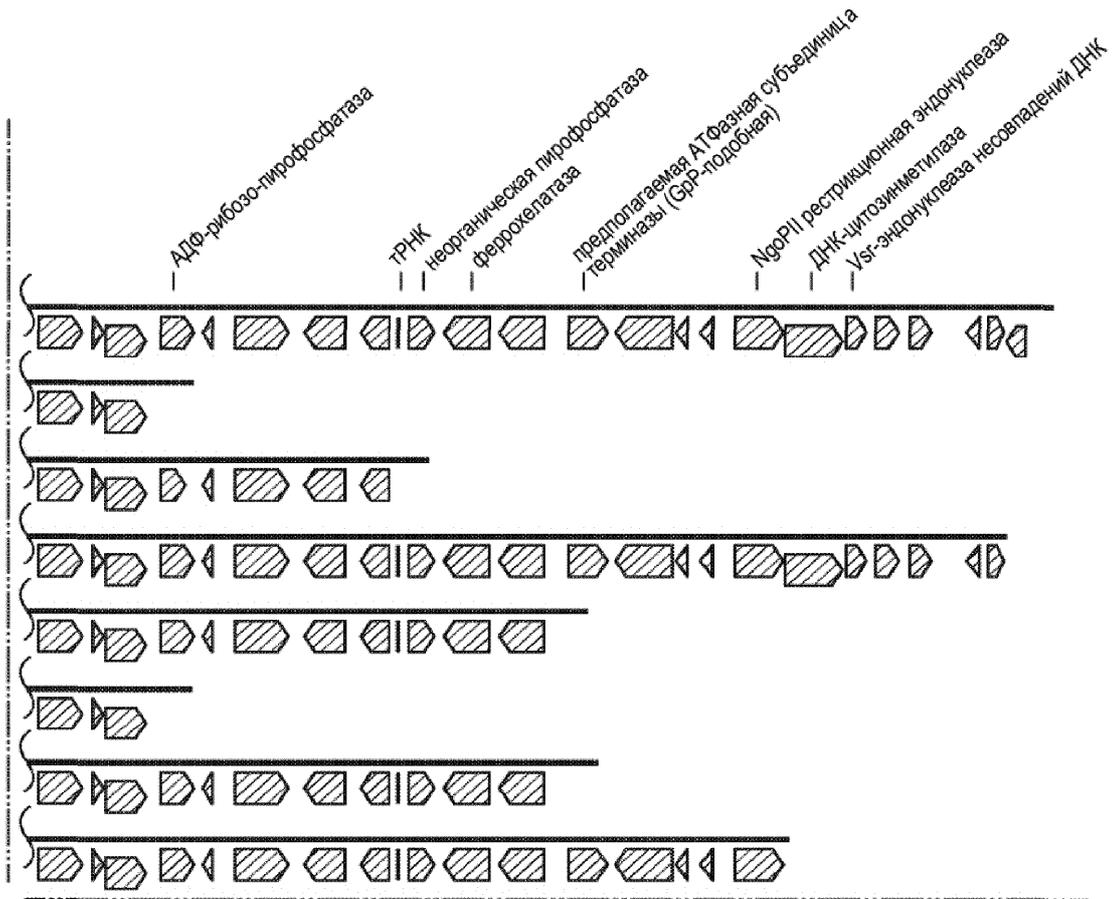
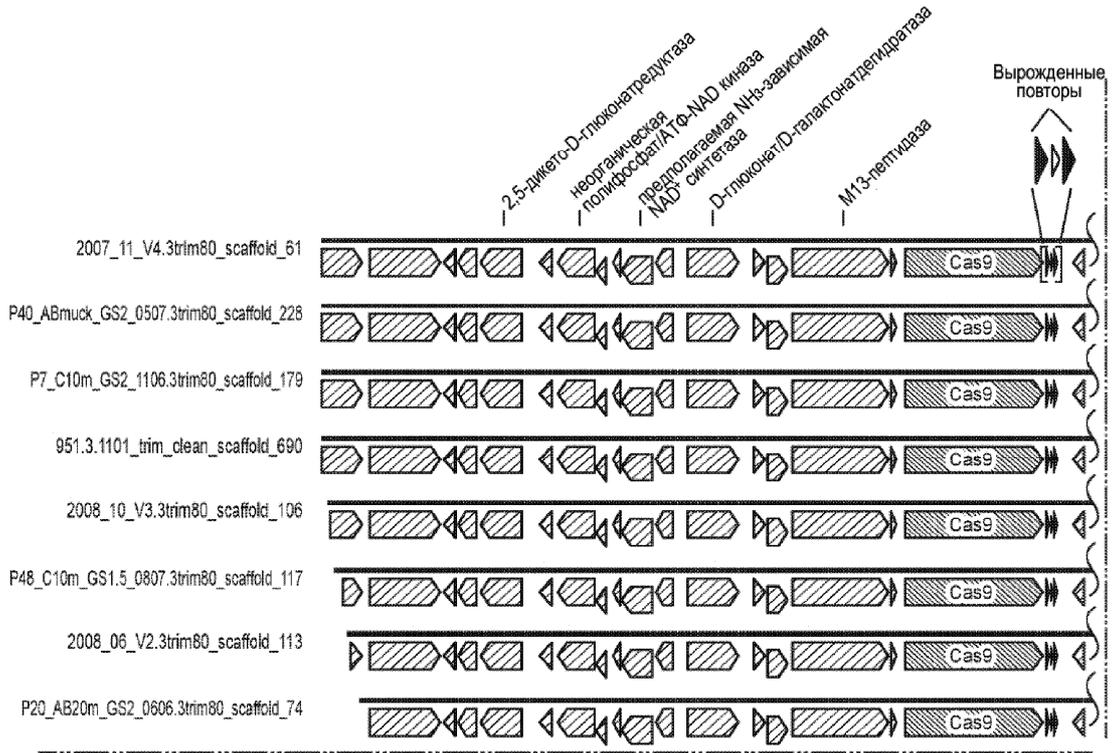


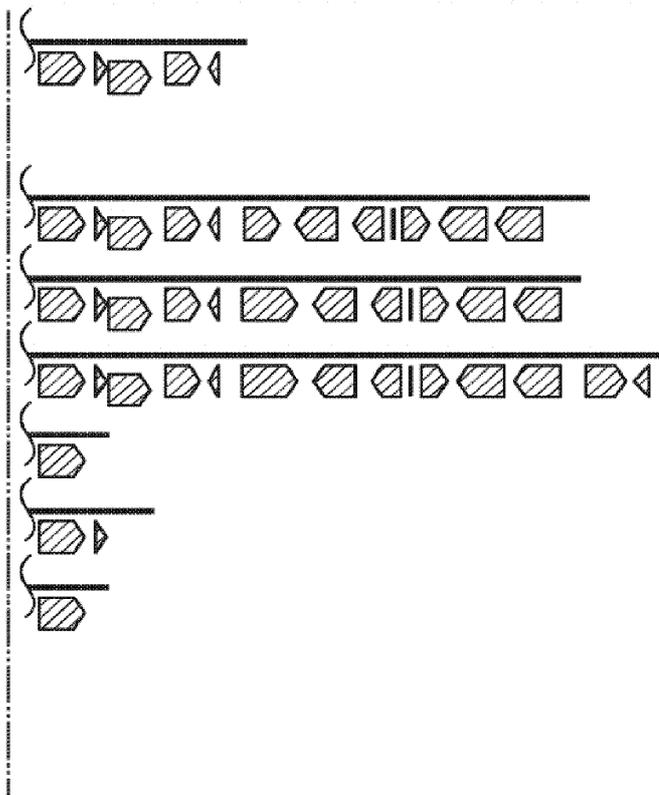
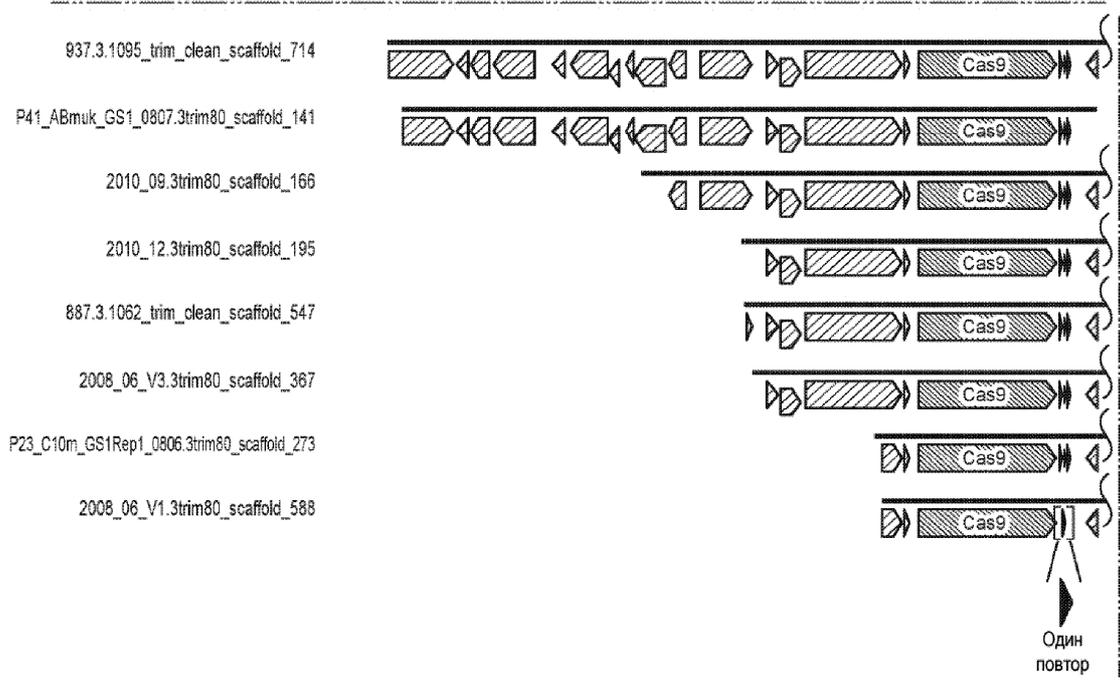
C



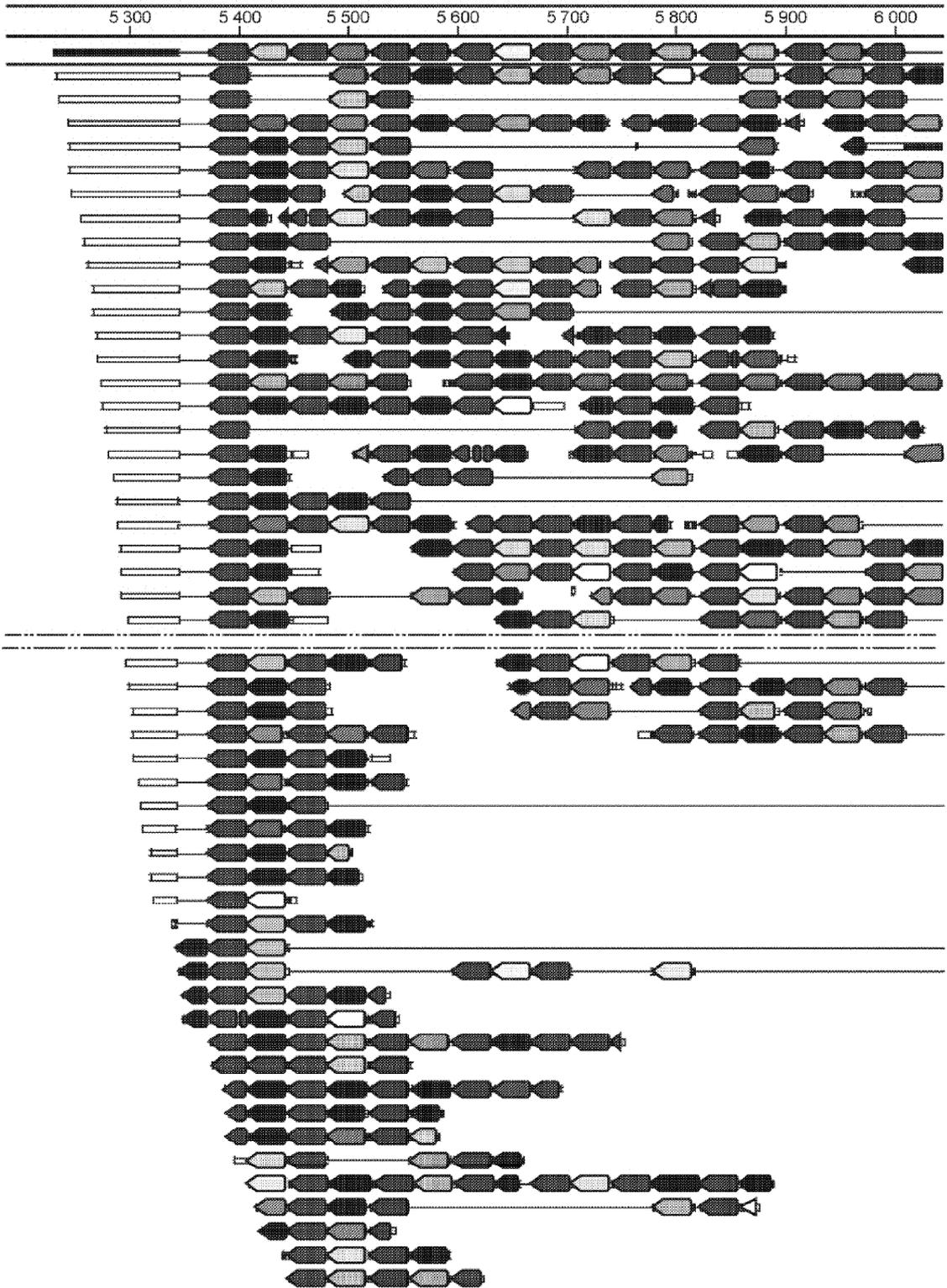
Фиг. 11

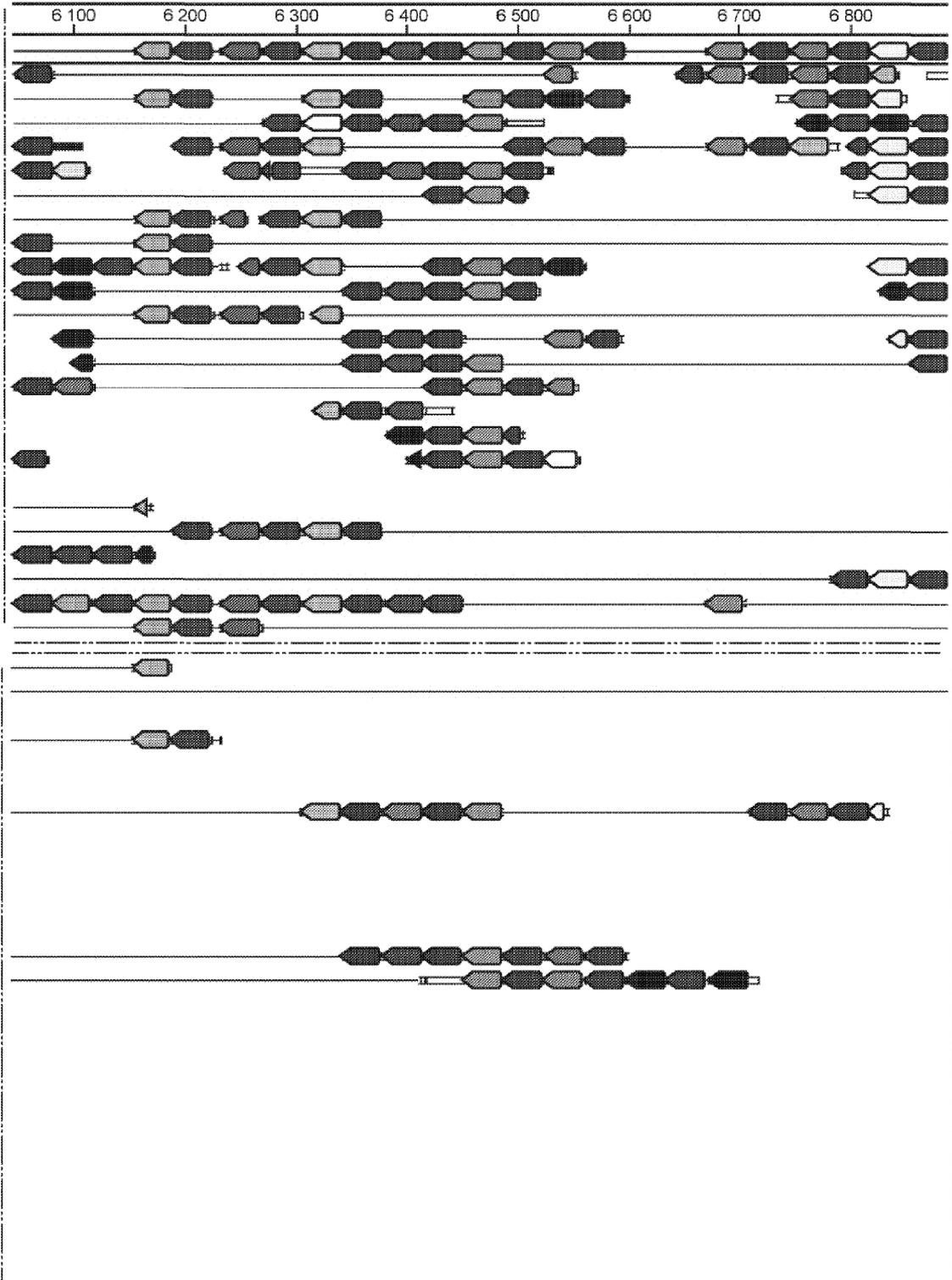


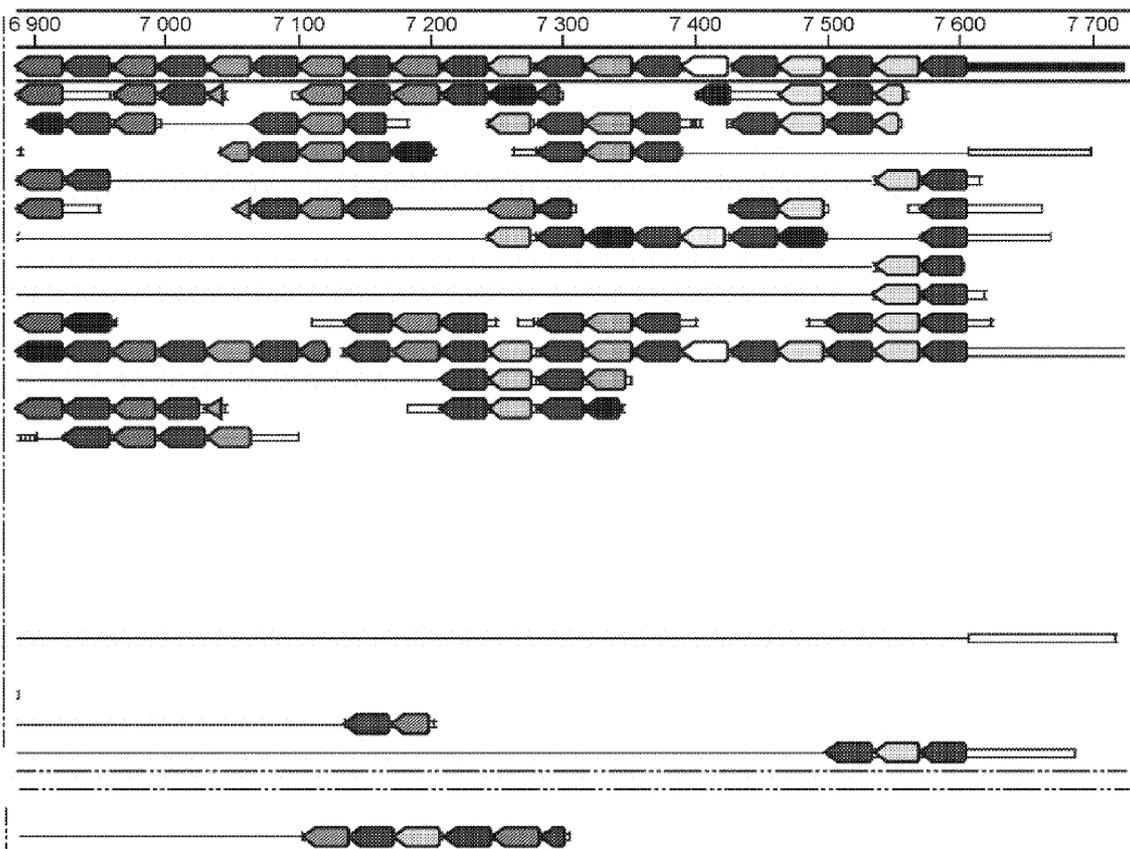




Фиг. 13



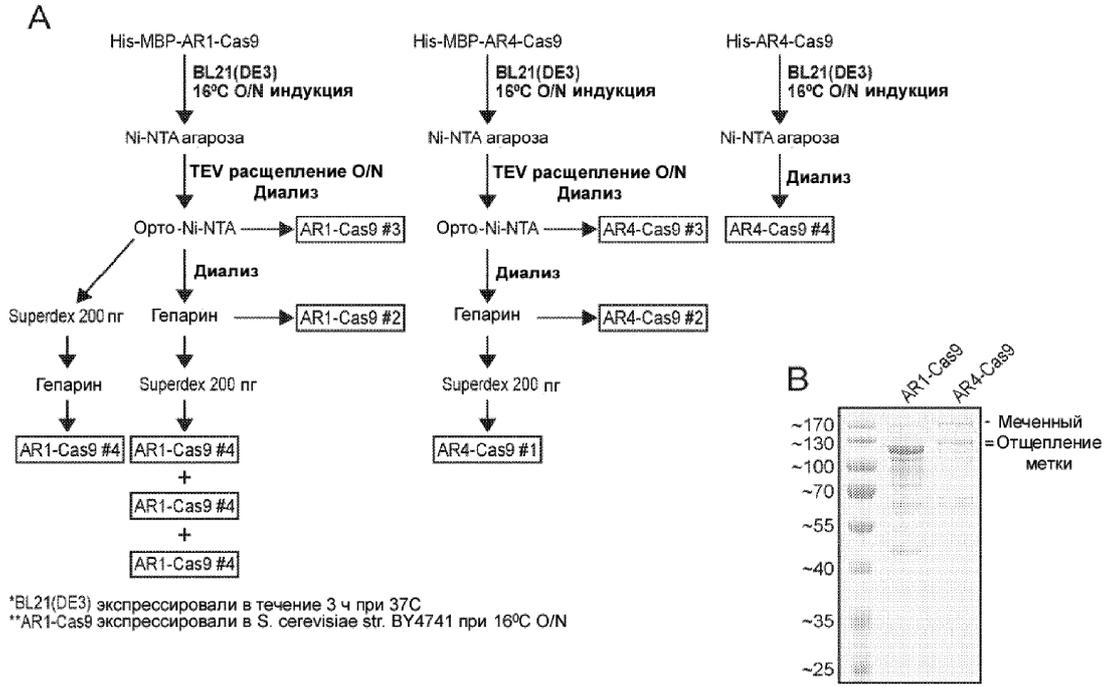




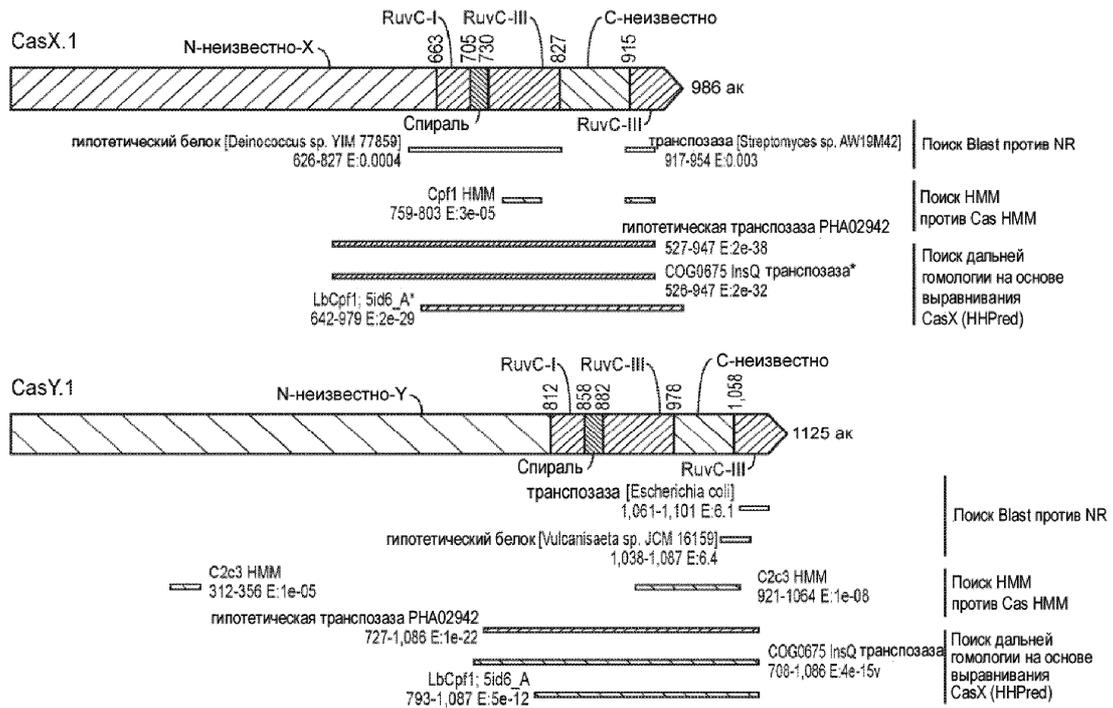
Фиг. 14



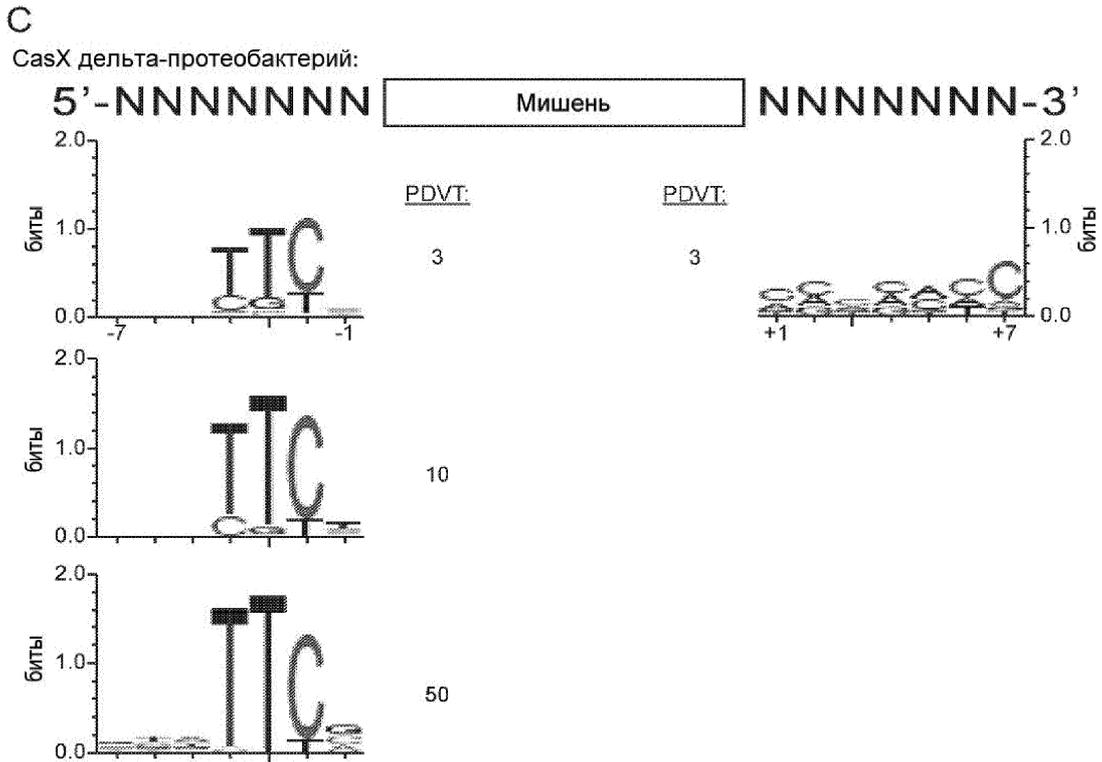
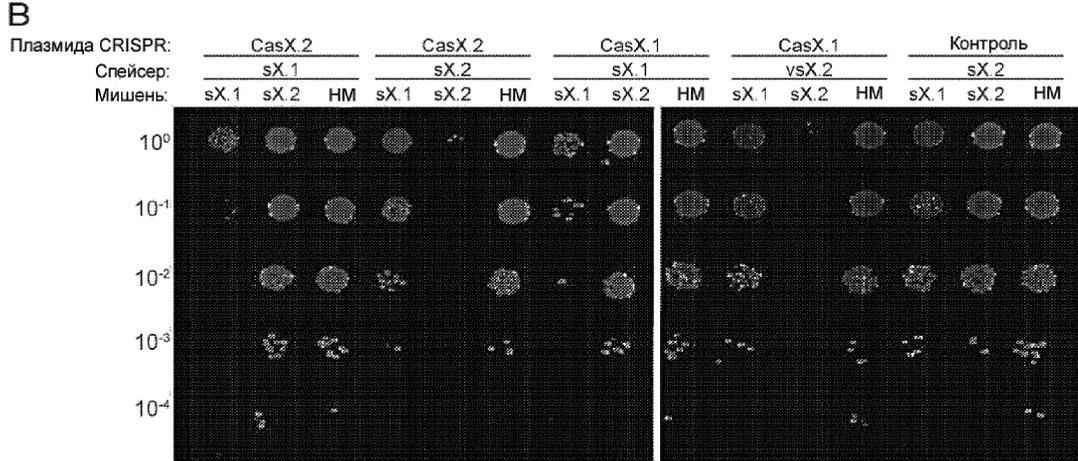
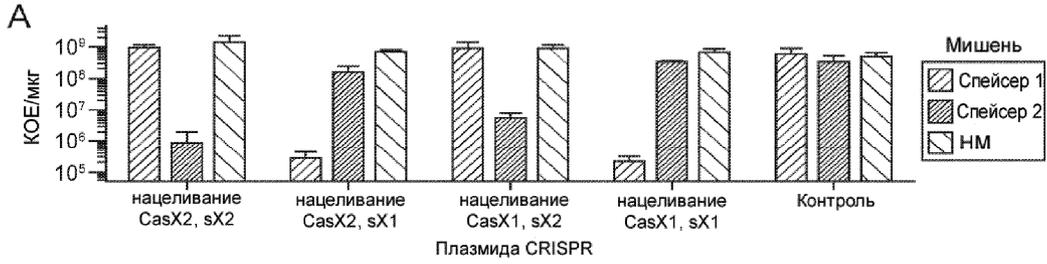




Фиг. 17



Фиг. 18





## In vitro условия расщепления для анализа Cas9 из ARMAN-1

Очистка белка	Буфер	Соль (мМ)	Металл	Направл.	Мишень	Температура
AR1-Cas9 #1	Трис pH 7.5	300	Mg <sup>2+</sup> Mn <sup>2+</sup> Zn <sup>2+</sup>	сгРНК сг:69 сг:69 сг:69	дцДНК оцДНК ДНК-пузырек оцРНК дцРНК	37
AR1-Cas9 #1	Трис pH 7.5	100-500	Mg <sup>2+</sup>	сг:69 сг:104 сг:179	дцДНК	37
AR1-Cas9 #1	Трис pH 7.5	300	Mg <sup>2+</sup> Mn <sup>2+</sup> Zn <sup>2+</sup>	сг:69 сг:104 сг:179	дцДНК	30-48
AR1-Cas9 #1	MOPS: pH 6 pH 6.5 pH 7.0 pH 7.5	300	Mg <sup>2+</sup>	сг:69 сг:104 сг:179	дцДНК	37
AR1-Cas9 #1	Цитрат pH 5 pH 5.5 pH 6	300	Mg <sup>2+</sup>	сг:69 сг:104 сг:179	дцДНК	37
AR1-Cas9 #1	Трис pH 7.5	300	Mg <sup>2+</sup> Mn <sup>2+</sup> Zn <sup>2+</sup>	сг:69 сг:104 сг:179	плазида	37-50
AR1-Cas9 #2	Трис pH 7.5	300	Mg <sup>2+</sup> Mn <sup>2+</sup> Zn <sup>2+</sup>	сг:69 сг:104 сг:179	дцДНК	37
AR1-Cas9 #3	Трис pH 7.5	300	Mg <sup>2+</sup> Mn <sup>2+</sup> Zn <sup>2+</sup>	сг:69 сг:104 сг:179	дцДНК	37
AR1-Cas9 #4	Трис pH 7.5	300	Mg <sup>2+</sup> Mn <sup>2+</sup> Zn <sup>2+</sup>	сг:69 сг:104 сг:179	дцДНК	37
AR1-Cas9 #5	Трис pH 7.5	300	Mg <sup>2+</sup> Mn <sup>2+</sup> Zn <sup>2+</sup>	сг:69 сг:104 сг:179	дцДНК	37
AR1-Cas9 #6	Трис pH 7.5	300	Mg <sup>2+</sup> Mn <sup>2+</sup> Zn <sup>2+</sup>	сг:69 сг:104 сг:179	оцДНК дцДНК	37
AR4-Cas9 #1	Трис pH 7.5	300	Mg <sup>2+</sup> Mn <sup>2+</sup> Zn <sup>2+</sup>	сгРНК- 122	дцДНК	37
AR4-Cas9 #2	Трис pH 7.5	300	Mg <sup>2+</sup> Mn <sup>2+</sup> Zn <sup>2+</sup>	сгРНК- 122	дцДНК	37
AR4-Cas9 #3	Трис pH 7.5	300	Mg <sup>2+</sup> Mn <sup>2+</sup> Zn <sup>2+</sup>	сгРНК- 122	дцДНК	37
AR4-Cas9 #4	Трис pH 7.5	300	Mg <sup>2+</sup> Mn <sup>2+</sup> Zn <sup>2+</sup>	сгРНК- 122	дцДНК	37

Фиг. 21

