

(19)



**Евразийское  
патентное  
ведомство**

(11) **038924**(13) **B1**(12) **ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ**

(45) Дата публикации и выдачи патента  
**2021.11.10**

(21) Номер заявки  
**201401319**

(22) Дата подачи заявки  
**2013.03.15**

(51) Int. Cl. **C12N 15/11** (2006.01)  
**C12N 15/63** (2006.01)  
**C07K 19/00** (2006.01)  
**C12N 5/10** (2006.01)

---

(54) **СПОСОБЫ И КОМПОЗИЦИИ РНК-СПЕЦИФИЧЕСКОЙ МОДИФИКАЦИИ ДНК-МИШЕНИ И РНК-СПЕЦИФИЧЕСКОЙ МОДУЛЯЦИИ ТРАНСКРИПЦИИ**

---

(31) **61/652,086; 61/716,256; 61/757,640;  
61/765,576**

(32) **2012.05.25; 2012.10.19; 2013.01.28;  
2013.02.15**

(33) **US**

(43) **2015.05.29**

(86) **PCT/US2013/032589**

(87) **WO 2013/176772 2013.11.28**

(71)(73) Заявитель и патентовладелец:

**ТЕ РИДЖЕНТС ОФ ТЕ  
ЮНИВЕРСИТИ ОФ КАЛИФОРНИЯ  
(US); ЮНИВЕРСИТИ ОФ ВЬЕНА  
(AT); ШАРПЕНТЬЕ ЭММАНЮЭЛЬ  
(DE)**

(72) Изобретатель:

**Дудна Дженифер А., Йинек Мартин  
(US), Шарпентье Эмманюэль (DE),  
Хылинский Кшиштоф (AT), Дудна  
Кейт Джеймс Хэррисон, Лим Вендел,  
Ци Лэй (US)**

(74) Представитель:

**Фелицына С.Б. (RU)**

(56) SASHITAL, DIPALI G. et al., "Mechanism of foreign DNA selection in a bacterial adaptive immune system", *Molecular Cell*, Epub. 19 April 2012, Vol. 46, No. 5, p. 606-615, ISSN 1097-2765. See the whole document

WIEDENHEFT, BLAKE et al., "RNA-guided genetic silencing systems in bacteria and archaea", *Nature*, 15 February 2012, Vol. 482, No. 7385, p. 331-338, ISSN 0028-0836. See the whole document

HALE, CARYN R. et al., "RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex", *Cell*, 25 November 2009, Vol. 139, No. 5, p. 945-956, ISSN 0092-8674. See the whole document

US-A1-20100076057

US-A1-20110189776

JINEK, MARTIN et al., "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity", *Science*, 17 August 2012, Vol. 337, No. 6096, p. 816-821, ISSN 0036-8075. See the whole document

(57) В изобретении предлагается РНК, нацеленная на ДНК, при этом РНК содержит специфическую последовательность и вместе с модифицирующим полипептидом обеспечивает сайт-специфическую модификацию ДНК-мишени и/или полипептида, ассоциированного с ДНК-мишенью. Настоящее изобретение дополнительно обеспечивает сайт-специфические модифицирующие полипептиды. Настоящее изобретение дополнительно обеспечивает способы сайт-специфической модификации ДНК-мишени и/или полипептида, ассоциированного с ДНК-мишенью. Настоящее изобретение обеспечивает способы модулирования транскрипции нуклеиновой кислоты-мишени в клетке-мишени, как правило, путем приведения нуклеиновой кислоты-мишени в контакт с ферментативно неактивным полипептидом Cas9 и РНК, нацеленной на ДНК. Также предложены наборы и композиции для осуществления указанных способов. Настоящее изобретение обеспечивает генетически модифицированные клетки, которые продуцируют Cas9; и Cas9 трансгенные многоклеточные организмы, не представляющие собой человека.

**B1****038924****038924****B1**

### **Перекрестная ссылка**

В заявке на данный патент заявляется приоритет по предварительным заявкам на выдачу патентов США с серийными номерами 61/652086, поданной 25 мая 2012 г., 61/716256, поданной 19 октября 2012 г., 61/757640, поданной 28 января 2013 г., и 61/765576, поданной 15 февраля 2013 г., каждая заявка из которых содержится в данном описании в виде ссылки во всей своей полноте.

### **Заявление относительно исследования, финансируемого из федерального бюджета**

Настоящее изобретение было сделано при поддержке правительства с помощью гранта № GM081879, полученного от National Institutes of Health. Правительство имеет определенные права на это изобретение.

### **Перечень последовательностей, который содержится в виде ссылки, обеспечивают в виде текстового файла**

Перечень последовательностей обеспечивают при этом в виде текстового файла, "BERK-187WO-SeqList\_ST25.txt", созданного 13 марта 2013 г. и имеющего размер 7645 Кб. Содержимое текстового файла содержится в данном документе в виде ссылки во всей своей полноте.

### **Уровень техники**

Около 60% бактерий и 90% архей обладают системами CRISPR (кластеризованные короткие палиндромные повторы, разделенные регулярными промежутками)/CRISPR-ассоциированной (Cas) системой для придания устойчивости к элементам чужеродной ДНК. Система CRISPR типа II из *Streptococcus pyogenes* содержит только один ген, кодирующий белок Cas9, и две РНК - зрелую РНК CRISPR (crРНК) и частично комплементарную трансдействующую РНК (tracrРНК) - которые являются необходимыми и достаточными для РНК-направленного сайленсинга чужеродных ДНК.

В последние годы созданные ферменты нуклеазы, предназначенные для направленного воздействия на специфические последовательности-мишени ДНК, привлекли значительное внимание в качестве мощных инструментов для генетической манипуляции над клетками и целыми организмами, и они позволяют осуществлять целевые делеции, замену и репарацию гена, а также вводить экзогенные последовательности (трансгены) в геном. Были разработаны две основные технологии создания сайт-специфических ДНК-нуклеаз, обе из которых основаны на конструировании гибридных ферментов эндонуклеаз, в которых неспецифичный к последовательности домен ДНК-эндонуклеазы слит с созданным доменом связывания ДНК. Тем не менее, нацеливание на каждый новый геномный локус требует создание нового фермента-нуклеазы, делая оба эти подхода требующими много времени и дорогостоящими. Кроме того, обе технологии страдают от ограниченной точности, что может привести к непрогнозируемым и внецелевым эффектам.

Систематический анализ геномов и генетическое перепрограммирование клеток предусматривает направленное воздействие на множество генов для экспрессии или репрессии. В настоящее время наиболее распространенным подходом для направленного воздействия на произвольные гены с целью регулирования является использование РНК-интерференции (РНКи). Этот подход имеет ограничения. Например, РНКи может проявлять значительные внецелевые эффекты и токсичность.

В данной области существует потребность в создании технологии, которая бы позволяла осуществлять точное нацеливание нуклеазной активности (или других белковых активностей) на различные месторасположения в пределах ДНК-мишени таким образом, который не требует создания нового белка для каждой новой последовательности-мишени. Кроме того, в данной области техники существует потребность в способах контроля экспрессии генов с минимальными внецелевыми эффектами.

### **Краткое описание изобретения**

Настоящее изобретение обеспечивает РНК, нацеленную на ДНК, которая содержит нацеливающую последовательность и, вместе с модифицирующим полипептидом, обеспечивает сайт-специфическую модификацию ДНК-мишени и/или полипептида, ассоциированного с ДНК-мишенью. Настоящее изобретение дополнительно обеспечивает сайт-специфические модифицирующие полипептиды. Настоящее изобретение дополнительно обеспечивает способы сайт-специфической модификации ДНК-мишени и/или полипептида, ассоциированного с ДНК-мишенью. Настоящее изобретение обеспечивает способы модулирования транскрипции нуклеиновой кислоты-мишени в клетке-мишени, как правило, включающий приведение в контакт нуклеиновой кислоты-мишени с ферментативно неактивным полипептидом Cas9 и РНК, нацеленной на ДНК.

Также обеспечивают наборы и композиции для осуществления способов. Настоящее изобретение обеспечивает генетически модифицированные клетки, которые продуцируют Cas9; и трансгенные многоклеточные организмы Cas9, которые не являются человеком.

### **Признаки изобретения**

Признаки настоящего изобретения включают РНК, нацеленную на ДНК, которая содержит (i) первый сегмент, содержащий нуклеотидную последовательность, которая является комплементарной последовательности в ДНК-мишени; и (ii) второй сегмент, который взаимодействует с сайт-специфическим модифицирующим полипептидом. В некоторых случаях первый сегмент содержит 8 нуклеотидов, которые имеют комплементарность 100% по отношению к последовательности в ДНК-мишени. В некоторых случаях второй сегмент содержит нуклеотидную последовательность с идентичностью по меньшей мере

60% на участке из по меньшей мере 8 последовательных нуклеотидов по отношению к любой из нуклеотидных последовательностей, приведенных в SEQ ID NO:431-682 (например, 431-562). В некоторых случаях второй сегмент содержит нуклеотидную последовательность с идентичностью по меньшей мере 60% на участке из по меньшей мере 8 последовательных нуклеотидов по отношению к любой из нуклеотидных последовательностей, приведенных в SEQ ID NO:563-682. В некоторых случаях сайт-специфический модифицирующий полипептид содержит аминокислотную последовательность с идентичностью по меньшей мере около 75% аминокислотной последовательности с аминокислотами 7-166 или 731-1003 аминокислотной последовательности Cas9/Csn1, изображенной на фиг. 3, или соответствующим участкам любой из аминокислотных последовательностей, приведенных в SEQ ID NO:1-256 и 795-1346.

Признаки настоящего изобретения включают полинуклеотид ДНК, который содержит нуклеотидную последовательность, кодирующую РНК, нацеленную на ДНК. В некоторых случаях рекомбинантный вектор экспрессии содержит полинуклеотид ДНК. В некоторых случаях нуклеотидная последовательность, кодирующая РНК, нацеленную на ДНК, функционально связана с промотором. В некоторых случаях промотор является индуцируемым промотором. В некоторых случаях нуклеотидная последовательность, кодирующая РНК, нацеленную на ДНК, дополнительно содержит сайт множественного клонирования. Признаки настоящего изобретения содержат клетку-хозяина, генетически модифицированную *in vitro*, которая содержит полинуклеотид ДНК.

Признаки настоящего изобретения включают рекомбинантный вектор экспрессии, который содержит (i) нуклеотидную последовательность, кодирующую РНК, нацеленную на ДНК, при том, что РНК, нацеленная на ДНК, содержит (a) первый сегмент, содержащий нуклеотидную последовательность, комплементарную последовательности в ДНК-мишени; и (b) второй сегмент, который взаимодействует с сайт-специфическим модифицирующим полипептидом; и (ii) нуклеотидную последовательность, кодирующую сайт-специфический модифицирующий полипептид, который содержит (a) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и (b) участок активности, который проявляет сайт-специфическую ферментативную активность, причем сайт ферментативной активности определяется РНК, нацеленной на ДНК.

Признаки настоящего изобретения включают рекомбинантный вектор экспрессии, который содержит (i) нуклеотидную последовательность, кодирующую РНК, нацеленную на ДНК, при том, что РНК, нацеленная на ДНК, содержит (a) первый сегмент, содержащий нуклеотидную последовательность, комплементарную последовательности в ДНК-мишени; и (b) второй сегмент, который взаимодействует с сайт-специфическим модифицирующим полипептидом; и (ii) нуклеотидную последовательность, кодирующую сайт-специфический модифицирующий полипептид, при том, что сайт-специфический модифицирующий полипептид содержит (a) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и (b) участок активности, который модулирует транскрипцию в пределах ДНК-мишени, причем сайт модулированной транскрипции в пределах ДНК-мишени определяется РНК, нацеленной на ДНК.

Признаки настоящего изобретения включают вариант сайт-специфического модифицирующего полипептида, который содержит (i) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК, при том, что РНК, нацеленная на ДНК, содержит нуклеотидную последовательность, комплементарную последовательности в ДНК-мишени; и (ii) участок активности, который проявляет сниженную сайт-специфическую ферментативную активность, причем сайт ферментативной активности определяется РНК, нацеленной на ДНК. В некоторых случаях вариант сайт-специфического модифицирующего полипептида содержит мутацию H840A в последовательности *S. pyogenes* SEQ ID NO:8 или соответствующую мутацию в любой из аминокислотных последовательностей, приведенных в SEQ ID NO:1-256 и 795-1346. В некоторых случаях вариант сайт-специфического модифицирующего полипептида содержит мутацию D10A в последовательности *S. pyogenes* SEQ ID NO:8 или соответствующую мутацию в любой из аминокислотных последовательностей, приведенных в SEQ ID NO:1-256 и 795-1346. В некоторых случаях вариант сайт-специфического модифицирующего полипептида содержит как (i) мутацию D10A в последовательности *S. pyogenes* SEQ ID NO:8 или соответствующую мутацию в любой из аминокислотных последовательностей, приведенных в SEQ ID NO:1-256 и 795-1346; так и (ii) мутацию H840A в последовательности *S. pyogenes* SEQ ID NO:8 или соответствующую мутацию в любой из аминокислотных последовательностей, приведенных в SEQ ID NO:1-256 и 795-1346.

Признаки настоящего изобретения включают вариант гибридного сайт-специфического модифицирующего полипептида, который содержит (i) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК, при том, что РНК, нацеленная на ДНК, содержит нуклеотидную последовательность, комплементарную последовательности в ДНК-мишени; и (ii) участок активности, который проявляет сайт-специфическую ферментативную активность, причем сайт ферментативной активности определяется РНК, нацеленной на ДНК. В некоторых случаях гибридный сайт-специфический модифицирующий полипептид содержит аминокислотную последовательность с идентичностью по меньшей мере около 75% аминокислотной последовательности с аминокислотами 7-166 или 731-1003 аминокислотной последовательности Cas9/Csn1, изображенной на фиг. 3, или с соответствующими участками лю-

бой из аминокислотных последовательностей, приведенных в SEQ ID NO:1-256 и 795-1346. В некоторых случаях РНК, нацеленная на ДНК, дополнительно содержит нуклеотидную последовательность с идентичностью по меньшей мере 60% на участке из по меньшей мере 8 последовательных нуклеотидов по отношению к любой из нуклеотидных последовательностей, представленных в виде SEQ ID NO:431-682 (например, SEQ ID NO:563-682). В некоторых случаях РНК, нацеленная на ДНК, дополнительно содержит нуклеотидную последовательность с идентичностью по меньшей мере 60% на участке из по меньшей мере 8 последовательных нуклеотидов по отношению к любой из нуклеотидных последовательностей, представленных в виде SEQ ID NO:431-562. В некоторых случаях ферментативная активность гибридного сайт-специфического модифицирующего полипептида изменяет ДНК-мишень. В некоторых случаях ферментативная активность гибридного сайт-специфического модифицирующего полипептида представляет собой нуклеазную активность, активность метилтрансферазы, активность деметилазы, активность репарации ДНК, активность повреждения ДНК, активность дезаминирования, активность дисмутазы, алкилирующую активность, активность депуринизации, активность окисления, активность формирования димера пиримидина, активность интегразы, активность транспозазы, активность рекомбиназы, полимеразную активность, лигазную активность, геликазную активность, активность фотолитазы или гликозилазную активность. В некоторых случаях ферментативная активность гибридного сайт-специфического модифицирующего полипептида представляет собой нуклеазную активность. В некоторых случаях нуклеазная активность вводит двухцепочечный разрыв в ДНК-мишень. В некоторых случаях ферментативная активность гибридного сайт-специфического модифицирующего полипептида изменяет полипептид-мишень, ассоциированный с ДНК-мишенью. В некоторых случаях ферментативная активность гибридного сайт-специфического модифицирующего полипептида является активностью метилтрансферазы, активностью деметилазы, ацетилтрансферазной активностью, активностью деацетилазы, киназной активностью, фосфатазной активностью, активностью убиквитинлигазы, деубиквитинирующей активностью, активностью аденилирования, активностью деаденилирования, активностью сумоилирования, активностью десумоилирования, активностью рибозилирования, активностью дерибозилирования, активностью миристоилирования или активностью демиристоилирования.

Признаки настоящего изобретения включают полинуклеотид, содержащий нуклеотидную последовательность, кодирующую гибридный сайт-специфический модифицирующий полипептид. В некоторых случаях полинуклеотид представляет собой полинуклеотид РНК. В некоторых случаях полинуклеотид представляет собой полинуклеотид ДНК. Признаки настоящего изобретения включают рекомбинантный вектор экспрессии, содержащий полинуклеотид. В некоторых случаях полинуклеотид функционально связан с промотором. В некоторых случаях промотор является индуцируемым промотором. Признаки настоящего изобретения включают *in vitro* генетически модифицированную клетку-хозяина, которая содержит полинуклеотид.

Признаки настоящего изобретения включают гибридный сайт-специфический модифицирующий полипептид, который содержит (i) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК, при том что РНК, нацеленная на ДНК, содержит нуклеотидную последовательность, комплементарную последовательности в ДНК-мишени; и (ii) участок активности, который модулирует транскрипцию в пределах ДНК-мишени, причем сайт модулированной транскрипции в пределах ДНК-мишени определяется РНК, нацеленной на ДНК. В некоторых случаях участок активности увеличивает транскрипцию в пределах ДНК-мишени. В некоторых случаях участок активности уменьшает транскрипцию в пределах ДНК-мишени.

Признаки настоящего изобретения включают генетически модифицированную клетку, содержащую рекомбинантный сайт-специфический модифицирующий полипептид, который содержит РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и участок активности, который проявляет сайт-специфическую ферментативную активность, причем сайт ферментативной активности определяется РНК, нацеленной на ДНК. В некоторых случаях сайт-специфический модифицирующий полипептид содержит аминокислотную последовательность с идентичностью по меньшей мере около 75% аминокислотной последовательности с аминокислотами 7-166 или 731-1003 аминокислотной последовательности Cas9/Csn1, изображенной на фиг. 3, или с соответствующими участками любой из аминокислотных последовательностей, приведенных в SEQ ID NO:1-256 и 795-1346. В некоторых случаях клетка выбрана из группы, которая состоит из клетки археи, бактериальной клетки, эукариотической клетки, одноклеточного эукариотического организма, соматической клетки, зародышевой клетки, стволовой клетки, клетки растения, клетки водорослей, клетки животного, клетки беспозвоночных, клетки позвоночных, клетки рыбы, клетки земноводного, клетки птицы, клетки млекопитающего, клетки свиньи, клетки коровы, клетки козы, клетки овцы, клетки грызуна, клетки крысы, клетки мыши, клетки примата, не являющегося человеком, и клетки человека.

Признаки настоящего изобретения включают трансгенный организм животного, не являющегося человеком, геном которого содержит трансген, который содержит нуклеотидную последовательность, кодирующую рекомбинантный сайт-специфический модифицирующий полипептид, который содержит (i) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и (ii) участок активности, который проявляет сайт-специфическую ферментативную активность, причем сайт фермента-

тивной активности определяется РНК, нацеленной на ДНК. В некоторых случаях сайт-специфический модифицирующий полипептид содержит аминокислотную последовательность с идентичностью по меньшей мере около 75% аминокислотной последовательности с аминокислотами 7-166 или 731-1003 аминокислотной последовательности Cas9/Csn1, изображенной на фиг. 3, или с соответствующими участками любой из аминокислотных последовательностей, приведенных в SEQ ID NO:1-256 и 795-1346. В некоторых случаях организм выбран из группы, которая состоит из археи, бактерии, эукариотического одноклеточного организма, водорослей, растения, животного, беспозвоночного, мухи, червя, кишечнополостного, позвоночного, рыбы, земноводного, птицы, млекопитающего, копытного, грызуна, крысы, мыши и примата, не являющегося человеком.

Признаки настоящего изобретения включают композицию, которая содержит (i) РНК, нацеленную на ДНК, или полинуклеотид ДНК, кодирующий ее, указанная РНК, нацеленная на ДНК, содержит (a) первый сегмент, содержащий нуклеотидную последовательность, комплементарную последовательности в ДНК-мишени; и (b) второй сегмент, который взаимодействует с сайт-специфическим модифицирующим полипептидом; и (ii) сайт-специфический модифицирующий полипептид или полинуклеотид, кодирующий его, указанный сайт-специфический модифицирующий полипептид содержит (a) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и (b) участок активности, который проявляет сайт-специфическую ферментативную активность, причем сайт ферментативной активности определяется РНК, нацеленной на ДНК. В некоторых случаях первый сегмент РНК, нацеленной на ДНК, содержит 8 нуклеотидов, которые имеют комплементарность 100% по отношению к последовательности в ДНК-мишени. В некоторых случаях РНК, нацеленная на ДНК, содержит нуклеотидную последовательность с идентичностью по меньшей мере 60% на участке из по меньшей мере 8 последовательных нуклеотидов по отношению к любой из нуклеотидных последовательностей, представленных в SEQ ID NO:431-682 (например, SEQ ID NO:563-682). В некоторых случаях РНК, нацеленная на ДНК, содержит нуклеотидную последовательность с идентичностью по меньшей мере 60% на участке из по меньшей мере 8 последовательных нуклеотидов по отношению к любой из нуклеотидных последовательностей, представленных в SEQ ID NO:431-562. В некоторых случаях сайт-специфический модифицирующий полипептид содержит аминокислотную последовательность с идентичностью по меньшей мере около 75% аминокислотной последовательности с аминокислотами 7-166 или 731-1003 аминокислотной последовательности Cas9/Csn1, изображенной на фиг. 3, или с соответствующими участками любой из аминокислотных последовательностей, приведенных в SEQ ID NO:1-256 и 795-1346. В некоторых случаях ферментативная активность модифицирует ДНК-мишень. В некоторых случаях ферментативная активность представляет собой нуклеазную активность, активность метилтрансферазы, активность деметилазы, активность репарации ДНК, активность повреждения ДНК, активность дезаминирования, активность дисмутазы, алкилирующую активность, активность депуринизации, активность окисления, активность формирования димера пиримидина, активность интегразы, активность транспозазы, активность рекомбиназы, полимеразную активность, лигазную активность, геликазную активность, активность фототиазы или гликозилазную активность. В некоторых случаях ферментативная активность представляет собой нуклеазную активность. В некоторых случаях нуклеазная активность вводит двухцепочечный разрыв в ДНК-мишень. В некоторых случаях ферментативная активность изменяет полипептид-мишень, ассоциированный с ДНК-мишенью. В некоторых случаях ферментативная активность является активностью метилтрансферазы, активностью деметилазы, ацетилтрансферазной активностью, активностью деацетилазы, киназной активностью, фосфатазной активностью, активностью убиквитинлигазы, деубиквитирующей активностью, активностью аденилирования, активностью деаденилирования, активностью сумоилирования, активностью десумоилирования, активностью рибозилирования, активностью дерибозилирования, активностью миристоилирования или активностью демиристоилирования. В некоторых случаях полипептид-мишень представляет собой гистон, а ферментативная активность представляет собой активность метилтрансферазы, деметилазную активность, ацетилтрансферазную активность, активность деацетилазы, активность киназы, активность фосфатазы, активность убиквитинлигазы или деубиквитирующую активность. В некоторых случаях РНК, нацеленная на ДНК, представляет собой двойную молекулу РНК, нацеленной на ДНК, и композиция содержит как РНК-нацеливающую, так и РНК-активатор, дуплекс-образующие сегменты которых комплементарны и гибридизуются с образованием второго сегмента РНК, нацеленной на ДНК. В некоторых случаях дуплекс-образующий сегмент РНК-активатора содержит нуклеотидную последовательность с идентичностью по меньшей мере 60% на участке из по меньшей мере 8 последовательных нуклеотидов по отношению к любой из нуклеотидных последовательностей, представленных в SEQ ID NO:431-682.

Признаки настоящего изобретения включают композицию, которая содержит (i) РНК, нацеленную на ДНК, по настоящему изобретению или полинуклеотид ДНК, кодирующий ее; и (ii) буфер для стабилизации нуклеиновых кислот. Признаки настоящего изобретения содержат композицию, которая содержит (i) сайт-специфический модифицирующий полипептид по настоящему изобретению или полинуклеотид, кодирующий его; и (ii) буфер для стабилизации нуклеиновых кислот и/или белков. Признаки настоящего изобретения содержат композицию, которая содержит (i) РНК, нацеленную на ДНК, или полинуклеотид ДНК, кодирующий ее, где РНК, нацеленная на ДНК, содержит (a) первый сегмент, содер-

жащий нуклеотидную последовательность, комплементарную последовательности в ДНК-мишени; и (b) второй сегмент, который взаимодействует с сайт-специфическим модифицирующим полипептидом; и (ii) сайт-специфический модифицирующий полипептид или полинуклеотид, кодирующий его, указанный сайт-специфический модифицирующий полипептид содержит (a) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и (b) участок активности, который модулирует транскрипцию в пределах ДНК-мишени, причем сайт модулированной транскрипция в пределах ДНК-мишени определяется РНК, нацеленной на ДНК. В некоторых случаях участок активности увеличивает транскрипцию в пределах ДНК-мишени. В некоторых случаях участок активности уменьшает транскрипцию в пределах ДНК-мишени. Признаки настоящего изобретения включают композицию, которая содержит (i) сайт-специфический модифицирующий полипептид или полинуклеотид, кодирующий его; и (ii) буфер для стабилизации нуклеиновых кислот и/или белков.

Признаки настоящего изобретения включают способ сайт-специфической модификации ДНК-мишени, при этом способ включает приведение ДНК-мишени в контакт с (i) РНК, нацеленной на ДНК, или полинуклеотидом ДНК, кодирующим то же самое, причем РНК, нацеленная на ДНК, содержит (a) первый сегмент, содержащий нуклеотидную последовательность, комплементарную последовательности в ДНК-мишени; и (b) второй сегмент, который взаимодействует с сайт-специфическим модифицирующим полипептидом; и (ii) сайт-специфический модифицирующий полипептид или полинуклеотид, кодирующий его, причем сайт-специфический модифицирующий полипептид содержит (a) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и (b) участок активности, который проявляет сайт-специфическую ферментативную активность. В некоторых случаях ДНК-мишень является внехромосомной. В некоторых случаях ДНК-мишень содержит последовательность РАМ комплементарной цепи, которая представляет собой 5'-ССУ-3', где У представляет собой любой нуклеотид ДНК, а У расположен непосредственно 5' в последовательности-мишени комплементарной цепи ДНК-мишени. В некоторых случаях ДНК-мишень является частью хромосомы *in vitro*. В некоторых случаях ДНК-мишень является частью хромосомы *in vivo*. В некоторых случаях ДНК-мишень является частью хромосомы в клетке. В некоторых случаях клетка выбрана из группы, которая состоит из клетки археи, бактериальной клетки, эукариотической клетки, одноклеточного эукариотического организма, соматической клетки, зародышевой клетки, стволовой клетки, клетки растения, клетки водорослей, клетки животного, клетки беспозвоночных, клетки позвоночных, клетки рыбы, клетки земноводного, клетки птицы, клетки млекопитающего, клетки свиньи, клетки коровы, клетки козы, клетки овцы, клетки грызуна, клетки крысы, клетки мыши, клетки примата, не являющегося человеком, и клетки человека. В некоторых случаях РНК, нацеленная на ДНК, содержит нуклеотидную последовательность с идентичностью по меньшей мере 60% на участке из по меньшей мере 8 последовательных нуклеотидов по отношению к любой из нуклеотидных последовательностей, представленных в SEQ ID NO:431-682 (например, SEQ ID NO:563-682). В некоторых случаях РНК, нацеленная на ДНК, содержит нуклеотидную последовательность с идентичностью по меньшей мере 60% на участке из по меньшей мере 8 последовательных нуклеотидов по отношению к любой из нуклеотидных последовательностей, представленных в SEQ ID NO:431-562. В некоторых случаях, ДНК-модифицирующий полипептид содержит аминокислотную последовательность с идентичностью по меньшей мере около 75% аминокислотной последовательности с аминокислотами 7-166 или 731-1003 аминокислотной последовательности Cas9/Csn1, изображенной на фиг. 3, или соответствующим участкам любой из аминокислотных последовательностей, приведенных в SEQ ID NO:1-256 и 795-1346. В некоторых случаях ферментативная активность модифицирует ДНК-мишень. В некоторых случаях ферментативная активность представляет собой нуклеазную активность, активность метилтрансферазы, активность деметилазы, активность репарации ДНК, активность повреждения ДНК, активность дезаминирования, активность дисмутазы, алкилирующую активность, активность депуринизации, активность окисления, активность формирования димера пиримидина, активность интегразы, активность транспозазы, активность рекомбиназы, полимеразную активность, лигазную активность, геликазную активность, активность фотолиазы или гликозилазную активность. В некоторых случаях ДНК-модифицирующая ферментативная активность представляет собой нуклеазную активность. В некоторых случаях, нуклеазная активность вводит двухцепочечный разрыв в ДНК-мишень. В некоторых случаях приведение в контакт происходит в условиях, которые позволяют осуществлять негомологичное соединение концов или репарирование по механизму гомологичной рекомбинации. В некоторых случаях способ дополнительно включает приведение ДНК-мишени в контакт с донорным полинуклеотидом, причем донорный полинуклеотид, участок донорного полинуклеотида, копия донорного полинуклеотида или участок копии донорного полинуклеотида интегрируется в ДНК-мишень. В некоторых случаях способ не включает приведение в контакт клетки с донорным полинуклеотидом, причем ДНК-мишень модифицируют таким образом, что удаляют нуклеотиды внутри ДНК-мишени. В некоторых случаях ферментативная активность изменяет полипептид-мишень, ассоциированный с ДНК-мишенью. В некоторых случаях ферментативная активность является активностью метилтрансферазы, активностью деметилазы, ацетилтрансферазной активностью, активностью деацетилазы, киназной активностью, фосфатазной активностью, активностью убиквитинлигазы, деубиквитинирующей активностью, активностью фосфатазации, активностью деаденилирования, активностью сумоилирования, активностью десумоилирования,

активностью рибозилирования, активностью дерибозилирования, активностью миристоилирования или активностью демиристоилирования. В некоторых случаях полипептид-мишень представляет собой гистон, а ферментативная активность представляет собой активность метилтрансферазы, деметилазную активность, ацетилтрансферазную активность, активность дезацетилазы, активность киназы, активность фосфатазы, активность убиквитинлигазы или деубиквитинирующую активность. В некоторых случаях комплекс дополнительно содержит РНК-активатор. В некоторых случаях РНК-активатор содержит нуклеотидную последовательность с идентичностью по меньшей мере 60% на участке из по меньшей мере 8 последовательных нуклеотидов по отношению к любой из нуклеотидных последовательностей, представленных в SEQ ID NO:431-682.

Признаки настоящего изобретения включают способ модулирования сайт-специфической транскрипции в пределах ДНК-мишени, при этом способ включает приведение ДНК-мишени в контакт с (i) РНК, нацеленной на ДНК, или полинуклеотидом ДНК, кодирующей ее, причем РНК, нацеленная на ДНК, содержит (a) первый сегмент, содержащий нуклеотидную последовательность, комплементарную последовательности ДНК-мишени; и (b) второй сегмент, который взаимодействует с сайт-специфическим модифицирующим полипептидом; и (ii) сайт-специфическим модифицирующим полипептидом или полинуклеотидом, кодирующим его, причем сайт-специфический модифицирующий полипептид содержит (a) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и (b) участок активности, который проявляет сайт-специфическую ферментативную активность. В некоторых случаях транскрипция в пределах ДНК-мишени растет. В некоторых случаях транскрипция в пределах ДНК-мишени уменьшается.

Признаки настоящего изобретения включают способ сайт-специфической модификации ДНК-мишени, при этом способ включает приведение ДНК-мишени в контакт с (i) РНК, нацеленной на ДНК, или полинуклеотидом ДНК, кодирующим то же самое, причем РНК, нацеленная на ДНК, содержит (a) первый сегмент, содержащий нуклеотидную последовательность, комплементарную последовательности в ДНК-мишени; и (b) второй сегмент, который взаимодействует с сайт-специфическим модифицирующим полипептидом; и (ii) сайт-специфическим модифицирующим полипептидом или полинуклеотидом, кодирующим его, причем сайт-специфический модифицирующий полипептид содержит (a) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и (b) участок активности, который модулирует транскрипцию в пределах ДНК-мишени. В некоторых случаях сайт-специфический модифицирующий полипептид увеличивает транскрипцию в пределах ДНК-мишени. В некоторых случаях сайт-специфический модифицирующий полипептид уменьшает транскрипцию в пределах ДНК-мишени.

Признаки настоящего изобретения включают способ промотирования сайт-специфического расщепления и модифицирования ДНК-мишени в клетке, при этом способ включает введение в клетку (i) РНК, нацеленной на ДНК, или полинуклеотида ДНК, кодирующего то же самое, причем РНК, нацеленная на ДНК, содержит (a) первый сегмент, содержащий нуклеотидную последовательность, комплементарную последовательности в ДНК-мишени; и (b) второй сегмент, который взаимодействует с сайт-специфическим модифицирующим полипептидом; и (ii) сайт-специфическим модифицирующим полипептидом или полинуклеотидом, кодирующим то же самое, причем сайт-специфический модифицирующий полипептид содержит (a) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и (b) участок активности, который проявляет нуклеазную активность, что создает двухцепочечный разрыв в ДНК-мишени; причем сайт двухцепочечного разрыва определяется РНК, нацеленной на ДНК, при этом приведение в контакт происходит в условиях, которые позволяют осуществлять негомологичное соединение концов или репарирование по механизму гомологичной рекомбинации, а ДНК-мишень расщепляют и вновь соединяют с получением модифицированной последовательности ДНК. В некоторых случаях способ дополнительно включает приведение ДНК-мишени в контакт с донорным полинуклеотидом, причем донорный полинуклеотид, участок донорного полинуклеотида, копия донорного полинуклеотида или участок копии донорного полинуклеотида интегрируется в ДНК-мишень. В некоторых случаях способ не включает приведение в контакт клетки с донорным полинуклеотидом, причем ДНК-мишень модифицируют таким образом, что удаляют нуклеотиды внутри ДНК-мишени. В некоторых случаях клетка выбрана из группы, которая состоит из клетки археи, бактериальной клетки, эукариотической клетки, одноклеточного эукариотического организма, соматической клетки, зародышевой клетки, стволовой клетки, клетки растения, клетки водорослей, клетки животного, клетки беспозвоночных, клетки позвоночных, клетки рыбы, клетки земноводного, клетки птицы, клетки млекопитающего, клетки свиньи, клетки коровы, клетки козы, клетки овцы, клетки грызуна, клетки крысы, клетки мыши, клетки примата, не являющегося человеком, и клетки человека. В некоторых случаях клетка находится *in vitro*. В некоторых случаях клетка находится *in vivo*.

Признаки настоящего изобретения включают способ получения у субъекта генетически модифицированной клетки, включающий (i) введение в клетку (i) РНК, нацеленной на ДНК, или полинуклеотида ДНК, кодирующего то же самое, причем РНК, нацеленная на ДНК, содержит (a) первый сегмент, содержащий нуклеотидную последовательность, комплементарную последовательности в ДНК-мишени; и (b) второй сегмент, который взаимодействует с сайт-специфическим модифицирующим полипептидом; и (ii)

сайт-специфического модифицирующего полипептида или полинуклеотида, кодирующего то же самое, причем сайт-специфический модифицирующий полипептид содержит (а) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и (b) участок активности, который проявляет нуклеазную активность, что создает двухцепочечный разрыв в ДНК-мишени; причем сайт двухцепочечного разрыва определяется РНК, нацеленной на ДНК, при этом приведение в контакт происходит в условиях, которые позволяют осуществлять негомологичное соединение концов или репарирование по механизму гомологичной рекомбинации, а ДНК-мишень расщепляют и вновь соединяют с получением модифицированной последовательности ДНК; таким образом, получая генетически модифицированную клетку; и (II) трансплантацию генетически модифицированной клетки указанному субъекту. В некоторых случаях способ дополнительно включает приведение в контакт клетки с донорным иолинуклеотидом, причем донорный полинуклеотид, участок донорного полинуклеотида, копия донорного полинуклеотида или участок копии донорного полинуклеотида интегрируется в ДНК-мишень. В некоторых случаях способ не включает приведение в контакт клетки с донорным полинуклеотидом, причем ДНК-мишень модифицируют таким образом, что удаляют нуклеотиды внутри ДНК-мишени. В некоторых случаях клетка выбрана из группы, которая состоит из клетки археи, бактериальной клетки, эукариотической клетки, одноклеточного эукариотического организма, соматической клетки, зародышевой клетки, стволовой клетки, клетки растения, клетки водорослей, клетки животного, клетки беспозвоночных, клетки позвоночных, клетки рыбы, клетки земноводного, клетки птицы, клетки млекопитающего, клетки копытного, клетки грызуна, клетки примата, не являющегося человеком, и клетки человека.

Признаки настоящего изобретения включают способ модифицирования ДНК-мишени в генетически модифицированной клетке, которая содержит нуклеотидную последовательность, кодирующую экзогенный сайт-специфический модифицирующий полипептид, включающий введение в генетически модифицированную клетку РНК, нацеленной на ДНК, или полинуклеотида ДНК, кодирующего ее, где (i) РНК, нацеленная на ДНК, содержит (а) первый сегмент, содержащий нуклеотидную последовательность, комплементарную последовательности в ДНК-мишени; и (b) второй сегмент, который взаимодействует с сайт-специфическим модифицирующим полипептидом; и (ii) сайт-специфический модифицирующий полипептид содержит (а) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и (b) участок активности, который проявляет нуклеазную активность. В некоторых случаях сайт-специфический модифицирующий полипептид содержит аминокислотную последовательность с идентичностью по меньшей мере около 75% аминокислотной последовательности с аминокислотами 7-166 или 731-1003 аминокислотной последовательности Cas9/Csn1, изображенной на фиг. 3, или с соответствующими участками любой из аминокислотных последовательностей, приведенных в SEQ ID NO:1-256 и 795-1346. В некоторых случаях клетка выбрана из группы, которая состоит из клетки археи, бактериальной клетки, эукариотической клетки, одноклеточного эукариотического организма, соматической клетки, зародышевой клетки, стволовой клетки, клетки растения, клетки водорослей, клетки животного, клетки беспозвоночных, клетки позвоночных, клетки рыбы, клетки земноводного, клетки птицы, клетки млекопитающего, клетки копытного, клетки грызуна, клетки примата, не являющегося человеком, и клетки человека. В некоторых случаях клетка находится *in vivo*. В некоторых случаях экспрессия сайт-специфического модифицирующего полипептида находится под контролем индуцируемого промотора. В некоторых случаях экспрессия сайт-специфического модифицирующего полипептида находится под контролем специфического к типу клетки промотора.

Признаки настоящего изобретения включают набор, содержащий: РНК, нацеленную на ДНК, или полинуклеотид ДНК, кодирующий ее; и реагент для растворения и/или разбавления. В некоторых случаях набор дополнительно содержит реагент, выбранный из группы, состоящей из буфера для введения в клетки РНК, нацеленной на ДНК, промывочного буфера, контрольного реагента, контрольного вектора экспрессии или полинуклеотида РНК, реагента для транскрибирования РНК, нацеленной на ДНК, из ДНК, а также их комбинаций.

Признаки настоящего изобретения включают набор, который содержит сайт-специфический модифицирующий полипептид по настоящему изобретению или полинуклеотид, кодирующий его; и реагент для растворения и/или разбавления. В некоторых случаях набор дополнительно содержит реагент, выбранный из группы, состоящей из буфера для введения в клетки сайт-специфического модифицирующего полипептида, промывочного буфера, контрольного реагента, контрольного вектора экспрессии или полинуклеотида РНК, реагента для *in vitro* получения сайт-специфического модифицирующего полипептида из ДНК, а также их комбинаций.

Признаки настоящего изобретения включают набор, который содержит сайт-специфический модифицирующий полипептид по настоящему изобретению или полинуклеотид, кодирующий его; и реагент для растворения и/или разбавления. Признаки настоящего изобретения включают набор, который содержит: РНК, нацеленную на ДНК, или полинуклеотид ДНК, кодирующий ее, причем РНК, нацеленная на ДНК, содержит (а) первый сегмент, содержащий нуклеотидную последовательность, комплементарную последовательности в ДНК-мишени; и (b) второй сегмент, который взаимодействует с сайт-специфическим модифицирующим полипептидом; и (ii) сайт-специфический модифицирующий полипептид или полинуклеотид, кодирующий его, указанный сайт-специфический модифицирующий поли-

пептид содержит (а) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и (b) участок активности, который проявляет сайт-специфическую ферментативную активность, причем сайт ферментативной активности определяется РНК, нацеленной на ДНК.

Признаки настоящего изобретения включают набор, который содержит (i) РНК, нацеленную на ДНК, или полинуклеотид ДНК, кодирующий ее, которая содержит (а) первый сегмент,- содержащий нуклеотидную последовательность, комплементарную последовательности в ДНК-мишени; и (b) второй сегмент, который взаимодействует с сайт-специфическим модифицирующим полипептидом; и (ii) сайт-специфический модифицирующий полипептид или полинуклеотид, кодирующий его, который содержит:

(а) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и

(b) участок активности, который модулирует транскрипцию в пределах ДНК-мишени, причем сайт модулированной транскрипции в пределах ДНК-мишени определяется РНК, нацеленной на ДНК.

Признаки настоящего изобретения включают набор, содержащий: (i) любой из рекомбинантных векторов экспрессии, приведенных выше; и (ii) реагент для растворения и/или разбавления. Признаки настоящего изобретения включают набор, содержащий: (i) любой из рекомбинантных векторов экспрессии, приведенных выше; и (ii) рекомбинантный вектор экспрессии, который содержит нуклеотидную последовательность, которая кодирует сайт-специфический модифицирующий полипептид, причем сайт-специфический модифицирующий полипептид содержит (а) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и (b) участок активности, который проявляет сайт-специфическую ферментативную активность, при том что сайт ферментативной активности определяется РНК, нацеленной на ДНК. Признаки настоящего изобретения включают набор, содержащий: (i) любой из рекомбинантных векторов экспрессии, приведенных выше; и (ii) рекомбинантный вектор экспрессии, который содержит нуклеотидную последовательность, кодирующую сайт-специфический модифицирующий полипептид, причем сайт-специфический модифицирующий полипептид содержит (а) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и (b) участок активности, который модулирует транскрипцию в пределах ДНК-мишени, причем сайт модулированной транскрипции в пределах ДНК-мишени определяется РНК, нацеленной на ДНК.

Признаки настоящего изобретения включают набор для нацеливания на ДНК-мишень, который содержит две или более РНК, нацеленные на ДНК, или полинуклеотиды ДНК, кодирующие то же самое, причем первый сегмент по меньшей мере одной из двух или более РНК, нацеленных на ДНК, отличается по меньшей мере одним нуклеотидом от первого сегмента по меньшей мере одного другого из двух или более РНК, нацеленных на ДНК.

#### Краткое описание графических материалов

Фиг. 1А-В обеспечивает схематическое изображение двух иллюстративных РНК, нацеленных на ДНК, по изобретению, каждая из которых связана с сайт-специфическим модифицирующим полипептидом и с ДНК-мишенью.

Фиг. 2 иллюстрирует "редактирование" ДНК-мишени посредством введения двухцепочечных разрывов ДНК, с использованием сайт-специфического модифицирующего полипептида Cas9/Csn1 и РНК, нацеленной на ДНК.

Фиг. 3А-В иллюстрируют последовательность аминокислот белка Cas9/Csn1 из *Streptococcus pyogenes* (SEQ ID NO:8). Cas9 имеет домены, гомологичные как эндонуклеазе HNH, так и эндонуклеазе RuvC. (А) Мотивы 1-4 надчеркнуты. (В) Домены 1 и 2 надчеркнуты.

Фиг. 4А-В иллюстрируют относительную идентичность (в %) между белками Cas9/Csn1 нескольких видов. (А) Идентичность последовательности по отношению к *Streptococcus pyogenes*. Например, домен 1 представляет собой аминокислоты 7-166, а домен 2 представляет собой аминокислоты 731-1003 Cas9/Csn1 из *Streptococcus pyogenes*, как проиллюстрировано на фиг. 3В. (В) Идентичность последовательности по отношению к *Neisseria meningitidis*. Например, домен 1 представляет собой аминокислоты 13-139, а домен 2 представляет собой аминокислоты 475-750 Cas9/Csn1 из *Neisseria meningitidis* (SEQ ID NO:79).

Фиг. 5 иллюстрирует множественное выравнивание последовательностей мотивов 1-4 белков Cas9/Csn1 различных несходных видов, выбранных из филогенетической таблицы на фиг. 32 (см. фиг. 32, 3А и табл. 1) (*Streptococcus pyogenes* (SEQ ID NO:8), *Legionella pneumophila* (SEQ ID NO:17), *Gamma proteobacterium* (SEQ ID NO:107), *Listeria innocua* (SEQ ID NO:3), *Lactobacillus gasseri* (SEQ ID NO:152), *Eubacterium rectale* (SEQ ID NO:99), *Staphylococcus lugdunensis* (SEQ ID NO:185), *Mycoplasma synoviae* (SEQ ID NO:22), *Mycoplasma mobile* (SEQ ID NO:16), *Wolinella succinogenes* (SEQ ID NO:10), *Flavobacterium columnare* (SEQ ID NO:235), *Fibrobacter succinogenes* (SEQ ID NO:121), *Bacteroides fragilis* (SEQ ID NO:21), *Acidothermus cellulolyticus* (SEQ ID NO:42) и *Bifidobacterium dentium* (SEQ ID NO:131)).

Фиг. 6А-В показывают выравнивание встречающихся в природе последовательностей *tracrP*РНК ("РНК-активатор") различных видов (*L. innocua* (SEQ ID NO:268); *S. pyogenes* (SEQ ID NO:267); *S. mutans* (SEQ ID NO:269); *S. thermophilus1* (SEQ ID NO:270); *M. mobile* (SEQ ID NO:274); *N. meningitidis* (SEQ ID NO:272); *P. multocida* (SEQ ID NO:273); *S. theimophilus2* (SEQ ID NO:271); и *S. pyogenes* (SEQ ID NO:267)). (А) множественное выравнивание последовательностей некоторых ортологов *tracrP*РНК (AlignX, пакет VectorNTI, Invitrogen) ассоциировано с локусом CRISPR/Cas схожей архитектуры и весьма близ-

кими последовательностями Cas9/Csn1. Черными четырехугольниками отмечены совпадающие нуклеотиды (B) некоторых ортологов *tracrP*HK (AlignX, пакет VectorNTI, Invitrogen), ассоциированных с локусом CRISPR/Cas схожей архитектуры и с неблизкородственными последовательностями Cas9/Csn1. Следует обратить внимание на сходство последовательностей ортологов *tracrP*HK *N. meningitidis* и *P. multocida*. Черными четырехугольниками отмечены совпадающие нуклеотиды. В случае дополнительных иллюстративных последовательностей РНК-активатора, см. SEQ ID NO:431-562.

Фиг. 7А-В иллюстрируют выравнивания встречающихся в природе дуплекс-образующих сегментов последовательностей *crP*HK ("РНК-нацеливающая") различных видов (*L. innocua* (SEQ ID NO://); *S. pyogenes* (SEQ ID NO://); *S. mutans* (SEQ ID NO://); *S. thermophilus1* (SEQ ID NO://); *C. jejuni* (SEQ ID NO://); *S. pyogenes* (SEQ ID NO://); *F. novicida* (SEQ ID NO://); *M. mobile* (SEQ ID NO://); *N. meningitidis* (SEQ ID NO://); *P. multocida* (SEQ ID NO://); и *S. thermophilus2* (SEQ ID NO://). (А) множественные выравнивания последовательностей иллюстративного дуплекс-образующего сегмента последовательностей РНК-нацеливающей (AlignX, пакет VectorNTI, Invitrogen) ассоциированы с локусами аналогичной архитектуры и последовательностями весьма подобными Cas9/Csn1. (В) множественные выравнивания последовательностей иллюстративного дуплекс-образующего сегмента последовательностей РНК-нацеливающей (AlignX, пакет VectorNTI, Invitrogen) ассоциированной с локусами различной архитектуры и различными последовательностями Cas9. Черными четырехугольниками отмечены совпадающие нуклеотиды. В случае иллюстративных дуплекс-образующих сегментов последовательностей РНК-активатора, см. SEQ ID NO:563-679.

Фиг. 8 иллюстрирует схему гибридизации встречающихся в природе дуплекс-образующих сегментов *crP*HK ("РНК-нацеливающая") с дуплекс-образующим сегментом соответствующего ортолога *tracrP*HK ("РНК-активатор"). Верхняя последовательность, РНК-нацеливающая; нижняя последовательность, дуплекс-образующий сегмент соответствующей РНК-активатора. Локус CRISPR относят к системе CRISPR/Cas (Nmeni/CASS4) типа II. Используется номенклатура в соответствии с базой данных CRISPR (CRISPR DB). 8. *pyogenes* (SEQ ID NO:// и //); *S. mutans* (SEQ ID NO:// и //); *S. thermophilus1* (SEQ ID NO:// и //); *S. thermophilus2* (SEQ ID NO:// и //); *L. innocua* (SEQ ID NO:// и //); *T. denticola* (SEQ ID NO:// и //); *N. meningitidis* (SEQ ID NO:// и //); *S. gordonii* (SEQ ID NO:// и //); *B. bifidum* (SEQ ID NO:// и //); *L. salivarius* (SEQ ID NO:// и //); *F. tularensis* (SEQ ID NO:// и //); и *L. pneumophila* (SEQ ID NO:// и //). Обратите внимание, что некоторые из видов содержат два локуса CRISPR типа II каждый. В случае дополнительных иллюстративных последовательностей РНК-активатора, см. SEQ ID NO:431-562. В случае иллюстративных дуплекс-образующих сегментов последовательностей РНК-активатора, см. SEQ ID NO:563-679.

Фиг. 9 иллюстрирует пример последовательностей *tracrP*HK (РНК-активатор) и *crP*HK (РНК-нацеливающая) двух видов. Степень взаимозаменяемости существует; например, белок Cas9/Csn1 *S. pyogenes* является функциональным с *tracrP*HK и *crP*HK, полученной из *L. innocua*. (I) Обозначает каноническую уотсон-криковскую пару оснований, в то время как (•) обозначает неоднозначную пару оснований G-U. "Изменяемые 20 нт" или "20 нт" представляют собой ДНК-нацеленный сегмент, который является комплементарным ДНК-мишени (эта область может быть вплоть до около 100 нт в длину). Также показана структура РНК, нацеленной на ДНК, в виде одиночной молекулы, которая содержит признаки РНК-нацеливающей и РНК-активатора. (Белковые последовательности Cas9/Csn1 самых разнообразных видов проиллюстрированы на фиг. 3 и приведены в виде SEQ ID NO:1-256 и 795-1346) *Streptococcus pyogenes*: сверху вниз: (SEQ ID NO://, //, //); *Listeria innocua*: сверху вниз: (SEQ ID NO://, //, //). Полученные последовательности являются неограничивающими примерами и предназначены для иллюстрации того, как РНК, нацеленные на ДНК, в виде одиночной молекулы, и РНК нацеленные на ДНК, в виде двух молекул, могут быть разработаны на основе существующих в природе последовательностей самых разнообразных видов. Различные примеры подходящих последовательностей широкого спектра видов приведены следующим образом (белок Cas9: SEQ ID NO:1-259; *tracrP*HK: SEQ ID NO:431-562 или их комплексы; *crP*HK: SEQ ID NO:563-679 или их комплексы, и пример РНК, нацеленных на ДНК, в виде одиночной молекулы: SEQ ID NO:680-682).

Фиг. 10А-Е иллюстрируют то, что Cas9 представляет собой эндонуклеазу ДНК, направляемую двумя молекулами РНК. Фиг. Е: сверху вниз, SEQ ID NO:278-280 и //.

Фиг. 11А-В иллюстрируют то, что Cas9 использует два нуклеазные домена для расщепления двух цепей ДНК-мишени.

Фиг. 12А-Е иллюстрируют то, что катализируемое Cas9 расщепление ДНК-мишени требует активирующий домен в *tracrP*HK и регулируется с помощью затравочной последовательности в *crP*HK. Фиг. 12С: сверху вниз, SEQ ID NO:278-280, и //; фиг. 12D: сверху вниз, SEQ ID NO:281-290; и фиг. 12Е: сверху вниз, SEQ ID NO:291-292, 283, 293-298.

Фиг. 13А-С иллюстрируют то, что РАМ требуется для разрешения расщепления ДНК-мишени с помощью комплекса Cas9- *tracrP*HK:*crP*HK.

Фиг. 14А-С иллюстрируют то, что Cas9 может быть запрограммирован с использованием отдельной созданной молекулы РНК, сочетающей признаки *tracrP*HK и *crP*HK. Химера А (SEQ ID NO:299); химера В (SEQ ID NO:300).

- Фиг. 15 иллюстрирует РНК-опосредованный иммунный путь CRISPR/Cas типа II.
- Фиг. 16А-В иллюстрируют очистку нуклеазы Cas9.
- Фиг. 17А-С иллюстрируют то, что Cas9, управляемый двойной-tracrPНК:crPНК, расщепляет протоспейсер плазмиды и олигонуклеотид ДНК. Фиг. 17В: сверху вниз, SEQ ID NO:301-303 и //; и фиг. 17С: сверху вниз, SEQ ID NO:304-306 и //.
- Фиг. 18А-В иллюстрируют то, что Cas9 является Mg<sup>2+</sup>-зависимой эндонуклеазой с 3'-5' экзонуклеазной активностью.
- Фиг. 19А-С иллюстрируют то, что расщепление Cas9 ДНК-мишени с помощью двойная-tracrPНК:crPНК, является сайт-специфическим; фиг. 19С: сверху вниз, SEQ ID NO:307-309, //, 337-339 и //.
- Фиг. 20А-В иллюстрируют то, что расщепление Cas9 ДНК-мишени, направляемое двойная-tracrPНК:crPНК, является быстрым и эффективным.
- Фиг. 21А-В иллюстрируют то, что HNH и RuvC-подобные домены направляют расщепление Cas9 комплементарной и некомплемментарной цепи ДНК соответственно.
- Фиг. 22 иллюстрирует то, что tracrPНК необходима для распознавания ДНК-мишени.
- Фиг. 23А-В иллюстрируют то, что минимальная область tracrPНК способна направлять управляемое двойная-tracrPНК:crPНК расщепление ДНК-мишени.
- Фиг. 24А-Д иллюстрируют то, что направляемое двойная-tracrPНК:crPНК расщепление ДНК-мишени с помощью Cas9 может быть видоспецифическим.
- Фиг. 25А-С иллюстрируют то, что затравочная последовательность в crPНК направляет управляемое двойная-tracrPНК:crPНК расщепление ДНК-мишени с помощью Cas9. Фиг. 25А: зонд 1 ДНК-мишени (SEQ ID NO:310); спейсер 4 crPНК (1-42) (SEQ ID NO:311); tracrPНК (15-89) (SEQ ID NO://). Фиг. 25В левая часть (SEQ ID NO:310).
- Фиг. 26А-С иллюстрируют то, что последовательность PAM имеет важное значение в случае расщепления ДНК протоспейсерной плазмиды с помощью Cas9-tracrPНК: crPНК и в случае Cas9-опосредованной интерференции плазмидной ДНК в бактериальных клетках. Фиг. 26В: сверху вниз, SEQ ID NO:312-314; и фиг. 26С: сверху вниз, SEQ ID NO:315-320.
- Фиг. 27А-С иллюстрируют то, что Cas9, управляемое отдельной гибридной РНК, имитирующих двойную-tracrPНК:crPНК, расщепляет протоспейсерную ДНК. Фиг. 27С: сверху вниз, SEQ ID NO:321-324.
- Фиг. 28А-Д иллюстрируют создание de novo гибридных РНК, специфических к последовательности гена зеленого флуоресцентного белка (GFP). Фиг. 28В: сверху вниз, SEQ ID NO:325-326. Фиг. 28С: последовательность-мишень GFP1 (SEQ ID NO:327); последовательность-мишень GFP2 (SEQ ID NO:328); последовательность-мишень GFP3 (SEQ ID NO:329); последовательность-мишень GFP4 (SEQ ID NO:330); последовательность-мишень GFP5 (SEQ ID NO:331); гибридная РНК GFP1 (SEQ ID NO:332); гибридная РНК GFP2 (SEQ ID NO:333); гибридная РНК GFP3 (SEQ ID NO:334); гибридная РНК GFP4 (SEQ ID NO:335); гибридная РНК GFP5 (SEQ ID NO:336).
- Фиг. 29А-Е иллюстрируют то, что коэкспрессия Cas9 и направляющей РНК в человеческих клетках генерирует разрывы двуцепочечной ДНК в локусе-мишени. Фиг. 29С: сверху вниз, SEQ ID NO:425-428.
- Фиг. 30А-В иллюстрируют то, что клеточные лизаты содержат активное Cas9:sgPНК и поддерживают сайт-специфическое расщепление ДНК.
- Фиг. 31А-В иллюстрируют то, что 3'-концевое удлинение конструкций sgPНК повышает сайт-специфический NHEJ-опосредованный мутагенез. Фиг. 31А: сверху вниз, SEQ ID NO:428-430.
- Фиг. 32А-В иллюстрируют филогенетическое дерево репрезентативных последовательностей Cas9 из различных организмов (А), а также архитектуры локуса Cas9 для основных групп дерева (В).
- Фиг. 33А-Е иллюстрируют архитектуру CRISPR-Cas типа II отдельных видов бактерий.
- Фиг. 34А-В иллюстрируют коэкспрессию tracrPНК и pre-crPНК отдельных систем CRISPR Cas типа II. Фиг. 34А: сверху вниз, SEQ ID NO://, //, //, //, //, //, //, //; фиг. 34В: сверху вниз, SEQ ID NO://, //, //, //.
- Фиг. 35 иллюстрирует выравнивание последовательности ортологов tracrPНК, демонстрируя разнообразие последовательностей tracrPНК.
- Фиг. 36А-Ф иллюстрируют экспрессию бактериальных ортологов tracrPНК и crPНК, обнаруженных с помощью глубокого секвенирования РНК.
- Фиг. 37А-О перечисляют все ортологи tracrPНК и зрелые crPНК, полученные с помощью секвенирования изученных бактериальных видов, в том числе координаты (представляющие интерес области) и соответствующие последовательности сДНК (от 5' до 3').
- Фиг. 38 А-В иллюстрируют таблицу видов бактерий, которые содержат локусы CRISPR-Cas типа II, характеризующиеся наличием геномной подписи cas9. Эти последовательности используют в случае филогенетического анализа.
- Фиг. 39А-В иллюстрируют устройство системы интерференции CRISPR (CRISPRi).
- Фиг. 40А-Е иллюстрируют то, что CRISPRi эффективно подавляет элонгацию и инициацию транскрипции.
- Фиг. 41А-В иллюстрируют то, что CRISPRi функционирует с помощью блокирования элонгации транскрипции.

Фиг. 42А-С иллюстрируют направленную специфичность системы CRISPRi.

Фиг. 43А-Е иллюстрируют характеристику факторов, которые влияют на эффективность сайленсинга.

Фиг. 44А-С иллюстрируют функциональное профилирование сложной регуляторной сети с использованием нокадауна гена CRISPRi.

Фиг. 45А-В иллюстрируют сайленсинг гена с использованием CRISPRi в клетках млекопитающего.

Фиг. 46 иллюстрирует механизм системы CRISPR типа II из *S. pyogenes*.

Фиг. 47А-В иллюстрируют кривые роста клеточных культур *E. coli*, совместно трансформированные с помощью dCas9 и sgPНК.

Фиг. 48 иллюстрирует то, что CRISPRi может подавлять экспрессию гена-репортера на мультико- нийной плазмиде.

Фиг. 49А-С иллюстрируют данные PНК-последовательность клеток с sgPНК, которые нацелены на различные гены.

Фиг. 50А-Е иллюстрируют эффекты сайленсинга sgPНК с прилегающими двойными ошибочными спариваниями.

Фиг. 51А-С иллюстрируют комбинаторные эффекты подавления с использованием двух sgPНК, для того, чтобы регулировать отдельный ген.

Фиг. 52 иллюстрирует то, что репрессия sgPНК является зависимой от локусов-мишеней и относительно расстояния от начала транскрипции.

Фиг. 53А-С иллюстрируют экспериментальные результаты, которые демонстрируют, что вариант сайт-специфического полипептида Cas9 (dCas9) работает в случае заявленных способов, при условии, что dCas9 имеет уменьшенную активность только в домене RuvC1 (например, D10A), только в домене HNH (например, H840A) или в обоих доменах (например, D10A и H840A).

Фиг. 54А-С перечисляют примеры подходящих партнеров по слиянию (или их фрагментов) в случае заявленного варианта сайт-специфического полипептида Cas9. Примеры содержат, но без ограничения им, перечисленное.

Фиг. 55А-Д иллюстрируют то, что гибридный сайт-специфический полипептид может быть использован для активации (увеличения) транскрипции в клетках человека.

Фиг. 56 иллюстрирует то, что гибридный сайт-специфический полипептид может быть использован: для репрессии (уменьшения) транскрипции в клетках человека.

Фиг. 57А, В иллюстрируют то, что искусственные последовательности, которые разделяют около 50% идентичности с встречающимися в природе trасPНК и sgPНК, могут функционировать с Cas9 для расщепления ДНК-мишени до тех пор, пока сохраняется структура белок-связывающего домена PНК, нацеленной на ДНК.

## Определения

### Часть I

Термины "полинуклеотид" и "нуклеиновая кислота", используемые в данном документе взаимозаменяемо, относятся к полимерной форме нуклеотидов любой длины, как рибонуклеотидов, так и дезоксирибонуклеотидов. Таким образом, этот термин содержит, но без ограничения ими, одно-, двух- или многоцепочечные ДНК или PНК, геномную ДНК, κДНК, гибриды ДНК-PНК или полимер, который содержит пуриновые и пиримидиновые основания или другие природные, химически или биохимически измененные, ненатуральные или дериватизированные нуклеотидные основания. "Олигонуклеотид" в целом относится к полинуклеотидам из от около 5 до около 100 нуклеотидов в форме одно- или двухцепочечной ДНК. Тем не менее, для целей данного раскрытия не существует верхнего предела длины олигонуклеотида. Олигонуклеотиды, также известные как "олигомеры" или "олиги", могут быть выделены из генов или химически синтезированы с помощью способов, известных в данной области техники. Термины "полинуклеотид" и "нуклеиновая кислота" следует понимать как те, которые содержат, как это применимо к описываемым вариантам реализации настоящего изобретения, одноцепочечные (например, смысловые и антисмысловые) и двухцепочечные полинуклеотиды.

"Структура типа "стебель-петля" относится к нуклеиновой кислоте, имеющей вторичную структуру, которая содержит область нуклеотидов, которые, как известно, или прогнозируется, образуют двойную цепь (участок стебля), который с одной стороны связан с областью, преимущественно, одноцепочечных нуклеотидов (участка петли). Термины структуры "шпильки" и "конъюгация в себе" также используют в данном документе для обозначения структур типа "стебель - петля". Подобные структуры хорошо известны в данной области техники и эти термины используют в соответствии с их известными значениями в данной области техники. Как известно в данной области техники, структура типа "стебель - петля" не требует точного спаривания оснований. Таким образом, стебель может содержать одно или несколько ошибочных спариваний оснований. В качестве альтернативы, образование пары оснований может быть точным, т.е. не содержать никаких ошибочных спариваний.

Под "гибридирующийся", "комплементарный" или "практически комплементарный" подразумевают то, что нуклеиновая кислота (например, PНК) содержит последовательность нуклеотидов, которая позволяет ей нековалентно связываться, т.е. образовывать пары оснований Уотсона-Крика и/или пары ос-

нований G/U, "отжигаться" или "гибридизоваться" с другой нуклеиновой кислотой последовательность-специфическим, антипараллельным способом (т.е. нуклеиновая кислота специфически связывается с комплементарной нуклеиновой кислотой) при соответствующих условиях *in vitro* и/или *in vivo* температуры и ионной силы раствора. Как известно в данной области техники, стандартное образование пары оснований по Уотсон-Крику содержит образование пары аденина (А) с тимидином (Т), образование пары аденина (А) с урацилом (U) и образование пары гуанина (G) с цитозином (С) [ДНК, РНК]. Кроме того, в этой области техники в случае гибридизации между двумя молекулами РНК (например, дцРНК) также известны пары оснований гуанина (G) с урацилом (U). Например, образование пары оснований GAJ является частично ответственным за вырожденность (т.е. избыточность) генетического кода в контексте спаривания оснований анти-кодона тРНК с кодонами в мРНК. В контексте данного описания, гуанин (G) из белок-связывающего сегмента (дуплекса дцРНК) молекулы РНК-мишени, нацеленной на ДНК, считается комплементарным урацилу (U), и наоборот. По существу, если пара оснований G/U может быть образована в данной нуклеотидной позиции белок-связывающего сегмента (дуплекса дцРНК) молекулы РНК-мишени, нацеленной на ДНК, положение не считается некомплементарным, но считается, что оно является комплементарным.

Условия гибридизации и промывания хорошо известны и приведены в качестве примера в Sambrook, J., Fritsch, E. F. и Maniatis, T. *Molecular Cloning: A Laboratory Manual, Second Edition*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (1989), частично Часть 11 и табл. 11.1 в вышеуказанном источнике; и Sambrook, J. и Russell, W., *Molecular Cloning: A Laboratory Manual, Third Edition*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (2001). Условия температуры и ионной силы определяют "жесткость" условий гибридизации.

Гибридизация требует того, чтобы обе нуклеиновые кислоты содержали комплементарные последовательности, хотя возможны ошибочные спаривания между основаниями. Условия, подходящие для гибридизации двух нуклеиновых кислот, зависят переменных, хорошо известных в данной области техники: длина нуклеиновых кислот и степень комплементарности. Чем больше степень комплементарности между двумя нуклеотидными последовательностями, тем больше значение температуры плавления ( $T_m$ ) для гибридов нуклеиновых кислот, которые имеют эти последовательности. В случае гибридизации нуклеиновых кислот с короткими участками комплементарности (например, комплементарность более чем 35 нуклеотидов или меньше, 30 или меньше, 25 или меньше, 22 или меньше, 20 или меньше, или 18 или меньше) положение ошибочных спариваний становится важным (см. Sambrook и соавт., выше, 11.7-11.8). Как правило, длина гибридизуемой нуклеиновой кислоты составляет по меньшей мере около 10 нуклеотидов. Иллюстративные минимальные длины в случае гибридизуемой нуклеиновой кислоты составляет: по меньшей мере около 15 нуклеотидов; по меньшей мере около 20 нуклеотидов; по меньшей мере около 22 нуклеотидов; по меньшей мере около 25 нуклеотидов; и по меньшей мере около 30 нуклеотидов). Кроме того, специалисту в данной области техники будет понятно, что температура и концентрация соли промывочного раствора может быть изменена по мере необходимости в зависимости от таких факторов, как длина области комплементарности и степень комплементарности.

В данной области техники является понятным, что последовательность полинуклеотида не обязательно должна быть на 100% комплементарна таковой его нуклеиновой кислоты-мишени, чтобы быть специфически гибридизоваться или поддаваться гибридизации. Кроме того, полинуклеотид может гибридизоваться над одним или несколькими сегментами таким образом, что промежуточные или соседние сегменты не участвуют в процессе гибридизации (например, структура петли или структура шпильки). Полинуклеотид может иметь по меньшей мере 70%, по меньшей мере 80%, по меньшей мере 90%, по меньшей мере 95%, по меньшей мере 99 или 100% комплементарность последовательности с областью-мишенью внутри последовательности нуклеиновой кислоты-мишени, к которой он является специфическим. Например, бессмысловая нуклеиновая кислота, в которой 18 из 20 нуклеотидов бессмыслового соединения являются комплементарными к области-мишени, и, следовательно, могут специфически гибридизоваться, является примером 90-процентной комплементарности. В этом примере, остальные некомплементарные нуклеотиды могут быть сгруппированы или представлены вперемежку с комплементарными нуклеотидами и не обязательно должно быть смежными друг с другом или с комплементарными нуклеотидами. Процент комплементарности между отдельными участками последовательностей нуклеиновых кислот в пределах нуклеиновых кислот может быть определен по стандартной методике с использованием программ BLAST (основные инструменты поиска локального выравнивания) и программ PowerBLAST, известных в данной области техники (Altschul и соавт., *J. Mol. Biol.*, 1990, 215, 403-410; Zhang и Madden, *Genome Res.*, 1997, 7, 649-656) или с помощью программы Gap (Wisconsin Sequence Analysis Package, Version 8 для Unix, Genetics Computer Group, University Research Park, Madison Wis.), используя настройки по умолчанию, которые используют алгоритм Смита и Уотермана (*Adv. Appl. Math.*, 1981, 2, 482-489).

Термины "пептид", "полипептид" и "белок" используются в данном документе взаимозаменяемо и относятся к полимерной форме аминокислот любой длины, которые могут содержать кодированные и не кодированные аминокислоты, химически или биохимически модифицированные или дериватизированные аминокислоты и полипептиды, которые имеют модифицированный остов пептида.

"Связывающий", как используют в данном документе (например, со ссылкой на РНК-связывающий домен полипептида), относится к нековалентному взаимодействию макромолекул (например, между белком и нуклеиновой кислотой). Вместе с тем, в состоянии нековалентного взаимодействия макромолекулы являются, как говорят, "ассоциированными", "взаимодействующими" или "связанными" (например, если говорят, что молекула X взаимодействует с молекулой Y, подразумевают то, что молекула X связывается с молекулой Y нековалентным способом). Не все компоненты связывающего взаимодействия должны быть последовательность-специфическими (например, контакты с фосфатными остатками в остове ДНК), но некоторые части связывающего взаимодействия могут быть последовательность-специфическими. Связывающие взаимодействия, как правило, характеризуются константой диссоциации (Kd) менее 10<sup>-6</sup> М, менее 10<sup>-7</sup> М, менее 10<sup>-8</sup> М, менее 10<sup>-9</sup> М, менее 10<sup>-10</sup> М, менее 10<sup>-11</sup> М, менее 10<sup>-12</sup> М, менее 10<sup>-13</sup> М, менее 10<sup>-14</sup> М или менее 10<sup>-15</sup> М. "Аффинность" соответствует силе связывания, увеличенная аффинность связывания коррелируется с более низким Kd.

Под "связывающем доменом" подразумевают белковый домен, который способен нековалентно связываться с другой молекулой. Связывающий домен может связываться с, например, молекулой ДНК (ДНК-связывающий белок), молекулой РНК (РНК-связывающий белок) и/или молекулой белка (белок-связывающий белок). В случае с белком, связывающим домен белка, он может связываться сам с собой (чтобы образовать гомодимеры, гомотримеры и т.д.) и/или он может связываться с одной или более молекулами другого белка или белков.

Термин "консервативная замена аминокислоты" относится к взаимозаменяемости в белках аминокислотных остатков, имеющих сходные боковые цепи. Например, группа аминокислот, которые имеют алифатические боковые цепи, состоит из глицина, аланина, валина, лейцина и изолейцина; группа аминокислот, которые имеют алифатические-гидроксильные боковые цепи, состоит из серина и треонина; группа аминокислот, боковые цепи которых содержат амид, состоит из аспарагина и глутамина; группа аминокислот, которые имеют ароматические боковые цепи, состоит из фенилаланина, тирозина и триптофана; группа аминокислот, которые имеют основные боковые цепи, состоит из лизина, аргинина и гистидина; группа аминокислот, которые имеют кислотные боковые цепи, состоит из глутамата и аспартата; и группа аминокислот, которые имеют серосодержащие боковые цепи, состоящие из цистеина и метионина. Иллюстративные консервативные замещающие группы аминокислот представляют собой: валин-лейцин-изолейцин, фенилаланин-тирозин, лизин-аргинин, аланин-валин и аспарагин-глутамин.

Полинуклеотид или полипептид имеет определенный процент "идентичности последовательности" к другому полинуклеотиду или полипептиду, что означает, что при выравнивании, этот процент оснований или аминокислот при сравнении двух последовательностей являются одинаковыми, и занимают ту же позицию. Идентичность последовательностей можно определить с помощью ряда различных способов. Для определения идентичности последовательности, последовательности могут быть выровнены с помощью различных способов и компьютерных программ (например, BLAST, T-COFFEE, MUSCLE, MAFFT и т.д.), доступных в интернете на сайтах, в том числе [ncbi.nlm.nih.gov/BLAST](http://ncbi.nlm.nih.gov/BLAST), [ebi.ac.uk/Tools/msa/tcoffee/](http://ebi.ac.uk/Tools/msa/tcoffee/), [ebi.ac.uk/Tools/msa/muscle/](http://ebi.ac.uk/Tools/msa/muscle/), [mafft.cbrc.jp/alignment/software/](http://mafft.cbrc.jp/alignment/software/). See, например, Altschul и соавт. (1990), *J. Mol. Biol.* 215:403-10.

Последовательность ДНК, которая "кодирует" определенную РНК, представляет собой последовательность нуклеиновой кислоты ДНК, которая транскрибируется в РНК. Полинуклеотид ДНК может кодировать РНК (мРНК), которая транслируется в белок, или полинуклеотид ДНК может кодировать РНК, которая не транслируется в белок (например, тРНК, рРНК или РНК, нацеленная на ДНК; также называемая "некодирующая" РНК или "псРНК").

"Кодирующая белок последовательность" или последовательность, которая кодирует определенный белок или полипептид, представляет собой последовательность нуклеиновой кислоты, которая транскрибируется в мРНК (в случае ДНК) и транслируется (в случае мРНК) в полипептид *in vitro* или *in vivo* при помещении под контроль соответствующих регуляторных последовательностей. Границы кодирующей последовательности определяют с помощью иницирующего кодона на 5'-конце (N-конце) и стоп-кодона остановки трансляции на 3'-конце (C-конце). Кодирующая последовательность может содержать, но без ограничения ими, кДНК из прокариотической или эукариотической мРНК, геномные последовательности ДНК из прокариотической или эукариотической ДНК и синтетические нуклеиновые кислоты. Стоп-кодон транскрипции обычно располагается с 3' стороны по отношению к кодирующей последовательности.

Как используется в данном документе, "последовательность промотора" представляет собой регуляторную область ДНК, способную связывать РНК-полимеразу и инициировать транскрипцию расположенной ниже (в 3' направлении) кодирующей или некодирующей последовательности. Для целей определения настоящего изобретения, последовательность промотора ограничена на ее 3'-конце сайтом инициации транскрипции и простирается выше (в 5' направлении), чтобы содержать минимальное число оснований или элементов, необходимых для инициации транскрипции на уровнях, детектируемых выше фона. В последовательности промотора возможно найти сайт инициации транскрипции, а также домены связывания белка, ответственные за связывание РНК-полимеразы. Эукариотические промоторы будут часто, но не всегда, содержать боксы "ТАТА" и боксы "САТ". Различные промоторы, в том числе инду-

цируемые промоторы, могут быть использованы для приведения в действие различных векторов по настоящему изобретению.

Промотор может быть постоянно активным промотором (т.е. промотором, который постоянно находится в активном/"ВКЛ" состоянии), он может быть индуцируемым промотором (т.е. промотором, состояние которого является активным/"ВКЛ" или неактивным/"ВЫКЛ", которым управляют с помощью внешнего сигнала, например, наличия определенной температуры, соединения или белка), он может быть пространственно ограниченным промотором (т.е. элементом управления транскрипции, энхансером и т.д.) (например, тканеспецифичный промотор, промотор, специфический к типу клеток и т.д.) и он может быть временно ограниченным промотором (т.е. промотор находится в состоянии "ВКЛ" или "ВЫКЛ" во время отдельных стадий эмбрионального развития или в течение определенных этапов биологического процесса, например, цикла фолликул волоса у мышей).

Подходящие промоторы могут быть получены из вирусов, и, следовательно, могут упоминаться как вирусные промоторы или они могут быть получены из любого организма, в том числе прокариотических или эукариотических организмов. Подходящие промоторы могут быть использованы для управления экспрессией с помощью любой РНК-полимеразы (например, pol I, pol II, pol III). Примеры промоторов включают, но не ограничиваются ими, ранний промотор SV40, промотор длинного концевой повтора вируса опухоли молочной железы мышей (LTR); основной поздний промотор аденовируса (Ad MLP); промотор вируса простого герпеса (HSV), промотор цитомегаловируса (CMV), такой как область немедленно-раннего промотора CMV (CMVIE), промотор вируса саркомы Рауса (RSV), небольшой ядерный промотор U6 человека (U6) (Miyagishi и соавт., *Nature Biotechnology* 20, 497-500 (2002)), усиленный промотор U6 (например, Xia и соавт., *Nucleic Acids Res.*; 2003 Sep 1; 31(17)), промотор H1 человека (H1) и подобные.

Примеры индуцируемых промоторов включают, но не ограничиваются ими, промотор РНК-полимеразы T7, промотор РНК-полимеразы T3, изопропил-бета-D-тиогактопиранозид (IPTG)-регулируемый промотор, индуцируемый лактозой промотор, промотор теплового шока, тетрациклин-регулируемый промотор, стероид-регулируемый промотор, металл-регулируемый промотор, промотор, регулируемый рецептором эстрогена и т.д. Индуцируемые промоторы могут регулироваться с помощью молекул, в том числе, но без ограничения ими, доксициклином; РНК-полимеразой, например, РНК-полимеразой T7; рецептором эстрогена; слиянием эстроген-рецептор; и т.д.

В некоторых вариантах реализации настоящего изобретения промотор представляет собой пространственно ограниченный промотор (т.е. специфический к типу клеток промотор, тканеспецифический промотор и т.д.), таким образом, что в многоклеточном организме промотор является активным в подмножестве определенных клеток (т.е. "ВКЛ"). Пространственно ограниченные промоторы также могут упоминаться как энхансеры, элементы контроля транскрипции, управляющие последовательности и т.д. Может быть использован любой пригодный пространственно ограниченный промотор, а выбор подходящего промотора (например, специфического к мозгу промотора, промотора, который стимулирует экспрессию в подгруппе нейронов, промотора, который запускает экспрессию в зародышевой линии, промотора, который запускает экспрессию в легких, промотора, который запускает экспрессию в мышцах, промотора, который запускает экспрессию в островковых клетках поджелудочной железы и т.д.) зависит от организма. Например, различные пространственно ограниченные промоторы, известны для растений, мух, червей, млекопитающих, мышей и т.д. Таким образом, пространственно ограниченный промотор может быть использован для того, чтобы регулировать экспрессию нуклеиновой кислоты, кодирующую сайт-специфический модифицирующий полипептид-мишень в широком множестве различных тканей и типов клеток, в зависимости от организма. Некоторые пространственно ограниченные промоторы также временно ограничены таким образом, что промотор находится в состоянии "ВКЛ" или "ВЫКЛ" во время определенных стадий эмбрионального развития или в течение определенных этапов биологического процесса (например, цикл фолликул волоса у мышей).

С целью иллюстрации, примеры пространственно ограниченных промоторов включают, но не ограничиваются ими, нейрон-специфические промоторы, адипоцит-специфические промоторы, кардиомиоцит-специфические промоторы, промоторы, специфические к гладким мышцам, промоторы, специфические к фоторецепторам и т.д. Нейрон-специфические пространственно ограниченные промоторы включают, но не ограничиваются ими, нейрон-специфический промотор енолазы (NSE) (см., например, EMBL HSENO2, X51956); промотор декарбоксилазы ароматических аминокислот (AADC); промотор нейрофиламента (см., например, GenBank HUMNFL, L04147); промотор синапсина (см., например, GenBank HUMSYNIB, M55301); промотор thy-1 (см., например, Chen и соавт. (1987) *Cell* 51:7-19; и Llewellyn, и соавт. (2010) *Nat. Med.* 16(10):1161-1166); промотор серотониновых рецепторов (см., например, GenBank S62283); промотор тирозингидроксилазы (TH) (см., например, Oh и соавт. (2009) *Gene Ther* 16:437; Sasaoka и соавт. (1992) *Mol Brain Res.* 16:274; Boundy и соавт. (1998) *J. Neurosci.* 18:9989; и Kaneda и соавт. (1991) *Neuron* 6:583-594); промотор GnRH (см., например, Radovick и соавт. (1991) *Proc. Natl. Acad. Sci. USA* 88:3402-3406); промотор L7 (см., например, Oberdick и соавт. (1990) *Science* 248:223-226); промотор DNMT (см., например, Bartge и соавт. (1988) *Proc. Natl. Acad. Sci. USA* 85:3648-3652); промотор энкефалина (см., например, Comb и соавт. (1988) *EMBO J.* 17:3793-3805); промотор основного белка миелина

(MBP); промотор Ca<sup>2+</sup>-кальмодулин-зависимой протеинкиназы II-альфа (CamKII $\alpha$ ) (см., например, Mayford и соавт. (1996) Proc. Natl. Acad. Sci. USA 93:13250; и Casanova и соавт. (2001) Genesis 31:37); энхансер CMV/промотор фактора роста тромбоцитов- $\beta$  (см., например, Liu и соавт. (2004) Gene Therapy 11:52-60); и подобные.

Адипоцит-специфические пространственно ограниченные промоторы включают, но не ограничиваются ими, промотор/энхансер гена aP2, например, область от -5,4 т. п. н. до +21 и.о. гена aP2 человека (см., например, Tozzo и соавт. (1997) Endocrinol. 138:1604; Ross и соавт. (1990) Proc. Natl. Acad. Sci. USA 87:9590; и Pavjani и соавт. (2005) Nat. Med. 11:797); промотор переносчика глюкозы-4 (GLUT4) (см., например, Knight и соавт. (2003) Proc. Natl. Acad. Sci. USA 100:14725); промотор транслоказы жирных кислот (FAT/CD36) (см., например, Kuriki и соавт. (2002) Biol. Pharm. Bull. 25:1476; и Sato и соавт. (2002) J. Biol. Chem. 277:15703); промотор стеарил-КоА-дегидрогеназы-1 (SCD1) (Tabor и соавт. (1999) J. Biol. Chem. 274:20603); промотор лептина (см., например, Mason и соавт. (1998) Endocrinol. 139:1013; и Chen и соавт. (1999) Biochem. Biophys. Res. Comm. 262:187); промотор адипонектина (см., например, Kita и соавт. (2005) Biochem. Biophys. Res. Comm. 331:484; и Chakrabarti (2010) Endocrinol. 151:2408); промотор адипина (см., например, Plan, и соавт. (1989) Proc. Natl. Acad. Sci. USA 86:7490); промотор резистина (см., например, Seo и соавт. (2003) Molec. Endocrinol. 17:1522); и подобные.

Кардиомиоцит-специфические пространственно ограниченные промоторы включают, но не ограничиваются ими, для управления последовательностями, полученными из следующих генов: легкую цепь миозина-2,  $\alpha$ -тяжелую цепь миозина, AE3, сердечный тропонин C, сердечный актин и т.п. Franz и соавт. (1997) Cardiovasc. Res. 35:560-566; Robbins и соавт. (1995) Ann. N.Y. Acad. Sci. 752:492-505; Linn и соавт. (1995) Circ. Res. 76:584-591; Parmacek и соавт. (1994) Mol. Cell. Biol. 14:1870-1885; Hunter и соавт. (1993) Hypertension 22:608-617; и Sartorelli и соавт. (1992) Proc. Natl. Acad. Sci. USA 89:4047-4051.

Специфические к гладким мышцам пространственно ограниченные промоторы включают, но не ограничиваются ими, промотор SM22a (см., например, Akyurek и соавт. (2000) Mol. Med. 6:983; и патент США № 7169874); промотор смутелина (см., например, WO 2001/018048); промотор  $\alpha$ -гладкомышечного актина; и подобные. Например, область в 0,4 т.п.н. промотора SM22 $\alpha$ , в пределах которой лежат два элемента CARG, были показаны для того, чтобы опосредовать экспрессию, специфическую к гладкомышечным клеткам сосудов (см., например, Kim, и соавт. (1997) Mol. Cell. Biol. 17, 2266-2278; Li, и соавт., (1996) J. Cell Biol. 132, 849-859; и Moessler, и соавт. (1996) Development 122, 2415-2425).

Специфические к фоторецепторам пространственно ограниченные промоторы включают, но не ограничиваются ими, промотор родопсина; промотор родопсинкиназа (Young и соавт. (2003) Ophthalmol. Vis. Sci. 44:4076); промотор гена бета фосфодиэстеразы (Nicoud и соавт. (2007) J. Gene Med. 9:1015); промотор гена пигментного ретинита (Nicoud 32 и соавт. (2007) ранее); энхансер гена межфоторецепторного ретинол-связывающего белка (IRBP) (Nicoud и соавт. (2007) ранее); промотор гена IRBP (Yokooyama и соавт. (1992) Exp Eye Res. 55:225); и подобные.

Термины "регуляторные последовательности ДНК", "контрольные элементы" и "регуляторные элементы", которые используют в данном документе взаимозаменяемо, относятся к последовательностям регуляции транскрипции и трансляции, таким как промоторы, энхансеры, сигналы полиаденилирования, терминаторы, сигналы деградации белка, и подобные, которые обеспечивают и/или регулируют транскрипцию некодирующей последовательности (например, РНК, нацеленной на ДНК) или кодирующей последовательности (например, сайт-специфического модифицирующего полипептида или полипептида Cas9/Csn1) и/или регулируют трансляцию закодированного полипептида.

Термин "встречающаяся в природе" или "немодифицированный", как используют в данном документе, применительно к нуклеиновой кислоте, полипептиду, клетке или организму, соответствует нуклеиновой кислоте, полипептиду, клетке или организму, которые обнаружены в природе. Например, полипептидная или полинуклеотидная последовательность, присутствующая в организме (в том числе вирусов), которая может быть выделена из природного источника и которая не была намеренно модифицирована человеком в лаборатории, является встречающейся в природе.

Термин "гибридный", как используют в данном документе применительно к нуклеиновой кислоте или полипептиду, относится к двум компонентам, которые определяют структурами, полученными из различных источников. Например, если "гибридный" используют в контексте гибридного полипептида (например, гибридный белок Cas9/Csn1), гибридный полипептид содержит аминокислотные последовательности, которые получены из различных полипептидов. Гибридный полипептид может содержать либо измененные, либо встречающиеся в природе полипептидные последовательности (например, в первую аминокислотную последовательность из модифицированного или немодифицированного белка Cas9/Csn1; и вторую аминокислотную последовательность, отличную от белка Csn1/Cas9). Аналогичным образом, "гибридный" в контексте полинуклеотида, кодирующего гибридный полипептид, содержит нуклеотидные последовательности, полученные из различных кодирующих областей (например, первой нуклеотидной последовательности, кодирующей модифицированный или немодифицированный белок Cas9/Csn1; и второй нуклеотидной последовательности, кодирующей полипептид, отличный от белка Cas9/Csn1).

Термин "гибридный полипептид" соответствует полипептиду, который сделан путем комбинации (т.е. "слиянием") двух отличных разделенных сегментов аминокислотной последовательности, обычно посредством вмешательства человека. Полипептид, который содержит гибридную аминокислотную последовательность, представляет собой гибридный полипептид. Некоторые гибридные полипептиды могут упоминаться как "слитые варианты".

"Гетерологичный", как используют в данном документе, означает нуклеотидную последовательность или полипептид, который не обнаружен в нативной нуклеиновой кислоте или белке соответственно. Например, в гибридном белке Cas9/Csn1, РНК-связывающий домен встречающегося в природе бактериального полипептида Cas9/Csn1 (или его варианта) может быть слит с гетерологичной полипептидной последовательностью (например, полипептидная последовательность белка, отличного от Cas9/Csn1, или полипептидная последовательность другого организма). Гетерологичная полипептидная последовательность может проявлять активность (например, ферментативную активность), что также будет продемонстрировано с помощью гибридного белка Cas9/Csn1 (например, активность метилтрансферазы, активность ацетилтрансферазы, киназная активность, активность убиквитинирования и т.д.). Гетерологичная полипептидная последовательность может быть связана с природной последовательностью нуклеиновой кислоты (или ее вариантом) (например, с помощью генной инженерии) для того, чтобы создать гибридную нуклеотидную последовательность, кодирующую гибридный полипептид. В качестве другого примера, в варианте слитого сайт-специфического полипептида Cas9, вариант сайт-специфического полипептида Cas9 может быть слит с гетерологичным полипептидом (например, полипептидом, отличным от Cas9), который проявляет активность, которая также будет проявляться слитым вариантом сайт-специфического полипептида Cas9. Гетерологичная последовательность нуклеиновой кислоты может быть связана с одним из вариантов сайт-специфического полипептида Cas9 (например, с помощью генной инженерии) для того, чтобы создать нуклеотидную последовательность, кодирующую слитый вариант сайт-специфического полипептида Cas9.

"Рекомбинантный", как используют в данном документе, означает, что конкретная нуклеиновая кислота (ДНК или РНК) является продуктом различных комбинаций клонирования, рестрикции, полимеразной цепной реакции (PCR) и/или этапов лигирования, в результате чего полученная конструкция имеет в структуре кодирующую или некодирующую последовательность, которая отличается от эндогенных нуклеиновых кислот, обнаруженных в природных системах. Последовательности ДНК, кодирующие полипептиды, могут быть собраны из фрагментов кДНК или из серии синтетических олигонуклеотидов для того, чтобы обеспечить синтетическую нуклеиновую кислоту, которая может быть экспрессирована из рекомбинантной транскрипционной единицы, содержащейся в клетке или в *in vitro* системе транскрипции и трансляции. Геномная ДНК, которая содержит соответствующие последовательности, также может быть использована при формировании рекомбинантного гена или транскрипционной единицы. Последовательности нетранслируемой ДНК могут присутствовать 5' или 3' от открытой рамки считывания, где такие последовательности не мешают манипулированию или экспрессии кодирующих областей, и действительно могут действовать с той целью, чтобы модулировать получение желаемого продукта с помощью различных механизмов (см. "регуляторные последовательности ДНК", выше). Кроме того, последовательности ДНК, кодирующие РНК (например, РНК, нацеленную на ДНК), которые не транслируются, также могут считаться рекомбинантными. Таким образом, например, термин "рекомбинантная" нуклеиновая кислота относится к нуклеиновой кислоте, не встречающейся в природе, например, полученной с помощью искусственной комбинации двух различных разделенных сегментов последовательности посредством вмешательства человека. Эту искусственную комбинацию часто осуществляют либо с помощью средств химического синтеза, либо с помощью искусственного манипулирования выделенными сегментами нуклеиновых кислот, например, с помощью способов генной инженерии. Подобное обычно осуществляют для замены кодона на кодон, кодирующий ту же самую аминокислоту, консервативную аминокислоту или неконсервативную аминокислоту. Как вариант, это выполняют для того, чтобы соединить вместе сегменты нуклеиновой кислоты с необходимыми функциями с получением необходимой комбинации функций. Эту искусственную комбинацию часто осуществляют либо с помощью средств химического синтеза, либо с помощью искусственного манипулирования выделенными сегментами нуклеиновых кислот, например, с помощью способов генной инженерии. Если рекомбинантный полинуклеотид кодирует полипептид, последовательность кодируемого полипептида может быть встречающейся в природе ("дикого типа") или может представлять собой вариант (например, мутант) встречающейся в природе последовательности. Таким образом, термин "рекомбинантный" полипептид не обязательно относится к полипептиду, последовательность которого, естественно, не встречается в природе. Вместо этого, "рекомбинантный" полипептид, кодируют с помощью рекомбинантной последовательностью ДНК, но последовательность полипептида может быть встречающейся в природе ("дикого типа") или не встречающейся в природе (например, вариант, мутант и т.д.). Таким образом, "рекомбинантный" полипептид является результатом вмешательства человека, но может быть встречающейся в природе аминокислотной последовательностью.

"Вектор" или "вектор экспрессии" представляет собой репликон, такой как плаزمид, фаг, вирус или космида, к которому другой сегмент ДНК, т.е. "вставка", может быть присоединен таким образом,

чтобы вызвать в клетке репликацию присоединенного сегмента.

"Кассета экспрессии" содержит кодирующую ДНК последовательность, функционально связанную с промотором. Термин "функционально связанный" относится к взаиморасположению, в котором описанные компоненты находятся в связи, позволяющей им функционировать предназначенным им способом. Например, промотор функционально связан с кодирующей последовательностью, при условии, что промотор влияет на его транскрипцию или экспрессию.

Термины "рекомбинантный вектор экспрессии" или "конструкция ДНК" используют в данном документе взаимозаменяемо для обозначения молекулы ДНК, которая содержит вектор и по меньшей мере одну вставку. Рекомбинантные векторы экспрессии, как правило, получают с целью экспрессии и/или репродуцирования вставки(ок) или с целью построения других рекомбинантных последовательностей нуклеотидов. Вставка(и) может быть или может не быть функционально связана с промоторной последовательностью и может быть или может не быть функционально связана с регуляторными последовательностями ДНК.

Клетка была "генетически модифицирована" или "трансформирована" или "трансфицирована" с помощью экзогенной ДНК, например, рекомбинантным вектором экспрессии, если такую ДНК вводят внутрь клетки. Наличие экзогенной ДНК приводит к постоянному или временному генетическому изменению. Трансформирующая ДНК может быть или может не быть интегрирована (ковалентно связана) в геном клетки. В прокариотах, дрожжах и клетках млекопитающих, например, трансформирующая ДНК, может быть сохранена на эписомальном элементе, таком как плазмида. По отношению к эукариотическим клеткам, стабильно трансформированная клетка представляет собой такую клетку, в которой трансформирующая ДНК интегрирована в хромосому таким образом, что она наследуется дочерними клетками посредством хромосомной репликации. Эта стабильность демонстрируется с помощью способности эукариотической клетки быть основателем клеточных линий или клонов, которые содержат популяцию дочерних клеток, содержащих трансформирующую ДНК. "Клон" представляет собой популяцию клеток, полученную из одной клетки или общего предка с помощью митоза. "Линия клеток" представляет собой клон первичной клетки, который способен к стабильному росту *in vitro* в течение многих поколений.

Подходящие способы генетической модификации (также упоминается как "трансформирование") включают, например, вирусную или бактериофаговую инфекцию, трансфекцию, конъюгацию, слияние протопластов, липофекцию, электропорацию, осаждение фосфатом кальция, опосредованную полиэтиленгликолином (PEI) трансфекцию, DEAE-декстран опосредованную трансфекцию, опосредованную липосомами трансфекцию, технологию генной пушки, осаждение фосфатом кальция, прямую микроинъекцию, опосредованную наночастицами доставку нуклеиновой кислоты (см., например, Panyam и COELBT., *Adv Drug Deliv Rev.* 2012 Sep 13. pii: S0169-409X(12)00283-9. doi: 10.1016/j.addr.2012.09.023) и подобные.

Выбор способа генетической модификации, как правило, зависит от типа клетки, которую подвергают трансформированию, и обстоятельств, при которых происходит трансформация (например, *in vitro*, *ex vivo* или *in vivo*). Общее обсуждение этих способов может быть найдено в Ausubel, и соавт., *Short Protocols in Molecular Biology*, 3rd ed., Wiley & Sons, 1995.

"ДНК-мишень", как используют в данном документе, представляет собой полинуклеотид ДНК, который содержит "сайт-мишень" или "последовательность-мишень". Термины "сайт-мишень", "последовательность-мишень" или "протоспейсер ДНК-мишень" используют в данном документе взаимозаменяемо для обозначения последовательности нуклеиновой кислоты, присутствующей в ДНК-мишени, с которой будет связываться ДНК-нацеленный сегмент РНК-мишени, нацеленной на ДНК (см. фиг. 1 и 39), при условии существования достаточных для связывания условий. Например, сайт-мишень (или последовательность-мишень) 5'-GAGCATATC-3' (SEQ ID NO://) в пределах ДНК-мишени подвергается действию (или связывается с помощью, или гибридизуется с, или является комплементарной) последовательности 5'-GAU AUGCUC-3' (SEQ ID NO://). Обычно в клетке присутствуют подходящие для связывания ДНК/РНК условия, в том числе, физиологические условия. Другие подходящие для связывания ДНК/РНК условия (например, условия в бесклеточной системе), известны в данной области; см., например, Sambrook, ранее. Цепь ДНК-мишени, которая является комплементарной и гибридизуется с РНК, нацеленной на ДНК, упоминается как "комплементарная цепь", а цепь ДНК-мишени, которая является комплементарной к "комплементарной цепи" (и, следовательно, не комплементарна РНК, нацеленной на ДНК), упоминается как "некомплементарная цепь" или "некомплементарная цепь" (см. фиг. 12).

Под "сайт-специфическим модифицирующим полипептидом", "РНК-связывающим сайт-специфическим полипептидом", "РНК-связывающим сайт-специфическим модифицирующим полипептидом" или "сайт-специфическим полипептидом" подразумевают полипептид, который связывает РНК и ориентирован на специфическую последовательность ДНК. Сайт-специфический модифицирующий полипептид, как описано в данном документе, нацелен на специфическую последовательность ДНК с помощью молекулы РНК, с которой он связан. Молекула РНК содержит последовательность, комплементарную к последовательности-мишени в пределах исследуемой ДНК, тем самым направляя связанный полипептид на специфическую локализацию в пределах ДНК-мишени (последовательности-мишени).

Под "расщеплением" подразумевают разрывание ковалентного остова молекулы ДНК. Расщепление может быть инициировано различными способами, в том числе, но без ограничения ими, ферментативным или химическим гидролизом фосфодиэфирной связи. Оба, и одноцепочечное расщепление, и двухцепочечное расщепление, являются возможными, а двухцепочечное расщепление может происходить в результате двух отдельных случаев одноцепочечного расщепления. Расщепление ДНК может привести к получению либо тупых концов, либо ступенчатых концов. В некоторых вариантах реализации настоящего изобретения комплекс, который содержит РНК, нацеленную на ДНК, и сайт-специфический модифицирующий полипептид, используют для целевого двухцепочечного расщепления ДНК.

"Нуклеаза" и "эндонуклеаза" в данном документе используют взаимозаменяемо для обозначения фермента, который обладает каталитической активностью для расщепления ДНК.

Под "доменом расщепления", "активным доменом" или "нуклеазным доменом" нуклеазы подразумевают полипептидную последовательность или домен в пределах нуклеазы, который обладает каталитической активностью для расщепления ДНК. Домен расщепления может содержаться в отдельной полипептидной цепи или активность расщепления может быть следствием ассоциации двух (или более) полипептидов. Отдельный домен нуклеазы может состоять из более чем одного выделенного участка аминокислот в пределах данного полипептида.

Молекула РНК, которая связывается с сайт-специфическим модифицирующим полипептидом и нацеливает полипептид на специфическую локализацию в пределах ДНК-мишень, в данном документе называется "РНК, нацеленная на ДНК", или "полинуклеотид РНК, нацеленной на ДНК" (также называемый в данном документе "направляющая" РНК" или "gРНК"). РНК, нацеленная на ДНК, состоит из двух сегментов, "нацеленного на ДНК сегмента" и "белок-связывающего сегмента". Под "сегментом" подразумевают сегмент/часть/область молекулы, например, непрерывный участок нуклеотидов в РНК. Сегмент может также означать область/часть комплекса таким образом, что сегмент может содержать области более чем одной молекулы. Например, в некоторых случаях белок-связывающий сегмент (описанный ниже) РНК, нацеленной на ДНК, представляет собой одну молекулу РНК и поэтому белок-связывающий сегмент содержит область этой молекулы РНК. В других случаях, белок-связывающий сегмент (описанный ниже) РНК, нацеленной на ДНК, состоит из двух отдельных молекул, которые гибридизуются в области комплементарности. В качестве иллюстративного, неограничивающего примера, белок-связывающий сегмент РНК, нацеленной на ДНК, которая содержит две отдельные молекулы, может содержать (i) 40-75 пар оснований первой молекулы РНК, которая имеет 100 пар оснований в длину; и (ii) 10-25 пар оснований второй молекулы РНК, которая имеет 50 пар оснований в длину. Определение "сегмент", если в конкретном контексте не определено иное, не ограничивается определенным количеством общих пар оснований, не ограничивается каким-либо конкретным числом пар оснований данной молекулы РНК, не ограничивается частным числом отдельных молекул в пределах комплекса и может содержать участки молекулы РНК, которые имеют любую общую длину и могут содержать или могут не содержать области с комплементарностью к другим молекулам.

Нацеленный на ДНК сегмент (или "нацеленная на ДНК последовательность") содержит нуклеотидную последовательность, которая комплементарна определенной последовательности в пределах ДНК-мишени (комплементарная цепь ДНК-мишени). Белок-связывающий сегмент (или "белок-связывающая последовательность") взаимодействует с сайт-специфическим модифицирующим полипептидом. Если сайт-специфический модифицирующий полипептид представляет собой родственный Cas9 или Cas9 полипептид (более подробно описанный ниже), сайт-специфическое расщепление ДНК-мишени происходит в локализациях, как (i) комплементарным спариванием оснований между РНК, нацеленной на ДНК, и ДНК-мишенью; так и (ii) коротким мотивом (обозначенным как смежный с протоспейсером мотив (РАМ)) в ДНК-мишени.

Белок-связывающий сегмент РНК-мишени, нацеленной на ДНК, содержит два комплементарных участка нуклеотидов, которые гибридизуются друг с другом, для того, чтобы сформировать дуплекс двухцепочечной РНК (дуплекс дцРНК).

В некоторых вариантах реализации настоящего изобретения нуклеиновая кислота-мишень (например, РНК, нацеленная на ДНК, нуклеиновая кислота, которая содержит нуклеотидную последовательность, кодирующую РНК, нацеленную на ДНК; нуклеиновая кислота, кодирующая сайт-специфический полипептид; и т.д.) содержит в себя модификацию или последовательность, которая обеспечивает дополнительный желательный признак (например, модифицированную или регулируемую стабильность; субклеточное нацеливание; слежение, например, флуоресцентную метку; связывающий сайт в случае белка или белкового комплекса; и т.д.). Неограничивающие примеры содержат 5'-кэп (например, 7-метилгуанилатный кэп (m7G)); 3'-полиаденилированный хвост (т.е. 3'-поли(А) хвост); последовательность рибосвитча (например, для обеспечения регулируемой стабильности и/или регулируемой доступности для белков и/или белковых комплексов); последовательность контроля устойчивости; последовательность, которая образует дуплекс дцРНК (т.е. шпильку)); модификацию или последовательность, которая нацеливает РНК на внутриклеточную локализацию (например, ядро, митохондрию, хлоропласты и подобные); модификацию или последовательность, которую обеспечивает возможность слежения (например, прямую конъюгацию с флуоресцентной молекулой, конъюгацию с фрагментом, который облег-

чает флуоресцентную детекцию, последовательность, которая позволяет осуществлять флуоресцентную детекцию и др.); модификацию или последовательность, которая обеспечивает связывающий сайт для белков (например, белков, которые действуют на ДНК, в том числе активаторы транскрипции, репрессоры транскрипции, метилтрансферазы ДНК, деметилазы ДНК, гистонацетилтрансферазы, гистондеацетилазы и подобные); и их комбинации.

В некоторых вариантах реализации настоящего изобретения РНК, нацеленная на ДНК, содержит дополнительный сегмент либо на 5'-, либо на 3'-конце, что обеспечивает любой из признаков, описанных выше. Например, подходящий третий сегмент содержит: 5'-кэп (например, 7-метилгуанилатный кэп (m7G)); 3'-полиаденилированный хвост (т.е. 3'-поли(А) хвост); последовательность рибосвитча (например, для обеспечения регулируемой стабильности и/или регулируемой доступности для белков и белковых комплексов); последовательность контроля устойчивости; последовательность, которая образует дуплекс дцРНК (т.е. шпильку)); модификацию или последовательность, которая нацеливает РНК на внутриклеточную локализацию (например, ядро, митохондрию, хлоропласты и подобные); модификацию или последовательность, которую обеспечивает возможность слежения (например, прямую конъюгацию с флуоресцентной молекулой, конъюгацию с фрагментом, который облегчает флуоресцентную детекцию, последовательность, которая позволяет осуществлять флуоресцентную детекцию и др.); модификацию или последовательность, которая обеспечивает связывающий сайт в случае белков (например, белков, которые действуют на ДНК, в том числе активаторы транскрипции, репрессоры транскрипции, метилтрансферазы ДНК, деметилазы ДНК, гистонацетилтрансферазы, гистондеацетилазы и подобные); и их комбинации.

РНК-мишень, нацеленная на ДНК, и сайт-специфический модифицирующий полипептид-мишень (т.е. сайт-специфический полипептид) образуют комплекс (т.е. связываются с помощью нековалентных взаимодействий). РНК, нацеленная на ДНК, обеспечивает комплексу специфичность к мишени за счет того, что содержит нуклеотидную последовательность, которая является комплементарной к последовательности ДНК-мишени. Сайт-специфический модифицирующий полипептид комплекса обеспечивает сайт-специфическую активность. Другими словами, сайт-специфический модифицирующий полипептид направляют к последовательности ДНК-мишени (например, последовательности-мишени в нуклеиновой кислоте, находящейся в составе хромосомы; последовательности-мишени в нуклеиновой кислоте, находящейся вне хромосомы, например, эписомной нуклеиновой кислоте, мини-кольце и т.д.; последовательности-мишени в митохондриальной нуклеиновой кислоте; последовательности-мишени в хлоропластной нуклеиновой кислоте; последовательности-мишени в плазмиде; и т.д.) в силу его ассоциации с белок-связывающим сегментом РНК, нацеленной на ДНК.

В некоторых вариантах реализации настоящего изобретения РНК-мишень, нацеленная на ДНК, содержит две отдельные молекулы РНК (полинуклеотиды РНК: "РНК-активатор" и "РНК-нацеливающий", см. ниже) и упоминается в данном описании как "РНК, нацеленная на ДНК, в виде двойной молекулы" или "РНК, нацеленная на ДНК, в виде двух молекул". В других вариантах реализации настоящего изобретения РНК, нацеленная на ДНК, представляет собой отдельную молекулу РНК (отдельный полинуклеотид РНК) и в данном документе соответствует "РНК, нацеленной на ДНК, в виде одиночной молекулы", "направляющей" РНК в виде одиночной молекулы" или "sgРНК. Термин "РНК, нацеленная на ДНК", или "gРНК" является обширным, соответствующим как РНК, нацеленной на ДНК, в виде двойной молекулы, так и РНК, нацеленной на ДНК, в виде одиночной молекулы (т.е. sgРНК).

Иллюстративная РНК, нацеленная на ДНК, в виде двойной молекулы, содержит crРНК-подобную ("РНК CRISPR", "РНК-нацеливающую", "crРНК" или "повтор crРНК") молекулу и соответствующую tracrРНК-подобную ("транседействующую РНК CRISPR", "РНК-активатор" или "tracrРНК") молекулу. crРНК-подобная молекула (РНК-нацеливающая) содержит как нацеленный на ДНК-сегмент (одноцепочечный) РНК, нацеленной на ДНК, так и участок нуклеотидов ("дуплекс-образующий сегмент"), который образует одну половину дуплекса дцРНК белок-связывающего сегмента РНК, нацеленной на ДНК. Соответствующая tracrРНК-подобная молекула (РНК-активатор) содержит участок нуклеотидов (дуплекс-образующий сегмент), который образует другую половину дуплекса дцРНК белок-связывающего сегмента РНК, нацеленной на ДНК.

Другими словами, участок нуклеотидов crРНК-подобной молекулы является комплементарным и гибридизуется с участком нуклеотидов tracrРНК-подобной молекулы с образованием дуплекса дцРНК белок-связывающего домена РНК, нацеленной на ДНК. Таким образом, каждая crРНК-подобная молекула, можно сказать, имеет соответствующую tracrРНК-подобную молекулу. crРНК-подобная молекула дополнительно обеспечивает одноцепочечный ДНК-нацеленный сегмент. Таким образом, crРНК-подобная и tracrРНК-подобная молекула (в качестве соответствующей пары) гибридизуются с образованием РНК, нацеленной на ДНК. Точная последовательность данной молекулы crРНК или tracrРНК характерна для видов, в которых находят молекулы РНК. Различные crРНК и tracrРНК проиллюстрированы соответствующими комплементарными парами на фиг. 8. РНК-мишень, нацеленная на ДНК, в виде двойной молекулы, может содержать любую соответствующую пару crРНК и tracrРНК. РНК-мишень, нацеленная на ДНК, в виде двойной молекулы, может содержать любую соответствующую пару crРНК и tracrРНК.

Термин "РНК-активатор" используют в данном документе для обозначения *trans*РНК-подобной молекулы РНК, нацеленной на ДНК, в виде двойной молекулы. Термин "РНК-нацеливающая" используют в данном документе для обозначения *cis*РНК-подобной молекулы РНК, нацеленной на ДНК, в виде двойной молекулы. Термин "дуплекс-образующий сегмент" используется в данном документе для обозначения участка нуклеотидов РНК-активатора или РНК-нацеливающей, который способствует образованию дуплекса *dc*РНК с помощью гибридизации с участком нуклеотидов соответствующей молекулы РНК-активатора или РНК-нацеливающей. Другими словами, РНК-активатор содержит дуплекс-образующий сегмент, который является комплементарным дуплекс-образующему сегменту соответствующей РНК-нацеливающей. Таким образом, РНК-активатор содержит дуплекс-образующий сегмент, тогда как РНК-нацеливающая содержит как дуплекс-образующий сегмент, так и ДНК-нацеленный сегмент РНК, нацеленной на ДНК. Таким образом, РНК-мишень, нацеленная на ДНК, в виде двойной молекулы, может состоять из любой соответствующей пары РНК-активатора и РНК-нацеливающей.

"Клетка-хозяин", как используют в данном документе, обозначает то, находящуюся *in vivo* или *in vitro* эукариотическую клетку, прокариотическую клетку (например, бактериальную клетку или клетку архей) или клетку из многоклеточных организмов (например, клеточную линию) культивируемую в одноклеточном состоянии, где эукариотические или прокариотические клетки могут быть, или были использованы, в качестве реципиентов нуклеиновой кислоты, и включает потомство исходной клетки, которая была трансформирована с помощью нуклеиновой кислоты. Понятно, что потомство отдельной клетки не обязательно может быть полностью идентичным по морфологии, геномному или общему комплементу ДНК в качестве дополнения исходного родителя, ввиду естественной, случайной или преднамеренной мутации. "Рекомбинантная клетка-хозяин" (также упоминается как "генетически модифицированная клетка-хозяин") является клеткой-хозяином, в которую вводят гетерологичную нуклеиновую кислоту, например, вектор экспрессии. Например, целевая бактериальная клетка-хозяин является генетически модифицированной бактериальной клеткой-хозяином в соответствии с введением в подходящую бактериальную клетку-хозяина экзогенной нуклеиновой кислоты (например, плазмиды или рекомбинантного вектора экспрессии), а целевая эукариотическая клетка-хозяин является генетически модифицированной эукариотической клеткой-хозяином (например, зародышевой клеткой млекопитающего) в соответствии с введением в подходящую эукариотическую клетку-хозяина экзогенной нуклеиновой кислоты.

Термин "стволовая клетка" используют в данном документе для обозначения клетки (например, растительной стволовой клетки, стволовой клетки позвоночного), которая обладает способностью как к самообновлению, так и генерированию дифференцированного типа клеток (см. Morrison и соавт. (1997) *Cell* 88:287-298). В контексте клеточного онтогенеза, прилагательное "дифференцированный" или "дифференцирование" является относительным понятием. "Дифференцированная клетка" является клеткой, которая продвинулась дальше по пути развития, чем клетка, с которой она сравнивается. Таким образом, плюрипотентные стволовые клетки (описанные ниже) могут дифференцироваться в линейно-специфические клетки-предшественники (например, мезодермальные стволовые клетки), которые, в свою очередь, могут дифференцироваться в клетки, которые дополнительно ограничены (например, предшественники нейронов), которые могут дифференцироваться в зрелые клетки (т.е. окончательно дифференцированные клетки, например, нейроны, кардиомиоциты и т.д.), которые играют характерную роль в определенном типе ткани, и могут сохранять, или могут не сохранять, способность к пролиферации в дальнейшем. Стволовые клетки могут быть охарактеризованы как присутствием специфических маркеров (например, белки, РНК и т.д.), так и отсутствием специфических маркеров. Стволовые клетки могут быть также определены с помощью функциональных анализов как *in vitro*, так и *in vivo*, в частности, анализов, относящихся к способности стволовых клеток давать начало потомству различных дифференцированных клеток.

Стволовые клетки, представляющие интерес, включают плюрипотентные стволовые клетки (PSC). Термин "плюрипотентная стволовая клетка" или "PSC" используют в данном документе для обозначения стволовой клетки, способной продуцировать все типы клеток организма. Таким образом, PSC может давать клетки всех зародышевых слоев организма (например, эндодермы, мезодермы и эктодермы позвоночного животного). Плюрипотентные клетки способны образовывать тератомы и вносить свой вклад в ткани эктодермы, мезодермы или энтодермы в живом организме. Плюрипотентные стволовые клетки растений способны давать все типы клеток растения (например, клетки корня, стебля, листа и т.д.).

PSC животных могут быть получены несколькими различными способами. Например, эмбриональные стволовые клетки (ESC) получают из внутренней клеточной массы эмбриона (Thomson и соавт., *Science*. 1998 Nov 6; 282(5391): 1145-7), в то время как индуцированные плюрипотентные стволовые клетки (iPSC) являются производными соматических клеток (Takahashi и соавт., *Cell*. 2007 Nov 30; 131(5):861-72; Takahashi и соавт., *Nat Protoc*. 2007; 2(12):3081-9; Yu и соавт., *Science*. 2007 Dec 21; 318(5858):1917-20. *Epub* 2007 Nov 20). Поскольку термин PSC относится к плюрипотентным стволовым клеткам, независимо от их происхождения, термин PSC охватывает термины ESC и iPSC, так же как и термин эмбриональные зародышевые стволовые клетки (EGSC), который представляет собой еще один пример PSC. PSC могут существовать в виде установленной клеточной линии, они могут быть получены непосред-

венно из первичной эмбриональной ткани или они могут быть получены из соматической клетки. PSC могут быть клетками-мишенями по способам, описанным в данном документе.

Под "эмбриональной стволовой клеткой" (ESC) подразумевают PSC, которую выделяют из эмбриона, как правило, из внутренней клеточной массы бластоцисты. Линии ESC приведены в NIH Human Embryonic Stem Cell Registry, например, hESBGN-01, hESBGN-02, hESBGN-03, hESBGN-04 (BresaGen, Inc.); HES-1, HES-2, HES-3, HES-4, HES-5, HES-6 (ES Cell International); Miz-hES1 (MizMedi Hospital-Seoul National University); HSF-1, HSF-6 (University of California at San Francisco); и H1, H7, H9, H13, H14 (Wisconsin Alumni Research Foundation (WiCell Research Institute)). Представляющие интерес стволовые клетки также содержат эмбриональные стволовые клетки других приматов, таких как стволовые клетки макаки-резуса и стволовые клетки марьшишки. Стволовые клетки могут быть получены от любых видов млекопитающих, например, человека, лошади, быка, свиньи, собаки, кошки, грызуна, например, мыши, крысы, хомяка, примата и т.д. (Thomson и соавт. (1998) *Science* 282:1145; Thomson и соавт. (1995) *Proc. Natl. Acad. Sci. USA* 92:7844; Thomson и соавт. (1996) *Biol. Reprod.* 55:254; Shambloott и соавт., *Proc. Natl. Acad. Sci. USA* 95:13726, 1998). В культуре, ESC обычно вырастают в виде плоских колоний с большими ядерно-цитоплазматическими коэффициентами, определенными границами и выделяющимися ядрышками. Кроме того, ESC экспрессируют SSEA-3, SSEA-4, TRA-1-60, TRA-1-81 и щелочную фосфатазу, но не SSEA-1. Примеры способов получения и характеристики ESC могут быть найдены, например, в патенте США № 7029913, патенте США № 5843780 и в патенте США № 6200806, описания которых содержатся в настоящем раскрытии в виде ссылки. Способы, в случае пролиферирующих hESC в недифференцированной форме, описывают в WO 99/20741, WO 01/51616 и WO 03/020920.

Под "эмбриональной зародышевой стволовой клеткой" (EGSC), "эмбриональной зародышевой клеткой" или "клеткой EG" подразумевают PSC, который является производным половых клеток и/или зародышевых клеток-предшественников, например, эмбриональных стволовых клеток, т.е. тех, которые станут спермой и яйцеклетками. Эмбриональные стволовые клетки (клетки EG), как полагают, обладают сходными с эмбриональными стволовыми клетками свойствами, как описано выше. Примеры способов получения и характеристики клеток EG могут быть найдены, например, в патенте США № 7153684; Matsui, Y., и соавт., (1992) *Cell* 70:841; Shambloott, M., и соавт. (2001) *Proc. Natl. Acad. Sci. USA* 98: 113; Shambloott, M., и соавт. (1998) *Proc. Natl. Acad. Sci. USA*, 95:13726; и Koshimizu, U., и соавт. (1996) *Development*, 122: 1235, раскрытия которых содержатся в настоящем документе в виде ссылки.

Под "индуцированной плюрипотентной стволовой клеткой" или "iPSC" подразумевают PSC, которая является производной от клетки, которая не является PSC (т.е. из клетки, которая является дифференцированной по отношению к PSC). iPSC могут быть получены из нескольких различных типов клеток, в том числе окончательно дифференцированных клеток. iPSC имеют морфологию, подобную клеткам ES, растут в виде плоских колоний с большими ядерно-цитоплазматическими коэффициентами, определенными границами и выделяющимися ядрами. Кроме того, iPSC экспрессируют один или несколько ключевых маркеров плюрипотентности, известных любому специалисту в данной области техники, в том числе, но без ограничения ими, щелочную фосфатазу, SSEA3, SSEA4, Sox2, Oct3/4, Nanog, TRA160, TRA181, TDGF 1, Dmmt3b, FoxD3, GDF3, Cyp26al, TERT и zfp42. Примеры способов получения и характеристики iPSC могут быть найдены, например, в публикациях патентов США № US20090047263, US20090068742, US20090191159, US20090227032, US20090246875 и US20090304646, раскрытия которых содержатся в настоящем документе в виде ссылки. Как правило, для создания IpSC соматические клетки обрабатывают факторами перепрограммирования (например, Oct4, SOX2, KLF4, MYC, Nanog, Lin28 и т.д.), известными в данной области техники для перепрограммирования соматических клеток, чтобы они стали плюрипотентными стволовыми клетками.

Под "соматической клеткой" подразумевают любую клетку в организме, которая, в отсутствие экспериментальной манипуляции, обычно не способна продуцировать все типы клеток организма. Другими словами, соматические клетки представляют собой клетки, которые дифференцировались в достаточной степени, и которые не будут в норме генерировать клетки всех трех зародышевых листков тела, т.е. эктодермы, мезодермы и энтодермы. Например, соматические клетки будут содержать как нейроны, так и нейронные клетки-предшественники, последние из которых могут быть способны естественным образом давать все, или некоторые из, типов клеток центральной нервной системы, но не могут дать начало клеткам клеточных линий мезодермы или энтодермы.

Под "митотической клеткой" подразумевают клетку, которая претерпевает митоз. Митоз представляет собой способ, с помощью которого эукариотическая клетка разделяет хромосомы в своем ядре на два идентичных набора в двух отдельных ядрах. Это, как правило, затем сопровождается цитокинезом, который делит ядра, цитоплазму, органеллы и клеточную мембрану на две клетки, которые содержат примерно равные доли этих клеточных компонентов.

Под "постмитотической клеткой" подразумевают клетку, которая вышла из митоза, т.е. она является "покоящейся", что подразумевает то, что она уже не подвергается делению. Это состояние покоя может быть временным, т.е. обратимым, или он может быть постоянным.

Под "мейотической клеткой" подразумевают клетку, которая претерпевает мейоз. Мейоз представляет собой способ, с помощью которого клетка разделяет свой ядерный материал с целью получения гамет

или спор. В отличие от митоза, в мейозе хромосомы претерпевают этап рекомбинации, в котором генетический материал перемешивается между хромосомами. Кроме того, в результате мейоза получаются четыре (генетически уникальные) гаплоидные клетки, по сравнению с двумя (генетически идентичными) диплоидными клетками, полученными после митоза.

Под "рекомбинацией" подразумевают способ обмена генетической информацией между двумя полинуклеотидами. При использовании в данном документе, "репарация по механизму гомологичной рекомбинации (HDR)" означает специализированную форму репарации ДНК, которая имеет место, например, при репарировании двунитевых разрывов в клетках. Этот процесс требует гомологии нуклеотидной последовательности, в нем используются "донорные" молекулы в качестве шаблона для репарации молекулы-"мишени" (т.е. той, в которой содержится двухцепочечный разрыв) и приводит к передаче генетической информации от донора к мишени. Репарация по механизму гомологичной рекомбинации может привести к изменению последовательности молекулы-мишени (например, вставке, делеции, мутации), если ДНК-мишень отличается от донорного полинуклеотида, частью или всей последовательности донорного полинуклеотида, которую вводят в ДНК-мишень. В некоторых вариантах реализации настоящего изобретения донорный полинуклеотид, участок донорного полинуклеотида, копию донорного полинуклеотида или участок копии донорного полинуклеотида интегрируют в ДНК-мишень.

Под "негомологичным соединением концов (NHEJ)" подразумевают репарация двухцепочечных разрывов в ДНК с помощью прямого лигирования концов разрыва друг с другом без необходимости гомологичной матрицы (в отличие от репарации по механизму гомологичной рекомбинации, которая требует гомологичную последовательность для того, чтобы осуществить репарацию). NHEJ часто приводит к потере (делеции) нуклеотидной последовательности вблизи сайта двухцепочечного разрыва.

Термины "лечение", "лечить" и подобные используют в данном документе для того, чтобы обозначать, как правило, получение необходимого фармакологического и/или физиологического эффекта. Эффект может быть профилактическим, в терминах полного или частичного предотвращения заболевания или его симптома, и/или может быть терапевтическим в терминах частичного или полного исцеления в случае заболевания и/или неблагоприятного эффекта, относящегося к этому заболеванию. "Лечение", как используют в данном документе, охватывает любое лечение заболевания или симптома у млекопитающего, и содержит (а) профилактику заболевания или симптома, наблюдающегося у субъекта, который может быть предрасположен к приобретению заболевания или симптома, но до сих пор не установленно как такового; (b) ингибирование заболевания или симптома, т.е. прекращение его развития; или (c) облегчение заболевания, т.е. ослабление симптомов заболевания. Терапевтическое средство может быть введено до, во время или после начала заболевания или травмы. Лечение текущего заболевания, при условии, что лечение стабилизирует или снижает нежелательные клинические симптомы пациента, представляет особый интерес. Подобное лечение желательно осуществляют перед полной потерей функции в пораженных тканях. Целевую терапию желательно вводить в течение симптоматической стадии заболевания, а в некоторых случаях, после симптоматической стадии заболевания.

Термины "индивидуум", "субъект", "хозяин" и "пациент" являются в данном документе взаимозаменяемыми и относятся к любому млекопитающему, особенно к человеку, для которого желательна диагностика, лечение или терапия.

Общие способы молекулярной и клеточной биохимии возможно найти в таких стандартных учебниках, как *Molecular Cloning: A Laboratory Manual*, 3rd Ed. (Sambrook и соавт., HarBor Laboratory Press 2001); *Short Protocols in Molecular Biology*, 4th Ed. (Ausubel и соавт. eds., John Wiley & Sons 1999); *Protein Methods* (Bollag и соавт., John Wiley & Sons 1996); *Nonviral Vectors for Gene Therapy* (Wagner и соавт. eds., Academic Press 1999); *Viral Vectors* (Kaplift & Loewy eds., Academic Press 1995); *Immunology Methods Manual* (I. Lefkovits ed., Academic Press 1997); и *Cell and Tissue Culture: Laboratory Procedures in Biotechnology* (Doyle & Griffiths, John Wiley & Sons 1998), раскрытия которых содержатся в настоящем документе в виде ссылки.

Перед дальнейшим описанием настоящего изобретения, должно быть понятно, что это изобретение не ограничено конкретными описанными вариантами реализации настоящего изобретения, тогда как таковые могут, разумеется, варьироваться. Следует также иметь в виду, что терминология, используемая в настоящем документе, существует для целей описания конкретных вариантов осуществления и не предназначена для ограничения, поскольку объем настоящего изобретения будет ограничен только прилагаемой формулой изобретения.

Когда указан диапазон значений, следует понимать, что каждое промежуточное значение, до десятых долей единицы нижнего предела, если из контекста явно не следует иное, между значениями верхнего и нижнего пределов этого диапазона и любое другое указанное значение, или промежуточное значение в этом установленном диапазоне входит в объем настоящего изобретения. Верхние и нижние пределы этих меньших диапазонов могут независимо быть включены в меньшие диапазоны, которые также входят в объем настоящего изобретения, с учетом любого конкретно исключенного предела в установленном интервале. Там, где установленный диапазон содержит один или оба предела, диапазоны, исключющие либо один из них, либо оба из этих пределов, также входят в объем настоящего изобретения.

Некоторые диапазоны представляют в данном документе с численными значениями, которые пред-

варяется термином "около". Термин "около" используют в данном документе для того, чтобы предусмотреть буквальное указание точного количества, которому оно предшествует, а также числа, которое является близким к, или приблизительным количеством, которому предшествует данный термин. При определении, является ли число близким к, или приблизительным по отношению к конкретному приведенному числу, близким или приблизительным неприведенным числом может быть такое число, которое, в том контексте, в котором оно представлено, обеспечивает существенный эквивалент отдельного приведенного числа.

Если не определено иное, все технические и научные термины, используемые в данном документе, имеют то же значение, которое обычно понимает специалист в данной области техники, к которой принадлежит настоящее изобретение. Хотя любые способы и материалы, подобные или эквивалентные тем, которые описаны в данном документе, также могут быть использованы при осуществлении или при тестировании настоящего изобретения, предпочтительные способы и материалы приведены ниже. Все публикации, упомянутые в данном документе, содержатся в данном документе в виде ссылки для раскрытия и описания способов и/или материалов, для раскрытия которых и цитируют публикации.

Все публикации и патенты, цитируемые в данном описании, содержатся в данном документе в виде ссылки таким образом, как если бы каждая отдельная публикация или патент были конкретно и индивидуально указаны как те, которые содержатся в виде ссылки и содержатся в данном документе в виде ссылки для раскрытия и описания способов и/или материалов из процитированных публикаций. Цитирование любой публикации осуществляют в случае ее раскрытия перед датой подачи и это не должно быть истолковано как признание того, что настоящее изобретение не имеет права датировать такую публикацию задним числом в соответствии с более ранней датой создания изобретения. Кроме того, даты публикации могут отличаться от фактических дат публикации, которые, возможно, должны быть подтверждены независимыми источниками.

Следует отметить, что, как используется в данном документе и в прилагаемой формуле изобретения, формы единственного числа содержат ссылки на множественное число, если из контекста явно не следует иное. Таким образом, например, ссылка на "полинуклеотид" содержит множество таких полинуклеотидов, а ссылка на "полипептид" содержит ссылку на один или несколько полипептидов и их эквивалентов, известных специалистам в данной области техники, и так далее. Также отмечают, что формула изобретения может быть составлена для того, чтобы исключить любой необязательный элемент. Таким образом, это заявление призвано служить в качестве antecedentного базиса для использования такой исключительной терминологии как "исключительно", "только" и подобных, в связи с перечислением пунктов формулы изобретения или использования "негативного" признака.

Следует понимать, что определенные признаки изобретения, которые для ясности описаны в контексте отдельных вариантов реализации настоящего изобретения, также могут быть представлены в комбинации в одном варианте реализации настоящего изобретения. Наоборот, различные признаки изобретения, которые для краткости описаны в контексте одного варианта реализации настоящего изобретения, также могут быть представлены отдельно или в любой подходящей подкомбинации. Все комбинации вариантов реализации настоящего изобретения, относящиеся к настоящему изобретению, в частности, входят в объем настоящего изобретения и описаны в данном документе так же, как если бы каждый комбинация была индивидуально и явно раскрыта. Кроме того, все подкомбинации различных вариантов реализации настоящего изобретения и их элементы также специально входят в объем настоящего изобретения и описаны в данном документе так же, как если бы каждая подкомбинация была индивидуально и явно раскрыта в данном документе.

Публикации, обсуждаемые в данном документе, предоставляются исключительно с целью их раскрытия перед датой подачи настоящей заявки. Ничто в данном документе не должно быть истолковано как признание того, что настоящее изобретение не имеет права датировать такую публикацию задним числом по отношению к предшествующему изобретению. Кроме того, даты публикации могут отличаться от фактических дат публикации, которые, возможно, должны быть подтверждены независимыми источниками.

## **Подробное описание**

### **Часть I**

Настоящее изобретение обеспечивает РНК, нацеленную на ДНК, которая содержит специфическую последовательность и, вместе с модифицирующим полипептидом, обеспечивает в сайт-специфическую модификацию ДНК-мишени и/или полипептида, ассоциированного с ДНК-мишенью. Настоящее изобретение дополнительно обеспечивает сайт-специфические модифицирующие полипептиды. Настоящее изобретение дополнительно обеспечивает способы сайт-специфической модификации ДНК-мишени и/или полипептида, ассоциированного с ДНК-мишенью. Настоящее изобретение обеспечивает способы модулирования транскрипции нуклеиновой кислоты-мишени в клетке-мишени, как правило, включающим контактирование нуклеиновой кислоты-мишени с ферментативно неактивным полипептидом Cas9 и РНК, нацеленной на ДНК. Также обеспечиваются наборы и композиции для осуществления способов. Настоящее изобретение обеспечивает генетически модифицированные клетки, которые продуцируют Cas9; и трансгенные многоклеточные организмы Cas9.

Нуклеиновые кислоты.

РНК, нацеленная на ДНК.

Настоящее изобретение обеспечивает РНК, нацеленную на ДНК, которая направляет активность ассоциированного полипептида (например, сайт-специфического модифицирующего полипептида) на специфическую последовательность-мишень в пределах ДНК-мишени. РНК, нацеленная на ДНК, содержит первый сегмент (также называемый в данном документе как "нацеленный на ДНК сегмент" или "нацеленная на ДНК последовательность") и второй сегмент (также называемый в данном документе как "белок-связывающий сегмент" или "белок-связывающая последовательность").

Нацеленный на ДНК сегмент РНК, нацеленной на ДНК.

Нацеленный на ДНК сегмент РНК, нацеленной на ДНК, содержит нуклеотидную последовательность, которая является комплементарной последовательности в ДНК-мишени. Другими словами, Нацеленный на ДНК сегмент РНК-мишени, нацеленной на ДНК, взаимодействует с ДНК-мишенью последовательности-специфическим образом посредством гибридизации (т.е. спаривания оснований). Таким образом, нуклеотидная последовательность нацеленного на ДНК сегмента может различаться и определять локализацию в пределах: ДНК-мишени, в которой будет взаимодействовать РНК, нацеленная на ДНК, и ДНК-мишень. Нацеленный на ДНК сегмент РНК-мишени, нацеленной на ДНК, может быть модифицирован (например, с помощью генной инженерии) с целью гибридизации с любой необходимой последовательностью в пределах ДНК-мишени.

Нацеленный на ДНК сегмент может иметь длину от около 12 до около 100 нуклеотидов. Например, нацеленный на ДНК сегмент может иметь длину от около 12 нуклеотидов (нт) до около 80 нт, от около 12 до около 50 нт, от около 12 до около 40 нт, от около 12 до около 30 нт, от около 12 до около 25 нт, от около 12 до около 20 нт или от около 12 до около 19 нт. Например, нацеленный на ДНК сегмент может иметь длину от около 19 до около 20 нт, от около 19 до около 25 нт, от около 19 до около 30 нт, от около 19 до около 35 нт, от около 19 до около 40 нт, от около 19 до около 45 нт, от около 19 до около 50 нт, от около 19 до около 60 нт, от около 19 до около 70 нт, от около 19 до около 80 нт, от около 19 до около 90 нт, от около 19 до около 100 нт, от около 20 до около 25 нт, от около 20 до около 30 нт, от около 20 до около 35 нт, от около 20 до около 40 нт, от около 20 до около 45 нт, от около 20 до около 50 нт, от около 20 до около 60 нт, от около 20 до около 70 нт, от около 20 до около 80 нт, от около 20 до около 90 нт или от около 20 до около 100 нт. Нуклеотидная последовательность (нацеленная на ДНК последовательность) нацеленного на ДНК сегмента, который комплементарен нуклеотидной последовательности (последовательности-мишени) ДНК-мишени, может иметь длину по меньшей мере около 12 нт. Например, нацеленная на ДНК последовательность нацеленного на ДНК сегмента, который комплементарен последовательности-мишени ДНК-мишени, может иметь длину по меньшей мере около 12 нт, по меньшей мере около 15 нт, по меньшей мере около 18 нт, по меньшей мере около 19 нт, по меньшей мере около 20 нт, по меньшей мере около 25 нт, по меньшей мере около 30 нт, по меньшей мере около 35 нт или по меньшей мере около 40 нт. Например, нацеленная на ДНК последовательность нацеленного на ДНК сегмента, который комплементарен последовательности-мишени ДНК-мишени, может иметь длину от около 12 нуклеотидов (нт) до около 80 нт, от около 12 до около 50 нт, от около 12 до около 45 нт, от около 12 до около 40 нт, от около 12 до около 35 нт, от около 12 до около 30 нт, от около 12 до около 25 нт, от около 12 до нт около 20 нт, от около 12 до около 19 нт, от около 19 до около 20 нт, от около 19 до около 25 нт, от около 19 до около 30 нт, от около 19 до около 35 нт, от около 19 до около 40 нт, от около 19 до около 45 нт, от около 19 до около 50 нт, от около 19 до около 60 нт, от около 20 до около 25 нт, от около 20 до около 30 нт, от около 20 до около 35 нт, от около 20 до около 40 нт, от около 20 до около 45 нт, от около 20 до около 50 нт или от около 20 до около 60 нт. Нуклеотидная последовательность (нацеленная на ДНК последовательность) нацеленного на ДНК сегмента, который комплементарен нуклеотидной последовательности (последовательности-мишени) ДНК-мишени, может иметь длину по меньшей мере около 12 нт.

В некоторых случаях нацеленная на ДНК последовательность нацеленного на ДНК сегмента, который комплементарен последовательности-мишени ДНК-мишени, имеет 20 нуклеотидов в длину. В некоторых случаях нацеленная на ДНК последовательность нацеленного на ДНК сегмента, который комплементарен последовательности-мишени ДНК-мишени, имеет 19 нуклеотидов в длину.

Процент комплементарности между нацеленной на ДНК последовательностью нацеленного на ДНК сегмента и последовательностью-мишенью ДНК-мишени может быть по меньшей мере 60% (например, по меньшей мере 65%, по меньшей мере 70%, по меньшей мере 75%, по меньшей мере 80%, по меньшей мере 85%, по меньшей мере 90%, по меньшей мере 95%, по меньшей мере 97%, по меньшей мере 98%, по меньшей мере 99 или 100%). В некоторых случаях процент комплементарности между нацеленной на ДНК последовательностью нацеленного на ДНК сегмента и последовательностью-мишенью ДНК-мишени составляет 100% в пределах семи последовательных нуклеотидов, крайних на 5'-конце последовательности-мишени комплементарной цепи ДНК-мишени. В некоторых случаях процент комплементарности между нацеленной на ДНК последовательностью нацеленного на ДНК сегмента и последовательности-мишени ДНК-мишени составляет по меньшей мере 60% в пределах 20 последовательных нуклеотидов. В некоторых случаях процент комплементарности между нацеленной на ДНК последователь-

ностью нацеленного на ДНК сегмента и последовательностью-мишенью ДНК-мишени составляет 100% в пределах четырнадцати последовательных крайних на 5'-конце нуклеотидов последовательности-мишени комплементарной цепи ДНК-мишени, и всего лишь 0% - в остальной части. В таком случае, нацеленная на ДНК последовательность может иметь 14 нуклеотидов в длину (см. фиг. 12D-E). В некоторых случаях процент комплементарности между нацеленной на ДНК последовательностью нацеленного на ДНК сегмента и последовательностью-мишенью ДНК-мишени составляет 100% в пределах семи последовательных нуклеотидов, крайних на 5'-конце последовательности-мишени комплементарной цепи ДНК-мишени, и всего лишь 0% - в остальной. В таком случае, нацеленная на ДНК последовательность может иметь 7 нуклеотидов в длину.

Белок-связывающий сегмент РНК, нацеленной на ДНК.

Белок-связывающий сегмент РНК-мишени, нацеленной на ДНК, взаимодействует с сайт-специфическим модифицирующим полипептидом. РНК-мишень, нацеленная на ДНК, направляет связанный полипептид к специфической нуклеотидной последовательности в пределах ДНК-мишени посредством вышеуказанного нацеленного на ДНК сегмента. Белок-связывающий сегмент РНК, нацеленной на ДНК, содержит два участка нуклеотидов, которые комплементарны друг другу. Комплементарные нуклеотиды белок-связывающего сегмента гибридизуются с образованием дуплекса двухцепочечной РНК (дцРНК) (см. фиг. 1A и 1B).

РНК, нацеленная на ДНК, в виде двойной молекулы состоит из двух отдельных молекул РНК. Каждая из двух молекул РНК РНК-мишени, нацеленной на ДНК, в виде двойной молекулы содержит участок нуклеотидов, которые комплементарные друг другу, таким образом комплементарные нуклеотиды из двух молекул РНК гибридизуются с образованием дуплекса двухцепочечной РНК в белок-связывающем сегменте (фиг. 1A).

В некоторых вариантах реализации настоящего изобретения дуплекс-образующий сегмент РНК-активатора по меньшей мере на 60% идентичен одной из молекул РНК-активатора (tracРНК), приведенной в SEQ ID NO:431-562 или к их комплементу, на участке из по меньшей мере 8 последовательных нуклеотидов. Например, дуплекс-образующий сегмент РНК-активатора (или ДНК, кодирующей дуплекс-образующий сегмент РНК-активатора) по меньшей мере на около 60% идентичен, по меньшей мере на около 65% идентичен, по меньшей мере на около 70% идентичен, по меньшей мере на около 75% идентичен, по меньшей мере на около 80% идентичен, по меньшей мере на около 85% идентичен, по меньшей мере на около 90% идентичен, по меньшей мере на около 95% идентичен, по меньшей мере на около 98% идентичен, по меньшей мере на около 99% идентичен, или на 100% идентичен одной из последовательностей tracРНК, приведенных в SEQ ID NO:431-562 или к их комплементу, на участке из по меньшей мере 8 последовательных нуклеотидов.

В некоторых вариантах реализации настоящего изобретения дуплекс-образующий сегмент РНК-нацеливающей является на по меньшей мере около 60% идентичным одной из последовательностей РНК-нацеливающей (сгРНК), приведенной в SEQ ID NO:563-679 или к их комплементу, на участке из по меньшей мере 8 последовательных нуклеотидов. Например, дуплекс-образующий сегмент РНК-нацеливающей (или ДНК, кодирующей дуплекс-образующий сегмент РНК-нацеливающей) по меньшей мере на около 65% идентичен, по меньшей мере на около 70% идентичен, по меньшей мере на около 75% идентичен, по меньшей мере на около 80% идентичен, по меньшей мере на около 85% идентичен, по меньшей мере на около 90% идентичен, по меньшей мере на около 95% идентичен, по меньшей мере на около 98% идентичен, по меньшей мере на около 99% идентичен, или на 100% идентичен одной из последовательностей сгРНК, приведенных в SEQ ID NO:563-679 или к их комплементу, на участке из по меньшей мере 8 последовательных нуклеотидов.

РНК нацеленная на ДНК, в виде двойной молекулы, может быть получена для того, чтобы обеспечить контролируемое (т.е. зависящее от условий) связывание РНК-нацеливающей с РНК-активатором. Из-за того, что РНК, нацеленная на ДНК, в виде двойной молекулы, является нефункциональной, за исключением тех случаев, когда и РНК-активатор, и РНК-нацеливающая связаны в функциональный комплекс с dCas9, РНК, нацеленная на ДНК, в виде двойной молекулы, может быть индуцируемой (например, индуцируемой лекарственным средством) за счет того, что связывание между РНК-активатором и РНК-нацеливающей является индуцируемым. В качестве неограничивающего примера, аптамеры РНК могут быть использованы для регулирования (т.е. контроля) связывания РНК-активатора с РНК-нацеливающей. Соответственно, РНК-активатор и/или РНК-нацеливающая могут содержать последовательность аптамера РНК.

Аптамеры РНК известны в данной области техники и представляют собой, как правило, синтетический вариант рибосвитча. Термины "аптамер РНК" и "рибосвитч" используют в данном документе взаимозаменяемо, для того, чтобы включить как синтетические, так и природные последовательности нуклеиновых кислот, которые обеспечивают индуцируемую регуляцию структуры (и, следовательно, доступность специфических последовательностей) молекулы РНК, частью которых они являются. Аптамеры РНК обычно содержат последовательность, которая складывается в определенную структуру (например, шпилька), которая специфически связывает конкретное лекарственное средство (например, малую молекулу). Связывание лекарственного средства вызывает структурные изменения в укладке РНК, изменяя

свойства нуклеиновой кислоты, частью которой является аптамер. В качестве неограничивающих примеров: (i) РНК-активатор с аптамером может быть не способным связываться с соответствующей РНК-нацеливающей, если аптамер не связан с соответствующим лекарственным средством; (ii) РНК-нацеливающая с аптамером, может быть неспособна связываться с соответствующей РНК-активатором, если аптамер не связан с соответствующим лекарственным средством; и (iii) РНК-нацеливающая и РНК-активатор, каждая из которых содержит отличный аптамер, связывающий различные лекарственные средства, могут быть неспособны связываться друг с другом, если не присутствуют оба лекарственных средства. Как показано на этих примерах, РНК, нацеленная на ДНК, в виде двойной молекулы, может быть создана таким образом, чтобы быть индуцируемой.

Примеры аптамеров и рибосвитчей можно найти, например, в Nakamura и соавт., *Genes Cells*. 2012 May; 17(5):344-64; Vavalle и соавт., *Future Cardiol*. 2012 May; 8(3):371-82; Citartan и соавт., *Biosens Bioelectron*. 2012 Apr 15; 34(1):1-11; и Liberman и соавт., *Wiley Interdiscip Rev РНК*. 2012 May-Jun; 3(3):369-84; все из которых содержатся в данном документе в виде ссылки во всей их полноте.

Неограничивающие примеры нуклеотидных последовательностей, которые могут содержаться в РНК, нацеленной на ДНК, в виде двойной молекулы, содержат любую из последовательностей, приведенных в SEQ ID NO:431-562, или их комплементы, сопряженные с любыми последовательностями, представленными в SEQ ID NO:563-679, или их комплементами, которые могут гибридизоваться с образованием белок-связывающего сегмента.

РНК-мишень, нацеленная на ДНК, в виде одиночной молекулы, состоит из двух участков нуклеотидов (РНК-нацеливающей и РНК-активатора), которые являются комплементарными друг другу, ковалентно связаны с помощью промежуточных нуклеотидов ("линкеры" или "линкерные нуклеотиды") и гибридизуются с образованием дуплекса двухцепочечной РНК (дуплекса дцРНК) белок-связывающего сегмента, что приводит к структуре типа "стебель-петля" (фиг. 1В). РНК-нацеливающая и РНК-активатор могут быть ковалентно связаны посредством 3'-конца РНК-нацеливающей и 5'-конца РНК-активатора. Как вариант, РНК-нацеливающая и РНК-активатор могут быть ковалентно связаны посредством 3'-конца РНК-нацеливающей и 5'-конца РНК-активатора.

Линкер РНК, нацеленной на ДНК, в виде одиночной молекулы, может иметь длину от около 3 нуклеотидов до около 100 нуклеотидов. Например, линкер может иметь длину от около 3 нуклеотидов (нт) до около 90 нт, от около 3 до около 80 нт, от около 3 до около 70 нт, от около до около 60 нт, от около 3 до около 50 нт, от около 3 до около 40 нт, от около 3 до около 30 нт, от около 3 до около 20 нт или от около 3 до около 10 нт. Например, линкер может иметь длину от около 3 до около 5 нт, от около 5 до около 10 нт, от около 10 до около 15 нт, от около 15 до около 20 нт, от около 20 до около 25 нт, от около 25 до около 30 нт, от около 30 до около 35 нт, от около 35 до около 40 нт, от около 40 до около 50 нт, от около 50 до около 60 нт, от около 60 до около 70 нт, от около 70 до около 80 нт, от около 80 до около 90 нт, или от около 90 до около 100 нт. В некоторых вариантах реализации настоящего изобретения линкер РНК, нацеленной на ДНК, в виде одиночной молекулы имеет 4 нт.

Иллюстративная РНК, нацеленная на ДНК, в виде одиночной молекулы, содержит два комплементарных участка нуклеотидов, которые гибридизуются с образованием дуплекса дцРНК. В некоторых вариантах реализации настоящего изобретения один из двух комплементарных участков нуклеотидов РНК, нацеленной на ДНК, в виде одиночной молекулы (или ДНК, кодирующей участок) является на по меньшей мере около 60% идентичной одной из молекул РНК-активатора (tracrРНК), приведенной в SEQ ID NO:431-562, или ее комплементу, на участке по меньшей мере из 8 последовательных нуклеотидов. Например, один из двух комплементарных участков нуклеотидов РНК, нацеленной на ДНК, в виде одиночной молекулы, (или ДНК, кодирующей участок) является по меньшей мере на около 65% идентичным, по меньшей мере на около 70% идентичным, по меньшей мере на около 75% идентичным, по меньшей мере на около 80% идентичным, по меньшей мере на около 85% идентичным, по меньшей мере на около 90% идентичным, по меньшей мере на около 95% идентичным, по меньшей мере на около 98% идентичным, по меньшей мере на около 99% идентичным или на 100% идентичным одной из последовательностей tracrРНК, приведенных в SEQ ID NO:431-562 или ее комплементу на участке из по меньшей мере 8 последовательных нуклеотидов.

В некоторых вариантах реализации настоящего изобретения один из двух комплементарных участков нуклеотидов РНК, нацеленной на ДНК, в виде одиночной молекулы (или ДНК, кодирующей участок) является на по меньшей мере около 60% идентичным одной из последовательностей РНК-нацеливающей (crРНК), приведенной в SEQ ID NO:563-679, или ее комплементу, на участке по меньшей мере из 8 последовательных нуклеотидов. Например, один из двух комплементарных участков нуклеотидов РНК, нацеленной на ДНК, в виде одиночной молекулы (или ДНК, кодирующей участок) является по меньшей мере на около 65% идентичным, по меньшей мере на около 70% идентичным, по меньшей мере на около 75% идентичным, по меньшей мере на около 80% идентичным, по меньшей мере на около 85% идентичным, по меньшей мере на около 90% идентичным, по меньшей мере на около 95% идентичным, по меньшей мере на около 98% идентичным, по меньшей мере на около 99% идентичным или на 100% идентичным одной из последовательностей crРНК, приведенных в SEQ ID NO:563-679 или ее комплементу на участке из по меньшей мере 8 последовательных нуклеотидов.

С помощью стандартных методов могут быть определены соответствующие встречающиеся в природе родственные пары crРНК и tracrРНК для SEQ ID NO:431-679, принимая во внимание название видов и образование пары оснований (в случае дуплекса дцРНК белок-связывающего домена) при определении соответствующих родственных пар (см. фиг. 8 в качестве неограничивающего примера).

По отношению к и РНК-мишени, нацеленной на ДНК, в виде одиночной молекулы, и РНК-мишени, нацеленной на ДНК, в виде двойной молекулы, фиг. 57 иллюстрирует тот факт, что искусственные последовательности, которые демонстрируют очень малую (около 50%) идентичность с встречающимися в природе tracrРНК и crРНК, могут функционировать с Cas9 для расщепления ДНК-мишени до тех пор, пока сохраняют структуру белок-связывающего домена РНК, нацеленной на ДНК. Таким образом, складчатую структуру встречающегося в природе белок-связывающего домена РНК, нацеленной на ДНК, следует учитывать при разработке искусственных белок-связывающих доменов (либо в виде двойных молекул, либо в виде одиночной молекулы). В качестве неограничивающего примера, функциональная искусственная РНК, нацеленная на ДНК, на фиг. 57 разрабатывали на основе структуры белок-связывающего сегмента встречающейся в природе ДНК-нацеливающей (например, в том числе то же количество пар оснований в дуплексе РНК и в том числе та же область "петли" во встречающейся в природе РНК). Исходя из того, что структуры могут быть легко получены средним специалистом в данной области техники для любой встречающейся в природе пары crРНК: tracrРНК любых видов (см. SEQ ID NO:431-679 в случае последовательностей crРНК и tracrРНК широкого спектра видов), искусственная РНК, нацеленная на ДНК, может быть создана для того, чтобы имитировать естественную структуру в случае данного вида при использовании Cas9 (или родственной Cas9, см. фиг. 32А) этих видов (см. фиг. 24D и связанные с ней детали в примере 1). Таким образом, подходящая РНК, нацеленная на ДНК, может быть искусственно созданной РНК (не встречающейся в природе), которая содержит белок-связывающий домен, созданный для того, чтобы имитировать структуру белок-связывающего домена встречающейся в природе РНК, нацеленной на ДНК (см. SEQ ID NO:431-679, принимая во внимание название видов при определении соответствующих родственных пар).

Белок-связывающий сегмент может иметь длину от около 10 нуклеотидов до около 100 нуклеотидов. Например, белок-связывающий сегмент может иметь длину от около 15 нуклеотидов (нт) до около 80 нт, от около 15 до около 50 нт, от около 15 до около 40 нт, от около 15 до около 30 нт или от около 15 до около 25 нт.

Также, по отношению как и к РНК-мишени, нацеленной на ДНК, в виде одиночной молекулы, так и к РНК-мишени, нацеленной на ДНК, в виде двойной молекулы дуплекс дцРНК белок-связывающего сегмента может иметь длину от около 6 пар оснований (п.о.) до около 50 п.о. Например, дуплекс дцРНК белок-связывающего сегмента может иметь длину от около 6 до около 40 п.о., от около 6 до около 30 п.о., от около 6 до около 25 п.о., от около 6 до около 20 п.о., от около 6 до около 15 п.о., от около 8 до около 40 п.о., от около 8 до около 30 п.о., от около 8 до около 25 п.о., от около 8 до около 20 п.о. или от около 8 до около 15 п.о. Например, дуплекс дцРНК белок-связывающего сегмента может иметь длину от около 8 до около 10 п.о., от около 10 до около 15 п.о., от около 15 до около 18 п.о., от около 18 до около 20 п.о., от около 20 до около 25 п.о., от около 25 до около 30 п.о., от около 30 до около 35 п.о., от "около 35 до около 40 п.о. или от около 40 до около 50 п.о. В некоторых вариантах реализации настоящего изобретения дуплекс дцРНК белок-связывающего сегмента имеет длину в 36 пар оснований. Процент комплементарности между нуклеотидными последовательностями, которые гибридизуются с образованием дуплекса дцРНК белок-связывающего сегмента, может составлять по меньшей мере около 60%. Например, процент комплементарности между нуклеотидными последовательностями, которые гибридизуются с образованием дуплекса дцРНК белок-связывающего сегмента, может быть по меньшей мере около 65%, по меньшей мере около 70%, по меньшей мере около 75%, по меньшей мере около 80%, по меньшей мере около 85%, по меньшей мере около 90%, по меньшей мере около 95%, по меньшей мере около 98% или по меньшей мере около 99%. В некоторых случаях процент комплементарности между нуклеотидными последовательностями, которые гибридизуются с образованием дуплекса дцРНК белок-связывающего сегмента составляет 100%.

Сайт-специфический модифицирующий полипептид.

РНК-мишень, нацеленная на ДНК, и сайт-специфический модифицирующий полипептид-мишень образуют комплекс. РНК, нацеленная на ДНК, обеспечивает целевую специфичность комплекса с помощью того, что содержит нуклеотидную последовательность, которая является комплементарной к последовательности ДНК-мишени (как отмечалось выше). Сайт-специфический модифицирующий полипептид комплекса обеспечивает сайт-специфическую активность. Другими словами, сайт-специфический модифицирующий полипептид направляют на последовательность ДНК (например, хромосомную последовательность или внехромосомную последовательность, например, эписомную последовательность, последовательность мини-кольца, митохондриальную последовательность, последовательность хлоропласта и т.д.) благодаря его ассоциации с, по меньшей мере, белок-связывающим сегментом РНК, нацеленной на ДНК, (описанной выше).

Сайт-специфический модифицирующий полипептид- по изобретению модифицирует ДНК-мишень (например, расщепление или метилирование ДНК-мишени) и/или полипептид, ассоциированный с ДНК-

мишенью (например, метилирование или ацетилование хвоста гистона). Сайт-специфический модифицирующий полипептид также упоминается в данном документе как "сайт-специфический полипептид" или "РНК-связывающий сайт-специфический модифицирующий полипептид".

В некоторых случаях сайт-специфический модифицирующий полипептид представляет собой встречающийся в природе модифицирующий полипептид. В других случаях, сайт-специфический модифицирующий полипептид не является встречающимся в природе полипептидом (например, гибридный полипептид, как описано ниже, или встречающийся в природе полипептид, который модифицирован, например, мутацией, делецией, вставкой).

Иллюстративные встречающиеся в природе сайт-специфические модифицирующие полипептиды приведены в SEQ ID NO:1-255 в виде неограничивающего и не исчерпывающего перечня встречающихся в природе эндонуклеаз Cas9/Csn1. Эти встречающиеся в природе полипептиды, как описано в данном документе, связываются с РНК, нацеленной на ДНК, и таким образом, направляются на специфическую последовательность в пределах ДНК-мишени и расщепляют ДНК-мишень с получением двухцепочечно-го разрыва. Сайт-специфический модифицирующий полипептид-мишень содержит два участка, РНК-связывающий участок и участок активности. В некоторых вариантах реализации настоящего изобретения сайт-специфический модифицирующий полипептид-мишень содержит (i) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК, причем РНК, нацеленная на ДНК, содержит нуклеотидную последовательность, комплементарную последовательности в ДНК-мишени; и (ii) участок активности, который проявляет сайт-специфическую ферментативную активность (например, активность для метилирования ДНК, активности для расщепления ДНК, активности для ацетилирования гистонов, активности для метилирования пистонов и т.д.), причем сайт ферментативной активности определяется РНК, нацеленной на ДНК.

В других вариантах реализации настоящего изобретения сайт-специфический модифицирующий полипептид-мишень содержит (i) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК, причем РНК, нацеленная на ДНК, содержит нуклеотидную последовательность, комплементарную последовательности в ДНК-мишени; и (ii) участок активности, который модулирует транскрипцию в ДНК-мишени (например, чтобы увеличить или уменьшить транскрипцию), причем сайт модулированной транскрипции в пределах ДНК-мишени определяется РНК, нацеленной на ДНК.

В некоторых случаях сайт-специфический модифицирующий полипептид-мишень имеет ферментативную активность, которая изменяет ДНК-мишень (например, нуклеазную активность, активность метилтрансферазы, активность деметилазы, активность репарации ДНК, активность повреждения ДНК, активность дезаминирования, дисмутаазную активность, активность алкилирования, активность депуринизации, окислительную активность, активность формирования димера пиримидина, активность интегразы, активность транспозазы, активность рекомбиназы, активность полимеразы, активность лигазы, геликазную активность, фотолиазную активность или гликозилазную активность).

В других случаях, сайт-специфический модифицирующий полипептид-мишень имеет ферментативную активность, которая изменяет полипептид (например, гистон), ассоциированный с ДНК-мишенью (например, активность метилтрансферазы, деметилазную активность, ацетилтрансферазную активность, активность деацетилазы, киназную активность, фосфатазную активность, активность убиквитинлигазы, деубиквитирующую активность, активность аденилирования, активность деаденилирования, активность сумоилирования, активность десумоилирования, активность рибозилирования, активность рибозилирования, активность миристоилирования или активность демиристоилирования).

Иллюстративные сайт-специфические модифицирующие полипептиды.

В некоторых случаях сайт-специфический модифицирующий полипептид содержит аминокислотную последовательность, имеющую идентичность по меньшей мере около 75%, по меньшей мере около 80%, по меньшей мере около 85%, по меньшей мере около 90%, по меньшей мере около 95%, по меньшей мере около 99 или 100% аминокислотной последовательности по с аминокислотами 7-166 или 731-1003 аминокислотной последовательности Cas9/Csn1, изображенной на фиг. 3, или с соответствующими участками любой из аминокислотных последовательностей, приведенных в SEQ ID NO:1-256 и 795-1346.

Модификации нуклеиновой кислоты.

В некоторых вариантах реализации настоящего изобретения нуклеиновая кислота по изобретению (например, РНК, нацеленная на ДНК) содержит одну или несколько модификаций, например, модификацию основания, модификацию остова, и т.д., чтобы обеспечить нуклеиновую кислоту с новым или улучшенным свойством (например, улучшенной стабильностью). Как известно в данной области техники, нуклеозид представляет собой комбинацию основание-сахар. Участок нуклеозида, который соответствует основанию, как правило, представляет собой гетероциклическое основание. Два наиболее распространенных класса таких гетероциклических оснований представляют собой пурины и пиримидины. Нуклеотиды представляют собой нуклеозиды, которые дополнительно содержат фосфатную группу, ковалентно связанную с сахарной частью нуклеозида. В случае тех нуклеозидов, которые содержат пентофуранозильный сахар, фосфатная группа может быть связана с 2', 3' или 5' гидроксильной группой сахара. В образующихся олигонуклеотидах фосфатные группы ковалентно связывают соседние нуклеозиды друг с другом, чтобы сформировать линейное полимерное соединение. В свою очередь, соответствующие кон-

цы этого линейного полимерного соединения могут быть дополнительно соединены с образованием циклического соединения, однако, линейные соединения, как правило, являются подходящими. Кроме того, линейные соединения могут иметь внутреннюю комплементарность оснований нуклеотидов и, следовательно, могут укладываться таким образом, чтобы получать двухцепочечное соединение полностью или частично. В пределах олигонуклеотидов, фосфатные группы, как правило, упоминаются как образование межнуклеозидного остова олигонуклеотида. Нормальная связь или остов РНК и ДНК представляет собой 3'-5' фосфодиэфирную связь.

Модифицированные остовы и модифицированные межнуклеозидные связи.

Примеры подходящих нуклеиновых кислот, которые содержат модификации, содержат нуклеиновые кислоты, содержащие модифицированные остовы или неприродные межнуклеозидные связи. Нуклеиновые кислоты (которые имеют модифицированные остовы) содержат те, которые сохраняют атом фосфора в остове, и те, которые не имеют атом фосфора в остове.

Подходящие модифицированные остовы олигонуклеотидов, которые содержат в себе атом фосфора, содержат, например, фосфоротиоаты, хиральные фосфоротиоаты, фосфородитиоаты, фосфотриэфиры, аминоалкилфосфотриэфиры, метильные и другие алкильные фосфонаты, в том числе 3'-алкиленфосфонаты, 5'-алкиленфосфонаты и хиральные фосфонаты, фосфинаты, фосфорамидаты, в том числе 3'-аминофосфорамидат и аминоалкилфосфорамидаты, фосфородиамидаты, тионофосфорамидаты, тионоалкилфосфонаты, тионоалкилфосфотриэфиры, селенофосфаты и боранофосфаты, которые имеют нормальные 3'-5' связи, их 2'-5' связанные аналоги, и те, которые имеют обратную полярность, причем одна или несколько межнуклеотидных связей представляет собой 3'-3', 5'-5' или 2'-2' связь. Подходящие олигонуклеотиды, которые имеют обратную полярность, содержат одинарную 3'-3' связь на крайней 3' межнуклеотидной связи, т.е. отдельный инвертированный нуклеозидный остаток, который может быть основным (нуклеосахарин отсутствует или имеет на его месте гидроксильную группу). Также содержатся различные соли (такие как, например, калия или натрия), смешанные соли и формы свободных кислот.

В некоторых вариантах реализации настоящего изобретения нуклеиновая кислота по изобретению содержит один или несколько фосфоротиоатных и/или межнуклеозидных связей с участием гетероатома, в частности  $-\text{CH}_2\text{-NH-O-CH}_2-$ ,  $-\text{CH}_2\text{-N(CH}_3\text{)-O-CH}_2-$  (известный как метилен (метилимно) или остов ММ1),  $-\text{CH}_2\text{-O-N(CH}_3\text{)-CH}_2-$ ,  $-\text{CH}_2\text{-N(CH}_3\text{)-N(CH}_3\text{)-CH}_2-$  и  $-\text{O-N(CH}_3\text{)-CH}_2\text{-CH}_2-$  (причем нативную фосфодиэфирную межнуклеотидную связь представляют как  $-\text{O-P(=O)(OH)-O-CH}_2-$ ). Межнуклеозидные связи типа ММ1 раскрыты в упомянутом выше патенте США № 5489677. Подходящие амидные межнуклеозидные связи раскрыты в патенте США № 5602240.

Также подходящими являются нуклеиновые кислоты, которые имеют морфолиновые остовные структуры, как описано в, например, патенте США № 5034506. Например, в некоторых вариантах реализации настоящего изобретения нуклеиновая кислота по изобретению содержит 6-членное кольцо морфолиновое кольцо вместо рибозного кольца. В некоторых из этих вариантов реализации настоящего изобретения фосфородиамидатную или отличную нефосфодиэфирную межнуклеозидную связь заменяет фосфодиэфирная связь.

Подходящие модифицированные полинуклеотидные остовы, которые не содержат в себе атом фосфора, имеют остовы, которые образованы короткой алкильной цепью или циклоалкильными межнуклеозидными связями, смешанными межнуклеозидными связями с участием гетероатома и алкила или циклоалкила, или одной или несколькими короткоцепочечными гетероатомными или гетероциклическими межнуклеозидными связями. Они содержат те, которые имеют морфолиновые связи (образованные частично из сахарной части нуклеозида); силоксановые остовы; сульфидные, сульфоксидные и сульфонные остовы; формацетильные и тиоформацетильные остовы; метиленовые формацетильные и тиоформацетильные остовы; рибоацетильные остовы; алкенсодержащие остовы; сульфаматные остовы; метилениминовые и метиленигидразиновые остовы; сульфонатные и сульфонамидные остовы; амидные остовы; и другие, которые имеют смешанные части компонентов N, O, S и  $\text{CH}_2$ .

Миметики.

Нуклеиновая кислота по изобретению может быть миметиком нуклеиновой кислоты. Термин "миметик", как он применяется по отношению к полинуклеотидам, включает полинуклеотиды, где только фуранозное кольцо или и фуранозное кольцо, и межнуклеотидную связь заменяют без-фуранозными группами, замену только фуранозного кольца также упоминают в уровне техники в качестве сахарного суррогата. Фрагмент гетероциклического основания или модифицированный фрагмент гетероциклического основания поддерживают для гибридизации с соответствующей нуклеиновой кислотой-мишенью. Подобную нуклеиновую кислоту, миметик полинуклеотида, которая, как было показано, имеет прекрасные характеристики гибридизации, упоминают как пептидо-нуклеиновую кислоту (PNA). В PNA, сахаростов полинуклеотида заменяют остовом, содержащим амид, в частности, с аминоэтилглициновым остовом. Нуклеотиды сохраняют и связывают непосредственно или опосредованно с аза-атомами азота в амидной части остова.

Один миметик полинуклеотида, который, как сообщалось, обладает превосходными свойствами гибридизации, представляет собой пептидо-нуклеиновую кислоту (PNA). Остов в соединениях PNA пред-

ставляет собой две или более связанные аминоэтилглициновые единицы, которые дают PNA остов, содержащий амид. Фрагменты гетероциклических оснований связывают непосредственно или опосредованно с аза-атомами азота в амидной части остова. Репрезентативные патенты США, которые описывают получение соединений PNA, включают, но не ограничиваются ими: пат. США № 5539082; 5714331; и 5719262.

Другой класс миметика полинуклеотида, который был изучен, основан на связанных морфолиновых единицах (морфолиновые нуклеиновые кислоты), имеющих гетероциклические основания, прикрепленные к морфолиновому кольцу. Описан ряд связывающих групп, которые связывают морфолиновые мономерные единицы в морфолиновую нуклеиновую кислоту. Один класс связывающих групп был выбран для того, чтобы дать неионное олигомерное соединение. Неионные олигомерные соединения на основе морфолина, с меньшей вероятностью, имеют нежелательные взаимодействия с клеточными белками. Полинуклеотиды на основе морфолина представляют собой неионные миметики олигонуклеотидов, которые, с меньшей вероятностью, образуют нежелательные взаимодействия с клеточными белками (Dwayne A. Braasch и David R. Corey, *Biochemistry*, 2002, 41(14), 4503-4510). Полинуклеотиды на основе морфолина раскрыты в патенте США № 5034506. Было подготовлено множество соединений в пределах морфолинового класса полинуклеотидов, которые имеют множество различных связывающих групп, соединяющих мономерные субъединицы.

Еще один класс миметика полинуклеотида упоминают как циклогексениловые нуклеиновые кислоты (CeNA). Фуранозное кольцо, которое обычно присутствует в молекуле ДНК/РНК, заменяют циклогексенильным кольцом. Мономеры фосфорамидит, защищенные DMT CeNA, получают и используют для синтеза олигомерного соединения с последующей классической химией фосфорамидита. Были получены и изучены полностью модифицированные с помощью CeNA олигомерные соединения и олигонуклеотиды, которые имеют определенные положения, модифицированные с помощью CeNA (см. Wang и соавт., *J. Am. Chem. Soc.*, 2000, 122, 8595-8602). В целом, включение мономеров CeNA в цепочку ДНК повышает стабильность гибрида ДНК/РНК. Олигоаденилаты CeNA формируют комплексы с комплементарными РНК и ДНК с аналогичной стабильностью к нативным комплексам. Изучение включения структур CeNA в природные структуры нуклеиновых кислот проводилось с помощью ЯМР и кругового дихроизма, для того, чтобы обеспечить более легкую конформационную адаптацию.

Еще одна, модификация содержит запертые нуклеиновые кислоты (LNA), в которых 2'-гидроксильная группа связана с атомом углерода сахарного кольца 4', тем самым образуя 2'-С, 4'-С-оксиметиленовую связь, что приводит к образованию бициклического сахарного фрагмента. Связь может быть метиленовой группой (-CH<sub>2</sub>-), которая связывает атом кислорода 2' и атом углерода 4', где n составляет 1 или 2 (Singh и соавт., *Chem. Commun.*, 1998, 4, 455-456). LNA и аналоги LNA характеризуются очень высокой термической стабильностью дуплекса с комплементарной ДНК и РНК (T<sub>m</sub>=3-10°C), устойчивостью по отношению к 3'-эксонуклеолитической деградации и хорошими свойствами растворимости. Были описаны активные и нетоксичные антисмысловые олигонуклеотиды, которые содержат LNA (Wahlestedt и соавт., *Proc. Natl. Acad. Sci. U.S.A.*, 2000, 97, 5633-5638).

Были описаны синтез и получение мономеров LNA аденина, цитозина, гуанина, 5-метил-цитозина, тимина и урацила, а также их олигомеризация и свойства распознавания нуклеиновых кислот (Koshkin и соавт., *Tetrahedron*, 1998, 54, 3607-3630). LNA и их получение описаны также в WO 98/39352 и WO 99/14226.

Модифицированные фрагменты сахара.

Нуклеиновая кислота по изобретению также может содержать один или несколько замещенных фрагментов сахара. Подходящие полинуклеотиды содержат замещающую сахар группу, выбранную из OH; F; O-, S- или N-алкил; O-, S- или N-алкенил; O-, S- или N-алкинил; или O-алкил-O-алкил, где алкил, алкенил и алкинил могут быть замещенным или незамещенным алкилом C<sub>1</sub>-C<sub>10</sub> или алкенилом и алкинилом C<sub>2</sub>-C<sub>10</sub>. Частично подходящими являются O((CH<sub>2</sub>)<sub>n</sub>O)<sub>m</sub>CH<sub>3</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 составляют от 1 до около 10. Другие подходящие полинуклеотиды содержат сахар-замещающую группу, выбранную из низшего алкила C<sub>1</sub>-C<sub>10</sub>, 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, аминоалкиламина, полиалкиламина, замещенного силила, РНК-расщепляющей группы, "репортерной" группы, интеркалятора, группы для совершенствования фармакокинетических свойств олигонуклеотида или группы для улучшения фармакодинамических свойств олигонуклеотида, и других заместителей, имеющих аналогичные свойства. Подходящая модификация содержит 2'-метоксиэтокси (2'-O-CH<sub>2</sub>CH<sub>2</sub>OCH<sub>3</sub>, так же известный как 2'-O-(2-метоксиэтил) или 2'-МОЕ) (Martin и соавт., *Helv. Chim. Acta*, 1995, 78, 486-504), т.е. алкоксиалкоксильную группу. Дополнительная подходящая модификация содержит 2'-диметиламинооксиэтокси, т.е. группу O(CH<sub>2</sub>)<sub>2</sub>ON(CH<sub>3</sub>)<sub>2</sub>, также известную как 2'-DMAOE, как описано в примерах ниже, и 2'-диметиламинооксиэтоксиэтокси (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 сахарзамещающие группы содержат метокси (-O-CH<sub>3</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'-сахарзамещающие

группы могут быть в арабино-положении (вверх) или рибо-положении (вниз). Подходящая модификация 2'-арабино представляет собой 2'-F. Подобные модификации также могут быть сделаны в других положениях олигомерного соединения, в частности, в 3'-положении сахара 3'-концевого нуклеозида или в 2'-5'-связанных олигонуклеотидах и в 5'-положении 5'-концевого нуклеотида. Олигомерные соединены вместо пентофуранозильного сахара могут также иметь такие миметики сахара, как циклобутильные фрагменты.

Модификации и замещения оснований.

Нуклеиновая кислота по изобретению также может содержать модификации или замены нуклеоснования (часто называемое в данной области техники просто как "основание"). Как используют в данном документе, "немодифицированные" или "природные" нуклеоснования содержат пуриновые основания аденин (А) и гуанин (G) и пиримидиновые основания тимин (Т), цитозин (С) и урацил (U). Модифицированные нуклеоснования содержат различные синтетические и природные нуклеоснования, 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-аминоаденин, 8-азагуанин и 8-азааденин, 7-деазагуанин и 7-деазааденин, 3-деазагуанин и 3-деазааденин. Дополнительные модифицированные нуклеоснования содержат трициклические пиримидины, такие как феноксазинцитидин (1H-пиримидо(5,4-b)(1,4)бензоксазин-2(3H)-он), фенотиазинцитидин (1H-пиримидо(5,4-b)(1,4)бензотиазин-2(3H)-он), G-тиски, такие как замещенный феноксазинцитидин (например, 9-(2-аминоэтокси)-H-пиримидо(5,4-(b)(1,4)бензоксазин-2(3H)-он), карбазолцитидин (2H-пиримидо(4,5-b)индол-2-он), пиридоиндолцитидин (H-пиридо(3',2':4,5)пироло(2,3-d)пиримидин-2-он).

Группы гетероциклических оснований могут также содержать те, в которых пуриновое или пиримидиновое основание заменяют другими гетероциклами, например 7-деза-аденином, 7-дезагуанозином, 2-аминопиридином и 2-пиридоном. Дополнительные нуклеоснования содержат те, которые описаны в патенте США № 3687808, раскрытом в The Concise Encyclopedia Of Polymer Science And Engineering, pages 858-859, Kroschwitz, J. I., ed. John Wiley & Sons, 1990, раскрытом English и соавт., *Angewandte Chemie, International Edition*, 1991., 30, 613, и раскрытом Sanghvi, Y. S., Chapter 15, *Antisense Research and Applications*, pages 289-302, Crooke, S. T. и Lebleu, B., ed., CRC Press, 1993. Некоторые из этих нуклеоснований являются полезными для увеличения аффинности связывания олигомерного соединения. К ним относятся 5-замещенные пиримидины, 6-азапиримидины и N-2, N-6 и O-6 замещенные пурины, в том числе 2-аминопропиладенин, 5-пропинилурацил и 5-пропинилцитозин. Замещения 5-метилцитозина, как было показано, увеличивает стабильность дуплекса нуклеиновой кислоты на 0,6-1,2°C (Sanghvi и соавт., eds., *Antisense Research and Applications*, CRC Press, Boca Raton, 1993, p. 276-278) и представляют собой подходящие нуклеотидные замещения, например, в сочетании с 2'-О-метоксиэтильными модификациями сахаров.

Конъюгаты.

Другая возможная модификации нуклеиновой кислоты по изобретению представляет собой химическое связывание с полинуклеотидом одного или нескольких фрагментов или конъюгатов, которые усиливают активность, клеточное распределение или клеточное поглощение олигонуклеотида. Эти фрагменты или конъюгаты могут содержать группы, содержащие сопряженные двойные связи, ковалентно связанные с функциональными группами, такими как первичные или вторичные гидроксильные группы. Группы, содержащие сопряженные двойные связи, включают, но не ограничиваются ими, интеркаляторы, репортерные молекулы, полиамины, полиамиды, полиэтиленгликоли, простые полиэфиры, группы, которые повышают фармакодинамические свойства олигомеров, и группы, которые повышают фармакокинетические свойства олигомеров. Подходящие сопряженные группы включают, но не ограничиваются ими, холестерин, липиды, фосфолипиды, биотин, феназин, долат, фенантридин, антрахинон, акридин, флуоресцеины, родамины, кумарины и красители. Группы, которые повышают фармакодинамические свойства, содержат группы, которые улучшают поглощение, повышают устойчивость к деградации и/или усиливают последовательность-специфическую гибридизацию с нуклеиновой кислотой-мишенью. Группы, которые повышают фармакокинетические свойства, содержат группы, улучшающие поглощение, распределение, метаболизм или экскрецию нуклеиновой кислоты по изобретению.

Сопряженные группы включают, но не ограничиваются ими, липидные группы, такие как фрагмент холестерина (Letsinger и соавт., *Proc. Natl. Acad. Sci. USA*, 1989, 86, 6553-6556), холевая кислота (Manoharan и соавт., *Bioorg. Med. Chem. Lett.*, 1994, 4, 1053-1060), простой тиоэфир, например, гексил-S-третилтиол (Manoharan и соавт., *Ann. N.Y. Acad. Sci.*, 1992, 660, 306-309; Manoharan и соавт., *Bioorg. Med. Chem. Lett.*, 1993, 3, 2765-2770), тиохолестерин (Oberhauser и соавт., *Nucl. Acids Res.*, 1992, 20, 533-538), алифатическую цепь, например, остатки додекандиола или ундецила (Saison-Behmoaras и соавт., *EMBO J.*, 1991, 10, 1111-1118; Kabanov и соавт., *FEBS Lett.*, 1990, 259, 327-330; Svinarchuk и соавт., *Bio-*

chimie, 1993, 75, 49-54), фосфолипид, например, ди-гексадецил-рац-глицерин или 1,2-ди-О-гексадецил-рац-глицеро-3-Н-фосфонат триэтиламмония (Manoharan и соавт., *Tetrahedron Lett.*, 1995, 36, 3651-3654; Shea и соавт., *Nucl. Acids Res.*, 1990, 18, 3777-3783), аполиамин или цепь полиэтиленгликоля (Manoharan и соавт., *Nucleosides & Nucleotides*, 1995, 14, 969-973) или адамантан-уксусную кислоту (Manoharan и соавт., *Tetrahedron Lett.*, 1995, 36, 3651-3654), палмитильный фрагмент (Mishra и соавт., *Biochim. Biophys. Acta*, 1995, 1264, 229-237) или октадециламин или фрагмент гексиламино-карбонил-оксихолестерина (Crooke и соавт., *J. Pharmacol. Exp. Ther.*, 1996, 277, 923-937).

Конъюгат может содержать "домен белковой трансдукции" или РТД (также известный как CPP - проникающий в клетку пептид), который может относиться к полипептиду, полинуклеотиду, углеводу, органическому или неорганическому соединению, которое облегчает перемещение липидного бислоя, мицеллы, клеточной мембраны, мембраны органеллы или мембраны везикулы. РТД, прикрепленный к другой молекуле, которая может варьироваться от небольшой полярной молекулы до большой макромолекулы и/или наночастицы, облегчает перемещение молекулы сквозь мембрану, например, перемещающейся из внеклеточного пространства во внутриклеточное пространство, или цитозола в пределах органеллы. В некоторых вариантах реализации настоящего изобретения РТД ковалентно связан с аминоконцом экзогенного полипептида (например, сайт-специфического модифицирующего полипептида). В некоторых вариантах реализации настоящего изобретения РТД ковалентно связан с карбоксильным концом экзогенного полипептида (например, сайт-специфического модифицирующего полипептида). В некоторых вариантах реализации настоящего изобретения РТД ковалентно связан с нуклеиновой кислотой (например, РНК, нацеленной на ДНК, полинуклеотидом, кодирующим РНК, нацеленную на ДНК, полинуклеотидом, кодирующим сайт-специфический модифицирующий полипептид, и т.д.). Иллюстративные РТД включают, но не ограничиваются ими, домен трансдукции минимального ундекапептидного белка (соответствующий остаткам 47-57 TAT HIV-1, содержащий YGRKKRRQRRR; SEQ ID NO:264); последовательность полиаргинина, содержащую количество аргинина, достаточного для прямого входа в ячейку (например, 3, 4, 5, 6, 7, 8, 9, 10 или 10-50 аргининов); домен VP22 (Zender и соавт. (2002) *Cancer Gene Ther.* 9(6):489-96); домен трансдукции белка *Drosophila Antennapedia* (Noguchi и соавт. (2003) *Diabetes* 52(7): 1732-1737); усеченный белок кальцитонина человека (Trehin и соавт. (2004) *Pharm. Research* 21:1248-1256); полилизин (Wender и соавт. (2000) *Proc. Natl. Acad. Sci. USA* 97:13003-13008); RRQRRTSKLMKR (SEQ ID NO:265); транспортан GWTLNSAGYLLGKINLKALAALAKKIL (SEQ ID NO:266); KALAWEAKLAKALAKALAKHLAKALAKALCKEA (SEQ ID NO:267); и RQIKI-WFQNRRMKWKK (SEQ ID NO:268). Иллюстративные РТД включают, но не ограничиваются ими, YGRKKRRQRRR (SEQ ID NO:264), RKKRRQRRR (SEQ ID NO:269); гомополимер аргинина из от 3 остатков аргинина до 50 остатков аргинина; иллюстративные аминокислотные последовательности домена РТД включают, но не ограничиваются ими, любую из следующих: YGRKKRRQRRR (SEQ ID NO:264); RKKRRQRRR (SEQ ID NO:270); YARAAARQARA (SEQ ID NO:271); THRLPRRRRRR (SEQ ID NO:272); и GGRRARRRRRR (SEQ ID NO:273). В некоторых вариантах реализации настоящего изобретения РТД представляет собой активируемый CPP (ACPP) (Aguilera и соавт. (2009) *Integr Biol (Camb)* June; 1(5-6): 371-381). ACPP содержит поликаатионный CPP (например, Arg9 или "R9"), соединенный посредством расщепляемого линкера с соответствующим полианионом (например, Glu9 или "E9"), что уменьшает суммарный заряд практически до нуля и, таким образом, ингибирует адгезию и поглощение в клетки. После расщепления линкера высвобождается полианион, локально обнаруживая полиаргинин и его адгезионную способность, таким образом, "активируя" ACPP для перемещения сквозь мембрану.

Иллюстративные РНК, нацеленные на ДНК.

В некоторых вариантах реализации настоящего изобретения подходящая РНК, нацеленная на ДНК, состоит из двух отдельных молекул полинуклеотида РНК. Первая из двух отдельных молекул полинуклеотидов РНК (РНК-активатор) содержит нуклеотидную последовательность, имеющую идентичность по меньшей мере около 60%, по меньшей мере около 65%, по меньшей мере около 70%, по меньшей мере около 75%, по меньшей мере около 80%, по меньшей мере около 85%, по меньшей мере около 90%, по меньшей мере около 95%, по меньшей мере около 98%, по меньшей мере около 99 или 100% нуклеотидной последовательности на участке из по меньшей мере 8 последовательных нуклеотидов по отношению к любой из нуклеотидных последовательностей, представленных в SEQ ID NO:431-562, или их комплементам. Вторая из двух отдельных молекул полинуклеотидов РНК (РНК-нацеливающая) содержит нуклеотидную последовательность, имеющую идентичность по меньшей мере около 60%, по меньшей мере около 65%, по меньшей мере около 70%, по меньшей мере около 75%, по меньшей мере около 80%, по меньшей мере около 85%, по меньшей мере около 90%, по меньшей мере около 95%, по меньшей мере около 98%, по меньшей мере около 99 или 100% нуклеотидной последовательности на участке из по меньшей мере 8 последовательных нуклеотидов по отношению к любой из нуклеотидных последовательностей, представленных в SEQ ID NO:563-679, или их комплементам.

В некоторых вариантах реализации настоящего изобретения подходящая РНК, нацеленная на ДНК, представляет собой отдельный полинуклеотид РНК и содержит первую нуклеотидную последовательность, имеющую идентичность по меньшей мере около 60%, по меньшей мере около 65%, по меньшей мере около 70%, по меньшей мере около 75%, по меньшей мере около 80%, по меньшей мере около 85%,

по меньшей мере около 90%, по меньшей мере около 95%, по меньшей мере около 98%, по меньшей мере около 99 или 100% нуклеотидной последовательности на участке из по меньшей мере 8 последовательных нуклеотидов по отношению к любой из нуклеотидных последовательностей, представленных в SEQ ID NO:431-562, и вторую нуклеотидную последовательность, имеющую идентичность по меньшей мере около 60%, по меньшей мере около 65%, по меньшей мере около 70%, по меньшей мере около 75%, по меньшей мере около 80%, по меньшей мере около 85%, по меньшей мере около 90%, по меньшей мере около 95%, по меньшей мере около 98%, по меньшей мере около 99 или 100% нуклеотидной последовательности на участке из по меньшей мере 8 последовательных нуклеотидов по отношению к любой из нуклеотидных последовательностей, представленных в SEQ ID NO:463-679.

В некоторых вариантах реализации настоящего изобретения РНК, нацеленная на ДНК, представляет собой РНК, нацеленную на ДНК, в виде двойной молекулы, а РНК-нацеливающая содержит последовательность 5'GUUUUAGAGCUA-3' (SEQ ID NO:679), связанную на своем 5'-конце с участком нуклеотидов, которые являются комплементарными ДНК-мишени. В некоторых вариантах реализации настоящего изобретения РНК, нацеленная на ДНК, представляет собой РНК, нацеленную на ДНК, в виде двойной молекулы, а РНК-нацеливающая содержит последовательность 5' UAGCAAGWAAAAUAAGGCUAGUCCG-3' (SEQ ID NO://).

В некоторых вариантах реализации настоящего изобретения РНК, нацеленная на ДНК, представляет собой РНК, нацеленную на ДНК, в виде одиночной молекулы, и содержит " последовательность 5'-GUUUUAGAGCUA-линкер-UAGCAAGUAAAAUAAGGCUAGUCCG-3', связанную на своем 5'-конце с участком нуклеотидов, которые являются комплементарными ДНК-мишени (где "линкер" обозначает любую линкерную нуклеотидную последовательность, которая может содержать любую нуклеотидную последовательность) (SEQ ID NO://). Другие иллюстративные РНК в виде, нацеленные на ДНК, в виде одиночной молекулы, включают те, что представлены в SEQ ID NO:680-682.

Нуклеиновые кислоты, кодирующие РНК, нацеленную на ДНК, по изобретению и/или сайт-специфический модифицирующий полипептид по изобретению.

Настоящее раскрытие обеспечивает нуклеиновую кислоту, содержащую нуклеотидную последовательность, кодирующую РНК, нацеленную на ДНК, по изобретению и/или сайт-специфический модифицирующий полипептид по изобретению. В некоторых вариантах реализации настоящего изобретения нуклеиновая кислота, кодирующая РНК, нацеленную на ДНК, по изобретению представляет собой вектор экспрессии, например, рекомбинантный вектор экспрессии.

В некоторых вариантах реализации настоящего изобретения способ включает приведение ДНК-мишени в контакт или введение в клетку (или популяцию клеток) одной или нескольких нуклеиновых кислот, содержащих нуклеотидные последовательности, кодирующие РНК, нацеленную на ДНК, и/или сайт-специфический модифицирующий полипептид. В некоторых вариантах реализации настоящего изобретения клетка, содержащая ДНК-мишень, находится *in vitro*. В некоторых вариантах реализации настоящего изобретения клетка, содержащая ДНК-мишень, находится *in vivo*. Подходящие нуклеиновые кислоты, содержащие нуклеотидные последовательности, кодирующие РНК, нацеленную на ДНК, и/или сайт-специфический модифицирующий полипептид, содержат векторы экспрессии, притом, что вектор экспрессии, содержащий нуклеотидную последовательность, кодирующую РНК, нацеленную на ДНК, и/или сайт-специфический модифицирующий полипептид, представляет собой "рекомбинантный вектор экспрессии".

В некоторых вариантах реализации настоящего изобретения рекомбинантный вектор экспрессии представляет собой вирусную конструкцию, например, рекомбинантную адено-ассоциированную вирусную конструкцию (см., например, патент США № 7078387), рекомбинантную аденовирусную конструкцию, рекомбинантную лентивирусную конструкцию, рекомбинантную ретровирусную конструкцию и т.д.

Подходящие векторы экспрессии включают, но не ограничиваются ими, вирусные векторы (например, вирусные векторы, основанные на вирусе осповакцины; полиовирусе; аденовирусе (см., например, Li и соавт., *Invest Ophthalmol Vis Sci* 35:2543 2549, 1994; Borrás и соавт., *Gene Ther* 6:515 524, 1999; Li и Davidson, *PNAS* 92:7700 7704, 1995; Sakamoto и соавт., *N 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); адено-ассоциированном вирусе (см., например, Ali и соавт., *Hum Gene Ther* 9:81 86, 1998, Flannery и соавт., *PNAS* 94:6916 6921, 1997; Bennett и соавт., *Invest Ophthalmol Vis Sci* 38:2857 2863, 1997; Jomary и соавт., *Gene Ther* 4:683 690, 1997, Rolling и соавт., *Hum Gene Ther* 10:641 648, 1999; Ali и соавт., *Hum Mol Genet* 5:591 594, 1996; Srivastava в WO 93/09239, Samulski и соавт., *J. Vir.* (1989) 63:3822-3828; Mendelson и соавт., *Virol.* (1988) 166:154-165; и Flotte и соавт., *PNAS* (1993) 90:10613-10617); SV40; вирусе простого герпеса; вирусе иммунодефицита человека (см., например, Miyoshi и соавт., *PNAS* 94:10319 23, 1997; Takahashi и соавт., *J Virol* 73:7812 7816, 1999); ретровирусном векторе (например, вирусе лейкоза мыши, вирусе некроза селезенки и векторах, полученных из ретровирусов, таких как вирус саркомы Рауса, вирус саромы Харви, вирус лейкоза птицы, лентивирус, вирус иммунодефицита человека, миелопролиферативный вирус саркомы и вирус опухоли молочной железы); и подобные.

Многочисленные подходящие векторы экспрессии известны специалистам в данной области техни-

ки, а многие из них коммерчески доступны. Следующие векторы приведены в качестве примера; в случае эукариотических клеток-хозяев: pXT1, pSG5<sub>==</sub> (Stratagene), pSVK3, pBPV, pMSG и pSVLSV40 (Pharmacia). Тем не менее, любой другой вектор может быть использован при условии, что он совместим с клеткой-хозяином.

В зависимости от используемой системы хозяин/вектор, любой из ряда подходящих регуляторных элементов транскрипции и трансляции, в том числе конститутивные и индуцируемые промоторы, энхансеры транскрипции, терминаторы транскрипции и т.д., может быть использован в векторе экспрессии (см., например, Bitter и соавт. (1987) *Methods in Enzymology*, 153:516-544).

В некоторых вариантах реализации настоящего изобретения нуклеотидная последовательность, кодирующая РНК, нацеленную на ДНК, и/или сайт-специфический модифицирующий полипептид, функционально связаны с регуляторным элементом, например, регуляторным элементом транскрипции, таким как промотор. Регуляторный элемент транскрипции может быть функциональным либо в эукариотической клетке, например, клетке млекопитающего; либо в прокариотической клетке (например, бактериальной клетке или клетке архей). В некоторых вариантах реализации настоящего изобретения нуклеотидная последовательность, кодирующая РНК, нацеленную на ДНК, и/или сайт-специфический модифицирующий полипептид, функционально связаны с множеством регуляторных элементов, которые позволяют осуществить экспрессию нуклеотидной последовательности, кодирующей РНК, нацеленную на ДНК, и/или сайт-специфический модифицирующий полипептид в прокариотических и эукариотических клетках.

Неограничивающие примеры подходящих эукариотических промоторов (промоторов, функциональных в эукариотической клетке) содержат таковые из немедленно-раннего цитомегаловируса (CMV), тимидинкиназы вируса простого герпеса (HSV), раннего и позднего SV40, длинных концевых повторов (LTR) из ретровируса и металлотионеина-I мыши. Выбор соответствующего вектора и промотора находится в пределах компетенции специалиста в данной области техники. Вектор экспрессии также может содержать сайт связывания рибосомы для инициации трансляции и терминатор транскрипции. С целью усиления экспрессии, вектор экспрессии также может содержать соответствующие последовательности. Вектор экспрессии также может содержать нуклеотидные последовательности, кодирующие белковые метки (например, метку 6xHis, гемагглютининовую метку, зеленый флуоресцентный белок и т.д.), которые присоединяют к сайт-специфическому модифицирующему полипептиду, что приводит к гибриднему полипептиду.

В некоторых вариантах реализации настоящего изобретения нуклеотидная последовательность, кодирующая РНК, нацеленную на ДНК, и/или сайт-специфический модифицирующий полипептид, функционально связана с индуцируемым промотором. В некоторых вариантах реализации настоящего изобретения нуклеотидная последовательность, кодирующая РНК, нацеленная на ДНК, и/или сайт-специфический модифицирующий полипептид, функционально связаны с конститутивным промотором.

Способы введения нуклеиновой кислоты в клетку-хозяина известны в данной области техники, и любой известный способ может быть использован для введения нуклеиновой кислоты (например, экспрессирующей конструкции) в клетку. Подходящие способы включают в себя, например, вирусную или бактериофаговую инфекцию, трансфекцию, конъюгацию, слияние протопластов, липофекцию, электропорацию, осаждение фосфатом кальция, опосредованную полиэтиленгликолем (PEG) трансфекцию, опосредованную DEAE-декстраном трансфекцию, опосредованную липосомами трансфекцию, технологию генной пушки, осаждение фосфатом кальция, прямую микроинъекцию, опосредованную наночастицами доставку нуклеиновой кислоты (см., например, Panyam и соавт., *Adv Drug Deliv Rev.* 2012 Sep 13. pii: S0169-409X(12)00283-9. doi: 10.1016/j.addr.2012.09.023) и подобные.

Гибридные полипептиды.

Настоящее изобретение обеспечивает гибридный сайт-специфический модифицирующий полипептид. Гибридный сайт-специфический модифицирующий полипептид по изобретению взаимодействует с (например, связывается с) РНК-мишенью, нацеленной на ДНК (описанной выше). РНК, нацеленная на ДНК, направляет гибридный сайт-специфический модифицирующий полипептид к последовательности-мишени в пределах ДНК-мишени (например, хромосомной последовательности или экстрахромосомной последовательности, например, эписомной последовательности, последовательности мини-кольца, митохондриальной последовательности, последовательности хлоропластов, и т.д.). Гибридный сайт-специфический модифицирующий полипептид-мишень модифицирует ДНК-мишень (например, расщепляет или метилирует ДНК-мишень) и/или полипептид, ассоциированный с ДНК-мишенью (например, метилирует или ацетирует хвост гистона).

Гибридный сайт-специфический модифицирующий полипептид по изобретению модифицирует ДНК-мишень (например, расщепляет или метилирует ДНК-мишень) и/или полипептид, ассоциированный с ДНК-мишенью (например, метилирует или ацетирует хвост гистона). Гибридный сайт-специфический модифицирующий полипептид также упоминается в данном документе как "гибридный сайт-специфический полипептид" или "гибридный РНК-связывающий сайт-специфический модифицирующий полипептид".

Гибридный сайт-специфический модифицирующий полипептид по изобретению содержит два уча-

стка, РНК-связывающий участок и участок активности. Гибридный сайт-специфический модифицирующий полипептид-мишень содержит аминокислотные последовательности, которые получают по меньшей мере из двух различных полипептидов. Гибридный сайт-специфический модифицирующий полипептид-мишень может содержать измененные и/или встречающиеся в природе полипептидные последовательности (например, первую аминокислотную последовательность из модифицированного или немодифицированного белка Cas9/Csn1; и вторую аминокислотную последовательность, отличную от белка Csn1/Cas9).

РНК-связывающий участок.

В некоторых случаях РНК-связывающий участок гибридного сайт-специфического модифицирующего полипептида-мишени представляет собой встречающийся в природе полипептид. В других случаях РНК-связывающий участок гибридного сайт-специфического модифицирующего полипептида-мишени не является встречающейся в природе молекулой (т.е. он модифицирован, например, мутацией, делецией, вставкой). Встречающиеся в природе представляющие интерес РНК-связывающие участки получают из сайт-специфических модифицирующих полипептидов, известных в данной области техники. Например, SEQ ID NO:1-256 и 795-1346 представляют собой неограничивающий и исчерпывающий перечень встречающихся в природе эндонуклеаз Cas9/Csn1, которые могут быть использованы в качестве сайт-специфических модифицирующих полипептидов. В некоторых случаях РНК-связывающий участок гибридного сайт-специфического модифицирующего полипептида-мишени содержит аминокислотную последовательность, обладающую идентичность по меньшей мере около 75%, по меньшей мере около 80%, по меньшей мере около 85%, по меньшей мере около 90%, по меньшей мере около 95%, по меньшей мере около 98%, по меньшей мере около 99 или 100% аминокислотной последовательности по отношению к РНК-связывающему участку полипептида, который имеет любую из аминокислотных последовательностей, приведенных в SEQ ID NO:1-256 и 795-1346).

В некоторых случаях сайт-специфический модифицирующий полипептид содержит аминокислотную последовательность, имеющую идентичность по меньшей мере около 75%, по меньшей мере около 80%, по меньшей мере около 85%, по меньшей мере около 90%, по меньшей мере около 95%, по меньшей мере около 99 или 100% аминокислотной последовательности с аминокислотами 7-166 или 731-1003 аминокислотной последовательности Cas9/Csn1, изображенной на фиг. 3, или с соответствующими участками любой из аминокислотных последовательностей, приведенных в SEQ ID NO:1-256 и 795-1346.

Участок активности.

В дополнение к РНК-связывающей части, гибридный сайт-специфический модифицирующий полипептид содержит "участок активности". В некоторых вариантах реализации настоящего изобретения участок активности гибридного сайт-специфического модифицирующего полипептида-мишени содержит встречающийся в природе участок активности сайт-специфического модифицирующего полипептида (например, эндонуклеазы Cas9/Csn1). В других вариантах реализации настоящего изобретения участок активности гибридного сайт-специфического модифицирующего полипептида-мишени содержит модифицированную (например, заменой, делецией, вставкой) аминокислотную последовательность (встречающегося в природе участка активности сайт-специфического модифицирующего полипептида). Встречающиеся в природе представляющие интерес участки активности получают из сайт-специфических модифицирующих полипептидов, известных в данной области техники. Например, SEQ ID NO:1-256 и 795-1346 обеспечивают неограничивающий и исчерпывающий перечень встречающихся в природе эндонуклеаз Cas9/Csn1, которые могут быть использованы в качестве сайт-специфических модифицирующих полипептидов. Участок активности гибридного сайт-специфического модифицирующего полипептида-мишени является изменяемым и может содержать любую гетерологичную полипептидную последовательность, которая может быть полезна в раскрытых в данном документе способах.

В некоторых вариантах реализации настоящего изобретения гибридный сайт-специфический модифицирующий полипептид-мишень содержит (i) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК, причем РНК, нацеленная на ДНК, содержит нуклеотидную последовательность, комплементарную последовательности ДНК-мишени; и (ii) участок активности, который проявляет сайт-специфическую ферментативную активность (например, активность метилирования ДНК, активность расщепления ДНК, активность ацетилирования гистонов, активность метилирования гистонов и т.д.), причем сайт ферментативной активности определяется РНК, нацеленной на ДНК.

В других вариантах реализации настоящего изобретения гибридный сайт-специфический модифицирующий полипептид-мишень содержит (i) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК, причем РНК, нацеленная на ДНК, содержит нуклеотидную последовательность, комплементарную последовательности ДНК-мишени; и (ii) участок активности, который модулирует транскрипцию в ДНК-мишени (например, чтобы увеличить или уменьшить транскрипцию), причем сайт модулированной транскрипции в пределах ДНК-мишени определяется РНК, нацеленной на ДНК.

В некоторых случаях участок активности гибридного сайт-специфического модифицирующего полипептида-мишени имеет ферментативную активность, которая изменяет ДНК-мишень (например, нуклеазную активность, активность метилтрансферазы, активность деметилазы, активность репарации ДНК, активность повреждения ДНК, активность дезаминирования, дисмутазную активность, активность алкилиро-

вание, активность депурирования, окислительную активность, активность формирования димера пиримидина, активность интегразы, активность транспозазы, активность рекомбиназы, активность полимеразы, активность лигазы, геликазную активность, фотолиазную активность или гликозилазную активность).

В отличных случаях, участок активности гибридного сайт-специфического модифицирующего полипептида-мишени имеет ферментативную активность (например, активность метилтрансферазы, активность деметилазы, ацетилтрансферазную активность, активность деацетилазы, киназную активность, фосфатазную активность, активность убиквитинлигазы, деубиквитирующую активность, активность аденилирования, активность деаденилирования, активность сумоилирования, активность десумоилирования, активность рибозилирования, активность дерибозилирования, активностью миристоилирования или активность демиристоилирования), которая изменяет полипептид, ассоциированный с ДНК-мишенью (например, гистон).

В некоторых случаях участок активности гибридного сайт-специфического модифицирующего полипептида-мишени проявляет ферментативную активность (как описано выше). В некоторых случаях участок активности гибридного сайт-специфического модифицирующего полипептида-мишени модулирует транскрипцию ДНК-мишени (как описано выше). Участок активности гибридного сайт-специфического модифицирующего полипептида-мишени является изменяемым и может содержать любую гетерологичную полипептидную последовательность, которая может быть полезна в раскрытых в данном документе способах.

Иллюстративные гибридные сайт-специфические модифицирующие полипептиды.

В некоторых вариантах реализации настоящего изобретения участок активности гибридного сайт-специфического модифицирующего полипептида содержит измененную форму белка Cas9/Csn1. В некоторых случаях измененная форма белка Cas9/Csn1 содержит замену аминокислоты (например, делецию, вставку или замещение), что снижает встречающуюся в природе нуклеазную активность белка Cas9/Csn1. Например, в некоторых случаях, измененная форма белка Cas9/Csn1 имеет менее 50%, менее 40%, менее 30%, менее 20%, менее 10%, менее 5% или менее 1% нуклеазной активности соответствующего поли пептида Cas9/Csn1 дикого типа. В некоторых случаях измененная форма полипептида Cas9/Csn1 не имеет никакой существенной нуклеазной активности.

В некоторых вариантах реализации настоящего изобретения измененная форма полипептида Cas9/Csn1. содержит мутацию D10A (аспартат на аланин в положении аминокислоты 10 SEQ ID NO:8) (или соответствующую мутацию любого из белков, представленных в SEQ ID NO:1-256 и 795-1346), которая может расщеплять комплементарную цепь, ДНК-мишени, но имеет пониженную способность расщеплять некомплемтарную цепь ДНК-мишени (см. фиг. 11). В некоторых вариантах реализации настоящего изобретения измененная форма полипептида Cas9/Csn1 содержит мутацию H840A (гистидин на аланин в положении аминокислоты 10 SEQ ID NO:8) (или соответствующую мутацию любого из белков, приведенных в SEQ ID NO:1-256 и 795-1346), которая может расщеплять некомплемтарную цепь ДНК-мишени, но имеет пониженную способность расщеплять комплементарную цепь ДНК-мишени (см. фиг. 11). В некоторых вариантах реализации настоящего изобретения измененная форма полипептида Cas9/Csn1 содержит и мутацию D10A, и мутацию H840A (или соответствующие мутации любого из белков, приведенных в SEQ ID NO:1-256 и 795-1346), причем полипептид имеет пониженную способность расщеплять и комплементарные, и некомплемтарные цепи ДНК-мишени. С целью достижения вышеуказанных эффектов, другие остатки могут быть мутировавшими (т.е. инактивированы одним или другим участком нуклеазы). В качестве неограничивающих примеров, остатки D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986 и/или A987 (или соответствующие мутации любого из белков, приведенные в SEQ ID NO:1-256 и 795-1346) могут быть изменены (т.е. заменены) (см. фиг. 3, фиг. 5, фиг. 11A и табл. 1 с получением дополнительной информации о сохранении аминокислотных остатков Cas9). Кроме того, подходящими являются мутации, отличные от замещения аланина. Для получения дополнительной важной информации.

В табл. 1 приведены 4 мотива, которые присутствуют в последовательностях Cas9 различных видов (см. также фиг. 3 и 5). Аминокислоты, перечисленные в данном документе, представляют собой Cas9 из *S. pyogenes* (SEQ ID NO:8).

Таблица 1

№ мотива	Мотив	Аминокислоты (№№ остатков)	Высококонсервативный
1	RuvC-подобный I	IGLDIGTNSVGWAVI (7-21) (SEQ ID NO:260)	D10, G12, G17
2	RuvC-подобный II	IVIEMARE (759-766) (SEQ ID NO:261)	E762
3	HNH-мотив	DVDHIVPQSFLKDDSIDNKVLT RSDKN (837-863) (SEQ ID NO:262)	H840, N854, N863
4	RuvC-подобный II	HNAHDAYL (982-989) (SEQ ID NO:263)	H982, H983, A984, D986, A987

В некоторых случаях гибридный сайт-специфический модифицирующий полипептид содержит аминокислотную последовательность, имеющую идентичность по меньшей мере около 75%, по меньшей мере около 80%, по меньшей мере около 85%, по меньшей мере около 90%, по меньшей мере около 95%,

по меньшей мере около 99 или 100% аминокислотной последовательности с аминокислотами 7-166 или 731-1003 аминокислотной последовательности Cas9/Csn1, изображенной на фиг. 3, или с соответствующими участками любой из аминокислотных последовательностей, приведенных в SEQ ID NO:1-256 и 795-1346. В некоторых случаях гибридный сайт-специфический модифицирующий полипептид содержит 4 мотива (как приведено в табл. 4 и проиллюстрировано на фиг. 3А и 5), аминокислотная последовательность каждого из которых имеет идентичность по меньшей мере около 75%, по меньшей мере около 80%, по меньшей мере около 85%, по меньшей мере около 90%, по меньшей мере около 95%, по меньшей мере около 99 или 100% аминокислотной последовательности по отношению к каждому из 4 мотивов, приведенных в табл. 1 (SEQ ID NO:260-263), или к соответствующим участкам любой из аминокислотных последовательностей, приведенных в SEQ ID NO:1-256 и 795-1346. В некоторых случаях гибридный сайт-специфический модифицирующий полипептид содержит аминокислотную последовательность, имеющую идентичность по меньшей мере около 75%, по меньшей мере около 80%, по меньшей мере около 85%, по меньшей мере около 90%, по меньшей мере около 95%, по меньшей мере около 99 или 100% аминокислотной последовательности с аминокислотами 7-166 или 731-1003 аминокислотной последовательности Cas9/Csn1, изображенной на фиг. 3, или к соответствующим участкам любой из аминокислотных последовательностей, приведенных в SEQ ID NO:1-256 и 795-1346.

В некоторых вариантах реализации настоящего изобретения участок активности сайт-специфический модифицирующий полипептид содержит гетерологичный полипептид, который имеет ДНК-модифицирующую активность и/или активность фактора транскрипции и/или ДНК-ассоциированную полипептид-модифицирующую активность. В некоторых случаях гетерологичный полипептид замещает участок полипептида Cas9/Csn1, который обеспечивает нуклеазную активность. В других вариантах реализации настоящего изобретения сайт-специфический модифицирующий полипептид-мишень содержит как участок полипептида Cas9/Csn1, который обычно обеспечивает нуклеазную активность (и этот участок может быть полностью активным или может вместо этого быть изменен таким образом, чтобы иметь менее 100% от соответствующей, активности дикого типа), так и гетерологичный полипептид. Другими словами, в некоторых случаях, гибридный сайт-специфический модифицирующий полипептид-мишень представляет собой слитый полипептид, содержащий как участок полипептида Cas9/Csn1, который обычно обеспечивает нуклеазную активность, так и гетерологичный полипептид. В других случаях, гибридный сайт-специфический модифицирующий полипептид-мишень представляет собой слитый полипептид, содержащий модифицированный (например, заменой, делецией, вставкой аминокислоты) вариант участка активности полипептида Cas9/Csn1 и гетерологичный полипептид. В других случаях, гибридный сайт-специфический модифицирующий полипептид-мишень представляет собой слитый полипептид, содержащий гетерологичный полипептид и РНК-связывающий участок встречающегося в природе или модифицированного сайт-специфического модифицирующего полипептида.

Например, в гибридном белке Cas9/Csn1, встречающийся в природе (или модифицированным, например, мутацией, делецией, вставкой) бактериальный полипептид Cas9/Csn1 может быть слит с гетерологичной полипептидной последовательностью (например, полипептидной последовательностью из белка, отличного от Cas9/Csn1 или полипептидной последовательности другого организма). Гетерологичная полипептидная последовательность может проявлять активность (например, ферментативную активность), что также будет продемонстрировано с помощью гибридного белка Cas9/Csn1 (например, активность метилтрансферазы, активность ацетилтрансферазы, активность киназы, активность убиквитинирования и т.д.). Гетерологичная полипептидная последовательность может быть связана с отличной последовательностью нуклеиновой кислоты (например, с помощью генной инженерии) для того, чтобы создать гибридную нуклеотидную последовательность, кодирующую гибридный полипептид. В некоторых вариантах реализации настоящего изобретения гибридный полипептид Cas9/Csn1 получают с помощью слияния полипептида Cas9/Csn1 (например, Cas9 дикого типа или вариант Cas9, например, Cas9 с пониженной или инактивированной нуклеазной активностью) с гетерологичной последовательностью, которая обеспечивает субклеточные локализации (например, клеточный сигнал внутриядерной локализации (NLS) в случае нацеливания на ядро; клеточный сигнал внутриядерной локализации в случае нацеливания на митохондрии; клеточный сигнал внутриядерной локализации в случае нацеливания на хлоропласты; сигнал удержания ER; и подобные). В некоторых вариантах реализации настоящего изобретения гетерологичная последовательность может обеспечить метку для простоты слежения или очистки (например, флуоресцентный белок, например, зеленый флуоресцентный белок (GFP), YFP, RFP, CFP, mCherry, tdTomato и подобные; метку HIS, например, метку 6xHis; метку гемагглютинаина (HA); метку FLAG; метку Мус; и подобные). В некоторых вариантах реализации настоящего изобретения гетерологичная последовательность может обеспечивать увеличение или уменьшение стабильности. В некоторых вариантах реализации настоящего изобретения гетерологичная последовательность может обеспечивать связывающий домен (например, для того, чтобы обеспечить способность гибридного полипептида Cas9 связываться с другим представляющим интерес белком, например, ДНК или белок, модифицирующий гистон, фактор транскрипции или репрессор транскрипции, белок рекрутинга и т.д.).

Примеры различных дополнительных подходящих партнеров по слиянию (или их фрагментов) в

случае целевого варианта сайт-специфического полипептида Cas9, включают, но не ограничиваются ими, те, которые приведены на фиг. 54.

Нуклеиновая кислота, кодирующая гибридный сайт-специфический модифицирующий полипептид по изобретению.

Настоящее изобретение обеспечивает нуклеиновую кислоту, содержащую нуклеотидную последовательность, кодирующую гибридный сайт-специфический модифицирующий полипептид. В некоторых вариантах реализации настоящего изобретения нуклеиновая кислота, содержащая нуклеотидную последовательность, кодирующую гибридный сайт-специфический модифицирующий полипептид по изобретению, представляет собой вектор экспрессии, например, рекомбинантный вектор экспрессии.

В некоторых вариантах реализации настоящего изобретения способ включает приведение ДНК-мишени в контакт или введение в клетку (или популяцию клеток) одной или нескольких нуклеиновых кислот, содержащих гибридный сайт-специфический модифицирующий полипептид. Подходящие нуклеиновые кислоты, содержащие нуклеотидные последовательности, кодирующие гибридный сайт-специфический модифицирующий полипептид, содержат векторы экспрессии, при том что вектор экспрессии, содержащий нуклеотидную последовательность, кодирующую гибридный сайт-специфический модифицирующий полипептид, представляет собой "рекомбинантный вектор экспрессии".

В некоторых вариантах реализации настоящего изобретения рекомбинантный вектор экспрессии представляет собой вирусную конструкцию, например, рекомбинантную адено-ассоциированную вирусную конструкцию (см., например, патент США № 7078387), рекомбинантную аденовирусную конструкцию, рекомбинантную лентивирусную конструкцию и т.д.

Подходящие векторы экспрессии включают, но не ограничиваются ими, вирусные векторы (например, вирусные векторы, основанные на вирусе осповакцины; полиовирусе; аденовирусе (см., например, Li и соавт., *Invest Ophthalmol Vis Sci* 35:2543-2549, 1994; Borrás и соавт., *Gene Ther* 6:515-524, 1999; Li и Davidson, *PNAS* 92:7700-7704, 1995; Sakamoto и соавт., *N 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); адено-ассоциированном вирусе (см., например, AI и соавт., *Hum Gene Ther* 9:81-86, 1998; Flannery и соавт., *PNAS* 94:6916-6921, 1997; Bennett и соавт., *Invest Ophthalmol Vis Sci* 38:2857-2863, 1997; Jomary и соавт., *Gene Ther* 4:683-690, 1997; Rolling и соавт., *Hum Gene Ther* 10:641-648, 1999; AI и соавт., *Hum Mol Genet* 5:591-594, 1996; Srivastava в WO 93/09239, Samulski и соавт., *J. Vir.* (1989) 63:3822-3828; Mendelson и соавт., *Virol.* (1988) 166:154-165; и Flotte и соавт., *PNAS* (1993) 90:10613-10617); SV40; вирусе простого герпеса; вирусе иммунодефицита человека (см., например, Miyoshi и соавт., *PNAS* 94:10319-23, 1997; Takahashi и соавт., *J Virol* 73:7812-7816, 1999); ретровирусном векторе (например, вирусе лейкоза мыши, вирусе некроза селезенки и векторах, полученных из ретровирусов, таких как вирус саркомы Рауса, вирус саркомы Харви, вирус лейкоза птицы, лентивирус, вирус иммунодефицита человека, миелопролиферативный вирус саркомы и вирус опухоли молочной железы); и подобных.

Многочисленные подходящие векторы экспрессии известны специалистам в данной области техники, а многие из них коммерчески доступны. Следующие векторы приведены в качестве примера; в случае эукариотических клеток-хозяев: pXT1, pSG5 (Stratagene), pSVK3, pBPV, pMSG и pSVLSV40 (Pharmacia). Тем не менее, любой другой вектор может быть использован при условии, что он совместим с клеткой-хозяином.

В зависимости от используемой системы хозяин/вектор, любой из ряда подходящих регуляторных элементов транскрипции и трансляции, в том числе конститутивных и индуцируемых промоторов, элементов энхансера транскрипции, терминаторов транскрипции и т.д., может быть использован в векторе экспрессии (см., например, Bitter и соавт. (1987) *Methods in Enzymology*, 153:516-544).

В некоторых вариантах реализации настоящего изобретения нуклеотидная последовательность, кодирующая гибридный сайт-специфический модифицирующий полипептид, функционально связана с регуляторным элементом, например, регуляторным элементом транскрипции, таким как промотор. Регуляторный элемент транскрипции может быть функциональным либо в эукариотической клетке, например, клетке млекопитающего; либо в прокариотической клетке (например, бактериальной клетке или клетке архей). В некоторых вариантах реализации настоящего изобретения нуклеотидная последовательность, кодирующая гибридный сайт-специфический модифицирующий полипептид, функционально связана с множеством регуляторных элементов, которые позволяют осуществить экспрессию нуклеотидной последовательности, кодирующей гибридный сайт-специфический модифицирующий полипептид и в прокариотических, и в эукариотических клетках.

Неограничивающие примеры подходящих эукариотических промоторов (промоторов, функциональных в эукариотической клетке) содержат таковые из немедленно-раннего цитомегаловируса (CMV), тимидинкиназы вируса простого герпеса (HSV), раннего и позднего SV40, длинных концевых повторов (LTR) из ретровируса и металлотионеина-I мыши. Выбор соответствующего вектора и промотора находится в пределах компетенции специалиста в данной области техники. Вектор экспрессии также может содержать сайт связывания рибосомы для инициации трансляции и терминатор транскрипции. Для усиления экспрессии вектор экспрессии также может содержать соответствующие последовательности. Вектор экспрессии также может содержать нуклеотидные последовательности, кодирующие белковые метки

(например, метка 6xHis, гемагглютининовая метка (НА), флуоресцентный белок (например, зеленый флуоресцентный белок; желтый флуоресцентный белок и т.д.) и т.д.), которые являются слитыми с гибридным сайт-специфическим модифицирующим полипептидом.

В некоторых вариантах реализации настоящего изобретения нуклеотидная последовательность, кодирующая гибридный сайт-специфический модифицирующий полипептид, функционально связана с индуцируемым промотором (например, промотором теплового шока, тетрациклин-регулируемым промотором, стероид-регулируемым промотором, металл-регулируемым промотором, регулируемым рецептором эстрогена промотором и т.д.). В некоторых вариантах реализации настоящего изобретения нуклеотидная последовательность, кодирующая гибридный сайт-специфический модифицирующий полипептид, функционально связана с пространственно ограниченным и/или временно ограниченным промотором (например, тканеспецифическим промотором, специфическим к типу клеток промотором, и т.д.). В некоторых вариантах реализации настоящего изобретения нуклеотидная последовательность, кодирующая сайт-специфический модифицирующий полипептид, функционально связана с конститутивным промотором.

Для введения нуклеиновой кислоты (например, экспрессирующей конструкции) в стволовую клетку или клетку-предшественника, могут быть использованы способы введения нуклеиновой кислоты в клетку-хозяина, известные в данной области техники. Подходящие способы включают в себя, например, вирусную или бактериофаговую инфекцию, трансфекцию, конъюгацию, слияние протопластов, липофекцию, электропорацию, осаждение фосфатом кальция, опосредованную полиэтиленгликолем (PEG) трансфекцию, опосредованную DEAE-декстраном трансфекцию, опосредованную липосомами трансфекцию, технологию генной пушки, осаждение фосфатом кальция, прямую микроинъекцию, опосредованную наночастицами доставку нуклеиновой кислоты (см., например, Panyam и соавт., *Adv Drug Deliv Rev.* 2012 Sep 13. pii: S0169-409X(12)00283-9. doi: 10.1016/j.addr.2012.09.023) и подобные.

Способы.

Настоящее изобретение обеспечивает способы модификации ДНК-мишени и/или ассоциированного с ДНК-мишенью полипептида. В общем, способ по изобретению включает приведение ДНК-мишени в контакт с комплексом ("нацеливающим комплексом"), комплекс которого содержит РНК, нацеленную на ДНК, и сайт-специфический модифицирующий полипептид.

Как было описано выше, РНК-мишень, нацеленная на ДНК, и сайт-специфический модифицирующий полипептид-мишень образуют комплекс. РНК, нацеленная на ДНК, обеспечивает целевую специфичность комплексу тем, что содержит нуклеотидную последовательность, которая является комплементарной к последовательности ДНК-мишени. Сайт-специфический модифицирующий полипептид комплекса обеспечивает сайт-специфическую активность. В некоторых вариантах реализации настоящего изобретения комплекс-мишень модифицирует ДНК-мишень, что приводит к, например, расщеплению ДНК, метилированию ДНК, повреждению ДНК, репарации ДНК и т.д. В других вариантах реализации настоящего изобретения комплекс-мишень модифицирует полипептид-мишень, ассоциированный с ДНК-мишенью (например, гистон, ДНК-связывающий белок и т.д.), что приводит к, например, метилированию гистонов, ацетилированию гистонов, убиквитинированию гистонов и подобному. ДНК-мишень может быть, например, депротенинизированной ДНК *in vitro*, хромосомной ДНК в клетках *in vitro*, хромосомной ДНК в клетках *in vivo*, и т.д.

В некоторых случаях сайт-специфический модифицирующий полипептид демонстрирует нуклеазную активность, которая расщепляет ДНК-мишень в тех последовательностях ДНК-мишени, которые определены областью комплементарности между РНК, нацеленной на ДНК, и ДНК-мишенью. В некоторых случаях, если сайт-специфический модифицирующий полипептид представляет собой родственный Cas9 или Cas9 полипептид, сайт-специфическое расщепление ДНК-мишени происходит в локациях, определенных с помощью как (i) комплементарности спариваемых оснований между РНК, нацеленной на ДНК, и ДНК-мишенью; так и (ii) короткого мотива [упоминающегося как смежный с протоспейсером мотив (РАМ)] в ДНК-мишени. В некоторых вариантах реализации настоящего изобретения (например, если используют Cas9 от *S. pyogenes* или близкородственные Cas9 (см. SEQ ID NO:1-256 и 795-1346)) последовательность РАМ некомплемментарной цепи представляет собой 5'-XGG-3', где X представляет собой любой нуклеотид ДНК, а X представляет собой непосредственно 3' конец последовательности-мишени некомплемментарной цепи ДНК-мишени (см. фиг. 10). Таким образом, последовательность РАМ комплементарной цепи представляет собой 5'-CCY-3', где Y представляет собой любой нуклеотид ДНК и Y представляет собой непосредственно 5'-конец последовательности-мишени комплементарной цепи ДНК-мишени (см. фиг. 10, где РАМ некомплемментарной цепи представляет собой 5'-GGG-3' и РАМ комплементарной цепи 5'-CCC-3'). В некоторых подобных вариантах реализации настоящего изобретения X и Y могут быть комплементарными друг другу, а парой оснований X-Y может быть любая пара оснований (например, X=C и Y=G; X=G и Y=C; X=A и Y=T, X=T и Y=A).

В некоторых случаях может быть преимущественным использовать различные белки Cas9 (т.е. белки Cas9 различных видов) в различных предусмотренных способах с целью получить выгоду от различных ферментативных свойств различных белков Cas9 (например, в случае различий по предпочтительной последовательности РАМ; в случае увеличенной или уменьшенной ферментативной активности, в

случае увеличенного или уменьшенного уровня клеточной токсичности; для изменения баланса между NHEJ, репарирования по механизму гомологичной рекомбинации, одноцепочечных разрывов, двухцепочечных разрывов и т.д.). Белки Cas9 различных видов (см. SEQ ID NO:1-256 и 795-1346) могут требовать различных последовательностей PAM в ДНК-мишени. Таким образом, в случае конкретного предпочтительного белка Cas9, требование последовательности PAM может отличаться от последовательности 5'-XGG-3', описанной выше.

В данном документе были определены многие ортологи Cas9 от широкого спектра видов, и эти белки разделяют лишь несколько идентичных аминокислот. Все идентифицированные ортологи Cas9 имеют одинаковую архитектуру домена с центральным доменом эндонуклеазы HNH и доменом расщепления RuvC/RNaseH (см. фиг. 3А, 3В, 5 и табл. 1). Белки Cas9 имеют 4 основных мотива с консервативной архитектурой. Мотивы 1, 2 и 4 представляют собой RuvC-подобные мотивы, в то время как мотив 3 представляет собой HNH-мотив. В некоторых случаях подходящий сайт-специфический модифицирующий полипептид содержит аминокислотную последовательность, имеющую четыре мотива, причем каждый из мотивов 1-4 имеет идентичность по меньшей мере около 75%, по меньшей мере около 80%, по меньшей мере около 85%, по меньшей мере около 90%, по меньшей мере около 95%, по меньшей мере около 99 или 100% аминокислотной последовательности по отношению к мотивам 1-4 аминокислотной последовательности Cas9/Csn1, изображенной на фиг. 3А (SEQ ID NO:260-263, соответственно, как приведено в табл. 1), или по отношению к соответствующим участкам в любой из аминокислотных последовательностей, приведенных в SEQ ID NO:1-256 и 795-1346 (см. фиг. 5 для выравнивания мотивов 1-4 из расходящихся последовательностей Cas9). В некоторых случаях подходящий сайт-специфический модифицирующий полипептид содержит аминокислотную последовательность, имеющую идентичность по меньшей мере около 75%, по меньшей мере около 80%, по меньшей мере около 85%, по меньшей мере около 90%, по меньшей мере около 95%, по меньшей мере около 99 или 100% аминокислотной последовательности по отношению к аминокислотам 7-166 или 731-1003 аминокислотной последовательности Cas9/Csn1, изображенной на фиг. 3, или по отношению к соответствующим участкам любой из аминокислотных последовательностей, приведенных в SEQ ID NO:1-256 и 795-1346. Любой белок Cas9, как определено выше, может быть использован в качестве сайт-специфического модифицирующего полипептида или в качестве части гибридного сайт-специфического модифицирующего полипептида по заявленным способам.

Нуклеазная активность расщепляет ДНК-мишень для того, чтобы получить двухцепочечные разрывы. Эти разрывы затем репарируют с помощью клетки одним из двух способов: негомологичного соединения концов и репарирования по механизму гомологичной рекомбинации (фиг. 2). В случае негомологичного соединения концов (NHEJ), двухцепочечный разрыв репарируют с помощью прямого лигирования разорванных концов друг с другом. Таким образом, ни один новый материал нуклеиновой кислоты не вставляют в сайт, хотя некоторые материалы нуклеиновой кислоты могут быть потеряны, что приводит к делеции. В репарировании по механизму гомологичной рекомбинации донорный полинуклеотид с гомологией к расщепленной последовательности ДНК-мишени используют в качестве матрицы в случае репарирования расщепленной последовательности ДНК-мишени, что приводит к передаче генетической информации от донорного полинуклеотида к ДНК-мишени. Как таковой, новый материал нуклеиновой кислоты может быть вставлен/скопирован в сайт. В некоторых случаях ДНК-мишень контактирует с донорным полинуклеотидом-мишенью. В некоторых случаях донорный полинуклеотид-мишень вводят в клетку-мишень. Модификации ДНК-мишени, в соответствии с NHEJ и/или репарированием по механизму гомологичной рекомбинации, приводят к, например, коррекции генов, замещение генов, транспозоновому мутагенезу, вставке трансгена, делеции нуклеотида, разрушению гена, мутации гена и т.д.

Соответственно, расщепление ДНК с помощью сайт-специфического модифицирующего полипептида может быть использовано для удаления материала нуклеиновой кислоты из последовательности ДНК-мишени (например, для того, чтобы разрушить ген, который производит клетки, чувствительные к инфекции (например, ген CCR5 или CXCR4, который производит Т-клетки, чувствительные к HIV-инфекции), чтобы удалить болезнетворные тринуклеотидные повторы в нейронах, для того, чтобы создать генные нокауты и мутации в качестве моделей в исследовании и т.д.) с помощью расщепления последовательности ДНК-мишени и позволяя клетке осуществлять репарирование последовательности в отсутствие экзогенно-обеспечиваемого донорного полинуклеотида. Таким образом, заявленные способы могут быть использованы для того, чтобы нокаутировать ген (приводящего к полному отсутствию транскрипции или измененной транскрипции) или для того, чтобы изменить (knock in или нокин) генетический материал в предпочтительном локусе ДНК-мишени.

В качестве альтернативы, если РНК, нацеленная на ДНК, и сайт-специфический модифицирующий полипептид вводят совместно в клетки с донорной полинуклеотидной последовательностью, которая содержит, по меньшей мере, сегмент с гомологией к последовательности ДНК-мишени, заявленные способы могут быть использованы с целью добавления, т.е. вставки или замены, материала нуклеиновой кислоты в последовательность ДНК-мишени (например, чтобы "изменить (knock in)" нуклеиновую кислоту, которая кодирует белок, мРНК, микроРНК и т.д.), с целью добавления метки (например, 6xHis, флуоресцентного белка (например, зеленого флуоресцентного белка; желтого флуоресцентного белка и

т.д.), гемагглютинаина (НА), FLAG и т.д.), с целью добавления регуляторной последовательности к гену (например, промотора, сигнала полиаденилирования, последовательности внутренней посадки рибосомы (IRES), белка 2A, иницирующего кодона, стоп-кодона, сигнала сплайсинга, сигнала локализации и т.д.), с целью изменения последовательности нуклеиновой кислоты (например, введения мутации) и подобных. Таким образом, комплекс, содержащий РНК, нацеленную на ДНК, и сайт-специфический модифицирующий полипептид, является полезным в любом приложении, *in vitro* или *in vivo*, в котором является желательным модифицирование ДНК сайт-специфическим, т.е. "направленным" способом, например, нокаутом гена, нокином гена, "редактированием" гена, транспозоновым мутагенезом и т.д., как используют в, например, генной терапии, например, для лечения заболевания или в качестве противовирусного, противопатогенного или противоракового терапевтического средства, производства генетически модифицированных организмов в сельском хозяйстве, крупномасштабного производства белков клетками для терапевтических, диагностических или исследовательских целей, индукции клеток iPS, биологического исследования, нацеливания на гены патогенов с целью делеции или замены и т.д.

В некоторых вариантах реализации настоящего изобретения сайт-специфический модифицирующий полипептид содержит измененную форму белка Cas9/Csn1. В некоторых случаях измененная форма белка Cas9/Csn1 содержит замену аминокислоты (например, делецию, вставку или замещение), что снижает встречающуюся в природе нуклеазную активность белка Cas9/Csn1. Например, в некоторых случаях, измененная форма белка Cas9/Csn1 имеет менее 50%, менее 40%, менее 30%, менее 20%, менее 10%, менее 5% или менее 1% нуклеазной активности соответствующего полипептида Cas9/Csn1 дикого типа. В некоторых случаях измененная форма полипептида Cas9/Csn1 не имеет никакой существенной нуклеазной активности. Если сайт-специфический модифицирующий полипептид является модифицированной формой полипептида Cas9/Csn1, который не имеет никакой существенной нуклеазной активности, он может упоминаться как "dCas9".

В некоторых вариантах реализации настоящего изобретения измененная форма полипептида Cas9/Csn1 содержит мутацию D10A (аспартат на аланин в положении аминокислоты 10 SEQ ID NO:8) (или соответствующую мутацию любого из белков, представленных в SEQ ID NO:1-256 и 795-1346), которая может расщеплять комплементарную цепь ДНК-мишени, но имеет пониженную способность расщеплять некомплемтарную цепь ДНК-мишени (таким образом, что приводит к одноцепочечному разрыву (SSB) вместо DSB; см. фиг. 11). В некоторых вариантах реализации настоящего изобретения измененная форма полипептида Cas9/Csn1 содержит мутацию H840A (гистидин на аланин в положении аминокислоты 840 SEQ ID NO:8) (или соответствующую мутацию любого из белков, представленных в SEQ ID NO:1-256 и 795-1346), которая может расщеплять комплементарную цепь ДНК-мишени, но имеет пониженную способность расщеплять комплементарную цепь ДНК-мишени (таким образом, что приводит к одноцепочечному разрыву (SSB) вместо DSB; см. фиг. 11). Использование варианта D10A или H840A Cas9 (или соответствующих мутаций в любом из белков, указанных как SEQ ID NO:1-256 и 795-1346) может изменить ожидаемый биологический результат, ввиду того, что негомологичное соединение концов (NHEJ) является гораздо более вероятным, при условии того, что присутствуют DSB, в отличие от SSB. Таким образом, в некоторых случаях, если является желательным снижение вероятности DSB (и, следовательно, уменьшение вероятности NHEJ), может быть использован вариант D10A или H840A Cas9. Другие остатки могут подвергаться мутации с целью достижения подобных эффектов (т.е. инактивирования одного или другого участка нуклеазы). В качестве неограничивающих примеров, остатки D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986 и/или A987 (или соответствующие мутации любого из белков, приведенных в SEQ ID NO:1-256 и 795-1346) могут быть изменены (т.е. заменены) (см. фиг. 3, 5, 11A и табл. 1 с получением дополнительной информации о сохранении аминокислотных остатков Cas9). Кроме того, подходящими являются мутации, отличные от замещения аланина. В некоторых вариантах реализации настоящего изобретения, при условии, что сайт-специфический полипептид (например, сайт-специфический модифицирующий полипептид) снижает каталитическую активность (например, когда белок Cas9 имеет D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986 и/или мутацию A987, например, D10A, G12A, G17A, E762A, H840A, N854A, N863A, H982A, H983A, A984A и/или D986A), полипептид может по-прежнему связываться с ДНК-мишенью сайт-специфическим образом (так как он все еще направляется в последовательность ДНК-мишени с помощью РНК, нацеленной на ДНК), пока он сохраняет способность взаимодействовать с РНК, нацеленной на ДНК.

В некоторых вариантах реализации настоящего изобретения измененная форма полипептида Cas9/Csn1 содержит и мутацию D10A, и мутацию H840A (или соответствующие мутации любого из белков, приведенных в SEQ ID NO:1-256 и 795-1346), причем полипептид имеет пониженную способность расщеплять и комплементарные, и некомплемтарные цепи ДНК-мишени. Другие остатки могут подвергаться мутации с целью достижения подобных эффектов (т.е. инактивирования одного или другого участка нуклеазы). В качестве неограничивающих примеров, остатки D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986 и/или A987 (или соответствующие мутации любого из белков, приведенные в SEQ ID NO:1-256 и 795-1346) могут быть изменены (т.е. заменены) (см. фиг. 3, 5, 11A и табл. 1 с получением дополнительной информации о сохранении аминокислотных остатков Cas9). Кроме того,

подходящими являются мутации, отличные от замещения аланина.

В некоторых вариантах реализации настоящего изобретения сайт-специфический модифицирующий полипептид содержит гетерологичную последовательность (например, слитую). В некоторых вариантах реализации настоящего изобретения гетерологичная последовательность может обеспечивать субклеточную локализацию сайт-специфического модифицирующего полипептида (например, клеточный сигнал внутриядерной локализации (NLS) в случае нацеливания на ядро; сигнал митохондриальной локализации в случае нацеливания на митохондрии; сигнал локализации хлоропласта в случае нацеливания на хлоропласты; сигнал удержания ER; и подобные). В некоторых вариантах реализации настоящего изобретения гетерологичная последовательность может обеспечить метку для простоты слежения или очистки (например, флуоресцентный белок, например, зеленый флуоресцентный белок (GFP), YFP, RFP, CFP, mCherry, tdTomato и подобные; метку HIS, например, метку 6xHis; метку гематоглиутинина (HA); метку FLAG; метку Мус; и подобные). В некоторых вариантах реализации настоящего изобретения гетерологичная последовательность может обеспечивать увеличение или уменьшение стабильности.

В некоторых вариантах реализации настоящего изобретения сайт-специфический модифицирующий полипептид-мишень может быть оптимизированы по кодонам. Этот тип оптимизации известен в данной области техники и предусматривает мутацию чужеродной ДНК для того, чтобы имитировать предпочтения кодонов предполагаемого организма-хозяина или клетки при кодировании того же белка. Таким образом, кодоны изменяют, но кодируемый белок остается неизменным. Например, если предназначенная клетка-мишень представляет собой клетку человека, Cas9, оптимизированный по кодонам человека (или вариант, например, ферментативно неактивный вариант) будет подходящим сайт-специфическим модифицирующим полипептидом (см. SEQ ID NO:256 в качестве примера). Любой подходящий сайт-специфический модифицирующий полипептид (например, любой Cas9, такой, как любая из последовательностей, представленных в SEQ ID NO:1-256 и 795-1346) может быть оптимизирована по кодонам. В качестве отличного неограничивающего примера, при условии, что предназначенная клетка-хозяин является клеткой мыши, то оптимизированный по кодонам мыши Cas9 (или вариант, например, ферментативно неактивный вариант) будет подходящим сайт-специфическим модифицирующим полипептидом. В то время как оптимизация кодонов не обязательна, она допустима и в некоторых случаях может быть предпочтительной.

В некоторых вариантах реализации настоящего изобретения РНК, нацеленную на ДНК, и сайт-специфический модифицирующий полипептид используют в качестве индуцируемой системы для прекращения экспрессии гена в бактериальных клетках. В некоторых случаях нуклеиновые кислоты, кодирующие соответствующую РНК, нацеленную на ДНК, и/или соответствующий сайт-специфический полипептид, вводят в хромосому клетки-мишени под контроль индуцируемого промотора. Когда РНК, нацеленную на ДНК, и/или сайт-специфический полипептид индуцируют, ДНК-мишень расщепляется (или модифицируется другим путем) в локализации, представляющей интерес (например, в гене-мишени на отдельной плазмиде), при условии, что и РНК, нацеленная на ДНК, и сайт-специфический модифицирующий полипептид присутствуют и образуют комплекс. Таким образом, в некоторых случаях, штаммы экспрессирующих бактерий создают так, чтобы они содержали последовательности нуклеиновых кислот, кодирующие соответствующий сайт-специфический модифицирующий полипептид в бактериальном геноме и/или соответствующую РНК, нацеленную на ДНК, на плазмиде (например, под контролем индуцируемого промотор), осуществляя эксперименты, в которых экспрессию любого гена-мишени (экспрессированного с отдельной плазмиды введенной в штамм) возможно контролировать с помощью индуцирования экспрессии РНК, нацеленной на ДНК, и сайт-специфического полипептида.

В некоторых случаях сайт-специфический модифицирующий полипептид имеет ферментативную активность, которая модифицирует ДНК-мишень отличными от введения двухцепочечных разрывов способами. Представляющая интерес ферментативная активность, которая может быть использована для модификации ДНК-мишени (например, с помощью слияния гетерологичного полипептида обладающего ферментативной активностью с сайт-специфическим модифицирующим полипептидом, таким образом, получая гибридный сайт-специфический модифицирующий полипептид), является, но без ограничения ими, активность метилтрансферазы, активность деметилазы, активность репарации ДНК, активность повреждения ДНК, активность дезаминирования, активность дисмутаза, алкилирующую активность, активность депуринизации, активность окисления, активность формирования димера пиримидина, активность интегразы, активность транспозазы, активность рекомбиназы, полимеразную активность, лигазную активность, геликазную активность, активность фотолиазы или гликозилазную активность). Как это считается в данной области техники, метилирование и деметилирование являются важными способами эпигенетического регулирования гена, в то время как повреждение ДНК и активность репарации являются важными для выживания клеток и для надлежащего поддержания генома в ответ на стрессовое воздействие окружающей среды.

Как таковые, способы в данном документе находят применение в эпигенетической модификации ДНК-мишени и могут быть использованы для того, чтобы контролировать эпигенетическую модификацию ДНК-мишени в любой локализации в ДНК-мишени с помощью генной инженерии необходимой комплементарной последовательности нуклеиновой кислоты в ДНК-нацеленном сегменте РНК, нацелен-

ной на ДНК. Способы данного документа также находят применение в намеренном и контролируемом повреждении ДНК в любой необходимой локализации в пределах ДНК-мишени. Способы в данном документе также находят применение в последовательность-специфическом и контролируемом репарировании ДНК в любой необходимой локализации в пределах ДНК-мишени. Способы направления ферментативной активности модификации ДНК-мишени на специфические локализации ДНК-мишени находят применение как в научно-исследовательских, так и в клинических применениях.

В некоторых случаях сайт-специфический модифицирующий полипептид обладает активностью, которая модулирует транскрипцию ДНК-мишени (например, в случае гибридного сайт-специфического модифицирующего полипептида, и т.д.). В некоторых случаях гибридные сайт-специфические модифицирующие полипептиды, содержащие гетерологичный полипептид, который обладает способностью увеличивать или уменьшать транскрипцию (например, транскрипционный активатор или полипептиды репрессора транскрипции), используют для увеличения или уменьшения транскрипции ДНК-мишени в специфической локализации в ДНК-мишени, которую направляют с помощью нацеленного на ДНК сегмента РНК, нацеленной на ДНК. Примеры исходных полипептидов для обеспечения гибридного сайт-специфического модифицирующего полипептида с модулирующей активностью транскрипции включают, но не ограничиваются ими, индуцируемые светом регуляторы транскрипции, чувствительные к небольшим молекулам/лекарственным средствам регуляторы транскрипции, факторы транскрипции, репрессоры транскрипции и т.д. В некоторых случаях заявленный способ используют для контроля экспрессии кодирующей РНК-мишени (кодирующий белок ген) и/или некодирующей РНК-мишени (например, тРНК, рРНК, мнРНК, миРНК, микроРНК, длинной некодирующей РНК и т.д.).

В некоторых случаях сайт-специфический модифицирующий полипептид имеет ферментативную активность, которая модифицирует полипептид, ассоциированный с ДНК (например, гистон). В некоторых вариантах реализации настоящего изобретения ферментативная активность является активностью метилтрансферазы, активностью деметилазы, ацетилтрансферазной активностью, активностью деацетилазы, киназной активностью, фосфатазной активностью, активностью убиквитинлигазы, например убиквитинизирующей (активностью), деубиквитинизирующей активностью, активностью аденилирования, активностью деаденилирования, активностью сумоилирования, активностью десумоилирования, активностью рибозилирования, активностью дерибозилирования, активностью миристоилирования, активностью демиристоилирования, гликозилирующей активностью (например, от O-GlcNAc трансферазы) или дегликозилирующей активностью. Ферментативные активности, приведенные в данном документе, катализируют ковалентные модификации белков. В данной области техники известно, что подобные модификации влияют на устойчивость или активность белка-мишени (например, фосфорилирование вследствие киназной активности, может стимулировать или подавлять активность белка в зависимости от белка-мишени). Особый интерес представляет тот случай, когда белковые мишени представляют собой гистоны. В данной области техники известно, что гистоновые белки связывают ДНК и образуют комплексы, известные как нуклеосомы. Гистоны могут быть модифицированы (например, с помощью метилирования, ацетилирования, убиквинилирования, фосфорилирования) для того, чтобы вызвать структурные изменения в окружающей ДНК, тем самым контролируя доступность потенциально больших участков ДНК для взаимодействующих факторов, таких как факторы транскрипции, полимеразы и подобные. Один гистон может быть изменен многими различными способами и в различных комбинациях (например, триметилирование лизина 27 гистона 3, H3K27 связывают с областями ДНК подавленной транскрипции, в то время как триметилирование лизина 4 гистона 3, H3K4, связывают с областями ДНК активной транскрипции). Таким образом, сайт-специфический модифицирующий полипептид с гистон-модифицирующей активностью находит применение в сайт-специфическом контроле структуры ДНК и может быть использован с целью изменения паттерна модификации гистонов в выбранной области ДНК-мишени. Такие способы находят применение и в исследовании, и в клинических применениях.

В некоторых вариантах реализации настоящего изобретения несколько РНК, нацеленных на ДНК, используют одновременно, для того, чтобы одновременно изменять различные локализации на той же ДНК-мишени или на различных ДНК-мишенях. В некоторых вариантах реализации настоящего изобретения две или более РНК, нацеленные на ДНК, нацелены на один и тот же ген, транскрипт или локус. В некоторых вариантах реализации настоящего изобретения две или более РНК, нацеленные на ДНК, нацелены на различные не связанные локусы. В некоторых вариантах реализации настоящего изобретения две или более РНК, нацеленные на ДНК, нацелены на различные, но связанные локусы.

В некоторых случаях сайт-специфический модифицирующий полипептид обеспечивают непосредственно в виде белка. В качестве отдельного неограничивающего примера, грибок (например, дрожжи) может быть трансформирован экзогенным белком и/или нуклеиновой кислотой с использованием трансформации сферопластов (см. Kawai и соавт., *Bioeng Bugs*. 2010 Nov-Dec; 1(6):395-403: "Transformation of *Saccharomyces cerevisiae* and other fungi: methods and possible underlying mechanism"; и Tanka и соавт., *Nature*. 2004 Mar 18; 428(6980):323-8: "Conformational variations in an infectious protein determine prion strain differences"; обе из которых содержатся в данном документе в виде ссылки во всей своей полноте). Таким образом, сайт-специфический модифицирующий полипептид (например, Cas9) может быть инкорпорирован в сферопласт (с или без нуклеиновой кислоты, кодирующей РНК, нацеленную на ДНК, и с

или без донорного полинуклеотида), а сферопласт может быть использован для введения содержимого в дрожжевую клетку. Сайт-специфический модифицирующий полипептид может быть введен в клетку (доставлен в клетку) любым пригодным способом; подобные способы известны специалистам с обычной квалификацией в данной области техники. В качестве другого неограничивающего примера, сайт-специфический модифицирующий полипептид может быть введен непосредственно в клетку (например, с или без нуклеиновой кислоты, кодирующей РНК, нацеленную на ДНК, и с или без донорного полинуклеотида), например клетка эмбриона данио-рерио, пронуклеус оплодотворенной яйцеклетки мыши и т.д.

Представляющие интерес клетки-мишени.

В некоторых из перечисленных выше приложениях, заявленные способы могут быть использованы для того, чтобы вызвать расщепление ДНК, модификацию ДНК и/или модуляцию транскрипции в митотических или постмитотических клетках *in vivo* и/или *ex vivo* и/или *in vitro* (например, с получением генетически модифицированных клеток, которые могут быть снова введены человеку). Поскольку РНК, нацеленная на ДНК, обеспечивает специфичность гибридизации с ДНК-мишенью, митотическая и/или постмитотическая представляющая интерес клетка по описанным способам включает клетку любого организма (например, бактериальную клетку, клетку архей, одноклеточный эукариотический организм, клетку растений, например, *Botryococcus braunii*, *Chlamydomonas reinhardtii*, *Nannochloropsis gaditana*, *Chlorella pyrenoidosa*, *Sargassum patens* C. Agardh, и подобные, клетку грибка (например, дрожжевую клетку), клетку животного, клетку беспозвоночного животного (например, плодовой мушки, кишечнополостной, иглокожего, нематоды, и т.д.), клетку позвоночного животного (например, рыбы, амфибии, рептилии, птицы, млекопитающего), клетку млекопитающего, клетку грызуна, клетку человека и т.д.).

Любой тип клеток может представлять интерес (например, стволовая клетка, например, эмбриональная стволовая (ES) клетка, индуцированная плюропотентная стволовая (iPS) клетка, половая клетка; соматическая клетка, например, фибробласт, гематопоэтическая клетка, нейрон, мышечная клетка, костная клетка, гепатоцит, панкреатическая клетка; *in vitro* или *in vivo* эмбриональная клетка эмбриона на любой стадии, например 1-, 2-, 4-, 8-клеточная и т.д. стадия эмбрионов данио-рерио; и т.д.). Клетки могут быть от установившихся клеточных линий или же они могут быть первичными клетками, причем "первичные клетки", "линии первичных клеток" и "первичные культуры" используют в данном документе взаимозаменяемо для обозначения клеток и клеточных культур, которые были получены от субъекта, и которым дали возможность расти *in vitro* в течение ограниченного числа пассажей, т.е. делений, культуры. Например, первичные культуры являются культурами, которые, возможно, были пассажированы 0, 1, 2, 4, 5, 10 или 15 раз, но недостаточное количество раз, чтобы пройти через стадию криза. Как правило, первичные клеточные линии по настоящему изобретению поддерживают в течение менее чем 10 пассажей *in vitro*. Клетки-мишени во многих вариантах реализации настоящего изобретения представляют собой одноклеточные организмы или растут в культуре.

Если клетки являются первичными клетками, они могут быть собраны от индивидуума любым пригодным способом. Например, лейкоциты могут быть собраны пригодным способом с помощью афереза, лейкоцитоза, разделения по градиенту плотности, и т.д., в то время как клетки тканей, таких как кожа, мышца, костный мозг, селезенка, печень, поджелудочная железа, легкое, кишечник, желудок и т.д., являются наиболее пригодными для сбора их с помощью биопсии. Соответствующий раствор может быть использован для дисперсии или суспензии собранных клеток. Подобный раствор, как правило, будет сбалансированным солевым раствором, например, физиологическим раствором, фосфатно-солевым буферным раствором (PBS), сбалансированным солевым раствором Хенкса, и т.д., как правило дополненным фетальной телячьей сывороткой или другим встречающимся в природе фактором, в сочетании с приемлемым буфером при низкой концентрации, обычно от 5-25 мМ. Пригодные буферы содержат HEPES, фосфатные буферы, лактатные буферы и т.д. Эти клетки могут быть использованы незамедлительно или они могут быть сохранены, заморожены в течение длительных периодов времени и могут быть разморожены и использованы повторно. В таких случаях, как правило, клетки будут заморожены в 10% DMSO, 50% сыворотки, 40% буферной среде или некотором отличном подобном растворе, как обычно используют в данной области техники для того, чтобы сохранить клетки при подобных низких температурах, и размораживают широко известным в данной области техники способом для размораживания замороженных культивируемых клеток.

Нуклеиновые кислоты, кодирующие РНК, нацеленную на ДНК, и/или сайт-специфический модифицирующий полипептид.

В некоторых вариантах реализации настоящего изобретения заявленный способ включает приведение ДНК-мишени в контакт с или введение в клетку (или популяцию клеток) одной или нескольких нуклеиновых кислот, содержащих нуклеотидные последовательности, кодирующие РНК, нацеленную на ДНК, и/или сайт-специфический модифицирующий полипептид и/или донорный полинуклеотид. Подходящие нуклеиновые кислоты, (содержащие нуклеотидные последовательности, кодирующие РНК, нацеленную на ДНК, и/или сайт-специфический модифицирующий полипептид, содержат векторы экспрессии, при том что вектор экспрессии, содержащий нуклеотидную последовательность, кодирующую РНК, нацеленную на ДНК, и/или сайт-специфический модифицирующий полипептид, представляет собой "рекомбинантный вектор экспрессии".

В некоторых вариантах реализации настоящего изобретения рекомбинантный вектор экспрессии представляет собой вирусную конструкцию, например, рекомбинантную адено-ассоциированную вирусную конструкцию (см., например, патент США № 7078387), рекомбинантную аденовирусную конструкцию, рекомбинантную лентивирусную конструкцию и т.д.

Подходящие векторы экспрессии включают, но не ограничиваются ими, вирусные векторы (например, вирусные векторы, основанные на вирусе осповакцины; полиовирусе; аденовирусе (см., например, Li и соавт., *Invest Ophthalmol Vis Sci* 35:2543-2549, 1994; Borrás и соавт., *Gene Ther* 6:515-524, 1999; Li и Davidson, *PNAS* 92:7700-7704, 1995; Sakamoto и соавт., *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); адено-ассоциированном вирусе (см., например, Ali и соавт., *Hum Gene Ther* 9:81-86, 1998; Flannery и соавт., *PNAS* 94:6916-6921, 1997; Bennett и соавт., *Invest Ophthalmol Vis Sci* 38:2857-2863, 1997; Jomary и соавт., *Gene Ther* 4:683-690, 1997; Rolling и соавт., *Hum Gene Ther* 10:641-648, 1999; Ali и соавт., *Hum Mol Genet* 5:591-594, 1996; Srivastava в WO 93/09239, Samulski и соавт., *J. Vir.* (1989) 63:3822-3828; Mendelson и соавт., *Virol.* (1988) 166:154-165; и Flotte и соавт., *PNAS* (1993) 90:10613-10617); SV40; вирусе простого герпеса; вирусе иммунодефицита человека (см., например, Miyoshi и соавт., *PNAS* 94:10319-23, 1997; Takahashi и соавт., *J Virol* 73:7812-7816, 1999); ретровирусном векторе (например, вирусе лейкоза мыши, вирусе некроза селезенки и векторах, полученных из ретровирусов, таких как вирус саркомы Рауса, вируса саркомы Харви, вирус лейкоза птицы, лентивирус, вирус иммунодефицита человека, миелопролиферативный вирус саркомы и вирус опухоли молочной железы); и подобные.

Многочисленные подходящие векторы экспрессии известны специалистам в данной области техники, а многие из них коммерчески доступны. Следующие векторы приведены в качестве примера; в случае эукариотических клеток-хозяев: pXT1, pSG5 (Stratagene), pSVK3, pBPV, pMSG и pSVLVS40 (Pharmacia). Тем не менее, любой другой вектор может быть использован при условии, что он совместим с клеткой-хозяином.

В некоторых вариантах реализации настоящего изобретения нуклеотидная последовательность, кодирующая РНК, нацеленную на ДНК, и/или сайт-специфический модифицирующий полипептид, функционально связаны с регуляторным элементом, например, регуляторным элементом транскрипции, таким как промотор. Регуляторный элемент транскрипции может быть функциональным либо в эукариотической клетке, например, клетке млекопитающего; либо в прокариотической клетке (например, в бактериальной клетке или клетке археи). В некоторых вариантах реализации настоящего изобретения нуклеотидная последовательность, кодирующая РНК, нацеленную на ДНК, и/или сайт-специфический модифицирующий полипептид, функционально связаны с множеством регуляторных элементов, которые обеспечивают экспрессию нуклеотидной последовательности, кодирующей РНК, нацеленную на ДНК, и/или сайт-специфический модифицирующий полипептид в прокариотических и эукариотических клетках.

В зависимости от используемой системы хозяин/вектор, любой из ряда подходящих элементов контроля транскрипции и трансляции, в том числе и конститутивные и индуцируемые промоторы, энхансеры транскрипции, терминаторы транскрипции и т.д., может быть использован в векторе экспрессии (например, промоторе U6, промоторе H1 и т.д.; см. выше) (см., например, Bitter и соавт. (1987) *Methods in Enzymology*, 153:516-544).

В некоторых вариантах реализации настоящего изобретения РНК, нацеленная на ДНК, и/или сайт-специфический модифицирующий полипептид могут быть представлены в виде РНК. В таких случаях, РНК, нацеленная на ДНК, и/или РНК, кодирующая сайт-специфический модифицирующий полипептид, могут быть получены с помощью непосредственного химического синтеза или могут быть транскрибированы *in vitro* из ДНК, кодирующей РНК, нацеленную на ДНК. Способы синтеза РНК из матрицы ДНК хорошо известны в данной области техники. В некоторых случаях РНК, нацеленная на ДНК, и/или РНК, кодирующая сайт-специфический модифицирующий полипептид, синтезируют *in vitro* с помощью фермента РНК-полимеразы (например, полимеразы T7, полимеразы T3, полимеразы SP6 и т.д.). После того, как была синтезирована, РНК может быть непосредственно связана с ДНК-мишенью или может быть введена в клетки с помощью любого из хорошо известных способов введения нуклеиновых кислот в клетки (например, микроинъекции, электропорации, трансфекции и т.д.).

Нуклеотиды, кодирующие РНК, нацеленную на ДНК (вводят либо в виде ДНК, либо в виде РНК), и/или сайт-специфический модифицирующий полипептид (вводят либо в виде ДНК, либо в виде РНК), и/или донорный полинуклеотид могут быть введены клеткам с использованием хорошо развитых способов трансфекции; см., например, Angel и Yanik (2010) *PLoS ONE* 5(7): e11756, и имеющихся в продаже реагентов TransMessenger® от Qiagen, Stemfect™ РНК Transfection Kit от Stemgent, и TransIT®-mРНК Transfection Kit от Minis Bio LLC. См. также Beumer и соавт. (2008) Efficient gene targeting in *Drosophila* by direct embryo injection with zinc-finger nucleases. *PNAS* 105(50): 19821-19826. Альтернативно, нуклеиновые кислоты, кодирующие РНК, нацеленную на ДНК, и/или сайт-специфический модифицирующий полипептид и/или гибридный сайт-специфический модифицирующий полипептид и/или донорный полинуклеотид могут быть представлены на векторах ДНК. Доступны многие векторы, например, плазмиды, космиды, мини-кольца, фаги, вирусы и т.д., полезные для передачи нуклеиновых кислот в клетки-мишени. Векторы, содержащие нуклеиновую кислоту(ы), могут быть сохранены эписомально, например,

такие как плазмиды, мини-кольца ДНК, вирусы, такие как цитомегаловирус, аденовирус, и т.д., или они могут быть интегрированы в геном клетки-мишени, посредством гомологичной рекомбинации или случайной интеграции, например, векторы, полученные из ретровируса, такие как MMLV, HIV-1, ALV и т.д.

Векторы могут быть введены непосредственно клеткам-мишеням. Другими словами, эти клетки контактируют с векторами, содержащими нуклеиновую кислоту, кодирующую РНК, нацеленную на ДНК, и/или сайт-специфический модифицирующий полипептид и/или гибридный сайт-специфический модифицирующий полипептид и/или донорный полинуклеотид, так, что векторы поглощаются клетками. Способы контактирования клеток с векторами нуклеиновых кислот, которые являются плазмидами, в том числе электропорация, трансфекция хлоридом кальция, микроинъекция и липофекция, хорошо известны в данной области техники. В случае доставки вирусного вектора, клетки контактируют с вирусными частицами, содержащими нуклеиновую кислоту, кодирующую РНК, нацеленную на ДНК, и/или сайт-специфический модифицирующий полипептид и/или гибридный сайт-специфический модифицирующий полипептид и/или донорный полинуклеотид. Ретровирусы, например, лентивирусы, являются особенно подходящими для способа по настоящему изобретению. Часто используемые ретровирусные векторы являются "дефектными", т.е. не в состоянии производить вирусные белки, необходимые для продуктивной инфекции. Предпочтительно, репликация вектора требует роста в линии клеток с дефектом упаковки. В случае генерации вирусных частиц, содержащих представляющие интерес нуклеиновые кислоты, ретровирусные нуклеиновые кислоты, содержащие нуклеиновую кислоту, упаковывают в вирусные капсиды с помощью линии клеток с дефектом упаковки. Различные линии клеток с дефектом упаковки обеспечивают различный оболочечный белок (экотропный, амфотропный или ксенотропный), который будет включен в капсид, этот оболочечный белок определяет для клеток специфичность вирусной частицы (экотропной в случае мышей и крыс; амфотропной в случае большинства типов клеток млекопитающих, в том числе человека, собаки и мыши, и ксенотропной в случае большинства типов клеток млекопитающих, за исключением мышинных клеток). Соответствующая линия клеток с дефектом упаковки может быть использована для того, чтобы упакованные вирусные частицы были нацелены на клетки. Способы введения ретровирусных векторов, содержащих нуклеиновую кислоту, кодирующую факторы перепрограммирования в линии клеток с дефектом упаковки и сбор вирусных частиц, которые создаются с помощью линий упаковки, хорошо известных в данной области техники. Нуклеиновые кислоты могут быть также введены с помощью непосредственной микроинъекции (например, инъекции РНК в эмбрион данио-рерио).

Векторы, используемые для введения в клетки-мишени нуклеиновых кислот, кодирующих РНК, нацеленную на ДНК, и/или сайт-специфический модифицирующий полипептид и/или гибридный сайт-специфический модифицирующий полипептид и/или донорный полинуклеотид, как правило, содержат промоторы, пригодные для запуска экспрессии, т.е. активации транскрипции представляющей интерес нуклеиновой кислоты. Другими словами, представляющая интерес нуклеиновая кислота будет функционально связана с промотором. Это может содержать повсеместно действующие промоторы, например, промотор CMV- $\beta$ -актина или индуцируемые промоторы, такие как промоторы, которые являются активными в отдельных популяциях клеток или те, которые отвечают на присутствие лекарственных средств, таких как тетрациклин. С помощью транскрипционной активации, предполагается, что транскрипция будет увеличена выше основных уровней в клетке-мишени по меньшей мере в около 10 раз, по меньшей мере в около 100 раз, а точнее по меньшей мере в около 1000 раз. Кроме того, векторы, используемые для введения РНК, нацеленной на ДНК, и/или сайт-специфического модифицирующего полипептида и/или гибридного сайт-специфического модифицирующего полипептида и/или донорного полинуклеотида клеткам-мишеням, могут содержать последовательности нуклеиновых кислот, которые кодируют селективируемые маркеры в клетках-мишенях, таким образом делая возможной идентификацию клеток, которые поглощают РНК, нацеленную на ДНК, и/или сайт-специфический модифицирующий полипептид и/или гибридный сайт-специфический модифицирующий полипептид и/или донорный полинуклеотид.

РНК, нацеленная на ДНК, и/или сайт-специфический модифицирующий полипептид и/или гибридный сайт-специфический модифицирующий полипептид вместо этого могут быть использованы для контактирования с ДНК или введены в клетки в виде РНК. Способы введения РНК в клетки известны в данной области техники и могут содержать, например, прямую инъекцию, трансфекцию или любой другой способ, используемый для введения ДНК.

Сайт-специфический модифицирующий полипептид-мишень вместо этого может быть предоставлен клеткам в виде полипептида. Такой полипептид может необязательно быть слитым с доменом полипептида, который увеличивает растворимость продукта. Домен может быть связан с полипептидом посредством определенного сайта расщепления протеазой, например, последовательностью TEV, которая расщепляется протеазой TEV. Линкер также может содержать одну или несколько гибких последовательностей, например, из от 1 до 10 остатков глицина. В некоторых вариантах реализации настоящего изобретения расщепление слитого белка осуществляют в буфере, который поддерживает растворимость продукта, например, в присутствии от 0,5 до 2 М мочевины, в присутствии полипептидов и/или полинуклеотидов, которые увеличивают растворимость, и подобных. Представляющие интерес домены содержат эндосомолитические домены, например, домен гриппа HA; и другие полипептиды, которые по-

могут в производстве, например, домена IF2, домена GST, домена GRPE и подобных. Полипептид может быть приготовлен для улучшения стабильности. Например, пептиды могут быть пегилированы, при том что полиэтиленоксигруппа обеспечивает улучшенное время жизни в кровотоке.

Дополнительно или альтернативно, сайт-специфический модифицирующий полипептид-мишень может быть слит с доменом полипептида проникающего вещества, чтобы способствовать поглощению клеткой. Множество доменов проникающего вещества известны в данной области техники и могут быть использованы в не-интегрирующих полипептидах по настоящему изобретению, в том числе пептидах, пептидомиметиках и непептидных носителях. Например, пептид проникающего вещества может быть производным от третьей альфа-спирали фактора транскрипции дрозифилы обыкновенной, именуемой пенетратин, которая содержит аминокислотную последовательность RQIKIWFQNRRMKWKK (SEQ ID NO://). В качестве другого примера, пептид проникающего вещества содержит аминокислотную последовательность основной области tat HIV-1, которая может содержать, например, аминокислоты 49-57 встречающегося в природе белка tat. Другие проникающие домены содержат поли-аргининовые мотивы, например, область аминокислот 34-56 белка rev HIV-1, нона-аргинин, окта-аргинин и подобные (см., например, Futaki и соавт. (2003) *Curr Protein Pept Sci.* 2003 Apr; 4(2): 87-9 и 446; и Wender и соавт. (2000) *Proc. Natl. Acad. Sci. U.S.A* 2000 Nov. 21; 97(24): 13003-8; опубликованные заявки на патенты США № 20030220334; 20030083256; 20030032593 и 20030022831, которые содержатся в данном документе исключительно в качестве ссылки для изложения транслокационных пептидов и пептоидов). Последовательность нона-аргинина (R9) является одной из наиболее эффективных РТД, которые были описаны (Wender и соавт. 2000; Uemura и соавт. 2002). Сайт, на котором осуществляют слияние, может быть выбран для того, чтобы оптимизировать характеристики биологической активности, секреции или связывания полипептида. Оптимальный сайт определяют с помощью стандартного экспериментирования.

Сайт-специфический модифицирующий полипептид по изобретению может быть получен *in vitro* или с помощью эукариотических клеток или с помощью прокариотических клеток и может быть дополнительно процессирован посредством разворачивания, например, тепловой денатурации, снижения DTT и т.д., и может быть в дальнейшем ренатурирован, с использованием способов, известных в данной области техники.

Представляющие интерес модификации, которые не изменяют первичную последовательность, содержат химическое модифицирование полипептидов, например, ацилирование, ацетилирование, карбоксилирование, амидирование и т.д. Кроме того, содержатся модификации гликозилирования, например, сделанные путем модификации паттернов гликозилирования полипептида во время его синтеза и процессинга или на дополнительных этапах процессинга; например, с помощью воздействия на полипептид ферментов, которые влияют на гликозилирование, например, ферментов гликозилирования или дегликозилирования у млекопитающих. Помимо этого, приведенное представляет собой последовательности, которые имеют фосфорилированные аминокислотные остатки, например, фосфотирозин, фосфосерин или фосфотreonин.

Кроме того, в настоящем изобретении предлагаются РНК, нацеленные на ДНК, и сайт-специфические модифицирующие полипептиды, которые были модифицированы с помощью обычных молекулярно-биологических способов и синтетической химии с тем, чтобы улучшить их устойчивость к протеолитической деградации, чтобы изменить специфичность последовательности-мишени для того, чтобы оптимизировать свойства растворимости, чтобы изменить белковую активность (например, активность модулирования транскрипции, ферментативную активность и т.д.) или для придания им более подходящих свойств для терапевтического применения. Аналоги таких полипептидов содержат те, которые содержат отличные от встречающихся в природе L-аминокислот, например, D-аминокислоты или встречающиеся в природе синтетические аминокислоты. D-аминокислоты могут быть заменены на некоторые или на все аминокислотные остатки.

Сайт-специфические модифицирующие полипептиды могут быть получены с помощью *in vitro* синтеза с использованием обычных способов, известных в данной области техники. Доступны различные коммерческие синтетические аппараты, например, автоматизированные синтезаторы от Applied Biosystems, Inc., Beckman, и т.д. С помощью синтезаторов, встречающиеся в природе аминокислоты могут быть замещены неприродными аминокислотами. Отдельная последовательность и способ подготовки будут определяться удобством, экономикой, необходимой чистотой и подобным.

При желании, различные группы могут быть введены в белок в ходе синтеза или во время экспрессии, которые позволяют связываться с другими молекулами или с поверхностью. Таким образом, могут быть использованы цистеины, для того, чтобы получить тиоэфиры, гистидины для связывания с комплексом металлического иона, карбоксильные группы для образования амидов или сложных эфиров, аминогруппы для образования амидов и подобные.

Сайт-специфические модифицирующие полипептиды могут быть также выделены и очищены в соответствии с обычными способами рекомбинантного синтеза. Из экспрессирующего хозяина может быть получен лизат, который затем очищают с помощью HPLC, эксклюзионной хроматографии, гель-электрофореза, аффинной хроматографии или другого способа очистки. По большей части, композиции, которые используют, будут содержать по меньшей мере 20% требуемого продукта по массе, обычно бо-

лее по меньшей мере около 75% по массе, предпочтительно, по меньшей мере около 95% по массе и, в терапевтических целях, как правило, по меньшей мере около 99,5% по массе, по отношению к загрязняющим веществам, ассоциированным со способом получения продукта и его очистки. Обычно процентное содержание будет основываться на общем количестве белка.

Чтобы индуцировать расщепление и рекомбинацию ДНК, или любую желаемую модификацию ДНК-мишени, или любую желаемую модификацию полипептида, ассоциированного с ДНК-мишенью, РНК, нацеленную на ДНК, и/или сайт-специфический модифицирующий полипептид и/или донорный полинуклеотид, независимо от того, будут ли они введены в качестве нуклеиновых кислот или полипептидов, предоставляют клеткам в течение от около 30 мин до около 24 ч, например 1, 1,5, 2, 2,5, 3, 3,5, 4, 5, 6, 7, 8, 12, 16, 18, 20 ч, или любого другого периода времени от около 30 мин до около 24 ч, которые могут быть повторены с частотой от около ежедневного до около каждые 4 дня, например каждые 1,5 дней, каждые 2 дня, каждые 3 дня или любая другая частота от около ежедневного до около каждые четыре дня. Средство(а) может быть предоставлено клеткам-мишеням один или несколько раз, например, один раз, дважды, три раза или более чем три раза, а клетки, возможно, инкубируют со средством(ами) в течение некоторого периода времени после каждого случая контактирования, например, 16-24 ч, после чего среду заменяют на свежую среду, а клетки культивируют дополнительно.

В тех случаях, в которых клетке предоставляют два или более различных комплекса нацеливания (например, две различные РНК, нацеленные на ДНК, которые являются комплементарными различным последовательностям в пределах одной и той же или различных ДНК-мишеней), комплексы могут быть предоставлены одновременно (например, в виде двух полипептидов и/или нуклеиновых кислот) или доставляются одновременно. Альтернативно, они могут быть предоставлены последовательно, например, комплекс нацеливания предоставляют первым, после чего, предоставляют второй комплекс нацеливания и т.д., или наоборот.

Как правило, эффективное количество РНК, нацеленной на ДНК, и/или сайт-специфического модифицирующего полипептида и/или донорного полинуклеотида предоставляют ДНК-мишени или клеткам для того, чтобы индуцировать расщепление. Эффективное количество РНК, нацеленной на ДНК, и/или сайт-специфического модифицирующего полипептида и/или донорного полинуклеотида представляет собой количество, необходимое для того, чтобы вызвать 2-кратное или более увеличение количества целевой модификации, наблюдаемое между двумя гомологичными последовательностями, по отношению к отрицательному контролю, например, клетке, которая контактирует с пустым вектором или нерелевантным полипептидом. То есть эффективное количество или доза РНК, нацеленной на ДНК, и/или сайт-специфического модифицирующего полипептида и/или донорного полинуклеотида будет индуцировать 2-кратное увеличение, 3-кратное увеличение, 4-кратное или более увеличение количества целевой модификации, наблюдаемой в области ДНК-мишени, в некоторых случаях, 5-кратное увеличение, 6-кратное или более увеличение, иногда в 7- или 8-кратное или более увеличение в размере наблюдаемой рекомбинации, например, увеличение в 10, 50 или 100 раз или более, в некоторых случаях, увеличение в 200, 500, 700 или 1000 раз или более, например 5000- или 10000-кратное увеличение в размере наблюдаемой рекомбинации. Количество целевой модификации может быть измерено любым пригодным способом. Например, конструкция молчащего репортера, содержащая последовательность, комплементарную нацеливающему сегменту (последовательность нацеливания) РНК, нацеленной на ДНК, окруженной повторяющимися последовательностями, которые при рекомбинировании будут воссоздавать нуклеиновую кислоту, кодирующую активный репортер, могут быть котрансфицированы в клетки, а количество белка репортера оценивают после контакта с РНК, нацеленной на ДНК, и/или сайт-специфическим модифицирующим полипептидом и/или донорным полинуклеотидом, например, через 2, 4, 8, 12, 24, 36, 48, 72 ч или более после контакта с РНК, нацеленной на ДНК, и/или сайт-специфическим модифицирующим полипептидом и/или донорным полинуклеотидом. В качестве отличного, более чувствительного анализа, например, степени рекомбинации в представляющей интерес области геномной ДНК, содержащей ДНК-последовательности-мишени, могут быть оценены с помощью ПЦР или Саузерн-гибридизации области после контактирования с РНК, нацеленной на ДНК, и/или сайт-специфическим модифицирующим полипептидом и/или донорным полинуклеотидом, например, через 2, 4, 8, 12, 24, 36, 48, 72 ч или более после контактирования с РНК, нацеленной на ДНК, и/или сайт-специфическим модифицирующим полипептидом и/или донорным полинуклеотидом.

Приведение в контакт клеток с РНК, нацеленной на ДНК, и/или сайт-специфическим модифицирующим полипептидом и/или донорным полинуклеотидом может происходить в любой культуральной среде и при любых условиях культивирования, которые способствуют выживанию клеток. Например, клетки могут быть суспендированы в любой подходящей питательной среде, которая является пригодной, например, модифицированная по способу Исков DMEM или RPMI 1640, дополненная фетальной телячьей сывороткой или инактивированная нагреванием козья сыворотка (приблизительно 5-10%), L-глутамин, тиол, особенно 2-меркаптоэтанол и антибиотики, например, пенициллин и стрептомицин. Культура может содержать факторы роста, к которым клетки являются чувствительными. Факторы роста, как определено в данном документе, представляют собой молекулы, способные стимулировать выживание, рост и/или дифференциацию клеток, либо в культуре, либо в интактной ткани, посредством спе-

цифических воздействий на трансмембранный рецептор. Факторы роста содержат полипептиды и отличные от полипептидов факторы. Условия, способствующие выживанию клеток, как правило, допускают негомологичное соединение концов и репарирование по механизму гомологичной рекомбинации.

В приложениях, в которых желательно встроить полинуклеотидную последовательность в последовательность ДНК-мишени, полинуклеотид, содержащий донорную последовательность, подлежащую вставке, также предоставляют клетке. Под "донорной последовательностью" или "донорным полинуклеотидом" подразумевают последовательность нуклеиновой кислоты, которая будет вставлена в сайт расщепления, индуцированного сайт-специфическим модифицирующим полипептидом. Донорный полинуклеотид содержит достаточную гомологию с геномной последовательностью в сайте расщепления, например, 70, 80, 85, 90, 95 или 100% гомологии с нуклеотидными последовательностями, фланкирующими сайт расщепления, например, в пределах около 50 или менее оснований сайта расщепления, например, в пределах около 30 оснований, в пределах около 15 оснований, в пределах около 10 оснований, в пределах около 5 оснований или фланкирующих непосредственно сайт расщепления для поддержки репарирования по механизму гомологичной рекомбинации между ним и геномной последовательностью, к которой обладает гомологией. Приблизительно 25, 50, 100 или 200 нуклеотидов или более 200 нуклеотидов гомологии последовательностей между донорной и геномной последовательностью (или любое целочисленное значение между 10 и 200 или более нуклеотидами) будут поддерживать репарирование по механизму гомологичной рекомбинации. Донорные последовательности могут быть любой длины, например, 10 нуклеотидов или более, 50 нуклеотидов или более, 100 нуклеотидов или более, 250 нуклеотидов или более, 500 нуклеотидов или более, 1000 или более нуклеотидов, 5000 нуклеотидов или более и т.д.

Донорная последовательность, как правило, не является идентичной геномной последовательностью, которую заменяет. Предпочтительно, донорная последовательность может содержать по меньшей мере одну или несколько точечных мутаций, вставок, делеций, инверсии или реаранжировок по отношению к геномной последовательности, при условии, что присутствует гомология, достаточная для поддержания репарирования по механизму гомологичной рекомбинации. В некоторых вариантах реализации настоящего изобретения донорная последовательность содержит негомологичную последовательность, фланкированную двумя областями гомологии, такими, что репарирование по механизму гомологичной рекомбинации между областью ДНК-мишени и двумя фланкирующими последовательностями приводит к вставке негомологичной последовательности в области-мишени. Донорные последовательности могут также содержать остов вектора, содержащий последовательности, которые являются негомологичными представляющими интерес областями ДНК и которые являются не предназначенными для вставки в представляющую интерес область ДНК. Как правило, гомологичная область(и) донорной последовательности будет иметь идентичность по меньшей мере 50% последовательности с геномной последовательностью, с которой рекомбинация является необходимой. В некоторых вариантах реализации настоящего изобретения присутствует идентичность 60, 70, 80, 90, 95, 98, 99 или 99,9% последовательности. В зависимости от длины донорного полинуклеотида, может присутствовать любое значение в пределах идентичности последовательности от 1 до 100%.

Донорная последовательность может содержать некоторые отличия в последовательности по сравнению с геномной последовательностью, например, сайты рестрикции, полиморфизмы нуклеотидов, селективные маркеры (например, гены устойчивости к лекарственным средствам, флуоресцентные белки, ферменты и т.д.), и т.д., которые могут быть использованы для оценки успешной вставки донорной последовательности в сайте расщепления или в некоторых случаях может быть использован для других целей (например, для обозначения экспрессии в геномном локусе-мишени). В некоторых случаях, если находятся в кодирующей области, подобные различия нуклеотидной последовательности не будут изменять аминокислотную последовательность или приводить к молчащим аминокислотным заменам (например, изменениям, которые не влияют на структуру или функцию белка). В качестве альтернативы, эти различия последовательностей могут содержать последовательности, фланкирующие рекомбинации, такие последовательности как FLP, IoxP или тому подобные, которые могут быть активированы в более позднее время для удаления последовательности маркера.

Донорная последовательность может быть предоставлена клетке в виде одноцепочечной ДНК, одноцепочечной РНК, двухцепочечной ДНК или двухцепочечной РНК. Она может быть введена в клетку в линейной или круглой форме. Если вводят в линейной форме, концы донорной последовательности могут быть защищены (например, от экзонуклеолитической деградации) с помощью способов, известных специалистам в данной области техники. Например, один или несколько дидезоксирибонуклеотидных остатков добавляют к 3'-концу линейной молекулы и/или самокомплементарных олигонуклеотидов лигируют к одному или обоим концам. Смотри, например, Chang и соавт. (1987) Proc. Natl. Acad. Sci. USA 84:4959-4963; Nehls и соавт. (1996) Science 272:886-889. Дополнительные способы защиты экзогенных полинуклеотидов от деградации, включают, но не ограничиваются ими, добавление концевой аминокислотной группы (аминогруппы) и использование модифицированных межнуклеотидных связей, таких как, например, фосфоротиоаты, фосфорамидаты и остатки O-метилрибозы или дезоксирибозы. В качестве альтернативы защите концов линейной донорной последовательности, последовательности, находящиеся вне областей

гомологии, могут быть удлинены, и эти последовательности могут быть разрушены, не влияя на рекомбинацию. Донорная последовательность может быть введена в клетку как часть молекулы вектора, которые имеют дополнительные последовательности, такие как, например, начало репликации, промоторы и гены, кодирующие устойчивость к антибиотикам. Более того, донорные последовательности могут быть введены в виде депротенизированной нуклеиновой кислоты, в виде нуклеиновой кислоты, образующей комплекс с таким средством, как липосома или поллоксамер, или могут быть доставлены с помощью вирусов (например, аденовирус, AAV), как описано выше в случае нуклеиновых кислот, кодирующих РНК, нацеленную на ДНК, и/или сайт-специфический модифицирующий полипептид и/или донорный полинуклеотид.

После описанных выше способов, представляющая интерес область ДНК может быть расщепленной и модифицированной, т.е. "генетически модифицированной", *ex vivo*. В некоторых вариантах реализации настоящего изобретения, например, при условии, что селектируемый маркер вставляют в представляющую интерес область ДНК, популяция клеток может быть обогащена теми, которые содержат генетическую модификацию путем отделения генетически модифицированных клеток от остальной популяции. Перед обогащением "генетически модифицированные" клетки могут составлять лишь около 1% клеточной популяции или более (например, 2% или более, 3% или более, 4% или более, 5% или более, 6% или более, 7% или более, 8% или более, 9% или более, 10% или более, 15% или более или 20% или более). Отделение "генетически модифицированных" клеток может быть достигнуто любым удобным способом разделения, подходящим для используемого селективного маркера. Например, если встраивают флуоресцентный маркер, клетки могут быть отделены с помощью активируемой флуоресценцией сортировки клетки, а если был, встраивают поверхностный маркер, клетки могут быть отделены от гетерогенной популяции с помощью способов аффинного разделения, например, магнитного разделения, аффинной хроматографии, "пэннинга" с аффинным реагентом, прикрепленным к твердой матрице, или другим удобным способом. Способы, обеспечивающие точное разделение, содержат активируемые флуоресценцией сортировщики клеток, которые могут иметь различную степень сложности, например, несколько цветовых каналов, каналы обнаружения рассеяния света на малых и больших углах, каналы сопротивления и т.д. Эти клетки могут быть выбраны против мертвых клеток с использованием красителей, ассоциированных с мертвыми клетками (например, пропидий йодид). Может быть использован любой способ, который не является чрезмерно вредным для жизнеспособности генетически модифицированных клеток. Таким способом достигают клеточные композиции, которые являются высокообогащенными клетками, содержащими модифицированную ДНК. Под "высокообогащенными" подразумевает то, что генетически модифицированные клетки будут составлять 70% или более, 75% или более, 80% или более, 135% или более, 90% или более клеточной композиции, например, около 95% или более, или 98% или более клеточной композиции. Другими словами, композиция может быть по существу чистой композицией генетически модифицированных клеток.

Генетически модифицированные клетки, полученные с помощью способов, описанных в данном документе, могут быть использованы без промедления. Альтернативно, клетки могут быть заморожены при температуре жидкого азота и храниться в течение длительных периодов времени, затем могут быть разморожены и способны быть использованными повторно. В таких случаях, клетки будут заморожены в 10% диметилсульфоксида (DMSO), 50% сыворотки, 40% буферной среды или некотором отличном подобном растворе, который обычно используют в данной области техники для того, чтобы сохранить клетки при подобных низких температурах, и размораживают широкоизвестным в данной области техники способом для размораживания замороженных культивируемых клеток.

Генетически модифицированные клетки могут быть культивированы в пробирке при различных условиях культивирования. Эти клетки могут расти в культуре, т.е. быть выращенными в условиях, обеспечивающих их пролиферацию. Культуральная среда может быть жидкой или полужидкой, например, содержащей агар, метилцеллюлозу и т.д.

Клеточная популяция может быть суспендирована в соответствующей питательной среде, такой как модифицированная по способу Исков DMEM или RPMI 1640, обычно с добавлением фетальной телячьей сыворотки (около 5-10%), L-глутамин, тиола, особенно 2-меркаптоэтанол и антибиотиков, например, пенициллина и стрептомицина. Культура может содержать факторы роста, к которым регуляторные Т-лимфоциты являются чувствительными. Факторы роста, как определено в данном документе, представляют собой молекулы, способные стимулировать выживание, рост и/или дифференциацию клеток, либо в культуре, либо в интактной ткани, посредством специфических воздействий на трансмембранный рецептор. Факторы роста содержат полипептиды и отличные от полипептида факторы.

Клетки, которые были генетически модифицированы таким образом, могут быть пересажены пациенту для таких целей, как генная терапия, например, для лечения заболевания или в качестве противовирусного, противопатогенного или противоракового терапевтического средства, для производства генетически модифицированных организмов в сельском хозяйстве или для биологических исследований. Субъект может быть новорожденным, молодой особью или взрослым. Особый интерес представляют собой субъекты-млекопитающие. Некоторые виды млекопитающих, которые можно лечить настоящими способами, представляют собой собак и кошек; животные семейства лошадиных; крупный рогатый скот; коз; и

т.д., и приматов, в особенности людей. В случае экспериментальных исследований могут быть использованы животные модели, в частности, мелкие млекопитающие (например, мышь, крыса, морская свинка, хомяк, зайцеобразные (например, кролик) и т.д.).

Клетки могут быть предоставлены субъекту отдельно или вместе с соответствующей подложкой или матрицей, например, для того, чтобы поддержать их рост и/или организацию в ткани, на которую их пересаживают. Как правило, может быть введено по меньшей мере  $1 \times 10^3$  клеток, например  $5 \times 10^3$  клеток,  $1 \times 10^4$  клеток,  $5 \times 10^4$  клеток,  $1 \times 10^5$  клеток,  $1 \times 10^6$  клеток или более. Клетки могут быть введены субъекту посредством любого из следующих путей: парентерального, подкожного, внутривенного, внутривенного, интраспинального, внутриглазного или в спинномозговую жидкость. Клетки могут быть введены с помощью инъекции, катетера или тому подобного. Примеры способов целенаправленной доставки, т.е. доставки в месте повреждения, содержат, например, способ посредством резервуар Оммайя, например, в случае интратекальной поставки (см., например, патенты США № 5222982 и 5385582, которые содержатся в данном документе в виде ссылки); с помощью болюсного введения, например, с помощью шприца, например, в сустав; с помощью непрерывной инфузии, например, канюли, например, с конвекцией (см., например, заявку на патент США № 20070254842, которая содержится в данном документе в виде ссылки.); или с помощью имплантации устройства, на котором были реверсивно прикреплены клетки (см., например, заявку на патент США № 20080081064 и 20090196903, которые содержатся в настоящем документе в виде ссылки). Клетки также могут быть введены в эмбрион (например, бластоцист) с целью создания трансгенных животных (например, трансгенной мыши).

Количество введений субъекту данного лечения может изменяться. Введение генетически модифицированных клеток субъекту может быть разовым событием; но в некоторых ситуациях, например, в ходе лечения, может вызвать улучшение в течение ограниченного периода времени и требует продолжающейся серии многократных введений. В других ситуациях, многократные введения генетически модифицированных клеток могут потребоваться перед тем, как наблюдается эффект. Точные протоколы зависят от заболевания или состояния, стадии заболевания и параметров отдельного субъекта, подлежащего лечению.

В других аспектах настоящего изобретения, РНК, нацеленная на ДНК, и/или сайт-специфический модифицирующий полипептид и/или донорный полинуклеотид используют для модификации клеточной ДНК *in vivo*, снова для таких целей, как генная терапия, например, для лечения заболевания или в качестве противовирусного, противопатогенного или противоракового терапевтического средства, для производства генетически модифицированных организмов в сельском хозяйстве или в случае биологических исследований. В этих *in vivo* вариантах реализации настоящего изобретения РНК, нацеленная на ДНК, и/или сайт-специфический модифицирующий полипептид и/или донорный полинуклеотид вводят непосредственно индивидууму. РНК, нацеленную на ДНК, и/или сайт-специфический модифицирующий полипептид и/или донорный полинуклеотид можно вводить с помощью любого из нескольких хорошо известных в данной области техники способов введения субъекту пептидов, небольших молекул и нуклеиновых кислот. РНК, нацеленная на ДНК, и/или сайт-специфический модифицирующий полипептид и/или донорный полинуклеотид могут содержаться в виде различных составов. В частности, РНК, нацеленная на ДНК, и/или сайт-специфический модифицирующий полипептид и/или донорный полинуклеотид по настоящему изобретению могут быть приготовлены в виде фармацевтических композиций с помощью комбинации подходящими фармацевтически приемлемыми носителями или разбавителями.

Фармацевтические препараты представляют собой композиции, которые содержат одну или более РНК, нацеленную на ДНК, и/или сайт-специфический модифицирующий полипептид и/или донорный полинуклеотид, присутствующий в фармацевтически приемлемом носителе. "Фармацевтически приемлемые основы" могут быть основами, утвержденными регулирующим органом федерального правительства или правительства штата или перечисленными в фармакопее США или другой общепризнанной фармакопее для применения в случае млекопитающих, таких как человек. Термин "основа" означает разбавитель, адъювант, наполнитель или носитель, с которым соединение по настоящему изобретению готовят для введения млекопитающему. Такие фармацевтические основы могут быть липидами, например, липосомами, например, липосомными дендримерами; жидкостями, такими как вода и масла, в том числе получаемыми из нефти, животного, растительного или синтетического происхождения, такими как арахисовое масло, соевое масло, минеральное масло, кунжутное масло и подобные, физиологический раствор; аравийская камедь, желатин, крахмальная паста, тальк, кератин, коллоидный диоксид кремния, мочевины и подобные. Кроме того, могут быть использованы вспомогательные, стабилизирующие, загущающие, смазывающие и красящие средства. Фармацевтические композиции могут быть приготовлены в виде препаратов в твердой, полужидкой, жидкой или газообразной формах, таких как таблетки, капсулы, порошки, гранулы, мази, растворы, суппозитории, инъекции, ингаляции, гели, микросферы и аэрозоли. Таким образом, введение РНК, нацеленной на ДНК, и/или сайт-специфического модифицирующего полипептида и/или донорного полинуклеотида может быть достигнуто различными способами, в том числе оральным, буккальным, ректальным, парентеральным, внутрибрюшинным, внутрикожным, трансдермальным, интратрахиальным, внутриглазным и т.д. введением. Действующее вещество может быть

системным после введения или может быть локализовано с помощью применения местного введения, интрамурального введения или использования имплантата, который действует с той целью, чтобы сохранить активную дозу в месте имплантации. Действующее вещество может быть приготовлено для немедленной активности или оно может быть приготовлено для замедленного высвобождения.

В случае некоторых условиях, особенно условий центральной нервной системы, это может быть необходимо для того, чтобы приготовить средства для преодоления гематоэнцефалического барьера (БВВ). Одна из стратегий доставки лекарственных средств через гематоэнцефалический барьер (БВВ) влечет за собой разрушение БВВ, либо осмотическими средствами, такими как маннит или лейкотриены, либо биохимически с использованием вазоактивных веществ, таких как брадикинин. Потенциал использования открытия БВВ для нацеливания специфических средств на опухоли головного мозга также является возможным. Разрушающее БВВ средство может быть введено совместно с терапевтическими композициями по настоящему изобретению, при условии, что композиции вводят с; помощью внутрисосудистой инъекции. Другие стратегии прохождения сквозь БВВ могут повлечь за собой использование эндогенных транспортных систем, в том числе опосредованного caveолином-1 трансцитоза, опосредованных носителей переносчиков, таких как глюкоза и аминокислотные носители, опосредованного рецептором трансцитоза в случае инсулина или трансферрина и активных эффлюксных переносчиков, таких как р-гликопротеин. Фрагменты активного транспорта могут быть также конъюгированы с терапевтическими соединениями для использования по настоящему изобретению, чтобы облегчить транспорт сквозь эндотелиальную стенку кровеносного сосуда. Кроме того, доставка терапевтических средств за БВВ может быть осуществлена с помощью целенаправленной доставки, например, с помощью интратекальной доставки, например, посредством резервуара Оммайя (см., например, патенты США № 5222982 и 5385582, которые содержатся в данном документе в виде ссылки); с помощью болюсной инъекции, например, с помощью шприца, например, интравитреально или интракраниально; с помощью непрерывной инфузии, например, канюли, например, с конвекцией (см., например, заявку на патент США № 20070254842, которая содержится в данном документе в виде ссылки); или с помощью устройства имплантации, на котором было реверсивно прикреплено средство (см., например, заявки на патент США № 20080081064 и 20090196903, которые содержатся в настоящем документе в виде ссылки).

Как правило, обеспечивают эффективное количество РНК, нацеленной на ДНК, и/или сайт-специфического модифицирующего полипептида и/или донорного полинуклеотида. Как обсуждалось выше по отношению к способам *ex vivo*, эффективное количество или эффективная доза РНК, нацеленной на ДНК, и/или сайт-специфического модифицирующего полипептида и/или донорного полинуклеотида *in vivo* представляет собой количество, необходимое для того, чтобы вызвать 2-кратное или более увеличение количества рекомбинации, наблюдаемой между двумя гомологичными последовательностями по отношению к отрицательному контролю, например, клетке, которая контактирует с пустым вектором или нерелевантным полипептидом. Количество рекомбинации может быть измерено любым удобным способом, например, как описано выше и известно в данной области техники. Расчет эффективного количества или эффективной дозы вводимой РНК, нацеленной на ДНК, и/или сайт-специфического модифицирующего полипептида и/или донорного полинуклеотида находится в пределах квалификации обычного специалиста в данной области техники, и будет обычной для специалистов в данной области техники. Конечное вводимое количество будет зависеть от пути введения и от природы расстройства или состояния, которое должно быть вылечено.

Эффективное количество, которое дают конкретному пациенту, будет зависеть от множества факторов, некоторые из которых будут отличаться от пациента к пациенту. Компетентный врач сможет определить эффективное количество терапевтического средства для введения пациенту с той целью, чтобы остановить или обратить прогрессирование болезненного состояния настолько, насколько это требуется. Используя данные о LD50 у животных и другую имеющуюся информацию о средстве, врач может определить максимальную безопасную дозу для человека в зависимости от способа введения. Например, внутривенно вводимая доза может быть больше, чем интратекально вводимая доза, с учетом большего объема жидкости, в которое вводят терапевтическую композицию. Аналогичным образом, композиции, которые быстро выводят из организма, могут быть введены в более высоких дозах или в многократных дозах, для того, чтобы поддерживать терапевтическую концентрацию. Используя обычную квалификацию, компетентный врач сможет оптимизировать дозировку конкретного терапевтического средства в ходе обычных клинических испытаний.

Для включения в лекарственное средство, РНК, нацеленная на ДНК, и/или сайт-специфический модифицирующий полипептид и/или донорный полинуклеотид могут быть получены из соответствующего коммерческого источника. В качестве общего предложения, общее фармацевтически эффективное количество РНК, нацеленной на ДНК, и/или сайт-специфического модифицирующего полипептида и/или донорного полинуклеотида вводят парентерально в дозе, лежащей в диапазоне, который может быть измерен с помощью ривой дозовой зависимости.

Медикаменты основанные на РНК, нацеленной на ДНК, и/или сайт-специфическом модифицирующем полипептиде и/или донорном полинуклеотиде, т.е. препараты РНК, нацеленной на ДНК, и/или сайт-специфического модифицирующего полипептида и/или донорного полинуклеотида, используемые для

терапевтического введения, должны быть стерильными. Стерильность легко достигается с помощью фильтрации через стерильные фильтрационные мембраны (например, мембраны 0,2 мкм). Терапевтические композиции обычно помещают в контейнер, имеющий стерильное входное отверстие, например, пакет для внутривенного раствора или флакон, имеющий пробку, прокалываемую иглой для подкожной инъекции. Медикаменты, основанные на РНК, нацеленной на ДНК, и/или сайт-специфическом модифицирующем полипептиде и/или донорном полинуклеотиде, могут храниться в виде однодозных или многодозных контейнерах, например, запечатанных ампулах или флаконах, в виде водного раствора или в виде лиофилизированного препарата для растворения. В качестве примера лиофилизированной композиции, 10-мл ампулы заполняют 5 мл стерилизованного фильтрацией 1% (масса/объем) водного раствора соединения, а полученную смесь лиофилизируют. Инфузионный раствор готовят путем ресуспендирования лиофилизированного соединения с помощью бактериостатической воды для инъекций.

Фармацевтические композиции могут содержать, в зависимости от желаемого препарата, фармацевтически приемлемые нетоксичные наполнители или разбавители, которые относят к носителям или обычно используемым для приготовления фармацевтических композиций для введения животным или человеку. Разбавитель выбирают так, чтобы не влиять на биологическую активность комбинации. Примерами таких разбавителей являются дистиллированная вода, забуференная вода, физиологический раствор, PBS, раствор Рингера, раствор декстрозы и раствор Хэнкса. Кроме того, фармацевтическая композиция или композиция может содержать другие носители, адъюванты или нетоксичные, нетерапевтические, неиммуногенные стабилизаторы, наполнители и подобные. Композиции могут также содержать дополнительные вещества для приближения к физиологическим условиям, таким как доведение pH, и буферные средства, средства, регулирующие токсичность, смачивающие агенты и детергенты.

Композиция также может содержать любой из множества стабилизирующих средств, таких как антиоксидант, например. При условии, что фармацевтическая композиция содержит полипептид, полипептид может быть образован в виде комплекса с различными хорошо известными соединениями, которые повышают *in vivo* стабильность полипептида или, иным образом, повышают его фармакологические свойства (например, увеличение времени полужизни полипептида, уменьшение его токсичности, повышение растворимости или поглощения). Примеры таких модификаций или комплексообразующих средств содержат сульфат, глюконат, цитрат и фосфат. Нуклеиновые кислоты или полипептиды композиции также могут образовывать комплексы с молекулами, которые повышают их характеристики *in vivo*. Такие молекулы содержат, например, углеводы, полиамины, аминокислоты, отличные пептиды, ионы (например, натрий, калий, кальций, магний, марганец) и липиды.

Дальнейшее руководство по отношению к составам, которые являются подходящими для различных типов введения, может быть найдены в Remington's Pharmaceutical Sciences, Mace Publishing Company, Philadelphia, Pa., 17th ed. (1985). Для краткого обзора способов доставки лекарственных средств, см. Langer, Science 249:1527-1533 (1990).

Фармацевтические композиции могут быть введены для профилактического и/или терапевтического лечения. Токсичность и терапевтическая эффективность активных ингредиентов могут быть определены в соответствии со стандартными фармацевтическими процедурами в клеточных культурах или на лабораторных животных, в том числе, но без ограничения ими, определение LD50 (дозы, летальной для 50% популяции) и ED50 (дозы, терапевтически эффективной для 50% популяции). Соотношение доз между токсичным и терапевтическим эффектом представляет собой терапевтический индекс, который выражают в виде отношения LD50/ED50. Предпочтительными являются способы лечения, которые демонстрируют большие терапевтические индексы.

Данные, полученные от клеточной культуры и/или в исследованиях на животных, могут быть использованы при определении диапазона доз для человека. Дозировка активного ингредиента обычно лежит в пределах диапазона циркулирующих концентраций, которые содержат ED50 с низкой токсичностью. Дозировка может изменяться в пределах этого диапазона в зависимости от используемой лекарственной формы и используемого способа введения.

Компоненты, используемые в приготовлении фармацевтических композиций, предпочтительно имеют высокую чистоту и, по существу, не содержат потенциально вредных примесей (например, по меньшей мере, класс National Food (NF), как правило, по меньшей мере, аналитической степени чистоты, и как правило, по меньшей мере, фармацевтического качества). Кроме того, композиции, предназначенные для использования *in vivo*, как правило, стерильны. По мере того, как данное соединение должно быть синтезировано перед использованием, полученный продукт, как правило, практически свободен от любых потенциально токсичных средств, особенно любых эндотоксинов, которые могут присутствовать в процессе синтеза или очистки. Композиции для парентерального введения являются также стерильными, изотоническими и, по существу, сделаны в условиях GMP.

Эффективное количество терапевтической композиции, данное конкретному пациенту, будет зависеть от множества факторов, некоторые из которых будут отличаться от пациента к пациенту. Компетентный врач сможет определить эффективное количество терапевтического средства для введения пациенту с той целью, чтобы остановить или обратить прогрессирование болезненного состояния настолько, насколько это требуется. Используя данные о LD50 у животных и другую имеющуюся информацию о

средстве, врач может определить максимальную безопасную дозу для человека в зависимости от способа введения. Например, внутривенно вводимая доза может быть больше, чем интратекально вводимая доза, с учетом большего объема жидкости, в которое вводят терапевтическую композицию. Аналогичным образом, композиции, которые быстро выводят из организма, могут быть введены в более высоких дозах или в многократных дозах, для того, чтобы поддерживать терапевтическую концентрацию. Используя обычную квалификацию, компетентный врач сможет оптимизировать дозировку конкретного терапевтического средства в ходе обычных клинических испытаний.

Генетически модифицированные клетки-хозяева.

Настоящее изобретение обеспечивает генетически модифицированные клетки-хозяева, в том числе выделенные генетически модифицированные клетки-хозяева, причем генетически модифицированная клетка-хозяин содержит (была генетически модифицирована с: 1) экзогенной РНК, нацеленной на ДНК; 2) экзогенной нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую РНК, нацеленную на ДНК; 3) экзогенным сайт-специфическим модифицирующим полипептидом (например, встречающийся в природе Cas9; модифицированным, т.е. мутировавшим или вариантным, Cas9; гибридным Cas9; и т.д.); 4) экзогенной нуклеиновой кислотой, содержащей нуклеотидную последовательность, кодирующую сайт-специфический модифицирующий полипептид; или 5) любой комбинацией приведенных выше. Генетически модифицированную клетку-мишень получают с помощью генетической модификации клетки-хозяина, например: 1) экзогенной РНК, нацеленной на ДНК; 2) экзогенной нуклеиновой кислотой, содержащей нуклеотидную последовательность, кодирующую РНК, нацеленную на ДНК; 3) экзогенным сайт-специфическим модифицирующим полипептидом; 4) экзогенной нуклеиновой кислотой, содержащей нуклеотидную последовательность, кодирующую сайт-специфический модифицирующий полипептид; или 5) любой комбинацией приведенных выше).

Все клетки, пригодные для того, чтобы быть клеткой-мишенью, также являются пригодными для того, чтобы быть генетически модифицированной клеткой-хозяином. Например, представляющей интерес генетически модифицированной клеткой-хозяином может быть клетка любого организма (например, бактериальная клетка, клеткой архей, клетка одноклеточного эукариотического организма, клетка растения, клетка водорослей, например, *Botryococcus braunii*, *Chlamydomonas reinhardtii*, *Nannochloropsis gaditana*, *Chlorella pyrenoidosa*, *Sargassum patens* С Agardh и подобных, клетка грибка (например, дрожжевой клеткой), клетка животного, клетка беспозвоночного животного (например, плодовой мушки, кишечнополостной, иглокожего, нематоды, и т.д.), клетка позвоночного животного (например, рыбы, амфибии, рептилии, птицы, млекопитающего), клетка млекопитающего (например, свиньи, коровы, козы, овцы, грызуна, крысы, мыши, не являющегося человеком примата, человека и т.д.), и т.д.

В некоторых вариантах реализации настоящего изобретения генетически модифицированная клетка-хозяин была генетически модифицирована экзогенной нуклеиновой кислотой, содержащей нуклеотидную последовательность, кодирующую сайт-специфический модифицирующий полипептид (например, встречающийся в природе Cas9; модифицированный, т.е. мутировавший или вариантный, Cas9; гибридный Cas9; и т.д.). Для модификаций ДНК генетически модифицированной клетки-хозяина на нее можно направленно воздействовать с помощью введения в клетку РНК, нацеленной на ДНК, (или ДНК, кодирующую РНК, нацеленную на ДНК, которая определяет геномную локализацию/последовательность, будет модифицированной) и необязательно донорной нуклеиновой кислотой, В некоторых вариантах реализации настоящего изобретения нуклеотидная последовательность, кодирующая сайт-специфический модифицирующий полипептид, функционально связана с индуцируемым промотором (например, промотором теплового шока, тетрациклин-регулируемым промотором, стероид-регулируемым промотором, металл-регулируемым промотором, регулируемым рецептором эстрогена промотором и т.д.). В некоторых вариантах реализации настоящего изобретения нуклеотидная последовательность, кодирующая сайт-специфический модифицирующий полипептид, функционально связана с пространственно ограниченным и/или временно ограниченным промотором (например, тканеспецифическим промотором, специфическим к типу клеток промотором и т.д.). В некоторых вариантах реализации настоящего изобретения нуклеотидная последовательность, кодирующая сайт-специфический модифицирующий полипептид, функционально связана с конститутивным промотором.

В некоторых вариантах реализации настоящего изобретения генетически модифицированная клетка-хозяин по изобретению находится *in vitro*. В некоторых вариантах реализации настоящего изобретения генетически модифицированная клетка-хозяин по изобретению находится *in vivo*. В некоторых вариантах реализации настоящего изобретения генетически модифицированная клетка-хозяин по изобретению представляет собой прокариотическую клетку или получена из прокариотической клетки. В некоторых вариантах реализации настоящего изобретения генетически модифицированная клетка-хозяин по изобретению представляет собой бактериальную клетку или получена из бактериальной клетки. В некоторых вариантах реализации настоящего изобретения генетически модифицированная клетка-хозяин по изобретению представляет собой клетку архей или получена из клетки архей. В некоторых вариантах реализации настоящего изобретения генетически модифицированная клетка-хозяин по изобретению представляет собой эукариотическую клетку или получена из эукариотической клетки. В некоторых вариантах реализации настоящего изобретения генетически модифицированная клетка-хозяин представляет собой клетку рас-

тения или получена из клетки растения. В некоторых вариантах реализации настоящего изобретения генетически модифицированная клетка-хозяин представляет собой клетку животного или получена из клетки животного. В некоторых вариантах реализации настоящего изобретения генетически модифицированная клетка-хозяин представляет собой клетку беспозвоночного или получена из клетки беспозвоночного. В некоторых вариантах реализации настоящего изобретения генетически модифицированная клетка-хозяин представляет собой клетку позвоночного или получена из клетки позвоночного. В некоторых вариантах реализации настоящего изобретения генетически модифицированная клетка-хозяин представляет собой клетку млекопитающего или получена из клетки млекопитающего. В некоторых вариантах реализации настоящего изобретения генетически модифицированная клетка-хозяин представляет собой клетку грызуна или получена из клетки грызуна. В некоторых вариантах реализации настоящего изобретения генетически модифицированная клетка-хозяин представляет собой клетку человека или получена из клетки человека.

Настоящее изобретение дополнительно предлагает потомство генетически модифицированной клетки по изобретению, причем потомство может содержать ту же экзогенную нуклеиновую кислоту или полипептид как и генетически модифицированная клетка по изобретению, из которой они были получены. Настоящее изобретение дополнительно обеспечивает композицию, содержащую целевую генетически модифицированную клетку-хозяина.

Генетически модифицированные стволовые клетки и генетически модифицированные клетки-предшественники.

В некоторых вариантах осуществления, целевая генетически модифицированная клетка-хозяин представляет собой генетически модифицированную стволовую клетку или клетку-предшественник. Подходящие клетки-хозяева содержат, например, стволовые клетки (стволовые клетки взрослых, эмбриональные стволовые клетки, iPS клеток и т.д.) и клетки-предшественника (например, кардиальные клетки-предшественники, нейронные клетки-предшественники и т.д.). Подходящие клетки-хозяева содержат стволовые клетки млекопитающих и клетки-предшественники, в том числе, например, стволовые клетки грызуна, клетки-предшественники грызунов, стволовые клетки человека, клетки-предшественники человека и т.д. Подходящие клетки-хозяева содержат *in vitro* клетки-хозяева, например, выделенные клетки-хозяева.

В некоторых вариантах реализации настоящего изобретения генетически модифицированная клетка-хозяин содержит экзогенную нуклеиновую кислоту РНК, нацеленной на ДНК. В некоторых вариантах реализации настоящего изобретения генетически модифицированная клетка-хозяин содержит экзогенную нуклеиновую кислоту, содержащую нуклеотидную последовательность, кодирующую РНК, нацеленную на ДНК. В некоторых вариантах реализации настоящего изобретения генетически модифицированная клетка-хозяин содержит экзогенный сайт-специфический модифицирующий полипептид (например, встречающийся в природе Cas9; модифицированный, т.е. мутировавший или вариантный, Cas9; гибридный Cas9; и т.д.). В некоторых вариантах реализации настоящего изобретения генетически модифицированная клетка-хозяин содержит экзогенную нуклеиновую кислоту, содержащую нуклеотидную последовательность, кодирующую сайт-специфический модифицирующий полипептид. В некоторых вариантах реализации настоящего изобретения генетически модифицированная клетка-хозяин содержит экзогенную нуклеиновую кислоту, содержащую нуклеотидную последовательность, кодирующую 1) РНК, нацеленную на ДНК, и 2) сайт-специфический модифицирующий полипептид.

В некоторых случаях сайт-специфический модифицирующий полипептид содержит аминокислотную последовательность, имеющую идентичность по меньшей мере около 75%, по меньшей мере около 80%, по меньшей мере около 85%, по меньшей мере около 90%, по меньшей мере около 95%, по меньшей мере около 99 или 100% аминокислотной последовательности по отношению к аминокислотам 7-166 или 731-1003 аминокислотной последовательности Cas9/Csn1, изображенной на фиг. 3, или по отношению к соответствующим участкам любой из аминокислотных последовательностей, приведенных в SEQ ID NO:1-256 и 795-1346.

Композиции.

Настоящее изобретение обеспечивает композицию, содержащую РНК, нацеленную на ДНК, и/или сайт-специфический модифицирующий полипептид. В некоторых случаях сайт-специфический модифицирующий полипептид представляет собой гибридный полипептид-мишень. Композиция по изобретению пригодна для осуществления способа по настоящему изобретению, например, способа сайт-специфической модификации ДНК-мишени; способа сайт-специфической модификации полипептида, ассоциированного с ДНК-мишенью; и т.д.

Композиции, содержащие РНК, нацеленную на ДНК.

Настоящее изобретение обеспечивает композицию, содержащую РНК-нацеленную на ДНК. Композиция может содержать, в дополнение к РНК, нацеленной на ДНК, одно или несколько из: солей, например, NaCl, MgCl<sub>2</sub>, KCl, MgSO<sub>4</sub> и т.д.; буферного средства, например трис-буфера, N-(2-гидроксиэтил)пиперазин-S'-(2-этансульфонової кислоты) (HEPES), 2-(N-морфолино)этансульфонової кислоты (MES), натриевой соли MES, 3-(N-морфолино)пропансульфонової кислоты (MOPS), N-трис[гидроксиэтил]метил-3-аминопропансульфонової кислоты (TAPS) и т.д.; солубилизирующего

средства; детергента, например, неионогенного детергента, такого как Tween-20, и т.д.; ингибитора нуклеазы; и подобных. Например, в некоторых случаях композиция по изобретению содержит РНК-мишень, нацеленную на ДНК, и буфер для стабилизации нуклеиновых кислот.

В некоторых вариантах реализации настоящего изобретения РНК, нацеленная на ДНК, присутствующая в композиции-мишени, является чистой, например, по меньшей мере около 75%, по меньшей мере около 80%, по меньшей мере около 85%, по меньшей мере около 90%, по меньшей мере около 95%, по меньшей мере около 98%, по меньшей мере около 99% или более 99% чистоты, где "% чистота" означает, что РНК, нацеленная на ДНК, на указанный процент свободна от других макромолекул или загрязняющих веществ, которые могут присутствовать в процессе получения РНК, нацеленной на ДНК.

Композиции, содержащие гибридный полипептид-мишень.

Настоящее изобретение обеспечивает композицию, содержащую гибридный полипептид-мишень. Композиция может содержать, в дополнение к гибриднему полипептиду-мишени, одно или более из солей, например, NaCl, MgCl<sub>2</sub>, KCl, MgSO<sub>4</sub> и т.д.; буферного средства, например, трис-буфера, HEPES, MES, натриевой соли MES, MOPS, TAPS и т.д.; солюбилизующего средства; детергента, например, неионогенного детергента, такого как Tween-20, и т.д.; ингибитора протеазы; восстанавливающего средства (например, дитиотреитола); и подобных.

В некоторых вариантах реализации настоящего изобретения гибридный полипептид-мишень, присутствующая в композиции-мишени, является чистым, например, с по меньшей мере около 75%, по меньшей мере около 80%, по меньшей мере около 85%, по меньшей мере около 90%, по меньшей мере около 95%, по меньшей мере около 98%, по меньшей мере около 99%, или более 99% чистоты, где "% чистота" означает то, что сайт-специфический модифицирующий полипептид на указанный процент свободен от других белков, отличных макромолекул или загрязняющих веществ, которые могут присутствовать в процессе получения гибридного полипептида.

Композиции содержащие РНК, нацеленную на ДНК, и/или сайт-специфический модифицирующий полипептид.

Настоящее изобретение относится к композиции, содержащей (i) РНК, нацеленную на ДНК, или полинуклеотид ДНК, кодирующий ее; и (ii) сайт-специфический модифицирующий полипептид или полинуклеотид, кодирующий его. В некоторых случаях сайт-специфический модифицирующий полипептид представляет собой гибридный сайт-специфический модифицирующий полипептид-мишень. В других случаях, сайт-специфический модифицирующий полипептид представляет собой встречающийся в природе сайт-специфический модифицирующий полипептид. В некоторых случаях сайт-специфический модифицирующий полипептид проявляет ферментативную активность, которая модифицирует ДНК-мишень. В других случаях, сайт-специфический модифицирующий полипептид проявляет ферментативную активность, которая модифицирует полипептид, ассоциированный с ДНК-мишенью. В еще одних случаях, сайт-специфический модифицирующий полипептид модулирует транскрипцию ДНК-мишени.

Настоящее изобретение обеспечивает композицию, содержащую: (i) РНК, нацеленную на ДНК, как описано выше, или полинуклеотид ДНК, кодирующий ее, где РНК, нацеленная на ДНК, содержит (а) первый сегмент, содержащий нуклеотидную последовательность, комплементарную последовательности в ДНК-мишени; и (b) второй сегмент, который взаимодействует с сайт-специфическим модифицирующим полипептидом; и (ii) сайт-специфический модифицирующий полипептид или полинуклеотид, кодирующий его, где сайт-специфический модифицирующий полипептид содержит (а) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и (b) участок активности, который проявляет сайт-специфическую ферментативную активность, причем сайт ферментативной активности определяется РНК, нацеленной на ДНК.

В некоторых случаях композиция по изобретению содержит: композицию, содержащую: (i) РНК-мишень, нацеленную на ДНК, где РНК, нацеленная на ДНК, содержит (а) первый сегмент, содержащий нуклеотидную последовательность, комплементарную последовательности в ДНК-мишени; и (b) второй сегмент, который взаимодействует с сайт-специфическим модифицирующим полипептидом; и (ii) сайт-специфический модифицирующий полипептид, где сайт-специфический модифицирующий полипептид содержит (а) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и (b) участок активности, который проявляет сайт-специфическую ферментативную активность, причем сайт ферментативной активности определяется РНК, нацеленной на ДНК.

В отличных вариантах реализации настоящего изобретения композиция по изобретению содержит (i) полинуклеотид, кодирующий РНК, нацеленную на ДНК, где РНК, нацеленная на ДНК, содержит (а) первый сегмент, содержащий нуклеотидную последовательность, комплементарную последовательности в ДНК-мишени; и (b) второй сегмент, который взаимодействует с сайт-специфическим модифицирующим полипептидом; и (ii) полипептид, кодирующий сайт-специфический модифицирующий полипептид, где сайт-специфический модифицирующий полипептид содержит (а) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и (b) участок активности, который проявляет сайт-специфическую ферментативную активность, причем сайт ферментативной активности определяется РНК, нацеленной на ДНК.

В некоторых вариантах реализации настоящего изобретения композиция по изобретению содержит

обе молекулы РНК, нацеленной на ДНК, в виде двойной молекулы. Таким образом, в некоторых вариантах реализации настоящего изобретения композиция-мишень содержит РНК-активатор, которая содержит дуплекс-образующий сегмент, комплементарный дуплекс-образующему сегменту РНК-нацеливающей (см. фиг. 1А). Дуплекс-образующие сегменты РНК-активатора и РНК-нацеливающей гибридизируются с образованием дуплекса дцРНК белок-связывающего сегмента РНК, нацеленной на ДНК. РНК-нацеливающая дополнительно обеспечивает ДНК-нацеленный сегмент (одноцепочечный) РНК, нацеленной на ДНК, и, следовательно, направляет РНК, нацеленную на ДНК, к специфической последовательности в пределах ДНК-мишени. В качестве неограничивающего примера, дуплекс-образующий сегмент РНК-активатора содержит нуклеотидную последовательность с идентичностью по меньшей мере около 70%, по меньшей мере около 80%, по меньшей мере около 90%, по меньшей мере около 95%, по меньшей мере около 98% или 100% с последовательностью 5'-UAGCAAGUUA AAAAU-3' (SEQ ID NO:562). В качестве другого неограничивающего примера, дуплекс-образующий сегмент РНК-активатора содержит нуклеотидную последовательность с идентичностью по меньшей мере около 70%, по меньшей мере около 80%, по меньшей мере около 90%, по меньшей мере около 95%, по меньшей мере около 98 или 100% с последовательностью 5'-GUUUUAGAGCUA-3' (SEQ ID NO:679).

Настоящее изобретение обеспечивает композицию, содержащую: (i) РНК, нацеленную на ДНК, или полинуклеотид ДНК, кодирующий ее, где РНК, нацеленная на ДНК, содержит (а) первый сегмент, содержащий нуклеотидную последовательность, комплементарную последовательности ДНК-мишени; и (b) второй сегмент, который взаимодействует с сайт-специфическим модифицирующим полипептидом; и (ii) сайт-специфический модифицирующий полипептид или полинуклеотид, кодирующий его, где сайт-специфический модифицирующий полипептид содержит (а) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и (b) участок активности, который модулирует транскрипцию в пределах ДНК-мишени, причем сайт модулированной транскрипции в пределах ДНК-мишени определяется РНК, нацеленной на ДНК.

Например, в некоторых случаях композиция по изобретению содержит (i) РНК, нацеленную на ДНК, где РНК, нацеленная на ДНК, содержит (а) первый сегмент, содержащий нуклеотидную последовательность, комплементарную последовательности в ДНК-мишени; и (b) второй сегмент, который взаимодействует с сайт-специфическим модифицирующим полипептидом; и (ii) сайт-специфический модифицирующий полипептид, где сайт-специфический модифицирующий полипептид содержит (а) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и (b) участок активности, который модулирует транскрипцию в пределах ДНК-мишени, причем сайт модулированной транскрипции в пределах ДНК-мишени определяется РНК, нацеленной на ДНК.

В качестве другого примера, в некоторых случаях композиция по изобретению содержит (i) полинуклеотид ДНК, кодирующий РНК, нацеленную на ДНК, где РНК, нацеленная на ДНК, содержит (а) первый сегмент, содержащий нуклеотидную последовательность, комплементарную последовательности в ДНК-мишени; и (b) второй сегмент, который взаимодействует с сайт-специфическим модифицирующим полипептидом; и (ii) полинуклеотид, кодирующий сайт-специфический модифицирующий полипептид, где сайт-специфический модифицирующий полипептид содержит (а) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и (b) участок активности, который модулирует транскрипцию в пределах ДНК-мишени, причем сайт модулированной транскрипции в пределах ДНК-мишени определяется РНК, нацеленной на ДНК.

Композиция-мишень может содержать, в дополнение к i) РНК, нацеленной на ДНК, или полинуклеотиду ДНК, кодирующему ее; и ii) сайт-специфическому модифицирующему полипептиду или полинуклеотиду, кодирующему его, одно или более из: солей, например, NaCl, MgCl<sub>2</sub>, KCl, MgSO<sub>4</sub> и т.д.; буферного средства, например, трис-буфера, HEPES, MES, натриевой соли MES, MOPS, TAPS и т.д.; солибутилизирующего средства; детергента, например, неионогенного детергента, такого как Tween-20, и т.д.; ингибитора протеазы; восстанавливающего средства (например, дитиотреитола); и подобных.

В некоторых случаях компоненты композиции являются индивидуально чистыми, например, каждый из компонентов является на по меньшей мере около 75%, по меньшей мере около 80%, по меньшей мере около 90%, по меньшей мере около 95%, по меньшей мере около 98%, по меньшей мере около 99% или по меньшей мере 99% чистым. В некоторых случаях отдельные компоненты композиции-мишени являются чистыми перед добавлением к композиции.

Например, в некоторых вариантах реализации настоящего изобретения сайт-специфический модифицирующий полипептид, присутствующий в композиции-мишени, является чистым, например, на по меньшей мере около 75%, по меньшей мере около 80%, по меньшей мере около 85%, по меньшей мере около 90%, по меньшей мере около 95%, по меньшей мере около 98%, по меньшей мере около 99% или более чем на 99% чистым, причем "% чистоты" означает, что сайт-специфический модифицирующий полипептид на указанный процент свободен от других белков (например, белков, отличных от сайт-специфического модифицирующего полипептида), других макромолекул или загрязняющих веществ, которые могут присутствовать в процессе получения сайт-специфического модифицирующего полипептида.

## Наборы.

Настоящее изобретение обеспечивает наборы для осуществления заявленного способа. Набор по изобретению может содержать один или более из сайт-специфического модифицирующего полипептида; нуклеиновой кислоты, содержащей нуклеотид, кодирующий сайт-специфический модифицирующий полипептид; РНК, нацеленной на ДНК; нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую РНК, наделенную на ДНК; РНК-активатора; нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую РНК-активатор; РНК-нацеливающей; и нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую РНК-нацеливающую. Сайт-специфический модифицирующий полипептид; нуклеиновая кислота, содержащая нуклеотид, кодирующий сайт-специфический модифицирующий полипептид; РНК, нацеленная на ДНК; нуклеиновая кислота, содержащая нуклеотидную последовательность, кодирующую РНК, нацеленную на ДНК; РНК-активатор; нуклеиновая кислота, содержащая нуклеотидную последовательность, кодирующую РНК-активатор; РНК-нацеливающая; и нуклеиновая кислота, содержащая нуклеотидную последовательность, кодирующую РНК-нацеливающую описаны детально выше. Набор может содержать комплекс, который содержит один или более из: сайт-специфического модифицирующего полипептида; нуклеиновой кислоты, содержащей нуклеотид, кодирующий сайт-специфический модифицирующий полипептид; РНК, наделенной на ДНК; нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую РНК, наделенную на ДНК; РНК-активатора; нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую РНК-активатор; РНК-нацеливающей; и нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую РНК-нацеливающую.

В некоторых вариантах реализации настоящего изобретения набор по изобретению содержит сайт-специфический модифицирующий полипептид или полинуклеотид, кодирующий его. В некоторых вариантах реализации настоящего изобретения сайт-специфический модифицирующий полипептид содержит (а) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и (b) участок активности, который модулирует транскрипцию в ДНК-мишени, в котором сайт модулированной транскрипции в пределах ДНК-мишени определяется РНК, нацеленной на ДНК. В некоторых случаях участок активности сайт-специфического модифицирующего полипептида проявляет сниженную или инактивированную нуклеазную активность. В некоторых случаях сайт-специфический модифицирующий полипептид представляет собой гибридный сайт-специфический модифицирующий полипептид.

В некоторых вариантах реализации настоящего изобретения набор содержит: сайт-специфический модифицирующий полипептид или полинуклеотид, кодирующий его и реагент для ресуспендирования и/или разбавления сайт-специфического модифицирующего полипептида. В некоторых вариантах реализации настоящего изобретения набор содержит нуклеиновую кислоту (например, ДНК, РНК), содержащую нуклеотид, кодирующий сайт-специфический модифицирующий полипептид. В некоторых вариантах реализации настоящего изобретения набор содержит: нуклеиновую кислоту (например, ДНК, РНК), содержащую нуклеотид, кодирующий сайт-специфический модифицирующий полипептид; и реагент для ресуспендирования и/или разбавления сайт-специфического модифицирующего полипептида.

Набор, содержащий сайт-специфический модифицирующий полипептид или полинуклеотид, кодирующий его, может далее содержать один или несколько дополнительных реагентов, причем подобные дополнительные реагенты могут быть выбраны из буфера для введения сайт-специфического модифицирующего полипептида в клетку; промывочного буфера; контрольного реагента; контрольного вектора экспрессии или полинуклеотида РНК; реагента для *in vitro* получения сайт-специфического модифицирующего полипептида из ДНК и подобных. В некоторых случаях сайт-специфический модифицирующий полипептид, который содержится в наборе, представляет собой гибридный сайт-специфический модифицирующий полипептид-мишень, как описано выше.

В некоторых вариантах реализации настоящего изобретения набор содержит РНК, нацеленную на ДНК, или полинуклеотид ДНК, кодирующий ее, где РНК, нацеленная на ДНК, содержит (а) первый сегмент, содержащий нуклеотидную последовательность, комплементарную последовательности в ДНК-мишени; и (b) второй сегмент, который взаимодействует с сайт-специфическим модифицирующим полипептидом. В некоторых вариантах реализации настоящего изобретения РНК, нацеленная на ДНК, дополнительно содержит третий сегмент (как описано выше). В некоторых вариантах реализации настоящего изобретения набор содержит (i) РНК, нацеленную на ДНК, или полинуклеотид ДНК, кодирующий ее, где РНК, нацеленная на ДНК, содержит (а) первый сегмент, содержащий нуклеотидную последовательность, комплементарную последовательности в ДНК-мишени; и (b) второй сегмент, который взаимодействует с сайт-специфическим модифицирующим полипептидом; и (ii) сайт-специфический модифицирующий полипептид или полинуклеотид, кодирующий его, где сайт-специфический модифицирующий полипептид содержит (а) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и (b) участок активности, который проявляет сайт-специфическую ферментативную активность, причем сайт ферментативной активности определяется РНК, нацеленной на ДНК. В некоторых вариантах реализации настоящего изобретения участок активности сайт-специфического модифицирующего полипептида не проявляет ферментативную активность (содержит инактивированную нуклеазу, например, посредством мутации). В некоторых случаях набор содержит РНК, нацеленную на ДНК, и

сайт-специфический модифицирующий полипептид. В других случаях, набор содержит (i) нуклеиновую кислоту, содержащую нуклеотидную последовательность, кодирующую РНК, нацеленную на ДНК; и (ii) нуклеиновую кислоту, содержащую нуклеотидную последовательность, кодирующую сайт-специфический модифицирующий полипептид.

В другом примере, набор может содержать: (i) РНК, нацеленную на ДНК, или полинуклеотид ДНК, кодирующий ее, которые содержат (a) первый сегмент, содержащий нуклеотидную последовательность, комплементарную последовательности ДНК-мишени; и (b) второй сегмент, который взаимодействует с сайт-специфическим модифицирующим полипептидом; и (ii) сайт-специфический модифицирующий полипептид или полинуклеотид, кодирующий его, которые содержат (a) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и (b) участок активности, который модулирует транскрипцию в пределах ДНК-мишени, причем сайт модулированной транскрипции в пределах ДНК-мишени определяется РНК, нацеленной на ДНК. В некоторых случаях набор содержит (i) РНК, нацеленную на ДНК; и сайт-специфический модифицирующий полипептид. В других случаях, набор содержит (i) нуклеиновую кислоту, содержащую нуклеотидную последовательность, кодирующую РНК, нацеленную на ДНК; и (ii) нуклеиновую кислоту, содержащую нуклеотидную последовательность, кодирующую сайт-специфический модифицирующий полипептид.

Настоящее изобретение обеспечивает набор, содержащий (1) рекомбинантный вектор экспрессии, который содержит (i) нуклеотидную последовательность, кодирующую РНК, нацеленную на ДНК, при том что РНК, нацеленная на ДНК, содержит (a) первый сегмент, содержащий нуклеотидную последовательность, комплементарную последовательности в ДНК-мишени; и (b) второй сегмент, который взаимодействует с сайт-специфическим модифицирующим полипептидом; и (ii) нуклеотидную последовательность, кодирующую сайт-специфический модифицирующий полипептид, который содержит (a) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и (b) участок активности, который проявляет сайт-специфическую ферментативную активность, причем сайт ферментативной активности определяется РНК, нацеленной на ДНК; и (2) реагент для растворения и/или разбавления вектора экспрессии.

Настоящее изобретение обеспечивает набор, содержащий (1) рекомбинантный вектор экспрессии, который содержит (i) нуклеотидную последовательность, кодирующую РНК, нацеленную на ДНК, при том что РНК, нацеленная на ДНК, содержит (a) первый сегмент, содержащий нуклеотидную последовательность, комплементарную последовательности в ДНК-мишени; и (b) второй сегмент, который взаимодействует с сайт-специфическим модифицирующим полипептидом; и (ii) нуклеотидную последовательность, кодирующую сайт-специфический модифицирующий полипептид, который содержит (a) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и (b) участок активности, который модулирует транскрипцию в пределах ДНК-мишени, причем сайт ферментативной активности определяется РНК, нацеленной на ДНК; и (2) реагент для растворения и/или разбавления рекомбинантного вектора экспрессии.

Настоящее изобретение обеспечивает набор, содержащий (1) рекомбинантный вектор экспрессии, который содержит нуклеиновую кислоту, содержащую нуклеотидную последовательность, которая кодирует РНК, нацеленную на ДНК, содержащую: (i) первый сегмент, содержащий нуклеотидную последовательность, комплементарную последовательности в ДНК-мишени; и (ii) второй сегмент, который взаимодействует с сайт-специфическим модифицирующим полипептидом; и (2) реагент для растворения и/или разбавления рекомбинантного вектора экспрессии. В некоторых вариантах реализации настоящего набора, указанный набор содержит рекомбинантный вектор экспрессии, содержащий нуклеотидную последовательность, кодирующую сайт-специфический модифицирующий полипептид, причем сайт-специфический модифицирующий полипептид содержит (a) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и (b) участок активности, который проявляет сайт-специфическую ферментативную активность, при том что сайт ферментативной активности определяется РНК, нацеленной на ДНК. В других вариантах реализации настоящего набора, указанный набор содержит рекомбинантный вектор экспрессии, содержащий нуклеотидную последовательность, кодирующую сайт-специфический модифицирующий полипептид, причем сайт-специфический модифицирующий полипептид содержит (a) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и (b) участок активности, который модулирует транскрипцию в пределах ДНК-мишени, при том что сайт ферментативной активности в пределах ДНК-мишени определяется РНК, нацеленной на ДНК.

В некоторых вариантах реализации любого из указанных выше наборов, набор содержит РНК-активатор или РНК-нацеливающую. В некоторых вариантах реализации настоящего изобретения любого из указанных выше наборов, набор содержит РНК, нацеленную на ДНК, в виде одиночной молекулы. В некоторых вариантах реализации настоящего изобретения любого из указанных выше наборов, набор содержит две или более двойных молекул РНК, нацеленных на ДНК, или две или более РНК, нацеленных на ДНК в виде одиночной молекулы. В некоторых вариантах реализации любого из указанных выше наборов, РНК, нацеленная на ДНК, (например, в том числе две или более РНК, нацеленные на ДНК) может быть обеспечена в виде совокупности (например, совокупности молекул РНК, совокупности молекул

ДНК, кодирующих РНК, нацеленную(ые) на ДНК и т.д.). Такие наборы могут быть полезными, например, для использования в сочетании с описанными выше генетически модифицированными клетками-хозяевами, содержащими сайт-специфический модифицирующий полипептид-мишень. В некоторых вариантах реализации настоящего изобретения любого из указанных выше наборов, набор дополнительно содержит донорный полинуклеотид для влияния на необходимую генетическую модификацию. Компоненты набора могут находиться в отдельных контейнерах; или могут быть объединены в одном контейнере.

Любой из вышеописанных наборов может дополнительно содержать один или более дополнительных реагентов, где такие дополнительные реагенты могут быть выбраны из буфера для разбавления; раствора для растворения; промывочного буфера; контрольного реагента; контрольного вектора экспрессии или полинуклеотида РНК; реагента для *in vitro* получения сайт-специфического модифицирующего полипептида из ДНК и подобные.

В дополнение к упомянутым выше компонентам, набор может дополнительно содержать инструкции по использованию компонентов набора при осуществлении заявленных способов. Инструкции для осуществления заявленных способов, как правило, отражены на соответствующем носителе информации. Например, инструкции могут быть напечатаны на подложке, такой как бумага или пластик, и т.д. Таким образом, инструкции могут присутствовать в наборах в виде вставки в пакет, в маркировке контейнера набора или его компонентов (т.е. ассоциированных с упаковкой или с частью упаковки) и т.д. В других вариантах реализации настоящего изобретения инструкции присутствуют в виде электронного файла для хранения данных, присутствующего на подходящем машиночитаемом носителе данных, например, CD-ROM, дискете, флэш-дискете и т.д. В еще одних вариантах реализации настоящего изобретения сами инструкции не присутствуют в наборе, но предусмотрены средства для получения инструкций из удаленного источника, например, через интернет. Пример этого варианта реализации настоящего изобретения представляет собой набор, который содержит веб-адрес, на котором инструкции могут быть просмотрены и/или с которого возможно загрузить инструкции. Как и инструкции, эти средства для получения инструкций записывают на подходящий носитель.

Генетически модифицированные организмы, не являющиеся человеком.

В некоторых вариантах реализации настоящего изобретения генетически модифицированная клетка-хозяин была генетически модифицирована экзогенной нуклеиновой кислотой, содержащей нуклеотидную последовательность, кодирующую сайт-специфический модифицирующий полипептид (например, встречающийся в природе Cas9; модифицированный, т.е. мутировавший или вариантный, Cas9; гибридный Cas9; и т.д.). Если такая клетка представляет собой эукариотический одноклеточный организм, то модифицированную клетку можно считать генетически модифицированным организмом. В некоторых вариантах реализации настоящего изобретения, такой отличный от человека генетически модифицированный организм представляет собой трансгенный по Cas9 многоклеточный организм.

В некоторых вариантах реализации настоящего изобретения генетически модифицированная клетка-хозяин, не являющаяся клеткой человека (например, клетка, которая была генетически модифицирована экзогенной нуклеиновой кислотой, содержащей нуклеотидную последовательность, кодирующую сайт-специфический модифицирующий полипептид, например, встречающийся в природе Cas9; модифицированный, т.е. мутировавший или вариантный, Cas9; гибридный Cas9; и т.д.) может породить генетически модифицированный, не принадлежащий к человеческому роду организм (например, мышь, рыбу, лягушку, муху, червя и т.д.). Например, если генетически модифицированная клетка-хозяин представляет собой плюрипотентную стволовую клетку (т.е. PSC) или зародышевую клетку (например, спермий, ооцит, и т.д.), весь генетически модифицированный организм может быть получен из генетически модифицированной клетки-хозяина. В некоторых вариантах реализации настоящего изобретения генетически модифицированная клетка-хозяин представляет собой плюрипотентную стволовую клетку (например, ESC, iPSC, плюрипотентная стволовая клетка растений и т.д.) или зародышевую клетку (например, клетку спермий, ооцит и т.д.), либо *in vivo*, либо *in vitro*, что может дать начало генетически модифицированному организму. В некоторых вариантах реализации настоящего изобретения генетически модифицированная клетка-хозяин представляет собой PSC позвоночного (например, ESC, iPSC и т.д.) и ее используют для получения генетически модифицированного организма (например, с помощью введения PSC в бластоцисты с получением химерного/мозаичного животного, которое затем может быть скрещено с получением нехимерных/немозаичных генетически модифицированных организмов; прививки в случае растений; и т.д.). Любой пригодный способ/протокол получения генетически модифицированного организма, в том числе описанные в данном документе способы, являются подходящими для получения генетически модифицированной клетки-хозяина, содержащей экзогенную нуклеиновую кислоту, которая содержит нуклеотидную последовательность, кодирующую сайт-специфический модифицирующий полипептид (например, встречающийся в природе Cas9; модифицированный, т.е. мутировавший или вариантный, Cas9; гибридный Cas9; и т.д.). Способы получения генетически модифицированных организмов известны в данной области техники. Например, см. Cho и соавт., *Curr Protoc Cell Biol.* 2009 Mar; Chapter 19: Unit 19.11: Generation of transgenic mice; Gama и соавт., *Brain Struct Funct.* 2010 Mar; 214(2-3):91-109. Epub 2009 Nov 25: Animal transgenesis: an overview; Husaini и соавт., *GM Crops.* 2011 Jun-Dec; 2(3): 150-

62. Epub 2011 Jun 1: Approaches for gene targeting and targeted gene expression in plants.

В некоторых вариантах реализации настоящего изобретения генетически модифицированный организм содержит клетку-мишень для способов по настоящему изобретению и, таким образом, может считаться источником клеток-мишеней. Например, если генетически модифицированная клетка, содержащая экзогенную нуклеиновую кислоту, которая содержит нуклеотидную последовательность, кодирующую сайт-специфический модифицирующий полипептид (например, встречающийся в природе Cas9; модифицировано, т.е. мутировавший или вариантный, Cas9; гибридный Cas9; и т.д.), используют, с получением генетически модифицированного организма, после чего клетки генетически модифицированного организма содержат экзогенную нуклеиновую кислоту, содержащую нуклеотидную последовательность, кодирующую сайт-специфический модифицирующий полипептид (например, встречающийся в природе Cas9; модифицированный, т.е. мутировавший или вариантный, Cas9; гибридный Cas9; и т.д.). В некоторых подобных вариантах реализации настоящего изобретения ДНК клетки или клеток генетически модифицированного организма можно направленно модифицировать с помощью введения в клетку или клетки РНК, нацеленной на ДНК, (или ДНК, кодирующей РНК, нацеленную на ДНК) и, необязательно, донорной нуклеиновой кислоты. Например, при введении РНК, нацеленной на ДНК, (или ДНК, кодирующей РНК, нацеленную на ДНК) в субпопуляцию клеток (например, клеток головного мозга, клеток кишечника, клеток почки, клеток легких, клеток крови и т.п.) генетически модифицированного организма можно направленно модифицировать ДНК таких клеток, где геномная локализация модификации будет зависеть от нацеленной на ДНК последовательности введенной РНК, нацеленной на ДНК.

В некоторых вариантах реализации настоящего изобретения генетически модифицированный организм является источником клеток-мишеней для способов по настоящему изобретению. Например, генетически модифицированный организм содержит клетки, которые являются генетически модифицированными экзогенной нуклеиновой кислотой, содержащей нуклеотидную последовательность, кодирующую сайт-специфический модифицирующий полипептид (например, встречающийся в природе Cas9; модифицированный, т.е. мутировавший или вариантный, Cas9; гибридный Cas9; и т.д.) и может служить источником генетически модифицированных клеток, например PSC (например, ESC, iPSC, спермиев, ооцитов и т.д.), нейронов, клеток-предшественников, кардиомиоцитов и т.д.

В некоторых вариантах реализации настоящего изобретения генетически модифицированная клетка-хозяин представляет собой PSC, которая содержит экзогенную нуклеиновую кислоту, содержащую нуклеотидную последовательность, кодирующую сайт-специфический модифицирующий полипептид (например, встречающийся в природе Cas9; модифицированный, т.е. мутировавший или вариантный, Cas9; гибридный Cas9; и т.д.). Таким образом, PSC может быть клеткой-мишенью, таким образом, что ДНК PSC можно направленно модифицировать с помощью введения в PSC РНК, нацеленной на ДНК, (или ДНК, кодирующей РНК, нацеленную на ДНК), и, необязательно, донорной нуклеиновой кислоты, а геномная локализация модификации будет зависеть от нацеленной на ДНК последовательности введенной РНК, нацеленной на ДНК. Таким образом, в некоторых вариантах реализации настоящего изобретения способы, описанные в данном документе, могут быть использованы для модификации ДНК (например, удаления и/или замены любой необходимой геномной локализации) PSC, полученной из генетически модифицированного организма. Такие модифицированные PSC могут затем использовать для того, чтобы получить организмы, обладающие и (i) экзогенной нуклеиновой кислотой, содержащей нуклеотидную последовательность, кодирующую сайт-специфический модифицирующий полипептид (например, встречающийся в природе Cas9; модифицированный, т.е. мутировавший или вариантный, Cas9; гибридный Cas9; и т.д.) и (ii) модификацией ДНК, которая была введена в PSC.

Экзогенная нуклеиновая кислота, содержащая нуклеотидную последовательность, кодирующую сайт-специфический модифицирующий полипептид (например, встречающийся в природе Cas9; модифицированный, т.е. мутировавший или вариантный, Cas9; гибридный Cas9; и т.д.) может быть под контролем (т.е. функционально связанным с) неизвестного промотора (например, при условии, что нуклеиновая кислота случайным образом интегрируется в геном клетки-хозяина) или может быть под контролем (т.е. функционально связанным с) известного промотора. Подходящие известные промоторы могут быть любыми известными промоторами и включают постоянно активные промоторы (например, промотор CMV), индуцируемые промоторы (например, промотор теплового шока, тетрациклин-регулируемый промотор, стероид-регулируемый промотор, металл-регулируемый промотор, регулируемый рецептором эстрогена промотор и т.д.), пространственно ограниченные и/или временно ограниченные промоторы (например, тканеспецифический промотор, специфический к типу клеток промотор и т.д.) и т.д.

Генетически модифицированный организм по изобретению (например, организм, клетки которого содержат нуклеотидную последовательность, кодирующую сайт-специфический модифицирующий полипептид, например, встречающийся в природе Cas9; модифицированный, т.е. мутировавший или вариантный, Cas9; гибридный Cas9; и т.д.) может быть любым организмом, в том числе, например, растением; водорослью; беспозвоночным (например, книдарией, иглокожим, червем, мухой и др.); позвоночным (например, рыбой (например, данио-рерио, рыбой фугу, золотой рыбкой и т.д.), амфибией (например, саламандрой, лягушкой и т.д.), рептилией, птицей, млекопитающим и т.д.); копытным (например, козой, свиньей, овцой, коровой и т.д.); грызуном (например, мышью, крысой, хомяком, морской свинкой); зай-

цеобразным (например, кроликом); и т.д.

В некоторых случаях сайт-специфический модифицирующий полипептид содержит аминокислотную последовательность, имеющую идентичность по меньшей мере около 75%, по меньшей мере около 80%, по меньшей мере около 85%, по меньшей мере около 90%, по меньшей мере около 95%, по меньшей мере около 99 или 100% аминокислотной последовательности с аминокислотами 7-166 или 731-1003 аминокислотной последовательности Cas9/Csn1, изображенной на фиг. 3, с соответствующими участками любой из аминокислотных последовательностей, приведенных в SEQ ID NO:1-256 и 795-1346.

Трансгенные не принадлежащие к человеческому роду животные.

Как описано выше, в некоторых вариантах реализации настоящего изобретения нуклеиновая кислота по изобретению (например, нуклеотидная последовательность, кодирующая сайт-специфический модифицирующий полипептид, например, встречающийся в природе Cas9; модифицированный, т.е. мутировавший или вариантный, Cas9; гибридный Cas9; и т.д.) или рекомбинантный вектор экспрессии по изобретению используют в качестве трансгена, для того, чтобы получить трансгенное животное, которое продуцирует сайт-специфический модифицирующий полипептид. Таким образом, настоящее изобретение дополнительно обеспечивает трансгенное не принадлежащее к человеческому роду животное, которое содержит трансген, содержащий нуклеиновые кислоты, которые содержат нуклеотидную последовательность, кодирующую сайт-специфический модифицирующий полипептид, например, встречающийся в природе Cas9; модифицированный, т.е. мутировавший или вариантный, Cas9; гибридный Cas9; и т.д., как описано выше. В некоторых вариантах реализации настоящего изобретения геном трансгенного не принадлежащего к человеческому роду животного содержит нуклеотидную последовательность по изобретению, кодирующую сайт-специфический модифицирующий полипептид. В некоторых вариантах реализации настоящего изобретения трансгенное не принадлежащее к человеческому роду животное является гомозиготным по генетической модификации. В некоторых вариантах реализации настоящего изобретения трансгенное не принадлежащее к человеческому роду животное является гетерозиготным по генетической модификации. В некоторых вариантах реализации настоящего изобретения трансгенное не принадлежащее к человеческому роду животное представляет собой позвоночное, например, рыбу (например, данио-рерио, золотую рыбку, рыбу фугу, пещерную рыбу и т.д.), амфибию (лягушку, саламандру и т.д.), птицу (например, курицу, индейку и т.д.), рептилию (например, змею, ящерицу и т.д.), млекопитающее (например, копытное, например, свинью, корову, козу, овцу и т.д.; зайцеобразное (например, кролика); грызуна (например, крысу, мышь); отличного от человека примата; и т.д.) и т.д.

Экзогенная нуклеиновая кислота, содержащая нуклеотидную последовательность, кодирующую сайт-специфический модифицирующий полипептид (например, встречающийся в природе Cas9; модифицированный, т.е. мутировавший или вариантный, Cas9; гибридный Cas9; и т.д.) может быть под контролем (т.е. функционально связанным с) неизвестного промотора (например, при условии, что нуклеиновая кислота случайным образом интегрируется в геном клетки-хозяина) или может быть под контролем (т.е. функционально связанным с) известного промотора. Подходящие известные промоторы могут быть любыми известными промоторами и включают постоянно активные промоторы (например, промотор CMV), индуцируемые промоторы (например, промотор теплового шока, тетрациклин-регулируемый промотор, стероид-регулируемый промотор, металл-регулируемый промотор, регулируемый рецептором эстрогена промотор и т.д.), пространственно ограниченные и/или временно ограниченные промоторы (например, тканеспецифический промотор, специфический к типу клеток промотор и т.д.) и т.д.

В некоторых случаях сайт-специфический модифицирующий полипептид содержит аминокислотную последовательность, имеющую идентичность по меньшей мере около 75%, по меньшей мере около 80%, по меньшей мере около 85%, по меньшей мере около 90%, по меньшей мере около 95%, по меньшей мере около 99 или 100% аминокислотной последовательности с аминокислотами 7-166 или 731-1003 аминокислотной последовательности Cas9/Csn1, изображенной на фиг. 3, или с соответствующими участками любой из аминокислотных последовательностей, приведенных в SEQ ID NO:1-256 и 795-1346.

Трансгенные растения.

Как описано выше, в некоторых вариантах реализации настоящего изобретения нуклеиновая кислота по изобретению (например, нуклеотидная последовательность, кодирующая сайт-специфический модифицирующий полипептид, например, встречающийся в природе Cas9; модифицированный, т.е. мутировавший или вариантный, Cas9; гибридный Cas9; и т.д.) или рекомбинантный вектор экспрессии по изобретению используются в качестве трансгена для того, чтобы получить трансгенное растение, которое продуцирует сайт-специфический модифицирующий полипептид. Таким образом, настоящее изобретение дополнительно обеспечивает трансгенное растение, которое содержит трансген, содержащий нуклеиновые кислоты, которые содержат нуклеотидную последовательность, кодирующую сайт-специфический модифицирующий полипептид, например, встречающийся в природе Cas9; модифицированный, т.е. мутировавший или вариантный, Cas9; гибридный Cas9; и т.д., как описано выше. В некоторых вариантах реализации настоящего изобретения геном трансгенного растения содержит нуклеиновую кислоту-мишень. В некоторых вариантах реализации настоящего изобретения трансгенное растение является гомозиготным по генетической модификации. В некоторых вариантах реализации настоящего изобретения трансгенное растение является гетерозиготным по генетической модификации.

Способы введения экзогенных нуклеиновых кислот в клетки растений хорошо известны в данной области техники. Такие растительные клетки считаются "трансформированными", как определено выше. Подходящие способы включают вирусную инфекцию (такие как вирусы с двухцепочечными ДНК), трансфекцию, конъюгацию, слияние протопластов, электропорацию, технологию генной пушки, осаждение фосфатом кальция, прямую микроинъекцию, технология нитевидных кристаллов из карбида кремния, опосредованную *Agrobacterium* трансформацию и подобные. Выбор способа, как правило, зависит от типа клетки, которую подвергают трансформации, и обстоятельств, при которых происходит трансформация (например, *in vitro*, *ex vivo* или *in vivo*).

Способы трансформации, основанные на использовании почвенной бактерии *Agrobacterium tumefaciens*, особенно подходят для введения молекулы экзогенной нуклеиновой кислоты в сосудистое растение. Форма *Agrobacterium* дикого типа содержит T1 (опухолеиндуцирующую) плазмиду, который направляет рост туморогенного корончатого галла на растениях-хозяевах. Перенос опухолеиндуцирующей области T-ДНК из T1 плазмиды в геном растения требует наличия генов вирулентности, кодируемых T1 плазмидой, а также T-ДНК границ, которые представляют собой совокупность прямых повторов ДНК, которые очерчивают область, которая должна быть перенесена. Вектор на основе *Agrobacterium* представляет собой модифицированную форму T1 плазмиды, в которой опухолеиндуцирующие функции заменяют на представляющую интерес последовательность нуклеиновой кислоты, для того, чтобы они были введены растению-хозяину.

При опосредованной *Agrobacterium* трансформации обычно используются коинтегрирующие векторы или бинарные векторные системы, в которых компоненты T1 плазмиды делят между вектором-помощником, который постоянно находится в хозяине *Agrobacterium* и несет гены вирулентности, а также шаттл-вектором, который содержит представляющий интерес ген, связанный с последовательностями T-ДНК. В данной области техники хорошо известно и коммерчески доступно множество бинарных векторов, например, от Clontech (Palo Alto, Calif.). Способы сокультивирования *Agrobacterium* с культивируемыми клетками растений или поврежденной тканью, такой как ткани листьев, корневые эксплантаты, многосемядольные, столовые части или клубни, например, также хорошо известны в данной области. См., например, Glick и Thompson, (ред.), "Methods in Plant Molecular Biology and Biotechnology, Boca Raton, Fla.: CRC Press (1993).

Опосредованная микрочастицами трансформация также может быть использована для получением трансгенного растения по изобретению. Этот способ, впервые описанный Klein и соавт. (Nature 327:70-73 (1987)), основан на использовании микрочастиц, таких как золото или вольфрам, которые покрыты желаемой молекулой нуклеиновой кислоты, что было сделано с помощью осаждения хлоридом кальция, спермидином или полиэтиленгликолем. С использованием соответствующего устройства, такого как BIOLISTIC PD-1000 (Biogard; Hercules Calif.), частицы микроснарядов разгоняют до высокой скорости для введения в ткани покрытосеменного растения.

Нуклеиновая кислота по изобретению может быть введена в растение таким образом, чтобы нуклеиновая кислота была способна войти в растительную клетку(и), например, посредством *in vivo* или *ex vivo* протокола. Под "in vivo" подразумевают то, что нуклеиновую кислоту вводят в тело живого организма растения, например, инфльтрацией. Под "ex vivo" подразумевают то, что клетки или экспланты модифицируют вне растения, а затем такие клетки или органы регенерируют в растения. Был описан ряд векторов, подходящих для стабильной трансформации растительных клеток или для создания трансгенных растений, в том числе те, которые описаны в Weissbach and Weissbach, (1989) Methods for Plant Molecular Biology Academic Press, и Gelvin и соавт., (1990) Plant Molecular Biology Manual, Kluwer Academic Publishers. Конкретные примеры содержат те, которые получают из T1 плазмиды *Agrobacterium tumefaciens*, а также те, которые раскрыты в Herrera-Estrella и соавт. (1983) Nature 303: 209, Bevan (1984) Nucl Acid Res. 12: 8711-8721, Klee (1985) Bio/Technolo 3: 637-642. В качестве альтернативы, отличные от T1 векторы могут быть использованы для переноса ДНК в растения и клетки с помощью способов доставки свободной ДНК. Используя эти способы, могут быть получены трансгенные растения, такие как пшеница, рис (Christou (1991) Bio/Technology 9:957-9 и 4462) и кукуруза (Gordon-Kamm (1990) Plant Cell 2: 603-618). Незрелые эмбрионы могут также быть хорошей тканью-мишенью у однодольных в случае использования способов прямой доставки ДНК с помощью генной пушки (Weeks и соавт. (1993) Plant Physiol 102: 1077-1084; Vasil (1993) Bio/Technolo 10: 667-674; Wan и Lemeaux (1994) Plant Physiol 104: 37-48 и *Agrobacterium*-опосредованного переноса ДНК (Ishida и соавт. (1996) Nature Biotech 14: 745-750). Иллюстративные способы введения ДНК в хлоропласты представляют собой биолистическую бомбардировку, трансформацию протопластов с помощью полиэтиленгликоля и микроинъекцию (Danieli и соавт., Nat. Biotechnol 16:345-348, 1998; Staub и соавт., Nat. Biotechnol 18: 333-338, 2000; O'Neill и соавт., Plant T. 3:729-738, 1993; Knoblauch и соавт., Nat. Biotechnol 17: 906-909; патенты США № 5451513, 5545817, 5545818 и 5576198; в межд. заявке № WO 95/16783; и в Boynton и соавт., Methods in Enzymology 217: 510-536 (1993), Svab и соавт., Proc. Natl. Acad. Sci. USA 90: 913-917 (1993) и McBride и соавт., Proc. Natl. Acad. Sci. USA 91: 7301-7305 (1994)). Любой вектор, подходящий для способов биолистической бомбардировки, трансформации протопластов с помощью полиэтиленгликоля и микроинъекции будет пригоден в качестве нацеливающего вектора в случае трансформации хлоропластов. Любой вектор двухцепочеч-

ной ДНК может быть использован в качестве вектора трансформации, особенно при условии, что в способе введения не используется *Agrobacterium*.

Растения, которые могут быть генетически модифицированы, включают зерновые, кормовые культуры, фрукты, овощи, масляничные семенные культуры, пальмы, лесные культуры и вьющиеся растения. Конкретные примеры растений, которые могут быть модифицированы, представляют собой следующие: кукуруза, банан, арахис, горох огородный, подсолнечник, томат, рапс, табак, пшеница, ячмень, овес, картофель, соя, хлопок, гвоздика, сорго, люпин и рис.

Помимо этого, предлагаются трансформированные растительные клетки, ткани, растения и продукты, которые содержат трансформированные растительные клетки. Особенностью трансформированных клеток по изобретению, тканей и продуктов, которые содержат то же самое, является присутствие нуклеиновой кислоты по изобретению, интегрированной в геном, и получение сайт-специфического модифицирующего полипептида с помощью растительных клеток, например, встречающегося в природе Cas9; модифицированного, т.е. мутировавшего или вариантного, Cas9; гибридного Cas9; и т.д. Рекомбинантные клетки растений по настоящему изобретению применимы в виде популяций рекомбинантных клеток, или в качестве ткани, семени, целого растения, стебля, фрукта, листа, корня, цветка, стебля, клубня, зерна, корма для животных, поля растений и подобных.

Нуклеиновая кислота, содержащая нуклеотидную последовательность, кодирующую сайт-специфический модифицирующий полипептид (например, встречающийся в природе Cas9; модифицированный, т.е. мутировавший или вариантный, Cas9; гибридный Cas9; и т.д.) может быть под контролем (т.е. функционально связанным с) неизвестного промотора (например, при условии, что нуклеиновая кислота случайным образом интегрируется в геном клетки-хозяина) или может быть под контролем (т.е. функционально связанным с) известного промотора. Подходящие известные промоторы могут быть любым известным промотором и включают постоянно активные промоторы, индуцируемые промоторы, пространственно ограниченные и/или временно ограниченные промоторы и т.д.

В некоторых случаях сайт-специфический модифицирующий полипептид содержит аминокислотную последовательность, имеющую идентичность по меньшей мере около 75%, по меньшей мере около 80%, по меньшей мере около 85%, по меньшей мере около 90%, по меньшей мере около 95%, по меньшей мере около 99 или 100% аминокислотной последовательности с аминокислотами 7-166 или 731-1003 аминокислотной последовательности Cas9/Csn1, изображенной на фиг. 3, или с соответствующими участками любой из аминокислотных последовательностей, приведенных в SEQ ID NO:1-256 и 795-1346.

Помимо этого, предлагается репродуктивный материал трансгенного растения по изобретению, причем репродуктивный материал включает семена, растения-потомки и клональный материал.

## Определения

### Часть II

Термин "встречающийся в природе" или "немодифицированный", используют в данном документе применительно к нуклеиновой кислоте, полипептиду, клетке или организму, которые соответствуют обнаруживаемым в природе нуклеиновой кислоте, полипептиду, клетке иди организму. Например, полипептидная или полинуклеотидная последовательность, присутствующая в организме (в том числе вирусов), которая может быть выделена из природного источника и которая не была преднамеренно модифицирована человеком в лаборатории, является встречающейся в природе.

"Гетерологичный", как используют в данном документе, означает нуклеотидную последовательность или полипептид, которые не обнаруживаются в нативной нуклеиновой кислоте или белке соответственно. Например, в слитом варианте сайт-специфического полипептида Cas9, вариант сайт-специфического полипептида Cas9 может быть слит с гетерологичным полипептидом (т.е. полипептидом, отличным от Cas9). Гетерологичный полипептид может проявлять активность (например, ферментативную активность), что также будет проявляться слитым вариантом сайт-специфического полипептида Cas9. Гетерологичная последовательность нуклеиновой кислоты может быть связана с одним из вариантов сайт-специфического полипептида Cas9 (например, с помощью генной инженерии) для того, чтобы создать нуклеотидную последовательность, кодирующую слитый вариант сайт-специфического полипептида Cas9.

Термин "гибридный полипептид" означает полипептид, который не является встречающимся в природе, например, полученным с помощью искусственной комбинации двух отличных разделенных сегментов аминокислотной последовательности посредством вмешательства человека. Таким образом, гибридный полипептид также является результатом вмешательства человека. Таким образом, полипептид, который содержит гибридную аминокислотную последовательность, представляет собой гибридный полипептид.

"Сайт-специфический полипептид" и "РНК-связывающий сайт-специфический полипептид", означают полипептид, который связывает РНК и нацелен на специфическую последовательность ДНК. Сайт-специфический полипептид, как описано в данном документе, нацелен на специфическую последовательность ДНК посредством молекулы РНК, с которой он связан. Молекула РНК содержит последовательность, комплементарную к последовательности-мишени в пределах исследуемой ДНК, тем самым направляя связанный полипептид на специфическую локализацию в пределах ДНК-мишени (последова-

тельность-мишень).

В некоторых вариантах реализации настоящего изобретения нуклеиновая кислота по изобретению (например, РНК, нацеленная на ДНК, нуклеиновая кислота, которая содержит нуклеотидную последовательность, кодирующую РНК, нацеленную на ДНК; нуклеиновая кислота, кодирующая сайт-специфический полипептид; и т.д.) содержит в себя модификацию или последовательность, которая обеспечивает дополнительный желательный признак (например, модифицированную или регулируемую стабильность; субклеточное нацеливание; слежение, например, флуоресцентную метку; связывающий сайт в случае белка или белкового комплекса; и т.д.). Неограничивающие примеры содержат 5' кэп (например, 7-метилгуанилатный кэп (m7G)); полиаденилированный 3'-конец (т.е. 3'-поли(А) конец); последовательность рибосвитча а (например, для обеспечения регулируемой стабильности и/или регулируемой доступности для белков и/или комплексов белков); модификация или последовательность, которая нацеливает РНК на внутриклеточную локализацию (например, к ядру, митохондрии, хлоропластам и подобным); модификация или последовательность, которую обеспечивают возможность слежения (например, прямая конъюгация с флуоресцентной молекулой, конъюгация с фрагментом, который облегчает флуоресцентную детекцию, последовательность, которая обеспечивает флуоресцентную детекцию и др.); модификация или последовательность, которая обеспечивает сайт связывания для белков (например, белки, которые действуют на ДНК, в том числе активаторы транскрипции, репрессоры транскрипции, метилтрансферазы ДНК, деметилазы ДНК, ацетилтрансферазы гистонов, гистондеацетилазы гистонов и подобные); и их комбинации.

В некоторых вариантах реализации настоящего изобретения РНК, нацеленная на ДНК, содержит дополнительный сегмент либо на 5'-, либо на 3'-конце, что обеспечивает любой из признаков, описанных выше. Например, подходящий третий сегмент может содержать 5' кэп (например, 7-метилгуанилатный кэп (m7G)); полиаденилированный 3'-конец (т.е. 3'-поли(А) конец); последовательность рибосвитча а (например, для обеспечения регулируемой стабильности и/или регулируемой доступности для белков и/или комплексов белков); модификацию или последовательность, которая нацеливает РНК к внутриклеточной локализации (например, к ядру, митохондрии, хлоропластам и подобным); модификация или последовательность, которая обеспечивает возможность слежения (например, прямая конъюгация с флуоресцентной молекулой, конъюгация с фрагментом, который облегчает флуоресцентную детекцию, последовательность, которая обеспечивает флуоресцентную детекцию и др.); модификацию или последовательность, которая обеспечивает сайт связывания для белков (например, белки, которые действуют на ДНК, в том числе активаторы транскрипции, репрессоры транскрипции, метилтрансферазы ДНК, деметилазы ДНК, ацетилтрансферазы гистонов, гистондеацетилазы гистонов и подобные); и их комбинации.

РНК-мишень, нацеленная на ДНК, и сайт-специфический полипептид по изобретению образуют комплекс (т.е. связываются с помощью нековалентных взаимодействий). РНК, нацеленная на ДНК, обеспечивает целевую специфичность комплексу тем, что содержит нуклеотидную последовательность, которая является комплементарной к последовательности ДНК-мишени. Сайт-специфический полипептид обеспечивает сайт-специфическую активность. Другими словами, сайт-специфический полипептид направляют к последовательности ДНК-мишени (например, последовательности-мишени в хромосомной нуклеиновой кислоте; последовательности-мишени во внехромосомной нуклеиновой кислоте, например, эписомной нуклеиновой кислоте, мини-кольце и т.д.; последовательности-мишени в митохондриальной нуклеиновой кислоте; последовательности-мишени в хлоропластной нуклеиновой кислоте; последовательности-мишени в плазмиде; и т.д.) в силу его ассоциации с белок-связывающим сегментом РНК, нацеленной на ДНК.

В некоторых вариантах реализации настоящего изобретения заявленная РНК, нацеленная на ДНК, содержит две отдельные молекулы РНК (полинуклеотиды РНК) и упоминается в данном описании как "РНК, нацеленная на ДНК, в виде двойной молекулы" или "РНК в виде двух молекул, нацеленная на ДНК". В других вариантах реализации настоящего изобретения заявленная РНК, нацеленная на ДНК, представляет собой одиночную молекулу РНК (отдельный полинуклеотид РНК) и в данном документе обозначается как "РНК, нацеленная на ДНК, в виде одиночной молекулы". Если не указано другое, термин "РНК, нацеленная на ДНК", является всеохватывающим, соответствующим как РНК, нацеленной на ДНК, в виде одиночной молекулы, так и РНК, нацеленным на ДНК, в виде двойной молекулы.

Заявленная РНК, нацеленная на ДНК, в виде двойной молекулы состоит из двух отдельных молекул РБК ("РНК-нацеливающей" и "РНК-активатора"). Каждая из двух молекул РНК в РНК-мишени, нацеленной на ДНК, в виде двойной молекулы, содержит участок нуклеотидов, которые комплементарны друг с другом таким образом, что комплементарные нуклеотиды двух молекул РНК гибридизуются с образованием дуплекса двухцепочечной РНК белок-связывающего сегмента.

Заявленная РНК, нацеленная на ДНК, в виде одиночной молекулы, состоит из двух участков нуклеотидов (РНК-нацеливающей и РНК-активатора), которые являются комплементарными друг другу, и ковалентно связаны посредством промежуточных нуклеотидов ("линкеров" или "линкерных нуклеотидов") и гибридизуются с образованием дуплекса двухцепочечной РНК (дуплекса дцРНК) белок-связывающего сегмента, с образованием структуры типа "стебель - петля". РНК-нацеливающая и РНК-активатор могут быть ковалентно связаны через 3\* конец РНК-нацеливающей и 5'-конец РНК-

активатора. Как вариант, РНК-нацеливающая и РНК-активатор могут быть ковалентно связаны посредством 3'-конца РНК-нацеливающей и 5\* конца РНК-активатора.

Иллюстративная РНК в виде двойной молекулы, нацеленная на ДНК, содержит сгРНК-подобную ("РНК CRISPR", "РНК-нацеливающая", "сгРНК" или "повтор сгРНК") молекулу и соответствующую tracrРНК-подобную ("трансдействующую РНК CRISPR", "РНК-активатор" или "tracrРНК") молекулу. сгРНК-подобная молекула (РНК-нацеливающая) содержит как нацеленный на ДНК-сегмент (одноцепочечный) РНК, нацеленной на ДНК, так и участок ("дуплекс-образующий сегмент") нуклеотидов, который образует одну половину дуплекса дцРНК белок-связывающего сегмента РНК, нацеленной на ДНК. Соответствующая tracrРНК-подобная молекула (РНК-активатор) содержит участок нуклеотидов (дуплекс-образующий сегмент), который образует другую половину дуплекса дцРНК белок-связывающего сегмента РНК, нацеленной на ДНК. Другими словами, участок нуклеотидов сгРНК-подобной молекулы является комплементарным и гибридизуется с участком нуклеотидов tracrРНК-подобной молекулы с образованием дуплекса дцРНК белок-связывающего домена РНК, нацеленной на ДНК. Таким образом, каждая сгРНК-подобная молекула, можно сказать, имеет соответствующую tracrРНК-подобную молекулу. сгРНК-подобная молекула дополнительно обеспечивает одноцепочечный ДНК-нацеленный сегмент. Таким образом, сгРНК-подобная и tracrРНК-подобная молекула (в качестве соответствующей пары) гибридизуются с образованием РНК, нацеленной на ДНК. Точная последовательность данной молекулы сгРНК или tracrРНК характерна для видов, в которых находят молекулы РНК.

Термин "РНК-активатор" используют в данном документе для обозначения tracrРНК-подобной молекулы РНК, нацеленной на ДНК, в виде двойной молекулы. Термин "РНК-нацеливающая" используют в данном документе для обозначения сгРНК-подобной молекулы РНК, нацеленной на ДНК, в виде двойной молекулы. Термин "дуплекс-образующий сегмент" используется в данном документе для обозначения участка нуклеотидов РНК-активатора или РНК-нацеливающей, который способствует образованию дуплекса дцРНК с помощью гибридизации с участком нуклеотидов соответствующей молекулы РНК-активатора или РНК-нацеливающей. Другими словами, РНК-активатор содержит дуплекс-образующий сегмент, который является комплементарным дуплекс-образующему сегменту соответствующего РНК-нацеливающей. Таким образом, РНК-активатор содержит дуплекс-образующий сегмент, тогда как РНК-нацеливающая содержит как дуплекс-образующий сегмент, так и нацеленный на ДНК сегмент РНК, нацеленной на ДНК. Таким образом, РНК-мишень, нацеленная на ДНК, в виде двойной молекулы может состоять из любой соответствующей пары РНК-активатора и РНК-нацеливающей.

РНК в виде двух молекул, нацеленная на ДНК, может быть получена для того, чтобы обеспечить контролируемое (т.е. зависимое от условий) связывание РНК-нацеливающей с РНК-активатором. Из-за того, что РНК в виде двойной молекулы, нацеленная на ДНК, является нефункциональной, если только и РНК-активатор, и РНК-нацеливающая не связаны в функциональный комплекс с dCas9, РНК в виде двойной молекулы, нацеленная на ДНК, может быть индуцируемой (например, индуцируемой лекарственным средством) за счет того, что связывание между РНК-активатором и РНК-нацеливающей может быть индуцируемым. В качестве неограничивающего примера, аптамеры РНК могут быть использованы для регулирования (т.е. контроля) связывания РНК-активатора с РНК-нацеливающей. Соответственно, РНК-активатор и/или РНК-нацеливающая могут содержать последовательность аптамера РНК.

Аптамеры РНК известны в данной области техники и представляют собой, как правило, синтетический вариант рибосвитча. Термины "аптамер РНК" и "рибосвитч" используют в данном документе взаимозаменяемо, для того, чтобы охватить как синтетические, так и природные последовательности нуклеиновых кислот, которые обеспечивают индуцируемую регуляцию структуры (и, следовательно, доступность специфических последовательностей) молекулы РНК, частью которых они являются. Аптамеры РНК обычно содержат последовательность, которая складывается в определенную структуру (например, шпилька), которая специфически связывает конкретное лекарственное средство (например, малую молекулу). Связывание лекарственного средства вызывает структурные изменения в укладке РНК, изменяющие свойства нуклеиновой кислоты, частью которой является аптамер. В качестве ограничивающих примеров: (i) РНК-активатор с аптамером не может быть способен связываться с соответствующей РНК-нацеливающей, если аптамер не связан с соответствующим лекарственным средством; (ii) РНК-нацеливающая с аптамером, возможно, не сможет связываться с соответствующим РНК-активатором, если аптамер не связан с соответствующим лекарственным средством; и (iii) РНК-нацеливающая и РНК-активатор, каждый из которых содержит отличный аптамер, связывающий отличное лекарственное средство, возможно, не смогут связаться друг с другом, если не присутствуют оба лекарственных средства. Как показано на этих примерах, РНК в виде двух молекул, нацеленная на ДНК, может быть создана так, чтобы быть индуцируемой.

Примеры аптамеров и рибосвитчей можно найти, например, в Nakamura и соавт., *Genes Cells*. 2012-May; 17(5):344-64; Vavalle и соавт., *Future Cardiol*. 2012 May; 8(3):371-82; Citartan и соавт., *Biosens Bioelectron*. 2012 Apr 15; 34(1):1-11; и Liberman и соавт., *Wiley Interdiscip Rev RNA*. 2012 May-Jun; 3(3):369-84; все из которых содержатся в данном документе в виде ссылки во всей их полноте.

Неограничивающие примеры нуклеотидных последовательностей, которые могут содержаться в РНК, нацеленной на ДНК, в виде двух молекул содержат РНК-нацеливающую (например, SEQ ID

NO:566-567), что может образовывать пару с дуплекс-образующим сегментом любого из РНК-активаторов, приведенных в SEQ ID NO:671-678.

Иллюстративная РНК, нацеленная на ДНК, в виде одиночной молекулы содержит два комплементарных участка нуклеотидов, которые гибридизуются с образованием дуплекса дцРНК. В некоторых вариантах реализации настоящего изобретения один из двух комплементарных участков нуклеотидов РНК, нацеленной на ДНК, в виде одиночной молекулы, (или ДНК, кодирующей участок) на по меньшей мере около 60% идентичен одной из последовательностей РНК-активатора (tracrРНК), представленных в SEQ ID NO:431-562, на участке из меньшей мере 8 последовательных нуклеотидов. Например, один из двух комплементарных участков нуклеотидов РНК, нацеленной на ДНК, в виде одиночной молекулы (или ДНК, кодирующей участок) является по меньшей мере на около 65% идентичным, по меньшей мере на около 70% идентичным, по меньшей мере на около 75% идентичным, по меньшей мере на около 80% идентичным, по меньшей мере на около 85% идентичным, по меньшей мере на около 90% идентичным, по меньшей мере на около 95% идентичным, по меньшей мере на около 98% идентичным, по меньшей мере на около 99% идентичным или на 100% идентичным одной из последовательностей tracrРНК, приведенных в SEQ ID NO:431-562 на участке из по меньшей мере 8 последовательных нуклеотидов.

В некоторых вариантах реализации настоящего изобретения один из двух комплементарных участков нуклеотидов РНК, нацеленной на ДНК, в виде одиночной молекулы (или ДНК, кодирующей участок) на по меньшей мере около 60% идентичен одной из последовательностей РНК-нацеливающей (crРНК), представленных в SEQ ID NO:563-679 на участке из меньшей мере 8 последовательных нуклеотидов. Например, один из двух комплементарных участков нуклеотидов РНК, нацеленной на ДНК, в виде одиночной молекулы, (или ДНК, кодирующей участок) является по меньшей мере на около 65% идентичным, по меньшей мере на около 70% идентичным, по меньшей мере на около 75% идентичным, по меньшей мере на около 80% идентичным, по меньшей мере на около 85% идентичным, по меньшей мере на около 90% идентичным, по меньшей мере на около 95% идентичным, по меньшей мере на около 98% идентичным, по меньшей мере на около 99% идентичным или на 100% идентичным одной из последовательностей crРНК, приведенных в SEQ ID NO:563-679 на участке из по меньшей мере 8 последовательных нуклеотидов.

Как указано выше, "клетка-хозяин", как используют в данном документе, обозначает находящуюся *in vivo* или *in vitro* эукариотическую клетку, прокариотическую клетку (например, бактериальную или архейную клетку) или клетку из многоклеточных организмов (например, клеточную линию), которую культивируют в одноклеточном состоянии, где эукариотические или прокариотические клетки могут быть или были использованы в качестве реципиентов для нуклеиновой кислоты, а также включает потомков исходной клетки, которая была трансформирована нуклеиновой кислотой. Понятно, что потомство отдельной клетки не обязательно полностью идентично родителю по морфологии, или по геномной или общей ДНК составляющей, ввиду естественной, случайной или преднамеренной мутации. "Рекомбинантная клетка-хозяин" (также упоминается как "генетически модифицированная клетка-хозяин") является клеткой-хозяином, в которую вводят гетерологичную нуклеиновую кислоту, например, вектор экспрессии. Например, заявленная бактериальная клетка-хозяин является генетически модифицированной бактериальной клеткой-хозяином в силу того, что в подходящую бактериальную клетку-хозяина ввели экзогенную нуклеиновую кислоты (например, плазмиды или рекомбинантного вектора экспрессии), а заявленная эукариотическая клетка-хозяин является генетически модифицированной эукариотической клеткой-хозяином (например, зародышевой клеткой млекопитающего) в силу того, что в подходящую эукариотическую клетку-хозяина ввели экзогенную нуклеиновую кислоту.

Определения, приведенные в "Определения. Часть I", также являются применимыми по отношению к настоящему разделу; см. "Определения. Часть I" для дополнительного уточнения терминов.

Перед дальнейшим описанием настоящего изобретения, должно быть понятно, что это изобретение не ограничено конкретными описанными вариантами реализации настоящего изобретения, тогда как таковые могут, конечно, варьировать. Следует также иметь в виду, что терминология, используемая в настоящем документе, существует для целей описания конкретных вариантов осуществления и не предназначена для ограничения, поскольку объем настоящего изобретения будет ограничен только прилагаемой формулой изобретения.

При условии, что указан диапазон значений, следует понимать, что каждое промежуточное значение, до десятых долей единицы нижнего предела между значениями верхнего и нижнего пределов этого диапазона и любое другое указанное значение, или промежуточное значение в этом установленном диапазоне входит в объем настоящего изобретения, если из контекста явно не следует иное. Верхние и нижние пределы этих меньших диапазонов могут независимо быть включены в меньшие диапазоны, которые также входят в объем настоящего изобретения, с учетом любого конкретно исключенного предела в установленном интервале. Там, где установленный диапазон содержит один или оба предела, диапазоны, включающие либо один из них, либо оба из этих пределов, также входят в объем настоящего изобретения.

Если не определено иное, все технические и научные термины, используемые в данном документе, имеют то же значение, которое обычно понимает специалист в данной области техники, к которой при-

надлежит настоящее изобретение. Хотя любые способы и материалы, подобные или эквивалентные тем, которые описаны в данном документе, также могут быть использованы при осуществлении или при тестировании настоящего изобретения, предпочтительные способы и материалы приведены ниже. Все публикации, упомянутые в данном документе, содержатся в данном документе в виде ссылки для раскрытия и описания способов и/или материалов, для раскрытия которых и цитируют публикации.

Следует отметить, что, как используется в данном документе и в прилагаемой формуле изобретения, формы единственного числа включают соответствия во множественном числе, если из контекста явно не следует иное. Так, например, ссылка на "ферментативно неактивный полипептид Cas9" включает множество подобных полипептидов, а ссылку на "нуклеиновую кислоту-мишень" включает ссылку на одну или несколько нуклеиновых кислот-мишеней и их эквиваленты, известные специалистам в данной области техники, и пр. Также отмечают, что формула изобретения может быть составлена так, чтобы исключить любой необязательный элемент. Таким образом, это заявление призвано служить в качестве antecedentного базиса для использования такой исключительной терминологии как "исключительно", "только" и подобные, в связи с перечислением пунктов формулы изобретения или использования "негативного" признака.

Следует понимать, что определенные признаки изобретения, которые для ясности описаны в контексте отдельных вариантов реализации настоящего изобретения, также могут быть представлены в комбинации в одном варианте реализации настоящего изобретения. Наоборот, различные признаки изобретения, которые для краткости описаны в контексте одного варианта реализации настоящего изобретения, также могут быть представлены отдельно или в любой подходящей подкомбинации. Все комбинации вариантов реализации настоящего изобретения, относящиеся к настоящему изобретению, в частности, входят в объем настоящего изобретения и описаны в данном документе так же, как если бы каждая комбинация была индивидуально и явно раскрыта. Кроме того, все подкомбинации различных вариантов реализации настоящего изобретения и их элементы также специально входят в объем настоящего изобретения и описаны в данном документе таким же образом же, как если бы каждая подкомбинация была индивидуально и явно раскрыта в данном документе.

Публикации, обсуждаемые в данном документе, предоставляются исключительно для их раскрытия перед датой подачи настоящей заявки. Ничто в данном документе не должно быть истолковано как признание того, что настоящее изобретение не имеет права датировать такую публикацию задним числом в связи с более ранним созданием настоящего изобретения. Кроме того, даты публикации могут отличаться от фактических дат публикации, которые, возможно, должны быть подтверждены независимыми источниками.

## **Подробное описание**

### **Часть II**

Настоящее изобретение обеспечивает способы модулирования транскрипции нуклеиновой кислоты-мишени в клетке-хозяине. Способы обычно включают приведение в контакт нуклеиновой кислоты-мишени с ферментативно неактивным полипептидом Cas9 и РНК, нацеленной на ДНК, в виде одиночной молекулы. Способы могут быть полезными в различных областях применения, что также является предусмотренным.

Способ модуляции транскрипции по настоящему изобретению преодолевает некоторые из недостатков способов, использующих РНКи. Способ модуляции транскрипции по настоящему изобретению находит широкое применение в самых различных областях, в том числе в научных исследованиях, при открытии новых лекарственных средств (например, путем высокопроизводительного скрининга), валидации цели, промышленных использования (например, при создании сельскохозяйственной культуры; при создании микробов и т.д.), диагностического использования, терапевтического использования и способов визуализации.

#### **Способы модуляции транскрипции**

Настоящее изобретение обеспечивает способ селективной модуляции транскрипции ДНК-мишени в клетке-хозяине. Способ обычно включает а) введение в клетку-хозяина: i) РНК, нацеленной на ДНК, или нуклеиновой кислоты, содержащую нуклеотидную последовательность, кодирующую РНК, нацеленную на ДНК; и ii) варианта сайт-специфического полипептида Cas9 ("вариант полипептида Cas9") или нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодируют вариант полипептида Cas9, причем вариант полипептида Cas9 демонстрирует пониженную эндонуклеазную активность.

РНК, нацеленная на ДНК, (также упоминаемая в данном документе как "сгРНК"; или "направляющая" РНК, или "gРНК") содержит i) первый сегмент, содержащий нуклеотидную последовательность, комплементарную последовательности-мишени в ДНК-мишени; ii) второй сегмент, который взаимодействует с сайт-специфическим полипептидом; и iii) терминатор транскрипции. Первый сегмент, содержащий нуклеотидную последовательность, комплементарную последовательности-мишени в ДНК-мишени, называется в данном документе как "нацеливающий сегмент". Второй сегмент, который взаимодействует с сайт-специфическим полипептидом, также упоминается в данном документе как "белок-связывающая последовательность" или "dCas9-связывающая шпилька" или "ручка dCas9". Под "сегментом" подразуме-

мевают сегмент/часть/участок молекулы, например, непрерывный участок нуклеотидов в РНК. Определение "сегмента", если иное не определено в конкретном контексте, не ограничивается определенным количеством общих пар оснований, и может содержать участки молекул РНК, которые имеют какую-либо общую длину и могут содержать или не содержать области комплементарности с другими молекулами. РНК, нацеленная на ДНК, в соответствии с настоящим изобретением может быть отдельной молекулой РНК (отдельным полинуклеотидом РНК), которая в данном документе может соответствовать "РНК, нацеленной на ДНК, в виде одиночной молекулы", "направляющей" РНК в виде одиночной молекулы или "sgРНК". РНК, нацеленная на ДНК, в соответствии с настоящим изобретением может содержать две молекулы РНК. Термин "РНК, нацеленная на ДНК", или "gРНК" является всеобъемлющим, и обозначает как РНК, нацеленную на ДНК, в виде двух молекул, так и РНК, нацеленную на ДНК, в виде одиночной молекулы (т.е. sgРНК).

Вариант сайт-специфического полипептида Cas9 содержит i) РНК-связывающий участок, который взаимодействует с РНК, нацеленной на ДНК; и ii) участок активности, который проявляет пониженную эндонуклеазную активность.

РНК, нацеленная на ДНК, и указанный вариант полипептида Cas9 образуют комплекс в клетке-хозяине; указанный комплекс избирательно модулирует транскрипцию ДНК-мишени в клетке-хозяине.

В некоторых случаях способ модуляции транскрипции по настоящему изобретению обеспечивает селективную модуляцию (например, снижение или увеличение) нуклеиновой кислоты-мишени в клетке-хозяине. Например, "селективное" снижение транскрипции нуклеиновой кислоты-мишени снижает транскрипцию нуклеиновой кислоты-мишени на по меньшей мере около 10%, по меньшей мере около 20%, по меньшей мере около 30%, по меньшей мере около 40%, по меньшей мере около 50%, по меньшей мере около 60%, по меньшей мере около 70%, по меньшей мере около 80%, по меньшей мере около 90% или более 90%, по сравнению с уровнем транскрипции нуклеиновой кислоты-мишени в отсутствие РНК, нацеленной на ДНК/комплекса варианта полипептида Cas9. Селективное снижение транскрипции нуклеиновой кислоты-мишени снижает транскрипцию нуклеиновой кислоты-мишени, но существенно не снижает транскрипцию нецелевой нуклеиновой кислоты-мишени, например, транскрипция нецелевой нуклеиновой кислоты уменьшается, если в общем, на менее чем 10%, по сравнению с уровнем транскрипции нецелевой нуклеиновой кислоты, в отсутствие РНК, нацеленной на ДНК/варианта полипептидного комплекса Cas9.

Увеличенная транскрипция.

"Селективно" увеличенная транскрипция ДНК-мишени может увеличивать транскрипцию ДНК-мишени в по меньшей мере около 1,1 раза (например, в по меньшей мере около 1,2 раза, по меньшей мере около 1,3 раза, по меньшей мере около 1,4 раза, по меньшей мере около 1,5 раза, по меньшей мере около 1,6 раза, по меньшей мере около 1,7 раза, по меньшей мере около 1,8 раза, по меньшей мере около 1,9 раза, по меньшей мере около 2 раза, по меньшей мере около 2,5 раза, по меньшей мере, около 3 раза, по меньшей мере около 3,5 раза, по меньшей мере около 4 раза, по меньшей мере около 4,5 раза, по меньшей мере около 5 раз, по меньшей мере около 6 раз, по меньшей мере около 7 раз, по меньшей мере около 8 раз, по меньшей мере около 9 раз, по меньшей мере около 10 раз, по меньшей мере около 12 раз, по меньшей мере около 15 раз или по меньшей мере около 20 раз), по сравнению с уровнем транскрипции ДНК-мишени в отсутствие РНК, нацеленной на ДНК/варианта полипептидного комплекса Cas9. Селективное увеличение транскрипции ДНК-мишени повышает транскрипцию ДНК-мишени, но существенно не повышает транскрипцию нецелевой ДНК, например, транскрипцию нецелевой ДНК увеличивают, в общем, в менее чем около 5 раз (например, менее чем около 4 раз, менее чем около 3 раз, менее чем около 2 раз, менее чем около 1,8 раз, менее чем около 1,6 раз, менее чем около 1,4 раз, менее чем около 1,2 раз или менее чем около 1,1 раз), по сравнению с уровнем транскрипции нецелевой ДНК в отсутствие РНК, нацеленной на ДНК/варианта полипептидного комплекса Cas9.

В качестве неограничивающего примера, увеличение может быть достигнуто с помощью слияния dCas9 с гетерологичной последовательностью. Подходящие партнеры по слиянию включают, но без ограничения ими, полипептид, который обеспечивает активность, которая опосредованно увеличивает транскрипцию, действуя непосредственно на ДНК-мишень или на полипептид (например, гистон или другой ДНК-связывающий белок), ассоциированный с ДНК-мишенью. Подходящие партнеры по слиянию включают, но без ограничения ими, полипептид, обеспечивающий активность метилтрансферазы, активность деметилазы, ацетилтрансферазную активность, активность деацетилазы, киназию активность, фосфатазную активность, активность убиквитинлигазы; дезубиквитирующую активность, активность аденилирования, активность деаденилирования, активность сумоилирования, активность десумоилирования, активность рибозилирования, активность дерибозилирования, активность миристоилирования или активность демиристоилирования.

Дополнительные подходящие партнеры по слиянию включают, но без ограничения ими, полипептид, который непосредственно обеспечивает увеличение транскрипции нуклеиновой кислоты-мишени (например, активатора транскрипции или его фрагмента, белка или его фрагмента, который привлекает активатор транскрипции, регулятор транскрипции, чувствительный к небольшой молекуле/лекарственному средству и т.д.).

Неограничивающий пример заявленного способа, где используется слитый белок dCas9 для повышения транскрипции в прокариоте, включает модификацию бактериальной одноклеточной (В1Н) или двухклеточной (В2Н) системы. В системе В1Н, ДНК-связывающий домен (ВЦ) сливается с бактериальным доменом активации транскрипции (AD, например, альфа-субъединица в РНК-полимеразы *Escherichia coli* (РНК полимеразы)). Таким образом, dCas9-мишень может быть слита с гетерологичной последовательностью, содержащей AD. При условии, что слитый белок dCas9 поступает на область расположенную выше промотора (направляется туда за счет РНК, нацеленной на ДНК), AD (например, РНК полимеразы) слитого белка dCas9 рекрутирует холофермент РНК полимеразы, что приводит к активации транскрипции. В системе В2Н, ВД не является непосредственно слитым с AD; вместо этого, их взаимодействие опосредовано белок-белковым взаимодействием (например, взаимодействием GAL11P-GAL4). Для того чтобы изменить подобную систему для использования в заявленных способах, dCas9" может быть слит с первой белковой последовательностью, что обеспечивает белок-белковое взаимодействие (например, дрожжи GAL11P и/или белок GAL4) и РНКА может быть слит со второй белковой последовательностью, что завершает белок-белковое взаимодействие (например, GAL4, если GAL11P сливаются с dCas9, GAL11P, если GAL4 сливаются с dCas9 и т.д.). Аффинность связывания между GAL11P и GAL4 повышает эффективность связывания и частоту запуска транскрипции.

Неограничивающий пример заявленного способа с использованием слитого белка dCas9 для повышения транскрипции у эукариот включает слияние dCas9 с доменом активации (AD) (например, GAL4, белков активации герпеса VP16 или VP64, субъединицы ядерного фактора NF-κB p65 человека и т.п.). Чтобы сделать данную систему индуцируемой, экспрессия слитого белка dCas9 может быть контролируемой с помощью индуцируемого промотора (например, Tet-On, Tet-OFF и т.д.). РНК, нацеленная на ДНК, может быть сконструирована для нацеливания на известные элементы, отвечающие за транскрипцию (например, промоторы, энхансеры и т.д.), известных вышерасположенных активирующих последовательностей (UAS), последовательностей неизвестной или известной функции, которые, как полагают, могут контролировать экспрессию, и т.д.

Дополнительные партнеры по слиянию.

Неограничивающие примеры партнеров по слиянию для достижения увеличенной или уменьшенной транскрипции проиллюстрированы на фиг. 54 и включают домены активатора транскрипции и репрессора транскрипции (например, ассоциированный с Kruppel бокс (KRAB или SKD); домен взаимодействия Mad mSIN3 (SID); домен репрессора ERF (ERD) и т.д.). В некоторых таких случаях, слитый белок dCas9 за счет РНК, нацеленной на ДНК, направлен в определенную локализацию (т.е. последовательность) в ДНК-мишени и оказывает локус-специфическое регулирование, такое как блокирование связывания РНК-полимеразы с промотором (что селективно ингибирует функцию активатора транскрипции) и/или модифицирует состояние локального хроматина (например, при условии, что используют слитую последовательность, которая изменяет ДНК-мишень или модифицирует полипептид, ассоциированный с ДНК-мишенью). В некоторых случаях изменения являются временными (например, репрессия или активация транскрипции). В некоторых случаях изменения являются наследуемыми (например, когда осуществляют эпигенетические модификации ДНК-мишени или белкам, ассоциированным с ДНК-мишенью, например, нуклеосомных гистонов).

В некоторых вариантах реализации настоящего изобретения гетерологичная последовательность может быть слита с С-концом полипептида dCas9. В некоторых вариантах реализации настоящего изобретения гетерологичная последовательность может быть слита с N-концом полипептида dCas9. В некоторых вариантах реализации настоящего изобретения гетерологичная последовательность может быть слита с внутренним участком (т.е. участком, отличным от N- или С-конца) полипептида dCas9.

Биологические эффекты способа с использованием слитого белка-мишени dCas9 могут быть обнаружены любым удобным способом (например, анализами экспрессии генов; анализами на основе хроматина, например, иммунопреципитацией хроматина (ChiP), анализа хроматина *in vivo* (CiA) и т.д.; и подобными).

В некоторых случаях заявленный способ включает использование двух или более различных РНК, нацеленных на ДНК. Например, две различные РНК, нацеленные на ДНК, могут быть использованы в одной клетке-хозяине, причем две различные РНК, нацеленные на ДНК, направлены на две различные последовательности-мишени в той же самой нуклеиновой кислоте-мишени.

Таким образом, например, заявленный способ модуляции транскрипции может дополнительно включать введение в клетку-хозяина второй РНК, нацеленной на ДНК, или нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую вторую РНК, нацеленную на ДНК, причем вторая РНК, нацеленная на ДНК, содержит i) первый сегмент, содержащий нуклеотидную последовательность, комплементарную второй последовательности-мишени в ДНК-мишени; ii) второй сегмент, который взаимодействует с сайт-специфическим полипептидом; и iii) терминатор транскрипции. В некоторых случаях использование двух различных РНК, нацеленных на ДНК, направленных на две различные специфические последовательности в той же нуклеиновой кислоты-мишени обеспечивают усиленную модуляцию (например, уменьшение или увеличение) транскрипции нуклеиновой кислоты-мишени.

В другом примере, две различные РНК, нацеленные на ДНК, могут быть использованы в одной

клетке-хозяине, причем две различные РНК, нацеленные на ДНК, нацелены на две различные последовательности-мишени. Таким образом, например, заявленный способ модуляции транскрипции может дополнительно включать введение в клетку-хозяина второй РНК, нацеленной на ДНК, или нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую вторую РНК, нацеленную на ДНК, причем вторая РНК, нацеленная на ДНК, содержит i) первый сегмент, содержащий нуклеотидную последовательность, комплементарную, последовательности-мишени в по меньшей мере второй ДНК-мишени; ii) второй сегмент, который взаимодействует с сайт-специфическим полипептидом; и iii) терминатор транскрипции.

В некоторых вариантах реализации настоящего изобретения заявленная нуклеиновая кислота (например, РНК, нацеленная на ДНК, например, РНК, нацеленная на ДНК, в виде одиночной молекулы, РНК-активатор, РНК-нацеливающая и т.д.; донорный полинуклеотид; нуклеиновая кислота, кодирующая сайт-специфический модифицирующий полипептид; и т.д.) содержит модификацию или последовательность, которая обеспечивает дополнительный желательный признак (например, измененную или регулируруемую стабильность; субклеточное нацеливание; слежение, например, флуоресцентную метку; сайт связывания белка или белкового комплекса; и т.д.). Неограничивающие примеры содержат 5' кэп (например, 7-метилгуанилатный кэп (m7G)); 3' полиаденилированный хвост (т.е. 3'-поли(А) хвост); последовательность рибосвитча (например, для обеспечения регулируемой стабильности и/или регулируемой доступности для белков и/или белковых комплексов); последовательность терминатора; последовательность, которая образует дуплекс дцРНК (т.е. шпильку)); модификация или последовательность, которая направляет РНК к внутриклеточной локализации (например, к ядру, митохондрию, хлоропластам и подобной); модификация или последовательность, которую обеспечивает возможность слежения (например, прямая конъюгация с флуоресцентной "молекулой, конъюгация с фрагментом, который облегчает флуоресцентную детекцию, последовательность, которая позволяет осуществлять флуоресцентную детекцию и др.); модификация или последовательность, которая обеспечивает связывающий сайт для белков (например, белков, которые действуют на ДНК, в том числе активаторы транскрипции, репрессоры транскрипции, метилтрансферазы ДНК, деметилазы ДНК, гистонацетилтрансферазы, гистондеацетилазы и подобные); и их комбинации.

ДНК-нацеленный сегмент.

ДНК-нацеленный сегмент (или "нацеленная на ДНК последовательность") РНК, нацеленной на ДНК, ("сРНК") содержит нуклеотидную последовательность, которая комплементарна определенной последовательности в пределах ДНК-мишени (комплементарная цепь ДНК-мишени).

Другими словами, ДНК-нацеленный сегмент РНК-мишени, нацеленной на ДНК, взаимодействует с ДНК-мишенью последовательность-специфическим образом посредством гибридизации (т.е. спаривания оснований). Таким образом, нуклеотидная последовательность нацеленного на ДНК сегмента может изменяться и определяет локализацию в пределах ДНК-мишени, которая будет взаимодействовать с РНК, нацеленной на ДНК, и ДНК-мишенью. ДНК-нацеленный сегмент РНК-мишени, нацеленной на ДНК, может быть модифицирован (например, с помощью генной инженерии) чтобы он гибридизовался с любой необходимой последовательностью в пределах ДНК-мишени.

ДНК-нацеленный сегмент может иметь длину от около 12 нуклеотидов до около 100 нуклеотидов. Например, ДНК-нацеленный сегмент может иметь длину от около 12 нуклеотидов (нт) до около 80 нт, от около 12 до около 50 нт, от около 12 до около 40 нт, от около 12 до около 30 нт, от около 12 до около 25 нт, от около 12 до около 20 нт или от около 12 до около 19 нт. Например, ДНК-нацеленный сегмент может иметь длину от около 19 до около 20 нт, от около 19 до около 25 нт, от около 19 до около 30 нт, от около 19 до около 35 нт, от около 19 до около 40 нт, от около 19 до около 45 нт, от около 19 до около 50 нт, от около 19 до около 60 нт, от около 19 до около 70 нт, от около 19 до около 80 нт, от около 19 до около 90 нт, от около 19 до около 100 нт, от около 20 до около 25 нт, от около 20 до около 30 нт, от около 20 до около 35 нт, от около 20 до около 40 нт, от около 20 до около 45 нт, от около 20 до около 50 нт, от около 20 до около 60 нт, от около 20 до около 70 нт, от около 20 до около 80 нт, от около 20 до около 90 нт или от около 20 до около 100 нт.

Нуклеотидная последовательность (нацеленная на ДНК последовательность) нацеленного на ДНК сегмента, который комплементарен нуклеотидной последовательности (последовательности-мишени) ДНК-мишени, может иметь длину по меньшей мере около 12 нт. Например, нацеленная на ДНК последовательность нацеленного на ДНК сегмента, который комплементарен последовательности-мишени ДНК-мишени, может иметь длину по меньшей мере около 12 нт, по меньшей мере около 15 нт, по меньшей мере около 18 нт, по меньшей мере около 19 нт, по меньшей мере около 20 нт, по меньшей мере около 25 нт, по меньшей мере около 30 нт, по меньшей мере около 35 нт или по меньшей мере около 40 нт. Например, нацеленная на ДНК последовательность нацеленного на ДНК сегмента, который комплементарен последовательности-мишени ДНК-мишени, может иметь длину от около 12 нуклеотидов (нт) до около 80 нт, от около 12 до около 50 нт, от около 12 до около 45 нт, от около 12 до около 40 нт, от около 12 до около 35 нт, от около 12 до около 30 нт, от около 12 до около 25 нт, от около 12 до нт около 20 нт, от около 12 до около 19 нт, от около 19 до около 20 нт, от около 19 до около 25 нт, от около 19 до около 30 нт, от около 19 до около 35 нт, от около 19 до около 40 нт, от около 19 до около 45 нт, от около

19 до около 50 нт, от около 19 до около 60 нт, от около 20 до около 25 нт, от около 20 до около 30 нт, от около 20 до около 35 нт, от около 20 до около 40 нт, от около 20 до около 45 нт, от около 20 до около 50 нт или от около 20 до около 60 нт. Нуклеотидная последовательность (нацеленная на ДНК последовательность) нацеленного на ДНК сегмента, который комплементарен нуклеотидной последовательности (последовательности-мишени) ДНК-мишени, может иметь длину по меньшей мере около 12 нт.

В некоторых случаях нацеленная на ДНК последовательность нацеленного на ДНК сегмента, который комплементарен последовательности-мишени ДНК-мишени, имеет 20 нуклеотидов в длину. В некоторых случаях нацеленная на ДНК последовательность нацеленного на ДНК сегмента, который комплементарен последовательности-мишени ДНК-мишени, имеет 19 нуклеотидов в длину.

Процент комплементарности между нацеленной на ДНК последовательностью нацеленного на ДНК сегмента и последовательностью-мишенью ДНК-мишени может быть по меньшей мере 60% (например, по меньшей мере 65%, по меньшей мере 70%, по меньшей мере 75%, по меньшей мере 80%, по меньшей мере 85%, по меньшей мере 90%, по меньшей мере 95%, по меньшей мере 97%, по меньшей мере 98%, по меньшей мере 99% или 100%). В некоторых случаях процент комплементарности между нацеленной на ДНК последовательностью нацеленного на ДНК сегмента и последовательностью-мишенью ДНК-мишени составляет 100% в семи последовательных крайних на 5'-конце нуклеотидов последовательности-мишени комплементарной цепи ДНК-мишени. В некоторых случаях процент комплементарности между нацеленной на ДНК последовательностью нацеленного на ДНК сегмента и последовательности-мишени ДНК-мишени составляет по меньшей мере 60% из около 20 последовательных нуклеотидов. В некоторых случаях процент комплементарности между нацеленной на ДНК последовательностью нацеленного на ДНК сегмента и последовательностью-мишенью ДНК-мишени составляет 100% из четырнадцати последовательных крайних на 5'-конце нуклеотидов последовательности-мишени комплементарной цепи ДНК-мишени, и всего лишь 0% - в остальной части. В таком случае, нацеленная на ДНК последовательность может быть 14 нуклеотидов в длину. В некоторых случаях процент комплементарности между нацеленной на ДНК последовательностью нацеленного на ДНК сегмента и последовательностью-мишенью ДНК-мишени составляет 100% в семи последовательных крайних на 5'-конце нуклеотидов последовательности-мишени комплементарной цепи ДНК-мишени, и всего лишь - 0% в остальной. В таком случае, нацеленная на ДНК последовательность может быть 7 нуклеотидов в длину.

Белок-связывающий сегмент.

Белок-связывающий сегмент (т.е. "белок-связывающая последовательность") РНК, нацеленной на ДНК, взаимодействует с вариантом сайт-специфического полипептида. При условии, что вариант сайт-специфического полипептида Cas9, вместе с РНК, нацеленной на ДНК, связывается с ДНК-мишенью, транскрипция ДНК-мишени уменьшается.

Белок-связывающий сегмент РНК, нацеленной на ДНК, содержит два комплементарных участка нуклеотидов, которые гибридизуются друг с другом, чтобы образовать дуплекс двухцепочечной РНК (дуплекс дцРНК).

Белок-связывающий сегмент РНК, нацеленной на ДНК, по настоящему изобретению содержит два участка нуклеотидов (РНК-нацеливающий и РНК-активатор), которые комплементарны друг другу, ковалентно связаны с помощью промежуточных нуклеотидов (например, в случае РНК, нацеленной на ДНК, в виде одиночной молекулы) ("линкеры" или "линкерные нуклеотиды") и гибридизуются с образованием дуплекса двухцепочечной РНК (дуплекс дцРНК или "dCas9-связывающая шпилька") белок-связывающего сегмента, что, таким образом, образуют структуру типа "стебель - петля". Эта структура типа "стебель - петля" схематически проиллюстрирована на фиг. 39А. РНК-нацеливающая и РНК-активатор могут быть ковалентно связаны через 3' конец РНК-нацеливающей и 5'-конец РНК-активатора. Как вариант, РНК-нацеливающая и РНК-активатор могут быть ковалентно связаны через 3' конец РНК-нацеливающей и 5'-конец РНК-активатора.

Белок-связывающий сегмент может иметь длину от около 10 нуклеотидов до около 100 нуклеотидов, например, от около 10 нуклеотидов (нт) до около 20 нт, от около 20 до около 30 нт, от около 30 до около 40 нт, от около 40 до около 50 нт, от около 50 до около 60 нт, от около 60 до около 70 нт, от около 70 до около 80 нт, от около 80 до около 90 нт или от около 90 до около 100 нт. Например, белок-связывающий сегмент может иметь длину от около 15 нуклеотидов (нт) до около 80 нт, от около 15 до около 50 нт, от около 15 до около 40 нт, от около 15 до около 30 нт или от около 15 до около 25 нт.

Дуплекс дцРНК белок-связывающего сегмента может иметь длину от около 6 пар оснований (п.о.) до около 50 п.о. Например, дуплекс дцРНК белок-связывающего сегмента может иметь длину от около 6 до около 40 п.о., от около 6 до около 30 п.о., от около 6 до около 25 п.о., от около 6 до около 20 п.о., от около 6 до около 15 п.о., от около 8 до около 40 п.о., от около 8 до около 30 п.о., от около 8 п.о. до около 25 п.о., от около 8 до около 20 п.о. или от около 8 до п.о. около 15 п.о. Например, дуплекс дцРНК белок-связывающего сегмента может иметь длину от около 8 до около 10 п.о., от около 10 до около 15 п.о., от около 15 до около 18 п.о., от около 18 до около 20 п.о., от около 20 до около 25 п.о., от около 25 до около 30 п.о., от около 30 до около 35 п.о., от около 35 до около 40 п.о. или от около 40 до около 50 п.о. В некоторых вариантах реализации настоящего изобретения дуплекс дцРНК белок-связывающего сегмента имеет длину в 36 пар оснований. Процент комплементарности между нуклеотидными последовательностями

стями, которые гибридизуются с образованием дуплекса дцРНК белок-связывающего сегмента, может составлять по меньшей мере около 60%. Например, процент комплементарности между нуклеотидными последовательностями, которые гибридизуются с образованием дуплекса дцРНК белок-связывающего сегмента, может составлять по меньшей мере около 65%, по меньшей мере около 70%, по меньшей мере около 75%, по меньшей мере около 80%, по меньшей мере около 85%, по меньшей мере около 90%, по меньшей мере около 95%, по меньшей мере около 98% или по меньшей мере около 99%. В некоторых случаях процент комплементарности между нуклеотидных последовательностей, которые гибридизуются с образованием дуплекса дцРНК белок-связывающего сегмента, составляет 100%.

Линкер может иметь длину от около 3 до около 100 нуклеотидов. Например, линкер может иметь длину от около 3 нуклеотидов (нт) до около 90 нт, от около 3 до около 80 нт, от около 3 до около 70 нт, от около 3 до около 60 нт, от около 3 до около 50 нт, от около 3 до около 40 нт, от около 3 до около 30 нт, от около 3 до около 20 нт или от около 3 до около 10 нт. Например, линкер может иметь длину от около 3 до около 5 нт, от около 5 до около 10 нт, от около 10 до около 15 нт, от около 15 до около 20 нт, от около 20 до около 25 нт, от около 25 до около 30 нт, от около 30 до около 35 нт, от около 35 до около 40 нт, от около 40 до около 50 нт, от около 50 до около 60 нт, от около 60 до около 70 нт, от около 70 до около 80 нт, от около 80 до около 90 нт, или от около 90 до около 100 нт. В некоторых вариантах реализации настоящего изобретения линкер РНК, нацеленной на ДНК, имеет 4 нт.

Неограничивающие примеры нуклеотидных последовательностей, которые могут содержаться в подходящем белок-связывающем сегменте (т.е. ручка dCas9), приведены в SEQ ID NO:563-682 (например, см. фиг. 8 и 9).

В некоторых случаях подходящий белок-связывающий сегмент содержит нуклеотидную последовательность, которая отличается на 1, 2, 3, 4 или 5 нуклеотидов от любой одной из перечисленных выше последовательностей.

Последовательность контроля стабильности (например, сегмент терминатора транскрипции).

Последовательность контроля устойчивости влияет на стабильность РНК (например, РНК, нацеленной на ДНК, РНК-нацеливающей, РНК-активатора и т.д.). Одним из примеров подходящей последовательности контроля устойчивости является сегмент терминатора транскрипции (т.е. стоп-кодон транскрипции). Сегмент терминатора транскрипции РНК-мишени, нацеленной на ДНК, может иметь общую длину от около 10 нуклеотидов до около 100 нуклеотидов, например, от около 10 нуклеотидов (нт) до около 20 нт, от около 20 до около 30 нт, от около 30 до около 40 нт, от около 40 до около 50 нт, от около 50 до около 60 нт, от около 60 до около 70 нт, от около 70 до около 80 нт, от около 80 до около 90 нт или от около 90 до около 100 нт. Например, сегмент терминатора транскрипции может иметь длину от около 15 нуклеотидов (нт) до около 80 нт, от около 15 до около 50 нт, от около 15 до около 40 нт, от около 15 до около 30 нт или от около 15 до около 25 нт.

В некоторых случаях стоп-кодон терминации представляет собой тот, который является функциональным в эукариотической клетке. В некоторых случаях стоп-кодон терминации представляет собой тот, который является функциональным в прокариотической клетке.

Неограничивающие примеры нуклеотидных последовательностей, которые могут содержаться в последовательности контроля устойчивости (например, стоп-кодоне транскрипции или в каком-либо сегменте РНК, нацеленной на ДНК, для того, чтобы обеспечить повышение стабильности), содержат последовательности, приведенные в SEQ ID NO:683-696 и, например, 5'-UAAUCCACAGCCGCCAGUCCGCGGCAUUUU-5' (SEQ ID NO:795) (ро-зависимый сайт терминации trp).

Дополнительные последовательности.

В некоторых вариантах реализации настоящего изобретения РНК, нацеленная на ДНК, содержит дополнительный сегмент либо на 5'-, либо на 3'-конце. Например, подходящий дополнительный сегмент содержит 5'-кэп (например, 7-метилгуанилатный кэп (m7G)); 3'-полиаденилированный хвост (т.е. 3'-поли(А) хвост); последовательность рибосвитча (например, для обеспечения регулируемой стабильности и/или регулируемой доступности для белков и белковых комплексов); последовательность, которая образует дуплекс дцРНК' (т.е. шпильку)); модификация или последовательность, которая направляет РНК к внутриклеточной локализации (например, к ядру, митохондрии, хлоропласту и подобному); модификация или последовательность, которая обеспечивает возможность для слежения (например, прямая конъюгация с флуоресцентной молекулой, конъюгация с фрагментом, который облегчает флуоресцентную детекцию, последовательность, которая позволяет осуществлять флуоресцентную детекцию и др.); модификация или последовательность, которая обеспечивает связывающий сайт для белков (например, белков, которые действуют на ДНК, в том числе активаторы транскрипции, репрессоры транскрипции, метилтрансферазы ДНК, деметилазы ДНК, гистонацетилтрансферазы, гистондеацетилазы и подобные); модификация или последовательность, которая обеспечивает увеличенную, уменьшенную и/или управляемую стабильность; и их комбинации.

Множественные одновременно используемые РНК, нацеленные на ДНК.

В некоторых вариантах реализации настоящего изобретения несколько РНК, нацеленных на ДНК, используют одновременно, для того, чтобы одновременно модулировать транскрипцию в различных ло-

кализациях одной и той же ДНК-мишени или различных ДНК-мишенях. В некоторых вариантах реализации настоящего изобретения две или более РНК, нацеленные на ДНК, нацелены на один и тот же ген, транскрипт или локус. В некоторых вариантах реализации настоящего изобретения две или более РНК, нацеленные на ДНК, нацелены на различные не связанные локусы. В некоторых вариантах реализации настоящего изобретения две или более РНК, нацеленные на ДНК, нацелены на различные, но связанные локусы.

Поскольку РНК, нацеленные на ДНК, являются малыми и имеют прочную структуру, они могут одновременно присутствовать на одном и том же векторе экспрессии, и могут даже находиться под одним и тем же транскрипционным контролем, если это является желательным. В некоторых вариантах реализации настоящего изобретения две или более (например, 3 или более, 4 или более, 5 или более, 10 или более, 15 или более, 20 или более, 25 или более, 30 или более, 35 или более, 40 или более, 45 или более или 50 или более) РНК, нацеленных на ДНК, одновременно экспрессируются в клетке-мишени (из одного и того же или различных векторов). Экспрессированные РНК, нацеленные на ДНК, могут быть распознаны различными способами с помощью белков dCas9 из различных бактерий, таких как *S. ruogenes*, *S. thermophilus*, *L. innocua* и *N. meningitidis*.

Для того чтобы экспрессировать несколько РНК, нацеленных на ДНК, может быть использована искусственная система процессинга РНК, опосредованная эндорибонуклеазой Csy4. Несколько РНК, нацеленных на ДНК, могут быть объединены в тандемный повтор в транскрипте-предшественнике (например, экспрессированном из промотора U6) и разделены Csy4-специфической последовательностью РНК. Коэкспрессированный белок Csy4 расщепляет транскрипт-предшественник на несколько РНК, нацеленных на ДНК. Преимущества использования системы процессинга РНК включают: во-первых, отсутствие необходимости использовать несколько промоторов; во-вторых, исходя из того, что все РНК, нацеленные на ДНК, процессируются из транскрипта-предшественника, их концентрации нормализованы для схожего dCas9-связывания.

Csy4 представляет собой небольшой белок эндорибонуклеазы (РНКаза), полученный из бактерий *Pseudomonas aeruginosa*. Csy4, в частности, распознает минимальные шпильки РНК из 17-п.о. и проявляет быстрое (<1 мин) и высокоэффективное (>99,9%) расщепление РНК. В отличие от большинства РНКаз, расщепленный фрагмент РНК остается стабильным и функционально активным. Расщепление РНК на сюнове Csy4 может быть использовано в искусственную систему процессинга РНК. В этой системе шпильки РНК из 17 п.о. вставляют между несколькими фрагментами РНК, которые транскрибируют в виде транскрипта-предшественника с одного промотора. Коэкспрессия Csy4 является эффективной в генерировании отдельных фрагментов РНК.

Сайт-специфический полипептид.

Как было замечено выше, РНК-мишень, нацеленная на ДНК, и вариант сайт-специфического полипептида Cas9 образуют комплекс. РНК, нацеленная на ДНК, обеспечивает целевую специфичность комплексу тем, что содержит нуклеотидную последовательность, которая является комплементарной к последовательности в ДНК-мишени.

Вариант сайт-специфического полипептида Cas9 имеет сниженную эндодезоксирибонуклеазную активность. Например, вариант сайт-специфического полипептида Cas9, пригодного для использования в способе модуляции транскрипции по настоящему раскрытию проявляет менее около 20%, менее около 15%, менее около 10%, менее около 5%, менее около 1% или менее около 0,1% эндодезоксирибонуклеазной активности полипептида Cas9 дикого типа, например, полипептида Cas9 дикого типа, содержащего аминокислотную последовательность, как проиллюстрировано на фиг. 3 (SEQ ID NO:8). В некоторых вариантах реализации настоящего изобретения вариант сайт-специфического полипептида Cas9 по существу не имеет детектируемой эндодезоксирибонуклеазной активности. В некоторых вариантах реализации настоящего изобретения, при условии, что сайт-специфический полипептид имеет сниженную каталитическую активность (например, при условии, что белок Cas9 имеет D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986 и/или мутацию A987, например, D10A, G12A, G17A, E762A, H840A, N854A, N863A, H982A, H983A, A984A и/или D986A), полипептид может по-прежнему связываться с ДНК-мишенью сайт-специфическим образом (так как он все еще направляется к последовательности ДНК-мишени с помощью РНК, нацеленной на ДНК), пока он сохраняет способность взаимодействовать с РНК, нацеленной на ДНК.

В некоторых случаях подходящий вариант сайт-специфического полипептида Cas9 содержит аминокислотную последовательность, имеющую идентичность по меньшей мере около 75%, по меньшей мере около 80%, по меньшей мере около 85%, по меньшей мере около 90%, по меньшей мере около 95%, по меньшей мере около 99 или 100% аминокислотной последовательности с аминокислотами 7-166 или 731-1003 аминокислотной последовательности Cas9/Csn1, изображенной на фиг. 3 (SEQ ID NO:8), или с соответствующими участками любой из аминокислотных последовательностей SEQ ID NO:1-256 и 795-1346.

В некоторых случаях вариант сайт-специфического полипептида Cas9 может расщеплять комплементарную цепь ДНК-мишени, но имеет сниженную способность расщеплять некомплемментарную цепь ДНК-мишени. Например, вариант сайт-специфического полипептида Cas9 может иметь мутацию (ами-

нокислотную замену), которая снижает функцию домена RuvC (например, "домена 1" на фиг. 3). В качестве неограничивающего примера, в некоторых случаях, вариант сайт-специфического полипептида Cas9 содержит мутацию D10A (аспартат на аланин) в аминокислотной последовательности, изображенной на фиг. 3 (или соответствующая мутация любой из аминокислотных последовательностей, приведенных в SEQ ID NO:1-256 и 795-1346).

В некоторых случаях вариант сайт-специфического полипептида Cas9 может расщеплять некомплементарную цепь ДНК-мишени, но имеет сниженную способность расщеплять комплементарную цепь ДНК-мишени. Например, вариант сайт-специфического полипептида Cas9 может иметь мутацию (аминокислотную замену), которая снижает функцию домена HNH (мотивов доменов RuvC/HNH/RuvC, "домена 2" фиг. 3). В качестве неограничивающего примера, в некоторых случаях, вариант сайт-специфического полипептида Cas9 содержит мутацию H840A (гистидин на аланин в положении аминокислоты 840 в SEQ ID NO:8) или соответствующую мутацию любой из аминокислотных последовательностей, приведенных в SEQ ID NO:1-256 и 795-1346).

В некоторых случаях вариант сайт-специфического полипептида Cas9 имеет пониженную способность расщеплять и комплементарные, и некомплементарные цепи ДНК-мишени. В качестве неограничивающего примера, в некоторых случаях, вариант сайт-специфического полипептида Cas9 содержит обе мутации, D10A и H840A, аминокислотной последовательности, изображенной на фиг. 3 (или соответствующую мутацию любой из аминокислотных последовательностей, приведенных в SEQ ID NO:1-256 и 795-1346).

Другие остатки могут подвергаться мутации с целью достижения подобных эффектов (т.е. инактивирования одного или другого участка нуклеазы). В качестве неограничивающих примеров, остатки D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986 и/или A987 (или соответствующие мутации любого из белков, приведенные в SEQ ID NO:1-256 и 795-1346) могут быть изменены (т.е. заменены) (см. фиг. 3, 5, 11A и табл. 1 с получением дополнительной информации о сохранении аминокислотных остатков Cas9). Кроме того, подходящими являются мутации, отличные от замещения аланина.

В некоторых случаях вариант сайт-специфического полипептида Cas9 представляет собой слитый полипептид ("вариант слитого полипептида Cas9"), т.е. слитый полипептид, содержащий i) вариант сайт-специфического полипептида Cas9; и ii) ковалентно связанный гетерологичный полипептид (также упоминающийся как "партнер по слиянию").

Гетерологичный полипептид может проявлять активность (например, ферментативную активность), которая также будет проявляться у варианта слитого полипептида Cas9 (например, активность метилтрансферазы, активность ацетилтрансферазы, киназная активность, активность убиквитинирования и т.д.). Гетерологичная полипептидная последовательность может быть связана с другой последовательностью нуклеиновой кислоты (например, с помощью генной инженерии), чтобы получить гибридную нуклеотидную последовательность, кодирующую гибридный полипептид. В некоторых вариантах реализации настоящего изобретения вариант слитого полипептида Cas9 получают с помощью слияния варианта полипептида Cas9 с гетерологичной последовательностью, которая обеспечивает субклеточные локализации (т.е. гетерологичная последовательность представляет собой последовательность субклеточной локализации, например, клеточный сигнал внутриядерной локализации (NLS) в случае нацеливания к ядру; клеточный сигнал митохондриальной локализации в случае нацеливания к митохондрии; клеточный сигнал хлоропластной локализации в случае нацеливания к хлоропласту; сигнал удержания ER; и подобные). В некоторых вариантах реализации настоящего изобретения гетерологичная последовательность может обеспечить метку (например, гетерологичная последовательность представляет собой детектируемую метку) для простоты слежения или очистки (например, флуоресцентный белок, например, зеленый флуоресцентный белок (GFP), YFP, RFP, CFP, mCherry, tdTomato и подобные; гистидиновую метку, например, метку 6xHis; метку гемагглютинаина (HA); метку FLAG; метку Мус; и подобные). В некоторых вариантах реализации настоящего изобретения гетерологичная последовательность может обеспечивать повышенную или пониженную стабильность (т.е. гетерологичная последовательность представляет собой пептид контроля стабильности, например, дегрон, который в некоторых случаях может быть контролируемым (например, чувствительным к температуре или последовательностью дегрона, контролируемой лекарственным средством, см. ниже). В некоторых вариантах реализации настоящего изобретения гетерологичная последовательность может обеспечивать повышенную или пониженную стабильность ДНК-мишени (т.е. гетерологичная последовательность представляет собой последовательность модуляции транскрипции, например, фактор/активатор транскрипции или его фрагмент, белок или его фрагмент, который привлекает фактор/активатор транскрипции, репрессор транскрипции или его фрагмент, белок или его фрагмент, который привлекает репрессор транскрипции, регулятор транскрипции, чувствительный к небольшой молекуле/лекарственному средству и т.д.). В некоторых вариантах реализации настоящего изобретения гетерологичная последовательность может обеспечивать связывающий домен (т.е. гетерологичная последовательность представляет собой белок-связывающую последовательность, например, для того, чтобы обеспечить способность гибридного полипептида Cas9 связываться с другим представляющим интерес белком, например, ДНК или белком, модифицирующий гистон, фактор транскрипции или репрессор транскрипции, белок рекрутинга и т.д.).

Подходящие партнеры по слиянию, которые обеспечивают увеличенную или уменьшенную стабильность, включают, но без ограничения ими, последовательности дегрона. Дегроны, знакомые обычному специалисту в данной области техники, будут представлять собой аминокислотные последовательности, которые контролируют стабильность белка, частью которых они являются. Например, стабильность белка, содержащего последовательность дегрона, управляется, по меньшей мере, частично, последовательностью дегрона. В некоторых случаях подходящий дегрон является конститутивным, так что дегрон оказывает свое влияние на стабильность белка независимо от контроля условиями эксперимента (т.е. дегрон не является индуцируемым лекарственным средством или температурой и т.д.) В некоторых случаях дегрон обеспечивает вариант полипептида Cas9 с контролируемой стабильностью, таким образом, что вариант полипептида Cas9 может быть "включен" (т.е. стабилен) или "выключен" (т.е. нестабилен, деградировавший) в зависимости от желаемых условий. Например, если дегрон является чувствительным к температуре дегроном, вариант полипептида Cas9 может быть функциональным (т.е. "включенным", стабильным) при температуре ниже пороговой (например, 42, 41, 40, 39, 38, 37, 36, 35, 34, 33, 32, 31, 30°C и т.д.), но нефункциональным (т.е. "выключенным", деградированным) при температуре выше пороговой температуры. В качестве другого примера, если дегрон представляет собой дегрон, индуцируемый лекарственным средством, присутствие или отсутствие лекарственного средства может рекрутировать белок с "выключенного" (т.е. нестабильного) состояния во "включенное" (т.е. стабильное) состояние или наоборот. Иллюстративный дегрон, индуцируемый лекарственным средством, является производным от белка FKBP12. Стабильность дегрона контролируют с помощью присутствия или отсутствия небольшой молекулы, которая связывается с дегроном.

Примеры подходящих дегронов включают, но не ограничиваются ими, дегроны, контролируемые с помощью Shield-1, DHFR, ауксины и/или температуру. Неограничивающие примеры подходящих дегронов известны в данной области техники (например, Dohmen и соавт., *Science*, 1994. 263(5151): p. 1273-1276: Heat-inducible degron: a method for constructing temperature-sensitive mutants; Schoeber и соавт., *Am J Physiol Renal Physiol*. 2009 Jan; 296(1):F204-II: Conditional fast expression and function of multimeric TRPV5 channels using Shield-1; Chu и соавт., *Bioorg Med Chem Lett*. 2008 Nov 15; 18(22):5941-4: Recent progress with FKBP-derived destabilizing domains; Kanemaki, Pflugers Arch. 2012 Dec 28: Frontiers of protein expression control with conditional degrons; Yang и соавт., *Mol Cell*. 2012 Nov 30; 48(4):487-8: Titivated for destruction: the methyl degron; Barbour и соавт., *Biosci Rep*. 2013 Jan 18; 33(1): Characterization of the bipartite degron that regulates ubiquitin-independent degradation of thymidylate synthase; и Greussing и соавт., *J Vis Exp*. 2012 Nov 10;(69): Monitoring of ubiquitin-proteasome activity in living cells using a Degron (dgn)-destabilized green fluorescent protein (GFP)-based reporter protein; все из которых содержатся в настоящем описании в виде ссылки во всей своей полноте).

Примеры последовательностей дегрона были хорошо изучены и исследованы в случае как клеток, так и животных. Таким образом, слияние dCas9 с последовательностью дегрона приводит к получению "настраиваемого" и "индуцируемого" полипептида dCas9. Любая из партнеров по слиянию, описанных в данном документе, может быть использован в любой желаемой комбинации. В качестве отдельного неограничивающего примера с целью проиллюстрировать указанный момент, слитый белок dCas9 может содержать последовательность для детекции YFP, последовательность дегрона для стабильности и последовательность активатора транскрипции для увеличения транскрипции ДНК-мишени. Кроме того, количество партнеров по слиянию, которые могут быть использованы в виде слитого белка dCas9, не ограничено. В некоторых случаях слитый белок dCas9 содержит одну или несколько (например, две или более, три или более, четыре или более, пять или более) гетерологичных последовательностей.

Подходящие партнеры по слиянию включают, но без ограничения ими, полипептид, обеспечивающий активность метилтрансферазы, активность деметилазы, ацетилтрансферазную активность, активность деацетилазы, киназную активность, фосфатазную активность, активность убиквитинлигазы, деубиквитинирующую активность, активность аденилирования, активность деаденилирования, активность сумоилирования, активность десумоилирования, активность рибозилирования, активность дерибозилирования, активность миристоилирования или активность демиристоилирования, любая из которых может быть направлена на модифицирование непосредственно ДНК (например, метилирование ДНК) или модифицирование ДНК-ассоциированного полипептида (например, гистона или ДНК-связывающего белка). Дополнительные подходящие партнеры по слиянию включают, но не ограничиваются ими, пограничные элементы (например, CTCF), белки и их фрагменты, которые обеспечивают периферическое рекрутирование (например, Lamin A, Lamin B и т.д.), и стыкующиеся с белком элементы (например, FKBP/FRB, Pill/Aby1 и т.д.).

Примеры различных дополнительных подходящих партнеров по слиянию (или их фрагментов) в случае варианта сайт-специфического полипептида Cas9, включают, но не ограничиваются ими, те, которые приведены на фиг. 54.

В некоторых вариантах реализации настоящего изобретения сайт-специфический модифицирующий полипептид может быть кодон-оптимизирован. Этот тип оптимизации известен в данной области техники и подразумевает мутацию ДНК, полученной извне, для того, чтобы имитировать предпочтения кодонов предполагаемого организма-хозяина или клетки при кодировании того же белка. Таким образом,

кодона изменяют, но кодируемый белок остается неизменным. Например, если предназначенная клетка-мишень представляет собой клетку человека, оптимизированный для приближения к кодам, используемым у человека dCas9 (или вариант dCas9), может быть подходящим сайт-специфическим модифицирующим полипептидом. В качестве отличного неограничивающего примера, если предназначенная клетка-хозяин является клеткой мыши, то оптимизированный для приближения к кодам, используемым у мыши Cas9 (или вариант, например, ферментативно неактивный вариант), может быть подходящим сайт-специфическим полипептидом Cas9. В то время как оптимизация кодона не обязательна, она является допустимой и в некоторых случаях может быть предпочтительной.

Клетки-хозяева.

Способ для модулирования транскрипции по настоящему изобретению может быть использован для того, чтобы индуцировать модуляцию транскрипции в митотических или постмитотических клетках *in vivo* и/или *ex vivo* и/или *in vitro*. Поскольку РНК, нацеленная на ДНК, обеспечивает специфичность за счет гибридизации с ДНК-мишенью, митотическая и/или постмитотическая клетка может быть любой из множества клеток-хозяев, причем подходящие клетки-хозяева включают, но без ограничения ими, бактериальную клетку; клетку археи, одноклеточный эукариотический организм, клетку растения, клетку водорослей, например, *Botryococcus braunii*, *Chlamydomonas reinhardtii*, *Nannochloropsis gaditana*, *Chlorella rupeoidosa*, *Sargassum patens*, *C. agardhii* и подобные, клетку грибка, клетку животного, клетку беспозвоночного животного (например, насекомого, кишечнополостного, иглокожего, нематоды, и т.д.), эукариотического паразита (например, малярийного паразита, например, *Plasmodium falciparum*; гельминта; и т.д.); клетку позвоночного животного (например, рыбы, амфибии, рептилии, птицы, млекопитающего), клетку млекопитающего, например, клетку грызуна, клетку человека, клетку не являющегося человеком примата и т.д. Подходящие клетки-хозяева включают встречающиеся в природе клетки; генетически модифицированные клетки (например, клетки, генетически модифицированные в лаборатории, например, с помощью "вмешательства человека"); и клетки, над которыми осуществляли манипуляции *in vitro* любым способом. В некоторых случаях клетка-хозяин является изолированной.

Любой тип клеток может представлять интерес (например, стволовая клетка, например, эмбриональная стволовая (ES) клетка, индуцированная плюрипотентная стволовая (iPS) клетка, половая клетка; соматическая клетка, например, фибробласт, гематопоэтическая клетка, нейрон, мышечная клетка, костная клетка, гепатоцит, панкреатическая клетка; *in vitro* или *in vivo* эмбриональная клетка эмбриона на любой стадии, например, 1-клеточная, 2-клеточная, 4-клеточная, 8-клеточная и т.д. стадия эмбрионов данио-рерио; и т.д.). Клетки могут быть от установившихся клеточных линий или же они могут быть первичными клетками, причем "первичные клетки", "линии первичных клеток" и "первичные культуры" используют в данном документе взаимозаменяемо для обозначения клеток и клеточных культур, которые были получены от субъекта, и которым дали возможность расти *in vitro* в течение ограниченного числа пассажей, т.е. делений, культуры. Например, первичные культуры содержат культуры, которые, возможно, были пассажированы 0, 1, 2, 4, 5, 10 или 15 раз, но недостаточное количество раз, чтобы пройти стадию криза. Первичные клеточные линии могут поддерживать в течение менее чем 10 пассажей *in vitro*. Клетки-мишени во многих вариантах реализации настоящего изобретения представляют собой одноклеточные организмы или растут в культуре.

Если клетки являются первичными клетками, подобные клетки могут быть получены от индивидуума любым пригодным способом. Например, лейкоциты могут быть собраны пригодным способом с помощью афереза, лейкоцитозереза, разделения по градиенту плотности, и т.д., в то время как клетки тканей, таких как кожа, мышца, костный мозг, селезенка, печень, поджелудочная железа, легкое, кишечник, желудок и т.д., являются наиболее пригодными для сбора их с помощью биопсии. Соответствующий раствор может быть использован для диспергирования или суспендирования собранных клеток. Подобный раствор, как правило, будет сбалансированным солевым раствором, например, физиологическим раствором, фосфатно-солевой буферный раствор (PBS), сбалансированный солевой раствор Хенкса, и т.д., удобно дополняют фетальной телячьей сыворотки или другими встречающимися в природе факторами, в сочетании с приемлемым буфером при: низкой концентрации, обычно от 5-25 мМ. Пригодные буферы содержат HEPES, фосфатные буферы, лактатные буферы и т.д. Эти клетки могут быть использованы незамедлительно или они могут быть сохранены, заморожены в течение длительных периодов времени и могут быть разморожены и использованы повторно. В таких случаях, клетки будут заморожены в 10% диметилсульфоксида (DMSO), 50% сыворотки, 40% буферной среды или некотором отличном подобном растворе, который обычно используют в данной области техники для того, чтобы сохранить клетки при подобных низких температурах, и размораживают широкоизвестным в данной области техники способом для размораживания замороженных культивируемых клеток.

Введение нуклеиновой кислоты в клетку-хозяина.

РНК, нацеленная на ДНК, или нуклеиновая кислота, содержащая кодирующую ее нуклеотидную последовательность, может быть введена в клетку-хозяина с помощью любого из различных хорошо известных способов. Аналогичным образом, при условии, что заявленный способ содержит введение в клетку-хозяина нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую вариант сайт-специфического полипептида Cas9, такая нуклеиновая кислота может быть введена в клет-

ку-хозяина с помощью любого из множества хорошо известных способов.

Для введения нуклеиновой кислоты (например, экспрессирующей конструкции) в стволовую клетку или клетку-предшественника могут быть использованы способы введения нуклеиновой кислоты в клетку-хозяина, известные в данной области техники. Подходящие способы включают в себя, например, вирусную или бактериофаговую инфекцию, трансфекцию, конъюгацию, слияние протопластов, липофекцию, электропорацию, осаждение фосфатом кальция, опосредованную полиэтиленгликолем (PEG) трансфекцию, опосредованную DEAE-декстраном трансфекцию, опосредованную липосомами трансфекцию, технологию генной пушки, осаждение фосфатом кальция, прямую микроинъекцию, опосредованную наночастицами доставку нуклеиновой кислоты (см., например, Panyam и соавт., *Adv Drug Deliv Rev.* 2012 Sep 13. pii: S0169-409X(12)00283-9. doi: 10.1016/j.addr.2012.09.023) и подобные.

Нуклеиновые кислоты.

Настоящее изобретение обеспечивает выделенную нуклеиновую кислоту, содержащую нуклеотидную последовательность, кодирующую РНК, нацеленную на ДНК. В некоторых случаях нуклеиновая кислота также содержит нуклеотидную последовательность, кодирующую вариант сайт-специфического полипептида Cas9.

В некоторых вариантах реализации настоящего изобретения заявленный способ включает введение в клетку-хозяина (или популяцию клеток-хозяев) одной или нескольких нуклеиновых кислот, содержащих нуклеотидные последовательности, кодирующие РНК, нацеленную на ДНК, и/или вариант сайт-специфического модифицирующего полипептида Cas9. В некоторых вариантах реализации настоящего изобретения клетка, содержащая ДНК-мишень, находится *in vitro*. В некоторых вариантах реализации настоящего изобретения клетка, содержащая ДНК-мишень, находится *in vivo*. Подходящие нуклеиновые кислоты, содержащие нуклеотидные последовательности, кодирующие РНК, нацеленную на ДНК, и/или сайт-специфический полипептид, включают векторы экспрессии, при том что вектор экспрессии, содержащий нуклеотидную последовательность, кодирующую РНК, нацеленную на ДНК, и/или сайт-специфический полипептид, представляет собой "рекомбинантный вектор экспрессии".

В некоторых вариантах реализации настоящего изобретения рекомбинантный вектор экспрессии представляет собой вирусную конструкцию, например, рекомбинантную адено-ассоциированную вирусную конструкцию (см., например, патент США № 7078387), рекомбинантную аденовирусную конструкцию, рекомбинантную лентивирусную конструкцию, рекомбинантную ретровирусную конструкцию и т.д.

Подходящие векторы экспрессии включают, но не ограничиваются ими, вирусные векторы (например, вирусные векторы, основанные на вирусе осповакцины; полиовирусе; аденовирусе (см., например, Li и соавт., *Invest Ophthalmol Vis Sci* 35:2543 2549, 1994; Borrás и соавт., *Gene Ther* 6:515 524, 1999; Li и Davidson, *PNAS* 92:7700 7704, 1995; Sakamoto и соавт., *N 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); адено-ассоциированном вирусе (см., например, Ali и соавт., *Hum Gene Ther* 9:81 86, 1998, Flannery и соавт., *PNAS* 94:6916 6921, 1997; Bennett и соавт., *Invest Ophthalmol Vis Sci* 38:2857 2863, 1997; Jomary и соавт., *Gene Ther* 4:683 690, 1997; Rolling и соавт., *Hum Gene Ther* 10:641 648, 1999; Ali и соавт., *Hum Mol Genet* 5:591 594, 1996; Srivastava в WO 93/09239, Samulski и соавт., *J. Vir.* (1989) 63:3822-3828; Mendelson и соавт., *Virol.* (1988) 166:154-165; и Flotte и соавт., *PNAS* (1993) 90:10613-10617); SV40; вирусе простого герпеса; вирусе иммунодефицита человека (см., например, Miyoshi и соавт., *PNAS* 94:10319 23, 1997; Takahashi и соавт., *J Virol* 73:7812 7816, 1999); ретровирусном векторе (например, вирусе лейкоза мыши, вирусе некроза селезенки и векторах, полученных из ретровирусов, таких как вирус саркомы Рауса, вируса саркомы Харви, вирус лейкоза птицы, лентивирус, вирус иммунодефицита человека, миелопролиферативный вирус саркомы и вирус опухоли молочной железы); и подобные.

Многочисленные подходящие векторы экспрессии известны специалистам в данной области техники, а многие из них коммерчески доступны. Следующие векторы приведены в качестве примера; в случае эукариотических клеток-хозяев: pXT1, pSG5 (Stratagene), pSVK3, pBPV, pMSG и pSVLSV40 (Pharmacia). Тем не менее, любой другой вектор может быть использован при условии, что он совместим с клеткой-хозяином.

В зависимости от используемой системы хозяин/вектор, любой из ряда подходящих регуляторных элементов транскрипции и трансляции, в том числе конститутивные и индуцируемые промоторы, энхансеры транскрипции, терминаторы транскрипции и т.д., может быть использован в векторе экспрессии (см., например, Bitter и соавт. (1987) *Methods in Enzymology*, 153:516-544).

В некоторых вариантах реализации настоящего изобретения нуклеотидная последовательность, кодирующая РНК, нацеленную на ДНК, и/или сайт-специфический полипептид Cas9, функционально связана с регуляторным элементом, например, регуляторным элементом транскрипции, таким как промотор. Регуляторный элемент транскрипции может быть функциональным либо в эукариотической клетке, например, клетке млекопитающего; либо в прокариотической клетке (например, бактериальной клетке или клетке археи). В некоторых вариантах реализации настоящего изобретения нуклеотидная последовательность, кодирующая РНК, нацеленную на ДНК, и/или сайт-специфический полипептид Cas9, функционально связана с множеством регуляторных элементов, которые обеспечивают экспрессию нуклеотид-

ной последовательности, кодирующей РНК, нацеленную на ДНК, и/или сайт-специфический полипептид в прокариотических и эукариотических клетках.

Промотор может быть постоянно активным промотором (т.е. промотором, который постоянно находится в активном/"ВКЛ" состоянии), он может быть индуцируемым промотором (т.е. промотором, состояние которого является активным/"ON" или неактивным/"ВЫКЛ", которым управляют с помощью внешнего сигнала, например, присутствия определенной температуры, соединения или белка), он может быть пространственно ограниченным промотором (т.е. элементом управления транскрипции, энхансером и т.д.) (например, тканеспецифичный промотор, промотор, специфический к типу клеток и т.д.) и он может быть временно ограниченным промотором (т.е. промотор находится в состоянии "ВКЛ" или "ВЫКЛ" во время отдельных стадий эмбрионального развития или в течение определенных этапов биологического процесса, например, цикла фолликул волоса у мышей).

Подходящие промоторы могут быть получены из вирусов, и, следовательно, могут упоминаться как вирусные промоторы, или они могут быть получены из любого организма, в том числе прокариотических или эукариотических организмов. Подходящие промоторы могут быть использованы для управления экспрессией с помощью любой РНК-полимеразы (например, pol I, pol II, pol III). Примеры промоторов включают, но не ограничиваются ими, ранний промотор SV40, промотор длинного концевго повтора вируса опухоли молочной железы мышей (LTR); основной поздний промотор аденовируса (Ad MLP); промотор вируса простого герпеса (HSV), промотор цитомегаловируса (CMV), такой как область немедленно раннего промотора CMV (CMVIE), промотор вируса саркомы Рауса (RSV), небольшой ядерный промотор U6 человека (U6) (Miyagishi и соавт., *Nature Biotechnology* 20, 497 - 500 (2002)), усиленный промотор U6 (например, Xia и соавт., *Nucleic Acids Res.* 2003 Sep 1; 31(17)), промотор H1 человека (H1) и подобные.

Примеры индуцируемых промоторов включают, но не ограничиваются ими, промотор РНК-полимеразы T7, промотор РНК-полимеразы T3, изопропил-бета-D-тиогалактопиранозид(IPTG)-регулируемый промотор, индуцируемый лактозой промотор, промотор теплового шока, тетрациклин-регулируемый промотор (например, Tet-ON, Tet-OFF и т.д.), стероид-регулируемый промотор, металл-регулируемый промотор, промотор, регулируемый рецептором эстрогена и т.д. Индуцируемые промоторы могут регулироваться с помощью молекул, в том числе, но без ограничения ими, доксициклином; РНК-полимеразой, например, РНК-полимеразой T7; рецептором эстрогена; слиянием с рецептором эстрогена; и т.д.

В некоторых вариантах реализации настоящего изобретения промотор представляет собой пространственно ограниченный промотор (т.е. специфический к типу клеток промотор, тканеспецифичный промотор и т.д.), таким образом, что в многоклеточном организме в подмножестве специфических клеток промотор является активным (т.е. "ВКЛ"). Пространственно ограниченные промоторы также могут обозначаться как энхансеры, элементы управления транскрипции, управляющие последовательности и т.д. Может быть использован любой пригодный пространственно ограниченный промотор, а выбор подходящего промотора (например, специфический к мозгу промотор, промотор, который стимулирует экспрессию в подгруппе нейронов, промотор, который запускает экспрессию в зародышевой линии, промотор, который запускает экспрессию в легких, промотор, который запускает экспрессию в мышцах, промотор, который запускает экспрессию в островковых клетках поджелудочной железы и т.д.) зависит от организма. Например, различные пространственно ограниченные промоторы, известные в случае растений, мух, червей, млекопитающих, мышей и т.д. Таким образом, пространственно ограниченный промотор может быть использован для того, чтобы регулировать экспрессию нуклеиновой кислоты, кодирующей сайт-специфический полипептид-мишень в широком множестве различных тканей и типов клеток, в зависимости от организма. Некоторые пространственно ограниченные промоторы также временно ограничены таким образом, что промотор находится в состоянии "ВКЛ" или "ВЫКЛ" во время определенных стадий эмбрионального развития или в течение определенных этапов биологического процесса (например, цикл фолликул волоса у мышей).

С целью иллюстрации, примеры пространственно ограниченных промоторов включают, но не ограничиваются ими, нейрон-специфические промоторы, адипоцит-специфические промоторы, кардиомиоцит-специфичные промоторы, промоторы, специфичные к гладким мышцам, промоторы, специфичные к фоторецепторам и т.д. Нейрон-специфические пространственно ограниченные промоторы включают, но не ограничиваются ими, нейрон-специфический промотор энолазы (NSE) (см., например, HSENO2 EMBL, X51956); промотор декарбоксилазы ароматических аминокислот (AADC); промотор нейрофиламента (см., например, GenBank HUMNFL, L04147); промотор синапсина (см., например, GenBank HUMSYNIB, M55301); промотор thy-1 (см., например, Chen и соавт. (1987) *Cell* 51:7-19; и Llewellyn, и соавт. (2010) *Nat. Med.* 16(10):1161-1166); промотор серотониновых рецепторов (см., например, GenBank S62283); промотор тирозингидроксилазы (TH) (см., например, Oh и соавт. (2009) *Gene Ther* 16:437; Sasaoka и соавт. (1992) *Mol. Brain Res.* 16:274; Boundy и соавт. (1998) *J. Neurosci.* 18:9989; и Kaneda и соавт. (1991) *Neuron* 6:583-594); промотор GnRH (см., например, Radovick и соавт. (1991) *Proc. Natl. Acad. Sci. USA* 88:3402-3406); промотор L7 (см., например, Oberdick и соавт. (1990) *Science* 248:223-226); промотор DNMT (см., например, Bartge и соавт. (1988) *Proc. Natl. Acad. Sci. USA* 85:3648-3652); промотор энкефа-

лина (см., например, Comb и соавт. (1988) EMBO J 17:3793-3805); промотор основного белка миелина (MBP); промотор Ca<sup>2+</sup>-кальмодулин-зависимой протеинкиназы II-альфа (CamKII $\alpha$ ) (см., например, Mayford и соавт. (1996) Proc. Natl. Acad. Sci. USA 93:13250; и Casanova и соавт. (2001) Genesis 31:37); промотор энхансера CMV/фактора роста тромбоцитов- $\beta$  (см., например, Liu и соавт. (2004) Gene Therapy 11:52-60); и подобные.

Адиipoцит-специфические пространственно ограниченные промоторы включают, но не ограничиваются ими, промотор/энхансер гена aP2, например, область от -5,4 т.п.н. до +21 п.о. гена aP2 человека (см., например, Tozzo и соавт. (1997) Endocrinol. 138:1604; Ross и соавт. (1990) Proc. Natl. Acad. Sci. USA 87:9590; и Ravjani и соавт. (2005) Nat. Med. 11:797); промотор переносчика глюкозы-4 (GLUT4) (см., например, Knight и соавт. (2003) Proc. Natl. Acad. Sci. USA 100:14725); промотор транслоказы жирных кислот (FAT/CD36) (см., например, Kuriki и соавт. (2002) Biol. Pharm. Bull. 25:1476; и Sato и соавт. (2002) J. Biol. Chem. 277:15703); промотор стеароил-CoA-десатуразы-1 (SCD1) (Tabog и соавт. (1999) J. Biol. Chem. 274:20603); промотор лептина (см., например, Mason и соавт. (1998) Endocrinol. 139:1013; и Chen и соавт. (1999) Biochem. Biophys. Res. Comm. 262:187); промотор адипонектина (см., например, Kita и соавт. (2005) Biochem. Biophys. Res. Comm. 331:484; и Chakrabarti (2010) Endocrinol. 151:2408); промотор адипсина (см., например, Platt и соавт. (1989) Proc. Natl. Acad. Sci. USA 86:7490); промотор резистина (см., например, Seo и соавт. (2003) Molec. Endocrinol. 17:1522); и подобные.

Кардиомиоцит-специфические пространственно ограниченные промоторы включают, но не ограничиваются ими, контрольные последовательности, полученные из следующих генов: легкую цепь миозина-2, а-тяжелую цепь миозина, AE3, сердечный тропонин C, сердечный актин и тому подобные. Franz и соавт. (1997) Cardiovasc. Res. 35:560-566; Robbins и соавт. (1995) Ann. N.Y. Acad. Sci. 752:492-505; Linn и соавт. (1995) Circ. Res. 76:584-591; Parmacek и соавт. (1994) Mol. Cell. Biol. 14:1870-1885; Hunter и соавт. (1993) Hypertension 22:608-617; и Sartorelli и соавт. (1992) Proc. Natl. Acad. Sci. USA 89:4047-4051.

Специфические к гладким мышцам пространственно ограниченные промоторы включают, но не ограничиваются ими, промотор SM22 $\alpha$  (см., например, Akyurek и соавт. (2000) Mol. Med. 6:983; и патент США № 7169874); промотор смутелина (см., например, WO 2001/018048); промотор  $\alpha$ -гладкомышечного актина; и подобные. Например, область в 0,4 т.п.н. промотора SM22 $\alpha$ , в пределах которой лежат два элемента CARG, были показаны для того, чтобы опосредовать экспрессию, специфическую к гладкомышечным клеткам сосудов (см., например, Kim, и соавт. (1997) Mol. Cell. Biol. 17, 2266-2278; Li, и соавт., (1996) J. Cell Biol. 132, 849-859; и Moessler, и соавт. (1996) Development 122, 2415-2425).

Специфические к фоторецепторам пространственно ограниченные промоторы включают, но не ограничиваются ими, промотор родопсина; промотор родопсинкиназа (Young и соавт. (2003) Ophthalmol. Vis. Sci. 44:4076); промотор гена бетафосфодиэстеразы (Nicoud и соавт. (2007) J. Gene Med. 9:1015); промотор гена пигментного ретинита (Nicoud и соавт. (2007) ранее); промотор гена межфоторецепторного ретинол-связывающего белка (IRBP) (Nicoud и соавт. (2007) ранее); промотор гена IRBP (Yokoyama и соавт. (1992) Exp Eye Res. 55:225); и подобные.

#### Библиотеки.

Настоящее изобретение обеспечивает библиотеку РНК, нацеленных на ДНК. Настоящее изобретение обеспечивает библиотеку нуклеиновых кислот, содержащих нуклеотиды, кодирующие РНК, нацеленные на ДНК. Библиотека нуклеиновых кислот, содержащая нуклеотиды, кодирующие РНК, нацеленные на ДНК, может включать библиотеку рекомбинантных векторов экспрессии, содержащих нуклеотиды, кодирующие РНК, нацеленные на ДНК.

Библиотека по изобретению может содержать от около 10 отдельных членов до 1012 отдельных членов; например, целевая библиотека может содержать от около 10 отдельных членов до 102 отдельных членов, от около 102 отдельных членов до около 103 отдельных членов, от около 103 отдельных членов до около 105 отдельных членов, от около 105 отдельных членов до около 107 отдельных членов, от около 107 отдельных членов до около 109 отдельных членов или от около 109 отдельных членов до около 1012 отдельных членов.

"Отдельный член" библиотеки по изобретению отличается от других членов библиотеки нуклеотидной последовательностью нацеленного на ДНК сегмента РНК, нацеленной на ДНК. Таким образом, например, каждый отдельный член библиотеки по изобретению может содержать такую же, или по существу такую же, нуклеотидную последовательность белок-связывающего сегмента, как и все другие члены библиотеки; и может содержать такую же или по существу такую же нуклеотидную последовательность сегмента терминации транскрипции, как и все другие члены библиотеки; но отличается от других членов библиотеки нуклеотидной последовательностью нацеленного на ДНК сегмента РНК, нацеленной на ДНК. Таким образом, библиотека может содержать элементы, которые связываются с различными нуклеиновыми кислотами-мишенями.

#### Полезность.

Способ модуляции транскрипции в соответствии с настоящим изобретением находит применение в различных областях применения, что также предусматривается. Применения включает применения для научных исследований; диагностические применения; промышленные применения; и применения для

лечения.

Применения для научных исследований содержат, например, определение эффекта уменьшения или увеличения транскрипции нуклеиновой кислоты-мишени на, например, развитие, обмен веществ, экспрессию последующих генов и подобные.

Геномный анализ с высокой производительностью может быть осуществлен с использованием заявленного способа модуляции транскрипции, в котором нужно изменять только ДНК-нацеленный сегмент РНК, нацеленной на ДНК, в то время как белок-связывающий сегмент и сегмент терминации может (в некоторых случаях) оставаться постоянным. Библиотека (например, библиотека по изобретению), содержащая множество нуклеиновых кислот, используемая в геномном анализе, содержит: промотор, функционально связанный с нуклеотидной последовательности, кодирующей РНК, нацеленную на ДНК, где каждая нуклеиновая кислота будет содержать различные ДНК-нацеленные сегменты, стандартный белок-связывающий сегмент и стандартный сегмент терминации транскрипции. Чип может содержать более  $5 \times 10^4$  уникальных РНК, нацеленных на ДНК". Все это может быть использовано для масштабного фенотипирования, картирование ген-функция и в метагеномном анализе.

Заявленные способы, раскрытые в данном документе, находят применение в области метаболической инженерии. Поскольку уровни транскрипции могут эффективно и прогнозируемо контролироваться с помощью создания соответствующей РНК, нацеленной на ДНК, как описано в данном документе, активность метаболических путей (например, биосинтетических путей) можно точно контролировать и настраивать с помощью регулирования уровня специфических ферментов (например, с помощью увеличенной или уменьшенной транскрипции) в представляющем интерес пути метаболизма. Представляющие интерес пути метаболизма включают те, которые используют для химического производства (химических продуктов тонкого органического синтеза, топлива, антибиотиков, токсинов, агонистов, антагонистов и т.п.) и/или производства лекарственных средств.

Представляющие интерес биосинтетические пути метаболизма включают, но не ограничиваются ими, (1) мевалонатный путь метаболизма (например, путь метаболизма HMG-CoA-редуктазы) (преобразует ацетил-КоА в диметилаллилпирофосфат (DMAPP) и изопентенилпирофосфат (IPP), которые используют для биосинтеза широкого разнообразия биомолекул, в том числе терпеноидов/изопреноидов), (2) немевалонатный путь метаболизма (например, "путь метаболизма 2-С-метил-D-эритритол-4-фосфат/1-дезоксид-Д-ксилозулозо-5-фосфата", "путь метаболизма MEP/DOXP" или "путь метаболизма DXR") (также получают DMAPP и IPP, вместо того, чтобы с помощью преобразования пирувата и глицеральдегид-3-фосфата в DMAPP и IPP по альтернативному пути метаболизма в мевалонатный путь метаболизма), (3) поликетидный синтетический путь метаболизма (получают множество поликетидов с помощью различных ферментов поликетидсинтаз. Поликетиды содержат встречающиеся в природе небольшие молекулы, используемые для химиотерапии (например, тетрациклин и макролиды), а промышленно важные поликетиды содержат рапамицин (иммунодепрессант), эритромицин (антибиотик), лова-статин (гиполипидемический препарат) и эпотилон В (противораковое лекарственное средство)), (4) пути метаболизма жирных кислот, (5) путь метаболизма ДАНР (3-дезоксид-Д-арабино-гептулозонат-7-фосфата), (6) пути метаболизма, в которых получают потенциальные биотоплива (такие как короткоцепочечные спирты и алкан, сложные метиловые эфиры жирных кислот и жирные спирты, изопреноиды и т.д.) и т.д.

Сети и каскады.

Способы, раскрытые в данном документе, могут быть использованы для создания интегральных схем (т.е. каскада или каскадов) управления. Например, РНК-мишень, нацеленная на ДНК/вариант сайт-специфического полипептида Cas9, может быть использована для управления (т.е. модулирования, например, увеличения, уменьшения) экспрессией другой РНК, нацеленной на ДНК, или другого варианта сайт-специфического полипептида Cas9. Например, первая РНК, нацеленная на ДНК, может быть предназначена для модуляции транскрипции второго гибридного полипептида dCas9 с функцией, которая отличается от первого варианта сайт-специфического полипептида Cas9 (например, активность метилтрансферазы, диметилазная активность, активность ацетилтрансферазы, деацетилазная активность и т.д.). Кроме того, поскольку различные белки dCas9 (например, полученные от различных видов) могут потребовать различную ручку Cas9 (т.е. белок-связывающего сегмента), второй гибридный полипептид dCas9 может быть получен из видов, отличных от первого полипептида dCas9, приведенного выше. Таким образом, в некоторых случаях, второй гибридный полипептид dCas9 может быть выбран таким образом, что он не может взаимодействовать с первой РНК, нацеленной на ДНК. В некоторых случаях второй гибридный полипептид dCas9 может быть выбран таким образом, что он может взаимодействовать с первой РНК, нацеленной на ДНК. В некоторых таких случаях, активности двух (или более) белков dCas9 могут конкурировать (например, если полипептиды имеют противоположные активности) или могут суммироваться (например, если полипептиды имеют схожие или синергетические активности). Кроме того, как отмечалось выше, любой из комплексов (например, РНК, нацеленная на ДНК/полипептид dCas9) в сети могут быть предназначены для управления другими РНК, нацеленными на ДНК, или полипептидами dCas9. Поскольку РНК-мишень, нацеленная на ДНК, и вариант сайт-специфического полипептида Cas9 по изобретению может быть направлен на любую желаемую последовательность ДНК,

описанные в данном документе способы могут быть использованы для того, чтобы контролировать и регулировать экспрессию любой необходимой мишени. Интегрированные сети (т.е. каскады взаимодействий), которые могут быть созданы, изменяются в диапазоне от очень простых до очень сложных, и не имеют ограничений.

В сети, отличающейся тем, что два или более компонента (например, РНК, нацеленные на ДНК, РНК-активаторы, РНК-нацеливающие или полипептиды dCas9), которые находятся под регулирующим контролем другого комплекса РНК, нацеленной на ДНК/полипептида dCas9, уровень экспрессии одного компонента сети может влиять на уровень экспрессии (например, может увеличивать или уменьшать экспрессию) другого компонента сети. С помощью этого механизма экспрессия одного компонента может влиять на экспрессию другого компонента в той же сети, а сеть может содержать сочетание компонентов, которые увеличивают экспрессию других компонентов, а также компоненты, которые уменьшают экспрессию других компонентов. Как будет легко понятно специалисту в данной области техники, приведенные выше примеры, в соответствии с которыми уровень экспрессии одного компонента может влиять на уровень экспрессии одного или нескольких различных компонентов, представлены в иллюстративных целях и не являются ограничивающими. Дополнительный уровень сложности может быть обязательно введен в сети, при условии, что один или несколько компонентов являются модифицированными (как описано выше) для того, чтобы им можно было манипулировать (т.е. под контролем условий эксперимента, например, регулированием температуры; контролем лекарственным средством, т.е. контроль, индуцируемый лекарственным средством; контролем освещением; и т.д.).

В качестве отдельного неограничивающего примера, первая РНК, нацеленная на ДНК, может связываться с промотором второй РНК, нацеленной на ДНК, которая контролирует экспрессию целевого терапевтического/метаболического гена. В таком случае, условная экспрессия первой РНК, нацеленной на ДНК, опосредованно активирует терапевтический/метаболический ген. Каскады РНК этого типа могут быть использованы, например, для легкого преобразования репрессора в активатор, и могут быть использованы для контролирования логики или динамики экспрессии гена-мишени.

Способ модуляции транскрипции по изобретению также может быть использован для изыскания новых лекарственных средств и подтверждение мишени.

Наборы.

Настоящее изобретение обеспечивает набор для осуществления заявленного способа. Набор по изобретению содержит а) РНК, нацеленную на ДНК, по настоящему изобретению или нуклеиновую кислоту, содержащую нуклеотидную последовательность, кодирующую РНК, нацеленную на ДНК, причем РНК, нацеленная на ДНК, содержит i) первый сегмент, содержащий нуклеотидную последовательность, которая является комплементарной второй последовательности-мишени в ДНК-мишени; ii) второй сегмент, который взаимодействует с сайт-специфическим полипептидом; и iii) терминатор транскрипции; и б) буфер. В некоторых случаях нуклеиновая кислота, содержащая нуклеотидную последовательность, кодирующую РНК, нацеленную на ДНК, дополнительно содержит нуклеотидную последовательность, кодирующую вариант сайт-специфического полипептида Cas9, который проявляет эндодезоксирибонуклеазную активность, сниженную по сравнению с Cas9 дикого типа.

В некоторых случаях набор по изобретению дополнительно содержит вариант сайт-специфического полипептида Cas9, который проявляет эндодезоксирибонуклеазную активность, сниженную по сравнению с Cas9 дикого типа.

В некоторых случаях набор по изобретению дополнительно содержит нуклеиновую кислоту, содержащую нуклеотидную последовательность, кодирующую вариант сайт-специфического полипептида Cas9, который проявляет эндодезоксирибонуклеазную активность, сниженную по сравнению с Cas9 дикого типа.

Мишень может дополнительно содержать один или несколько дополнительных реагентов, причем подобные дополнительные реагенты могут быть выбраны из буфера; промывочного буфера; контрольного реагента; контрольного вектора экспрессии или полинуклеотида РНК; реагента для *in vitro* получения варианта сайт-специфического полипептида Cas9 из ДНК; и подобные. В некоторых случаях вариант сайт-специфического полипептида Cas9, который содержится в наборе по изобретению, представляет собой гибридный вариант сайт-специфического полипептида Cas9, как описано выше.

Компоненты набора могут находиться в отдельных контейнерах или могут быть объединены в одном контейнере.

В дополнение к упомянутым выше компонентам, набор по изобретению может дополнительно содержать инструкции по использованию компонентов набора при осуществлении заявленных способов. Инструкции для осуществления заявленных способов, как правило, отражены на соответствующем носителе информации. Например, инструкции могут быть, напечатаны на подложке, такой как бумага или пластик, и т.д. Таким образом, инструкции могут присутствовать в наборах в виде вставки в пакет, в маркировке контейнера набора или его компонентов (т.е. ассоциированных с упаковкой или с частью упаковки) и т.д. В других вариантах реализации настоящего изобретения инструкции присутствуют в виде электронного файла для хранения данных, присутствующего на подходящем машиночитаемом носителе данных, например, CD-ROM, дискете, флэш-диске и т.д. В еще одних вариантах реализации на-

стоящего изобретения инструкции как таковые не содержатся в наборе, предусмотрены средства для получения инструкций из удаленного источника, например, через интернет. Пример этого варианта реализации настоящего изобретения представляет собой набор, который содержит веб-адрес, на котором инструкции могут быть просмотрены и/или с которого возможно загрузить инструкции. Как и инструкции, эти способы получения инструкций записывают на подходящий носитель.

### Примеры

Следующие примеры излагают таким образом, чтобы обеспечить специалистам с обычной квалификацией в данной области техники полное раскрытие и описание того, как создать и использовать настоящее изобретение, и не предназначены для ограничения объема того, что авторы изобретения считают своим изобретением, и они не подразумевают, что все приведенные ниже эксперименты это единственно возможные или все осуществленные эксперименты. Были предприняты усилия, чтобы обеспечить точность применительно к используемым числам (например, количествам, температуре и т.д.), но некоторые экспериментальные ошибки и отклонения должны учитываться. Если не указано иное, части являются массовыми частями, молекулярная масса представляет собой среднемассовую молекулярную массу, температуру измеряют в градусах Цельсия, а давление равно или близко к атмосферному. Стандартные сокращения могут быть использованы, например, п.о.: пара(ы) оснований; т.п.н.: тысяча(и) пар нуклеотидов; пкл: пиколитр(ы); с: секунда(ы); мин: минута(ы); ч: час(ы); ак: аминокислота(ы); нт: нуклеотид(ы); в/м: внутримышечный(о); в/б: внутрибрюшинный(о); п/к: подкожный(о); и подобные.

Пример 1. Использование Cas9 для того, чтобы изменить ДНК-мишень.

Материалы и методы.

Бактериальные штаммы и условия культивирования.

*Streptococcus pyogenes*, культивировали в среде THY (Todd Hewitt Broth (ТНВ, Bacto, Becton Dickinson) с добавлением 0,2% дрожжевого экстракта (Oxoid)) или на TSA (триптиказносоевый агар, BBL, Becton Dickinson) с добавлением 3% овечьей крови, инкубировали при 37°C в атмосфере с добавлением 5% CO<sub>2</sub> без встряхивания. *Escherichia coli* культивировали в среде Luria-Bertani (LB) и агаре, инкубировали при 37°C при встряхивании. При необходимости, подходящие антибиотики добавляли в среду в следующих конечных концентрациях: ампициллин, 100 мкг/мл в случае *E.coli*; хлорамфеникол, 33 мкг/мл в случае *Escherichia coli*; канамицин, 25 мкг/мл в случае *E.coli* и 300 мкг/мл в случае *S. pyogenes*. Бактериальный рост клеток периодически контролировали путем измерения оптической плотности аликвот культуры при 620 нм с использованием микропланшетного-ридера (SLT Spectra Reader).

Трансформация бактериальных клеток.

Трансформация клеток *E.coli* плазмидной ДНК осуществляют в соответствии со стандартным протоколом метода теплового шока. Трансформацию *S. pyogenes* осуществляют таким образом, как описано выше, с некоторыми изменениями. Анализ трансформации выполняли для мониторинга влияния *in vivo* активности CRISPR/Cas на поддержание плазмиды, и осуществляли, по существу, как описано ранее. В нескольких словах, электрокомпетентные клетки *S. pyogenes* выравнивали до достижения одинаковой плотности клеток и подвергали электропорации с 500 нг плазмидной ДНК. Каждое трансформирование высевает от двух до трех раз и эксперимент осуществляют три раза, независимо с различными партиями компетентных клеток для статистического анализа. Эффективности трансформации рассчитывают как КОЕ (колониобразующие единицы) на мкг ДНК. Трансформирования контроля осуществляют со стерильной водой и вектором остова pEC85.

Манипуляции с ДНК.

Манипуляции с ДНК, в том числе подготовку образца ДНК, амплификацию, расщепление, лигирование, очистку, электрофорез на агарозном геле, осуществляют в соответствии со стандартными способами с незначительными изменениями. Плазмиды протоспейсера в случае *in vitro* расщепления и анализы трансформирования *S. pyogenes* осуществляли так, как описано ранее (4). Дополнительные плазмиды протоспейсера на основе pUC19 для анализов расщепления *in vitro* получают с помощью лигирования прогибридизовавшихся олигонуклеотидов между расщепленными сайтами EcoRI и BamHI в pUC19. Плазмида, содержащая ген GFP, была описана ранее (41). Наборы (Qiagen) используют для очистки ДНК и получения плазмиды. Мутагенез плазмиды осуществляют с использованием набора QuikChange® II XL (Stratagene) или набора сайт-специфического мутагенеза QuikChange (Agilent). Синтетические олигонуклеотиды и РНК поставляли VBC-Biotech Services, Sigma-Aldrich и Integrated DNA Technologies.

Олигонуклеотиды для матриц *in vitro* транскрипции.

Матрицы для *in vitro* транскрибируемого tracrPHK CRISPR типа II и crPHK *S. pyogenes* (1kz tracrPHK - ПЦР на хр. ДНК SF370; для crPHK - гибридизация двух олигонуклеотидов)

T7-tracrPHK (75 нт)

OLEC1521 (F 5' tracrPHK): SEQ ID NO:340

OLEC1522 (R 3' tracrPHK): SEQ ID NO:341

T7-crPHK (матрица)

OLEC2176 (F crPHK-sp1): SEQ ID NO:342

OLEC2178 (R crPHK-sp1): SEQ ID NO:343

OLEC2177 (F crPHK-sp2): SEQ ID NO:344

OLEC2179 (R crPHK-sp2): SEQ ID NO:345

Матрицы для *in vitro* транскрибируемого tracrPHK *N.meningitidis* и сконструированного crPHK-sp2 (для tracrPHK - ПЦР на хр. ДНК Z2491; для crPHK -гибридизация двух олигонуклеотидов)

T7-tracrPHK

OLEC2205 (F прогнозируемый 5'): SEQ ID NO:346

OLEC2206 (R прогнозируемый 3'): SEQ ID NO:347

T7-crPHK (матрица)

OLEC2209 (F sp2(speM) + повтор N.m.): SEQ ID NO:348

OLEC2214 (R sp2(speM) + повтор N.m.): SEQ ID NO:349

Матрицы для *in vitro* транскрибируемого tracrPHK *L. innosua* и сконструированного crPHK-sp2 (для tracrPHK - ПЦР на хр. ДНК Clp11262; для crPHK - гибридизация двух олигонуклеотидов)

T7-tracrPHK

OLEC2203 (F прогнозируемый 5'): SEQ ID NO:350

OLEC2204 (R прогнозируемый 3'): SEQ ID NO:351

T7-crPHK (матрица)

OLEC2207 (F sp2(speM) + повтор L.in.): SEQ ID NO:352

OLEC2212 (R sp2(speM) + повтор L.in.): SEQ ID NO:353

Олигонуклеотиды для построения плазмид с протоспейсером для *in vitro* и *in vivo* исследований.

Плазмиды для speM (спейсер 2 (CRISPR типа II-A, SF370; анализ профага протоспейсера Ø8232.3 от MGAS8232) *in vitro* и в *S. ruogenes* (матрица: хр. ДНК MGAS8232 или плазмиды, содержащие фрагменты speM)

pEC287

OLEC1555 (F speM): SEQ ID NO:354

OLEC1556 (R speM): SEQ ID NO:355

pEC488

OLEC2145 (F speM): SEQ ID NO:356

OLEC2146 (R speM): SEQ ID NO:357

pEC370

OLEC1593 (A22G протоспейсера 2 F pEC488): SEQ ID NO:358

OLEC1594 (A22G протоспейсера 2 R pEC488): SEQ ID NO:359

pEC371

OLEC1595 (T10C протоспейсера 2 F pEC488): SEQ ID NO:360

OLEC1596 (T10C протоспейсера 2 R pEC488): SEQ ID NO:361

pEC372

OLEC2185 (T7A протоспейсера 2 F pEC488): SEQ ID NO:362

OLEC2186 (T7A протоспейсера 2 R pEC488): SEQ ID NO:363

pEC373

OLEC2187 (A6T протоспейсера 2 F pEC488): SEQ ID NO:364

OLEC2188 (A6T протоспейсера 2 R pEC488): SEQ ID NO:365

pEC374

OLEC2235 (A5T протоспейсера 2 F pEC488): SEQ ID NO:366

OLEC2236 (A5T протоспейсера 2 R pEC488): SEQ ID NO:367

pEC375

OLEC2233 (A4T протоспейсера 2 F pEC488): SEQ ID NO:368

OLEC2234 (A4T протоспейсера 2 R pEC488): SEQ ID NO:369

pEC376

OLEC2189 (A3T протоспейсера 2 F pEC488): SEQ ID NO:370

OLEC2190 (A3T протоспейсера 2 R pEC488): SEQ ID NO:371

pEC377

OLEC2191 (PAM G1C протоспейсера 2 F pEC488): SEQ ID NO:372

OLEC2192 (PAM G1C протоспейсера 2 R pEC488): SEQ ID NO:373

pEC378

OLEC2237 (PAM GG1 протоспейсера 2 F pEC488, 2CC): SEQ ID NO:374

OLEC2238 (PAM GG1 протоспейсера 2 R pEC488, 2CC): SEQ ID NO:375

Плазмиды для SPy\_0700 (спейсер 1 (CRISPR типа II-A, SF370; анализ профага протоспейсера 0370.1 от SF370) *in vitro* и в *S. pyogenes* ((матрица: хр. ДНК SF370 или плазмиды, содержащие фрагменты SPy\_0700)

pEC489

OLEC2106 (F Spy\_0700): SEQ ID NO:376

OLEC2107 (R Spy\_0700): SEQ ID NO:377

pEC573

OLEC2941 (F PAM TG1, 2GG): SEQ ID NO:378

OLEC2942 (R PAM TG1, 2GG): SEQ ID NO:379

Олигонуклеотиды для проверки конструкций плазмиды и сайтов расщепления с помощью секвенирования

ColE1 (pEC85)

oliRN228 (секвенирование R): SEQ ID NO:380

speM (pEC287)

OLEC1557 (секвенирование F): SEQ ID NO:381

OLEC1556 (секвенирование R): SEQ ID NO:382

repDEG-pAMbeta1 (pEC85)

OLEC787 (секвенирование F): SEQ ID NO:383

Олигонуклеотиды для *in vitro* анализов расщепления

crPHK

Спейсер 1 crPHK (1-42): SEQ ID NO:384

Спейсер 2 crPHK (1-42): SEQ ID NO:385

Спейсер 4 crPHK (1-42): SEQ ID NO:386

Спейсер 2 crPHK (1-36): SEQ ID NO:387

Спейсер 2 crPHK (1-32): SEQ ID NO:388

Спейсер 2 crPHK (11-42): SEQ ID NO:389

tracrPHK

(4-89): SEQ ID NO:390

(15-89): SEQ ID NO:391

(23-89): SEQ ID NO:392

(15-53): SEQ ID NO:393

(15-44): SEQ ID NO:394

(15-36): SEQ ID NO:395

(23-53): SEQ ID NO:396

(23-48): SEQ ID NO:397

(23-44): SEQ ID NO:398

(1-26): SEQ ID NO:399

## гибридные РНК

Спейсер 1 – химера А: SEQ ID NO:400

Спейсер 1 – химера В: SEQ ID NO:401

Спейсер 2 – химера А: SEQ ID NO:402

Спейсер 2 – химера В: SEQ ID NO:403

Спейсер 4 – химера А: SEQ ID NO:404

Спейсер 4 – химера В: SEQ ID NO:405

GFP1: SEQ ID NO:406

GFP2: SEQ ID NO:407

GFP3: SEQ ID NO:408

GFP4: SEQ ID NO:409

GFP5: SEQ ID NO:410

Олигонуклеотиды ДНК в качестве субстратов для анализов расщепления (протоспейсер выделен жирным шрифтом, PAM подчеркнут)

протоспейсер 1 - комплементарный – ДТ: SEQ ID NO:411

протоспейсер 1 - некомплементарный – ДТ: SEQ ID NO:412

протоспейсер 2 - комплементарный – ДТ: SEQ ID NO:413

протоспейсер 2 - некомплементарный – ДТ: SEQ ID NO:414

протоспейсер 4 - комплементарный – ДТ: SEQ ID NO:415

протоспейсер 4 - некомплементарный – ДТ: SEQ ID NO:416

протоспейсер 2 - комплементарный – PAM1: SEQ ID NO:417

протоспейсер 2 - некомплементарный – PAM1: SEQ ID NO:418

протоспейсер 2 - комплементарный – PAM2: SEQ ID NO:419

протоспейсер 2 - некомплементарный – PAM2: SEQ ID NO:420

протоспейсер 4 - комплементарный – PAM1: SEQ ID NO:421

протоспейсер 4 - некомплементарный – PAM1: SEQ ID NO:422

протоспейсер 4 - комплементарный – PAM2: SEQ ID NO:423

протоспейсер 4 - некомплементарный – PAM2: SEQ ID NO:424

## In vitro транскрипция и очистка РНК.

РНК транскрибировали in vitro с использованием T7 Flash in vitro Transcription Kit (Epicentre, компании Illumina) и матриц ДНК, созданных с помощью ПЦР, содержащих последовательность промотора T7. РНК очищают на геле и проверяют качество перед использованием. Праймеры, используемые для получения матриц РНК из *S. pyogenes* SF370, *Listeria innocua* Clip 11262 и *Neisseria meningitidis* A Z2491, описаны выше.

## Очистка белка.

Последовательность, кодирующую Cas9 (остатки 1-1368), амплифицируют с помощью ПЦР из геномной ДНК *S. pyogenes* SF370 и встраивают в выбранный вектор экспрессии на основе pET с использованием безлигазного клонирования (LIC). Полученная слитая конструкция содержит N-концевую белковую метку, связывающую гексагистидин и мальтозу (His6-MBP), после которой расположена пептидная последовательность, содержащая сайт расщепления протеазы вируса табачного травления (TEV). Белок экспрессируют в штамме Rosetta 2 BL21 E.coli (DE3) (EMD Biosciences), который выращивают в среде 2xTY при 18°C в течение 16 ч после индукции с 0,2 мМ IPTG. Белок очищают с помощью комбинирования этапов аффинной, ионообменной и гель-хроматографии. В нескольких словах, клетки лизируют в 20 мМ Tris, pH 8,0, 500 мМ NaCl, 1 мМ TCEP (с добавлением смеси ингибиторов протеаз (Roche)) в гомогенизаторе (Avestin). Отдельными порциями очищенный лизат связывают с агарозой Ni-NTA (Qiagen). Смола тщательно промывают 20 мМ Tris, pH 8,0, 500 мМ NaCl и связанный белок элюируют 20 мМ Tris, pH 8,0, 250 мМ NaCl, 10% глицерина. Аффинную метку His6-MBP удаляют с помощью расщепления протеазой TEV, в то время как белок подвергают диализу в течение ночи против 20 мМ HEPES, pH 7,5, 150 мМ KCl, 1 мМ TCEP, 10% глицерина. Расщепленный белок Cas9 отделяют от слитой метки очисткой на 5 мл колонке SP Sepharose HiTrap (GE Life Sciences), элюируют с линейным градиентом 100 мМ - 1 М KCl. Белок дополнительно подвергают очистке с помощью гель-хроматографии на колонке Superdex 200 16/60 в 20 мМ HEPES pH 7,5, 150 мМ KCl и 1 мМ TCEP. Элюированный белок концентрируют до 8 мг/мл, быстро замораживают в жидком азоте и хранят при -80°C. Точечные мутанты Cas9 D10A, H840A и D10A/H840A получают с использованием набора сайт-специфического мутагенеза QuikChange (Agilent) и подтверждают с помощью секвенирования ДНК. Белки очищают в том же порядке, что и в случае белка Cas9 дикого типа.

Ортологи Cas9 из *Streptococcus thermophilus* (LMD-9, YP\_820832.1), *L. innocua* (Clip11262, NP\_472073.1), *Campylobacter jejuni* (подвид, *jejuni* NCTC 11168, YP002344900.1) и *N. meningitidis* (Z2491, YP\_002342100.1) экспрессируют в клетках pLysS Rosetta BL21 (DE3) (Novagen) в виде слитых белков His6-MBP (*N. meningitidis* и *C. jejuni*), His6-Thioredoxin (*L. innocua*) и His6-GST (*S. thermophilus*) и по существу очищают таким же образом, как в случае Cas9 *S. pyogenes*, но со следующими изменениями. В связи с большим количеством совместно очищаемых нуклеиновых кислот, все четыре белка Cas9 очищают с помощью дополнительного этапа очистки с гепарином на сефарозе перед гель-фильтрацией, элюируя связанный белок с линейным градиентом 100 мМ-2 М KCl. Это успешно удаляет загрязнение нуклеиновой кислоты из белков *C. jejuni*, *N. meningitidis*, и *L. innocua*, но не удается удалить совместно очищаемые нуклеиновые кислоты из препарата Cas9 *S. thermophilus*. Все белки концентрируют до 1-8 мг/мл в 20 мМ HEPES, pH 7,5, 150 мМ KCl и 1 мМ TCEP, быстро замораживают в жидком N<sub>2</sub> и хранят при -80°C.

Анализ расщепления ДНК плазмиды.

Синтетическую или транскрибируемую *in vitro* tracrPНК и crPНК предварительно гибридизуют перед реакцией с помощью нагревания до 95°C и медленного охлаждения до комнатной температуры. Нативную или расщепленную путем рестрикции линейризованную плазмидную ДНК (300 нг (~8 нМ)) инкубируют в течение 60 мин при 37°C с очищенным белком Cas9 (50-500 нМ) и дуплексом tracrPНК:crPНК (50-500 нМ, 1:1) в буфере, расщепляющем плазмиду Cas9 (20 мМ HEPES pH 7,5, 150 мМ KCl, 0,5 мМ DTT, 0,1 мМ EDTA), с или без 10 мМ MgCl<sub>2</sub>. Реакции останавливают загрузочным буфером 5X DNA, содержащим 250 мМ EDTA, продукт разделяют с помощью электрофореза на 0,8 или 1% агарозном геле и визуализируют путем окрашивания бромистым этидием. В случае анализов расщепления мутанта Cas9, реакции останавливают загрузочным буфером 5X SDS (30% глицерин, 1,2% SDS, 250 мМ EDTA) перед загрузкой на агарозный гель.

Анализ металл-зависимого расщепления.

Протоспейсер 2 плазмидной ДНК (5 нМ) инкубируют в течение 1 ч при 37°C с Cas9 (50 нМ), предварительно инкубированном с 50 нМ tracrPНК:crPНК-sp2 в буфере расщепления (20 мМ HEPES, pH 7,5, 150 мМ KCl, 0,5 мМ DTT, 0,1 мМ EDTA), с добавлением 1, 5 или 10 мМ MgCl<sub>2</sub>, 1 или 10 мМ MnCl<sub>2</sub>, CaCl<sub>2</sub>, ZnCl<sub>2</sub>, CoCl<sub>2</sub>, NiSO<sub>4</sub> или CuSO<sub>4</sub>. Реакцию останавливают с помощью добавления загрузочного буфера 5X SDS (30% глицерина, 1,2% SDS, 250 мМ EDTA), продукт разделяют с помощью электрофореза на 1% агарозном геле и визуализируют с помощью окрашивания бромистым этидием.

Однооборотный анализ.

Cas9 (25 нМ) предварительно инкубируют 15 мин при 37°C в буфере расщепления (20 мМ HEPES pH 7,5, 150 мМ KCl, 10 мМ MgCl<sub>2</sub>, 0,5 мМ DTT, 0,1 мМ EDTA) с дуплексными tracrPНК:crPНК-sp2 (25 нМ, 1:1) или обеими РНК (25 нМ), предварительно не гибридизованными, а реакцию начинают добавлением протоспейсера 2 плазмидной ДНК (5 нМ). Реакционную смесь инкубируют при 37°C. В определенные интервалы времени образцы отбирают из реакции, для остановки реакции добавляют загрузочный буфер 5X SDS (30% глицерина, 1,2% SDS, 250 мМ EDTA), а за расщеплением наблюдают с помощью электрофореза в 1% агарозном геле и окрашивают бромистым этидием. То же самое делают в случае кинетики одного оборота без предварительной инкубации Cas9 и РНК, причем протоспейсер 2 плазмидной ДНК (5 нМ) смешивают в буфере расщепления с дуплексом tracrPНК:crPНК-sp2 (25 нМ) или обеими РНК (25 нМ), предварительно не гибридизованными, а реакцию начинают добавлением Cas9 (25 нМ). Процент расщепления анализируют с помощью денситометрии и строят график зависимости среднего значения из трех независимых экспериментов в зависимости от времени. Рассчитывают данные, которые были получены с помощью нелинейного анализа регрессии, и скорости расщепления (кнабл [мин<sup>-1</sup>]).

Мультиоборотный анализ.

Cas9 (1 нМ) предварительно инкубируют в течение 15 мин при 37°C в буфере расщепления (20 мМ HEPES pH 7,5, 150 мМ KCl, 10 мМ MgCl<sub>2</sub>, 0,5 мМ DTT, 0,1 мМ EDTA) с предварительно ренатурированными tracrPНК:crPНК-sp2 (1 нМ, 1:1). Реакцию начинают с помощью добавления протоспейсера 2 плазмидной ДНК (5 нМ). В определенные интервалы времени отбирают пробы, а реакцию останавливают добавлением загрузочного буфера 5X SDS (30% глицерина, 1,2% SDS, 250 мМ EDTA). Продукт реакции расщепления разделяют с помощью электрофореза на 1% агарозном геле, окрашивают бромистым этидием, а процент расщепления анализируют с помощью денситометрии. Результаты четырех независимых экспериментов представляют в зависимости от времени (мин).

Анализ расщепления олигонуклеотида ДНК.

Олигонуклеотиды ДНК (10 пмоль) метят радиоактивной меткой путем инкубирования с 5 единицами полинуклеотидкиназы T4 (New England Biolabs) и ~3-6 пмоль (~20-40 мКи) [ $\gamma$ -<sup>32</sup>P]-АТФ (Promega) в рабочем буфере IX полинуклеотидкиназы T4 при 37°C в течение 30 мин в 50 мкл реакции. После тепловой инактивации (65°C в течение 20 мин), продукт реакции очищают на колонке Illustra MicroSpin G-25 (GE Healthcare), чтобы удалить не встроившуюся метку. Субстраты дуплекса (100 нМ) получают с помощью гибридизации меченых олигонуклеотидов с эквимольными количествами немеченого комплементарного олигонуклеотида при 95°C в течение 3 мин, с последующим медленным охлаждением до

комнатной температуры. В случае анализов расщепления *tracr*РНК и *cr*РНК гибридизуются при нагревании в течение 30 с до 95°C с последующим медленным охлаждением до комнатной температуры. Cas9 (конечная концентрация 500 нМ) предварительно инкубируют с прогибридизованным дуплексом *tracr*РНК:*cr*РНК (500 нМ) в буфере для анализа расщепления (20 мМ HEPES, pH 7,5, 100 мМ KCl, 5 мМ MgCl<sub>2</sub>, 1 мМ DTT, 5% глицерина) в общем объеме 9 мкл. Реакции инициируют добавлением 1 мкл ДНК-мишени (10 нМ) и инкубируют в течение 1 ч при 37°C. Реакции останавливают добавлением 20 мкл загружаемого красителя (5 мМ EDTA, 0,025% SDS, 5% глицерина в формамиде) и нагревают до 95°C в течение 5 мин. Продукты расщепления разделяют на 12% денатурирующем полиакриламидном геле, содержащем 7 М мочевины и визуализируют с помощью формирования изображения на люминесцентном фосфорном покрытии (Storm, GE Life Sciences). Требования тестирования РАМ в анализах расщепления (фиг. 13b) осуществляют с использованием субстратов дуплекса ДНК, которые предварительно гибридизуют и очищают на 8% нативном акриламидном геле, а затем метят радиоактивной меткой на обоих 5'-концах. Реакции организуют и анализируют так, как было описано выше.

Анализы изменения электрофоретической подвижности.

Дуплексы ДНК-мишени образуют с помощью смешивания каждой цепи (10 нмоль) в деионизированной воде, нагревая до 95°C в течение 3 мин и медленно охлаждают до комнатной температуры. Все ДНК очищают на 8% нативных гелях, содержащих 1X TBE. Полосы ДНК визуализируют с помощью UV-оттенения, вырезают и элюируют с помощью кусочков замачивающего геля в обработанной DEPC H<sub>2</sub>O. Элюированную ДНК осаждают этанолом и растворяют в обработанной DEPC H<sub>2</sub>O. Образцы ДНК метят на 5'-конце [ $\gamma$ -<sup>32</sup>P]-АТФ с использованием полинуклеотидкиназы T4 (New England Biolabs) в течение 30 мин при 37°C. РНК денатурируют теплом при 65°C в течение 20 мин, а неинкорпорированную радиоактивную метку удаляют с использованием колонки Illustra MicroSpin G-25 (GE Healthcare). Анализы связывания осуществляют в буфере, содержащем 20 мМ HEPES pH 7,5, 100 мМ KCl, 5 мМ MgCl<sub>2</sub>, 1 мМ DTT и 10% глицерина в общем объеме 10 мкл. Двухнаправленные мутанты Cas9 D10A/H840A программируют эквимольными количествами предварительно денатурированного дуплекса *tracr*РНК:*cr*РНК и титруют от 100 пМ до 1 мкМ. Меченую радиоактивной меткой ДНК добавляют до конечной концентрации 20 пМ. Образцы инкубируют в течение 1 ч при 37°C и растворяют при 4°C на 8% нативном полиакриламидном геле, содержащем 1X TBE и 5 мМ MgCl<sub>2</sub>. Гели сушат и ДНК визуализируют с помощью формирования изображения на люминесцентном фосфорном покрытии.

In silico анализ ДНК и белковых последовательностей.

Пакет Vector NTI (Invitrogen) используют для анализа последовательности ДНК (Vector NTI) и сравнительного анализа последовательностей белков (AlignX).

Виртуальное моделирование структуры РНК и совместная укладка.

Виртуальные прогнозы осуществляют с использованием алгоритмов пакета Vienna RNA (42, 43). Вторичные структуры РНК и модели совместной укладки прогнозируют с уложенной РНК и совместно уложенной РНК, соответственно, и визуализируют с помощью VARNA (44).

Результаты.

У бактерий и архей имеются опосредованные РНК адаптивные системы обороны, называемые кластерными короткими палиндромными повторами, разделенными регулярными промежутками (CRISPR)/CRISPR-ассоциированный (CAS), которые защищают организмы от вторжения вирусов и плазмид (1-3). Было продемонстрировано, что в некоторых из этих систем, зрелая *cr*РНК, которая спарена с транс-действующей *tracr*РНК (*tracr*РНК), образует структуру из двух РНК, которая направляет CRISPR-ассоциированный белок Cas9 для того, чтобы ввести двухцепочечные (DS) разрывы в ДНК-мишени. На сайтах, комплементарных последовательности направляющей *cr*РНК, нуклеазный домен HNH Cas9 расщепляет комплементарную цепочку, в то время как RuvC-подобный домен Cas9 расщепляет некомплемментарную цепочку. Когда двойная-*tracr*РНК:*cr*РНК, создают в виде единой РНК-химеры, она также направляет последовательность-специфическое расщепление дцДНК Cas9. В этих исследованиях было выявлено семейство эндонуклеаз, которые используют двойную РНК для сайт-специфического расщепления ДНК и показана возможность использования системы для РНК-программируемого редактирования генома.

Системы защиты CRISPR/Cas основаны на малых РНК для последовательность-специфического обнаружения и сайленсинга чужеродных нуклеиновых кислот. Системы CRISPR/Cas состоят из генов *cas*, организованных в оперон(ы) и набора(ов) CRISPR, состоящий из геном-специфических последовательностей (называемых спейсеры), вперемешку с одинаковыми повторами (1-3). CRISPR/Cas-опосредованный иммунитет включает три этапа. В адаптивной фазе, бактерия и архея, несущая один или несколько локусов CRISPR, отвечают на вирусную или плазмидную стимуляцию путем интеграции коротких фрагментов чужеродной последовательности (протоспейсеров) в хромосому хозяина на проксимальном конце повтора CRISPR (1-3). В фазах экспрессии и интерференции транскрипция элемента повторяющегося спейсера в молекулы предшественника РНК CRISPR (*pre-cr*РНК) с последующим ферментативным расщеплением дает короткие *cr*РНК, которые могут образовывать пары с комплементарными последовательностями протоспейсера инвазивных вирусных или плазмидных мишеней (4-11). Распозна-

вание мишени сгРНК направляет сайленсинг чужеродных последовательностей с помощью белков Cas, которые функционируют в комплексе с сгРНК (10, 12-20).

Существуют три типа систем CRISPR/Cas (21-23). Системы типа I и III имеют некоторые всеобъемлющие функции: специализированные эндонуклеазы Cas процессируют pre-сгРНК, и каждая зрелая сгРНК собирается в большой мульти-Cas белковый комплекс, способный распознавать и расщеплять нуклеиновые кислоты, комплементарные сгРНК. В отличие от этого, системы II типа процессируют preсгРНК по другому механизму, в котором транс-действующая сгРНК (tracсгРНК), комплементарная повторяющимся последовательностям в триггерах pre-сгРНК, процессируется с помощью двухцепочечной (ds) РИКазы III РНК-с специфической рибонуклеазы в присутствии белка Cas9 (ранее Csn1) (фиг. 15) (4, 24). Cas9 считается единственным белком, отвечающим на сгРНК-направленный сайленсинг чужеродной ДНК (25-27).

Нами было продемонстрировано, что в системах II типа белки Cas9 представляют собой семейство ферментов, которые требуют образования структуры, содержащей комплементарные основания, образованной между активирующей tracсгРНК и направляющей сгРНК, для того, чтобы расщеплять дцДНК-мишень. Сайт-специфическое расщепление происходит в локализациях, определенных с помощью как комплементарности спариваемых оснований между сгРНК и ДНК-мишенью протоспейсера, так и короткого мотива [упоминающегося как смежный с протоспейсером мотив (РАМ)], непосредственно соприкасающегося с комплементарной областью в ДНК-мишени. Настоящее исследование дополнительно демонстрирует то, что семейство эндонуклеаз Cas9 может быть запрограммировано с помощью одиночных молекул РНК для расщепления определенных сайтов ДНК, тем самым способствуя развитию простой и универсальной РНК-управляемой системы для получения разрывов в дцДНК, для направленного воздействия на и редактирования генома.

Cas9 представляет собой ДНК-эндонуклеазу, управляемую двумя РНК.

Как это предполагалось, Cas9, отличительный белок системы типа II, принимает участие и в созревании сгРНК, и в управляемой сгРНК - ДНК интерференции (фиг. 15) (4, 25-27). Cas9 участвует в созревании сгРНК (4), но его непосредственное участие в разрушении ДНК-мишени не было исследовано. Для того чтобы проверить, может ли, и если да то каким образом, Cas9 расщеплять ДНК-мишень, авторы использовали систему сверхэкспрессии для того, чтобы очистить белок Cas9, полученный из патогена *Streptococcus pyogenes* (фиг. 16, см. дополнительные материалы и методы), и проверили его способность расщеплять плазмидную ДНК или дуплекс олигонуклеотидов, содержащий последовательность протоспейсера, комплементарную зрелой сгРНК, и настоящую РАМ. Было обнаружено, что одна зрелая сгРНК была неспособна направлять Cas9-катализируемое расщепление плазмидной ДНК (фиг. 10А и 17А). Однако добавление tracсгРНК, которая может спариваться с повторяющимися последовательностями сгРНК и имеет важное значение для созревания сгРНК в этой системе, стимулирует Cas9 к расщеплению плазмидной ДНК (фиг. 10А и 17А). Реакция расщепления требует и магния, и присутствия последовательности сгРНК, комплементарной ДНК; сгРНК способна образовывать пару с tracсгРНК, но при наличии случайной последовательности, связывающей ДНК-мишень, не поддерживает расщепление плазмиды, катализируемое Cas9 (фиг. 10А, 17А, по сравнению сгРНК-sp2 с сгРНК-sp1, и фиг. 18А). Были получены аналогичные результаты с использованием короткого линейного двухцепочечного(дц) ДНКсубстрата (фиг. 10В и 17, В и С). Таким образом, транс-действующая tracсгРНК представляет собой небольшую некодирующую РНК с двумя важными функциями: запуск процессинга pre-сгРНК с помощью фермента РНКазы III (4), а после этого активация сгРНК-направленного расщепления ДНК с помощью Cas9.

Расщепление как плазмиды, так и короткой линейной дцДНК с помощью tracсгРНК:сгРНК-управляемого Cas9, является сайт-специфическим (фиг. 10, С-Е, и фиг. 19 А и В). Расщепление плазмидной ДНК дает тупые концы в положении на три пары оснований выше последовательности РАМ (фиг. 10С и Е и 19А и С) (26). Аналогичным образом, в пределах коротких дуплексов дцДНК, цепь ДНК, которая является комплементарной последовательности, связывающей мишень, в сгРНК (комплементарная цепь) расщепляют в положении на три пары оснований выше последовательности РАМ (фиг. 10D и Е, а также фиг. 19В и С). Некомплементарная цепь ДНК расщепляется в одном или нескольких сайтах в пределах трех-восьми пар оснований выше последовательности РАМ. Дополнительное исследование показано, что некомплементарную цепь сначала расщепляют эндонуклеолитически, а после этого укорачивают с помощью 3'-5'-экзонуклеазной активности (фиг. 18В). Скорости расщепления с помощью Cas9 в условиях одного оборота составляет от 0,3 до 1 мин<sup>-1</sup>, сопоставимые с таковыми в случае рестрикционной эндонуклеазы (фиг. 20А), в то время как инкубация комплекса Cas9-tracсгРНК:сгРНК дикого типа (ДТ) с пятикратным молярным избытком ДНК субстрата указывает на то, что Cas9, управляемый двойной-РНК, представляет собой мультиоборотный фермент (фиг. 20В). В отличие от комплекса каскада CRISPR типа I (18), Cas9 расщепляет и линейаризованные, и суперспиральные плазмиды (фиг. 10А и 11А). Таким образом, инвазивная плаزمида может, в принципе, быть расщеплена несколько раз с помощью белков Cas9, запрограммированных различными сгРНК.

Фиг. 10. (А) Cas9 программируют 42-нуклеотидной сгРНК-sp2 (сгРНК, содержащей последовательностью спейсера 2) в присутствии или в отсутствие 75-нуклеотидной tracсгРНК. Комплекс добавляют к кольцевой или линейаризованной с помощью XhoI плазмидной ДНК, содержащей последовательность,

комплементарную спейсеру 2, и функциональной PAM. crPНК-sp1, контроль специфичности; M, маркер ДНК; т.п.о., тысяч пар оснований. См. фиг. 17А. (B) Cas9 программируют с помощью crPНК-sp2 и tracrPНК (нуклеотиды 4-89). Комплекс инкубируют с двух- и одноцепочечными ДНК, несущими последовательность, комплементарную спейсеру 2 и функциональной PAM (4). Комплементарные или некомплементарные цепи ДНК метят радиоактивной меткой на 5'-конце и гибридизуют с не меченной парной цепью, нт, нуклеотиды. См. фиг. 17, B и C. (C) Секвенирование продуктов расщепления с фиг. 10А. Терминация удлинения праймера в реакции секвенирования указывает положение сайта расщепления. Выступающий конец А на 3'-конце (звездочки) представляет собой артефакт реакции секвенирования. См. фиг. 19, А и C. (D) Продукты расщепления на фиг. 10В анализируют вместе с маркерами длины, помеченными на 5'-конце, полученными из комплементарных и некомплементарных цепей дуплекса ДНК-мишени. M, маркер; P, продукт расщепления. См. фиг. 19, B и C. (E) Схематическое представление tracrPНК, crPНК-sp2 и последовательности ДНК протоспейсера 2. Представлены области crPНК, комплементарные tracrPНК (линия сверху) и ДНК протоспейсера (подчеркнутые). Последовательность PAM отмечены; сайты расщепления, отображенные в (C) и (D), представлены с помощью заполненных белым стрелок (C), заполненных черным стрелок [(D), комплементарная цепь] и черной полосы [(D), некомплементарная цепь].

Фиг. 15 иллюстрирует РНК-опосредованный иммунный путь CRISPR/Cas типа II. Этапы экспрессии и интерференции проиллюстрированы на графических материалах. Локусы CRISPR/Cas типа II состоят из оперона четырех генов, кодирующих белки Cas9, Cas1, Cas2 и Csn2, повтора CRISPR, состоящего из лидерной последовательности, за которой следуют идентичные повторы (черные прямоугольники) вперемешку с уникальными геном-специфическими спейсерами (ромбы) и последовательностью, кодирующей транс-действующую tracrPНК. В данном документе представлен локус CRISPR/Cas тип II S. ruogenes SF370 (Учетный номер NC\_002737) (4). Указаны экспериментально подтвержденные промоторы и терминаторы транскрипции этого локуса (4). Повтор CRISPR транскрибируют в виде предшественника молекулы РНК CRISPR (pre-crPНК), которая претерпевает процесс созревания, специфический для систем типа II (4). В S. ruogenes SF370, tracrPНК транскрибируют в виде двух первичных транскриптов в 171 и 89 нт в длину, которые имеют комплементарность к каждому повтору pre-crPНК. Первое событие процессинга содержит образование пар tracrPНК с pre-crPНК, образуя дуплекс РНК, который распознается и расщепляется с помощью конститутивной эндорибонуклеазы РНКазы III в присутствии белка Cas9. Опосредованное РНКазой III расщепление дуплекса РНК формирует 75-нуклеотидную процессированную tracrPНК и 66-нуклеотидные промежуточные crPНК, состоящие из центральной области, содержащей последовательность одного спейсера, окруженного участками повторяющихся последовательностей. Второе событие процессинга, опосредованное неизвестной рибонуклеазой(ами), приводит к образованию зрелой crPНК из 39-42 нуклеотидов в длину, состоящей из 3'-концевой направляющей последовательности, полученной из спейсера и 3'-концевой последовательности, полученной из повтора. После первого и второго событий процессинга, зрелая tracrPНК остается в паре со зрелыми crPНК и связывается с белком Cas9. В этом тройном комплексе, двойная структура tracrPНК:crPНК действует как "направляющая" РНК, которая направляет эндонуклеазу Cas9 к ее ДНК-мишени. Распознавание мишени с помощью комплекса Cas9-tracrPНК:crPНК инициируют с помощью сканирования инвазивной молекулы ДНК на гомологию между последовательностью протоспейсера в ДНК-мишени и последовательностью, полученной из спейсера, в crPНК. В дополнение к комплементарности протоспейсера ДНК и спейсера crPНК, направленное воздействие на ДНК требует присутствие короткого мотива (NGG, где N может быть любым нуклеотидом), примыкающего к протоспейсеру (примыкающий к протоспейсеру мотив - PAM). После образования пары между двойной-РНК и последовательностью протоспейсера, формируют R-петли, а впоследствии Cas9 вводит двухцепочечной разрыв (DSB) в ДНК. Расщепление ДНК-мишени с помощью Cas9 требует наличия двух каталитических доменов в белке. В определенном относительно PAM месте домен HNH расщепляет комплементарную цепь ДНК, в то время как RuvC-подобный домен расщепляет некомплементарную цепь.

Фиг. 16. (A) Cas9 S. ruogenes экспрессируют в E.coli в виде слитого белка, содержащего N-концевую метку His6-MBP и очищают с помощью комбинации этапов афинной, ионообменной и гель-хроматографии. Аффинную метку удаляют с помощью расщепления протеазой TEV после этапа аффинной очистки. Представлена хроматограмма конечного этапа гель-хроматографии на колонке Superdex 200 (16/60). Cas9 элюируют в виде отдельного мономерного пика, свободного от загрязняющих нуклеиновых кислот, о чем судят исходя из соотношения поглощений при 280 и 260 нм. Вставка; элюированные фракции разделяют с помощью SDS-PAGE в 10%-ном полиакриламидном геле и окрашивают с помощью SimplyBlue Safe Stain (Invitrogen). (B) анализ SDS-PAGE очищенных ортологов Cas9. Ортологи Cas9 очищают таким образом, как описано в Дополнительных материалах и методах. 2,5 мкг каждого очищенного Cas9 анализируют на 4-20% градиентном полиакриламидном геле и окрашивают с помощью SimplyBlue Safe Stain.

Фиг. 17 (также см. фиг. 10). Последовательность протоспейсера 1 происходит от SPy\_0700 SF370 (M1) S. ruogenes, мишени crPНКspl SF370 S. ruogenes (4). В данном случае последовательность протоспейсера 1 обрабатывают с помощью изменения PAM из нефункциональной последовательности (TTG) в

функциональную (TGG). Последовательность протоспейсера 4 происходит от MGAS10750\_Spy1285 MGAS10750 *S. pyogenes* (M4), мишени *crPHK-sp4* SF370 *S. pyogenes* (4). (A) Расщеплением плазмидной ДНК протоспейсера 1 управляют с помощью родственных дуплексов *tracrPHK:crPHK*. Продукты расщепления разделяют с помощью электрофореза в агарозном геле и визуализируют с помощью окрашивания бромидом этидия. М, маркер ДНК; размеры фрагментов указаны в парах оснований. (B) Расщеплением олигонуклеотида ДНК протоспейсера 1 управляют с помощью родственных дуплексов *tracrPHK:crPHK-sp1*. Продукты расщепления разделяют с помощью денатурирующих электрофореза в полиакриламидном геле и визуализируют с помощью формирования изображения на люминесцентном фосфорном покрытии. Размеры фрагментов указаны в нуклеотидах. (C) Расщеплением олигонуклеотида ДНК протоспейсера 4 управляют с помощью родственных дуплексов *tracrPHK:crPHK-sp4*. Продукты расщепления разделяют с помощью денатурирующих электрофореза в полиакриламидном геле и визуализируют с помощью формирования изображения на люминесцентном фосфорном покрытии. Размеры фрагментов указаны в нуклеотидах. (A, B, C) Эксперименты (A) осуществляют таким образом, как указано на фиг. 10A; в (B) и в (C) таким образом, как указано на фиг. 10B. (B, C) Схема взаимодействия *tracrPHK:crPHK* ДНК-мишени проиллюстрирована ниже. Области *crPHK*, комплементарные *tracrPHK* и ДНК протоспейсера, надчеркнуты и подчеркнуты соответственно. Метят последовательность РАМ.

Фиг. 18 (также см. фиг. 10). (A) Плазмидную ДНК протоспейсера 2 инкубируют с Cas9, образовавшем комплекс с *tracrPHK:crPHK-sp2*, в присутствии различных концентраций Mg<sup>2+</sup>, Mn<sup>2+</sup>, Ca<sup>2+</sup>, Zn<sup>2+</sup>, Co<sup>2+</sup>, Ni<sup>2+</sup> или Cu<sup>2+</sup>. Продукты расщепления разделяют с помощью электрофореза в агарозном геле и визуализируют с помощью окрашивания бромидом этидия. Указаны плазмидные формы. (B) Дуплекс олигонуклеотида ДНК протоспейсера 4, содержащий мотив РАМ, ренатурируют и очищают на геле перед тем, как метят радиоактивной меткой на обоих 5'-концах. Дуплекс (конечная концентрация 10 нМ) инкубируют с Cas9, программированным с помощью *tracrPHK* (23-89 нуклеотидов) и *crPHKsp4* (конечная концентрация 500 нМ, 1:1). В указанные моменты времени (мин), 10 мкл аликвоты реакции расщепления гасят формамидным буфером, содержащим 0,025% SDS и 5 мМ EDTA, и анализируют с помощью электрофореза в денатурирующем полиакриламидном геле таким образом, как указано на фиг. 10B. Указаны размеры в нуклеотидах.

Фиг. 19. (A) Картирование расщепления протоспейсера 1 плазмидной ДНК. Продукты расщепления с фиг. 17A анализируют с помощью секвенирования таким образом, как указано на фиг. 10C. Следует отметить, что выступающий конец А на 3'-конце (звездочка) представляет собой артефакт реакции секвенирования. (B) Картирование расщепления протоспейсера 4 плазмидной ДНК. Продукты расщепления с фиг. 17C анализируют с помощью денатурирующего электрофореза в полиакриламидном геле вместе с маркерами длины, помеченными на 5'-концах, и полученных из комплементарных и некомплементарных цепей дуплекса ДНК протоспейсера 4. М, маркер; Р, продукт расщепления. Дорожки 1-2: комплементарная цепь. Дорожки 3-4: некомплементарная цепь. Размеры фрагментов указаны в нуклеотидах. (C) Схематические изображения *tracrPHK*, *crPHK-sp1* и последовательности ДНК протоспейсера 1 (сверху) и *tracrPHK*, *crPHKsp4* и последовательности ДНК протоспейсера 4 (снизу). *tracrPHK:crPHK* образуют двойную структуру РНК, направленную на комплементарный протоспейсер ДНК, посредством образования пар *crPHK*-протоспейсер ДНК. Области *crPHK*, комплементарные *tracrPHK* и ДНК протоспейсера, надчеркнуты и подчеркнуты соответственно. Сайты расщепления в комплементарных и некомплементарных нитях ДНК, картированные в (A) (верх) и (B) (низ) обозначают стрелками (A и B, комплементарная цепь) и черной полосой (B, некомплементарная цепь) над последовательностями соответственно.

Фиг. 20. (A) Кинетические свойства отдельного оборота Cas9 при различных условиях предварительного отжига РНК и предварительной инкубации белка-РНК. Плазмидную ДНК протоспейсера 2 инкубируют с либо Cas9, предварительно инкубированным с предварительно прогибридизовавшимися *tracrPHK:crPHK-sp2* (○), Cas9, предварительно не инкубированным с предварительно прогибридизовавшимися *tracrPHK:crPHK-sp2* (●), Cas9, предварительно инкубированным с предварительно не ренатурированным *tracrPHK:crPHK-sp2* (◊) или Cas9, предварительно инкубированным с предварительно не прогибридизовавшимися РНК (■). Активность расщепления контролируют в зависимости от времени и анализируют с помощью электрофореза на агарозном геле с последующим окрашиванием бромидом этидия. Средний процент расщепления, полученный в трех независимых экспериментах, строят относительно времени (мин) и аппроксимируют нелинейной регрессией. Рассчитанные скорости расщепления (кнабл.) приведены в таблице. Полученные результаты свидетельствуют о том, что связывание Cas9 с РНК не является ограничивающей скоростью в условиях испытания. Указаны плазмидные формы. Полученные значения кнабл. сравнимы с таковыми в случае рестрикционных эндонуклеаз, которые, как правило, имеют порядок 1-10 в минуту (45-47). (B) Cas9 является мультиоборотной эндонуклеазой. Cas9, загруженный с соединенными в дуплекс *tracrPHK:crPHK-sp2* (1 нМ, 1:1:1 - указаны серой линии на графике), инкубируют с 5-кратным избытком нативной плазмидной ДНК протоспейсера 2. Расщепление контролируют с помощью взятия проб из реакционной смеси в определенные интервалы времени (от 0 до 120 мин) с последующим анализом на агарозном геле (вверху) и определением количества продуктов расщепления (нМ) (внизу). Приведены стандартные отклонения трех независимых экспериментов. В исследо-

ванном промежутке времени 1 нМ Cas9 способно расщеплять ~2,5 нМ плазмидной ДНК.

Каждый домен нуклеазы Cas9 расщепляет одну цепь ДНК. Cas9 содержит домены, гомологичные обоим эндонуклеазам, HNH и RuvC (фиг. 11А и 3) (21-23, 27, 28). Создают и очищают варианты Cas9, содержащие инактивирующие точечные мутации в каталитических остатках либо HNH-, либо RuvC-подобных доменов (фиг. 11А и 3) (23, 27). Инкубирование этих белков вариантного Cas9 с нативной плазмидной ДНК продемонстрировало, что белки мутантного Cas9, управляемого двойной-РНК, приводит к образованию разорванных по одной цепи кольцевых плазмид, в то время как комплекс белка Cas9-tracrPНК:crPНК ДТ дает линейный продукт ДНК (фиг. 10А и 11А и фиг. 17А и 25А). Этот результат указывает на то, что каждый из HNH и RuvC-подобных доменов Cas9 расщепляет одну цепь плазмидной ДНК. Для того чтобы определить, какую цепь ДНК-мишени расщепляют с помощью каждого каталитического домена Cas9, инкубируют мутантные комплексы Cas9-tracrPНК:crPНК с короткими субстратами дцДНК, в которых либо комплементарную, либо некомплементарную цепь метят радиоактивной меткой на ее 5'-конце. Полученные продукты расщепления указывают на то, что домен HNH Cas9 расщепляет комплементарную цепь ДНК, в то время как RuvC-подобный домен Cas9 расщепляет некомплементарную цепь ДНК (фиг. 11В и 21В).

Фиг. 11(А) (верх). Схематическое изображение структуры домена Cas9, где показаны положения мутаций домена. D10A, Asp 10→A1a10; H840A; His840→A1a840. Комплексы белков ДТ или нуклеазных мутантных белков Cas9 с tracrPНК:crPНК-sp2 анализируют на эндонуклеазную активность таким образом, как на фиг. 10А. (В) Комплексы Cas9 ДТ или мутантов нуклеазных доменов с tracrPНК и crPНК-sp2 проверяют на активность таким образом, как на фиг. 10В.

Фиг. 3 иллюстрируют аминокислотную последовательность Cas9 из *S. pyogenes* (SEQ ID NO:8). Белки Cas9/Csn1 из различных разнообразных видов имеют 2 домена, которые содержат мотивы, гомологичные и HNH, и RuvC эндонуклеазам. (А) Мотивы 1-4 (количество мотивов отмечают на левой стороне последовательности) приведены для Cas9/Csn1*S. pyogenes*. Три предполагаемых RuvC-подобных мотива (1, 2, 4) и предполагаемых мотива HNH (3) подчеркнуты. Остатки Asp10 и His840, которые заменяют на Ala, в этом исследовании выделяют звездочкой над последовательностью. Подчеркнутые остатки являются высоко консервативными среди белков Cas9 различных видов. Мутации в подчеркнутых остатках, скорее всего, повлияют на активность Cas9. Следует отметить, что в настоящем исследовании сопряжение двух нуклеазоподобных активностей иллюстрируют экспериментально (фиг. 11 и 21). (В) Домены 1 (аминокислоты 7-166) и 2 (аминокислоты 731-1003), которые содержат мотивы 1-4, проиллюстрированы для Cas9/Csn1*S. pyogenes*. Для дополнительной информации см. табл. 1 и фиг. 5.

Фиг. 21. Расщепление ДНК протоспейсера с помощью родственных мутантов tracrPНК:crPНК-направляемого Cas9 содержит мутации в HNH- или RuvC-подобном домене. (А) Расщепление протоспейсера 1 плазмидной ДНК. Эксперимент осуществляют таким образом, как на фиг. 11А. Конформации плазмидной ДНК и размеры указаны в парах оснований. (В) Расщепление олигонуклеотида ДНК протоспейсера 4. Эксперимент осуществляют таким образом, как на фиг. 11В. Указаны размеры в нуклеотидах.

Требования двойной-РНК для связывания и расщепления ДНК-мишени.

Для связывания ДНК-мишени и/или для стимулирования активности нуклеазы Cas9 ниже по ходу транскрипции от сайта распознавания мишени может потребоваться tracrPНК. Чтобы различать эти возможности, использовали анализ сдвига электрофоретической подвижности для мониторинга связывания ДНК-мишени каталитически неактивным Cas9 в присутствии или в отсутствие crPНК и/или tracrPНК. Добавление tracrPНК существенно повышает связывание ДНК-мишени с помощью Cas9, в то время как с одной Cas9, либо с Cas9-crPНК наблюдается слабое ДНК-специфическое связывание (фиг. 22). Это указывает на то, что tracrPНК требуется для распознавания ДНК-мишени, возможно, с помощью правильного ориентирования crPНК в случае взаимодействия с комплементарной цепью ДНК-мишени. Прогнозируемая вторичная структура tracrPНК:crPНК содержит область спаренных оснований между 22 нуклеотидами на 3'-конце crPНК и сегментом около 5'-конца зрелой tracrPНК (фиг. 10Е). Это взаимодействие создает структуру, в которой 20 5'-концевых нуклеотидов crPНК, последовательности которых варьируют в различных crPНК, доступны для связывания ДНК-мишени. Основная часть tracrPНК ниже области образования пары оснований crPНК является свободной и может образовать дополнительную структуру(ы) РНК и/или взаимодействовать с Cas9 или сайтом ДНК-мишени. Чтобы определить, является ли необходимой полная длина tracrPНК для сайт-специфического Cas9-катализируемого расщепления ДНК, анализировали комплексы Cas9-tracrPНК:crPНК, воссозданные с использованием полноразмерной зрелой (42-нуклеотидной) crPНК и различных усеченных форм tracrPНК, у которых отсутствовали 5'- или 3'-концевые последовательности. Способность этих комплексов к расщеплению тестировали с использованием короткой дцДНК-мишени. Существенно усеченная версия tracrPНК, содержащая нуклеотиды 23-48 нативной последовательности, способна поддерживать активное, управляемое двойной РНК, катализируемое Cas9 расщепление ДНК (фиг. 12А и С и фиг. 23А и В). Укороченные с любого конца crPНК, показали, что расщепление, катализируемое Cas9, в присутствии tracrPНК может быть инициировано с помощью crPНК, не содержащей 10 3'-концевых нуклеотидов (фиг. 12В и С). В отличие от этого, 10-нуклеотидная делеция с 5'-конца crPНК отменяет расщепление ДНК с помощью Cas9 (фиг. 12В). Также анализируют ортологи Cas9 из различных видов бактерий на их способность поддерживать расщепление

ДНК *S. pyogenes*, управляемое *tracrPНК:crPНК*. В отличие от близкородственных ортологов Cas9 *S. pyogenes*, более отдаленно связанные ортологи не являются функциональными в реакции расщепления (фиг. 24). Точно так же Cas9 *S. pyogenes*, управляемый дуплексами *tracrPНК:crPНК*, происходящими из более отдаленных систем, не являются способными эффективно расщеплять ДНК (фиг. 24). Специфичность видов управляемого двойной-РНК расщепления ДНК указывает на коэволюцию Cas9, *tracrPНК* и повтора *crPНК*, также как существование еще неизвестной структуры и/или последовательности в двойной-РНК, которая является критической для образования тройного комплекса со специфическими ортологами Cas9.

Для исследования требований к последовательности протоспейсера в случае иммунитета CRISPR/Cas типа II в бактериальных клетках, анализировали поддержание ряда плазмидных ДНК, содержащих протоспейсер с однонуклеотидной мутацией после их трансформации в *S. pyogenes* и могут ли эти плазмиды расщепляться с помощью Cas9 *in vitro*. В отличие от точечных мутаций, введенных на 5'-конце протоспейсера, мутации в области, близкой к PAM и сайтам расщепления Cas9, не допустимы *in vivo* и приводит к снижению эффективности расщепления плазмиды *in vitro* (фиг. 12D). Результаты согласуются с предыдущим отчетом об "ускользнувших" мутантах протоспейсера, выбранных в системе CRISPR типа II от *C. thermophilus in vivo* (27, 29). Кроме того, полученные результаты по поддержанию и расщеплению плазмиды указывают на то, что существует область "затравки", расположенная на 3'-конце последовательности протоспейсера, которая критична в для взаимодействия с *crPНК* и последующего расщепления с помощью Cas9. В подтверждение этого мнения, Cas9 усиливает гибридизацию комплементарной цепи ДНК с *crPНК*; это повышение было самым сильным в 3'-концевой области последовательности, специфической к *crPНК* (фиг. 25A-C). В подтверждение этого вывода, непрерывный участок из по меньшей мере 13 спаренных оснований между *crPНК* и сайтом ДНК-мишени, расположенный проксимально от PAM, является необходимым для эффективного расщепления мишени, в то время как допускается вплоть до шести расположенных подряд ошибочных спариваний в 5'-концевой области протоспейсера (фиг. 12E). Эти данные напоминают наблюдавшиеся ранее требования к затравочной последовательности, распознающей нуклеиновую кислоту-мишени в белках Argonaute (30, 31) и комплексах Cascade и CRISPR Csy (13, 14).

Фиг. 12. (А) Комплексы Cas9-*tracrPНК:crPНК* были восстановлены с использованием 42-нуклеотидной *crPНК-sp2* и процессированных конструкций *tracrPНК* и их анализировали на активность расщепления таким образом, как на фиг. 10В. (В) Cas9, запрограммированный с полноразмерными *tracrPНК* и усеченными *crPНК-sp2*, анализируют на активность таким образом, как в (А). (С) Минимальные области *tracrPНК* и *crPНК*, способные направлять Cas9-опосредованное расщепление ДНК (заштрихованная область). (D) Плазмиды, содержащие ДТ или мутантные последовательности протоспейсера 2 с указанными точечными мутациями, расщепляют *in vitro* с помощью программируемого Cas9 таким образом, как на фиг. 10А, и используют для анализов трансформирования ДТ- или *pre-crPНК*-дефицитного *S. pyogenes*. Эффективность трансформирования рассчитывают как колониеобразующие единицы (КОЕ) на мкг плазмидной ДНК. "Усы" представляют SD для трех биологических повторов. (E) Плазмиды, содержащие ДТ и мутантные вставки протоспейсера 2 с различной степенью ошибочных спариваний *crPНК*-ДНК-мишени (внизу), расщепляют *in vitro* с помощью программируемого Cas9 (вверху). Реакции расщепления дополнительно расщепляют XmnI. Фрагменты из 1880 и 800 п.о. представляют собой образованные с помощью Cas9 продукты расщепления, M, маркер ДНК.

Фиг. 22. Анализы сдвига электрофоретической подвижности выполняют с использованием дуплекса ДНК-мишени протоспейсера 4 и Cas9 (содержащего инактивирующие нуклеазный домен мутации D10A и H840), отдельно или в присутствии *crPНК-sp4*, *tracrPНК* (75 нт) или обоих. Дуплекс ДНК-мишени метят радиоактивной меткой на обоих 5'-концах. Cas9 (D10/H840A) и комплексы титруют от 1 нМ до 1 мМ. Связывание анализируют с помощью электрофореза в 8% нативном полиакриламидном геле и визуализируют с помощью формирования изображения на люминесцентном фосфорном покрытии. Обратите внимание, что в сам по себе Cas9 связывает ДНК-мишень с умеренной аффинностью. Это связывание не зависит от добавления *crPНК*, из чего можно предположить, что это представляет собой неспецифичное к последовательности взаимодействие с дцДНК. Кроме того, это взаимодействие может быть вытеснено с помощью одной *tracrPНК*, в отсутствие *crPНК*. В присутствии и *crPНК* и *tracrPНК*, связывание ДНК-мишени значительно усиливается и дает образцы с отличной электрофоретической подвижностью, что свидетельствует о специфичности распознавания ДНК-мишени.

Фиг. 23. Фрагмент *tracrPНК* охватывающий часть спаренной области *crPНК* и нижележащую область является достаточным для прямого расщепления олигонуклеотида ДНК протоспейсера с помощью Cas9. (А) Расщепление олигонуклеотида ДНК протоспейсера 1 и (В) расщепление олигонуклеотида ДНК протоспейсера 4 с помощью Cas9 направляется зрелой соответствующей *crPНК* и различными фрагментами *tracrPНК*. (А, В) Размеры указаны в нуклеотидах.

Фиг. 24. Так же, как и Cas9 из *S. pyogenes*, близкородственные ортологи Cas9 из грамм-положительной бактерии *L. innocua* и *S. thermophilus* расщепляют ДНК протоспейсера, при условии, что нацелены с помощью *tracrPНК:crPНК* из *S. pyogenes*. Тем не менее, расщепление ДНК не столь близкородственными ортологами Cas9 из грамотрицательных бактерий *C. jejuni* и *N. meningitidis* в тех же усло-

виях не наблюдается. *Spy*, *S. pyogenes* SF370 (Учетный номер NC\_002737); *Sth*, *S. thermophilus* LMD-9 (ортолог Cas9 STER\_1477; Учетный номер NC\_008532); *Lin*, *L. innocua* Clip11262 (Учетный номер NC\_003212); *Cje*, *C. jejuni* NCTC 11168 (Учетный номер NC\_002163); *Nme*, *N. meningitidis* A Z2491 (Учетный номер NC\_003116). (А) Расщепление плазмидной ДНК протоспейсера. Плазмидную ДНК протоспейсера 2 (300 нг) расщепляют с помощью различных ортологов Cas9 (500 нМ), направляемых гибридными дуплексами *tracrPНК:crPНК-sp2* (500 нМ, 1:1) из различных видов. Чтобы сконструировать дуплексы РНК, определили предположительные последовательности *tracrPНК* из *L. innocua* и *N. meningitidis* на основе ранее опубликованных данных нозерн-блоттинга (4). Дуплексы двухцепочечных гибридных РНК состоят из видоспецифичной *tracrPНК* и гетерологичной *crPНК*. Гетерологичную последовательность *crPНК* конструируют так, чтобы она на 5'-конце содержала ДНК-нацеленную последовательность *sp2 S. pyogenes* слитую на 3'-конце с *tracrPНК*-связывающей повторяющейся последовательностью *L. innocua* или *N. meningitidis*. Ортологи Cas9 из *S. thermophilus* и *L. innocua*, но не из *N. meningitidis* или *C. jejuni*, могут управляться *tracrPНК:crPНК-sp2 S. pyogenes* с целью расщепления плазмидной ДНК протоспейсера 2, хотя и с несколько сниженной эффективностью. Аналогичным образом, гибридный *tracrPНК:crPНК-sp2 L. innocua* может направлять Cas9 *S. pyogenes* на расщепление ДНК-мишени с высокой эффективностью, в то время как гибрид *tracrPНК:crPНК-sp2 N. meningitidis* стимулирует только небольшую активность расщепления ДНК с помощью Cas9 *S. pyogenes*. В качестве контроля, ортологи Cas9 *N. meningitidis* и *L. innocua* расщепляют плазмидную ДНК протоспейсера 2, при условии, что управляются с помощью родственного гибридного *tracrPНК:crPНК-sp2*. Следует отметить, что, как упоминалось выше, последовательность *tracrPНК N. meningitidis* является лишь предполагаемой, и она до сих пор не подтверждена с помощью секвенирования РНК. Таким образом, низкая эффективность расщепления может быть результатом либо низкой активности ортологов Cas9, либо использования неоптимальной созданной последовательности *tracrPНК*. (В) Расщепление олигонуклеотида ДНК протоспейсера. 5'-концевой меченый радиоактивной меткой олигонуклеотид комплементарной цепи (10 нМ), предварительно прогибридизованный с немеченым олигонуклеотидом некомплементарной цепи (протоспейсер 1) (10 нМ) (слева), или 5'-концевой меченый радиоактивной меткой олигонуклеотид комплементарной цепи (10 нМ), предварительно прогибридизованный с немеченым олигонуклеотидом комплементарной цепи (10 нМ) (справа) (протоспейсер 1), расщепляли с помощью различных ортологов Cas9 (500 нМ), направляемых дуплексом *tracrPНК:crPНК-sp1* из *S. pyogenes* (500 нМ, 1:1). Ортологи Cas9 из *S. thermophilus* и *L. innocua*, но не из *N. meningitidis* или *C. jejuni* могут направляться с помощью двойной-РНК *S. pyogenes*, для расщепления олигонуклеотида ДНК протоспейсера, хотя и с пониженной эффективностью. Следует отметить, что сайт расщепления на комплементарной цепи ДНК одинаковый в случае всех трех ортологов. Расщепление некомплементарной цепи происходит в различных положениях. (С) Идентичность аминокислотной последовательности ортологов Cas9. Ортологи Cas9 *S. pyogenes*, *S. thermophilus* и *L. innocua* имеют высокую процентную долю идентичных аминокислот. В отличие от этого, белки Cas9 *C. jejuni* и *N. meningitidis* отличаются и последовательностью и длиной (на ~300-400 аминокислот короче). (D) Совместные укладки созданных видоспецифических гетерологичных последовательностей *crPНК* с соответствующими ортологами *tracrPНК* из *S. pyogenes* (экспериментально подтвержденные, (4)), *L. innocua* (предполагаемая) или *N. meningitidis* (предполагаемая). *tracrPНК*; фрагменты спейсера 2 *crPНК*; и повторяющиеся фрагменты *crPНК* отслеживают и метят. Гибридные дуплексы *tracrPНК:crPНК-sp2 L. innocua* и *S. pyogenes* имеют очень схожие структурные характеристики, хоть и отличающиеся от гибрида *tracrPНК:crPНК N. meningitidis*. Вместе с данными расщепления, описанными выше в (А) и (В), спрогнозированная совместная укладка указывает на то, что видоспецифичное расщепление ДНК-мишени с помощью Cas9-*tracrPНК:crPНК* диктуется все еще неизвестной структурной особенностью дуплекса *tracrPНК:crPНК*, который распознают, в частности, родственные ортологи Cas9. Предполагается, что видоспецифическое расщепление, наблюдаемое в (А) и (В), происходит на уровне связывания Cas9 с *dual-tracrPНК:crPНК*. Двойная-РНК, направляющая расщеплением ДНК-мишени Cas9, может быть видоспецифической. В зависимости от степени разнообразия/эволюции среди белков Cas9 и дуплексов *tracrPНК:crPНК*, ортологи Cas9 и двойной-РНК являются частично взаимозаменяемыми.

Фиг. 25. Серии 8-нуклеотидных зондов ДНК комплементарны областям *crPНК*, содержащим ДНК-нацеленную область и *tracrPНК*-связывающую область, анализировали на их способность к гибридизации с *crPНК* в контексте дуплекса *tracrPНК:crPНК* и тройного комплекса Cas9-*tracrPНК:crPНК*. (А) Схематическое изображение последовательностей зондов ДНК, используемых в анализе, и их связывающих сайтов в *crPНК-sp4*. (В-С) Анализ сдвига электрофоретической подвижности зондов ДНК-мишени с *tracrPНК:crPНК-sp4* или Cas9-*tracrPНК:crPНК-sp4*. В эксперименте используют конструкцию *tracrPНК(15-89)*. Связывание дуплексов или комплексов с олигонуклеотидами-мишенями ДНК анализируют на 16% нативном полиакриламидном геле и визуализируют с помощью формирования изображения на люминесцентном фосфорном покрытии.

Мотив короткой последовательности диктует образование R-петли.

Во многих системах CRISPR/CAS, показано, что в распознавание "своего" по сравнению с "несвоим" вовлечен короткий мотив последовательности, который фиксирован в чужеродном геноме, именуемый PAM (27, 29, 32-34). Мотивы PAM имеют лишь несколько пар оснований в длину, а их точная по-

следовательность и положение изменяются в зависимости от типа системы CRISPR/CAS (32). В системе *S. pyogenes* типа II, PAM соответствует консенсусной последовательности NGG, содержащий две пары оснований G:C, которые находятся на одну пару оснований ниже от последовательности связывания crPНК, в пределах ДНК-мишени (4). Анализы трансформации показали, что мотив GG имеет важное значение для удаления плазмидной ДНК протоспейсера с помощью CRISPR/CAS в бактериальных клетках (фиг. 26А), что согласуется с предыдущими наблюдениями в *S. thermophilus* (27). Мотив также имеет важное значение для *in vitro* расщепления плазмиды протоспейсера с помощью tracrPНК:crPНК-направляемого Cas9 (фиг. 26В). Для определения роли PAM в расщеплении ДНК-мишени с помощью комплекса Cas9-tracrPНК:crPНК, исследуют ряд дуплексов дцНК, содержащих мутации в последовательности PAM в комплементарных или некомплементарных нитях, или обоих (фиг. 13А). Анализы расщепления, использующие эти субстраты, показали, что Cas9-катализируемое расщепление ДНК является особенно чувствительным к мутациям в последовательности PAM в некомплементарной цепи ДНК, в отличие от систем CRISPR/CAS типа I где распознается PAM комплементарной цепи (18, 34). Расщепление одноцепочечных ДНК-мишеней не зависит от мутаций мотива PAM. Это наблюдение показывает, что мотив PAM требуется только в контексте дцДНК-мишени и, таким образом, возможно необходимо для того, чтобы разрешить раскручивание дуплекса, инвазию цепи и формирование структуры R-петли. Когда использовали различные пары crPНК-ДНК-мишень (crPНК-sp4 и ДНК протоспейсера 4), выбранные в связи с присутствием канонической PAM, не присутствующей в ДНК-мишени протоспейсера 2, обнаружили, что оба нуклеотиды G PAM являются необходимыми для эффективного Cas9-катализируемого расщепления ДНК (фиг. 13В и фиг. 26С). Чтобы определить, играет ли PAM непосредственную роль в рекрутинге комплекса Cas9-tracrPНК:crPНК к соответствующему сайту ДНК-мишени, анализируют аффинности связывания комплекса с последовательностями ДНК-мишени, с помощью анализа по сдвигу подвижности на нативном геле (фиг. 13С). Мутация любого G в последовательности PAM существенно снижает сродство Cas9-tracrPНК:crPНК к ДНК-мишени. Это указывает на роль последовательности PAM в связывании ДНК-мишени с помощью Cas9.

Фиг. 13. (А) Cas9, запрограммированный двойной РНК, исследуют на активность таким образом, как и на фиг. 10В. ДТ и мутантные последовательности PAM в ДНК-мишенях указывают линиями. (В) Дуплексы ДНК-мишени протоспейсера 4 (меченные на обоих 5'-концах), содержащие ДТ и мутантные мотивы PAM, инкубируют с Cas9, запрограммированным tracrPНК:crPНК-sp4 (нуклеотиды от 23 до 89). В указанные моменты времени (в минутах), аликваты реакции расщепления отбирают и анализируют таким образом, как на фиг. 10В. (С) Анализы сдвига электрофоретической подвижности осуществляют с использованием РНК-запрограммированного Cas9 (D10A/H840A) и дуплексов ДНК-мишени протоспейсера 4 [такого же как и в (В)], содержащих ДТ, и мутировавших мотивов PAM. Комплекс Cas9 (D10A/H840A)-РНК титруют от 100 мкМ до 1 мМ.

Фиг. 26. (А) Мутации последовательности PAM в плазмидной ДНК протоспейсера 2 отменяют интерференцию поддержания плазмиды с помощью системы CRISPR/Cas типа II в бактериальных клетках. Плазмиды протоспейсера 2 дикого типа с функциональной или мутантной PAM трансформируют в мутант дикого типа (штамм SF370, также под названием EC904) и rge-crPНК-дефицитных мутант (EC1479) *S. pyogenes*, как и на фиг. 12D. Мутации PAM являются неприемлемыми для системы CRISPR/Cas типа II *in vivo*. Приведены значения и стандартные отклонения трех биологических повторов. (В) Мутации последовательности PAM в плазмидной ДНК протоспейсера отменяют расщепление с помощью Cas9-tracrPНК:crPНК. Плаزمида протоспейсера 2 дикого типа с функциональной или мутантной PAM подвергается расщеплению Cas9, как и на фиг. 10А. Мутантные плазмиды PAM не расщепляются с помощью комплекса Cas9-tracrPНК:crPНК. (С) Мутации канонической последовательности PAM отменяют интерференцию поддержания плазмиды системы CRISPR/Cas типа II в бактериальных клетках. Плазмиды протоспейсера 2 дикого типа с функциональной или мутантной PAM расщепляются Cas9, запрограммированным tracrPНК и crPНК-sp2. Реакции расщепления осуществляют в присутствии рестрикционной эндонуклеазы XmnI для визуализации продуктов расщепления Cas9 в виде двух фрагментов (~1880 и ~800 п.о.). Размеры фрагментов указаны в парах оснований. Cas9 может быть запрограммирован одиночной гибридной РНК. Исследование вероятной вторичной структуры дуплекса tracrPНК:crPНК (фиг. 10Е и 12С) привело к предположению, что функции, необходимые для сайт-специфического расщепления катализируемой Cas9 ДНК, могут быть собраны в одной гибридной РНК. Хотя механизм выбора цели tracrPНК:crPНК в природе и работает эффективно, возможность управления Cas9, одной РНК, является достаточно привлекательной в связи с его потенциальной полезностью для запрограммированного расщепления ДНК и редактирования генома (фиг. 1А-В). Авторы изобретения сконструировали два варианта гибридных РИСК, содержащих последовательность распознавания мишени на 5'-конце с последующей структурой шпильки, сохранившей то взаимодействие спаренных оснований, которое происходит между tracrPНК и crPНК (фиг. 14А). Этот одноцепочечный транскрипт эффективно сливает 3'-конец crPНК с 5'-концом tracrPНК, тем самым имитируя структуру двойной-РНК, необходимую для управления сайт-специфическим расщеплением ДНК с помощью Cas9. В анализах расщепления с использованием плазмидной ДНК наблюдают, что более длинные гибридные РНК в состоянии направлять расщепление ДНК, катализируемое Cas9, аналогично тому, что наблюдают в случае процессированного дуплекса

tracrRNA:crRNA (фиг. 14А и 27, А и С). Более короткая гибридная РНК не работает эффективно в этом анализе, подтверждая, что нуклеотиды, которые находятся в пределах 5 до 12 положений от области спаривания оснований tracrRNA:crRNA, являются важными для эффективного связывания Cas9 и/или распознавания цели. Аналогичные результаты получены в анализах расщепления с помощью короткой дцДНК в качестве субстрата, что дополнительно указывает на то, что положение сайта расщепления в ДНК-мишени является идентичным тому, которое наблюдают с использованием двойной tracrRNA: crRNA в качестве направляющей (фиг. 14В и 27, В и С). И, наконец, чтобы установить, является ли конструкция гибридной РНК универсальной, создали пять различных гибридных направляющих РНК для нацеливания к части гена, кодирующего зеленый флуоресцентный белок (GFP) (фиг. 28, А-С), и исследовали их эффективность против плазмиды, несущей GFP, кодирующей последовательность *in vitro*. Во всех пяти случаях, Cas9, запрограммированный этими гибридными РНК, эффективно расщепляет плазмиду с соответствующим сайтом-мишенью (фиг. 14С и 28D), указывая, что рациональная конструкция РНК является функциональной и может, в принципе, обеспечить нацеливание на любую представляющую интерес последовательность ДНК с несколькими ограничениями помимо присутствия динуклеотида GG, прилегающего к последовательности-мишени.

Фиг. 1. РНК, нацеленная на ДНК, содержит одноцепочечный "ДНК-нацеленный сегмент" и "белок-связывающий сегмент", которые содержат участок двухцепочечной РНК. (А) РНК, нацеленная на ДНК, может содержать две отдельные молекулы РНК (которую обозначают "двойной молекуле" или РНК "в виде двух молекул", нацеленная на ДНК). РНК, нацеленная на ДНК, в виде двойной молекулы содержит "РНК-нацеливающую" и "РНК-активатор". (В) РНК, нацеленная на ДНК, может быть в виде одной молекулы РНК (которую обозначают РНК, нацеленная на ДНК, "в виде одиночной молекулы"). РНК, нацеленная на ДНК, в виде одиночной молекулы содержит "нуклеотиды линкера".

Фиг. 14. (А) Плазмиду, несущую последовательность-мишень протоспейсера 4 и РАМ ДТ, расщепляют с помощью Cas9, запрограммированного дуплексом tracrRNA (4-89):crRNA-sp4 или *in vitro*-транскрибируемой гибридной РНК, построенных с помощью присоединения 3'-конца crRNA к 5'-концу tracrRNA с тетрапетлей GAAA. Реакции расщепления анализируют с помощью рестрикционного картирования с XmnI. Представлены последовательности гибридных РНК А и В, где ДНК-направленные последовательности (подчеркнуты), производные повторов crRNA (надчеркнуты), и последовательности, производные tracrRNA (пунктирно надчеркнуты). (В) Реакции расщепления дуплекса ДНК протоспейсера 4 осуществляют таким образом, как на фиг. 10В. (С) Пять гибридных РНК, созданных для нацеливания на ген GFP, используют для программирования Cas9 с целью расщепления плазмиды, содержащей ген GFP. Реакции расщепления плазмиды осуществляют таким образом, как на фиг. 12Е, за исключением того, что плазмидную ДНК картируют рестрикцией с AvrII после расщепления Cas9.

Фиг. 27. (А) Одиночная гибридная РНК направляет Cas9-катализируемое расщепление соответствующего протоспейсера (протоспейсер 1 и протоспейсер 2) плазмидной ДНК. Реакции расщепления осуществляют в присутствии рестрикционной эндонуклеазы XmnI для визуализации продуктов расщепления Cas9 в виде двух фрагментов (~1880 и ~800 п.о.). Размеры фрагментов указаны в парах оснований. (В) Одиночная гибридная РНК направляет Cas9-катализируемое расщепление соответствующего протоспейсера (протоспейсер 1 и протоспейсер 2) в олигонуклеотиде ДНК. Размеры фрагментов указаны в нуклеотидах. (С) Схематическое изображение гибридных РНК, используемых в эксперименте. Последовательности химерных РНК А и В представлены с 5' последовательностью нацеленной на ДНК протоспейсера crRNA (подчеркнута), tracrRNA-связывающей последовательностью crRNA (надчеркнута) и производной tracrRNA (пунктирно надчеркнута) последовательностью.

Фиг. 28. (А) Схематическое изображение плазмиды pCFJ127 экспрессирующей GFP. Участок-мишень открытой рамки считывания GFP указывают черной стрелкой. (В) Крупный план последовательности области-мишени. Последовательности, на которые нацелены гибридные РЖ, показаны серыми полосами. Динуклеотиды РАМ выделены рамкой. Уникальный сайт рестрикции SalI расположен на 60 п.о. выше локуса-мишени. (С) Слева: Последовательности ДНК-мишени показаны вместе с прилегающими к ним мотивами РАМ. Справа: Последовательности гибридных "направляющих" РНК. (D) pCFJ127 расщепляют с помощью Cas9, запрограммированного гибридными GFP1-5 РНК, как указано. Плазмиду дополнительно расщепляют с помощью SalI, а реакции анализируют с помощью электрофореза на 3% агарозном геле и визуализируют с помощью окрашивания красителем SYBR Safe.

#### Выводы.

Был определен механизм интерференции ДНК с участием структуры двойной-РНК, которая направляет эндонуклеазу Cas9 для того, чтобы ввести сайт-специфические двухцепочечные разрывы в ДНК-мишень. Белок Cas9, управляемый tracrRNA:crRNA, использует различные эндонуклеазные домены (HNH и RuvC-подобные домены) для того, чтобы расщеплять две цепи ДНК-мишени. Распознавание мишени с помощью Cas9 требует как затравочную последовательность в crRNA, так и содержащую динуклеотид GG последовательность РАМ, смежную с crRNA-связывающей областью в ДНК-мишени.

Дополнительно авторы изобретения показали, что эндонуклеаза Cas9 может быть запрограммирована на мишень-направляющей РНК, созданной как единый транскрипт, и расщепляют любую представляющую интерес последовательность дцДНК. Система является эффективной, универсальной и про-

граммируемой путем изменения последовательности, связывающей с ДНК-мишенью, в направляющей гибридной РНК. Нуклеазы с цинковыми пальцами и эффекторные нуклеазы, подобные активатору транскрипции, привлекают значительный интерес в качестве искусственных ферментов, созданных для манипулирования геномами (35-38). Это представляет альтернативную методологию, основанную на Cas9, запрограммированной РНК, которая облегчает нацеливание на ген и редактирование генома.

Цитируемые ссылки.

1. B. Wiedenheft, S. H. Sternberg, J. A. Doudna, *Nature* 482, 331 (2012).
2. D. Bhaya, M. Davison, R. Barrangou, *Annu. Rev. Genet.* 45, 273 (2011).
3. M. P. Terns, R. M. Terns, *Curr. Opin. Microbiol.* 14, 321 (2011).
4. E. Deltcheva и соавт., *Nature* 471, 602 (2011).
5. J. Carte, R. Wang, H. Li, R. M. Terns, M. P. Terns, *Genes Dev.* 22, 3489 (2008).
6. R. E. Haurwitz, M. Jinek, B. Wiedenheft, K. Zhou, J. A. Doudna, *Science* 329, 1355 (2010).
7. R. Wang, G. Preamplume, M. P. Terns, R. M. Terns, H. Li, *Structure* 19, 257 (2011).
8. E. M. Gesner, M. J. Schellenberg, E. L. Garside, M. M. George, A. M. Macmillan, *Nat. Struct. Mol. Biol.* 18, 688 (2011).
9. A. Hatoum-Aslan, I. Maniv, L. A. Marraffini, *Proc. Natl. Acad. Sci. U.S.A.* 108, 21218 (2011).
10. S. J. J. Brouns и соавт., *Science* 321, 960 (2008).
11. D. G. Sashital, M. Jinek, J. A. Doudna, *Nat. Struct. Mol. Biol.* 18, 680 (2011).
12. N. G. Lintner и соавт., *J. Biol. Chem.* 286, 21643 (2011).
13. E. Semenova и соавт., *Proc. Natl. Acad. Sci. U.S.A.* 108, 10098 (2011).
14. B. Wiedenheft и соавт., *Proc. Natl. Acad. Sci. U.S.A.* 108, 10092 (2011).
15. B. Wiedenheft и соавт., *Nature* 477, 486 (2011).
16. C. R. Hale и соавт., *Cell* 139, 945 (2009).
17. J. A. L. Howard, S. Delmas, I. Ivančić-Baće, E. L. Bolt, *Biochem. J.* 439, 85 (2011).
18. E. R. Westra и соавт., *Mol. Cell* 46, 595 (2012).
19. C. R. Hale и соавт., *Mol. Cell* 45, 292 (2012).
20. J. Zhang и соавт., *Mol. Cell* 45, 303 (2012).

21. K. S. Makarova и соавт., *Nat. Rev. Microbiol.* 9, 467 (2011).
22. K. S. Makarova, N. V. Grishin, S. A. Shabalina, Y. I. Wolf, E. V. Koonin, *Biol. Direct* 1, 7 (2006).
23. K. S. Makarova, L. Aravind, Y. I. Wolf, E. V. Koonin, *Biol. Direct* 6, 38 (2011).
24. S. Gottesman, *Nature* 471, 588 (2011).
25. R. Barrangou и соавт., *Science* 315, 1709 (2007).
26. J. E. Garneau и соавт., *Nature* 468, 67 (2010).
27. R. Sapranaukas и соавт., *Nucleic Acids Res.* 39, 9275 (2011).
28. G. K. Taylor, D. F. Heiter, S. Pietrovski, B. L. Stoddard, *Nucleic Acids Res.* 39, 9705 (2011).
29. H. Deveau и соавт., *J. Bacteriol.* 190, 1390 (2008).
30. B. P. Lewis, C. B. Burge, D. P. Bartel, *Cell* 120, 15 (2005).
31. G. Hutvagner, M. J. Simard, *Nat. Rev. Mol. Cell Biol.* 9, 22 (2008).
32. F. J. M. Mojica, C. Díez-Villaseñor, J. García-Martínez, C. Almendros, *Microbiology* 155, 733 (2009).
33. L. A. Marraffini, E. J. Sontheimer, *Nature* 463, 568 (2010).
34. D. G. Sashital, B. Wiedenheft, J. A. Doudna, *Mol. Cell* 46, 606 (2012).
35. M. Christian и соавт., *Genetics* 186, 757 (2010).
36. J. C. Miller и соавт., *Nat. Biotechnol.* 29, 143 (2011).
37. F. D. Urnov, E. J. Rebar, M. C. Holmes, H. S. Zhang, P. D. Gregory, *Nat. Rev. Genet.* 11, 636 (2010).
38. D. Carroll, *Gene Ther.* 15, 1463 (2008).
39. J. Sambrook, E. F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, ed. 2, 1989).
40. M. G. Caparon, J. R. Scott, Genetic manipulation of pathogenic streptococci. *Methods Enzymol.* 204, 556 (1991). doi:10.1016/0076-6879(91)04028-M Medline
41. C. Frøkjær-Jensen и соавт., Single-copy insertion of transgenes in *Caenorhabditis elegans*. *Nat. Genet.* 40, 1375 (2008). doi:10.1038/ng.248 Medline
42. R. B. Denman, Using RNAFOLD to predict the activity of small catalytic RNA. *Biotechniques* 15, 1090 (1993). Medline
43. I. L. Hofacker, P. F. Stadler, Memory efficient folding algorithms for circular RNA secondary structures. *Bioinformatics* 22, 1172 (2006). doi:10.1093/bioinformatics/btl023 Medline
44. K. Darty, A. Denise, Y. Ponty, VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* 25, 1974 (2009). doi:10.1093/bioinformatics/btp250 Medline

Пример 2. РНК-программированное редактирование генома в клетках человека.

Данные, представленные ниже, иллюстрируют то, что Cas9 может быть экспрессирован и локализован в ядре клеток человека, и что он собирается с направляющей РНК в виде одиночной молекулы ("sgРНК"; сосредотачивает в себе функции, необходимые для связывания как Cas9, и распознавания сайта ДНК-мишени) в клетке человека. Эти комплексы могут создавать двухцепочечные разрывы и стимулировать репарирование путем соединения негомологичных концов (NHEJ) в геномной ДНК в сайте, комплементарном последовательности sgРНК, активность для которой требуется как Cas9, так и sgРНК. Удлинение последовательности РНК на ее 3'-конце повышает активность нацеливания на ДНК в живых клетках. Кроме того, эксперименты с использованием экстрактов из трансфицированных клеток показали, что сборка sgРНК с Cas9 представляет собой ограничивающий фактор в случае Cas9-опосредованного расщепления ДНК. Эти результаты показывают, что РНК-запрограммированное редактирование генома работает в живых клетках и *in vivo*.

Материалы и методы.

Дизайн и конструкция плазмиды.

Последовательность, кодирующая Cas9 *Streptococcus pyogenes* (остатки 1-1368), слитая с эпитопом HA (аминокислотная последовательность DAYPYDVPDYASL (SEQ ID NO:274)), клеточным сигналом

внутриядерной локализации (аминокислотная последовательность PKKKRKVEDPKKKRKVD (SEQ ID NO:275)) кодон-оптимизирована для экспрессии в человеке и синтезирована с помощью GeneArt. Последовательность ДНК представляет собой SEQ ID NO:276, а последовательность белка представляет собой SEQ ID NO:277. Безлигазное клонирование (LIC) используют для вставки этой последовательности в векторы pcDNA3.1-производный GFP и LIC mCherry (векторы 6D и 6B соответственно получают от UC Berkeley MacroLab), что приводит к слиянию Cas9-HA-NLS-GFP и Cas9-HA-NLS-mCherry, экспрессированным под контролем промотора CMV. "Направляющие" sgРНК экспрессируют с использованием вектора экспрессии pSilencer 2.1-U6 puro (life Technologies) и pSuper (Oligoengine). Конструкции экспрессии РНК получают с помощью гибридизации комплементарных олигонуклеотидов с образованием РНК-кодирующей последовательности ДНК и лигирования прогибридизованного фрагмента ДНК между сайтами BamHI и HindIII в pSilencer 2.1-U6 puro и сайтами BglII и HindIII в pSuper.

Условия культивирования клеток и трансфекции ДНК.

Клетки HEK293T поддерживают в минимальной эссенциальной среде Игла, модифицированной по способу Дульбекко (DMEM), с добавлением 10% фетальной бычьей сыворотки (FBS) в 37°C увлажненном инкубаторе с 5% CO<sub>2</sub>. Клетки временно трансфицируют плазмидной ДНК с использованием либо XtremeGENE DNA Transfection Reagent (Roche), либо Turbofect Transfection Reagent (Thermo Scientific) согласно рекомендуемым протоколам. Вкратце, клетки HEK293T трансфицируют при 60-80% конфлюентности в 6-луночных планшетах с использованием 0,5 мкг плазмиды экспрессии Cas9 и 2,0 мкг плазмиды экспрессии РНК. Эффективности трансфекции оценивают на уровне 30-50% в случае Turbofect (фиг. 29E и 37A-B) и 80-90% в случае Xtremegene (фиг. 31B), на основе фракции GFP-положительных клеток, наблюдаемых с помощью флуоресцентной микроскопии. Через 48 ч после трансфекции клетки промывают фосфатно-солевым буфером (PBS) и лизируют с помощью применения 250 мкл буфера для лизиса (20 mM Hepes pH 7,5, 100 mM хлорида калия (KCl), 5 mM хлорида магния (MgCl<sub>2</sub>), 1 mM дитиотреитола (DTT), 5% глицерина, 0,1% тритона X-100, с добавлением смеси ингибиторов протеаз Роща), а затем встряхивают в течение 10 мин при 4°C. Полученный клеточный лизат разделяют на аликвоты для дополнительного анализа. Геномную ДНК выделяют из 200 мкл клеточного лизата с помощью DNeasy Blood и Tissue Kit (Qiagen) в соответствии с протоколом производителя.

Анализ экспрессии Cas9 путем вестерн-блоттинга.

HEK293T, трансфицированный плазмидой экспрессии Cas9-HA-NLS-GFP, собирают и лизируют через 48 ч после трансфекции, как описано выше. 5 мкл лизата разгоняют электрофорезом на 10% полиакриламидном геле SDS, переносят на мембрану PVDF и изучают с помощью HRP-конъюгированным анти-HA антитела в качестве зонда (Sigma, разведение 1:1000 в 1× PBS).

Анализ Surveyor.

Контрольный анализ осуществляют таким образом, как описано ранее [10,11,12]. Вкратце, locus клатрина легкой цепи (CLTA) человека амплифицируют путем ПЦР из 200 нг геномной ДНК с использованием высокоточной полимеразы, Herculase II Fusion ДНК Polymerase (Agilent Technologies) и прямого праймера 5'-GCAGCAGAAGAAGCCTTTGT-3' (SEQ ID NO://) и обратного праймера 5'-TTCCCTCTCTCCCTCCTCTC-3' (SEQ ID NO://). 300 нг 360 п.о. ампликона затем денатурируют с помощью нагревания до 95°C и медленно заново регидризируют с использованием термоблока для того, чтобы случайным образом прогибридизировать цепи дикого типа и мутантных ДНК. Затем образцы инкубируют с нуклеазой Cel-1 (Surveyor Kit, Transgenomic) в течение 1 ч при 42°C. Cel-1 распознает и расщепляет спирали ДНК, содержащие ошибочные спаривания (гибридизация дикий тип:мутант). Продукты расщепления нуклеазой Cel-1 разделяют на 10% акриламидном геле; и визуализируют с помощью окрашивания красителем SYBR Safe (Life Technologies). Количественное определение полос расщепления осуществляют с использованием программного обеспечения ImageLab (Bio-Rad). Процент расщепления определяют с помощью деления средней интенсивности продуктов расщепления (160-200 п.н.) на сумму интенсивностей нерасщепленного продукта ПЦР (360 п.н.) и продукта расщепления.

In vitro транскрипция.

Направляющую РНК транскрибируют in vitro с использованием рекомбинантной РНК-полимеразы T7 и матрицы ДНК, полученной с помощью регидризации комплементарных синтетических олигонуклеотидов, как описано выше. РНК очищают с помощью электрофореза в денатурирующем акриламидном геле с 7M мочевиной, осаждают этанолом и растворяют в обработанной DEPC воде.

Анализ нозерн-блоттинг.

РНК очищают от клеток HEK293T с использованием набора для изолирования small-RNA miRvana (Ambion). Для каждого образца, 800 нг РНК разделяют на 10% мочевино-PAGE геле после денатурации в течение 10 мин при 70°C в буфере загрузки РНК (0,5X TBE (pH 7,5), 0,5 мг/мл бромфенолового синего, 0,5 мг ксилонцианола и 47% формамида). После электрофореза при 10 Вт в 0,5X буфере TBE, пока бромфеноловый синий краситель не достигнет дна геля, образцы переносят электроблоттингом на мембрану Nytran при 20 Вольт в течение 1,5 ч в 0,5X TBE. Перенесенные РНК шивают на мембране Nytran в UV-Crosslinker (Stratagene) и предварительно гибридизуют при 45°C в течение 3 ч в буфере, содержащем 40% формамида, 5x SSC, 3x Derrhardt (0,1% каждого из фикола, поливинилпирролидона и BSA) и

200 мкг/мл ДНК спермы лосося. Предварительно гибридизованные мембраны инкубируют в течение ночи в предгибризационном буфере с добавлением олигонуклеотида, меченого  $^{32}\text{P}$  на 5'-конце антисмысловой ДНК, при 1 миллионе имп. в минуту/мл. После нескольких промываний в буфере SSC (окончательное промывание в 0.2X SSC), мембраны анализируют путем формирования изображения на люминесцентном фосфорном покрытии.

Анализ расщепления *in vitro*.

Клеточные лизаты получают, как описано выше, и инкубируют с донорной плазмидой CLTA-RFP [10]. Реакции расщепления осуществляют в общем объеме 20 мкл, содержащем 10 мкл лизата, 2 мкл буфера расщепления 5x (100 мМ HEPES pH 7,5, 500 мМ KCl, 25 мМ MgCl<sub>2</sub>, 5 мМ DTT, 25% глицерина) и 300 нг плазмиды. Там, где это указано, в реакции добавляют 10 пмоль *in vitro* транскрибируемой sgРНК CLTA1. Реакции расщепления осуществляют в общем объеме 20 мкл, содержащем 10 мкл лизата, 2 мкл буфера расщепления 5x (100 мМ HEPES pH 7,5, 500 мМ KCl, 25 мМ MgCl<sub>2</sub>, 5 мМ DTT, 25% глицерина) и 300 нг плазмиды. Там, где это указано, в реакции добавляют 10 пмоль *in vitro* транскрибируемой: sgРНК CLTA1. Реакции инкубируют при 37°C в течение 1 ч, а затем проводят рестрикцию 10 ед. XhoI (NEB) в течение дополнительных 30 мин при 37°C. Реакции останавливают добавлением Proteinase K (Thermo Scientific) и инкубируют при 37°C в течение 15 мин. Продукты расщепления анализируют с помощью электрофореза на 1% агарозном геле и окрашивают красителем SYBR Safe. Наличие фрагментов из ~2230 и ~3100 п.о. свидетельствует о Cas9-опосредованном расщеплении.

Результаты.

Для того чтобы проверить, может ли Cas9 быть запрограммирован для расщепления геномной ДНК в живых клетках, Cas9 совместно экспрессируют вместе с sgРНК, сконструированным для нацеливания на ген легкой цепи клатрина (CLTA) человека. Ранее на геномный локус CLTA направленно воздействовали и отредактировали с использованием ZFN [10]. Сначала исследовали экспрессию кодон-оптимизированной в отношении кодонов человека версии белка Cas9 *Streptococcus pyogenes* и sgРНК в клетках HEK293T человека. Белок Cas9 в 160 кДа экспрессируют в виде слитого белка, содержащего эпитоп HA, клеточный сигнал внутриядерной локализации (NLS) и зеленый флуоресцентный белок (GFP), прикрепленный к С-концу Cas9 (фиг. 29А). Анализ клеток, трансфицированных вектором, кодирующим слитый с GFP Cas9, показал значительную экспрессию Cas9 и локализацию в ядре (фиг. 29В). Вестерн-блоттинг подтвердил, что белок Cas9 экспрессируют в основном без изменений в экстрактах из этих клеток (фиг. 29А). Чтобы запрограммировать Cas9, экспрессируют sgРНК, содержащую 5'-концевую 20-нуклеотидную последовательность, комплементарную последовательности ДНК-мишени, а также 42-нуклеотидной 3'-концевую структуру типа "стебель - петля", необходимые для связывания Cas9 (фиг. 29С). Эта 3'-концевая последовательность соответствует минимальной структуре типа "стебель - петля", которую ранее используют для программирования Cas9 *in vitro* [9]. Экспрессией этой sgРНК управляют с помощью промотора U6 (РНК-полимеразы III), человека [11]. Анализ нозерн-блоттинг РНК, выделенной из клеток, трансфицированных вектором экспрессии, содержащим sgРНК, под контролем промотора U6, продемонстрировал, что sgРНК является действительно экспрессируемой, а ее стабильность повышают за счет присутствия Cas9 (фиг. 29D).

Фиг. 29 показывает, что совместная экспрессия Cas9 и направляющей РНК в клетках человека создает двуниевые разрывы ДНК в локусе-мишени. (А) Верх: схематическое изображение экспрессирующей конструкции Cas9-HA-NLS-GFP. Низ: лизат из клеток HEK293T, трансфицированных плазмидой, экспрессирующей Cas9, анализируют с помощью Вестерн-блоттинга с использованием анти-HA антитела. (В) Флуоресцентная микроскопия клеток HEK293T, экспрессирующих Cas9-HA-NLS-GFP. (С) Конструкция направляющей РНК в виде одиночной молекулы (sgРНК, т.е. РНК, нацеленной на ДНК, в виде одиночной молекулы), нацеленной на локус CLTA человека. Верх; схематическое изображение сайта-мишени sgРНК в экзоне 7 гена CLTA человека. Последовательность-мишень, которая гибридизуется с направляющим сегментом sgРНК CLTA1, обозначена как "sgРНК CLTA1". Динуклеотид GG протоспейсер, соседствующий с мотивом (PAM), отмечают стрелкой. Черные линии обозначают ДНК-связывающие области белка управления ZFN. Стоп-кодон трансляции открытой рамки считывания CLTA отмечают пунктирной линией для сравнения. Середина; схематическое изображение конструкции экспрессирующей sgРНК. РНК экспрессируют под контролем промотора Pol III U6 и поли(Т) тракта, который служит в качестве сигнала терминатора транскрипции Pol III. Низ; расщепление ДНК-мишени, управляемое sgРНК, с помощью Cas9. SgРНК состоит из 20-нуклеотидного 5'-концевого направляющего сегмента с последующим 42-нуклеотидной структурой типа "стебель - петля", необходимой для связывания Cas9. Cas9-опосредованное расщепление двух цепей ДНК-мишени происходит при раскручивании ДНК-мишени и формировании дуплекса между управляющим сегментом sgРНК и ДНК-мишенью. Это зависит от наличия мотива PAM (подходящего для того, чтобы использовать Cas9, например, динуклеотид GG, см. пример 1 выше) ниже последовательности-мишени в ДНК-мишени. Следует отметить, что последовательность-мишень инвертируют по отношению к верхней диаграмме. (D) Анализ экспрессии sgРНК в клетках HEK293T путем нозерн-блоттинга. (E) Нуклеазный анализ Surveyor геномной ДНК, выделенной из клеток HEK293T, экспрессирующих Cas9 и/или sgРНК CLTA. Конструкцию ZFN, ранее ис-

пользуемую для нацеливания на локуса CLTA [10], используют в качестве положительного контроля для обнаружения DSB-индуцированной репарации ДНК с помощью негомологичного соединения концов.

Далее был исследован один из сайт-специфических DSB, которые получают в клетках HEK293T, трансфицированных Cas9-NA-NLS-mCherry и sgРНК CLTA1. Чтобы сделать это, искали небольшие вставки и делеции в локусе, получаемые в результате несовершенной репарации путем DSB-индуцированного NHEJ с использованием анализа нуклеазы Surveyor [12]. Область геномной ДНК, на которую нацелена Cas9:sgРНК, амплифицируют с помощью PCR, а полученные продукты денатурируют и заново гибридизуют. Заново гибридизованные продукты PCR инкубируют с эндонуклеазой Cel-1, распознающей ошибочные спаривания и разделяют на акриламидном геле для идентификации полос расщепления Cel-1. Поскольку репарация ДНК с помощью NHEJ, как правило индуцируется DSB (двунитиевого разрыва), позитивный сигнал в тесте Surveyor указывает на то, что расщепление геномной ДНК имело место. С помощью этого анализа было обнаружено расщепление локуса CLTA в положении, на которое была нацелена sgРНК к CLTA1 (фиг. 29E). Пара ZFN, которые нацелены на соседний сайт в локусе CLTA, в этих экспериментах обеспечивает положительный контроль [10].

Для того чтобы определить, экспрессия либо Cas9, либо sgРНК представляет собой ограничивающий фактор в наблюдаемых реакциях редактирования генома, лизаты, полученные из трансфицированных клеток, инкубируют с плазмидной ДНК, несущей фрагмент гена CLTA, на который нацелена sgРНК CLTA1. Расщепление плазмидной ДНК не наблюдалось при инкубировании с лизатом, приготовленным из клеток, трансфицированных только вектором экспрессии Cas9-NA-NLS-GFP, в соответствии с результатами анализа Surveyor. Тем не менее, активное расщепление плазмиды обнаруживают при условии, что в лизат добавляют *in vitro* транскрибированные sgРНК CLTA1 (фиг. 30A). Эти результаты показывают, что ограничивающим фактором для функционирования Cas9 в клетках человека может быть сборка с sgРНК. Эту возможность проверяют непосредственно с помощью анализа расщепления плазмиды в лизатах из клеток, трансфицированных таким же образом, как и ранее, в присутствии и в отсутствие добавленной экзогенной sgРНК. Следует отметить, что при условии, что экзогенную sgРНК добавляют в лизат из клеток, трансфицированных обоими векторами экспрессии Cas9 и sgРНК, наблюдают существенное увеличение активности расщепления ДНК (фиг. 30B). Этот результат показывает, что ограничивающим фактором для функционирования Cas9 в клетках HEK293T является экспрессия sgРНК или ее добавление к Cas9.

Фиг. 30 иллюстрирует то, что клеточные лизаты, содержащие активную Cas9: sgРНК, поддерживают сайт-специфическое расщепление ДНК. (А) Лизаты из клеток, трансфицированных плазмидой(ами), указанными слева, инкубируют с плазмидной ДНК, содержащей РАМ и последовательность-мишень, комплементарную sgРНК CLTA1; где указано, что в реакцию добавляют 10 пмоль *in vitro* транскрибированной sgРНК CLTA1; вторичное расщепление с XhoI дает фрагменты в ~2230 и ~3100 п.о., свидетельствующих о прошедшем Cas9-опосредованном расщеплении. Контрольная реакция с использованием лизата из клеток, трансфицированных конструкции экспрессирующей ZFN, демонстрирует фрагменты несколько другого размера, отражающие смещение сайта-мишени ZFN по отношению к сайту-мишени CLTA1. (В) Лизаты клеток, трансфицированных с плазмидой экспрессии Cas9-GFP, и, если указано, плазмидой экспрессии sgРНК CLTA1, инкубируют с плазмидной ДНК-мишенью, также как (А) в отсутствие или в присутствии *in vitro*-транскрибируемой sgРНК CLTA1.

В качестве средства повышения сборки Cas9:sgРНК в живых клетках, авторы затем исследовали эффект удлинения предполагаемой Cas9-связывающей области направляющей РНК. Две новые версии sgРНК CLTA1 создают так, чтобы они содержали дополнительные шесть или двенадцать пар оснований в спирали, что имитирует взаимодействия спаренных оснований между crРНК и tracrРНК (фиг. 31A). Кроме того, 3'-конец направляющей РНК удлиняют на пять нуклеотидов, на основе нативной последовательности из tracrРНК *s. pyogenes* [9]. Векторы, кодирующие эти удлиненные на 3'-sgРНК, под контролем промоторов либо Pol III U6, либо H1, трансфицируют в клетки вместе с вектором экспрессии Cas9-NA-NLS-GFP, а сайт-специфическое расщепление генома исследуют с использованием анализа Surveyor (фиг. 31B). Результаты подтвердили, что расщепление требует как Cas9, так и sgРНК CLTA1, но не происходит при условии, что отдельно экспрессируются либо Cas9, либо sgРНК. Кроме того, наблюдают значительно увеличившуюся частоту NHEJ, что обнаруживается с помощью расщепления нуклеазой Cel-1, в то время как частота мутагенеза NHEJ, получаемая с контрольной парой ZFN, практически не изменяется.

Фиг. 31 иллюстрирует то, что 3' удлинение конструкций sgРНК повышает сайт-специфический NHEJ-опосредованный мутагенез. (А) Конструкцию для экспрессии sgРНК CLTA1 (вверху) создают так, чтобы генерировать транскрипты, содержащие оригинальную Cas9-связывающую последовательность (v1.0) или дуплексы dcРНК, удлиненные на 4 пары оснований (v2.1) или 10 пар оснований (v2.2). (В) Нуклеазный анализ Surveyor геномной ДНК, выделенной из клеток HEK293T, экспрессирующих Cas9 и/или sgРНК CLTA v1.0, v2.1 или v2.2. Конструкцию ZFN, ранее используемую для нацеливания на локус CLTA [10], используют в качестве положительного контроля для обнаружения DSB-индуцированной репарации ДНК с помощью негомологичного соединения концов.

Эти результаты обеспечивают основу для применения Cas9 в качестве простого молекулярного ин-

струмента для редактирования генома. Важной функцией этой системы является возможность программировать Cas9 несколькими sgРНК в одной и той же клетке, либо для повышения эффективности нацеливания на один локус или для нацеливания на нескольких локусов одновременно. Подобные стратегии найдут широкое применение в полногеномных экспериментах и масштабных научно-исследовательских работах, таких как разработка моделей мультигенных заболеваний.

Пример 3. Семейства tracrРНК и Cas9 системы иммунитета CRISPR-Cas типа II.

Авторы изобретения проводили поиск всех существующих в настоящее время предполагаемых локусов CRISPR-Cas типа II, в общедоступных бактериальных геномах, путем скрининга для обнаружения последовательностей, гомологичных Cas9, характерного белка системы типа II. Авторы изобретения построили (филогенетическое дерево на основе множественного выравнивания последовательностей выявленных ортологов Cas9. Длину повтора CRISPR и организация гена оперонов cas, ассоциированных систем типа II, проанализировали в различных подгруппах Cas9. Предложена подклассификация локусов типа II, которые дополнительно разделены на подгруппы с учетом выбранных 75 репрезентативных ортологов Cas9. Затем прогнозируют последовательности tracrРНК, в основном, с помощью извлечения последовательностей повторов CRISPR и скрининга антиповторов, в пределах или в непосредственной близости от генов cas и повторов CRISPR отдельных локусов типа II. Осуществляют сравнительный анализ последовательностей и прогнозируемых структур выбранных ортологов tracrРНК. Наконец, определяют экспрессию и профили процессинга tracrРНК и crРНК пяти видов бактерий.

Материалы и методы.

Бактериальные штаммы и условия культивирования.

Следующие среды используют для роста бактерии на чашках петри: TSA (триптиказосоевый агар, Trypticase™ Soy Agar (TSA II) BD BBL, Becton Dickinson) с добавлением 3% овечьей крови в случае *S. mutans* (UA159) и агар ВНІ (сердечно-мозговой экстракт, BD Vacto™ Brain Heart Infusion, Becton Dickinson) в случае *L. innocua* (Clip 11262). При культивировании в жидких культурах, среда THY (Todd Hewitt Broth (ТНВ, Vacto, Becton Dickinson) с добавлением 0,2% дрожжевого экстракта (Servabacter®) используют в случае *S. mutans*, бульон ВНІ в случае *L. innocua*, жидкую среду ВНІ, содержащую 1% витаминную смесь VX (Difco, Becton Dickinson) в случае *N. meningitidis* (A Z2491), Broth МН (Mueller Hinton Broth, Oxoid), в том числе 1% витаминную смесь VX в случае *C. jejuni* (NCTC 11168; ATCC 700819) и TSB (Tryptic Soy Broth, BD BBL™ Trypticase™ Soy Broth) в случае *F. novicida* (U112). *S. mutans* инкубируют при 37°C, 5% CO<sub>2</sub> без встряхивания. Штаммы *L. innocua*, *N. meningitidis* и *F. novicida* выращивают в аэробных условиях при 37°C при встряхивании. *C. jejuni* выращивают при 37°C в микроаэрофильных условиях с использованием атмосферы CampyGen (Oxoid). Рост бактериальных клеток наблюдается с последующим измерением оптической плотности культур при 620 нм (OD<sub>620</sub> нм) через регулярные промежутки времени с помощью микропланшетного-ридера (BioTek PowerWave™).

Секвенирование бактериальных библиотек малых РНК.

*C. jejuni* NCTC 11168 (ATCC 700819), *F. novicida* U112, *L. innocua* Clip11262, *N. meningitidis* A Z2491 и *S. mutans* UA159 культивируют до середины логарифмической фазы роста, и тотальную РНК экстрагируют с помощью TRIzol (Sigma-Aldrich). 10 мкг общего количества РНК каждого штамма обрабатывают ДНКазой TURBO™ (Ambion) для того, чтобы удалить остатки геномной ДНК. Рибосомальные РНК удаляют с помощью использования Ribo-Zero™ rRNA Removal Kits® в случае грамм-положительных и грамм-отрицательных бактерий (Epicentre), в соответствии с инструкциями изготовителя. После очистки набором RNA Clean & Concentrator™-5 (Zymo Research), библиотеки получают с использованием ScriptMiner™ Small RNA-Seq Library Preparation Kit (Multiplex, совместимый с Illumina®), в соответствии с инструкциями изготовителя. РНК обрабатывают с помощью Tobacco Acid Phosphatase (TAP) (Epicentre). Колонки от RNA Clean & Concentrator™-5 (Zymo Research) используют для последующей очистки РНК, а Phusion® High-Fidelity DNA Polymerase (New England Biolabs) используют для ПЦР-амплификации. Специфические определяемые пользователем штрих-коды добавляют к каждой библиотеке (RNA-Seq Barcode Primers (Illumina®-совместимый) Epicentre), а образцы секвенируют в устройстве Next Generation Sequencing (CSF NGS Unit; в интернете по адресу "csf." после "ac.at") Vienna Biocenter, Вена, Австрия (одностороннее секвенирование Illumina).

Анализ данных секвенирования tracrРНК и crРНК.

Считываемые фрагменты при секвенировании РНК разделяют с помощью инструмента illumina2bam и обрабатывают с помощью (i) удаления адаптера последовательностей Illumina (cutadapt 1.0) и (ii) удаления 15 нуклеотидов на 3'-конце, для того, чтобы улучшить качество считываний. После удаления считываемых фрагментов меньших, чем 15 нт, считываемые фрагменты кДНК выравнивают с их соответствующим геномом с использованием Bowtie, допуская 2 ошибочные спаривания: *C. jejuni* (GenBank: NC\_002163), *F. novicida* (GenBank: NC\_008601), *N. meningitidis* (GenBank: NC\_003116), *L. innocua* (GenBank: NC003212) и *S. mutans* (GenBank: NC\_004350). Покрытие считываемых фрагментов рассчитывают для каждого положения нуклеотида отдельно для обеих цепей ДНК с использованием BEDTools-Version-2.15.0. Создают файл нормализованных изменений, содержащий покрытие в считываемых фрагментах на миллион (сфнм) и визуализируют с помощью инструмента Integrative Genomics Viewer (IGV)

("www.", а затем "broadinstitute.org/igv/") (фиг. 36). С использованием SAMTools flagstat80, долю отображенных считываемых фрагментов рассчитывают из, в общей сложности, 9914184 отображенных считываемых фрагментов в случае *C. jejuni*, 48205 считываемых фрагментов в случае *F. novicida*, 13110087 считываемых фрагментов в случае *N. meningitidis*, 161865 считываемых фрагментов в случае *L. innocua* и 1542239 считываемых фрагментов в случае *S. mutans*. Файл, содержащий количество считываемых фрагментов, который начинается (5') и заканчивается (3') в каждом отдельном положении нуклеотидов, создают и визуализируют в IGV. В случае каждого ортолога *tracrPНК* и *crPНК*, общее число полученных считываемых фрагментов рассчитывают с использованием SAMtools.

Анализ последовательности Cas9, множественное выравнивание последовательностей и конструкция дерева.

Программу Position-Specific Iterated (PSI)-BLAST используют для выявления гомологов семейства Cas9 в базе NCBI "non redundant database". Последовательности короче 800 аминокислот отбрасывают. Программу BLASTClust, настроенную с отсечением длины покрытия на 0,8 и порогом оценки покрытия (вес в битах делят на длину выравнивания) 0,8, используют для кластеризации оставшихся последовательностей (фиг. 38). Эта процедура дает 78 кластеров (48 из них были представлены только одной последовательностью). Один (или редко несколько представителей) выбирают из каждого кластера и множественное выравнивание этих последовательностей строят с помощью программы MUSCLE с параметрами по умолчанию, после чего корректируют вручную на основе локальных выравниваний, полученных с использованием программ PSI-BLAST и HHpred. Несколько последовательностей являются невыравниваемыми и также исключены из окончательных выравниваний. Уверенно выровненные блоки с 272 информационными позициями используют для реконструкции дерева методом максимального правдоподобия с помощью программы FastTree с параметрами по умолчанию: эволюционная модель JTT, дискретная гамма модель с 20 нормированными категориями. Эту же программу используют для расчета значений исходной загрузки.

Фиг. 38 иллюстрирует последовательности, которые были сгруппированы в соответствии с программой кластеризации BLASTclust. В случае анализа BLASTclust (см. Материалы и методы), выбирают только последовательности длиннее 800 аминокислот. Используют репрезентативные штаммы, несущие гены ортологи *cas9*. Некоторые последовательности не группируют, но проверяют в качестве последовательностей Cas9, в связи с наличием консервативных мотивов и/или других генов *cas* в их непосредственной близости.

Анализ локусов CRISPR-Cas.

Повторяющиеся последовательности CRISPR выявляют в базе данных CRISPRdb или прогнозируют с использованием инструмента CRISPRFinder (Grissa I и соавт., BMC Bioinformatics 20.07; 8:172; Grissa I и соавт., Nucleic Acids Res 2007). Гены CAS определяют с использованием алгоритма BLASTp и/или проверяют с помощью базы данных KEGG (в интернете на "www.", после чего kegg.jp/).

Прогнозирование *in silico* и анализ ортологов *tracrPНК*.

Предполагаемые антиповторы идентифицируют с использованием программного обеспечения Vector NTI® (Invitrogen) посредством скрининга дополнительных выродженных повторяющихся последовательностей, которые не принадлежат к блоку повтор-спейсер на обеих нитях соответствующих геномов, допуская вплоть до 15 ошибочных спариваний. Промоторы транскрипции и ро-независимые терминаторы прогнозируют с помощью программы BDGP Neural Network Promoter Prediction ("www.", а затем fruitfly.org/seq\_tools/promoter.html) и программного обеспечения TransTermHP соответственно. Множественные выравнивания последовательностей осуществляют с помощью программы MUSCLE, с параметрами по умолчанию. Выравнивания анализируют на присутствие мотивов консервативных структур с использованием алгоритма RNAalifold пакета Vienna RNA 2.0.

Результаты.

Системы CRISPR-Cas типа II широко распространены в бактерии.

В дополнение к *tracrPНК*-кодирующей ДНК и блоку повтор-спейсер, локусы CRISPR-Cas типа II, как: правило, состоят из примерно трех - четырех генов *cas*, организованных в оперон (фиг. 32A-B). Cas9 представляет собой характерной белок для типа II и участвует в этапах экспрессии и интерференции. CAS1 и CAS2 представляют собой ядерные белки, которые являются общими для всех систем CRISPR-Cas и участвуют в приобретении спейсера. Csn2 и Cas4 присутствуют только в подмножестве систем типа II и, как предполагают, играют роль в адаптации. Чтобы получить максимальное количество локусов CRISPR-Cas типа II, содержащих *tracrPНК*, сначала скринируют общедоступные геномы на предмет последовательностей, гомологичных уже аннотированным белкам Cas9. 235 ортологи Cas9 выявляют в 203 видах бактерий. Набор из 75 разнообразных последовательностей, репрезентативных для всех найденных ортологов Cas9, отбирают для дополнительного анализа (фиг. 32, 38, а также "Материалы и методы").

Фиг. 32 иллюстрирует (A) филогенетическое дерево репрезентативных последовательностей Cas9 из различных организмов, а также (B) репрезентативную архитектуру локуса Cas9. Указывают исходные значения, рассчитанные для каждого узла. Те же цветные ветви представляют выбранные подкластеры подобных ортологов Cas9. Длину повтора CRISPR в нуклеотидах, средний размер белка Cas9 в аминокислотах (aa) и архитектуру консенсусного локуса демонстрируют для каждого подкластера. \* -

gi|116628213; \*\* - gi|116627542; † - gi|34557790; ‡ - gi|34557932. Type II-A характеризуют с помощью cas9 - csx12, cas1, cas2, cas4. Type II-B характеризуют с помощью cas9, cas1, cas2 после чего с помощью варианта csn2. Type II-C характеризуют с помощью консервативного оперона cas9, cas1, cas2 (см. также фиг. 38).

Далее, проводят множественное выравнивание последовательностей выбранных ортологов Cas9. Сравнительный анализ демонстрирует высокое разнообразие в аминокислотной композиции и размере белка. Ортологи Cas9 разделяют только несколько одинаковых аминокислот, но все найденные последовательности имеют одинаковую архитектуру домена с центральным эндонуклеазным доменом HNH и расщепленным доменом RuvC/PHKseH. Длины белков Cas9 изменяются в диапазоне от 984 (*Campylobacter jejuni*) до 1629 (*Francisella novicida*) аминокислот с обычными размерами ~1100 или ~1400 аминокислот. В связи с большим разнообразием последовательностей Cas9, особенно по длине междоменных областей, выбирают только хорошо выровненные, информативные позиции полученного выравнивания, для конструирования филогенетического дерева анализируемых последовательностей (фиг. 32 и "Материалы и методы"). Ортологи Cas9 группируют в три основные, монофилетические кластеры с некоторыми отдельными последовательностями. Наблюдаемая топология дерева Cas9 хорошо согласуется с действующей классификацией локусов типа II, с предварительно определенными типом II-A и типом II-B, образующими отдельные, монофилетические кластеры. Для дополнительной характеристики кластеров, подробно рассматривают композиции оперона cas и последовательности повтора CRISPR всех перечисленных штаммов.

Субгруппирование Cas9 отражает разнообразие в архитектуре локусов CRISPR-Cas типа II.

Более глубокий анализ выбранных локусов типа II продемонстрировал, что кластеризация последовательностей ортолога Cas9 коррелирует с разнообразием в длине повтора CRISPR. Для большинства систем CRISPR-Cas типа II, длина повтора составляет 36 нуклеотидов (нт) с некоторыми вариациями в случае двух подкластеров дерева Cas9. В кластере типа II-A (фиг. 32), который содержит локусы, кодирующие длинный ортолог Cas9, который ранее назывался Csx12, повторы CRISPR имеют 37 нуклеотидов в длину. Небольшой подкластер состоит из последовательностей, полученных из бактерий, принадлежащих к *Bacteroidetes* phylum (фиг. 32), и характеризуются необычайно длинными повторами CRISPR, до 48 нуклеотидов в длину. Кроме того, было замечено, что субгруппирование последовательностей Cas9 коррелирует с различными архитектурами оперона cas, как показано на фиг. 32. Третий основной кластер (фиг. 32) и внешние локусы (фиг. 32) состоят в основном из минимального оперона, состоящего из генов cas9, cas1 и cas2, за исключением некоторых неполных локусов, которые рассматривают ниже. Все другие локусы из двух первых основных кластеров связаны с четвертым геном, главным образом cas4, специфическим для типа II-A или csn2-подобным, специфическим для типа II-B (фиг. 32). Были определены гены, кодирующие короткие варианты белка Csn2, Csn2a, в пределах локусов, схожих с CRISPR01 *S. rugosus* типа II-B и CRISPR3 *S. thermophilus* (фиг. 32). Обнаружено, что более длинный вариант Csn2, Csn2b ассоциирован с локусами, схожими с CRISPR1 *S. thermophilus* типа II-B (фиг. 32). Интересно, что было определены дополнительные предполагаемые гены cas, кодирующие белки, без очевидного сходства последовательности с ранее описанными вариантами Csn2. Один из этих охарактеризованных белков ассоциируется исключительно с локусами видов *Mycoplasma* типа II-B (фиг. 32 и 33). Два других закодированы в локусах видов *Staphylococcus* типа II-B (фиг. 33). Во всех случаях разнообразие архитектуры оперона cas, соответствует субгруппированию последовательностей Cas9. Эти характеристики вместе с общей топологией дерева Cas9, разделенные у трех основных, различных, монофилетических кластеров, что позволило предложить новое, дополнительное разделение системы CRISPR-Cas типа II на три подтипа. Тип II связывают с Csx12-подобным Cas9 и Cas4, тип II-B связывают с Csn2-подобным, а тип II-C содержит только минимальный набор генов cas9, cas1 и cas2, как показано на фиг. 32.

Фиг. 33 иллюстрирует архитектуру CRISPR-Cas типа II отдельных видов бактерий. Вертикальные полосы отмечают группы локусов, которые кодируют ортологи Cas9, принадлежащие к одному и тому же подкластеру схемы (по сравнению с фиг. 32). Горизонтальная черная полоса, лидерная последовательность; черные прямоугольники и ромбы, блок повтор-спейсер. Предполагаемые антиповторы показаны с помощью стрелок, указывающих направление предполагаемой транскрипции ортолога tracrPHK. Отметим, что в случае локусов, которые не были проверены экспериментально, считается, что блок повтор-спейсер CRISPR считают в данном документе, транскрибируется с той же цепи, что и оперон cas. Соответственно, указывают направление транскрипции предполагаемого ортолога tracrPHK.

In silico прогнозирование новых ортологов tracrPHK.

Локусы типа II, выбранные ранее на основе 75 репрезентативных ортологов Cas9, скринировали на присутствие потенциальных ортологов tracrPHK. Предыдущий анализ, проведенный на ограниченном количестве последовательностей tracrPHK, показал, что ни последовательности tracrPHK, ни их локализация в пределах локусов CRISPR-Cas, как оказалось, не являются консервативными. Тем не менее, как упоминалось выше, tracrPHK также характеризуются последовательностью антиповтора, способной к спариванию оснований с каждым из повторов pre-crPHK для образования дуплексов повтора tracrPHK:precrPHK, которые расщепляются с помощью PHKазы III в присутствии Cas9. Для того чтобы

прогнозировать новые *tracrPНК*, используют эти характеристики и применяют следующий процесс: (i) скрининг на потенциальные антиповторы (образование пары оснований последовательности с повторами CRISPR) в локусах CRISPR-Cas, (ii) выбор антиповторов, расположенных в межгенных областях, (iii) проверка спаривания оснований антиповтор:повтор CRISPR, и (iv) прогнозирование промоторов и розависимых терминаторов транскрипции, ассоциированных с выявленными *tracrPНК*.

С целью скрининга предполагаемых антиповторов, авторы выявили последовательности повторов в базе данных CRISPRdb или, если информация не была доступна, спрогнозировали последовательности повторов, используя программное обеспечение CRISPRfinder. В предыдущем исследовании авторы изобретения экспериментально продемонстрировали, что направление транскрипции блока повтор-спейсер, по сравнению с опероном *cas*, отличается среди локусов. В данном документе секвенирование РНК подтверждает это наблюдение. В некоторых из анализируемого локусов, а именно в *F. novicida*, *N. meningitidis* и *C. jejuni*, блок повтор-спейсер транскрибируется в противоположном направлении оперона *cas* (см. раздел "Глубокое секвенирование РНК подтверждает экспрессию новых ортологов *tracrPНК*" и фиг. 33 и 34), в то время как в *S. ruogenes*, *S. mutans*, *S. thermophilus* и *L. innosua*, блок и оперон *cas* транскрибируются в одном и том же направлении. Это представляет собой единственные данные об экспрессии блока повтор-спейсер типа II, доступные на сегодняшний день. Чтобы спрогнозировать, направление транскрипции других блоков повтор-спейсер, авторы учли предыдущее наблюдение, что последние повторы блоков, как правило, мутируют. Это замечание согласуется с текущей моделью приобретения спейсера, в которой, как правило, первый повтор блока дублируют в процессе вставки спейсерной последовательности на фазе адаптации. В случае 37 блоков повтор-спейсер, возможно определить мутировавший повтор в предполагаемом конце блоков. Было отмечено, что предполагаемая ориентация транскрипции в случае блока повтор-спейсер *N. meningitidis* и *C. jejuni* может быть противоположной ориентации, определенной экспериментально (секвенирование РНК и анализ нозерн-блоттинг). Также как предполагаемая ориентация не одинакова в пределах кластеров, а, в большинстве случаев, можно было обнаружить потенциальные промоторы на обоих концах блоков, с учетом этого считается, что транскрипция блоков повтор-спейсер происходит в том же направлении, что и транскрипцию оперона *cas*, если не подтверждено обратное.

Фиг. 34 иллюстрирует коэкспрессию *tracrPНК* и *pre-crPНК* из выбранных систем CRISPR Cas типа II. Проиллюстрирована архитектура локусов CRISPR с подтвержденными положениями и направлениями транскрипции *tracrPНК* и *pre-crPНК*. Верхние последовательности, повторы *pre-crPНК*; нижние последовательности, образование пары оснований последовательностей *tracrPНК* с повторами *crPНК*. Предполагаемые сайты процессинга РНК, как показано с помощью секвенирования РНК, указывают стрелками. В случае каждого локуса, размеры стрелки представляют относительные количества полученных концов 5' и 3' (см. также фиг. 37).

Фиг. 37 перечисляет все ортологи *tracrPНК* и зрелые *crPНК*, полученные с помощью секвенирования изученных бактериальных видов, в том числе координаты (представляющие интерес области) и соответствующие последовательности с ДНК (от 5' до 3'). Стрелки указывают направление транскрипции (цепь). Приведено количество считываемых фрагментов кДНК (рассчитываемое с использованием SAMtools), покрытие (в процентах от картированных считываемых фрагментов) и преобладающих концов, ассоциированных с каждым транскриптом. Отображают количества считываемых фрагментов, начинающихся или заканчивающихся в каждом положении нуклеотида около 5'- и 3'-концов каждого транскрипта. Указывают размеры каждой из зрелых форм *crPНК*. Номер, присвоенный каждому виду *crPНК*, соответствует положению последовательности спейсера в *pre-crPНК*, и соответствии с CRISPRdb. Номер, присвоенный каждому виду *tracrPНК*, соответствует различным формам одного и того же транскрипта.

Затем скринировали выбранные локусы CRISPR-Cas, в том числе последовательности, расположенные на 1 т.п.о. против хода транскрипции и по ходу транскрипции в обеих нитях, на наличие возможных последовательностей повторов, которые не относят к блоку повтор-спейсер, допускающих до 15 ошибочных спариваний. В среднем, обнаруживают от одной до трех вырожденных повторяющихся последовательностей на локус, который может соответствовать антиповторам ортологов *tracrPНК*, и выбирают последовательности, расположенные в пределах межгенных областей. Предполагаемые антиповторы находят в четырех типичных локализациях: выше гена *cas9*, в области между *cas9* и *cas1* и выше или ниже по ходу транскрипции от блока повтор-спейсер (фиг. 33). Для каждой полученной последовательности, проверяли степень спаривания оснований, образованной между повтором и антиповтором (фиг. 44) с помощью прогнозирования возможного взаимодействия РНК:РНК, особенно фокусируясь на кандидатах с более длинной и областью полной комплементарности, образующем оптимальную двухцепочечную структуру в случае обработки РНКазой III. Чтобы предсказать промоторы и терминаторы транскрипции, фланкирующие антиповтор, устанавливают предполагаемые сайты начала и терминации транскрипции, которые будут содержаться в пределах области, расположенной максимально на 200 нт против хода транскрипции и на 100 нт по ходу транскрипции от последовательности антиповтора, соответственно, на основе предыдущих наблюдений. Как упоминалось выше, отсутствует экспериментальная информация по направлению транскрипции большинства блоков повтор-спейсер систем типа II. In silico алгоритмы прогнозирования промотора часто дают ложноположительные результаты и указывают на пред-

полагаемые промоторы, которые могли бы приводить к транскрипции блоков повтор-спейсер с обеих цепей. В некоторых случаях невозможно спрогнозировать терминаторы транскрипции, хотя экспрессия ортолога *tracrPHK* может быть подтверждена экспериментально, о чем свидетельствует локус *S. jejuni* (см. пункт "Глубокое секвенирование PHK подтверждает экспрессию новых ортологов *tracrPHK*"). Авторы изобретения предлагаем рассматривать прогнозирование промотора и терминатора транскрипции только в качестве поддерживающего, но не обязательного, этапа руководства, описанного выше.

Фиг. 44 иллюстрирует спрогнозированное образование пары оснований повтор *pre-crPHK*:антиповтор *tracrPHK* отдельных видов бактерий. b) Локусы CRISPR относят к системе CRISPR-Cas (Nmeni/CASS4) типа II. Используется номенклатура в соответствии с базой данных CRISPR (CRISPRdb). Следует обратить внимание, что LMD-9 *S. thermophilus* и *W. succinogenes* содержат два локуса типа II. c) Верхнюю последовательность, консенсусную последовательность повтора *pre-crPHK* (от 5' к 3'); нижнюю последовательность, последовательность гомолога *tracrPHK* гибридузуют с повтором (антиповтор; от 3' к 5'). Следует отметить, что данная последовательность повтора основана на предположении, что блок повтор-спейсер CRISPR транскрибируют той же цепи, что и оперон *cas*. В случае последовательностей, которые были подтверждены экспериментально в данном исследовании, данные секвенирования PHK принимают во внимание для того, чтобы определить образование пары оснований. См. фиг. 33. d) Два возможных антиповтора идентифицируют в *F. tularensis* подвида *novicida*, *W. succinogenes* и локус гамма протеобактерии HTCC5015 типа II. Образование пары верхней последовательности, антиповтор в пределах предполагаемой лидерной последовательности; образование пары нижней последовательности, антиповтор по ходу транскрипции от блока повтор-спейсер. См. фиг. 33. e) Два возможных антиповтора идентифицируют в локусе *S. wadsworthensis* типа II-A. Образование пары верхней последовательности, антиповтор; образование пары нижней последовательности, антиповтор в пределах предполагаемой лидерной последовательности. См. фиг. 33. f) Два возможных антиповтора идентифицируют в локусе *L. gasseri* типа II-A. Образование пары верхней последовательности, антиповтор против хода транскрипции от *cas9*; образование пары нижней последовательности, антиповтор между генами *cas9* и *cas1*. См. фиг. 33. g) Два возможных антиповтора идентифицируют в локусе *S. jejuni* типа II-C. Образование пары верхней последовательности, антиповтор против хода транскрипции от *cas9*; образование пары нижней последовательности, антиповтор по ходу транскрипции от блока повтор-спейсер. См. фиг. 33. h) Два возможных антиповтора идентифицируют в локусе *R. rubrum* типа II-C. Образование пары верхней последовательности, антиповтор по ходу транскрипции от блока повтор-спейсер; образование пары нижней последовательности, антиповтор против хода транскрипции от *cas1*. См. фиг. 33.

Множество ортологов *tracrPHK*.

Авторы изобретения спрогнозировали предполагаемые ортологи *tracrPHK* в 56 локусах из 75, выбранных ранее. Результаты прогнозирования проиллюстрированы на фиг. 33. Как уже упоминалось, направление транскрипции *tracrPHK*, указанное на этой фигуре, является гипотетическим и основано на указанном направлении транскрипции блока повтор-спейсер. Как отмечалось ранее, последовательности, кодирующие предполагаемые ортологи *tracrPHK*, были найдены выше, в пределах и ниже от оперона *cas*, а также выше блока повтор-спейсер, в том числе предполагаемых лидерных последовательностей, обычно встречающихся в локусах типа II (фиг. 33). Тем не менее, нами было отмечено, что антиповторы подобной локализации в пределах локусов CRISPR-Cas могут транскрибироваться в различных направлениях (как это наблюдается при сравнении, например, *Lactobacillus rhamnosus* и *Eubacterium rectale* или *Mycoplasma mobile* и *S. pyogenes* или *N. meningitidis*) (фиг. 33). Следует отметить, что локусы, сгруппированные в пределах одного и того же подкластера дерева *Cas9*, имеют общую архитектуру с точки зрения положения *tracrPHK*-кодирующего гена. Определяют антиповторы вокруг блока повтор-спейсер в локусах типа II-A, и, в основном, ниже от гена *cas9* в типах II-B и II-C с несколькими важными исключениями в случае предполагаемой *tracrPHK*, расположенной между *cas9* и *cas1* в трех различных подкластерах типа II-B.

Некоторые локусы CRISPR-Cas типа II имеют дефектные блоки повтор-спейсер и/или ортологи *tracrPHK*.

В случае шести локусов типа II (*Fusobacterium nucleatum*, *Aminomonas paucivorans*, *Helicobacter mustelae*, *Azospirillum* sp., *Prevotella ruminicola* и *Akkermansia muciniphila*) определяют потенциальные антиповторы со слабым спариванием оснований с последовательностью повтора или расположенными в пределах открытых рамок считывания. Следует отметить, что в этих локусах слабый антиповтор в пределах открытой рамки считывания гена кодирует предполагаемую АТФазу в *A. paucivorans*, сильный антиповтор в пределах первых 100 нт гена *cas9* в *Azospirillum* sp. Определяют B510 и сильное перекрытие антиповтора как с *cas9*, так и с *cas1* в *A. muciniphila* (фиг. 33). В случае двенадцати дополнительных локусов (*Peptoniphilus duerdenii*, *Coprococcus catus*, *Acidaminococcus intestini*, *Catenibacterium mitsuokai*, *Staphylococcus pseudintermedius*, *Ilyobacter polytropus*, *Elusimicrobium minutum*, *Bacteroides fragilis*, *Acidothermus cellulolyticus*, *Corynebacterium diptheriae*, *Bifidobacterium longum* и *Bifidobacterium dentium*) какой-либо предполагаемый анти-повтор не был обнаружен. Не существует доступной информации об экспрессии *pre-crPHK* и процессинга этих локусах CRISPR-Cas. Таким образом, функциональность систем типа II в отсутствие четко определенного ортолога *tracrPHK* еще предстоит изучить. В случае семи

проанализированных локусов не был выявлен какой-либо блок повтор-спейсер (*Parasutterella excrementihominis*, *Bacillus cereus*, *Ruminococcus albus*, *Rhodopseudomonas palustris*, *Nitrobacter hamburgensis*, *Bradyrhizobium* sp. и *Prevotella micans*) (фиг. 33) и в трех из них (*Bradyrhizobium* Sp. ВТАil, *N. hamburgensis* и *B. cereus*) обнаруживают единственный ген *cas9* и никаких других генов *cas* в непосредственной близости не найдено. Для этих трех локусов, авторы не смогли спрогнозировать какую-либо последовательность малой РНК против хода транскрипции или по ходу транскрипции от гена *cas9*. В случае *R. albus* и *P. excrementihominis*, геномный контиг, содержащий *cas9*, является слишком коротким, чтобы осуществить прогнозирование блока повтор-спейсер.

Глубокое секвенирование РНК подтверждает экспрессию новых ортологов *tracr*РНК.

Для того чтобы проверить спрогнозированные *in silico* *tracr*РНК и определить совместно процессируемые структуры *tracr*РНК:*pre-cr*РНК, РНК из некоторых грамположительных (*S. mutans* и *L. innocua*) и грамотрицательных (*N. meningitidis*, *C. jejuni* и *F. novicida*) бактерий анализируют с помощью глубокого секвенирования. Получают последовательности ортологов *tracr*РНК и процессированные *cr*РНК (фиг. 36 и фиг. 37). В соответствии с ранее опубликованными данными дифференциального секвенирования *tracr*РНК в *S. ruogenes26*, ортологи *tracr*РНК широко представлены в библиотеках, начиная от 0,08 до 6,2% от общего количества картированных считываемых фрагментов. Процессированные *tracr*РНК также присутствуют в большем количестве, чем первичные транскрипты, от 66 до более чем 95% от общего количества считываемых фрагментов *tracr*РНК (фиг. 36 и 37).

Фиг. 36 иллюстрирует экспрессию бактериальных ортологов *tracr*РНК и *cr*РНК, обнаруженных с помощью глубокого секвенирования РНК. Профили экспрессии ортологов *tracr*РНК и *cr*РНК отдельных бактериальных штаммов представляют по соответствующим геномам с помощью гистограммы (изображения, снятые с помощью инструмента Integrative Genomics Viewer (IGV)). *Campylobacter jejuni* (GenBank: NC\_002163), *Francisella novicida* (GenBank: NC\_008601), *Neisseria meningitidis* (GenBank: NC\_003116), *Listeria innocua* (GenBank: NC\_003212) и *Streptococcus mutans* (GenBank: NC004350). Даны геномные координаты. а) Покрытие последовательности рассчитывают с использованием BEDTools-Version-2.15.0 (Масштаб приведен в считываемых фрагментах на миллион). б) Указано распределение считываемых фрагментов, которое начинается (5') и заканчивается (3') в каждой позиции нуклеотида (Масштаб приведен в количестве считываемых фрагментов). Верхние части соответствуют транскриптам с положительно-полярной цепью, а нижние части соответствуют транскриптам с отрицательно-полярной цепью. Величины отрицательного покрытия и пики, представленные ниже осей, демонстрируют транскрипцию с отрицательно-полярной цепью генома. Преобладающие 5'- и 3'-концы считываемых фрагментов строят для всех РНК. Следует отметить, что с учетом низкого качества библиотеки сДНК *L. innocua*, считываемые фрагменты укорачивают в случае *cr*РНК и наблюдают накопление считываемых фрагментов на 3'-конце *tracr*РНК, по-видимому, ввиду деградации РНК.

Для оценки 5'-концов первичных транскриптов *tracr*РНК, авторы изобретения проанализировали относительное содержание всех 5'-концов считываемых фрагментов *tracr*РНК и выявили наиболее выделяющиеся считываемые фрагменты против хода транскрипции или в непосредственной близости к 5'-концу прогнозируемой последовательности антиповтора. 5'-концы ортологов *tracr*РНК дополнительно подтверждают с использованием алгоритма прогнозирования промотора. Определенные 5'-концы *tracr*РНК из *S. mutans*, *L. innocua* и *N. meningitidis* коррелируют как с *in silico* прогнозированием, так и с анализом нозерн-блоттинг экспрессии *tracr*РНК26. Самый выделяющийся 5'-конец *tracr*РНК *C. jejuni* идентифицируют в середине последовательности антиповтора. На 5 нуклеотидов выше обнаружен дополнительный предполагаемый 5'-конец, что коррелирует с *in silico* прогнозированием и обеспечивает более длинную последовательность взаимодействия с последовательностью повтора CRISPR. Авторы изобретения выявили относительно небольшое количество считываемых фрагментов из библиотеки *F. novicida*, что соответствует почти исключительно обработанным транскриптам. Анализ очень небольшого количества считываемых фрагментов первичных транскриптов обеспечивает 5'-конец, который соответствует вполне определенным *in silico* спрогнозированным промоторам. Исследование с использованием нозерн-блоттинга *tracr*РНК *F. novicida* дополнительно подтверждает правильность прогнозирования, демонстрируя низкую численность транскриптов около 90 нуклеотидов в длину. Результаты приведены в табл. 2. Для всех исследованных видов, кроме *N. meningitidis*, первичные транскрипты *tracr*РНК определяют в качестве отдельных видов малых РНК, от 75 до 100 нуклеотидов в длину. В случае *N. meningitidis*, обнаруживают преобладающую первичную форму *tracr*РНК из ~110 нт и предполагаемый более длинный транскрипт из ~170 нт, представленный очень низким количеством считываемых фрагментов и обнаруженный ранее в слабой полосе при нозерн-блот анализе.

Отдельные ортологи tracrPHK

Штаммы	Транскрипт	5'-конец <sup>b</sup>			3'-конец <sup>c</sup>	Длина (нт)
		PHK-посл.		Прогнозированный		
		Первый считываемый фрагмент	Наиболее выделяющийся			
SF370 <i>S. pyogenes</i>	первичный	-	854546	-	854376	171
	первичный	-	854464	-		89
	процессированный	-	854450	-		~75
NCTC 11168 <i>C. jejuni</i>	первичный	1455497	1455502	1455497	1455570	~75
	процессированный	-	1455509	-		~60
Clp 11262 <i>L. innocua</i>	первичный	2774774	2774774	2774773	2774863	~90
	процессированный	-	2774788	-		~75
UA 159 <i>S. mutans</i>	первичный	1335040	1335040	1355039	1335141	~100
	процессированный	-	1335054 1335062	-		~85 ~80
A Z2491 N. <i>meningitidis</i>	первичный	614158	614162	614154	614333	~175
	первичный	614223	614225	614223		~110
	процессированный	-	614240	-		~90
U112 <i>F. novicida</i>	первичный	817144	-	817145 817154	817065	~80
	процессированный	-	817138 817128	-		~75 ~65
LMD-9 <i>S. thermophilus</i>	первичный	-	-	1384330	1384425	~95
	первичный	-	-	646654	646762	~110
Pm70 <i>P. multocida</i>	первичный	-	-	1327287	1327396	~110
163K <i>M. mobile</i>	первичный	-	-	49470	49361	~110

<sup>a</sup> Ортологи tracrPHK *S. thermophilus*, *P. multocida* и *M. mobile* прогнозируют in silico.

<sup>b</sup> PHK-посл., полученная с помощью секвенирования PHK (табл.S3); первый считываемый фрагмент, первое положение на 5'-конце выявленное с помощью секвенирования; наиболее выделяющийся, самый распространенный 5'-конец, согласно данным PHK-секвенирования; предсказанный, предсказанный in silico сайт начала транскрипции; подчеркнуты 5'-концы, выбранных для выравнивания первичной tracrPHK.

<sup>c</sup> Рассчитанный 3'-конец в соответствии с данными PHK-секвенирования. и предполагаемого терминатора транскрипции.

Сайты коэкспрессии tracrPHK и pre-crPHK лежат в области антиповтор:повтор.

Исследуют процессированные транскрипты tracrPHK путем анализа 5'-концов многокопийных tracrPHK в пределах прогнозированной последовательности антиповтора и 5'-концов многокопийной зрелой crPHK (фиг. 34 и 45). Для всех видов определяют главные 5'-концы ортологов tracrPHK, которые могут возникнуть в результате совместного процессинга дуплексов повторов tracrPHK:crPHK с помощью PHKазы III. Также определяют процессированные 5'-концы crPHK, которые, скорее всего, получают из второго события созревания с помощью предполагаемой обрезки, что согласуется с предыдущими наблюдениями. Стоит отметить, что в близкородственных парах PHK *S. pyogenes*, *S. mutans* и *L. innocua* наблюдают тот же сайт процессинга вокруг пары оснований G:C в середине последовательности антиповтора. И в *S. mutans*, и в *L. innocua*, обнаруживают дополнительные выступающие 5'-концы tracrPHK и 3'-концы crPHK, из чего можно предположить, что происходит дополнительная обрезка дуплекса tracrPHK:crPHK, где 3'-конец crPHK укорачивается дополнительно к уже упоминавшейся обрезке 5'-конца, после катализируемого PHKазой III первого события процессинга. Кроме того, в *C. jejuni* обнаруживают только небольшое количество 3'-концов crPHK, которые могут соответствовать структурам процессируемым PHKазой III и выявили соответствующие 5'-концы прошедшей процессинг tracrPHK. Таким образом, предполагаемая обрезка, дуплексов tracrPHK:crPHK после первоначального расщепления PHKазой III может приводить к образованию более короткой части, производной повтора, в зрелой crPHK, с получением более коротких дуплексов tracrPHK:crPHK, стабилизированных тройным образованием пары оснований G:C в случае взаимодействия с Cas9 эндонуклеазой и последующим расщеплением ДНК-мишеней. Дуплекс PHK *N. meningitidis*, как предполагают, будет процессироваться в двух основных сайтах ближе к 3'-концу от повтора CRISPR, что приводит к образованию длинной части, производной от повтора, в зрелой crPHK и стабильного взаимодействия PHK:PHK, несмотря на центральное выпетливание в пределах дуплекса. Интересно, что дуплекс tracrPHK:pre-crPHK *F. novicida*, кажется, расщепляется в пределах области низкой комплементарности и некоторые из выявленных широко распространенных 5'-концов tracrPHK указывают на возможную дополнительную обрезку ее без сопутствующей обрезки crPHK. Различия в размерах первичных транскриптов и в локализация сайтов процессинга приводят к

различным длинам прошедшей процессинг *tracr*PHK в пределах от ~65 до 85 нт. Координаты и размеры транскриптов наиболее выделяющихся из обработанной *tracr*PHK приведены в табл. 2 и на фиг. 37. Наблюдаемые структуры обработки *tracr*PHK и *cr*PHK хорошо согласуются с предложенной ранее моделью двух событий созревания. Предполагаемая дополнительная обрезка некоторых из 5'-концов *tracr*PHK и 3'-концов *cr*PHK может происходить во время второго события созревания или, альтернативно, быть артефактом подготовки библиотеки кДНК или секвенирования РНК. Природа этого процессинга требует дополнительных исследований.

Последовательности ортологов *tracr*PHK являются весьма разнообразными. Также определяют сходства последовательностей отдельных ортологов *tracr*PHK. Проводят множественные выравнивания последовательностей первичных транскриптов *tracr*PHK *S. pyogenes* (только форма из 89 нт) *S. mutans*, *L. innocua* и *N. meningitidis* (только форма из 110 нт), *S. thermophilus*, *P. multocida* и *M. mobile* (табл. 2, фиг. 35). Наблюдают большое разнообразие в последовательностях *tracr*PHK, но значительную консервативность последовательностей из близкородственных локусов CRISPR-Cas. *tracr*PHK от *L. innocua*, *S. pyogenes*, *S. mutans* и *S. thermophiles* имеют идентичность в среднем 77%, а *tracr*PHK от *N. meningitidis* и *P. multocida* имеют идентичность 82%, согласно парным выравниванием. Средняя идентичность анализируемых последовательностей *tracr*PHK составляет 56%, что сравнимо с идентичностью случайных последовательностей РНК. Это наблюдение дополнительно подтверждает, что прогнозирование ортологов *tracr*PHK на основе сходства последовательности может быть выполнено только в случае близкородственных локусов. Авторы изобретения также осуществляли поиск возможных консервативных структур *tracr*PHK, но не нашли какое-либо значительное сходство, кроме одного совместного изменения и консервативной структуры терминатора транскрипции (фиг. 35).

Фиг. 35 иллюстрирует разнообразие последовательности ортологов *tracr*PHK. Множественное выравнивание последовательности *tracr*PHK. *S. thermophilus* и *S. thermophilus*2, *tracr*PHK, ассоциированная с ортологами Cas9 SEQ ID NO:41 и SEQ ID NO:40 соответственно. Черный, высококонсервативный; темно-серый, консервативный; светло-серый, малоконсервативный. Прогнозируемые консенсусные структуры изображают в верхней части выравнивания. Стрелки указывают нуклеотидные ковариации. *Tracr*PHK *S. pyogenes* SF370, *S. mutans* UA159, *L. innocua* Clip 11262, *C. jejuni* NCTC 11168, *F. novicida* U112 и *N. meningitidis* A Z2491 подтверждают секвенированием РНК и анализом нозерн-блоттинг. *Tracr*PHK *S. thermophiles* LMD-9 подтверждают анализом нозерн-блоттинг. *Tracr*PHK Pm70 *P. multocida* подтверждают высоким сходством локуса CRISPR-Cas с таковым из *N. meningitidis* A Z2491. *Tracr*PHK 163K *M. mobile* подтверждают *in silico*, исходя из однозначных прогнозирований промотора и терминатора транскрипции.

Пример 4. Перепрофилирование CRISPR в качестве РНК-управляемой платформы для последовательность-специфичного контроля экспрессии генов.

Регуляция гена-мишени в общегеномном масштабе представляет собой мощную стратегию для исследования, вмешательства и создания клеточных систем. Был разработан новый способ контроля за экспрессией гена, основанный на Cas9, РНК-управляемой ДНКэндонуклеазы из системы CRISPR типа II. Этот пример демонстрирует, что каталитически мертвый Cas9, не имеющий эндонуклеазной активности, при совместной экспрессии с "направляющей" РНК, формирует комплекс распознавания ДНК, который может специфически помешать элонгации транскрипции, связыванию РНК-полимеразы или связыванию факторов транскрипции. Эта система, называемая интерференцией CRISPR (CRISPRi), может эффективно подавлять экспрессию нацеленных генов в *Escherichia coli* без каких-либо обнаруживаемых внецелевых эффектов. CRISPRi может быть использована для подавления нескольких генов-мишеней одновременно, и ее эффекты являются обратимыми. Кроме того, система может быть приспособлена для репрессии генов в клетках млекопитающих. Эта РНК-управляемая платформа распознавания ДНК обеспечивает простой подход к избирательному вмешательству в экспрессию гена в общегеномном масштабе.

Материалы и методы.

Штаммы и среда.

Штамм 1655 K-12 *Escherichia coli* используют в качестве штамма-хозяина для измерений флуоресценции *in vivo*. Штамм, производный MG1655 *E. coli*, который эндогенно экспрессирует вариант РНК полимеразы с меткой эпитопа 3x-FLAG, прикрепленной к С-концу субъединицы RpoC, используют во всех экспериментах секвенирования. Богатую на EZ среду определенного состава (EZ-RDM, Teknoka) используют в качестве питательной среды в случае анализов флуоресценции *in vivo*. Генетическую трансформацию и верификацию преобразования проводят с использованием стандартных протоколов, с использованием генов AmpR, CmR или KanR в качестве селективных маркеров.

Конструкция плазмиды и клонирование гена *E. coli*.

Гены Cas9 и dCas9 клонируют из ранее описанного вектора pMJ806 и pMJ841 соответственно. Гены амплифицируют с помощью ГЦР и встраивают в вектор, содержащий ангидротетрациклин(аТс)-индуцируемый промотор PLtetO-1, хлорамфеникол-селектируемый маркер и точку начала репликации p15A. Матрицу *sg*PHK клонируют в вектор, содержащий минимальный синтетический промотор (J23119) с аннотированным сайтом инициации транскрипции, селективным маркером устойчивости к ампициллину и точкой начала репликации ColE1. Обратную ПЦР используют для получения кассет *sg*PHK с новыми

ми комплементарными областями из 20 п.о. Чтобы вставить флуоресцентные репортерные гены в геномы *E.coli*, ген флуоресценции сначала клонируют в исходный вектор, который затем амплифицируют с помощью ПЦР, чтобы получить линейаризованные фрагменты ДНК, содержащие 573' UTR последовательности *nsfA*, флуоресцентный ген и выбираемый маркер *KanR*. Штамм *E.coli* MG1655 трансформируют с температурно-чувствительной плазмидой *pKD46*, которая содержит  $\lambda$ -красные рекомбинантные белки (*Exo*, *Beta* и *Gamma*). Клеточные культуры выращивают при 30°C до оптической плотности (600 нм) ~0,5 и добавляют 0,2% арабинозы для того, чтобы вызвать экспрессию  $\lambda$ -красных рекомбинантных белков в течение 1 ч. Клетки собирают при 4°C и используют для трансформации линейаризованных фрагментов ДНК с помощью электропорации. Клетки, которые содержат правильные вставки генома, отбирают с использованием 50 мкг/мл канамицина.

Проточная цитометрия и анализ.

Штаммы культивируют в EZ-RDM, содержащей 100 мкг/мл карбенициллина и 34 мкг/мл хлорамфеникола в 2 мл 96-луночные глубоколоночные планшеты (Costar 3960) в течение ночи при 37°C и 1200 об/мин. 1 мкл этой культуры, полученной за эту ночь, затем добавляют к 249 мкл свежей EZ-RDM с теми же концентрациями антибиотика с 2 мкМ аТс, добавленной для индуцирования получения белка dCas9. При условии, что клетки выращивают до середины фазы логарифмического роста (~4 ч), уровни флуоресценции белка определяют с использованием проточного цитометра LSRII (BD Biosciences), оснащенного пробоотборником с высокой пропускной способностью. Клетки отбирают с низкой скоростью потока до тех пор, пока не соберут по меньшей мере 20000 клеток. Данные анализируют с помощью FCS Express (De Novo Software) с помощью селекции на многоугольной области, содержащей 60% клеточной популяции в диаграмме прямое светорассеяние - боковое светорассеяние. В случае каждого эксперимента, измеряют три культуры, а их стандартное отклонение указывают в величине ошибки.

Анализ  $\beta$ -галактозидазы.

Чтобы осуществить анализ  $\beta$ -галактозидазы, 1 мкл культуры, полученной за ночь таким образом, как описано выше, добавляют к 249 мкл свежей EZ-RDM с теми же концентрациями антибиотика с 2 мкМ аТс, с или без 1 мМ изопропил- $\beta$ -D-1-тиогалактопиранозида (IPTG). Клетки выращивают до середины фазы логарифмического роста. Активность *LacZ* из 100 мкл этой культуры измеряют с использованием набора для анализа  $\beta$ -галактозидазы дрожжей (Pierce) в соответствии с инструкциями.

Выделение и очистка тотальной РНК.

Для каждого образца моноклональную культуру *E.coli* выращивают при 37°C с OD (600 нм) 0,1 в 500 мл EZ-RDM к ранней фазе логарифмического роста (OD 0,45±0,05), после чего клетки собирают с помощью фильтрации сквозь 0,22 мкм нитроцеллюлозные фильтры (GE) и замораживают в жидком азоте для того, чтобы одновременно полностью остановить транскрипцию. Замороженные клетки (100 мкг) измельчают на мельнице-смесителе Qiagen TissueLyser II 6 раз при 15 Гц в течение 3 мин в присутствии 500 мкл замороженного буфера для лизиса (20 мМ трис pH 8, 0,4% тритон X-100, 0,1% NP-40, 100 мМ NH<sub>4</sub>Cl, 50 ед./мл SUPERase•In (Ambion) и 1× смесь ингибиторов протеаз (Complete, без EDTA, Roche) с добавлением 10 мМ MnCl<sub>2</sub> и 15 мкМ ингибитора транскрипции Tagetin (Epicentre).

Лизат ресуспендируют на льду с помощью пипетки. ДНКазу I RQ1 (в общем 110 ед., Promega) добавляют и инкубируют в течение 20 мин на льду. Реакцию гасят с EDTA (25 мМ в конце), а лизат осветляют при 4°C с помощью центрифугирования при 20000 g в течение 10 мин. Лизат загружают в колонку PD MiniTrap G-25 (GE Healthcare) и элюируют с буфером для лизиса, дополненного 1 мМ EDTA.

Очистка тотальной мРНК.

Тотальную РНК очищают из очищенного лизата с использованием набора miRNeasy (Qiagen). 1 мкг РНК в 20 мкл 10 мМ трис pH 7 смешивают с равным объемом 2× щелочного раствора для фрагментации (2 мМ EDTA, 10 мМ Na<sub>2</sub>CO<sub>3</sub>, 90 мМ NaHCO<sub>3</sub>, pH 9,3) и инкубируют в течение ~25 мин при 95°C с целью получения фрагментов в диапазоне от 30 до 100 нт. Реакцию фрагментации останавливают добавлением 0,56 мл ледяного раствора для осаждения (300 мМ NaOAc pH 5,5 плюс GlycoBlue (Ambion)), а РНК очищают с помощью стандартного осаждения изопропанолом. Фрагментированную мРНК затем дефосфорилируют в реакции 50 мкл с 25 ед. PNK T4 (NEB) в 1× буфере PNK (без АТФ) плюс 0,5 ед. SUPERase•In и осаждают с GlycoBlue посредством стандартных способов осаждения изопропанолом.

Очистка растущей цепи РНК.

В случае растущей цепи РНК, очищенный лизат добавляют в 0,5 мл аффинного геля M2 анти-FLAG (Sigma Aldrich), как описано ранее. Аффинный гель дважды промывают буфером для лизиса, добавляют 1 мМ EDTA перед инкубированием с очищенным лизатом при 4°C в течение 2,5 ч с нутацией. Иммунопреципитацию промывают 4×10 мл с лизирующим буфером с добавлением 300 мМ KCl, а связанную РНК полимеразу элюируют дважды буфером для лизиса с добавлением 1 мМ EDTA и 2 мг/мл пептида 3x-FLAG (Sigma Aldrich). Растущую цепь РНК очищают из элюата с использованием набора miRNeasy (Qiagen) и конвертируют в ДНК с использованием ранее установленного протокола библиотека генерации.

Получение библиотеки ДНК и секвенирования ДНК.

Библиотеку ДНК секвенируют на Illumina HiSeq 2000. Считываемые фрагменты обрабатывают с использованием пакета HTSeq Python и другого пользовательского программного обеспечения, написанного в Python. 3'-конец секвенированного транскрипта выравнивают относительно референсного генома с использованием Bowtie ("bowtie-bio" перед ".sourceforge.net") и профилей RNAP, созданных в MochiView ("johnsonlab.ucsf" перед ".edu/mochi.html").

Дизайн и конструкция плазмиды для CRISPRi в клетках человека.

Последовательность, кодирующую кодон-оптимизированный для экспрессии у млекопитающих Cas9 Streptococcus pyogenes (DNA 2.0), сливают с тремя С-концевыми последовательностями ядерной локализации SV40 (NLS) или с tagBFP, окруженным двумя NLS. С использованием стандартного клонирования независимого от лигирования, клонируют эти два слитых белка в MSCV-Puro (Clontech). Направляющие sgРНК экспрессируют с использованием вектора экспрессии на основе лентивирусного U6 из pSico, который коэкспрессирует mCherry от промотора CMV. Плазмиды экспрессии sgРНК клонируют с помощью вставки прогибридизованных праймеров в вектор экспрессии на основе лентивирусной U6, которую расщепляют с помощью BstXI и XhoI.

Клеточные культуры, трансфекция ДНК и измерения флуоресценции в случае CRISPRi в клетках человека.

Клетки HEK293 поддерживают в среде Игла, модифицированной по способу Дульбекко, (DMEM) в 10% FBS, 2 mM глутамин, 100 ед./мл стрептомицин и 100 мкг/мл пенициллин. HEK293 заражают ретровирусом MSCV, экспрессирующим GFP, с использованием стандартных протоколов и сортируют при помощи проточной цитометрии с использованием Aria2 FACS BD для стабильной экспрессии GFP. Экспрессирующие GFP клетки HEK293 временно трансфицируют с помощью реагента для трансфекции TransIT-LT1 (Miras) по рекомендуемому производителем протоколом в 24-луночных планшетах с использованием 0,5 мкг плазмиды экспрессии dCas9 и 0,5 мкг экспрессионной плазмиды РНК (с 0,25 мкг репортерной плазмиды GFP в случае фиг. 45B). Через 72 ч после трансфекции, клетки обрабатывают трипсином до суспензии отдельных клеток. Вектор U6 содержит конститутивный промотор CMV, запускающий ген mCherry. Экспрессию GFP анализируют с использованием механизма FACS LSR II BD с помощью селекции положительных популяций mCherry (> 10 раз ярче mCherry по сравнению с клетками отрицательного контроля).

Созданные РНК.

Структуры sgРНК, используемые на фигурах: только 20 нуклеотидная совпадающая область (ДНК-нацеленный сегмент) (если не указано иное):

mRFP-специфические sgРНК приведены на фиг. 40C (SEQ ID NO:741-746);

промотор-специфические sgРНК приведены на фиг. 40D (SEQ ID NO:747-751);

последовательность промотора-мишени на фиг. 40D (SEQ ID NO:752);

mRFP-специфические sgРНК приведены на фиг. 43B (SEQ ID NO:753-760);

sfGFP-специфическая sgРНК (gfp) приведена на фиг. 42B (SEQ ID NO:761);

sfGFP-специфические sgРНК приведены на фиг. 43B (SEQ ID NO:762-769);

двойные-sgРНК-специфические эксперименты на фиг. 43F и 51 (SEQ ID NO:770-778);

Lac-оперон-специфические sgРНК приведены на фиг. 44B (SEQ ID NO:779-787); и

EGFP-специфические sgРНК приведены на фиг. 45 (SEQ ID NO:788-794).

Таблица 3

Последовательности, используемые на фигурах примера 4 (перечислены выше)

Последовательность	SEQ ID NO:	Последовательность	SEQ ID NO:	Последовательность	SEQ ID NO:
T1	741	R2	771	ср	783
T2	742	R3	772	суа	784
T3	743	R4	773	сайт А	785
NT1	744	R5	774	сайт О	786
NT2	745	R6	775	сайт Р	787
NT3	746	R7	776	eT1	788
P1	747	R8	777	eT2	789
P2	748	R9	778	eNT1	790
P3	749	lacZ	779	eNT2	791
P4	750	lacI	780	eNT3	792
P5	751	lacY	781	eNT4	793
R1	770	lacA	782	eNT5	794

Результаты.

Система CRISPR (кластерные короткие палиндромные повторы, разделенные регулярными промежутками) обеспечивает новую потенциальную платформу для направленной регуляции генов. Около 40% бактерий и 90% архей обладают CRISPR/CRISPR-ассоциированными (Cas) системами для придания устойчивости к элементам чужеродной ДНК. Системы CRISPR используют малые спаренные РНК для того, чтобы нацелиться на и расщепить чужеродные элементы ДНК последовательность-специфическим

способом. Существуют различные системы CRISPR у различных организмов, а одна из простейших является системой CRISPR типа II из *Streptococcus pyogenes*: только один ген, кодирующий белок Cas9, и две РНК, зрелая РНК CRISPR (crРНК) и частично комплементарная трансдействующая РНК (tracrРНК), являются необходимыми и достаточными для РНК-управляемого подавления чужеродных ДНК (фиг. 46). Созревание crРНК требует tracrРНК и РНКазы III. Тем не менее, это требование возможно обойти с помощью создания небольшой "направляющей" РНК (sgРНК), содержащей сконструированную шпильку, которая имитирует комплекс tracrРНК-crРНК. Образование пары оснований между sgРНК и ДНК-мишенью вызывает двухцепочечные разрывы (DSB), в соответствии с эндонуклеазной активностью Cas9. Специфичность связывания определяется как образованием пар sgРНК-ДНК, так и коротким мотивом ДНК (мотив, соседствующий с протоспейсером или PAM, последовательность: NGG) прилегающий к комплементарной области ДНК. Таким образом, система CRISPR требует только минимальный набор из двух молекул - белка Cas9 и sgРНК, и, следовательно, имеет потенциал, чтобы использоваться в качестве независимой от хозяина ген-направленной платформы. Было показано, что Cas9/CRISPR может быть использован в случае сайт-селективного РНК-направленного редактирования генома (фиг. 39А).

Фиг. 46 иллюстрирует механизм системы CRISPR типа II из *S. pyogenes*. Система состоит из набора CRISPR-ассоциированных (Cas) белков и локуса CRISPR, который содержит блок последовательностей повтора-спейсера. Все повторы одинаковы, а все спейсеры являются разными и комплементарными последовательностям ДНК-мишени. Когда клетку заражают чужеродными элементами ДНК, локус CRISPR будет транскрибироваться в длинный транскрипт-предшественник, который будет расщепляться на более мелкие фрагменты. Расщепление опосредуется антисмысловой трансдействующей РНК (tracrРНК) и РНКазой III хозяина. После расщепления, один белок, Cas9, распознается и связывается с расщепленной формой crРНК. Cas9 направляет crРНК к ДНК и сканирует молекулу ДНК. Комплекс стабилизируется посредством образования пар между crРНК и ДНК-мишенью. В этом случае, Cas9 вызывает двухцепочечные разрывы ДНК благодаря своей нуклеазной активности. Это, как правило, удаляет соответствующие молекулы ДНК, а клетки приобретают иммунитет к отдельным популяциям ДНК.

Фиг. 39 иллюстрирует устройство системы интерференции CRISPR (CRISPRi). (А) Минимальная система интерференции состоит из одного белка и созданной химеры sgРНК. Химера sgРНК состоит из трех доменов (область в рамке): 20-нуклеотидная (нт) комплементарная область для специфического связывания ДНК, 42-нт для связывания Cas9 (ручка Cas9) и 40-нт терминатор транскрипции, полученный от *S. pyogenes*. Белок Cas9 дикого типа имеет нуклеазную активность. Белок dCas9 не имеет нуклеазную активность. (В) Белок Cas9 дикого типа связывается с sgРНК и образует комплекс белок-РНК. Комплекс связывается со специфическими мишенями ДНК благодаря образованию уотсон-криковских пар между sgРНК и мишенью ДНК. В случае Cas9 дикого типа, ДНК будет расщепляться ввиду нуклеазной активности белка Cas9. В случае если нуклеаза Cas9 дефектна, то комплекс нарушает соответствующую транскрипцию.

Минимальная система CRISPRi, состоящая из одного белка и РНК, может эффективно подавлять инициацию транскрипции и элонгацию.

Для реализации такой платформы CRISPRi в *E.coli*, ген Cas9 *S. pyogenes* дикого типа и sgРНК экспрессируют из бактериальных векторов, для того чтобы определить, может ли система нарушать экспрессию гена в локусе-мишене (фиг. 40А). Система CRISPR *S. pyogenes* является ортогональной к нативной системе *E.coli*. Белок Cas9 экспрессируют с ангидротетрациклин(aTc)-индуцируемого промотора на плазмиде, содержащей точку начала репликации p15A и sgРНК, экспрессируют с минимального конститутивного промотора на плазмиде, содержащей точку начала репликации ColE1. В качестве альтернативной стратегии, использовали каталитически мертвый мутант Cas9 (dCas9), который является дефектным в расщеплении ДНК и было показано, что эта форма Cas9 всё еще действует в качестве простого РНК-направляемого ДНК-связывающего комплекса.

Фиг. 40 иллюстрирует то, как CRISPRi эффективно подавляет элонгацию и инициацию транскрипции. (А) Система CRISPRi состоит из индуцируемого белка Cas9 и созданной химеры sgРНК. dCas9 содержит мутации в нуклеазных доменах RuvC1 и HNH. Химера sgРНК содержит три функциональные домена, как проиллюстрировано на фиг. 1. (В) Последовательность созданного sgРНК (NT1) и ДНК-мишени. NT1 нацелен на нематричную цепь ДНК mRFP-кодирующей области. Показана только область, окружающая мотив спаривания оснований (20-нт). Нуклеотиды, образующие пары оснований пронумерованы и dCas9-связывающая шпилька надчеркнута. Последовательность PAM подчеркнута. (С) CRISPRi блокирует элонгацию транскрипции цепь-специфическим образом. Синтетическую репортерную систему, основанную на флуоресценции, содержащую mRFP-кодирующий ген, встраивали в геном *E.coli* 1655 (локус nsfA). Шесть sgРНК, которые связываются либо с обеими цепями матричной ДНК, либо с нематричной цепью ДНК, коэкспрессируют с белком dCas9, с их влияние на mRFP-мишень, измеряли с помощью *in vivo* флуоресцентного анализа. Только sgРНК, которые связываются с нематричной цепью ДНК, демонстрируют сайленсинг (10~300 раз). В контрольном образце показана флуоресценция клеток с белком dCas9, но без sgРНК. (D) CRISPRi блокирует инициацию транскрипции. Пять sgРНК создают таким образом, чтобы они связывались с различными областями по всему промотору *E.coli* (J23119). Сайт начала транскрипции обозначают как +1. Пунктирным овалом отмечают исходный комплекс РНК

полимеразы, который охватывает область в 75 п.о. от -55 до +20. Только sgPНК, нацеленные на области внутри исходного комплекса PНК полимеразы, демонстрирует репрессию (P1-P4). В отличие от блокирования элонгации транскрипции, сайленсинг не зависит от того, на какую цепь ДНК воздействуют. (E) Регулирование CRISPRi является обратимым. И dCas9, и sgPНК (NT1) находятся под контролем аТс-индуцируемого промотора. Клеточные культуры поддерживают в течение экспоненциальной фазы. В момент  $T = 0,1$  мкМ аТс, добавляют в клетки с  $OD = 0,001$ . Репрессии mRFP-мишени начинается в течение 10 мин. Сигнал флуоресценции ослабевает таким образом, что это соответствует росту клеток, что свидетельствует о том, что это ослабление связано с делением клетки. Через 240 мин, флуоресценция достигает полностью подавленного уровня. При  $T = 370$  мин, аТс вымывают из питательной среды, а клетки разбавляют обратно до  $OD = 0,001$ . Флуоресценция, начинает увеличиваться через 50 мин, и требуется около 300 мин для того, чтобы она доросла до того же уровня, что и в положительном контроле. Положительный контроль: всегда без индуктора; отрицательный контроль: всегда с 1 мкМ индуктора аТс. Результаты флуоресценции в 2С, 2D, 2E представляют собой средние значения и SEM по меньшей мере трех биологических повторов. См. также фиг. 47 и 48.

Каждая из молекул sgPНК, коэкспрессированных с Cas9, состоит из трех доменов: 20-нуклеотидной (нт) мишень-специфической комплементарной области, 42-нт шпильки, связывающей Cas9 (ручка Cas9) и 40-нт терминатора транскрипции, полученного от *S. ruogenes* (фиг. 40B). Репортерную систему на основе красного флуоресцентного белка (mRFP) встраивают в геном MG1655 *E.coli*.

Коэкспрессия белка Cas9 дикого типа и sgPНК (NT1), нацеленной на последовательность, кодирующую mRFP, резко сокращает эффективность трансформации, вероятно, ввиду Cas9-индуцированных двухцепочечных разрывов в геноме (фиг. 47A). Секвенирование нескольких выживших колоний продемонстрировало, что у всех них имеются перестройки последовательности вокруг сайта-мишени mRFP в геноме, предполагая, что имеется значительная селекция против экспрессии Cas9 дикого типа и sgPНК, нацеленной на последовательность хозяина. Мутантный ген dCas9 (нерасщепляющий), который содержит две мутации подавления нуклеазных доменов RuvC1 и HNH (D10A и H841A), устраняют эту летальность, что подтверждается эффективностью трансформации и скоростью роста *E.coli* (фиг. 47A и B).

Фиг. 47 связана с фиг.40 и иллюстрирует кривые роста клеточных культур *E.coli*, котрансформированных dCas9 и sgPНК. (A) Эффективность трансформации в случае трансформации клеток *E.coli* с двумя плазмидами. Одна плазида содержит sgPНК, которая нацелена на геномную копию mRFP, а другая плазида содержит Cas9 дикого типа или dCas9. Котрансформирование Cas9 дикого типа и sgPНК является высокотоксичным, что может быть ослаблено с помощью dCas9. (B) sgPНК (NT1) создают для нацеливания на кодирующую последовательность mRFP. Коэкспрессия dCas9 и sgPНК практически не имеет никакого влияния на скорость роста клеток, предполагая, что взаимодействие dCas9-sgPНК с ДНК является достаточно сильным для того, чтобы блокировать PНК-полимеразу, но не ДНК-полимеразу или репликацию клеток. Результаты представляют средние значения и SEM по меньшей мере трех независимых экспериментов.

Чтобы проверить, может ли комплекс dCas9:sgPНК высокоэффективно репрессировать экспрессию генов, sgPНК, комплементарные различным областям кодирующей mRFP последовательности, связываться либо с матричной цепью ДНК, либо нематричной цепью ДНК. Результаты показали, что sgPНК, нацеленные на нематричную цепь ДНК, демонстрируют эффективное подавление гена (от 10 до 300-кратной репрессии), в то время как таковые, ориентированные на матричную цепь, демонстрируют незначительный эффект (фиг. 40C). Система проявляет подобные эффекты репрессии генов, которые находятся в пределах генома *E.coli* или на многокопийной плазмиде (фиг. 48). Кроме того, нацеливание на промоторную область также приводит к эффективному сплейсингу гена (фиг. 40D). Нацеливание sgPНК на бок -35, значительно нокаутирует экспрессию гена (P1, ~100-кратная репрессия), в то время как нацеливание на другие прилегающие области демонстрирует ослабленный эффект (P2-P4). Нацеливание на последовательности около 100 б.п. против хода транскрипции не демонстрируют никаких эффектов (P5). В отличие от нацеливания на кодирующую последовательность, при нацеливании на промотор, эффективность сайленсинга является независимой от цепи ДНК; нацеливание на матричную или нематричную цепи является одинаково эффективным (P2 и P3).

Фиг. 48 связана с фиг. 40C и иллюстрирует то, что CRISPRi может подавлять экспрессию репортерного гена на многокопийной плазмиде. Ген mRFP клонируют на плазмиде p15A. Наличие dCas9 и mRFP-специфической sgPНК (NT1) значительно подавляет mRFP (в ~300 раз). Эффект репрессии является аналогичным тому, который наблюдается когда mRFP находится в геноме (фиг. 40C). Сайленсинг является эффективным только при условии, что sgPНК действует на нематричную цепь ДНК, а не на матричную цепь ДНК (T1). Кроме того, сайленсинг является высокоспецифичным, так как 3 GFP-специфических sgPНК (gfp) не демонстрируют никакого влияния на экспрессию mRFP. Результаты флуоресценции представляют средние значения и SEM по меньшей мере трех биологических повторов.

Нокдаун гена CRISPRi является индуцируемым и обратимым.

В отличие от способов нокаута гена, одно из преимуществ использования нокдауна экспрессии гена с помощью CRISPRi заключается в том, что это вмешательство является обратимым. Чтобы проверить является ли регулирование CRISPRi индуцируемым и впоследствии обратимым, как dCas9, так и mRFP-

специфическую sgPНК (NT1) помещают под контроль aTc-индуцируемого промотора и измеряют CRISPRi-опосредованную регуляцию mRFP в ответ на индукторы в зависимости от времени (фиг. 40E). В нулевой момент времени, в клеточную культуру, которая выросла до ранней экспоненциальной фазы без индукторов, добавляют 1 мкМ aTc. Данные показали, что система может быстро реагировать на присутствие индукторов - флуоресцентный сигнал репортерного белка начал снижаться в течение 10 мин после добавления молекулы индуктора. Поскольку белок mRFP является стабильным, скорость уменьшения сигнала флуоресценции ограничена разбавлением белка ввиду роста клеток, что видно по тому, что в аналогичное время происходит удвоение клеток и потеря флуоресцентного сигнала наполовину (оба ~36 мин). В 240 мин все клетки равномерно подавляют до того же уровня, что и отрицательный контроль. В 420 мин, индуктор вымывают из среды роста и клетки разводят обратно до более низкой OD. После задержки в 50 мин, флуоресценция mRFP начинает расти. Потребовалось в общем 300 мин для флуоресценции в одиночной клетке, для того, чтобы увеличиться до такого же уровня, что и в положительном контроле. Задержку в 50 мин, скорее всего, определяется отклонением скорости оборота dCas9/sgPНК, вызванном разбавлением при росте клеток и делении. Таким образом, эти результаты демонстрируют, что эффекты сайленсинга у dCas9-sgPНК является индуцируемым и обратимым.

Секвенирование нативного транскрипта удлиннения (NET-Seq) подтверждает, что CRISPRi функционирует, блокируя транскрипцию.

dCas9, казалось, функционирует в качестве PНК-управляемого ДНК-связывающего комплекса, который может блокировать связывание PНК-полимеразы во время элонгации транскрипции. В силу того, что нематричная цепь ДНК имеет ту же идентичность последовательности, что и транскрибируемая мPНК, и что только sgPНК, которые связываются с нематричной цепью ДНК, проявляют эффект сайленсинга, остается возможность того, что комплекс dCas9: sgPНК взаимодействует с мPНК и изменяет его трансляцию или стабильность. Чтобы отличить эти возможности, недавно описанный подход секвенирования нативного удлиняющего транскрипта (NET-seq) применяют к *E. coli*, что может быть использована для глобального профилирования положений удлиненных PНК-полимеразой и контролирования влияния комплекса dCas9:sgPНК на транскрипцию. В этом способе NET-seq, систему CRISPRi трансформируют в штамм, производный от *E. coli* MG1655, который содержит FLAG-меченную PНК полимеразу. CRISPRi содержит sgPНК (NT1), которая связывается с областью, кодирующей mRFP. *In vitro* иммуноочистка меченой PНК полимеразы с последующим секвенированием транскриптов растущей цепи, ассоциированных с удлинением PНК полимеразой, позволяет различать сайты временной остановки PНК полимеразы.

Эти эксперименты демонстрируют, что sgPНК индуцирует сильную временную остановку транскрипции выше локуса-мишени sgPНК (фиг. 41A). Расстояние между сайтом временной остановки и сайтом-мишенью составляет 19 п.о., что находится в отличном соответствии с ранее представленным расстоянием в ~18 п.о. между включением нуклеотида PНК полимеразой и ее передней частью. Этот вывод согласуется с механизмом CRISPRi, в котором блокирование транскрипции связано с физическим столкновением между удлиняющей PНК полимеразой и комплексом dCas9:sgPНК (фиг. 41B). Связывание комплекса dCas9:sgPНК с матричной цепью имеет небольшой репрессивный эффект, на основании чего предполагают, что PНК полимеразы способна считать фрагмент через комплекс в данной конкретной ориентации. В этом случае, sgPНК, которая сталкивается с PНК полимеразой, может быть раскручена за счет геликазной активности PНК полимеразы. Эти эксперименты продемонстрировали, что CRISPRi использует PНК для непосредственной блокировки транскрипции. Этот механизм отличается от такового в случае PНКи, где для нокдауна экспрессии генов требуется разрушение уже транскрибированных матричных PНК, перед их трансляцией.

Фиг. 41 иллюстрирует то, что CRISPRi функционирует с помощью блокирования элонгации транскрипции. (A) FLAG-меченные молекулы PНК полимеразы иммунопреципитируют и секвенируют ассоциированные с ней транскрипты растущей цепи мPНК. Верхняя часть иллюстрирует результаты секвенирования транскрипта растущей цепи mRFP в клетках без sgPНК, а нижняя часть иллюстрирует результаты в клетках с sgPНК. Ей присутствии sgPНК, значительную временную остановку транскрипции наблюдают на 19 п.о. выше сайта-мишени, после чего количество считываемых фрагментов, подвергающихся секвенированию, резко падает. (B) Предполагается, что механизм CRISPRi основан на физическом столкновении между PНК полимеразой и dCas9-sgPНК. Расстояние от центра PНК полимеразы до ее переднего края составляет ~19 и.о., что хорошо соответствует измеренному расстоянию между сайтом временной остановки транскрипции и 3'-концом области образования пар с sgPНК. Временно остановленная PНК полимеразы прекращает элонгацию транскрипции когда наталкивается на помеху в виде dCas9-sgPНК.

Сайленсинг гена, направленный CRISPRi sgPНК является высокоспецифическим.

Чтобы оценить специфичность CRISPRi в общегеномном масштабе, осуществляют полное секвенирование транскриптома способом дробовика (PНК-сек) dCas9-трансформированных клеток с и без ко-экспрессии sgPНК (фиг. 42A). В присутствии sgPНК, нацеленной на mRFP (NT1), транскрипт mRFP является единственным геном, количество которого уменьшилось. Никакие другие гены не демонстрируют значительное изменение экспрессии при добавлении sgPНК, в пределах ошибок секвенирования. Также

осуществляют секвенирование РНК в клетках с различными sgРНК, которые нацелены на различные гены. Ни один из этих экспериментов не продемонстрировал существенных изменений генов, кроме гена-мишени (фиг. 49). Таким образом, управляемое sgРНК нацеливание на ген и его регулирование является весьма специфичным и не имеет значительных внецелевых эффектов.

Фиг. 42 иллюстрирует специфичность нацеливания системы CRISPRi. (А) Секвенирование мРНК в масштабе генома (РНК-seq) подтверждает, что нацеливание CRISPRi не имеет внецелевых эффектов. Используют NT1 sgРНК, которая связывается с кодирующей области mRFP. Выделены гены dCas9, mRFP и sfGFP. (В) Несколько sgРНК могут независимо подавлять два репортера флуоресцентных белков в той же ячейке. Каждая sgРНК специфично подавляет свой ген, но не отличный ген. Когда присутствуют обе sgРНК, подавляются оба гена. Величины ошибки представляют SEM по меньшей мере трех биологических повторов. (С) Микроскопические изображения, полученные при использовании двух sgРНК с целью управления двумя флуоресцентными белками. Верхняя часть иллюстрирует светлопольные изображения клеток *E.coli*, средняя часть иллюстрирует канал RFP, а нижняя иллюстрирует панель GFP. Коэкспрессия одной sgРНК и dCas9 подавляет только свой флуоресцентный белок, но не отличный. Эффект нокадауна является сильным, так как флуоресценция в клетках, содержащих определенный подавленный флуоресцентный белок, практически не наблюдается. Масштабная линейка, 10 мкм. Контролем являются клетки без каких-либо репортерных флуоресцентных белков. Результаты флуоресценции представляют средние значения и SEM по меньшей мере трех биологических повторов. См. также фиг. 49.

Фиг. 49 связана с фиг. 42А и иллюстрирует данные РНК-сек клеток с sgРНК, которые нацелены на различные гены. (А) (+/-) sgРНК, которая нацелена на промотор эндогенного гена *lacI* в *E.coli*. Аналогичную *lacI*-нацеливающую sgРНК используют таким образом, как на фиг. 44А. (В) (+/-) 1 мМ IPTG в случае клеток без авто-ингибированной sgРНК (sgРНК, подавляет свой собственный промотор). (С) (+/-) sgРНК, которая нацелена на эндогенный ген *lacZ* в *E.coli*. Аналогичную sgРНК, нацеленную на *lacZ* используют таким образом, как на фиг. 44А. 1 мМ IPTG также добавляют к клеткам с sgРНК, нацеленной на *lacZ*.

CRISPRi могут быть использованы для одновременного регулирования множества генов.

Система CRISPRi позволяет независимо, без помех, управлять несколькими генами. Создают двухцветную систему флуоресцентных-репортеров, основанную на mRFP и sfGFP. Создают две sgРНК с различными комплементарными областями для каждого гена. Экспрессия каждой sgРНК подавляет только свой ген и не оказывает никакого влияния на другой. Коэкспрессия двух sgРНК нокаутирует оба гена (фиг. 42В и С). Эти результаты свидетельствуют о том, что sgРНК-управляемое нацеливание является специфичным, где специфичность диктуется только идентичностью его последовательности с мишенью, и на нее не влияет присутствие других sgРНК. Такое поведение делает возможным мультиплексное управление несколькими генами одновременно с помощью CRISPRi.

Факторы, которые определяют эффективность сайленсинга у CRISPRi.

Для того чтобы найти детерминанты эффективности нацеливания CRISPRi, исследуют влияние длины, комплементарности последовательности и положение на эффективность подавления (фиг. 43А). Как предложено на фиг. 40С, локализация последовательности-мишени sgРНК в гене имеет важное значение для эффективности. sgРНК дополнительно конструировали так, чтобы они покрывали всю длину кодирующих областей как mRFP, так и sfGFP (Supplemental Data в случае последовательностей sgРНК). Во всех случаях, репрессия обратно коррелирует с расстоянием от мишени до сайта начала транскрипции (фиг. 43В). Сильную линейную корреляцию наблюдают в случае mRFP. Подобную, но немного более слабую корреляцию, наблюдают при условии, что в качестве мишени используют sfGFP, возможно, указывая на различные кинетики РНК-полимеразы в различных точках в элонгации этого гена.

SgРНК содержит 20-п.о. область, комплементарную мишени. Чтобы определить важность этой области спаривания оснований, изменяют длину sgРНК NT1 (фиг. 43С). В то время как удлинение области на 5'-конце не влияет на сайленсинг, усечение области серьезно уменьшает репрессию. Минимальная длина области спаривания оснований, необходимая для сайленсинга гена, составляет 12 п.о., где дополнительное усечение ведет к полной потере функции. Одиночные мутации вводят в область спаривания оснований NT1 sgРНК и исследуют суммарный эффект на сайленсинг. Основываясь на результатах, можно выделить три субобласти, каждая с отдельным вкладом в общее связывание и сайленсинг (фиг. 43D). Любая одиночная мутация первых 7 нуклеотидов резко уменьшает репрессию, предполагая, что эта последовательность представляет собой "затравочную область" для связывания, как было отмечено выше, как в случае систем CRISPR типа I, так и в случае типа II. Прилегающие нуклеотиды также мутировали попарно (фиг. 43Е и 50). В большинстве случаев, относительная активность репрессии, вызванная двойной мутацией, усилилась по сравнению с эффектом, вызванным одиночной мутацией, предполагая независимые отношения между ошибочными спариваниями. Кроме того, в соответствии с предыдущими результатами, касающимися важности последовательности PAM, неправильный PAM полностью отменяет сайленсинг даже при наличии области идеального спаривания в 20-п.о. (фиг. 43Е). Таким образом, специфичность системы CRISPRi определяют совместно с PAM (2-п.о.) и по меньшей мере 12-п.о. участка sgРНК-ДНК, площадь которой достаточно велика, чтобы покрывать большинство бактериальных геномов в уникальных сайтах-мишенях.

Исследуют две sgРНК, обе из которых нацелены на один и тот же ген (фиг. 43F и 51). В зависимости от взаимного расположения множества sgРНК, наблюдают различные комбинаторные эффекты. Объединение двух sgРНК, каждая из которых характеризуется примерно 300-кратной репрессией, позволяет в целом усилить сайленсинг вплоть до тысячи раз. Объединение двух более слабых sgРНК (~5-кратный) демонстрирует мультипликативные эффекты при совместном использовании. Подавляющие комбинаторные эффекты наблюдают при условии использования двух sgРНК, чьи мишени перекрываются. Это существует, вероятно, ввиду конкуренции обеих sgРНК при связывании с той же областью.

Фиг. 43 иллюстрирует характеристику факторов, которые влияют на эффективность сайленсинга. (А) Эффекты подавления измеряют для sgРНК с различными локусами-мишенями в одном и том же гене (расстояние от иницирующего трансляцию кодона) и sgРНК с различными длинами области спаривания оснований, нацеленных на один локус-мишень (на основе NT1). (В) Эффективность сайленсинга обратно коррелирует с расстоянием мишени от иницирующего трансляцию кодона (оранжевый -mRFP и зеленый - sfGFP). Относительную активность репрессии рассчитывают путем нормализации репрессии каждой sgРНК к таковой для sgРНК, с наиболее высокой кратностью изменения репрессии. Величины ошибки представляют SEM трех биологических повторов. (С) Длина области образования уотсон-криковских пар между sgРНК и ДНК-мишенью влияет на эффективность репрессии. Все удлинения области спаривания оснований усиливает сайленсинг, а усеечения резко сокращают реессию. Минимальная длина области образования пары для детектируемой репрессии составляет 12 п.о. Величины ошибки представляют SEM трех биологических повторов. (D) Отдельные ошибочные спаривания вводят в каждый нуклеотид sgРНК (NT1, фиг. 40B), и измеряют как эти отдельные ошибочные спаривания, влияют на эффективность репрессии. Были выделены три суб-области с различным влиянием на общее подавление. Они продемонстрировали скачкообразную функцию. Первая область из 7 нуклеотидов имеет решающее значение для сайленсинга, и, вероятно, представляет собой область "затравки" для "поиска" sgРНК той области ДНК-мишени, с которой она связывается. Последовательность PAM (NGG) является необходимой для подавления. Величины ошибки представляют SEM трех биологических повторов. (E) Эффекты сайленсинга для sgРНК с соседними двойными ошибочными спариваниями. Показана относительная активность репрессии sgРНК с одним ошибочным спариванием, где положение ошибочного спаривания отмечено внизу. Показана экспериментально измеренная активность sgРНК с двумя ошибочными спариваниями. Активность, рассчитанная с помощью умножения эффектов двух sgРНК с одним ошибочным образованием пары, показана белым и обозначена "Рассч". В большинстве случаев, активность подавления sgРНК с двумя ошибочными спариваниями оказалась равна простому умножению активностей sgРНК с одним ошибочным образованием пары (за исключением фиг. 50B), что указывает на независимые отношения между отдельными ошибочными спариваниями. Величины ошибки представляют SEM трех биологических повторов. (F) Комбинаторные эффекты сайленсинга от использования двойных sgРНК для нацеливания на один ген mRFP. С использованием двух sgРНК, которые нацелены на один и тот же ген, общий эффект нокдауна может быть увеличен до почти 1000-кратного. При условии, что две sgРНК связываются с перекрывающимися последовательностями того же гена, репрессия увеличивается. При условии, что две sgРНК нацелены на перекрывающиеся области, репрессия уменьшается. Величины ошибки представляют SEM трех биологических повторов.

Фиг. 50 связана с фиг. 43E и иллюстрирует эффекты сайленсинга у sgРНК с соседними двойными ошибочными спариваниями. Показана относительная активность репрессии sgРНК с одним ошибочным спариванием, где положение ошибочного спаривания отмечено внизу. Также показана экспериментально измеренная активность sgРНК с двумя ошибочными спариваниями. Активность, рассчитанная с помощью умножения эффектов двух sgРНК с одним ошибочным спариванием, показана белым и помечена "Рассч". Результаты флуоресценции представляют средние значения и SEM по меньшей мере трех биологических повторов.

Фиг. 51 связана с фиг. 43E и иллюстрирует комбинаторные эффекты сайленсинга, с использованием двух sgРНК для того, чтобы регулировать один ген. Во всех случаях, пересекющиеся sgРНК демонстрируют усиливающиеся эффекты подавления, а перекрытие sgРНК демонстрирует сниженный эффект. Комбинаторный эффект не зависит от того, нацелена ли sgРНК на матричные, или на нематричные цепи ДНК. Результаты флуоресценции представляют средние значения и SEM трех биологических повторов.

Исследование эндогенной регуляторной сети с использованием нокдауна генов с помощью CRISPRi. Систему CRISPRi затем используют в качестве платформы нокдауна гена для исследования эндогенных генных сетей. Известные ранее способы исследования микробных генных сетей в основном основаны на трудоемкой и дорогостоящей геномной инженерии и процедурах нокаута. Напротив, нокдаун гена с помощью CRISPRi требуют только сконструировать и синтезировать небольшую sgРНК, содержащую 20-п.о., комплементарную к области желаемого гена. Чтобы продемонстрировать это, CRISPRi используют для создания нокдаунных штаммов *E.coli* с помощью разработки sgРНК, предназначенных для систематического нарушения генов, которые являются частью хорошо изученного пути метаболизма лактозы *E.coli* (фиг. 44A). Анализы с β-галактозидазой осуществляли для измерения экспрессии LacZ в нокдаунных штаммах нокдауна, с и без изопропил-β-D-1-тио-галактопиранозида (IPTG), химического

вещества, которое подавляет *lac*-репрессор (*LacI*). В клетках дикого типа, добавление IPTG индуцирует экспрессию *LacZ*. Результаты демонстрируют, что *lacZ*-специфическая sgPНК может значительно подавлять экспрессию *LacZ* (фиг. 44В). С другой стороны, нацеливание sgPНК на ген *lacI* приводит к активации экспрессии *LacZ*, даже в отсутствие IPTG, как можно было бы ожидать в случае сайленсинга прямого репрессора экспрессии *LacZ*.

Известно, что cAMP-CRP представляет собой важный активатор экспрессии *LacZ* посредством связывания с цис-регулирующим сайтом выше от промотора (сайт А).

Соответственно, sgPНК, которая была нацелена на ген *ср* или на сайт А в промоторе *LacZ*, приводит к репрессии, демонстрируя возможность связи регулятора с его цис-регуляторной последовательностью, с использованием CRISPRi. Нацеливание на ген аденилатцилазы (*суа*), который является необходимым для получения cAMP, которая делает CRP более эффективным в промоторе *LacZ*, приводит только лишь к частичной репрессии. Добавление 1 мМ cAMP в среду роста, компенсировало эффекты для нокдауна *суа*, но не для нокдауна *ср*, предполагая, что *суа* представляет собой косвенный регулятор *LacZ*. Кроме того, нацеливание sgPНК на цис-регуляторный сайт *Lad* (сайт О) приводит к ингибированию, по-видимому, потому что комплекс Cas9, связывающийся на этом сайте, стерически блокирует PНК-полимеразу, имитируя поведение репрессора транскрипции *Lad*. Нацеливание на известный сайт, связывания PНК полимеразы (сайт Р), также блокирует экспрессию. В целом эти исследования демонстрируют то, что способ нокдауна гена, основанный на CRISPRi, обеспечивает быстрый и эффективный подход для анализа регуляторных функций (активации или репрессии) генов и цис элементов в комплексных регуляторных сетях (фиг. 44С).

Фиг. 44 иллюстрирует функциональное профилирование сложной регуляторной сети с использованием нокдауна гена CRISPRi. (А) sgPНК конструируют и используют для того, чтобы нокаутировать гены (*суа*, *ср*, *lacI*, *lacZ*, *lacY*, *lacA*) в пути метаболизма *lac*, или блокировать сайты оператора транскрипции (A/P/O). *LacI* представляет собой репрессор оперона *lacZYA* и связывается с сайтом оператора транскрипции (сайт О). Ген *LacZ* кодирует фермент, который катализирует преобразование лактозы в глюкозу. Несколько трансдействующих генов хозяина, таких как *суа* и *ср*, участвуют в активации системы *lacZYA*. Комплекс cAMP-CRP связывается с сайтом оператора транскрипции (сайт А) и рекрутирует PНК-полимеразу, связывающуюся с сайтом Р, что инициирует транскрипцию *lacZYA*. IPTG, химическое вещество, которое подавляет функцию *Lad*, индуцирует экспрессию *LacZ*. (В) Анализ β-галактозидазы нокдаунных штаммов без (белый) и с (серый) IPTG. Контроль демонстрирует, что клетки дикого типа без вмешательства CRISPRi могут быть индуцированы с помощью добавления IPTG. SgPНК, которая нацелена на *LacZ*, значительно подавляет экспрессию *LacZ*, даже в присутствии IPTG. При условии, что мишенью является *Lad*, экспрессия *LacZ* является высокой, даже без IPTG. Нацеливание на гены *суа* и *ср* приводит к снижению уровня экспрессии *LacZ* в присутствии IPTG. Присутствие 1 мМ cAMP спасает нокдаун *суа*, но не нокдаун *ср*. Блокирование сайтов оператора транскрипции приводит к репрессии *LacZ*, предполагая, что они являются важными цис-действующими регуляторными сайтами в случае *LacZ*. В процессе вмешательства, показано снижение (стрелки вниз) и увеличение (до стрелки) экспрессии *LacZ*. Величины ошибки представляют SEM трех биологических повторов. (С) Эксперименты с нокдауном позволяют профилировать роли регуляторов регуляторной цепи *lac*. Данные представлены на 2-D графике, с осью x, отражающей активность *LacZ* без IPTG, и осью y, отражающей ее активность с IPTG. Распространение овалов вдоль каждой оси демонстрирует стандартные отклонения. Результаты анализа β-галактозидазы представляют средние значения и SEM трех биологических повторов. Для данных PНК-сек для нацеливания на *Lad* и *LacZ*, см. также фиг. 49.

CRISPRi может нокаутировать экспрессию гена-мишени в клетках человека. Чтобы изучить общность подхода CRISPRi, использующего комплекс dCas9-sgPНК для подавления транскрипции, систему исследуют в клетках HEK293 млекопитающего. Белок dCas9 кодон-оптимизируют, сливают с тремя копиями последовательности ядерной локализации (NLS), и экспрессируют из ретровирусного вектора Murine Stem Cell Virus (MSCV). Такое же устройство sgPНК, как приведено на фиг. 40В, используют для того, чтобы экспрессировать с промотора U6 PНК-полимеразы III. Репортерную клеточную линию HEK293, экспрессирующую EGFP под промотором SV40, создают с помощью вирусной инфекции. При использовании sgPНК (eNT2), которая нацелена на нематричную цепь ДНК области, кодирующей EGFP, наблюдают умеренный, но воспроизводимый нокдаун экспрессии генов (46% репрессии, фиг. 45А). Репрессии зависят как от белка dCas9, так и от sgPНК, подразумевая то, что репрессия была вызвана комплексом dCas9-sgPНК и PНК-управляемого нацеливанием. Та же самая sgPНК проявляет лучшую репрессию того же гена, при условии, что он временно экспрессирован из плазмиды (63% репрессии, фиг. 52). В соответствии с бактериальной системой, только sgPНК, нацеленные на нематричную цепь, проявляют репрессию. Такие факторы, как расстояние между началом транскрипции и локальным состоянием хроматина, могут быть критическими параметрами, определяющими эффективность репрессии (фиг. 52). Оптимизация экспрессии dCas9 и sgPНК, стабильности, ядерной локализации и взаимодействия позволит дополнительно повысить эффективность CRISPRi в клетках млекопитающих.

Фиг. 45 иллюстрирует то, что CRISPRi может подавлять экспрессию гена в клетках человека. (А)

Система CRISPRi в клетках HEK293. Кассету экспрессии SV40-EGFP встраивают в геном посредством ретровирусной инфекции. Белок dCas9 является кодон-оптимизированным и слитым с тремя копиями последовательности NLS. sgРНК экспрессируют из вектора U6 РНК-полимеразы III. Котрансфекция dCas9 и sgРНК (eNT2), которые нацелены на нематричную цепь EGFP, снижают флуоресценцию (~46%), в то время как экспрессия либо dCas9, либо sgРНК отдельно не дает никакого эффекта. (B) dCas9:sgРНК-опосредованная репрессия зависит от локусов-мишеней. Создают семь sgРНК, нацеленные на различные области последовательности, кодирующих EGFP, на матричной или нематричной цепи. Только eNT2 и eNT5 демонстрируют умеренную реессию. Результаты: флуоресценции в 7A и 7B представляют средние значения и ошибку двух биологических повторов.

Фиг. 52 связана с фиг. 45 и иллюстрирует то, что репрессия sgРНК зависит от локусов-мишеней и относительно расстояния от начала транскрипции. Ту же sgРНК используют для подавления того же гена EGFP с разными промоторами. Комплексы Cas9/sgРНК подавляют транскрипцию из временно трансфектированной плазмидной ДНК. Уровень репрессии транскрипции был немного лучше (63%), чем это наблюдается в случае геномных генов, а процент GFP-отрицательных клеток увеличился в присутствии sgРНК. Локус-мишень имеет различное расстояние от начала транскрипции. В то время как SV40-EGFP демонстрирует реессию, LTR-EGFP не имеет эффекта. Результаты флуоресценции представляют средние значения и ошибку двух биологических повторов.

CRISPRi эффективно и селективно подавляет транскрипцию генов-мишеней.

Система CRISPRi является относительно простой платформой для направленной регуляции гена. CRISPRi не зависит от присутствия сложных факторов хозяина, но вместо этого требует только белок dCas9 и "направляющие" РНК, и, таким образом, является гибкой и легко конструируемой системой. Система может эффективно подавлять гены в бактерии. Сайленсинг является очень эффективным, так как не были обнаружены внецелевые эффекты. Кроме того, эффективность нокдауна может быть настроена путем изменения локусов-мишеней и степени спаривания оснований между sgРНК и геном-мишенью. Это делает возможным создание ряда аллелей гипоморфов - особенность, которая является особенно полезной в случае изучения основных генов. Система действует, непосредственно блокируя транскрипцию таким образом, что может быть легко запрограммирована с помощью проектирования sgРНК. Механистически, это отличается от сайленсинга на основе РНКи, который требует разрушения уже транскрибированных мРНК.

Кроме того, комплексы dCas9:sgРНК могут также модулировать транскрипцию посредством нацеливания на основные цис-действующие мотивы в пределах любого промотора, стерически блокируя их связывание с соответствующими трансдействующими факторами транскрипции. Таким образом, в дополнение к его использованию в качестве инструмента для нокдауна гена, CRISPRi может быть использован для функционального картирования промоторов и других регуляторных модулей генома.

CRISPRi пригоден для анализа в масштабах генома и регулирования. Способ CRISPRi основан на использовании РНК, нацеленных на ДНК, и только нацеленный на ДНК сегмент должен быть создан для специфических генов-мишеней. С развитием технологий крупномасштабного синтеза олигонуклеотидов ДНК, получение больших наборов олигонуклеотидов, которые содержат уникальные 20-п.о. области для нацеливания на геном, стало быстрым и недорогим. Эти олигонуклеотидные библиотеки могут позволить направленно воздействовать на большое количество отдельных генов, чтобы определить функцию гена или направленно воздействовать на пары генов для картирования генетических взаимодействий. Кроме того, CRISPRi может быть использован для одновременного модулирования экспрессии больших наборов генов, тогда как малый размер sgРНК позволяет объединить несколько элементов в один и тот же вектор экспрессии.

CRISPRi обеспечивает новые инструменты для работы с микробными геномами. Поскольку платформа CRISPRi является компактной и автономной, она может быть адаптирована для различных организмов. CRISPRi является мощным инструментом для изучения не модельных организмов, для которых способы геной инженерии являются не очень хорошо развитыми, в том числе патогенов или промышленно полезных организмов. В отличие от большинства эукариот, большинство бактерий лишены механизма РНКи. Как следствие, регулирование эндогенных генов с использованием разработанных синтетических РНК в настоящее время ограничено. CRISPRi может обеспечить РНКи-подобный способ возмущающего влияния на гены в микробах.

CRISPRi в качестве платформы для создания сетей, регулирующих транскрипцию CRISPRi, может быть использован как гибкий каркас для создания сетей, регулирующих транскрипцию. Платформа CRISPRi, ввиду того, что она, по существу, является РНК-направляемым ДНК-связывающим комплексом, а также обеспечивает гибкий каркас для управления сложной регуляторной машинерией клетки в специфических сайтах генома. Помимо просто блокирования транскрипции генов-мишеней, является возможным соединить белок dCas9 с различными регуляторными доменами, чтобы модулировать различные биологические процессы и генерировать различные функциональные результаты (например, активацию транскрипции, модификацию хроматина).

В системе CRISPRi возможно внедрить несколько sgРНК в цикл транскрипции, где одна вышерасположенная sgРНК будет контролировать экспрессию других нижерасположенных. Тогда как молекулы

РНК в микроорганизмах имеют тенденцию быть недолговечными, авторы подозревают, что генетические программы, регулируемые sgРНК, могут продемонстрировать быструю кинетику, отличную от схем, содержащих медленные процессы, такие как экспрессия и деградация белка. Таким образом, система CRISPRi представляет собой общую платформу генетического программирования, подходящую для различных медико-биологических исследований и клинических использований, в том числе функционального профилирования в масштабах генома, микробной метаболической инженерии и перепрограммирования клеток.

Пример 5. Химерный сайт-специфический полипептид может быть использован для модулирования (активации или репрессии) транскрипции в клетках человека.

Было продемонстрировано, что в клетках человека слитый белок, содержащий каталитически неактивный Cas9, и домен-активатор или домен-репрессор могут увеличить или уменьшить транскрипцию из ДНК-мишени соответственно.

Фиг. 55. Сливают гуманизированный каталитически неактивный Cas9 с доменом VP64 активатора транскрипции. (А) Для проверки эффективности активации генов, с использованием этой системы, вставляют GAL4 UAS индуцируемый промотор, который управляет GFP, в геном HEK293 (клетки культуры тканей человека). (В) Промотор UAS GAL4 может быть индуцирован в присутствии белка GAL4, полученного из дрожжей. Слияние dCas9-VP64 может эффективно активировать UAS GAL4 в 20 раз в присутствии родственной "направляющей" РНК, которая связывается с областью UAS GAL4. (С) Микроскопические изображения в случае активации dCas9-VP64. (D) Данные проточной цитометрии в случае активации dCas9-VP64.

Фиг. 56. Сливают гуманизированный каталитически неактивный Cas9 с доменом репрессора транскрипции KRAB. (Верх) Создают 10 "направляющих" РНК, которые нацелены на хорошо изученный ранний промотор SV40, и одну "направляющую" РНК, которая направлена на область, кодирующую EGFP. (Низ) С использованием негибридного dCas9, наблюдают 2~3-кратную реессию в случае gРНК P9 и NT2. Эта эффективность была значительно повышена, когда использовался слитый dCas9-KRAB. Например, при использовании dCas9-KRAB, P9 и NT2 демонстрируют 20-кратную и 15-кратную реессию соответственно. Кроме того, P1-P6 показали значительное снижение экспрессии, при условии, что используют слитый белок, и ограниченную реессию, при условии, что используют нехимерный dCas9.

Пример 6. Cas9 может использовать искусственные "направляющие" РНК, не существующие в природе, чтобы осуществить расщепление ДНК-мишени.

Искусственную crРНК и искусственную tracrРНК создают на основе белок-связывающего сегмента встречающихся в природе транскриптов crРНК и tracrРНК *S. pyogenes*, модифицированных для имитации асимметричного выпетливания в пределах естественного дуплекса crРНК:tracrРНК *S. pyogenes* (см. выпетливание в белок-связывающем домене кик искусственных (вверху), так и природных молекул (внизу) РНК, изображенных на фиг. 57А). Искусственная последовательность tracrРНК имеет менее 50% идентичности с естественной tracrРНК. Прогнозируемая вторичная структура белок-связывающего дуплекса crРНК:tracrРНК является одинаковой для обеих пар РНК, но прогнозируемая структура остальной части РНК значительно отличается.

Фиг. 57 иллюстрирует то, что искусственные последовательности, которые разделяют очень небольшую (идентичность около 50%) с встречающимися в природе tracrРНК и crРНК, могут функционировать с Cas9 для расщепления ДНК-мишени до тех пор, пока сохраняется структура белок-связывающего домена РНК, нацеленной на ДНК. (А) Совместная укладка tracrРНК и crРНК *S. pyogenes* и искусственных tracrРНК и crРНК. (В) Комбинацию Cas9 *S. pyogenes* и ортологов tracrРНК:crРНК используют для анализов расщепления плазмидной ДНК. *Spy* - *S. pyogenes*, *Lin* - *L. innocua*, *Nme* - *N. meningitidis*, *Pmu* - *P. multocida*. *S. pyogenes* Cas9 может управляться некоторыми, но не всеми, ортологами tracrРНК: crРНК, встречающимися в природе в отдельных видах бактерий. Примечательно, что Cas9 *S. pyogenes* может управляться парой искусственных tracrРНК:crРНК, которая была создана на основе структуры белка-связывающего сегмента встречающейся в природе РНК, нацеленной на ДНК, с использованием последовательности, совершенно несвязанной с системой CRISPR.

Используемая искусственная "tracrРНК" (активатор РНК) представляет собой 5'-GUUUUCCCUUUUCAAGAAUCUCCUGGGCACCUAUCUUCUAGGUGCCCUCCCUUGUUU-AAACCUGACCAGUUAACCGGCUGGUUAGGUUUUU-3' (SEQ ID NO:1347). Используемая искусственная "crРНК" (нацеливающая РНК) представляет собой: 5'-GAGAUUUUAUGAAAAGGGAAC-3' (SEQ ID NO:1348).

Пример 7. Генерирование трансгенных, не принадлежащих к человеческому роду организмов.

Экспрессирование Cas9 у трансгенной мыши (либо без изменений, либо измененных таким образом, что уменьшается ферментативная активность, либо модифицировано в виде слитого белка для любых целей, очерченных выше) получают с использованием пригодного способа, известного обычному специалисту в данной области техники (например, (i) нокаут гена локуса-мишени (например, ROSA 26) в эмбриональных стволовых клетках мыши (клетка ES), а затем инъекция их в бластоцисту и получение химерных мышей; (ii) инъекции случайно интегрирующего трансгена в пронуклеус оплодотворенной яйцеклетки мыши с последующим имплантацией яйцеклетки псевдобеременной самке; и т.д.). Белок

Cas9 находится под контролем промотора, который экспрессируется, по меньшей мере, в эмбриональных стволовых клетках и может находиться дополнительно под временным или тканеспецифическим контролем (например, индуцируемая лекарственным средством, контролируемая Cre/Lox промоторная система и т.д.), После того, как получают линии трансгенных мышей, экспрессирующих Cas9, выделяют и культивируют эмбриональные стволовые клетки, а в некоторых случаях клетки ES замораживают для дальнейшего использования. Поскольку выделенные клетки ES экспрессируют Cas9 (а в некоторых случаях экспрессия находится под временным управлением (например, индуцируемая лекарственным средством), новый нокаут или нокин в клетках (и, следовательно, мыши) быстро получают в любом желаемом локусе в геноме с помощью введения соответствующей спроектированной РНК, нацеленной на ДНК, которая направляет Cas9 на определенный предпочтительный локус. Такая система и многие ее варианты используют для создания новых генетически модифицированных организмов по любому предпочтительному локусу. При условии, что модифицированный Cas9 используют для модулирования транскрипции и/или модифицирования ДНК и/или модифицирования полипептидов, ассоциированных с ДНК, сами клетки ES (или какие-либо дифференцированные клетки, полученные из клеток ES (например, вся мышь, дифференцированная клеточная линия и т.д.)) используют для изучения свойств любого предпочтительного гена (или любого предпочтительного продукта экспрессии, или любого предпочтительного геномного локуса) просто с помощью введения соответствующей РНК, нацеленной на ДНК, в необходимую клетку, экспрессирующую Cas9.

Хотя настоящее изобретение описывают со ссылкой на конкретные варианты его реализации, специалистам в данной области техники должно быть понятно, что могут быть сделаны различные изменения, а эквиваленты могут быть заменены без отхода от истинной сущности и объема настоящего изобретения. Кроме того, многие модификации могут быть сделаны для того, чтобы адаптировать конкретную ситуацию, материал, композицию вещества, способ, технологический этап или этапы по отношению к объекту, сущности и объему настоящего изобретения. Все подобные модификации находятся в пределах объема прилагаемой формулы изобретения.

#### ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Способ направления полипептида Cas9 на последовательность-мишень в ДНК-мишени, приводящего к связыванию ДНК-мишени, или связыванию и модификации ДНК-мишени, или связыванию ДНК-мишени и модуляции транскрипции с ДНК-мишени, включающий приведение ДНК-мишени в контакт с комплексом, содержащим:

(a) указанный полипептид Cas9 и

(b) одиночную молекулу РНК, нацеленную на ДНК, содержащую:

(i) нацеливающий на ДНК сегмент, содержащий нуклеотидную последовательность, которая является комплементарной указанной последовательности-мишени в ДНК-мишени; и

(ii) белок-связывающий сегмент, который взаимодействует с указанным полипептидом Cas9, где белок-связывающий сегмент содержит два комплементарных участка нуклеотидов, которые гибридизуются друг с другом, для того, чтобы сформировать дуплекс двухцепочечной РНК (дцРНК), где указанный дуплекс дцРНК содержит комплементарные нуклеотиды tracrPНК и РНК CRISPR (crPНК),

где указанные два комплементарных участка нуклеотидов ковалентно связаны с помощью промежуточных нуклеотидов.

2. Способ направления полипептида Cas9 на последовательность-мишень в ДНК-мишени, приводящего к связыванию ДНК-мишени, или связыванию и модификации ДНК-мишени, или связыванию и модуляции транскрипции с ДНК-мишени, включающий приведение ДНК-мишени в контакт с комплексом, содержащим:

(a) указанный полипептид Cas9 и

(b) РНК, нацеленную на ДНК, содержащую:

(i) нацеливающий на ДНК сегмент, содержащий нуклеотидную последовательность, которая является комплементарной указанной последовательности-мишени в ДНК-мишени; и

(ii) белок-связывающий сегмент, который взаимодействует с указанным полипептидом Cas9, где белок-связывающий сегмент содержит два комплементарных участка нуклеотидов, которые гибридизуются друг с другом, для того, чтобы сформировать дуплекс двухцепочечной РНК (дцРНК), где указанный дуплекс дцРНК содержит комплементарные нуклеотиды tracrPНК и РНК CRISPR (crPНК),

где указанные два комплементарных участка нуклеотидов ковалентно не связаны с помощью промежуточных нуклеотидов,

где указанное приведение в контакт происходит вне клетки.

3. Способ направления полипептида Cas9 на последовательность-мишень в ДНК-мишени, приводящего к связыванию ДНК-мишени, или связыванию и модификации ДНК-мишени, или связыванию и модуляции транскрипции с ДНК-мишени, включающий приведение ДНК-мишени в контакт с комплексом, содержащим:

(a) указанный полипептид Cas9 и

(b) РНК, нацеленную на ДНК, содержащую:

(i) нацеливающий на ДНК сегмент, содержащий нуклеотидную последовательность, которая является комплементарной указанной последовательности-мишени в ДНК-мишени; и

(ii) белок-связывающий сегмент, который взаимодействует с указанным полипептидом Cas9, где белок-связывающий сегмент содержит два комплементарных участка нуклеотидов, которые гибридизуются друг с другом, для того, чтобы сформировать дуплекс двухцепочечной РНК (дцРНК), где указанный дуплекс дцРНК содержит комплементарные нуклеотиды tracrРНК и РНК CRISPR (crРНК),

где указанные два комплементарных участка нуклеотидов ковалентно не связаны с помощью промежуточных нуклеотидов,

где указанное приведение в контакт происходит в одноклеточном эукариотическом организме, клетке растения, клетке из беспозвоночного животного или клетке из позвоночного животного.

4. Способ по любому из предшествующих пунктов, где указанный дуплекс дцРНК имеет длину от 8 пар оснований (п.о.) до 30 п.о. или от 8 до 10 п.о.

5. Способ по любому из предшествующих пунктов, где процент комплементарности между нуклеотидными последовательностями, которые гибридизуются с образованием дуплекса дцРНК белок-связывающего сегмента, более 70%.

6. Способ по п.1, где ДНК-мишень находится в бактериальной клетке, клетке архей, одноклеточном эукариотическом организме, клетке растения, клетке беспозвоночного или в клетке позвоночного.

7. Способ по п.6, где ДНК-мишень находится в одноклеточном эукариотическом организме, клетке растения, клетке беспозвоночного или клетке позвоночного.

8. Способ по пп.1, 2 или 3, где ДНК-мишень представляет собой хромосомную ДНК.

9. Способ по любому из предшествующих пунктов, где РНК, нацеленная на ДНК, содержит одно или более из модифицированного нуклеотидного основания, модифицированного остова или не природной межнуклеозидной связи, модифицированного сахарного фрагмента, запертой нуклеиновой кислоты и пептидо-нуклеиновой кислоты.

10. Способ по любому из пп.1-8, где РНК, нацеленная на ДНК:

(i) содержит не природную межнуклеозидную связь, которая включает один или более из фосфориата, фосфорамидата, нефосфодиэфира, гетероатома, хирального фосфориата, фосфородитиата, фосфотриэфира, аминоалкилфосфотриэфира, 3'-алкиленфосфоната, 5'-алкиленфосфоната, хирального фосфоната, фосфината, 3'-аминофосфорамидата, аминоалкилфосфорамидата, фосфородиамидата, тионофосфорамидата, тионоалкилфосфоната, тионоалкилфосфотриэфира, селенофосфата и боранофосфата; или

(ii) содержит один или более из (a) не природной межнуклеозидной связи, выбранной из фосфориата, связи обратной полярности, нуклеозидной связи с удаленным азотистым основанием; (b) запертой нуклеиновой кислоты (LNA); и (c) модифицированного фрагмента сахара, выбранного из 2'-О-метоксиэтоксиды, 2'-О-метил и 2'-фтор; или

(iii) содержит один или более модифицированных фрагментов сахара, выбранных из 2'-О-(2-метоксиэтил), 2'-диметиламинооксиэтоксиды, 2'-диметиламинооксиэтоксиды, 2'-О-метил и 2'-фтор; или

(iv) содержит основание нуклеотида, включающее одно или более из 5-метилцитозина, 5-гидроксиметилцитозина, ксантина, гипоксантина, 2-аминоаденина, 6-метил производных аденина, 6-метил производных гуанина, 2-пропил производных аденина, 2-пропил производных гуанина, 2-тиоурацила, 2-тиотимина, 2-тиоцитозина, 5-пропинилурацила, 5-пропинилцитозина, 6-азоурацила, 6-азоцитозина, 6-азотимина, псевдоурацила, 4-тиоурацила, 8-галогеноаденина, 8-аминоаденина, 8-тиоладенина, 8-тиоалкиладенина, 8-гидроксиладенина, 8-галогеногуанина, 8-аминогуанина, 8-тиолгуанина, 8-тиоалкилгуанина, 8-гидроксиалкилгуанина, 5-галогеноцитозина, 5-бромцитозина, 5-трифторметилцитозина, 5-замещенного урацила, 5-замещенного цитозина, 7-метилгуанина, 7-метиладенина, 2-Ф-аденина, 2-аминоаденина, 8-азагуанина, 8-азааденина, 7-деазагуанина, 7-деазааденина, 3-деазагуанина, 3-деазааденина, трициклического пиримидина, феноксазинцитидина, фенотиазинцитидина, замещенного феноксазинцитидина, карбазолцитидина, пиридоиндолцитидина, 7-деазагуанозина, 2-аминопиридина, 2-пиридоина, 5-замещенного пиримидина, 6-азапиримидина, N-2, N-6 или O-6-замещенного пурина, 2-аминопропиладенина, 5-замещенного пиримидина, 6-азапиримидина, N-2, N-6 или O-6-замещенного пурина, 2-аминопропиладенина, 5-пропинилурацила и 5-пропинилцитозина.

11. Способ по любому из пп.1-10, в котором РНК, нацеленная на ДНК, конъюгирована с фрагментом, выбранным из полиамина, полиамида, полиэтиленгликоля, полиэфира, холестерина фрагмента, хлестероидной кислоты, тиоэфира, тиохолестерина, алифатической цепи, фосфолипида, адамантан-уксусной кислоты, палмитильного фрагмента, октадециламина или фрагмента гексиламино-карбонил-оксистерина, биотина, феназина, фолата, фенантридина, антрахинона, акридина, флуоресцеина, родамина, кумарина, группы, которая улучшает поглощение, повышает устойчивость к деградации и/или усиливает последовательность-специфическую гибридизацию, и группы, которая улучшает поглощение, распределение, метаболизм или экскрецию.

12. Способ по любому из пп.1-10, в котором РНК, нацеленная на ДНК, конъюгирована с красителем.

13. Способ по п.1 или 3, где приведение в контакт включает введение в клетку:

(a) указанного полипептида Cas9 или полинуклеотида, кодирующего указанный полипептид Cas9, и  
 (b) указанной РНК, нацеленной на ДНК, или нуклеиновой кислоты, кодирующей указанную РНК, нацеленную на ДНК.

14. Способ по п.13, где полинуклеотид, кодирующий указанный полипептид Cas9 и/или нуклеиновая кислота, кодирующая указанную РНК, нацеленную на ДНК, представляет собой рекомбинантный экспрессионный вектор.

15. Способ по п.14, где рекомбинантный экспрессионный вектор является вирусным вектором.

16. Способ по п.13, где способ дополнительно включает введение донорного полинуклеотида в клетку, где ДНК-мишень редактируют посредством вставки последовательности указанного донорного полинуклеотида в расщепленную цепь ДНК-мишени.

17. Способ по любому из предшествующих пунктов, где аминоконец полипептида Cas9 ковалентно связан с доменом белковой трансдукции, где указанный домен белковой трансдукции способствует перемещению полипептида Cas9 из цитозоля во внутреннее пространство органеллы клетки.

18. Способ по любому из пп.1-16, где карбоксильный конец полипептида Cas9 ковалентно связан с доменом белковой трансдукции, где указанный домен белковой трансдукции способствует перемещению полипептида Cas9 из цитозоля во внутреннее пространство органеллы клетки.

19. Способ по любому из пп.1, 6 и 7, где указанные промежуточные нуклеотиды составляют от 3 до 5 нуклеотидов.

20. Способ по любому из предшествующих пунктов, в котором указанное направление полипептида Cas9 к последовательности-мишени в ДНК-мишени приводит к связыванию и модификации ДНК-мишени.

21. Способ по п.20, в котором указанная модификация представляет собой расщепление ДНК-мишени.

22. Способ по любому из предшествующих пунктов, в котором нуклеотидная последовательность, комплементарная последовательности-мишени в ДНК-мишени, имеет длину от 18 до 25 нуклеотидов.

23. Композиция для направления полипептида Cas9 на последовательность-мишень в ДНК-мишени, приводящего к связыванию ДНК-мишени, или связыванию и модификации ДНК-мишени, или связыванию ДНК-мишени и модуляции транскрипции с ДНК-мишени, содержащая:

(a) указанный полипептид Cas9 или полинуклеотид, кодирующий указанный полипептид Cas9, и

(b) одиночную молекулу РНК, нацеленную на ДНК, или нуклеиновую кислоту, кодирующую указанную РНК, нацеленную на ДНК, причем указанная одиночная молекула РНК, нацеленная на ДНК, содержит:

(i) нацеливающий на ДНК сегмент, содержащий нуклеотидную последовательность, которая является комплементарной последовательности-мишени в ДНК-мишени; и

(ii) белок-связывающий сегмент, который взаимодействует с указанным полипептидом Cas9, где белок-связывающий сегмент содержит два комплементарных участка нуклеотидов, которые гибридизуются друг с другом, для того, чтобы сформировать дуплекс двухцепочечной РНК (дцРНК), где указанный дуплекс дцРНК содержит комплементарные нуклеотиды tracrPНК и РНК CRISPR (crPНК), и

где указанные два комплементарных участка нуклеотидов ковалентно связаны с помощью промежуточных нуклеотидов.

24. Композиция для направления полипептида Cas9 на последовательность-мишень в ДНК-мишени, приводящего к связыванию ДНК-мишени, или связыванию и модификации ДНК-мишени, или связыванию ДНК-мишени и модуляции транскрипции с ДНК-мишени, содержащая сконструированный и/или не происходящий из природы комплекс, содержащий:

(a) указанный полипептид Cas9 или полинуклеотид, кодирующий указанный полипептид Cas9, и

(b) РНК, нацеленную на ДНК, которая содержит:

(i) нацеливающий на ДНК сегмент, содержащий нуклеотидную последовательность, которая является комплементарной последовательности-мишени в ДНК-мишени; и

(ii) белок-связывающий сегмент, который взаимодействует с указанным полипептидом Cas9, где белок-связывающий сегмент содержит два комплементарных участка нуклеотидов, которые гибридизуются друг с другом, для того, чтобы сформировать дуплекс двухцепочечной РНК (дцРНК), где указанный дуплекс дцРНК содержит комплементарные нуклеотиды tracrPНК и РНК CRISPR (crPНК),

где указанные два комплементарных участка нуклеотидов ковалентно не связаны с помощью промежуточных нуклеотидов,

где РНК, нацеленная на ДНК, содержит одно или более из модифицированного нуклеотидного основания, модифицированного остова или не природной межнуклеозидной связи, модифицированного сахарного фрагмента, запертой нуклеиновой кислоты и пептидо-нуклеиновой кислоты.

25. Композиция для направления полипептида Cas9 на последовательность-мишень в ДНК-мишени, приводящего к связыванию ДНК-мишени, или связыванию и модификации ДНК-мишени, или связыванию ДНК-мишени и модуляции транскрипции с ДНК-мишени, содержащая сконструированный и/или не происходящий из природы комплекс, содержащий:

(a) указанный полипептид Cas9 или полинуклеотид, кодирующий указанный полипептид Cas9, и

(b) РНК, нацеленную на ДНК, или нуклеиновую кислоту, кодирующую указанную РНК, нацеленную на ДНК, причем указанная РНК, нацеленная на ДНК, содержит:

(i) нацеливающий на ДНК сегмент, содержащий нуклеотидную последовательность, которая является комплементарной последовательности-мишени в ДНК-мишени; и

(ii) белок-связывающий сегмент, который взаимодействует с указанным полипептидом Cas9; где белок-связывающий сегмент содержит два комплементарных участка нуклеотидов, которые гибридизуются друг с другом, для того, чтобы сформировать дуплекс двухцепочечной РНК (дцРНК), где указанный дуплекс дцРНК содержит комплементарные нуклеотиды tracrРНК и РНК CRISPR (crРНК),

где указанные два комплементарных участка нуклеотидов ковалентно не связаны с помощью промежуточных нуклеотидов.

26. Композиция по пп.23, 24 или 25, где указанный дуплекс дцРНК имеет длину от 8 пар оснований (п.о) до 30 п.о. или от 8 до 10 п.о.

27. Композиция по пп.23, 24, 25 или 26, где процент комплементарности между нуклеотидами, которые гибридизуются с образованием дуплекса дцРНК белок-связывающего сегмента, составляет более чем 70%.

28. Композиция по п.23 или 24, где ДНК-мишень находится в бактериальной клетке, клетке архей, одноклеточном эукариотическом организме, клетке растений, клетке беспозвоночного животного или клетке позвоночного животного.

29. Композиция по любому из пп.23-27, где ДНК-мишень находится в одноклеточном эукариотическом организме, клетке растений, клетке беспозвоночного животного или клетке позвоночного животного.

30. Композиция по любому из пп.23-27, где ДНК-мишень является хромосомной ДНК.

31. Композиция по любому из пп.23-30, где аминоконец полипептида Cas9 ковалентно связан с доменом белковой трансдукции, где указанный домен белковой трансдукции способствует перемещению полипептида Cas9 из цитозоля во внутреннее пространство органеллы клетки.

32. Композиция по любому из пп.23-30, где карбоксильный конец полипептида Cas9 ковалентно связан с доменом белковой трансдукции, где указанный домен белковой трансдукции способствует перемещению полипептида Cas9 из цитозоля во внутреннее пространство органеллы клетки.

33. Композиция по любому из пп.23 и 25-32, где РНК, нацеленная на ДНК, содержит одно или более из модифицированного нуклеотидного основания, модифицированного остова или неприродной межнуклеозидной связи, модифицированного сахарного фрагмента, запертой нуклеиновой кислоты и пептидо-нуклеиновой кислоты.

34. Композиция по п.24 или 33, где РНК, нацеленная на ДНК:

(i) содержит неприродную межнуклеозидную связь, которая включает один или более из фосфориата, фосфорамидата, нефосфодиэфира, гетероатома, хирального фосфориата, фосфородитиата, фосфотриэфира, аминоалкилфосфотриэфира, 3'-алкиленфосфоната, 5'-алкиленфосфоната, хирального фосфоната, фосфината, 3'-аминофосфорамидата, аминоалкилфосфорамидата, фосфородиамидата, тионофосфорамидата, тионоалкилфосфоната, тионоалкилфосфотриэфира, селенофосфата и боранофосфата; или

(ii) содержит один или более из (a) неприродной межнуклеозидной связи, выбранной из фосфориата, связи обратной полярности, нуклеозидной связи с удаленным азотистым основанием; (b) запертой нуклеиновой кислоты (LNA); и (c) модифицированного фрагмента сахара, выбранного из 2'-О-метоксиэтокси, 2'-О-метил и 2'-фтор; или

(iii) содержит один или более модифицированных фрагментов сахара, выбранных из 2'-О-(2-метоксиэтил), 2'-диметиламинооксиэтокси, 2'-диметиламиноэтоксиэтокси, 2'-О-метил и 2'-фтор; или

(iv) содержит основание нуклеотида, включающее одно или более из 5-метилцитозина, 5-гидроксиметилцитозина, ксантина, гипоксантина, 2-аминоаденина, 6-метил производных аденина, 6-метил производных гуанина, 2-пропил производных аденина, 2-пропил производных гуанина, 2-тиоурацила, 2-тиотимина, 2-тиоцитозина, 5-пропинилурацила, 5-пропинилцитозина, 6-азоурацила, 6-азоцитозина, 6-азотимина, псевдоурацила, 4-тиоурацила, 8-галогеноаденина, 8-аминоаденина, 8-тиоаденина, 8-тиоалкиладенина, 8-гидроксиладенина, 8-галогеногуанина, 8-аминогуанина, 8-тиолгуанина, 8-тиоалкилгуанина, 8-гидроксилагуанина, 5-галогеноцитозина, 5-бромцитозина, 5-трифторметилцитозина, 5-замещенного урацила, 5-замещенного цитозина, 7-метилгуанина, 7-метиладенина, 2-Ф-аденина, 2-аминоаденина, 8-азагуанина, 8-азааденина, 7-деазагуанина, 7-деазааденина, 3-деазагуанина, 3-деазааденина, трициклического пиримидина, феноксазинцитидина, фенотиазинцитидина, замещенного феноксазинцитидина, карбазолцитидина, пиридоиндолцитидина, 7-деазагуанозина, 2-аминопиридина, 2-пиридоина, 5-замещенного пиримидина, 6-азапиримидина, N-2, N-6 или O-6-замещенного пурина, 2-аминопропиладенина, 5-замещенного пиримидина, 6-азапиримидина, N-2, N-6 или O-6-замещенного пурина, 2-аминопропиладенина, 5-пропинилурацила и 5-пропинилцитозина.

35. Композиция по любому из пп.23-34, в которой РНК, нацеленная на ДНК, конъюгирована с фрагментом, выбранным из полиамина, полиамида, полиэтиленгликоля, полиэфира, холестерина фрагмента, хольстеринового фрагмента, холевой кислоты, тиоэфира, тиохлестерина, алифатической цепи, фосфолипида, адамантанкислоты, палмитильного фрагмента, октадециламина или фрагмента гексиламино-карбонил-

оксистерина, биотина, феназина, фолата, фенантридина, антрахинона, акридина, флуоресцеина, родамина, кумарина, группы, которая улучшает поглощение, повышает устойчивость к деградации и/или усиливает последовательность-специфическую гибридизацию, и группы, которая улучшает поглощение, распределение, метаболизм или экскрецию.

36. Композиция по любому из пп.23-34, где РНК, нацеленная на ДНК, конъюгирована с красителем.

37. Композиция по любому из пп.23-36, где полинуклеотид, кодирующий указанный полипептид Cas9 и/или нуклеиновая кислота, кодирующая указанную РНК, нацеленную на ДНК, представляет собой рекомбинантный экспрессионный вектор.

38. Композиция по п.37, где рекомбинантный экспрессионный вектор является вирусным вектором.

39. Композиция по п.23, где указанные промежуточные нуклеотиды составляют от 3 до 5 нуклеотидов.

40. Композиция по любому из пп.23-39, где указанная композиция предназначена для направления полипептида Cas9 к последовательности-мишени в ДНК-мишени, что приводит к связыванию и модификации ДНК-мишени.

41. Композиция по п.40, где указанная модификация представляет собой расщепление ДНК-мишени.

42. Композиция по любому из пп.23-41, отличающаяся тем, что нуклеотидная последовательность, комплементарная последовательности-мишени в ДНК-мишени, имеет длину от 18 до 25 нуклеотидов.

43. Композиция по любому из пп.23-42, отличающаяся тем, что композиция является лиофилизированной.

44. Композиция по любому из пп.23-42, дополнительно содержащая один или несколько из следующих компонентов: ингибитор нуклеазы, буферный агент, детергент, полиамин, адъювант, смачивающий агент, стабилизирующий агент, антиоксидант и комплексообразователь.

45. Композиция по любому из пп.23-42, дополнительно содержащая ингибитор нуклеазы.

46. Композиция по любому из пп.23-42, дополнительно содержащая буферный агент.

47. Одиночная молекула РНК, нацеленная на ДНК, для направления полипептида Cas9 на последовательность-мишень в ДНК-мишени, приводящего к связыванию ДНК-мишени, или связыванию и модификации ДНК-мишени, или связыванию ДНК-мишени и модуляции транскрипции с ДНК-мишени, где одиночная молекула РНК, нацеленная на ДНК, содержит:

(i) нацеленный на ДНК сегмент, содержащий нуклеотидную последовательность, которая является комплементарной указанной последовательности-мишени в ДНК-мишени; и

(ii) связывающий белок сегмент, который взаимодействует с белком Cas9, где связывающий белок сегмент содержит два комплементарных участка нуклеотидов, которые гибридизуются друг с другом с образованием дуплекса двухцепочечной РНК (дцРНК), где указанный дцРНК дуплекс содержит комплементарные нуклеотиды tracrRNA и CRISPR РНК (crRNA), и где указанные два комплементарных участка нуклеотидов ковалентно связаны с помощью промежуточных нуклеотидов.

48. Одиночная молекула РНК, нацеленная на ДНК, по п.47, где указанный дуплекс дцРНК имеет длину от 8 пар оснований (п.о) до 30 п.о. или от 8 до 10 п.о.

49. Одиночная молекула РНК, нацеленная на ДНК, по п.47 или 48, где процент комплементарности между нуклеотидами, которые гибридизуются с образованием дуплекса дцРНК белок-связывающего сегмента, составляет более чем 70%.

50. Одиночная молекула РНК, нацеленная на ДНК, по пп.47, 48 или 49, где ДНК-мишень находится в бактериальной клетке, клетке архей, одноклеточном эукариотическом организме, клетке растений, клетке беспозвоночного животного или клетке позвоночного животного.

51. Одиночная молекула РНК, нацеленная на ДНК, по пп.47, 48 или 49, где ДНК-мишень является хромосомной ДНК.

52. Одиночная молекула РНК, нацеленная на ДНК, по любому из пп.47-51, где одиночная молекула РНК, нацеленная на ДНК, содержит одно или более из модифицированного нуклеотидного основания, модифицированного остова или неприродной межнуклеозидной связи, модифицированного сахарного фрагмента, запертой нуклеиновой кислоты и пептидо-нуклеиновой кислоты.

53. Одиночная молекула РНК, нацеленная на ДНК, по любому из пп.47-51, где одиночная молекула РНК, нацеленная на ДНК:

(i) содержит неприродную межнуклеозидную связь, которая включает один или более из фосфотиата, фосфорамидата, нефосфодиэфира, гетероатома, хирального фосфотиата, фосфородитиата, фосфотриэфира, аминоалкилфосфотриэфира, 3'-алкиленфосфоната, 5'-алкиленфосфоната, хирального фосфоната, фосфината, 3'-аминофосфорамидата, аминоалкилфосфорамидата, фосфородиамидата, тионофосфорамидата, тионоалкилфосфоната, тионоалкилфосфотриэфира, селенофосфата и боранофосфата; или

(ii) содержит один или более из (а) неприродной межнуклеозидной связи, выбранной из фосфотиата, связи обратной полярности, нуклеозидной связи с удаленным азотистым основанием; (b) запертой нуклеиновой кислоты (LNA); и (с) модифицированного фрагмента сахара, выбранного из 2'-О-метоксиэтокси, 2'-О-метил и 2'-фтор; или

(iii) содержит один или более модифицированных фрагментов сахара, выбранных из 2'-О-(2-

метоксиэтил), 2'-диметиламинооксиэтокси, 2'-диметиламиноэтоксиэтокси, 2'-О-метил и 2'-фтор; или

(iv) содержит основание нуклеотида, включающее одно или более из 5-метилцитозина, 5-гидроксиметилцитозина, ксантина, гипоксантина, 2-аминоаденина, 6-метил производных аденина, 6-метил производных гуанина, 2-пропил производных аденина, 2-пропил производных гуанина, 2-тиоурацила, 2-тиотимина, 2-тиоцитозина, 5-пропинилурацила, 5-пропинилцитозина, 6-азоурацила, 6-азоцитозина, 6-азотимина, псевдоурацила, 4-тиоурацила, 8-галогеноаденина, 8-аминоаденина, 8-тиоладенина, 8-тиоалкиладенина, 8-гидроксиладенина, 8-галогеногуанина, 8-аминогуанина, 8-тиолгуанина, 8-тиоалкилгуанина, 8-гидроксиалкилгуанина, 5-галогеноцитозина, 5-бромцитозина, 5-трифторметилцитозина, 5-замещенного урацила, 5-замещенного цитозина, 7-метилгуанина, 7-метиладенина, 2-Ф-аденина, 2-аминоаденина, 8-азагуанина, 8-азааденина, 7-деазагуанина, 7-деазааденина, 3-деазагуанина, 3-деазааденина, трициклического пиримидина, феноксазинцитидина, фенотиазинцитидина, замещенного феноксазинцитидина, карбазолцитидина, пиридоиндолцитидина, 7-деазагуанозина, 2-аминопиридина, 2-пиридоина, 5-замещенного пиримидина, 6-азапиримидина, N-2, N-6 или O-6-замещенного пурина, 2-аминопропиладенина, 5-замещенного пиримидина, 6-азапиримидина, N-2, N-6 или O-6-замещенного пурина, 2-аминопропиладенина, 5-пропинилурацила и 5-пропинилцитозина.

54. Одиночная молекула РНК, нацеленная на ДНК, по любому из пп.47-53, где одиночная молекула РНК, нацеленная на ДНК, конъюгирована с фрагментом, выбранным из полиамина, полиамида, полиэтиленгликоля, полиэфира, холестерина фрагмента, холевой кислоты, тиоэфира, тиохолестерина, алифатической цепи, фосфолипида, адамантан-уксусной кислоты, палмитильного фрагмента, октадециламина или фрагмента гексиламино-карбонил-оксихолестерина, биотина, феназина, фолата, фенантридина, антрахинона, акридина, флуоресцеина, родамина, кумарина, группы, которая улучшает поглощение, повышает устойчивость к деградации и/или усиливает последовательность-специфическую гибридизацию, и группы, которая улучшает поглощение, распределение, метаболизм или экскрецию.

55. Одиночная молекула РНК, нацеленная на ДНК, по любому из пп.47-53, где одиночная молекула РНК, нацеленная на ДНК, конъюгирована с красителем.

56. Одиночная молекула РНК, нацеленная на ДНК, по любому из пп.47-55, отличающаяся тем, что указанные промежуточные нуклеотиды составляют от 3 до 5 нуклеотидов.

57. Одиночная молекула РНК, нацеленная на ДНК, по любому из пп.47-56, отличающаяся тем, что указанная нацеленная на ДНК одиночная молекула РНК предназначена для направления полипептида Cas9 к последовательности-мишени в ДНК-мишени, приводя к связыванию и модификации ДНК-мишени.

58. Одиночная молекула РНК, нацеленная на ДНК, по п.57, где указанная модификация представляет собой расщепление ДНК-мишени.

59. Одиночная молекула РНК, нацеленная на ДНК, по любому из пп.47-58, отличающаяся тем, что нуклеотидная последовательность, комплементарная последовательности-мишени в ДНК-мишени, имеет длину от 18 до 25 нуклеотидов.

60. Одиночная молекула РНК, нацеленная на ДНК, по любому из пп.47-59, отличающаяся тем, что указанная одиночная молекула РНК, нацеленная на ДНК, лиофилизирована.

61. Одиночная молекула РНК, нацеленная на ДНК, по любому из пп.47-59, отличающаяся тем, что указанная одиночная молекула РНК, нацеленная на ДНК, присутствует в композиции, содержащей одно или несколько из следующего: ингибитор нуклеазы, буферный агент, детергент, полиамин, адъювант, смачивающий агент, стабилизирующий агент, антиоксидант и комплексообразователь.

62. Одиночная молекула РНК, нацеленная на ДНК, по любому из пп.47-59, отличающаяся тем, что указанная одиночная молекула РНК, нацеленная на ДНК, присутствует в композиции, содержащей ингибитор нуклеазы.

63. Одиночная молекула РНК, нацеленная на ДНК, по любому из пп.47-59, отличающаяся тем, что указанная одиночная молекула РНК, нацеленная на ДНК, присутствует в композиции, содержащей буферный агент.

64. Композиция для направления полипептида Cas9 на последовательность-мишень в ДНК-мишени, приводящего к связыванию ДНК-мишени, или связыванию и модификации ДНК-мишени, или связыванию ДНК-мишени и модуляции транскрипции с ДНК-мишени, содержащая одну или несколько нуклеиновых кислот, где указанные одна или несколько нуклеиновых кислот содержат:

(a) первую нуклеотидную последовательность, кодирующую одиночную молекулу РНК, нацеленную на ДНК, содержащую:

(i) нацеленный на ДНК сегмент, содержащий нуклеотидную последовательность, которая является комплементарной последовательности-мишени в ДНК-мишени; и

(ii) связывающий белок сегмент, который взаимодействует с указанным полипептидом Cas9, где связывающий белок сегмент содержит два комплементарных участка нуклеотидов, которые гибридизуются друг с другом с образованием дуплекса двухцепочечной РНК (дцРНК), где указанный дцРНК дуплекс содержит комплементарные нуклеотиды tracrNK и CRISPR РНК (crРНК), и где указанные два комплементарных участка нуклеотидов ковалентно связаны с помощью промежуточных нуклеотидов,

где первая нуклеотидная последовательность, кодирующая указанную РНК, нацеленную на ДНК,

функционально связана с промотором, и

(b) вторую нуклеотидную последовательность, кодирующую полипептид Cas9, где нуклеотидная последовательность, кодирующая указанный полипептид Cas9, функционально связана с промотором.

65. Композиция по п.64, где указанные нуклеиновые кислоты представляют собой один или несколько рекомбинантных экспрессионных векторов.

66. Композиция по п.64, где один или несколько рекомбинантных экспрессионных векторов являются одним или несколькими вирусными векторами.

67. Композиция по пп.64-66, где вторая нуклеотидная последовательность также кодирует домен белковой трансдукции, который ковалентно связан с карбоксильным концом или аминоконцом полипептида Cas9, где домен белковой трансдукции способствует перемещению полипептида Cas9 из цитозоля во внутреннее пространство органеллы клетки.

68. Композиция по любому одному из пп.64-67, где указанный дуплекс дцРНК имеет длину от 8 пар оснований (п.о) до 30 п.о. или от 8 до 10 п.о.

69. Композиция по любому одному из пп.64-68, где процент комплементарности между нуклеотидами, которые гибридизуются с образованием дуплекса дцРНК белок-связывающего сегмента, составляет более чем 70%.

70. Композиция по любому из пп.64-69, где ДНК-мишень находится в бактериальной клетке, клетке архей, одноклеточном эукариотическом организме, клетке растений, клетке беспозвоночного животного или клетке позвоночного животного.

71. Композиция по любому из пп.64-70, где ДНК-мишень является хромосомной ДНК.

72. Композиция по любому из пп.64-71, отличающаяся тем, что указанные промежуточные нуклеотиды составляют от 3 до 5 нуклеотидов.

73. Композиция по любому из пп.64-72, отличающаяся тем, что предназначена для направления полипептида Cas9 к последовательности-мишени в ДНК-мишени, что приводит к связыванию и модификации ДНК-мишени.

74. Композиция по п.73, отличающаяся тем, что указанная модификация представляет собой расщепление ДНК-мишени.

75. Композиция по любому из пп.64-74, отличающаяся тем, что нуклеотидная последовательность, комплементарная последовательности-мишени в ДНК-мишени, имеет длину от 18 до 25 нуклеотидов.

76. Композиция по любому из пп.64-75, отличающаяся тем, что композиция является лиофилизированной.

77. Композиция по любому из пп.64-75, дополнительно содержащая один или несколько из следующих компонентов: ингибитор нуклеазы, буферный агент, детергент, полиамин, адъювант, смачивающий агент, стабилизирующий агент, антиоксидант и комплексообразователь.

78. Композиция по любому из пп.64-75, дополнительно содержащая ингибитор нуклеазы.

79. Композиция по любому из пп.64-75, дополнительно содержащая буферный агент.

80. Набор для направления полипептида Cas9 на последовательность-мишень в ДНК-мишени, приводящего к связыванию ДНК-мишени, или связыванию и модификации ДНК-мишени, или связыванию ДНК-мишени и модуляции транскрипции с ДНК-мишени, содержащий:

(a) указанный полипептид Cas9 или нуклеиновую кислоту, содержащую нуклеотидную последовательность, кодирующую указанный полипептид Cas9, и

(b) одиночную молекулу РНК, нацеленную на ДНК, или нуклеиновую кислоту, содержащую нуклеотидную последовательность, кодирующую указанную одиночную молекулу РНК, нацеленную на ДНК, причем указанная одиночная молекула РНК, нацеленная на ДНК, содержит:

(i) нацеленный на ДНК сегмент, содержащий нуклеотидную последовательность, которая является комплементарной последовательности-мишени в ДНК-мишени; и

(ii) связывающий белок сегмент, который взаимодействует с указанным полипептидом Cas9, где связывающий белок сегмент содержит два комплементарных участка нуклеотидов, которые гибридизуются друг с другом с образованием дуплекса двухцепочечной РНК (дцРНК), где указанный дцРНК дуплекс содержит комплементарные нуклеотиды tracrRNA и CRISPR РНК (crRNA), и где указанные два комплементарных участка нуклеотидов ковалентно связаны с помощью промежуточных нуклеотидов, и, где (a) и (b) находятся в одном или разных контейнерах.

81. Набор для направления полипептида Cas9 на последовательность-мишень в ДНК-мишени, приводящего к связыванию ДНК-мишени, или связыванию и модификации ДНК-мишени, или связыванию ДНК-мишени и модуляции транскрипции с ДНК-мишени, содержащий:

(a) указанный полипептид Cas9 или нуклеиновую кислоту, содержащую нуклеотидную последовательность, кодирующую указанный полипептид Cas9, и

(b) РНК, нацеленную на ДНК, или нуклеиновую кислоту, содержащую нуклеотидную последовательность, кодирующую указанную РНК, нацеленную на ДНК, причем указанная молекула РНК, нацеленная на ДНК, содержит:

(i) нацеленный на ДНК сегмент, содержащий нуклеотидную последовательность, которая является комплементарной последовательности-мишени в ДНК-мишени; и

(ii) связывающий белок сегмент, который взаимодействует с указанным полипептидом Cas9, где связывающий белок сегмент содержит два комплементарных участка нуклеотидов, которые гибридизуются друг с другом с образованием дуплекса двухцепочечной РНК (дцРНК), где указанный дцРНК дуплекс содержит комплементарные нуклеотиды tracrРНК и CRISPR РНК (crРНК),

где указанные два комплементарных участка нуклеотидов не связаны ковалентно с помощью промежуточных нуклеотидов, и

где (a) и (b) находятся в одном или разных контейнерах.

82. Набор по п.80 или 81, где ДНК-мишень находится в бактериальной клетке, клетке архей, одно-клеточном эукариотическом организме, клетке растений, клетке беспозвоночного животного или клетке позвоночного животного.

83. Набор по п.80 или 81, где ДНК-мишень является хромосомной ДНК.

84. Набор по любому одному из пп.80-83, где указанная нуклеиновая кислота, содержащая нуклеотидную последовательность, кодирующий указанный полипептид Cas9, и/или указанная нуклеиновая кислота, содержащая нуклеотидную последовательность, кодирующую указанную одиночную молекулу РНК, нацеленную на ДНК, представляет собой рекомбинантный экспрессионный вектор.

85. Набор по п.84, где рекомбинантный экспрессионный вектор является вирусным вектором.

86. Набор по любому одному из пп.80-85, где карбоксильный конец или аминоконец полипептида Cas9 ковалентно связан с доменом белковой трансдукции, где указанный домен белковой трансдукции способствует перемещению полипептида Cas9 из цитозоля во внутреннее пространство органеллы клетки.

87. Набор по любому из пп.80-86, где в любом из пунктов РНК, нацеленная на ДНК, содержит одно или более из модифицированного нуклеотидного основания, модифицированного остова или неприродной межнуклеозидной связи, модифицированного сахарного фрагмента, запертой нуклеиновой кислоты и пептидо-нуклеиновой кислоты.

88. Набор по любому из пп.80-86, где РНК, нацеленная на ДНК:

(i) содержит неприродную межнуклеозидную связь, которая включает один или более из фосфориата, фосфорамидата, нефосфодиэфира, гетероатома, хирального фосфориата, фосфородитиата, фосфотриэфира, аминоалкилфосфотриэфира, 3'-алкиленфосфоната, 5'-алкиленфосфоната, хирального фосфоната, фосфината, 3'-аминофосфорамидата, аминоалкилфосфорамидата, фосфородиамидата, тионофосфорамидата, тионоалкилфосфоната, тионоалкилфосфотриэфира, селенофосфата и боранофосфата; или

(ii) содержит один или более из (a) неприродной межнуклеозидной связи, выбранной из фосфориата, связи обратной полярности, нуклеозидной связи с удаленным азотистым основанием; (b) запертой нуклеиновой кислоты (LNA); и (c) модифицированного фрагмента сахара, выбранного из 2'-О-метоксиэтоксид, 2'-О-метил и 2'-фтор; или

(iii) содержит один или более модифицированных фрагментов сахара, выбранных из 2'-О-(2-метоксиэтил), 2'-диметиламинооксиэтоксид, 2'-диметиламинооксиэтоксид, 2'-О-метил и 2'-фтор; или

(iv) содержит основание нуклеотида, включающее одно или более из 5-метилцитозина, 5-гидроксиметилцитозина, ксантина, гипоксантина, 2-аминоаденина, 6-метил производных аденина, 6-метил производных гуанина, 2-пропил производных аденина, 2-пропил производных гуанина, 2-тиоурацила, 2-тиотимина, 2-тиоцитозина, 5-пропинилурацила, 5-пропинилцитозина, 6-азоурацила, 6-азоцитозина, 6-азотимина, псевдоурацила, 4-тиоурацила, 8-галогеноаденина, 8-аминоаденина, 8-тиоаденина, 8-тиоалкиладенина, 8-гидроксиладенина, 8-галогеногуанина, 8-аминогуанина, 8-тиолгуанина, 8-тиоалкилгуанина, 8-гидроксилагуанина, 5-галогеноцитозина, 5-бромцитозина, 5-трифторметилцитозина, 5-замещенного урацила, 5-замещенного цитозина, 7-метилгуанина, 7-метиладенина, 2-Ф-аденина, 2-аминоаденина, 8-азагуанина, 8-азааденина, 7-дезагуанина, 7-дезааденина, 3-дезагуанина, 3-дезааденина, трициклического пиримидина, феноксазинцитидина, фенотиазинцитидина, замещенного феноксазинцитидина, карбазолцитидина, пиридоиндолцитидина, 7-дезагуанозина, 2-аминопиридина, 2-пиридоина, 5-замещенного пиримидина, 6-азапиримидина, N-2, N-6 или O-6-замещенного пурина, 2-аминопропиладенина, 5-замещенного пиримидина, 6-азапиримидина, N-2, N-6 или O-6-замещенного пурина, 2-аминопропиладенина, 5-пропинилурацила и 5-пропинилцитозина.

89. Набор по любому из пп.80-88, в котором РНК, нацеленная на ДНК, конъюгирована с фрагментом, выбранным из полиамина, полиамида, полиэтиленгликоля, полиэфира, холестерина фрагмента, хлестериновой кислоты, тиоэфира, тиохлестерина, алифатической цепи, фосфолипида, адамантан-уксусной кислоты, палмитильного фрагмента, октадециламина или фрагмента гексиламино-карбонил-оксихлестерина, биотина, феназина, фолата, фенантридина, антрахинона, акридина, флуоресцеина, родамина, кумарина, группы, которая улучшает поглощение, повышает устойчивость к деградации и/или усиливает последовательность-специфическую гибридизацию, и группы, которая улучшает поглощение, распределение, метаболизм или экскрецию.

90. Набор по любому из пп.80-88, в котором РНК, нацеленная на ДНК, конъюгирована с красителем.

91. Набор по п.80, отличающийся тем, что указанные промежуточные нуклеотиды составляют от 3 до 5 нуклеотидов.

92. Набор по любому из пп.80-91, отличающийся тем, что набор предназначен для направления полипептида Cas9 к последовательности-мишени в ДНК-мишени, что приводит к связыванию и модификации ДНК-мишени.

93. Набор по п.92, отличающийся тем, что указанная модификация представляет собой расщепление ДНК-мишени.

94. Набор по любому из пп.80-93, отличающийся тем, что нуклеотидная последовательность, комплементарная последовательности-мишени в ДНК-мишени, имеет длину от 18 до 25 нуклеотидов.

95. Набор по любому из пп.80-94, в котором (а), или (б), или (а) и (б) лиофилизированы.

96. Набор по любому из пп.80-95, дополнительно содержащий одно или несколько из ингибитора нуклеазы, буферного агента, детергента, полиамина, адьюванта, смачивающего агента, стабилизирующего агента, антиоксиданта и комплексообразователя.

97. Набор по любому из пп.80-95, дополнительно содержащий ингибитор нуклеазы.

98. Набор по любому из пп.80-95, дополнительно содержащий буферный агент.

99. Генетически модифицированная эукариотическая клетка, содержащая одно или более из следующего:

(а) одиночную молекулу РНК, нацеленную на ДНК, содержащую:

(i) нацеленный на ДНК сегмент, содержащий нуклеотидную последовательность, которая является комплементарной последовательности-мишени в ДНК-мишени, и

(ii) связывающий белок сегмент, который взаимодействует с указанным полипептидом Cas9, где связывающий белок сегмент содержит два комплементарных участка нуклеотидов, которые гибридизуются друг с другом с образованием дуплекса двухцепочечной РНК (дцРНК), где указанный дцРНК дуплекс содержит комплементарные нуклеотиды tracrPНК и CRISPR РНК (сгРНК), где указанные два комплементарных участка нуклеотидов ковалентно связаны с помощью промежуточных нуклеотидов, и

(b) полипептид Cas9 и/или нуклеиновую кислоту, содержащую нуклеотидную последовательность, которая кодирует полипептид Cas9.

100. Генетически модифицированная эукариотическая клетка, содержащая одно или более из следующего:

(а) молекулу РНК, нацеленную на ДНК, содержащую:

(i) нацеленный на ДНК сегмент, содержащий нуклеотидную последовательность, которая является комплементарной последовательности-мишени в ДНК-мишени, и

(ii) связывающий белок сегмент, который взаимодействует с указанным полипептидом Cas9, где связывающий белок сегмент содержит два комплементарных участка нуклеотидов, которые гибридизуются друг с другом с образованием дуплекса двухцепочечной РНК (дцРНК), где указанный дцРНК дуплекс содержит комплементарные нуклеотиды tracrPНК и CRISPR РНК (сгРНК), где указанные два участка комплементарных нуклеотидов не связаны ковалентно промежуточными нуклеотидами, и

(b) полипептид Cas9 и/или нуклеиновую кислоту, содержащую нуклеотидную последовательность, которая кодирует полипептид Cas9.

101. Генетически модифицированная эукариотическая клетка по п.99 или 100, где указанный дуплекс дцРНК имеет длину от 8 пар оснований (п.о) до 30 п.о. или от 8 до 10 п.о.

102. Генетически модифицированная эукариотическая клетка по любому одному из пп.99-101, где процент комплементарности между нуклеотидами, которые гибридизуются с образованием дуплекса дцРНК белок-связывающего сегмента, составляет более чем 70%.

103. Генетически модифицированная эукариотическая клетка по любому из пп.99-102, где ДНК-мишень является хромосомной ДНК указанной эукариотической клетки.

104. Генетически модифицированная эукариотическая клетка по любому из пп.99-103, где карбоксильный конец или аминоконец полипептида Cas9 ковалентно связан с доменом белковой трансдукции, где домен белковой трансдукции способствует перемещению полипептида Cas9 из цитозоля во внутреннее пространство органеллы клетки.

105. Генетически модифицированная эукариотическая клетка по любому из пп.99-104, где РНК, нацеленная на ДНК, содержит одно или более из модифицированного нуклеотидного основания, модифицированного остова или неприродной межнуклеозидной связи, модифицированного сахарного фрагмента, запертой нуклеиновой кислоты и пептидо-нуклеиновой кислоты.

106. Генетически модифицированная эукариотическая клетка по любому из пп.99-104, где РНК, нацеленная на ДНК:

(i) содержит неприродную межнуклеозидную связь, которая включает один или более из фосфотиата, фосфорамидата, нефосфодиэфира, гетероатома, хирального фосфотиата, фосфородитиата, фосфотриэфира, аминоалкилфосфотриэфира, 3'-алкиленфосфоната, 5'-алкиленфосфоната, хирального фосфоната, фосфината, 3'-аминофосфорамидата, аминоалкилфосфорамидата, фосфородиамидата, тионофосфорамидата, тионоалкилфосфоната, тионоалкилфосфотриэфира, селенофосфата и боранофосфата; или

(ii) содержит один или более из (а) неприродной межнуклеозидной связи, выбранной из фосфотиата, связи обратной полярности, нуклеозидной связи с удаленным азотистым основанием; (b) запертой нуклеиновой кислоты (LNA); и (с) модифицированного фрагмента сахара, выбранного из 2'-О-

метоксиэтокси, 2'-0-метил и 2'-фтор; или

(iii) содержит один или более модифицированных фрагментов сахара, выбранных из 2'-О-(2-метоксиэтил), 2'-диметиламинооксиэтокси, 2'-диметиламиноэтоксиэтокси, 2'-О-метил и 2'-фтор; или

(iv) содержит основание нуклеотида, включающее одно или более из 5-метилцитозина, 5-гидроксиметилцитозина, ксантина, гипоксантина, 2-аминоаденина, 6-метил производных аденина, 6-метил производных гуанина, 2-пропил производных аденина, 2-пропил производных гуанина, 2-тиоурацила, 2-тиотимина, 2-тиоцитозина, 5-пропинилурацила, 5-пропинилцитозина, 6-азоурацила, 6-азоцитозина, 6-азотимина, псевдоурацила, 4-тиоурацила, 8-галогеноаденина, 8-аминоаденина, 8-тиоладенина, 8-тиоалкиладенина, 8-гидроксиладенина, 8-галогеногуанина, 8-аминогуанина, 8-тиолгуанина, 8-тиоалкилгуанина, 8-гидроксилгуанина, 5-галогеноцитозина, 5-бромцитозина, 5-трифторметилцитозина, 5-замещенного урацила, 5-замещенного цитозина, 7-метилгуанина, 7-метиладенина, 2-Ф-аденина, 2-аминоаденина, 8-азагуанина, 8-азааденина, 7-деазагуанина, 7-деазааденина, 3-деазагуанина, 3-деазааденина, трициклического пиримидина, феноксазинцитидина, фенотиазинцитидина, замещенного феноксазинцитидина, карбазолцитидина, пиридоиндолцитидина, 7-деазагуанозина, 2-аминопиридина, 2-пиридона, 5-замещенного пиримидина, 6-азапиримидина, N-2, N-6 или O-6-замещенного пурина, 2-аминопропиладенина, 5-замещенного пиримидина, 6-азапиримидина, N-2, N-6 или O-6-замещенного пурина, 2-аминопропиладенина, 5-пропинилурацила и 5-пропинилцитозина.

107. Генетически модифицированная эукариотическая клетка по любому из пп.99-106, где РНК, нацеленная на ДНК, конъюгирована с фрагментом, выбранным из полиамина, полиамида, полиэтиленгликоля, полиэфира, холестеринового фрагмента, хелевой кислоты, тиоэфира, тиохолестерина, алифатической цепи, фосфолипида, адамантан-уксусной кислоты, палмитильного фрагмента, октадециламина или фрагмента гексиламино-карбонил-оксихолестерина, биотина, феназина, фолата, фенантридина, антрахинона, акридина, флуоресцеина, родамина, кумарина, группы, которая улучшает поглощение, повышает устойчивость к деградации и/или усиливает последовательность-специфическую гибридизацию, и группы, которая улучшает поглощение, распределение, метаболизм или экскрецию.

108. Генетически модифицированная эукариотическая клетка по любому из пп.99-106, где РНК, нацеленная на ДНК, конъюгирована с красителем.

109. Способ по любому одному из пп.1-22, где ДНК-мишень редактируют путем негомологичного соединения концов и репарирования по механизму гомологичной рекомбинации.

110. Способ по п.14, где рекомбинантный экспрессионный вектор выбирают из группы, состоящей из плазмид, космид, мини-колец, фагов и вирусных векторов.

111. Способ по п.15, где вирусный вектор выбирают из группы, состоящей из вирусных векторов, основанных на вирусе коровьей оспы, полиовирусе, ретровирусе, лентивирусе, аденовирусе, аденоассоциированном вирусе и вирусе простого герпеса.

112. Способ по любому одному из пп.1-22 и 109-111, в котором указанные комплементарные нуклеотиды tracrRNA и crRNA гибридизуются с образованием от 8 до 15 пар оснований.

113. Способ по любому одному из пп.1-22 и 109-111, в котором указанные комплементарные нуклеотиды tracrRNA и crRNA с образованием от 15 до 18 пар оснований.

114. Способ по любому одному из пп.1-22 и 109-113, где полипептид Cas9 включает одну или несколько мутаций в домене RuvC и/или в домене HNH, которые ведут к сниженной нуклеазной активности по сравнению с соответствующим полипептидом Cas9 дикого типа.

115. Композиция по п.37, где рекомбинантный экспрессионный вектор выбирают из группы, состоящей из плазмид, космид, мини-колец, фагов и вирусных векторов.

116. Композиция по п.38, где вирусный вектор выбирают из группы, состоящей из вирусных векторов, основанных на вирусе коровьей оспы, полиовирусе, ретровирусе, лентивирусе, аденовирусе, аденоассоциированном вирусе и вирусе простого герпеса.

117. Композиция по любому одному из пп.23-46 и 115-116, в которой указанные комплементарные нуклеотиды tracrRNA и crRNA гибридизуются с образованием от 8 до 15 пар оснований.

118. Композиция по любому одному из пп.23-46 и 115-116, в которой указанные комплементарные нуклеотиды tracrRNA и crRNA гибридизуются с образованием от 15 до 18 пар оснований.

119. Композиция по любому одному из пп.23-46 и 115-118, где полипептид Cas9 включает одну или несколько мутаций в домене RuvC и/или в домене HNH, которые ведут к сниженной нуклеазной активности по сравнению с соответствующим полипептидом Cas9 дикого типа.

120. Одиночная молекула РНК, нацеленная на ДНК, по любому из пп.47-63, где указанные комплементарные нуклеотиды tracrRNA и crRNA гибридизуются с образованием от 8 до 15 пар оснований или гибридизуются с образованием от 15 до 18 пар оснований.

121. Одиночная молекула РНК, нацеленная на ДНК, по любому из пп.47-63 и 120, где ДНК-мишень представлена в одноклеточном эукариотическом организме, растительной клетке, клетке беспозвоночного животного или клетке позвоночного животного.

122. Композиция по п.65, где один или несколько рекомбинантных экспрессионных векторов выбраны из группы, состоящей из плазмид, космид, мини-колец, фагов и вирусных векторов.

123. Композиция по п.66, где один или несколько вирусных векторов выбраны из группы, состоя-

шей из вирусных векторов, основанных на вирусе коровьей оспы, полиовирусе, ретровирусе, лентивирусе, аденовирусе, адено-ассоциированном вирусе и вирусе простого герпеса.

124. Композиция по любому одному из пп.64-79 и 122, 123, где указанные комплементарные нуклеотиды tracrRNA и crRNA гибридизуются с образованием от 8 до 15 пар оснований.

125. Композиция по любому одному из пп.64-79 и 122, 123, где указанные комплементарные нуклеотиды tracrRNA и crRNA гибридизуются с образованием от 15 до 18 пар оснований.

126. Композиция по любому одному из пп.64-79 и 122-125, где ДНК-мишень представлена в одно-клеточном эукариотическом организме, растительной клетке, клетке беспозвоночного животного или клетке позвоночного животного.

127. Композиция по любому одному из пп.64-79 и 122-126, где полипептид Cas9 включает одну или несколько мутаций в домене RuvC и/или в домене HNH, которые ведут к сниженной нуклеазной активности по сравнению с соответствующим полипептидом Cas9 дикого типа.

128. Набор по п.84, где рекомбинантный экспрессионный вектор выбран из группы, состоящей из плазмид, космид, мини-колец, фагов и вирусных векторов.

129. Набор по п.85, где вирусный вектор выбран из группы, состоящей из вирусных векторов, основанных на вирусе коровьей оспы, полиовирусе, ретровирусе, лентивирусе, аденовирусе, адено-ассоциированном вирусе и вирусе простого герпеса.

130. Набор по любому одному из пп.80-98 и 128-129, где указанные комплементарные нуклеотиды tracrRNA и crRNA гибридизуются с образованием от 8 до 15 пар оснований.

131. Набор по любому одному из пп.80-98 и 128-129, где указанные комплементарные нуклеотиды tracrRNA и crRNA гибридизуются с образованием от 15 до 18 пар оснований.

132. Набор по любому одному из пп.80-98 и 128-131, где ДНК-мишень представлена в одноклеточном эукариотическом организме, растительной клетке, клетке беспозвоночного животного или клетке позвоночного животного.

133. Набор по любому одному из пп.80-98 и 128-132, где полипептид Cas9 включает одну или несколько мутаций в домене RuvC и/или в домене HNH, которые ведут к сниженной нуклеазной активности по сравнению с соответствующим полипептидом Cas9 дикого типа.

134. Генетически модифицированная эукариотическая клетка по любому из пп.99-108, где указанные комплементарные нуклеотиды tracrRNA и crRNA гибридизуются с образованием от 8 до 15 пар оснований.

135. Генетически модифицированная эукариотическая клетка по любому из пп.99-108, где указанные комплементарные нуклеотиды tracrRNA и crRNA гибридизуются с образованием от 15 до 18 пар оснований.

136. Генетически модифицированная эукариотическая клетка по любому из пп.99-108 и 134-135, где полипептид Cas9 включает одну или несколько мутаций в домене RuvC и/или в домене HNH, которые ведут к сниженной нуклеазной активности по сравнению с соответствующим полипептидом Cas9 дикого типа.

137. Применение композиции по любому из пп.23-46 и 115-119 для нокаута гена, нокина гена, редактирования гена или транспозонового мутагенеза гена.

138. Применение одноцепочечной РНК, нацеленной на ДНК, по любому одному из пп.47-63 и 120-121 для нокаута гена, нокина гена, редактирования гена или транспозонового мутагенеза гена.

139. Применение композиции по любому одному из пп.64-79 и 122-127 для нокаута гена, нокина гена, редактирования гена или транспозонового мутагенеза гена.

140. Применение набора по любому одному из пп.80-98 и 128-133 для нокаута гена, нокина гена, редактирования гена или транспозонового мутагенеза гена.

141. Применение генетически модифицированной эукариотической клетки по любому одному из пп.99-108 и 134-136 для нокаута гена, нокина гена, редактирования гена или транспозонового мутагенеза гена.

142. Композиция для направления полипептида Cas9 на последовательность-мишень в ДНК-мишени, приводящего к связыванию ДНК-мишени, или связыванию и модификации ДНК-мишени, или связыванию ДНК-мишени и модуляции транскрипции с ДНК-мишени, содержащая:

(а) гибридный полипептид Cas9 или нуклеиновую кислоту, кодирующую указанный гибридный полипептид Cas9, где указанный гибридный полипептид Cas9 содержит модифицированный белок Cas9, имеющий одну или несколько мутаций в домене RuvC и/или в домене HNH, которые вызывают уменьшенную нуклеазную активность по сравнению с соответствующим диким типом Cas9, и содержит гетерологичный полипептид; и

б) РНК, нацеленную на ДНК, или нуклеиновую кислоту, кодирующую указанную РНК, нацеленную на ДНК, причем указанная РНК, нацеленная на ДНК, содержит:

(i) нацеливающий на ДНК сегмент, содержащий нуклеотидную последовательность, которая является комплементарной последовательности-мишени в ДНК-мишени; и

(ii) белок-связывающий сегмент, который взаимодействует с указанным гибридным полипептидом Cas9, где белок-связывающий сегмент содержит два комплементарных участка нуклеотидов, которые

гибридизуются друг с другом, для того, чтобы сформировать дуплекс двухцепочечной РНК (дцРНК), где указанный дуплекс дцРНК содержит комплементарные нуклеотиды tracrRNA и РНК CRISPR (crRNA).

143. Композиция по п.142, где указанный гетерологичный полипептид представляет собой белковую метку.

144. Композиция по п.127, где белковая метка выбрана из группы, состоящей из метки bHis, геммагглютининовой метки и зеленого флуоресцентного белка (GFP).

145. Композиция по п.142, где указанный гетерологичный полипептид:

(i) имеет ДНК-модифицирующую активность, или

(ii) демонстрирует способность по увеличению или уменьшению транскрипции, или

(iii) имеет ферментативную активность, которая модифицирует полипептид, связанный с ДНК.

146. Композиция по п.145, где указанный гетерологичный полипептид:

(i) имеет ДНК-модифицирующую активность, выбранную из активности метилтрансферазы, активности деметилазы, активности репарации ДНК, активности повреждения ДНК, активности дезаминирования, активности дисмутаза, алкилирующей активности, активности депуринизации, активности окисления, активности формирования димера пиримидина, активности интегразы, активности транспозазы, активности рекомбиназы, полимеразной активности, лигазной активности, геликазной активности, активности фототиазы и гликозилазной активности; или

(ii) является полипептидом, активирующим транскрипцию, или полипептидом, подавляющим транскрипцию; или

(iii) имеет ферментативную активность, которая модифицирует полипептид, связанный с ДНК, которая выбрана из гистон-модифицирующей активности, активности метилтрансферазы, диметилазной активности, активности ацетилтрансферазы, деацетилазной активности, активности киназы, активности фосфатазы, убиквитинлигазной активности, деубиквитинизирующей активности, активности аденилирования, деаденилирующей активности, активности сумоилирования, активности десумоилирования, активности рибозилирования, активности дерибозилирования, активности миристоилирования, активности демиристоилирования, гликозилирующей активности и дегликозилирующей активности.

147. Композиция по любому одному из пп.142-146, где домен трансдукции белка ковалентно соединен с карбоксильным концом или аминоконцом гибридного полипептида Cas9, где указанный домен трансдукции белка способствует проникновению гибридного полипептида Cas9 из цитозоля в клеточную органеллу.

148. Композиция по любому одному из пп.142-147, где нуклеиновая кислота, кодирующая гибридный полипептид Cas9, включает нуклеотидную последовательность, которая (i) кодирует гибридный полипептид Cas9 и (ii) функционально соединена с промотором.

149. Композиция по п.148, где промотор функционален в эукариотической клетке.

150. Композиция по любому одному из пп.142-149, где указанные комплементарные нуклеотиды tracrRNA и crRNA гибридизуются с образованием от 8 до 15 пар оснований.

151. Композиция по любому одному из пп.142-149, где указанные комплементарные нуклеотиды tracrRNA и crRNA гибридизуются с образованием от 15 до 18 пар оснований.

152. Композиция по любому одному из пп.142-151, в которой РНК, нацеленная на ДНК, представляет собой двумолекулярную РНК, нацеленную на ДНК, и содержит две отдельные молекулы РНК, каждая из которых содержит один из двух участков нуклеотидов, которые гибридизуются с образованием дуплекса дцРНК.

153. Композиция по любому одному из пп.142-152, где ДНК-мишень представлена в одноклеточном эукариотическом организме, растительной клетке, клетке беспозвоночного животного или клетке позвоночного животного.

154. Композиция по любому одному из пп.142-153, где гибридный полипептид Cas9 включает одну или несколько мутаций в домене RuvC и/или в домене HNH, которые ведут к сниженной нуклеазной активности по сравнению с соответствующим полипептидом Cas9 дикого типа.

155. Генетически модифицированная эукариотическая клетка, включающая композицию по любому одному из пп.142-154.

156. Способ направления полипептида Cas9 на последовательность-мишень в ДНК-мишени, приводящего к связыванию ДНК-мишени, или связыванию и модификации ДНК-мишени, или связыванию ДНК-мишени и модуляции транскрипции с ДНК-мишени, включающий приведение ДНК-мишени в контакт с комплексом, включающим:

(a) гибридный полипептид Cas9, который содержит модифицированный белок Cas9, имеющий одну или несколько мутаций в домене RuvC и/или в домене HNH, которые вызывают уменьшенную нуклеазную активность по сравнению с соответствующим диким типом Cas9, и содержит гетерологичный полипептид, и

(b) РНК, нацеленную на ДНК, содержащую:

(i) нацеливающий на ДНК сегмент, содержащий нуклеотидную последовательность, которая является комплементарной последовательности-мишени в ДНК-мишени; и

(ii) белок-связывающий сегмент, который взаимодействует с указанным гибридным полипептидом

Cas9, где белок-связывающий сегмент содержит два комплементарных участка нуклеотидов, которые гибридизуются друг с другом, для того, чтобы сформировать дуплекс двухцепочечной РНК (дцРНК), где указанный дуплекс дцРНК содержит комплементарные нуклеотиды tracrRNA и РНК CRISPR (crRNA).

157. Способ по п.156, где гетерологичный полипептид представляет собой белковую метку.

158. Способ по п.157, где белковая метка выбрана из группы, состоящей из метки 6xHis, геммагглютининовой метки и зеленого флуоресцентного белка (GFP).

159. Способ по п.156, где гетерологичный полипептид:

- (i) имеет ДНК-модифицирующую активность, или
- (ii) демонстрирует способность по увеличению или уменьшению транскрипции, или
- (iii) имеет ферментативную активность, которая модифицирует полипептид, связанный с ДНК.

160. Способ по п.159, где указанный гетерологичный полипептид:

(i) имеет ДНК-модифицирующую активность, выбранную из активности метилтрансферазы, активности деметилазы, активности репарации ДНК, активности повреждения ДНК, активности дезаминирования, активности дисмутазы, алкилирующей активности, активности депуринизации, активности окисления, активности формирования димера пиримидина, активности интегразы, активности транспозазы, активности рекомбиназы, полимеразной активности, лигазной активности, геликазной активности, активности фототилазы и гликозилазной активности; или

(ii) является полипептидом, активирующим транскрипцию, или полипептидом, подавляющим транскрипцию; или

(iii) имеет ферментативную активность, которая модифицирует полипептид, связанный с ДНК, которая выбрана из гистон-модифицирующей активности, активности метилтрансферазы, диметилазной активности, активности ацетилтрансферазы, деацетилазной активности, активности киназы, активности фосфатазы, убиквитинлигазной активности, деубиквитинизирующей активности, активности аденилирования, деаденилирующей активности, активности сумоилирования, активности десумоилирования, активности рибозилирования, активности дерибозилирования, активности миристоилирования, активности демиростоилирования, гликозилирующей активности и дегликозилирующей активности.

161. Способ по любому из пп.156-160, где домен трансдукции белка ковалентно соединен с карбоксильным концом или аминоконцом гибридного полипептида Cas9, где указанный домен трансдукции белка способствует проникновению гибридного полипептида Cas9 из цитозоля в клеточную органеллу.

162. Способ по любому из пп.156-161, где нуклеиновая кислота, кодирующая гибридный полипептид Cas9, включает нуклеотидную последовательность, которая (i) кодирует гибридный полипептид Cas9 и (ii) функционально соединена с промотором.

163. Способ по п.162, где промотор функционален в эукариотической клетке.

164. Способ по любому из пп.156-163, где указанные комплементарные нуклеотиды tracrRNA и crRNA гибридизуются с образованием от 8 до 15 пар оснований.

165. Способ по любому из пп.156-163, где указанные комплементарные нуклеотиды tracrRNA и crRNA гибридизуются с образованием от 15 до 18 пар оснований.

166. Способ по любому из пп.156-165, в котором РНК, нацеленная на ДНК, представляет собой двумолекулярную РНК, нацеленную на ДНК, и содержит две отдельные молекулы РНК, каждая из которых содержит один из двух участков нуклеотидов, которые гибридизуются с образованием дуплекса дцРНК.

167. Способ по любому из пп.156-166, где ДНК-мишень представлена в одноклеточном эукариотическом организме, растительной клетке, клетке беспозвоночного животного или клетке позвоночного животного.

168. Способ по любому из пп.156-167, где гибридный полипептид Cas9 включает одну или несколько мутаций в домене RuvC и/или в домене HNH, которые ведут к сниженной нуклеазной активности по сравнению с соответствующим полипептидом Cas9 дикого типа.

169. Способ по любому из пп.156-168, где приведение в контакт включает введение в клетку по меньшей мере:

(a) указанного гибридного полипептида Cas9 или нуклеиновой кислоты, кодирующей указанный гибридный полипептид Cas9, и

(b) указанной РНК, нацеленной на ДНК, или нуклеиновой кислоты, кодирующей указанную РНК, нацеленную на ДНК.

170. Способ по п.169, где способ далее включает введение донорного полинуклеотида в клетку, и последовательность донорного полинуклеотида включается в ДНК-мишень.

171. Нуклеиновая кислота, кодирующая одиночную молекулу ДНК-нацеливающую РНК для направления полипептида Cas9 к последовательности-мишени в ДНК-мишени, приводящего к связыванию ДНК-мишени, или к связыванию и модификации ДНК-мишени, или к связыванию и модуляции транскрипции с ДНК-мишени, при этом одиночная молекула РНК, нацеленная на ДНК, включает:

(i) нацеливающий на ДНК сегмент, содержащий нуклеотидную последовательность, которая является комплементарной последовательности-мишени в ДНК-мишени; и

(ii) белок-связывающий сегмент, который взаимодействует с указанным полипептидом Cas9, где

белок-связывающий сегмент содержит два комплементарных участка нуклеотидов, которые гибридизуются друг с другом, для того, чтобы сформировать дуплекс двухцепочечной РНК (дцРНК), где указанный дуплекс дцРНК содержит комплементарные нуклеотиды tracrРНК и РНК CRISPR (crРНК) и где указанных два комплементарных участка нуклеотидов ковалентно связаны посредством промежуточных нуклеотидов.

172. Нуклеиновая кислота по п.171, отличающаяся тем, что процент комплементарности между нуклеотидами, которые гибридизуются с образованием дуплекса дцДНК белок-связывающего сегмента, превышает 70%.

173. Нуклеиновая кислота по п.171 или 172, отличающаяся тем, что ДНК-мишень присутствует в бактериальной клетке, архейной клетке, одноклеточном эукариотическом организме, растительной клетке, клетке беспозвоночного животного или клетке из позвоночного животного.

174. Нуклеиновая кислота по п.171 или 172, отличающаяся тем, что ДНК-мишень присутствует в одноклеточном эукариотическом организме, растительной клетке, клетке беспозвоночного животного или клетке позвоночного животного.

175. Нуклеиновая кислота по любому из пп.171-174, отличающаяся тем, что ДНК-мишень представляет собой хромосомную ДНК.

176. Нуклеиновая кислота по любому из пп.171-175, где нуклеиновая кислота представляет собой рекомбинантный экспрессионный вектор, выбранный из группы, состоящей из плазмид, космид, мини-колец, фаговых и вирусных векторов.

177. Нуклеиновая кислота по любому из пп.171-175, отличающаяся тем, что нуклеиновая кислота представляет собой вирусный вектор, выбранный из группы, состоящей из вирусных векторов, основанных на вирусе коровьей оспы, полиовирусе, ретровирусе, лентивирусе, аденовирусе, аденоассоциированном вирусе и вирусе простого герпеса.

178. Нуклеиновая кислота по любому из пп.171-177, где указанные комплементарные нуклеотиды tracrРНК и crРНК гибридизуются с образованием от 8 до 15 пар оснований.

179. Нуклеиновая кислота по любому из пп.171-177, где указанные комплементарные нуклеотиды tracrРНК и crРНК гибридизуются с образованием от 15 до 18 пар оснований.

180. Нуклеиновая кислота по любому из пп.171-179, где указанные промежуточные нуклеотиды составляют от 3 до 5 нуклеотидов.

181. Нуклеиновая кислота по любому из пп.171-180, отличающаяся тем, что указанная нуклеиновая кислота предназначена для направления полипептида Cas9 к последовательности-мишени в ДНК-мишени, приводя к связыванию и модификации ДНК-мишени.

182. Нуклеиновая кислота по п.181, где указанная модификация представляет собой расщепление ДНК-мишени.

183. Нуклеиновая кислота по любому из пп.171-182, отличающаяся тем, что нуклеотидная последовательность, комплементарная последовательности-мишени в ДНК-мишени, имеет длину от 18 до 25 нуклеотидов.

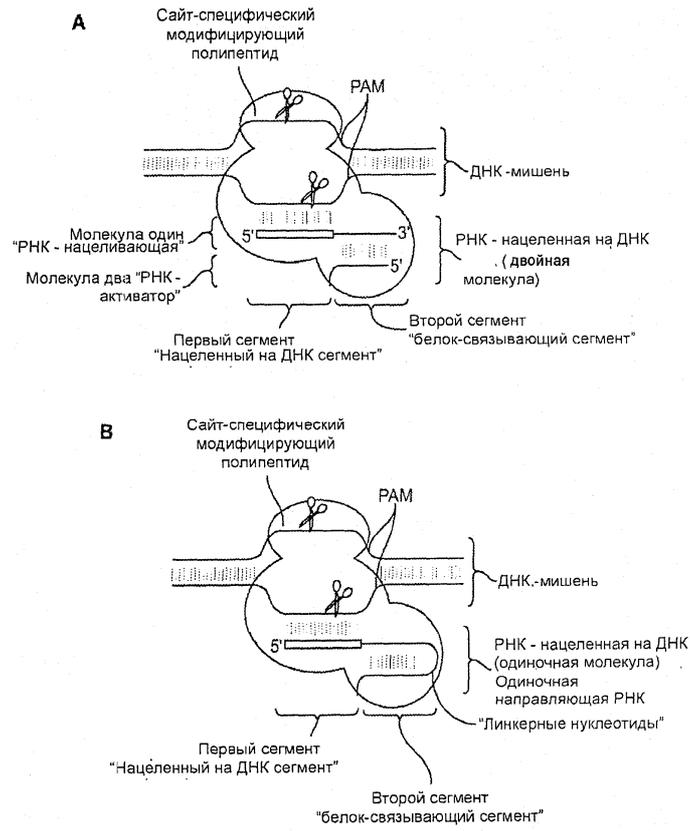
184. Нуклеиновая кислота по любому из пп.171-183, где указанная нуклеиновая кислота лиофилирована.

185. Нуклеиновая кислота по любому из пп.171-183, где указанная нуклеиновая кислота присутствует в композиции, содержащей одно или несколько из следующего: ингибитор нуклеазы, буферный агент, детергент, полиамин, адъювант, смачивающий агент, стабилизирующий агент, антиоксидант и комплексообразователь.

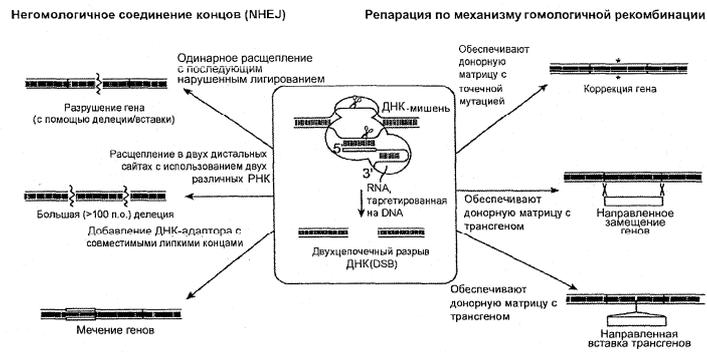
186. Нуклеиновая кислота по любому из пп.171-183, где указанная нуклеиновая кислота присутствует в композиции, содержащей ингибитор нуклеазы.

187. Нуклеиновая кислота по любому из пп.171-183, где указанная нуклеиновая кислота присутствует в композиции, содержащей буферный агент.

188. Применение нуклеиновой кислоты по любому из пп.135-143 для нокаута гена, нокина гена, редактирования гена или мечения гена.



Фиг. 1



Фиг. 2

**Cas9/Csn1 Streptococcus pyogenes**

Мотивы

1 MDKKYSIGLDIGTNSVGWAVITDDYKVPSSKLLKGLGNTDRHGIIKKNLIGALL  
 FDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLLEE  
 SFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLADSTDKVDLRLIYL  
 ALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASRVDA  
 KAILSARLSKSRRLLENLIAQLPGEKKNLFGNLIASLGLTPNFKSNFDLAED  
 AKLQLSKDYYDDDLNLAQIGDQYADLFLAAKNLSDATLLSDILRVNSEITK  
 APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGG  
 ASQEEFYKFIKPILEKMDGTEELLAKLNRDLLRKQRTFDNGSIPYQIHLGEL  
 HAILRRQEDFYFLKDNREKIEKILTRIPYVYVGLARGNSRFAMWTRKSEE  
 TITPWNFEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSHLLYEYFTVYN  
 ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIEC  
 FDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIVLTLTFED  
 REMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI  
 LDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDLSHEHIANLAGS  
 2 PAIKKILQTVKVVDELVKVMGRHKPENIVEMARENQTTQKGQKNSRERM  
 KRIEEGKELGSDILKEYPVENTQLQNEKLYLYLQNGRDMYVDQELDINRL  
 3 SDYD<sup>\*</sup>V<sup>\*</sup>D<sup>\*</sup>H<sup>\*</sup>I<sup>\*</sup>V<sup>\*</sup>P<sup>\*</sup>Q<sup>\*</sup>S<sup>\*</sup>F<sup>\*</sup>L<sup>\*</sup>K<sup>\*</sup>D<sup>\*</sup>S<sup>\*</sup>I<sup>\*</sup>D<sup>\*</sup>N<sup>\*</sup>K<sup>\*</sup>V<sup>\*</sup>L<sup>\*</sup>T<sup>\*</sup>R<sup>\*</sup>S<sup>\*</sup>D<sup>\*</sup>K<sup>\*</sup>N<sup>\*</sup>R<sup>\*</sup>G<sup>\*</sup>K<sup>\*</sup>S<sup>\*</sup>D<sup>\*</sup>N<sup>\*</sup>V<sup>\*</sup>P<sup>\*</sup>S<sup>\*</sup>E<sup>\*</sup>E<sup>\*</sup>V<sup>\*</sup>V<sup>\*</sup>K<sup>\*</sup>M<sup>\*</sup>K<sup>\*</sup>N<sup>\*</sup>Y<sup>\*</sup>W<sup>\*</sup>  
 RQLLNAKLITQRKFDNLTKAERGGSEL<sup>\*</sup>DKVGF<sup>\*</sup>IKR<sup>\*</sup>QLV<sup>\*</sup>ET<sup>\*</sup>RQ<sup>\*</sup>IT<sup>\*</sup>KH<sup>\*</sup>VA<sup>\*</sup>Q<sup>\*</sup>ILD  
 4 SRMNTKYDENDKLI<sup>\*</sup>REVR<sup>\*</sup>VIT<sup>\*</sup>LK<sup>\*</sup>SLV<sup>\*</sup>SD<sup>\*</sup>FR<sup>\*</sup>K<sup>\*</sup>DF<sup>\*</sup>Q<sup>\*</sup>F<sup>\*</sup>Y<sup>\*</sup>K<sup>\*</sup>VE<sup>\*</sup>IN<sup>\*</sup>NY<sup>\*</sup>HH<sup>\*</sup>AH<sup>\*</sup>DAY<sup>\*</sup>  
 LNAVGTALIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYS  
 NIMNFFKTEITLANGEIRKPLIETNGETGEIVWDKGRDFATVRKVL<sup>\*</sup>SMP<sup>\*</sup>QV<sup>\*</sup>  
 NIVK<sup>\*</sup>TEV<sup>\*</sup>QTGGFSKESILPKRNSDKLIARKKDWDPK<sup>\*</sup>YGG<sup>\*</sup>FDS<sup>\*</sup>P<sup>\*</sup>T<sup>\*</sup>VAYS<sup>\*</sup>VL  
 VVAKVEK<sup>\*</sup>GKSK<sup>\*</sup>LK<sup>\*</sup>SVK<sup>\*</sup>ELLGITIMERS<sup>\*</sup>SFEK<sup>\*</sup>DPID<sup>\*</sup>FLEAK<sup>\*</sup>GYKEVR<sup>\*</sup>KDL<sup>\*</sup>IKL  
 PKYSLFE<sup>\*</sup>ENGR<sup>\*</sup>KRMLASAGELQKGNELALPSKYVNF<sup>\*</sup>FLY<sup>\*</sup>LASH<sup>\*</sup>YEK<sup>\*</sup>LK<sup>\*</sup>GSP  
 EDNEQ<sup>\*</sup>QLFVE<sup>\*</sup>QH<sup>\*</sup>KHYL<sup>\*</sup>DEIIEQISEFSKRVILADANL<sup>\*</sup>DKVL<sup>\*</sup>SAY<sup>\*</sup>NK<sup>\*</sup>HR<sup>\*</sup>DK<sup>\*</sup>PI  
 REQAENI<sup>\*</sup>IHL<sup>\*</sup>FTL<sup>\*</sup>NL<sup>\*</sup>GAPAA<sup>\*</sup>FKY<sup>\*</sup>FD<sup>\*</sup>TT<sup>\*</sup>IDR<sup>\*</sup>KRY<sup>\*</sup>TST<sup>\*</sup>KEV<sup>\*</sup>L<sup>\*</sup>DAT<sup>\*</sup>L<sup>\*</sup>IHQ<sup>\*</sup>SIT<sup>\*</sup>GLY<sup>\*</sup>  
 ETRIDLSQLGGD

**Cas9/Csn1 Streptococcus pyogenes**

Домены

1 MDKKYSIGLDIGTNSVGWAVITDDYKVPSSKLLKGLGNTDRHGIIKKNLIGALL  
 LFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLLEE  
 ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLADSTDKVDLRLI  
 YLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASRV  
 DAKAILSARLSKSRRLLENLIAQLPGEKKNLFGNLIASLGLTPNFKSNFDLA  
 EDAKLQLSKDYYDDDLNLAQIGDQYADLFLAAKNLSDATLLSDILRVNSEI  
 TKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYID  
 GGASQEEFYKFIKPILEKMDGTEELLAKLNRDLLRKQRTFDNGSIPYQIHL  
 GELHAILRRQEDFYFLKDNREKIEKILTRIPYVYVGLARGNSRFAMWTRK  
 SEETITPWNFEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSHLLYEYFTV  
 YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKI  
 ECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIVLTLTF  
 EDREMIERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG  
 KTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDLSHEHIANLA  
 2 GSPAIK<sup>\*</sup>K<sup>\*</sup>G<sup>\*</sup>IL<sup>\*</sup>Q<sup>\*</sup>T<sup>\*</sup>V<sup>\*</sup>K<sup>\*</sup>V<sup>\*</sup>V<sup>\*</sup>DEL<sup>\*</sup>V<sup>\*</sup>K<sup>\*</sup>V<sup>\*</sup>M<sup>\*</sup>GR<sup>\*</sup>H<sup>\*</sup>K<sup>\*</sup>PEN<sup>\*</sup>IVEMARENQTTQKGQKNSR  
 ERMKRIEEGKELGSDILKEYPVENTQLQNEKLYLYLQNGRDMYVDQEL  
 DINRLSDYD<sup>\*</sup>V<sup>\*</sup>D<sup>\*</sup>H<sup>\*</sup>I<sup>\*</sup>V<sup>\*</sup>P<sup>\*</sup>Q<sup>\*</sup>S<sup>\*</sup>F<sup>\*</sup>L<sup>\*</sup>K<sup>\*</sup>D<sup>\*</sup>S<sup>\*</sup>I<sup>\*</sup>D<sup>\*</sup>N<sup>\*</sup>K<sup>\*</sup>V<sup>\*</sup>L<sup>\*</sup>T<sup>\*</sup>R<sup>\*</sup>S<sup>\*</sup>D<sup>\*</sup>K<sup>\*</sup>N<sup>\*</sup>R<sup>\*</sup>G<sup>\*</sup>S<sup>\*</sup>D<sup>\*</sup>N<sup>\*</sup>V<sup>\*</sup>P<sup>\*</sup>S<sup>\*</sup>E<sup>\*</sup>E<sup>\*</sup>V<sup>\*</sup>V<sup>\*</sup>K<sup>\*</sup>  
 MKNYWRQLLNAKLITQRKFDNLTKAERGGSEL<sup>\*</sup>DKVGF<sup>\*</sup>IKR<sup>\*</sup>QLV<sup>\*</sup>ET<sup>\*</sup>RQ<sup>\*</sup>IT<sup>\*</sup>  
 KHVAQILDSRMNTKYDENDKLI<sup>\*</sup>REVR<sup>\*</sup>VIT<sup>\*</sup>LK<sup>\*</sup>SLV<sup>\*</sup>SD<sup>\*</sup>FR<sup>\*</sup>K<sup>\*</sup>DF<sup>\*</sup>Q<sup>\*</sup>F<sup>\*</sup>Y<sup>\*</sup>K<sup>\*</sup>VE<sup>\*</sup>IN  
 NYHHAHDAYLNAVGTALIKKYPKLESEFVYGDYKVDVRKMIKSEQEIG  
 KATAKYFFYSNIMNFFKTEITLANGEIRKPLIETNGETGEIVWDKGRDFATV  
 RKVL<sup>\*</sup>SMP<sup>\*</sup>QV<sup>\*</sup>NIVK<sup>\*</sup>TEV<sup>\*</sup>QTGGFSKESILPKRNSDKLIARKKDWDPK<sup>\*</sup>YGG<sup>\*</sup>  
 DSPTVAYS<sup>\*</sup>VLVAKVEK<sup>\*</sup>GKSK<sup>\*</sup>LK<sup>\*</sup>SVK<sup>\*</sup>ELLGITIMERS<sup>\*</sup>SFEK<sup>\*</sup>DPID<sup>\*</sup>FLEAK<sup>\*</sup>  
 GYKEVR<sup>\*</sup>KDL<sup>\*</sup>IKL<sup>\*</sup>PKYSLFE<sup>\*</sup>ENGR<sup>\*</sup>KRMLASAGELQKGNELALPSKYVNF<sup>\*</sup>FLY<sup>\*</sup>  
 LASH<sup>\*</sup>YEK<sup>\*</sup>LK<sup>\*</sup>GSPEDNEQ<sup>\*</sup>QLFVE<sup>\*</sup>QH<sup>\*</sup>KHYL<sup>\*</sup>DEIIEQISEFSKRVILADANL<sup>\*</sup>DKVL  
 SAYNK<sup>\*</sup>HR<sup>\*</sup>DK<sup>\*</sup>PIREQAENI<sup>\*</sup>IHL<sup>\*</sup>FTL<sup>\*</sup>NL<sup>\*</sup>GAPAA<sup>\*</sup>FKY<sup>\*</sup>FD<sup>\*</sup>TT<sup>\*</sup>IDR<sup>\*</sup>KRY<sup>\*</sup>TST<sup>\*</sup>KEV<sup>\*</sup>L<sup>\*</sup>  
 D<sup>\*</sup>AT<sup>\*</sup>L<sup>\*</sup>IHQ<sup>\*</sup>SIT<sup>\*</sup>GLY<sup>\*</sup>ETRIDLSQLGGD

Фиг. 3

**A.** Идентичности последовательности по отношению к *Pyogenes Cas9/Csn1 S.*

Виды	Стандартная последовательность	Идентичности последовательности – выравнивание MUSCLE		
		Полноразмерный	Домен 1	Домен 2
<i>Streptococcus pyogenes</i> M1 GAS	NP_269215	100,0	100,0	100,0
<i>Streptococcus pyogenes</i> MGAS5005	YP_282132.1	99,9	99,4	100,0
<i>Listeria innocua</i> Clip11262	NP_472073	54,3	60,0	64,9
<i>Pasteurella multocida</i> subsp. <i>multocida</i> str. Pm70	NP_246064.1	19,7	29,0	25,9
<i>Streptococcus thermophilus</i> LMD-9 Csn1-A	YP_820832	59,2	75,6	72,4
<i>Streptococcus thermophilus</i> LMD-9 Csn1-B	YP_820161.1	20,6	27,3	26,8
<i>Neisseria meningitidis</i> Z2491	YP_002342100.1	20,2	33,6	28,1
<i>Streptococcus mutans</i> UA159	NP_721764	64,9	78,1	74,1
<i>Streptococcus gordonii</i> str. Challis substr. CH1	YP_001450662.1	19,8	28,2	27,0
<i>Campylobacter jejuni</i> subsp. <i>jejuni</i> NCTC 11168	YP_002344900.1	19,6	30,3	26,3
<i>Treponema denticola</i> ATCC 35405	NP_970941	32,5	47,3	38,8

**B.** Идентичности последовательности по отношению к *Meningitidis Cas9/Csn1 N.*

Виды	Стандартная последовательность	Идентичности последовательности – выравнивание MUSCLE		
		Полноразмерный	Домен 1	Домен 2
<i>Streptococcus pyogenes</i> M1 GAS	NP_269215	20,2	33,6	28,1
<i>Streptococcus pyogenes</i> MGAS5005	YP_282132.1	20,3	34,5	28,1
<i>Listeria innocua</i> Clip11262	NP_472073	18,8	33,6	25,9
<i>Pasteurella multocida</i> subsp. <i>multocida</i> str. Pm70	NP_246064.1	64,3	72,1	69,0
<i>Streptococcus thermophilus</i> LMD-9 Csn1-A	YP_820832	19,6	35,3	25,9
<i>Streptococcus thermophilus</i> LMD-9 Csn1-B	YP_820161.1	25,8	35,7	35,1
<i>Neisseria meningitidis</i> Z2491	YP_002342100.1	100,0	100,0	100,0
<i>Streptococcus mutans</i> UA159	NP_721764	19,2	36,1	25,5
<i>Streptococcus gordonii</i> str. Challis substr. CH1	YP_001450662.1	25,3	37,5	35,8
<i>Campylobacter jejuni</i> subsp. <i>jejuni</i> NCTC 11168	YP_002344900.1	34,7	45,0	41,0
<i>Treponema denticola</i> ATCC 35405	NP_970941	18,8	31,5	25,5

Фиг. 4

	Мотив 1	Мотив 2	Мотив 4
	*		
<i>S. pyogenes</i>	... IGLDIGTNSVGVAVI ...	... IVIEMARE ...	... HHAHDAYL ...
<i>L. pneumophila</i>	... IGIIDLGGKFTGVCLS ...	... MMQRLAYE ...	... SHAIDATL ...
<i>G. proteobacterium</i>	... IAIIDLAKFTGVVALY ...	... IIEHIARK ...	... SHVVDVAVC ...
<i>L. innocua</i>	... IGLDIGTNSVGVAVL ...	... IVVEMARE ...	... HHAHDAYL ...
<i>L. gasseri</i>	... VGLDVGTNSCGWVAM ...	... IAIEFTRD ...	... HHAIDAYL ...
<i>E. rectale</i>	... LALDIGIASVGVAIL ...	... IVIEMPRD ...	... HHAVDAML ...
<i>S. lugdunensis</i>	... IGLDIGITSVGYGLI ...	... IIELEARE ...	... HHAEDALI ...
<i>M. synoviae</i>	... IGFDLGVASVGSIV ...	... VVIEMARE ...	... HHAVDASI ...
<i>M. mobile</i>	... IGLDLGIASVGVCLIT ...	... IVVEVTRS ...	... HHAEDAYF ...
<i>W. succinogenes</i>	... LGVDLGISSLGWAIV ...	... VHFELARE ...	... HHAVDAIL ...
<i>F. columnare</i>	... IGLDLGTNSIGWAIR ...	... IHIEEMARE ...	... HHTIDAIT ...
<i>F. succinogenes</i>	... IGLDLGTNSIGWAVV ...	... IHLELGRD ...	... HHAMDATV ...
<i>B. fragilis</i>	... IGLDLGTNSIGWALV ...	... IRVELARE ...	... HHAMDALT ...
<i>A. cellulolyticus</i>	... LGVDVGSERSIGLAAY ...	... IVVELARG ...	... HHAVDAYV ...
<i>B. dentium</i>	... IGIIDVGLMSVGLAAI ...	... VQIEHVRE ...	... HHAVDAAV ...

Мотив 3

	Мотив 3
	*
<i>S. pyogenes</i>	... DVDHIVPQSFLKD-----DSIDNKVLRSDKN...
<i>L. pneumophila</i>	... EIDHIYPRSLSKKHFGVIFNSEVNLIYCQQGN...
<i>G. proteobacterium</i>	... EIDHIYPRSLTGRTKKTVFENSEANLIYCQSKGN...
<i>L. innocua</i>	... DIDHIVPQSFTID-----NSIDNLVLTSSAGN...
<i>L. gasseri</i>	... DIDHILPQSFTIKD-----DSLNRVIVKKAQN...
<i>E. rectale</i>	... EIDHIYPRSFISFD-----DARSNKVLVYRSEN...
<i>S. lugdunensis</i>	... EVDHII PRSVSFD-----NSYHNKVLVKQSEN...
<i>M. synoviae</i>	... EIDHVI PYSKASD-----DSWPNKLLVKKSTN...
<i>M. mobile</i>	... DIDHIVPRSFISFD-----DSFNSLVI VNKLDN...
<i>W. succinogenes</i>	... EIDHILPRRSAD-----DSFANKVLCCLARAN...
<i>F. columnare</i>	... DIEHTI PRSISQD-----NSQMNKTLCSLKFN...
<i>F. succinogenes</i>	... EIEHVI PQSLYFD-----DSFNSKVICRAEVN...
<i>B. fragilis</i>	... DIEHIIPQARLFD-----DSFNSKTL EARSVN...
<i>A. cellulolyticus</i>	... ELDHIVPRTDGG-----SNRHENLAITCGACN...
<i>B. dentium</i>	... EMDHIVPRKGVGS-----TNTRVNLAACAACN...

Фиг. 5



Штамм	КОЛИЧЕСТВО CRISPR	Идентификатор <sup>a</sup> CRISPR CASS4	Спаривание оснований ПОВТОР CRISPR:траcrPHK
<i>Streptococcus pyogenes</i> SF370	2	NC_002737_1	5' GUUUUAG--AGCUAUGCUGUUUUGAUGGJCCCAAAC 3'     *                                   3' AAAUUGAACGUAUCGACAAAACUACCAAGGUUGUU 5'
<i>Streptococcus mutans</i> UA159	1	NC_004350_1	GUUUUAG--AGCUGUGUUUUCCAAUGGUCCAAAAC     *                                   AAAUUGAACGACACAACAAGCUUACCAAGGUUGUG
<i>Streptococcus thermophilus</i> LMD-9	3	NC_008532_5	GUUUUAG--AGCUGUGUUUUCCAAUGGUCCAAAAC     *                                   AAAUUGAGCGACACAACAAGCUUACCAAGGUUGUG
		NC_008532_2	GUUUUUGUACUCU--CAAGUUUAAGUAACUGUACAAC     *                                   AAACAUCGAGACGUUCUAAAUUCUUGACACAUC
<i>Listeria innocua</i> Clp11262	1	NC_003212_2	GUUUUAG--AGCUAUGUUAUUUUGAUGGCUAACAAAAC     *                                   AAAUUGAACGUAUCAAAUAAAACUUAUGAUUGUUAUA
<i>Treponema denticola</i> ATCC35405	1	NC_002967_1	GUUUGAG--AGUUGUGUAUUUAAGUAGUUCUCAAAC     *                                   AACUUGAGCAACACAUUAAAUUACCUAGAUUUUA
<i>Neisseria meningitidis</i> Z2491	2	NC_003116_10	GUUUGAGCUCCUUUCUUAUUUCCGAGUGCUUACA *  UAAACUUGUCCCAAGUAUAAAGGUCACCGUGUA
<i>Streptococcus gordonii</i> str. Challis substr. CH1	1	NC_009785_2	GUUUUUGUACUCU--CAAGUUUAAGUAACUGUACAAC     *   AAACAUCGAGACGUUCUAAAUUACUUGACACAUC
<i>Bifidobacterium bifidum</i> S17	1	NC_014616_1	GUUCCA-AUGCCUGUCAGAUCAAUGACUUUGACCAC *   AGUUAUUACGACAGUCCAGUUAUCUGAACUAGUA
<i>Lactobacillus salivarius</i> UCC118	1	NC_007929_1	GUUCAGAAAGUAGUUAAAUCAAUAGGUUAAGACC  *                                       AAGUUGAGUCUACAAUUUAGUUAUCUCCAGUUUUGG CUAACGUAAGUUUACCAAAUUAUUCAGCAACUGAAC    * **                                    UUGUGUUAUGUUAUGGUUAUUAAGAUUGUUGACUUG
<i>Francisella tularensis</i> subsp. novicida U112	2	NC_008601_1 <sup>c</sup>	CUAACGUAAGUUUACCAAAUUAUUA--GCAACUGAAC    *  UUUAAUUAUUAACUAGGUUAUUAUUAUACG-AGACAUUA
<i>Legionella pneumophila</i> str. Paris	1	NC_006368_1	CCAAUUAUCCUCUUAUUAUUAUUAUUAUUAUUAUUA  AUUUAUUCUUUAAGUAGAUUUAAGCUAUGG-GACUUUA

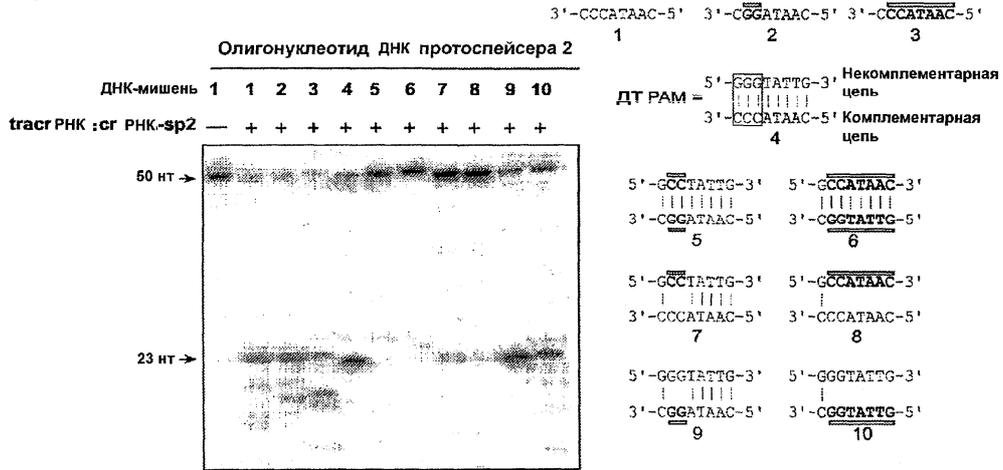
Фиг. 8



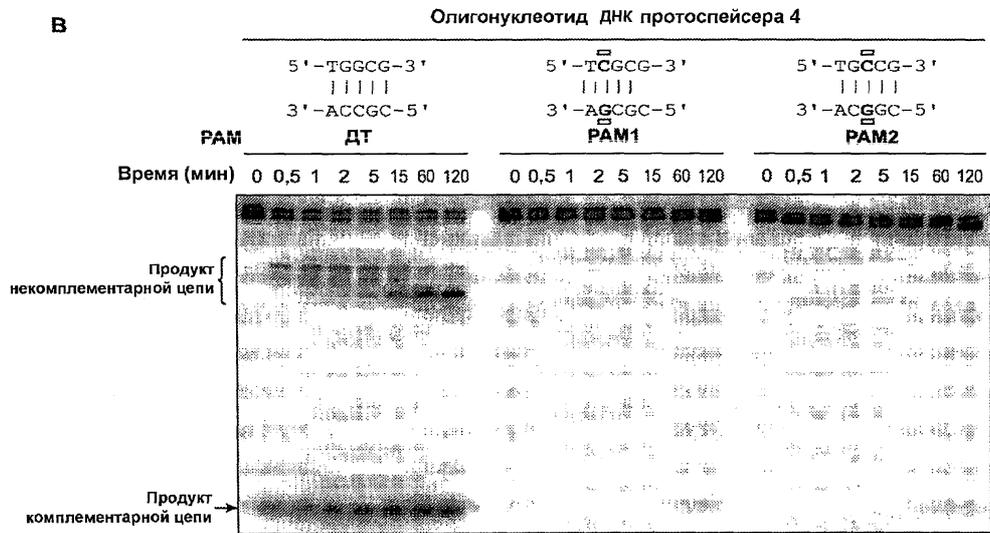




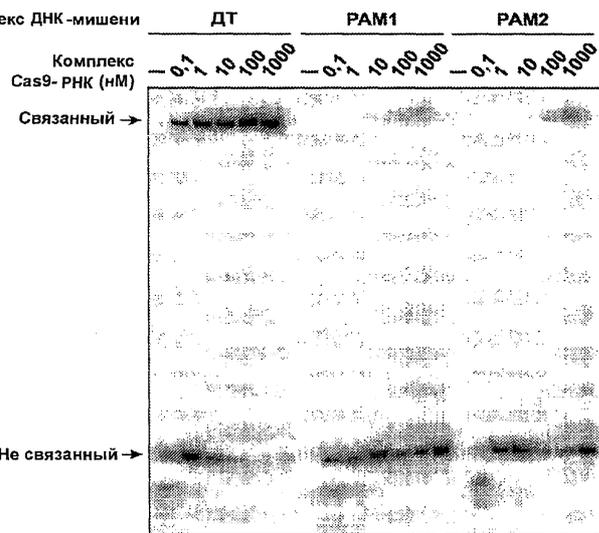
A



B

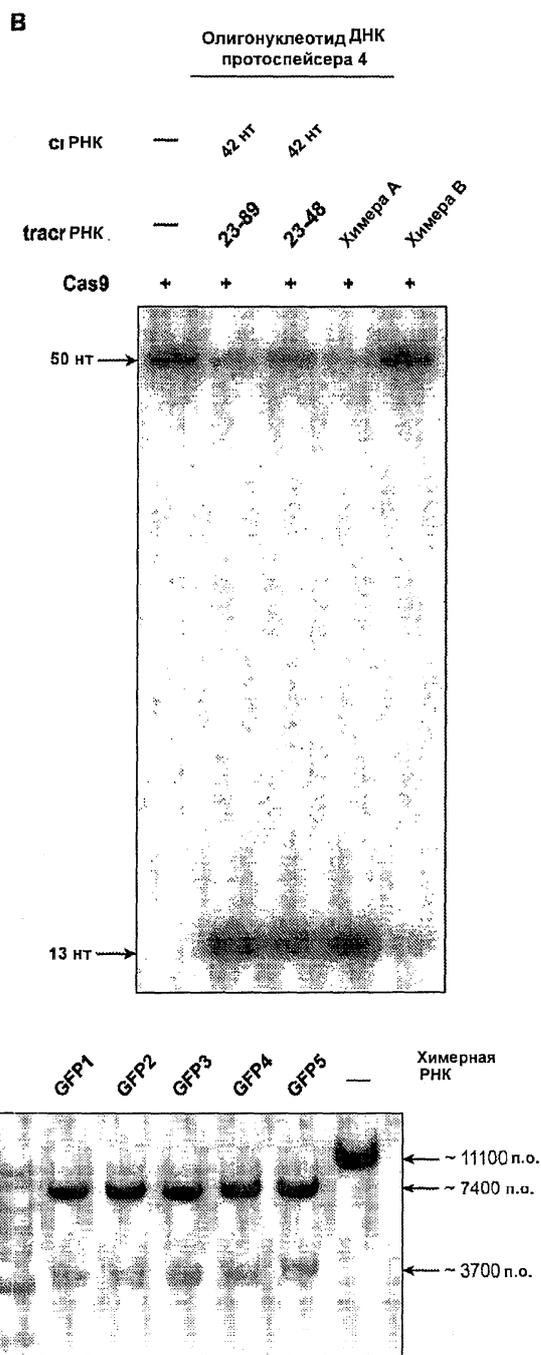


C

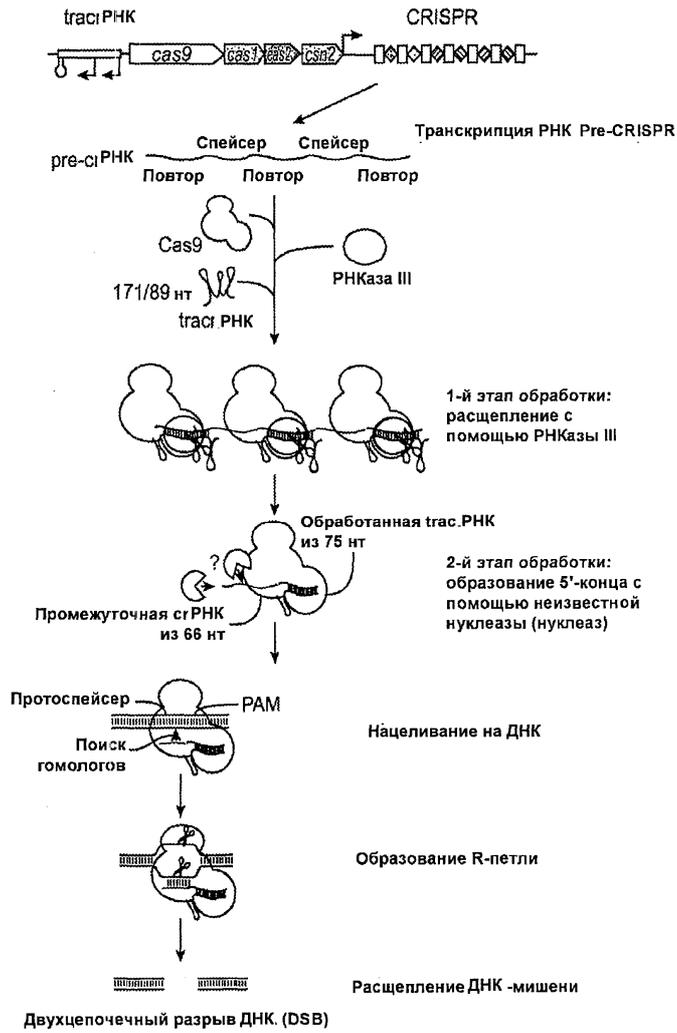


Фиг. 13

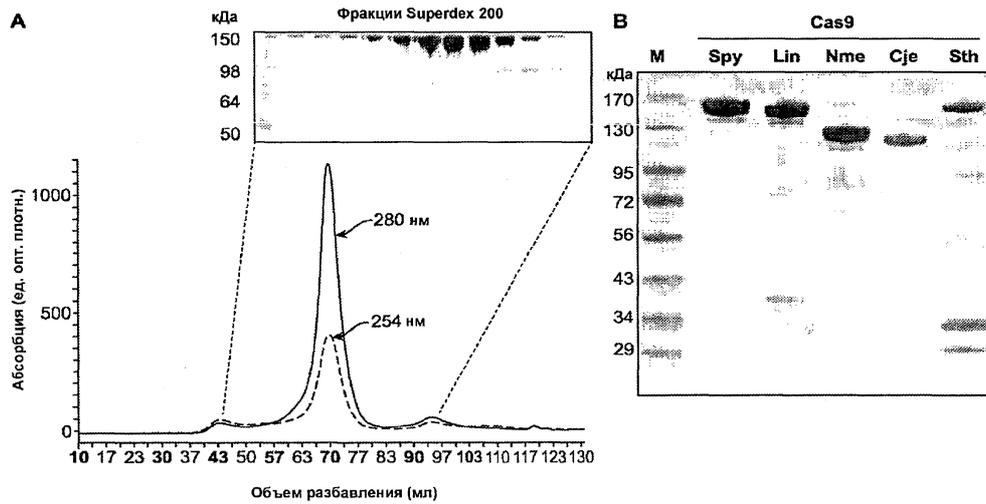




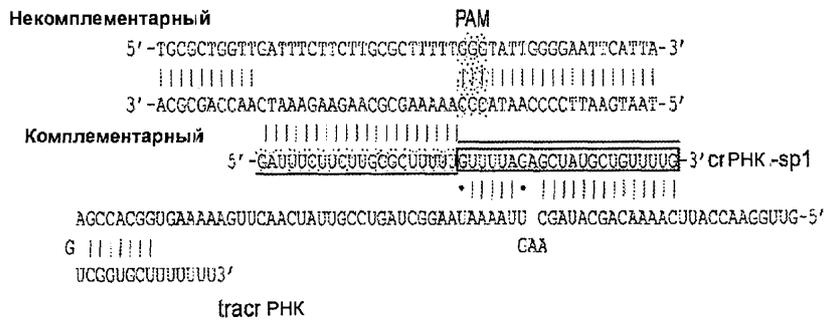
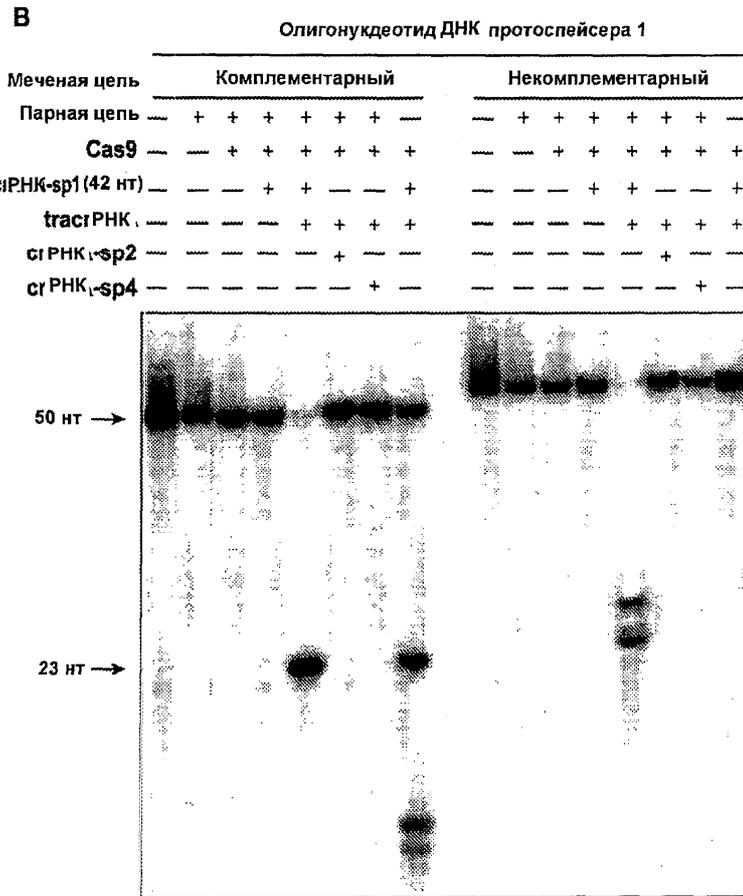
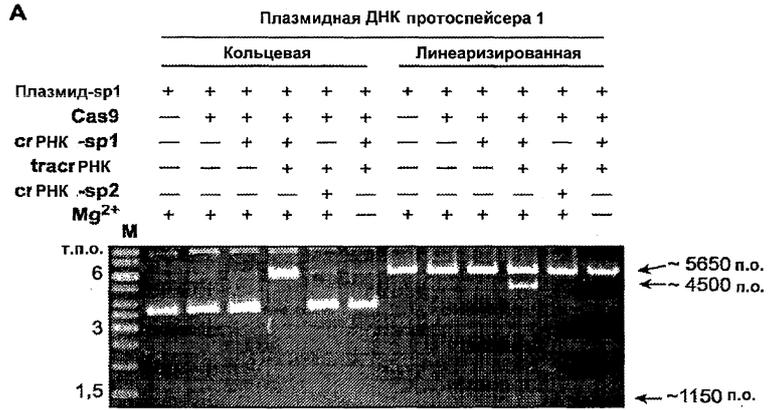
Фиг. 14



Фиг. 15



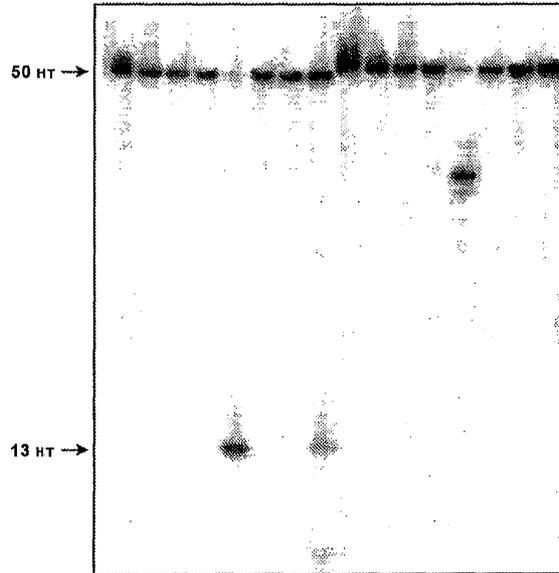
Фиг. 16



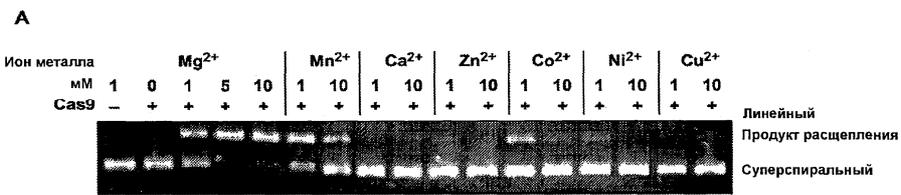
**C**

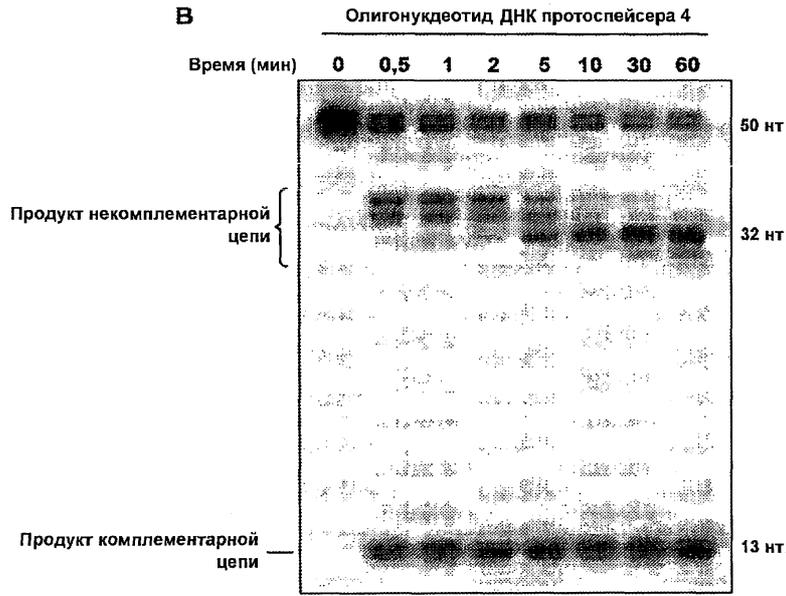
Олигонуклеотид ДНК протоспейсера 4

Меченая цепь	Комплементарный										Некомплементарный														
Парная цепь	-	+	+	+	+	+	+	+	-	-	+	+	+	+	+	+	-	-	+	+	+	+	+	+	-
<b>Cas9</b>	-	+	+	+	+	+	+	+	-	-	+	+	+	+	+	+	-	-	+	+	+	+	+	+	-
сi РНК-sp4 (42 нт)	-	-	+	+	-	-	+	-	-	-	-	+	+	-	-	+	-	-	-	+	+	-	-	+	-
tracrРНК <sub>1</sub>	-	-	-	-	+	+	+	+	-	-	-	-	-	+	+	+	-	-	-	-	+	+	+	+	-
сi РНК <sub>1</sub> -sp2	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-
сi РНК <sub>1</sub> -sp1	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-

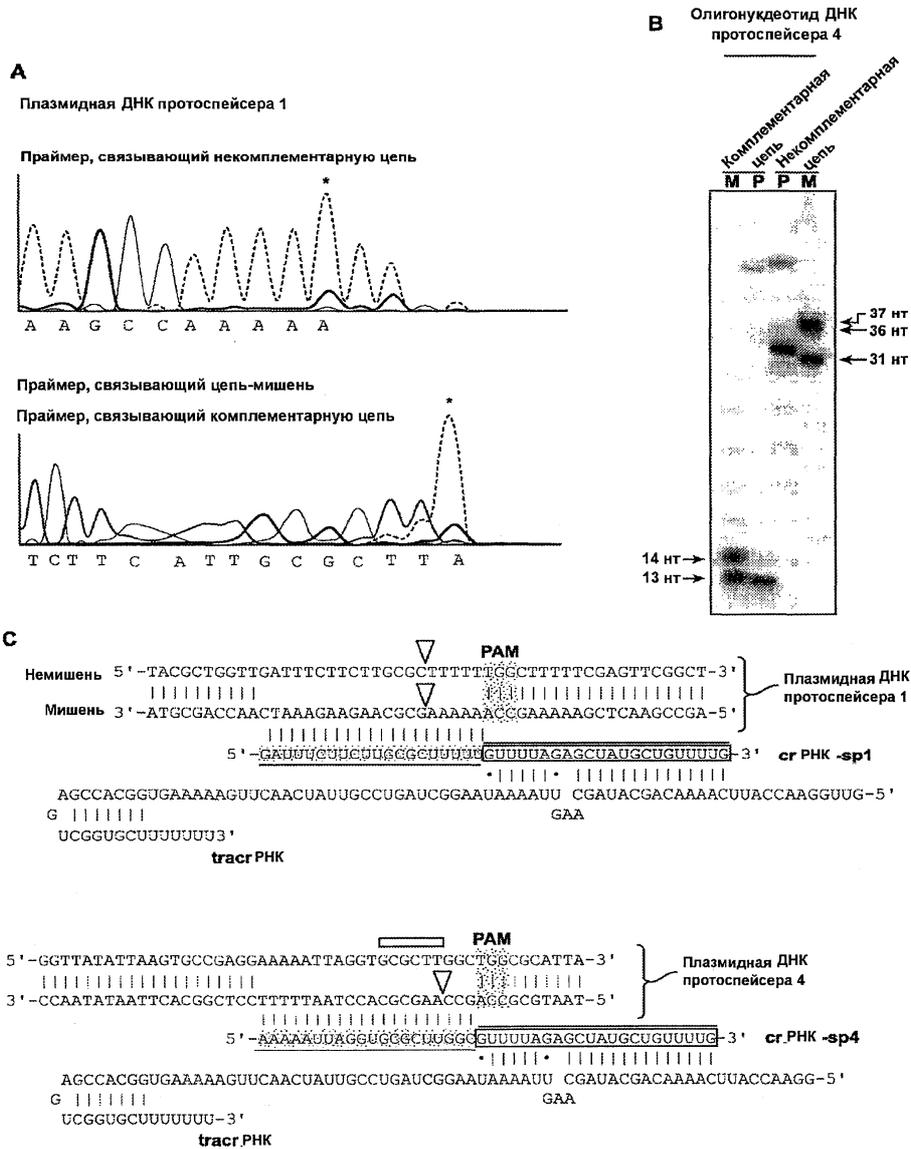


Фиг. 17

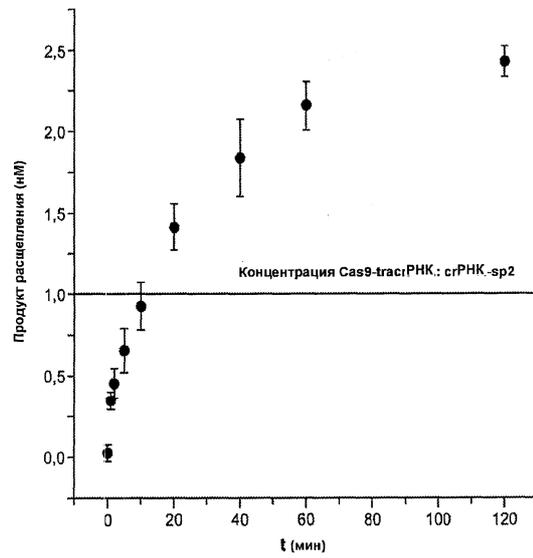
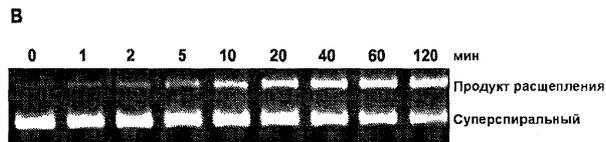
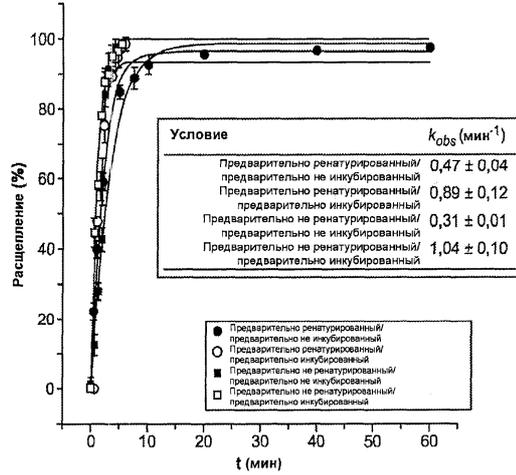
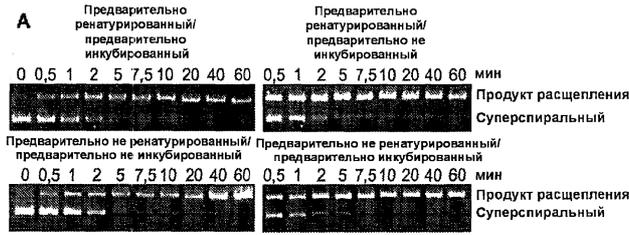




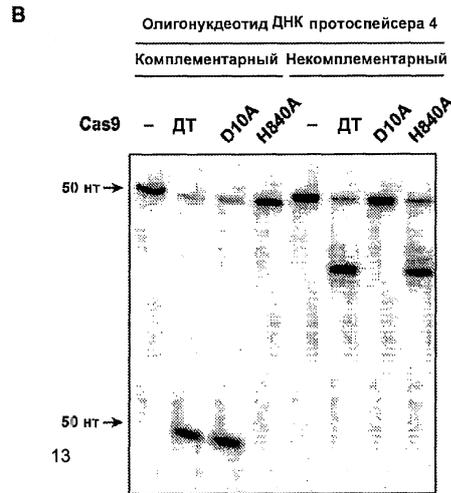
Фиг. 18



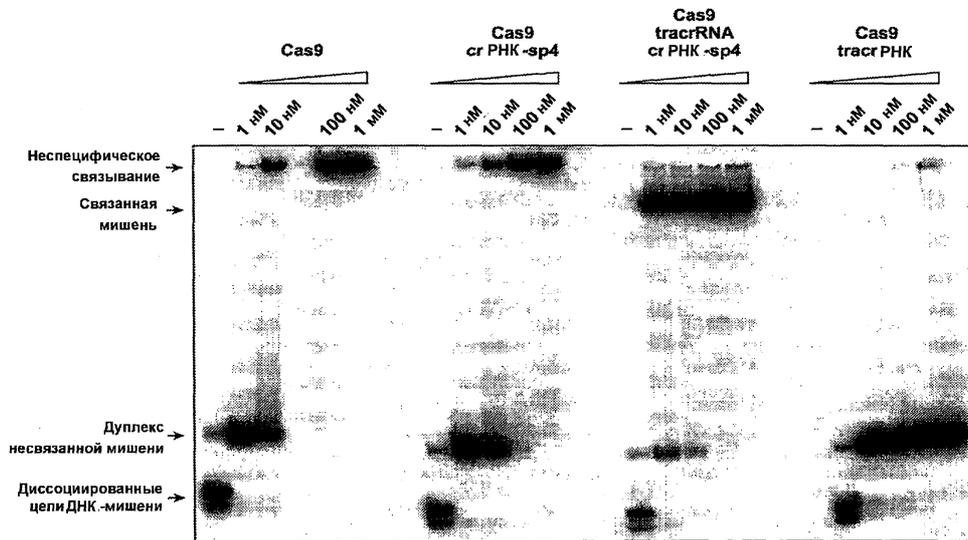
Фиг. 19



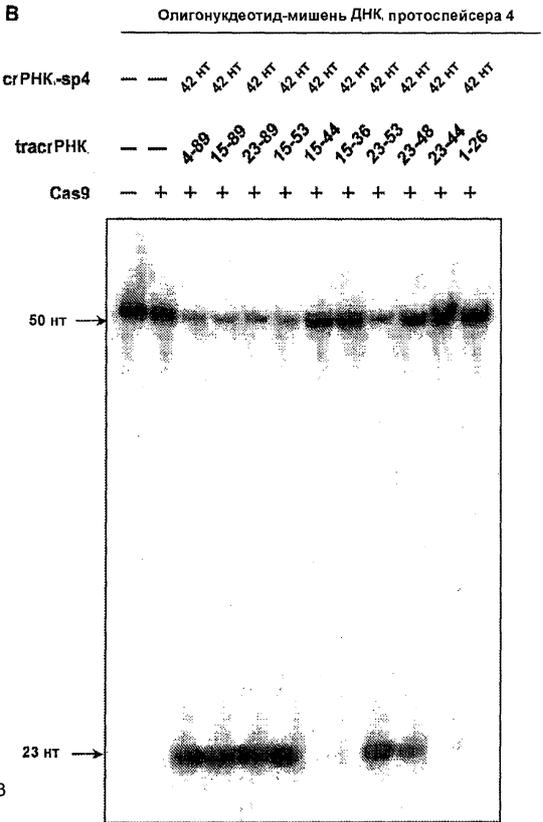
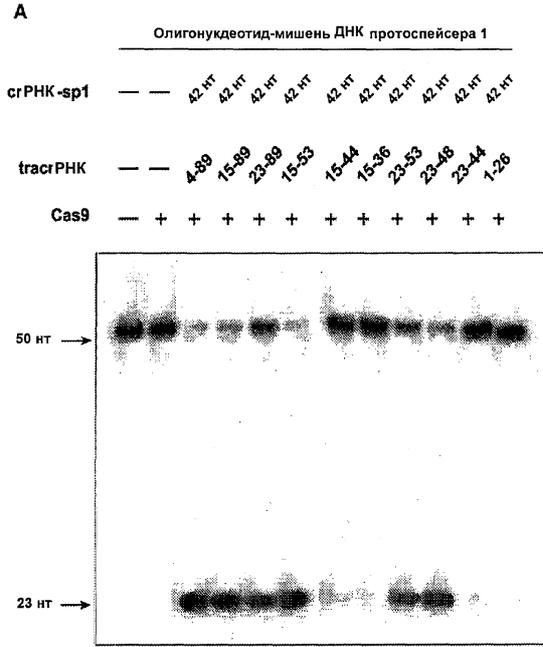
Фиг. 20



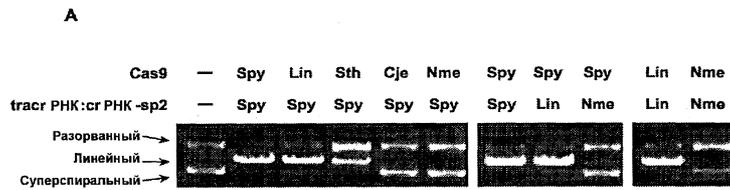
Фиг. 21

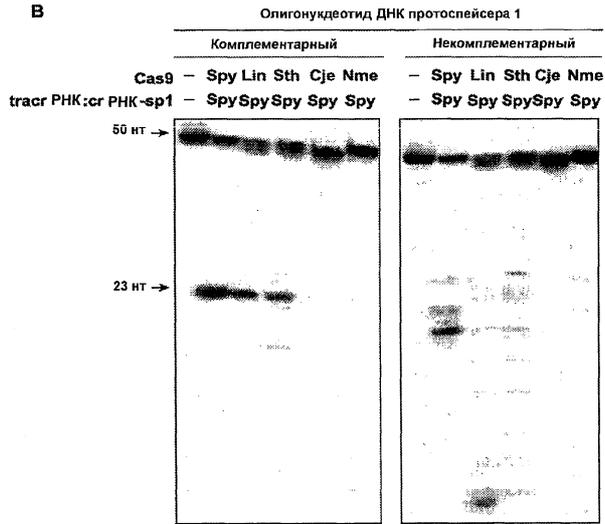


Фиг. 22



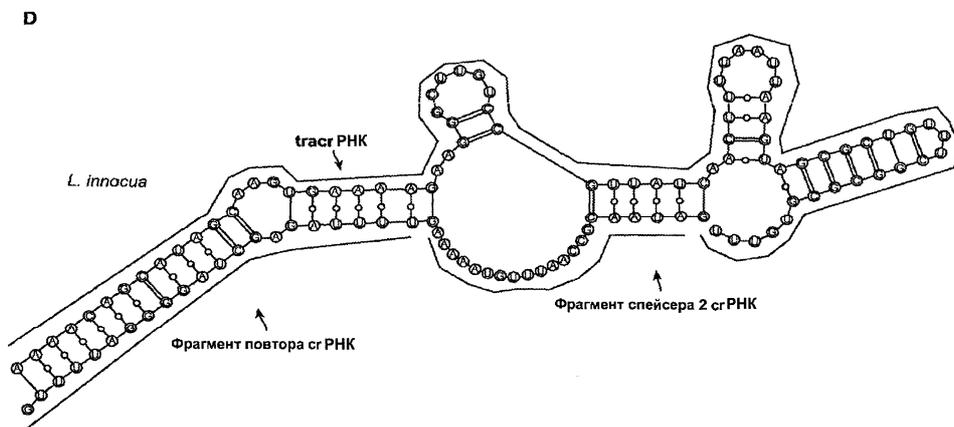
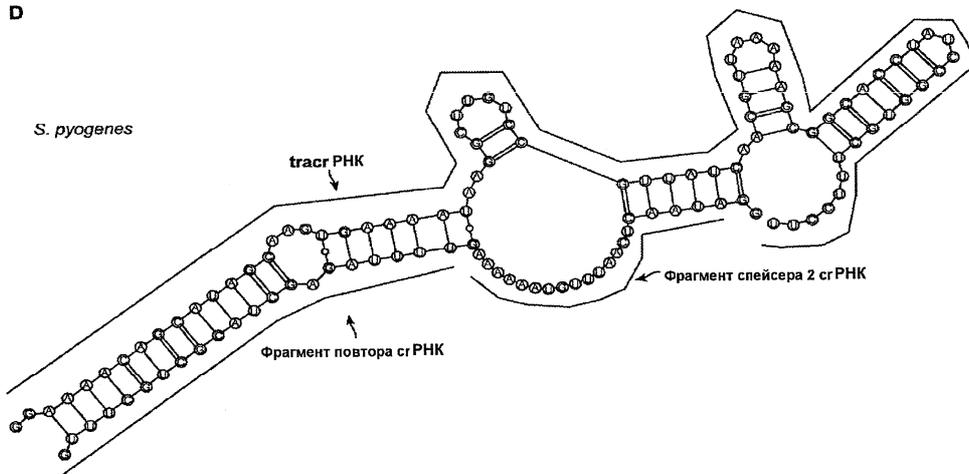
Фиг. 23

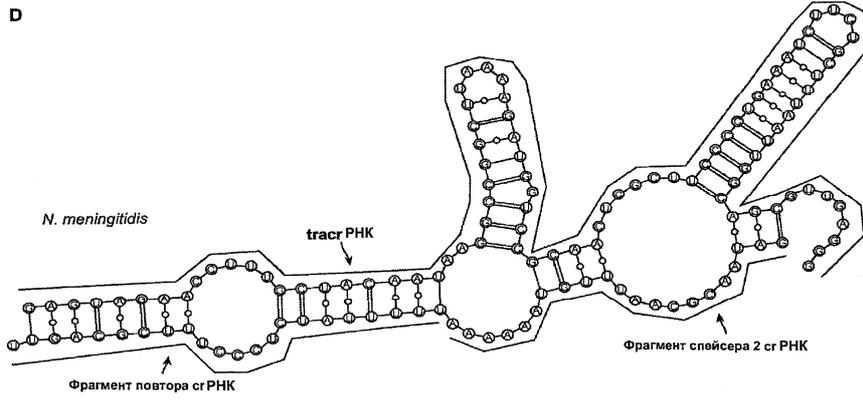




**C**

	<i>S. pyogenes</i>	<i>L. innocua</i>	<i>S. thermophilus</i>	<i>C. jejuni</i>	<i>N. meningitidis</i>
<i>S. pyogenes</i>	x	54	58	16	16
<i>L. innocua</i>	54	x	52	15	13
<i>S. thermophilus</i>	58	52	x	16	15
<i>C. jejuni</i>	16	15	16	x	32
<i>N. meningitidis</i>	16	14	15	32	x

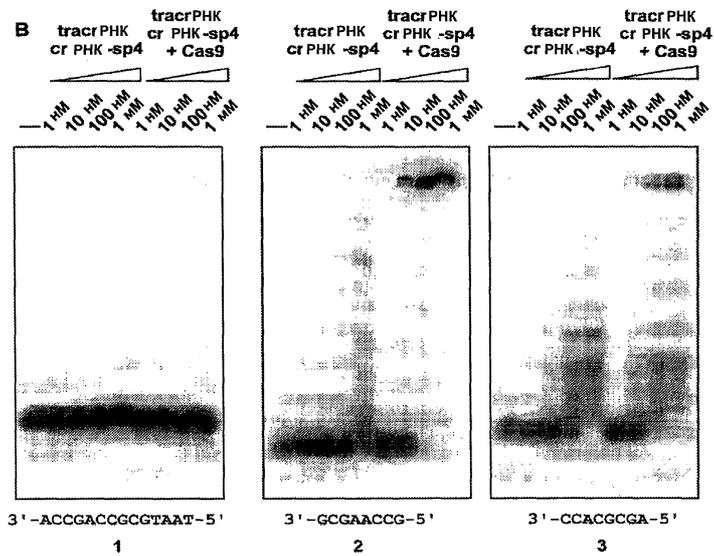
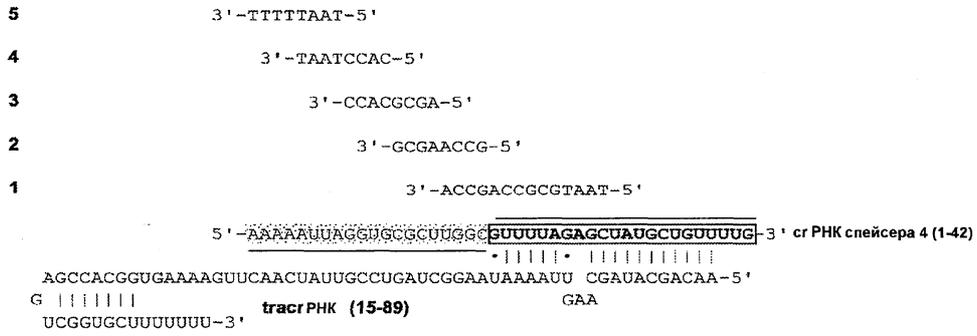


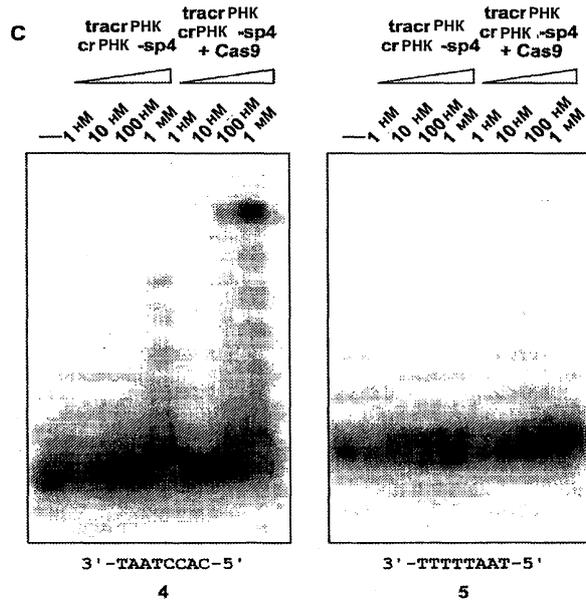


Фиг. 24

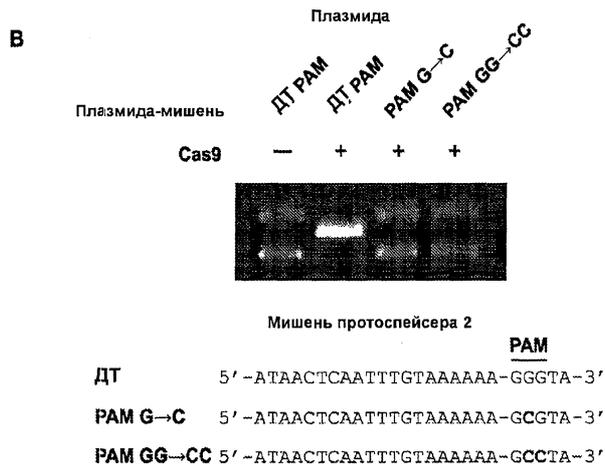
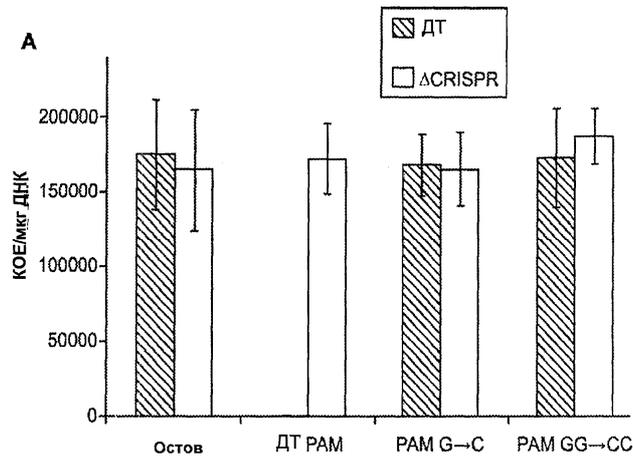
**A**

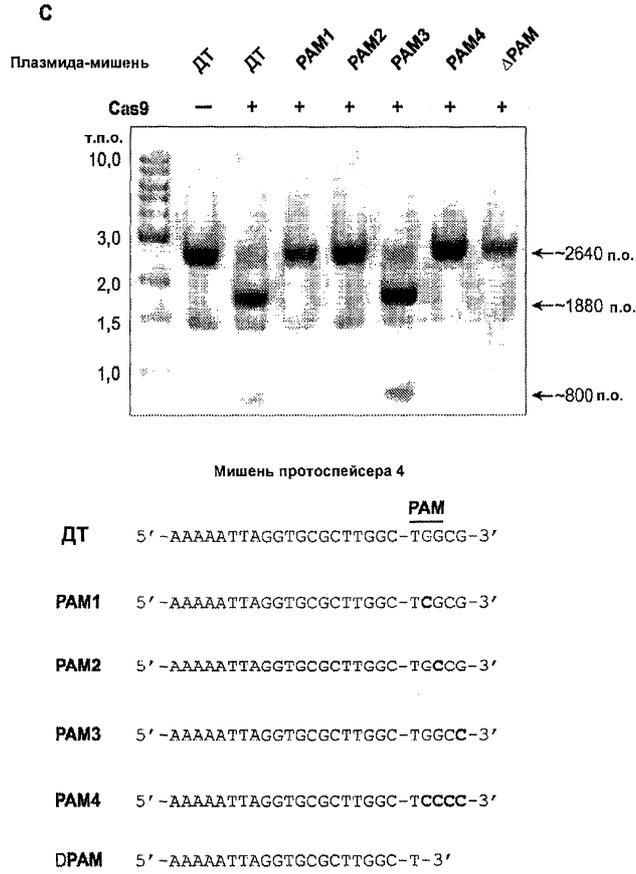
Зона ДНК-мишени



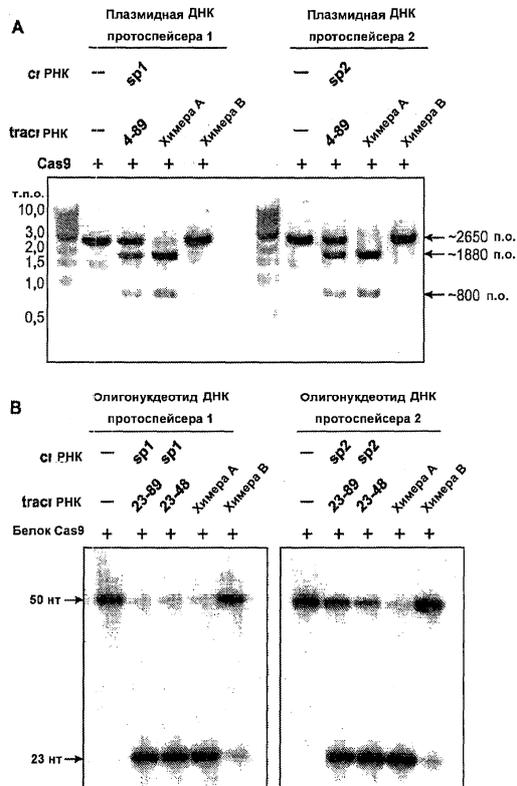


Фиг. 25





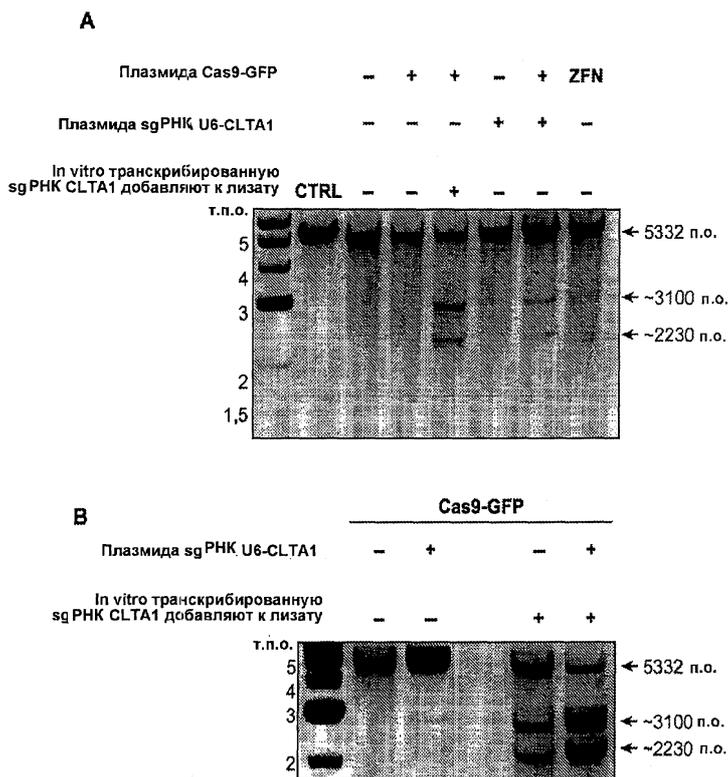
Фиг. 26



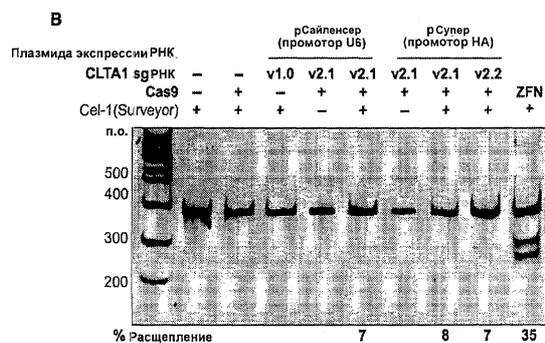
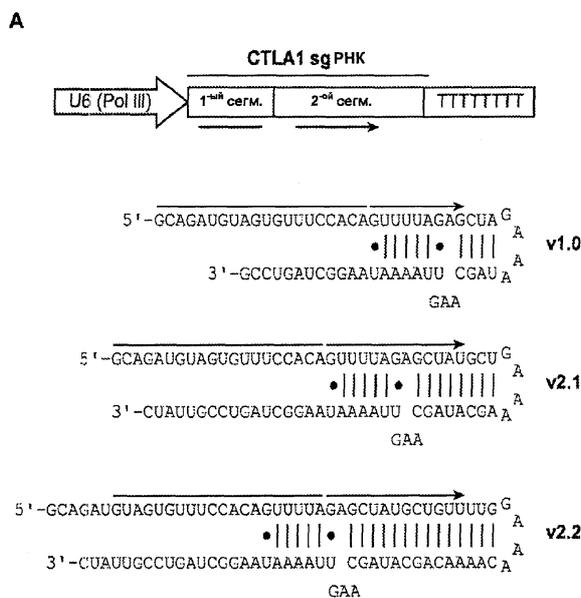




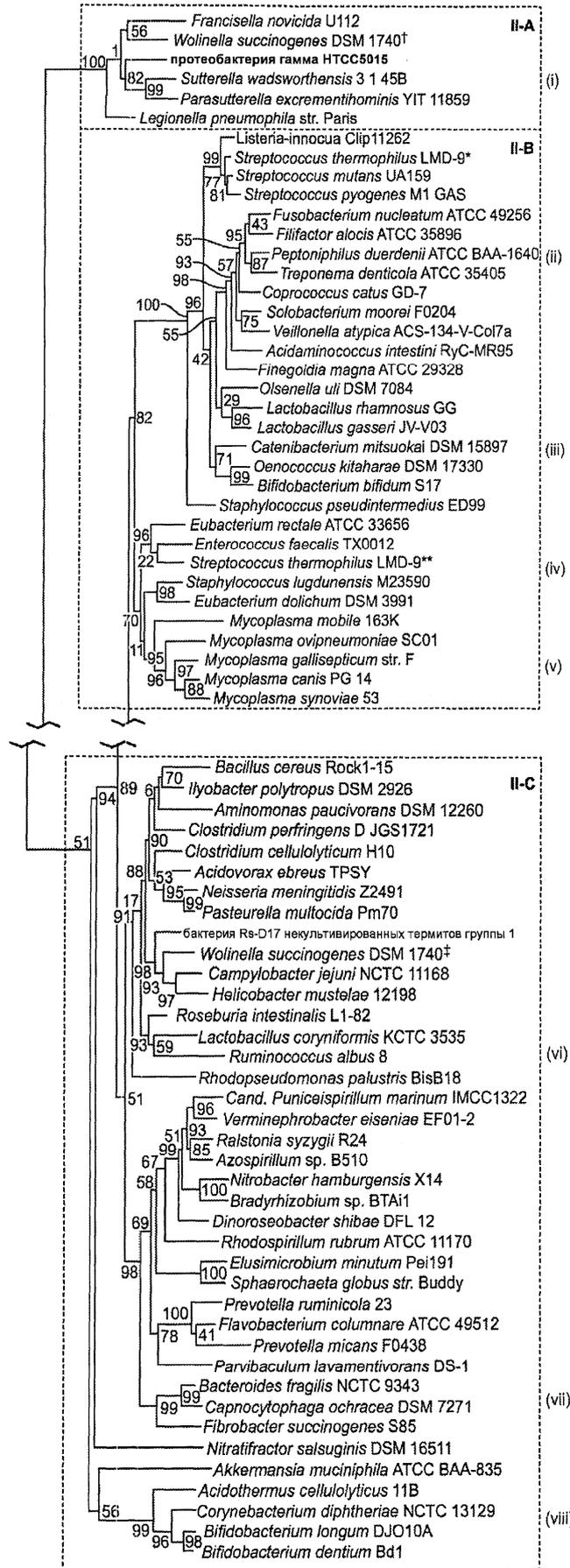




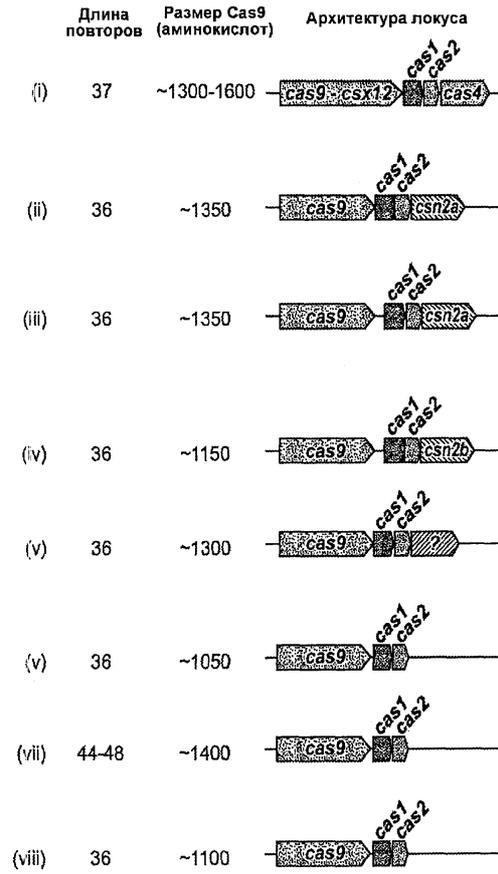
Фиг. 30



Фиг. 31



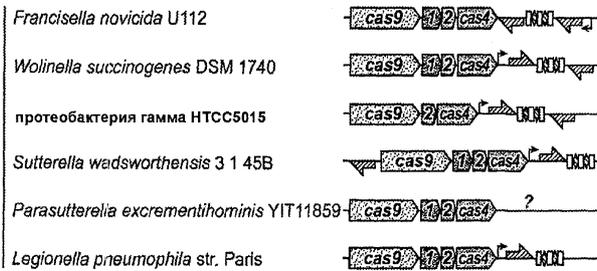
Фиг. 32А



Фиг. 32В

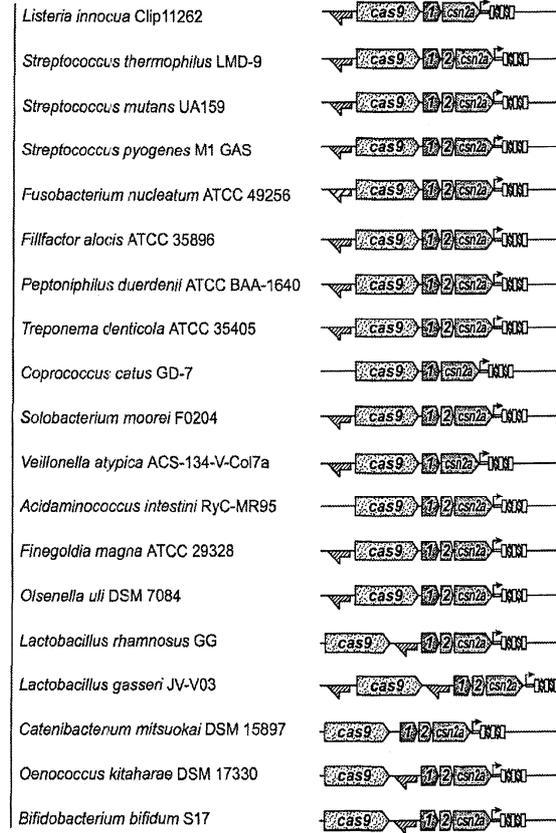
**A**

Тип II-A

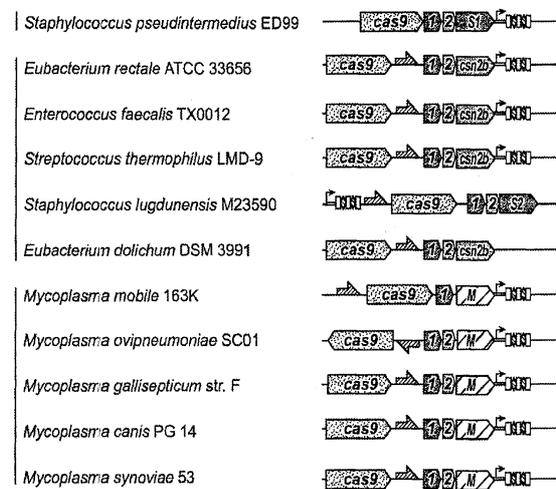


**B**

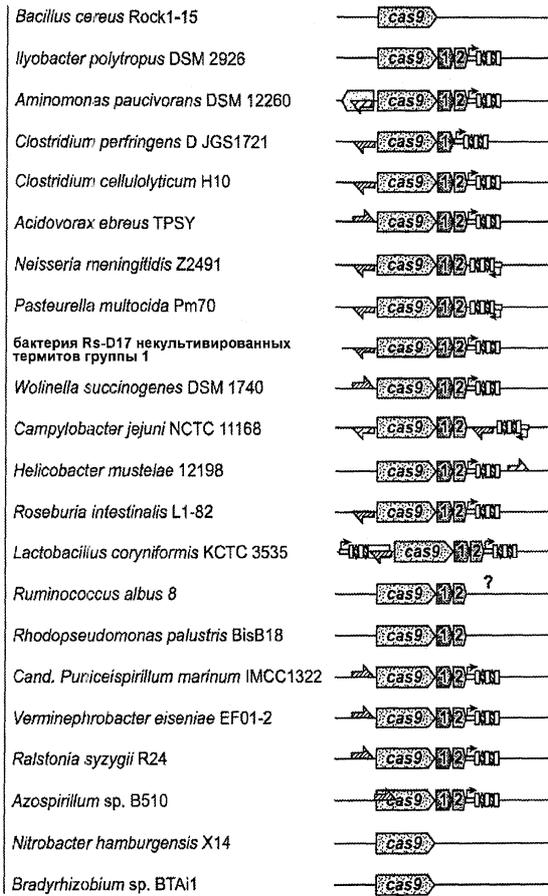
Тип II-B

**C**

Тип II-B (Продолжение)

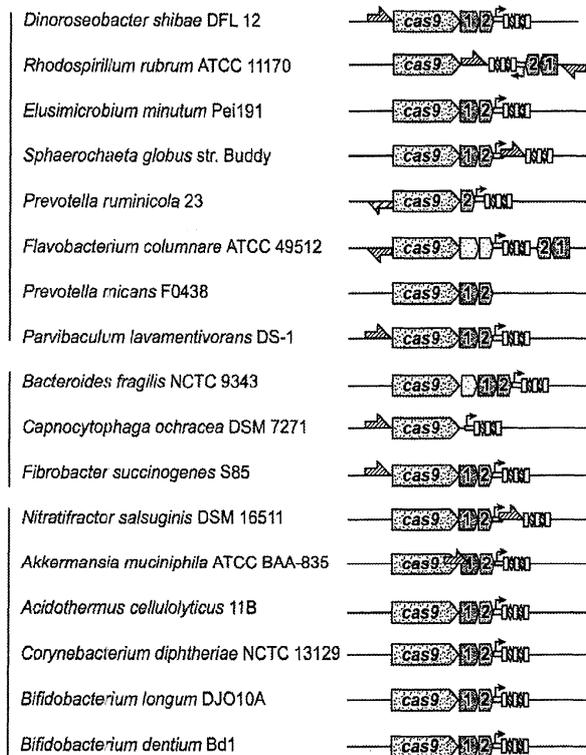


## D Тип II-C



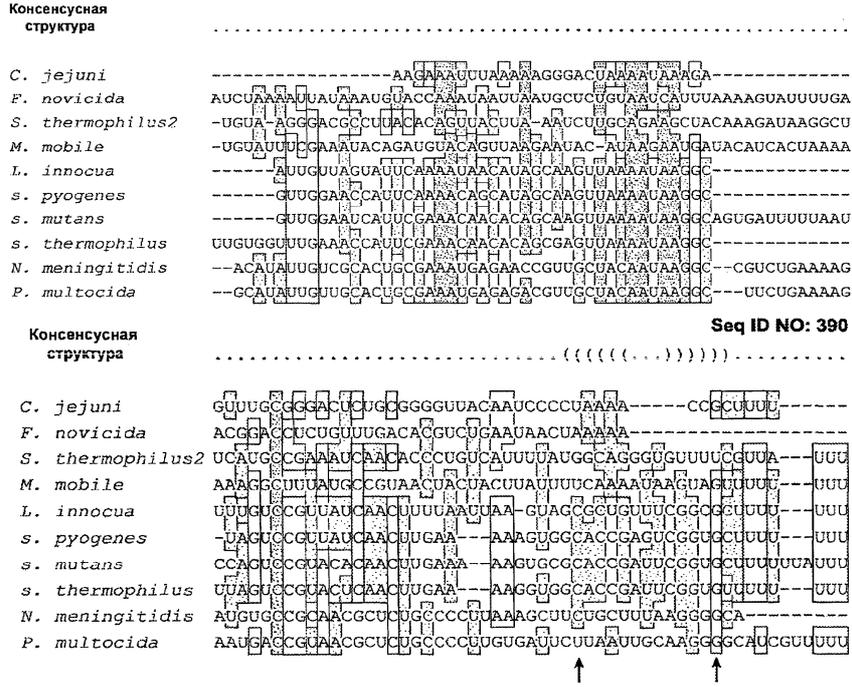
## E

## Тип II-C (Продолжение)

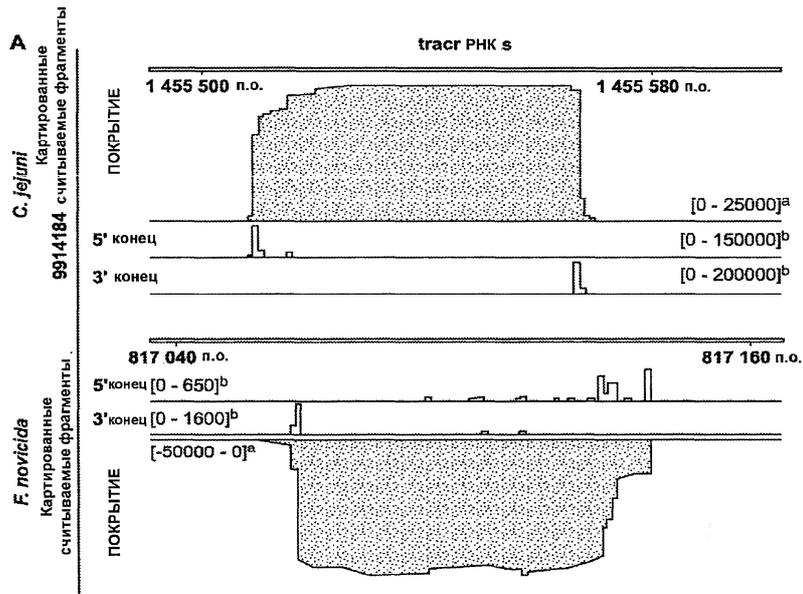


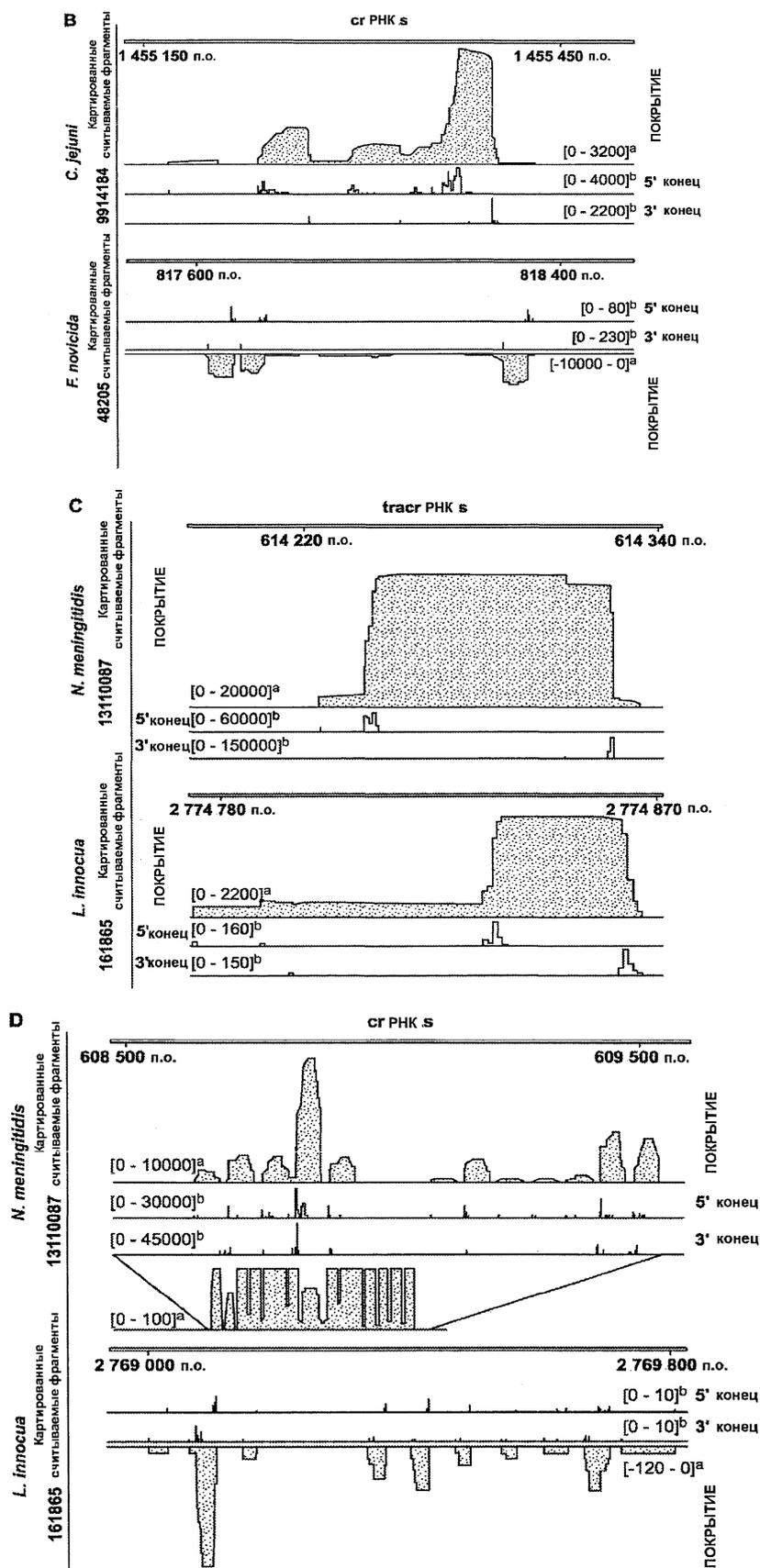
Фиг. 33

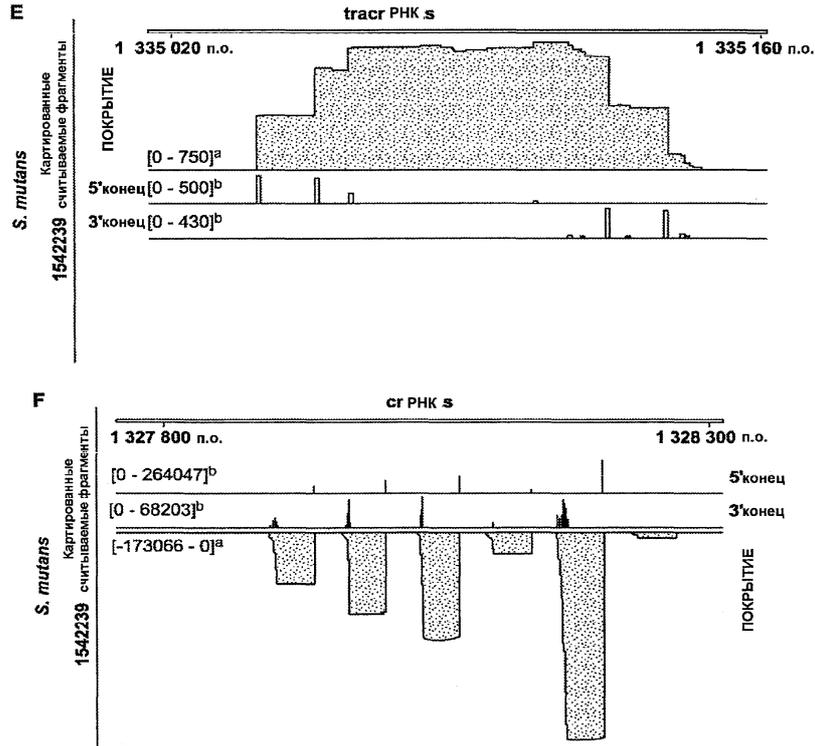




Фиг. 35







Фиг. 36

**А**

сРНК	Нить	Размер зрелой формы	Представляющая интерес область	Считываемые фрагменты	Покрывание (%)	Последовательность	Количество 5'-концевых считываемых фрагментов	Количество 3'-концевых считываемых фрагментов
<b><i>S. jejuni</i> NCTC 11168 (NC_002163.1), общее количество картированных считываемых фрагментов: 9914184</b>								
cr РНК 4 спейсера	→	сРНК1	36 1455167 1455202 781	0,0079	AGTTTTTAAAGAGCTTGGCGGTTGTTTAGT CCCTTTT	1455162 A 1 1455198 T 5 1455165 G 2 1455199 C 5 1455166 C 3 1455200 C 16 1455167 A 419 1455201 C 72 1455168 G 19 1455202 T 385 1455169 T 7 1455203 T 82 1455170 T 75 1455204 T 30 1455171 T 24 1455205 T 112 1455172 T 9 1455206 T 5 1455173 T 11 1455174 A 4		
		сРНК2	38 1455231 1455268 2658 29 1455240 1455268	0,0268	CAAAGTTTCATTAGTTGAATTTAACTGTTTTA GTCCCTTTT	1455227 T 4 1455264 T 3 1455228 T 1 1455265 C 14 1455230 C 37 1455266 C 60 1455231 C 206 1455267 C 240 1455232 A 72 1455268 T 1364 1455233 A 34 1455269 T 348 1455234 A 59 1455270 T 29 1455235 G 280 1455271 T 44 1455236 T 168 1455272 T 15 1455237 T 130 1455273 A 2 1455238 T 36 1455239 C 259 1455240 A 317 1455241 T 148 1455242 T 298 1455243 A 157 1455244 G 40		

**B**

sPHK	Нить	Размер зрелой формы	Представляющая интерес область	Считываемый фрагмент	Покрывание (%)	Последовательность	Количество 3'-концевых считываемых фрагментов	Количество 3'-концевых считываемых фрагментов
<b>C. jejuni NCTC 11168 (NC 002163.1), общее количество картированных считываемых фрагментов: 9914184</b>								
crPHK 4 спейсера	crPHK3	35	1455300	1455334	4729	0,0477	AAGAATGAGGATGATGATATTTTACAGTTTTA GTCCCTTTTT	1455292 T 16 1455329 G 21 1455293 C 62 1455330 T 479 1455294 T 5 1455331 C 23 1455295 A 12 1455332 C 48 1455296 C 23 1455333 C 179 1455297 A 367 1455334 T 1226 1455298 A 143 1455335 T 220 1455299 G 553 1455336 T 54 1455300 A 1014 1455337 T 49 1455301 A 415 1455338 T 55 1455302 T 242 1455339 A 1 1455303 G 589 1455304 A 55 1455305 G 2
							crPHK4	26

**C**

sPHK	Нить	Размер зрелой формы	Представляющая интерес область	Считываемый фрагмент	Покрывание (%)	Последовательность	Количество 3'-концевых считываемых фрагментов	Количество 3'-концевых считываемых фрагментов
<b>C. jejuni NCTC 11168 (NC 002163.1), общее количество картированных считываемых фрагментов: 9914184</b>								
tracrPHK →	tracrPHK1	65	1455502	1455566	833829	8,4105	AAGAAATTTAAAAGCGACTAAAATAAAGAGT TTGCGGGACCTCGCGGGTTACAATCCCCCTAA AACCGCTTTT	1455496 T 1 1455565 G 12371 1455497 A 31 1455566 C 713292 1455498 A 27 1455567 T 74594 1455499 G 24 1455568 T 10412 1455500 A 19 1455569 T 9580 1455501 A 232 1455570 T 426 1455502 A 435 1455571 A 91 1455503 T 369 1455572 A 97 1455504 T 253 1455573 A 542 1455505 T 85 1455506 A 65 1455507 A 193 1455508 A 33472 1455509 A 615001 1455510 A 131879 1455511 G 16444 1455512 G 9390 1455513 G 1053

**D**

sPHK	Нить	Размер зрелой формы	Представляющая интерес область	Считываемый фрагмент	Покрывание (%)	Последовательность	Количество 3'-концевых считываемых фрагментов	Количество 3'-концевых считываемых фрагментов						
<b>F. novicida U112 (NC 008601.1), общее количество картированных считываемых фрагментов: 48205</b>														
crPHK 13 спейсеров	crPHK5	52	817556	817607	117	0,2427	ATAACTCGACCAGATATTGACAAAGTTTCAGT TGCTGAATTAATTGGTAAACCT	817612 T 10 817557 A 3 817611 T 10 817556 C 55 817607 A 53 817555 T 13 817606 T 7 817553 C 12 817552 T 1						
							crPHK6	56	817627	817682	116	0,2406	GGCAGTTTGTATGGTCATATAGGAGTGT CAGTTGCTGAATTAATTGGTAAACCT	817685 A 4 817629 A 4 817682 G 23 817628 A 3 817681 G 13 817627 C 53 817679 A 2 817626 T 20 817677 G 4
													crPHK7	66
	crPHK9	53	817845	817897	24	0,0498								
	tracrPHK ←	tracrPHK1	74	817065	817138	2808	5,8251	GTACCAATAAATAAAGCTCTGTAATCATTA AAAGTATTTTGAACGGACCTCTGTTTGACACG TCTGAATAACTAAAAA	817140 A 2 817066 A 28 817139 T 2 817065 C 1523 817138 G 615 817064 T 440 817136 A 9 817063 A 7 817135 C 2 817062 A 10 817134 C 37 817061 A 19 817133 A 26 817060 A 14 817132 A 7 817059 A 6 817131 A 328 817058 G 5 817130 T 355 817057 C 32 817129 A 165 817055 A 10 817128 A 484 817127 T 24 817126 T 55 817125 A 32 817124 A 19					

**E**

sPHK	Нить	Размер зрелой формы	Представляющая интерес область	Покрываемость (%)	Покрываемость (%)	Последовательность	Количество 5'-концевых считываемых фрагментов	Количество 3'-концевых считываемых фрагментов
<i>N. meningitidis</i> AZ2491 (NC_003116.1), общее количество картированных считываемых фрагментов: 13110087								
crPHK → 16 сплайсеров	crPHK1	48	608456	608503	1346	0,0103	TATCCATTCCAGCCGGAAATTAAGTTGTAGC	608453 T 30 608500 T 25
							TCCCTTTCATTTCCGCACT	608454 G 105 608501 C 77
								608455 T 64 608502 G 175
								608456 A 259 608503 C 293
								608457 T 4 608504 A 78
								608458 C 21 608505 G 505
								608459 C 175 608507 T 1
								608460 A 118
								608461 T 39
								608517 T 5 608564 T 8
								608518 C 41 608565 T 16
								608519 T 7 608566 T 8
								608520 G 61 608567 C 31
								608521 C 21 608568 G 101
								608522 C 44 608569 C 173
								608523 T 21 608570 A 43
	608524 T 31 608571 G 6							
	608525 T 17 608572 T 2							

**F**

sPHK	Нить	Размер зрелой формы	Представляющая интерес область	Покрываемость (%)	Покрываемость (%)	Последовательность	Количество 5'-концевых считываемых фрагментов	Количество 3'-концевых считываемых фрагментов
<i>N. meningitidis</i> AZ2491 (NC_003116.1), общее количество картированных считываемых фрагментов: 13110087								
crPHK3	50	608586	608635	12402	0,0946	TAAAGGTTTCTGTGCGACCCGAATGTTGTAG	608583 T 5 608631 T 75	
						TCCCTTTCATTTCCGCACT	608584 G 3 608632 T 114	
							608585 G 513 608633 C 414	
							608586 T 2243 608634 G 1510	
							608587 A 2091 608635 C 2219	
							608588 A 188 608636 A 297	
							608589 A 233 608637 G 52	
							608590 G 744 608638 T 8	
							608591 G 1152	
							608646 T 7 608697 T 295	
							608647 C 127 608698 T 569	
							608648 T 203 608699 C 1167	
							608649 T 377 608700 G 4910	
							608650 T 2751 608701 C 4968	
							608651 A 477 608702 A 764	
							608652 A 1334 608703 G 62	
	608653 C 11694 608704 T 4							
	608654 T 3665							
	608655 T 362							

**G**

sPHK	Нить	Размер зрелой формы	Представляющая интерес область	Покрываемость (%)	Покрываемость (%)	Последовательность	Количество 5'-концевых считываемых фрагментов	Количество 3'-концевых считываемых фрагментов					
<i>N. meningitidis</i> AZ2491 (NC_003116.1), общее количество картированных считываемых фрагментов: 13110087													
crPHK5	49	608719	608767	28747	0,2193	AACCCACTAAATTTTGCATAATGCGGTTGTAGC	608717 G 20 608764 C 426						
						TCCCTTTCATTTCCGCACT	608718 C 1171 608765 C 1316						
							608719 A 8128 608766 G 3040						
							608720 A 116 608767 C 6431						
							608721 C 59 608768 A 1049						
							608722 C 420 608769 G 102						
							608723 C 1089 608770 T 11						
							608724 A 557						
							608781 G 42 608830 T 2295						
							608782 C 630 608831 C 7060						
							608783 C 10039 608832 G 13864						
							608784 T 26026 608833 C 44472						
							608785 T 11430 608834 A 6853						
							608786 T 8248 608835 G 802						
							608787 T 3648 608836 T 69						
						crPHK6	50	608784	608835	121014	0,9231	TTTTTTGTACTGTTTGAACGAGTTGTAG	608781 G 42 608830 T 2295
TCCCTTTCATTTCCGCACT	608782 C 630 608831 C 7060												
	608783 C 10039 608832 G 13864												
	608784 T 26026 608833 C 44472												
	608785 T 11430 608834 A 6853												
	608786 T 8248 608835 G 802												
	608787 T 3648 608836 T 69												
crPHK7	52	608848	608899	24611	0,1877							TTCTGTTTCAGATAGCAACGCACTAGTTGTAG	608846 A 28 608896 T 369
												AGCTCCCTTTCATTTCCGCACT	608847 A 631 608897 C 1114
													608848 T 11039 608898 G 3441
													608849 T 753 608899 C 9015
													608850 C 239 608900 A 1558
													608851 G 2205 608901 G 137
													608852 T 1910 608902 T 8
													608853 T 826
												84	608916
						TCCCTTTCATTTCCGCACT	608894 G 3 608997 A 2						
							608895 G 2 608998 C 2						
							608896 A 266 608999 T 259						
							608897 T 17 609000 G 171						
							608898 A 96 609001 G 24						
							608899 T 26 609004 T 16						

**H**

sPHK	Нить	Размер зрелой формы	Представляющая интерес область	Покрываемость (%)	Покрываемость (%)	Последовательность	Количество 5'-концевых считываемых фрагментов	Количество 3'-концевых считываемых фрагментов					
<i>N. meningitidis</i> AZ2491 (NC_003116.1), общее количество картированных считываемых фрагментов: 13110087													
crPHK10	49	609049	609097	5027	0,0383	CTTTTGTATTGATTCAGAGTGTGTGTAGC	609046 T 89 609094 T 81						
						TCCCTTTCATTTCCGCACT	609047 C 37 609095 C 218						
							609048 G 647 609096 G 845						
							609049 C 1614 609097 C 1237						
							609050 T 583 609098 A 219						
							609051 T 208 609099 G 36						
							609052 T 104 609101 G 7						
							609053 T 18						
						crPHK11	52	609112	609163	22711	0,1732	ATTCTGTCGATGATGGAACCTCGAGCATGTGTAG	609109 A 58 609160 T 540
												AGCTCCCTTTCATTTCCGCACT	609110 G 331 609161 C 1263
													609111 T 99 609162 G 2850
													609112 A 10627 609163 C 4882
													609113 T 1234 609164 A 552
													609114 T 66 609165 G 52
													609115 C 191 609166 T 15
													609116 G 4906
	609117 T 136												
crPHK12	52	609178	609229	5067	0,0386							TAGCCAGTGTAAACCCGACCCGCTTGTGTAG	609175 G 4 609226 T 57
												AGCTCCCTTTCATTTCCGCACT	609176 G 3 609227 C 199
													609177 G 42 609228 G 576
													609178 T 897 609229 C 1348
													609179 A 72 609230 A 321
													609180 G 584 609231 G 33
													609181 C 154 609232 T 2
							609182 C 314						
							609183 A 38						

**I**

sPHK	Нить	Размер зрелой формы	Представляющая интерес область	Ссылка на файл	Покрываемость (%)	Последовательность	Количество 3'-концевых считываемых фрагментов	Количество 3'-концевых считываемых фрагментов					
<b><i>N. meningitidis</i> AZ2491 (NC_003116.1), общее количество картированных считываемых фрагментов: 13110087</b>													
crPHK13		51	609245	609295	4666	0,0356	ATAGAAATACATACGCCGACTAATTAGTTCTGCTCCCTTTCTCATTTCGCAGT	609243	A	6	609292	T	65
								609244	A	62	609293	C	201
								609245	A	1311	609294	G	475
								609246	T	141	609295	C	1039
								609247	A	116	609296	A	229
609248	G	44	609297	G	25								
<b>609298 T 1</b>													
crPHK14		51	609311	609361	7147	0,0545	TTTTTGTAAATTGTCTGCCTTTTATTAGTTGTA GCTCCCTTTCTCATTTCGCAGT	609308	T	12	609358	T	136
								609309	T	22	609359	C	442
								609310	C	207	609360	G	858
								609311	T	1190	609361	C	2335
								609312	T	774	609362	A	540
								609313	T	577	609363	G	69
								609314	T	195	609364	T	5
								609315	T	37			
crPHK15		50	609378	609427	49818	0,3800	ACGGCGGAACCATGCCACAAAACGTTTAGCTCCCTTTCTCATTTCGCAGT	609375	C	319	609424	T	532
								609376	C	7253	609425	C	1414
								609377	C	7249	609426	G	2553
								609378	A	19015	609427	C	7448
								609379	C	547	609428	A	1854
								609380	G	307	609429	G	210
								609381	G	190	609430	T	10

**J**

sPHK	Нить	Размер зрелой формы	Представляющая интерес область	Ссылка на файл	Покрываемость (%)	Последовательность	Количество 3'-концевых считываемых фрагментов	Количество 3'-концевых считываемых фрагментов								
<b><i>N. meningitidis</i> AZ2491 (NC_003116.1), общее количество картированных считываемых фрагментов: 13110087</b>																
crPHK16		50	609444	609493	42398	0,3234	ATAATAACCCAAATACACGATGTTAAGTTGTA GCTCCCTTTCTCATTTCGCAGT	609441	A	8	609490	T	579			
								609442	A	548	609491	C	2197			
								609443	A	3686	609492	G	6466			
								609444	A	4638	609493	C	13217			
								609445	T	1428	609494	A	2438			
								609446	A	2861	609495	G	200			
								609447	A	1322	609496	T	12			
								609448	T	287						
								<b>614158 C 1 614319 G 311</b>								
								tracrPHK →		163	614162	614324	208318	1,5890	TTGCTTATTATATAACAAATAGATTATTGAC TTATCATCTCACACGGCTAGAAATCCCAAACA TATTGTCGCACCTGCGAATGAGAACCCTTGCT ACAATAAGCCCTGTGAAAGATGTGCCGCAA CGCTGCCCCCTTAAAGCTTCTGCTTAAAGGG GCATCGTTTAT	614161
614162	T	2	614321	G	258											
614164	G	1	614322	G	150											
614223	A	596	614323	C	41244											
614224	C	39	614324	A	128531											
614225	A	14761	614325	T	167											
614226	T	169	614326	C	4197											
614227	A	276	614327	G	36											
614239	C	208	614328	T	122											
614240	G	42037	614329	T	865											
614241	A	39965	614330	T	1132											
614242	A	22890	614331	A	1061											
614243	A	51186	614332	T	2788											
614244	T	1771	614333	T	3759											
614245	G	4192	614334	T	77											
614246	A	488	614335	C	18											
			614336	G	3											
			614337	G	3											
			614338	T	1											

**K**

sPHK	Нить	Размер зрелой формы	Представляющая интерес область	Ссылка на файл	Покрываемость (%)	Последовательность	Количество 3'-концевых считываемых фрагментов	Количество 3'-концевых считываемых фрагментов												
<b><i>L. innocua</i> Clp11262 (NC_003212.1), общее количество картированных считываемых фрагментов: 161866 (замечание: низкое качество библиотек RNA)</b>																				
crPHK ← 10 сплайсеров		35	2769606	2769640	2	0,0012	GGTAACCTTTCCTAGGATAGTTTATAGAGCTAT GTT	2769640	G	1	2769606	T	1							
								2769561	C	1	2769540	T	1							
								2769560	A	1										
								crPHK2	22	2769540	2769561	2	0,0012	CATTATGTTTTAGAGCTATGTT	2769491	G	3	2769468	G	1
															2769428	T	5	2769407	A	1
															2769427	T	2	2769406	T	1
								2769405	T	1										
								2769403	T	1										
								2769402	G	1										
								crPHK4	27	2769402	2769428	7	0,0043	TTATAGTTTTAGAGCTATGTTATTTTG	2769362	T	3	2769339	T	1
2769360	A	2	2769337	T	2															
2769164	T	1	2769143	A	1															
3769163	A	1	2769142	T	1															
crPHK5	26	2769337	2769362	5	0,0031	TAAATGTTTTAGAGCTATGTTATTTT	2769101	T	6	2769079	T	4								
							2769100	T	1	2769078	T	1								
							2769099	C	4	2769075	T	2								
							2769098	A	3	2769073	T	3								
							2769097	T	3	2769072	G	6								
							2769096	G	2											
							2769027	G	1	2769000	T	1								
							crPHK8		23	2769142	2769164	2	0,0012	TACAAGTTTTAGAGCTATGTTAT	2774774	A	34	2774861	T	2
2774783	A	1	2774862	T	47															
2774786	C	1	2774863	T	150															
2774787	A	1	2774864	T	67															
2774788	A	22	2774865	T	30															
2774794	C	1	2774866	G	15															
2774795	A	1	2774867	T	6															
2774796	T	5																		
2774797	A	2																		
2774799	C	5																		
2774801	A	1																		
crPHK9	30	2769072	2769101	19	0,0017	TTCATGTTGTTTTAGAGCTATGTTATTTTG	2774774	A	34	2774861	T	2								
							2774783	A	1	2774862	T	47								
							2774786	C	1	2774863	T	150								
crPHK10		28	2769000	2769027	19	0,0017	GTTTTAGAGCTATGCTATTCGAATACT	2774774	A	34	2774861	T	2							
								2774783	A	1	2774862	T	47							
								2774786	C	1	2774863	T	150							
								2774787	A	1	2774864	T	67							
								2774788	A	22	2774865	T	30							
								2774794	C	1	2774866	G	15							
								2774795	A	1	2774867	T	6							
								2774796	T	5										
								2774797	A	2										
								2774799	C	5										
2774801	A	1																		

**L**

sPHK	Нить	Размер зрелой формы	Представляющая интерес область	Ссылка на файл	Покрываемость (%)	Последовательность	Количество 3'-концевых считываемых фрагментов	Количество 3'-концевых считываемых фрагментов					
<b><i>L. innocua</i> Clp11262 (NC_003212.1), общее количество картированных считываемых фрагментов: 161866 (замечание: низкое качество библиотек RNA)</b>													
tracrPHK →		90	2774774	2774863	367	0,2267	ATTGTTAGTATTCACAAATAACATAGCAAGTTA AATGAAGGCTTTGTCGGTTATCAACTTTAAT TAAGTAGCGCTGTTCGGCGCTTTTTT	2774774	A	34	2774861	T	2
								2774783	A	1	2774862	T	47
								2774786	C	1	2774863	T	150
								2774787	A	1	2774864	T	67
								2774788	A	22	2774865	T	30
								2774794	C	1	2774866	G	15
								2774795	A	1	2774867	T	6
								2774796	T	5			
								2774797	A	2			
								2774799	C	5			
2774801	A	1											

**M**

sРНК	Нить	Размер зрелой формы	Представляющая интерес область	Покрываемый фрагмент	Покрываемость (%)	Последовательность	Количество 5'-концевых считываемых фрагментов	Количество 3'-концевых считываемых фрагментов
<b>S. mutans UA159 (NC 004350.2), общее количество картированных считываемых фрагментов: 1542239</b>								
crРНК 5 спейсеров	← crРНК1	38	1328162	1328199	267104	17,3192	GCCAATTAATTAATATGGTGAGTTFITAGAGCTG <u>EGTTGTTTCGA</u>	1328201 A 8 1328166 G 18547
								1328200 C 178 1328165 T 41345
								1328199 G 264047 1328164 T 53386
								1328198 C 191 1328163 G 9084
								1328197 C 167 1328162 T 59197
								1328196 A 117 1328161 T 28333
								1328160 T 4240
								1328159 C 18236
								1328158 G 26742
								1328157 A 5573
								1328156 A 17
	crРНК2	36	1328098	1328133	26578	1,7233	GCTAGCCGAGTGTAGTCTCTGTTTITAGAGCTG <u>EGTTGTTTCGA</u>	1328135 C 4 1328101 T 37
								1328134 A 13 1328100 G 800
								1328133 G 25656 1328099 T 6395
								1328132 C 212 1328098 T 11256
								1328131 T 25 1328097 G 301
								1328130 A 62 1328096 T 1453
								1328129 G 25 1328095 T 1755
								1328094 T 447
								1328093 C 1302
								1328092 G 1996
								1328091 A 670
								1328090 A 5

**N**

sРНК	Нить	Размер зрелой формы	Представляющая интерес область	Покрываемый фрагмент	Покрываемость (%)	Последовательность	Количество 5'-концевых считываемых фрагментов	Количество 3'-концевых считываемых фрагментов
<b>S. mutans UA159 (NC 004350.2), общее количество картированных считываемых фрагментов: 1542239</b>								
	crРНК3	34	1328034	1328134	138067	8,9567	TGTTGTCTCATCATAGTTAGGTTTITAGAGCTG <u>EGTTGTTTCGA</u>	1328069 G 4 1328036 G 2880
								1328068 C 66 1328035 T 733
								1328067 T 134361 1328034 G 68203
								1328066 G 609 1328033 T 37212
								1328065 T 321 1328032 T 24528
								1328031 G 1003
								1328030 T 889
								1328029 T 610
								1328028 T 124
								1328027 C 371
								1328026 G 546
								1328025 A 145
	crРНК4	35	1327967	1328001	104705	6,7892	CAATTAGACAAATAGACAAACCTTTITAGAGCTG <u>EGTTGTTTCGA</u>	1328003 T 8 1328969 T 348
								1328002 T 36 1328968 G 8606
								1328001 C 101017 1328967 T 59863
								1328000 A 681 1328966 T 25845
								1327999 A 216 1328965 G 756
								1328964 T 3053
								1328963 T 1433
								1328962 T 1602
								1328961 C 864
								1328960 G 839
								1328959 A 322

**O РНК**

sРНК	Нить	Размер зрелой формы	Представляющая интерес область	Покрываемый фрагмент	Покрываемость (%)	Последовательность	Количество 5'-концевых считываемых фрагментов	Количество 3'-концевых считываемых фрагментов
<b>S. mutans UA159 (NC 004350.2), общее количество картированных считываемых фрагментов: 1542239</b>								
	crРНК5	37	1327899	1327935	63999	4,1497	TTCGCACATGACTTGCACAGTTFITAGAGCTG <u>EGTTGTTTCGA</u>	1327940 A 4 1327902 G 902
								1327937 T 6 1327901 T 3296
								1327936 A 17 1327900 T 12381
								1327935 T 62587 1327899 G 19843
								1327934 T 1029 1327898 T 2864
								1327933 C 108 1327897 T 14180
								1327932 G 19 1327896 T 2061
								1327931 G 36 1327895 C 2079
								1327894 G 5259
								1327893 A 630
tracrРНК →	tracrРНК1	102	1335040	1335141	1299	0,0842	CTTGGATCATTCGAACAACACACCAAGTTA AATFAGGCAGTGTATTTAATCCAGTCCGTA CACAACTTGAAAAGTGGCCACCCGATTCGGTG CTTTTTTATTT	1335038 G 1 1335140 T 13
	tracrРНК2	88	1335054	1335141				1335040 G 466 1335141 T 64
	tracrРНК3	80	1335062	1335141				1335041 T 6 1335142 T 29
								1335042 T 1 1335143 A 13
								1335051 T 4 1335144 T 6
								1335053 G 3 1335145 T 13
								1335054 A 415 1335146 T 1
								1335055 A 2 1335149 T 1
								1335057 C 1
								1335058 A 1
								1335062 C 186
								1335063 A 15

Фиг. 37

A

Кластер	SEQ ID NO:
1	2, 3, 4, 5, 6, 7, 8, 15, 23, 24, 25, 36, 37, 38, 39, 41, 71, 74, 105, 116, 136, 138, 166, 177, 180, 183, 193, 204, and 232
2	83, 75, 156, 96, 121, 235, 208, 127, 182, 134, 119, 246, 153, 202
3	101, 168, 48, 226, 216, 210, 120, 102, 176, 57, 108, 79, 1, 245
4	219, 135, 53, 62, 240, 165, 217, 82, 212, 19, 40, 18, 194
5	84, 21, 150, 221, 111, 76, 47, 59, 77, 112, 198, 147
6	90, 91, 214, 92, 152, 98, 243, 197, 32, 227, 162
7	103, 187, 223, 151, 158, 126
8	88, 167, 13, 164, 184, 123
9	58, 73, 195, 148, 31, 33
10	206, 188, 211, 161, 205, 44

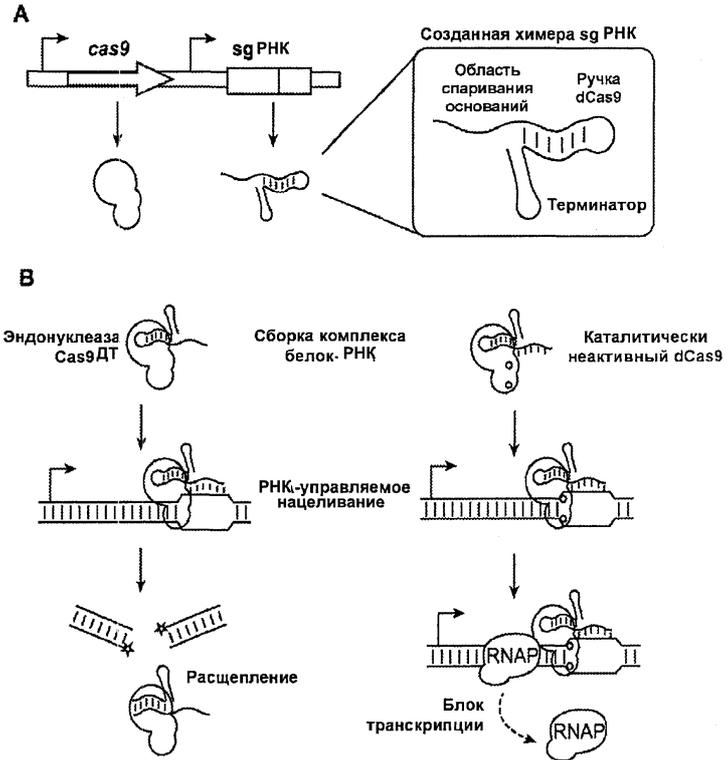
Кластер	SEQ ID NO:
11	50, 54, 78, 106, 174
12	209, 220, 146, 157
13	70, 154, 100, 117
14	128, 144, 118, 129
15	131, 66, 149, 145
16	89, 169, 163
17	141, 49, 72
18	196, 114, 86

Кластер	SEQ ID NO:
19	55, 27, 215
20	228, 234
21	160, 213
22	207, 237
23	230, 94
24	200, 247
25	133, 143
26	64, 68
27	20, 45
28	60, 56
29	99, 52

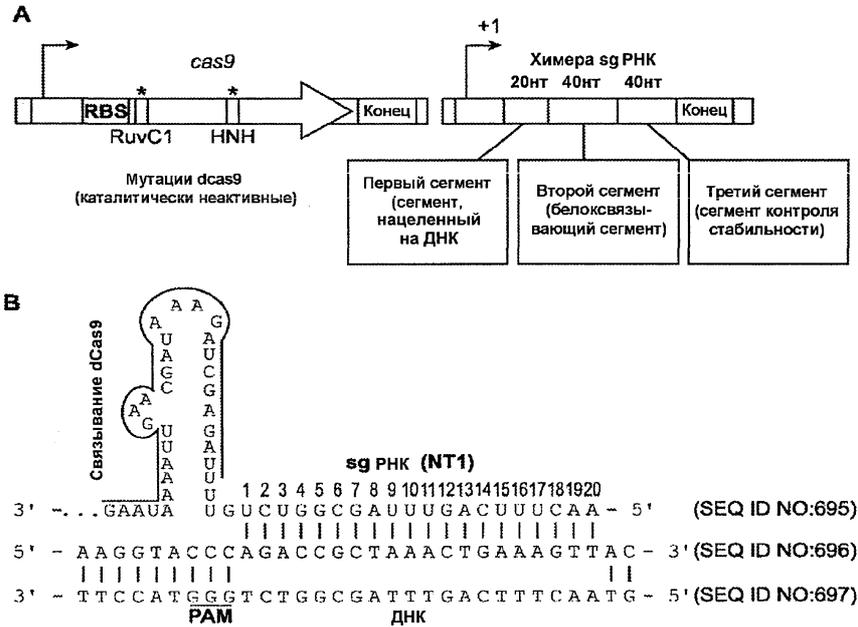
B

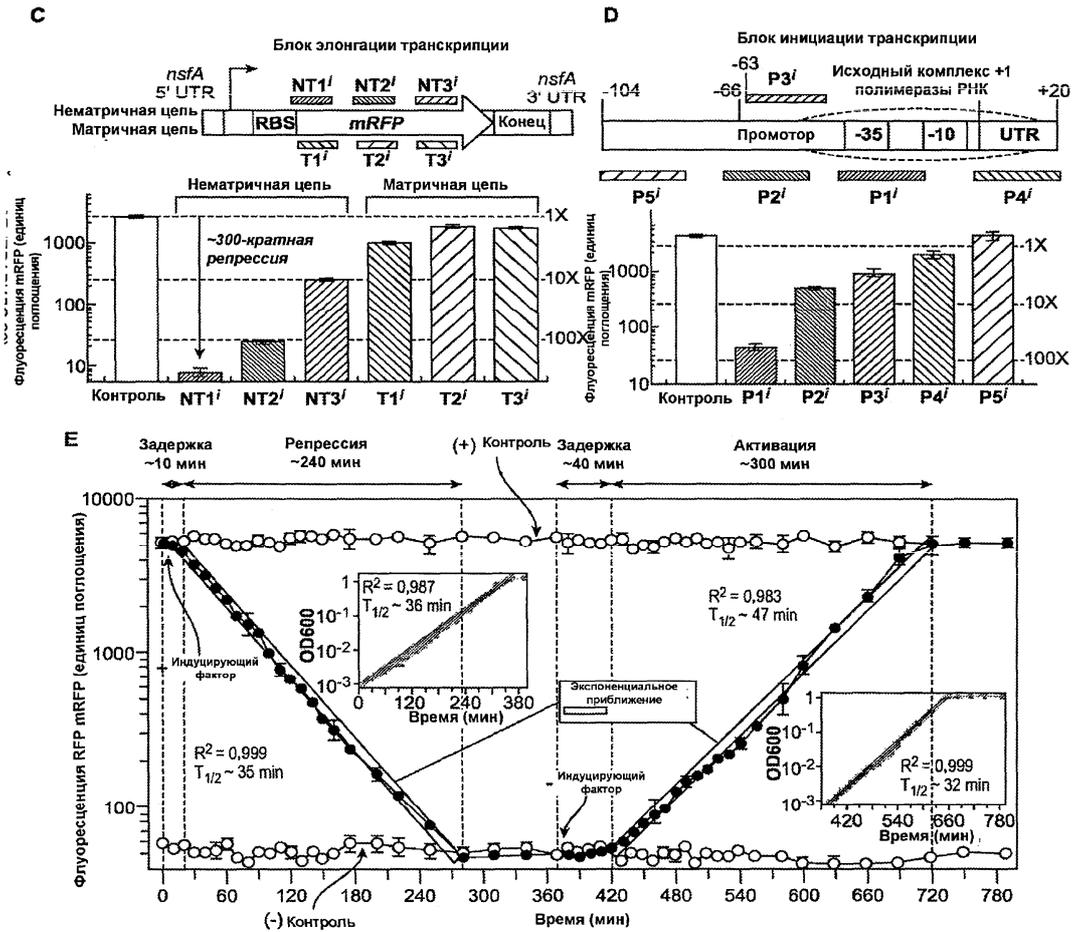
Кластер	SEQ ID NO:	Кластер	SEQ ID NO:	Кластер	SEQ ID NO:
30	244, 185	45	233	59	218
31	43	46	122	60	65
32	189	47	16	61	171
33	170	48	242	62	97
34	11	49	203	63	63
35	107	50	26	64	46
36	14	51	137	65	225
37	236	52	199	66	10
38	12	53	34	67	173
39	17	54	201	68	51
40	239	55	178	69	142
41	61	56	42	70	69
42	85	57	190	71	28
43	191	58	81	72	139
44	22			73	80
				74	172
				75	115
				76	229
				77	175
				78	181

Фиг. 38

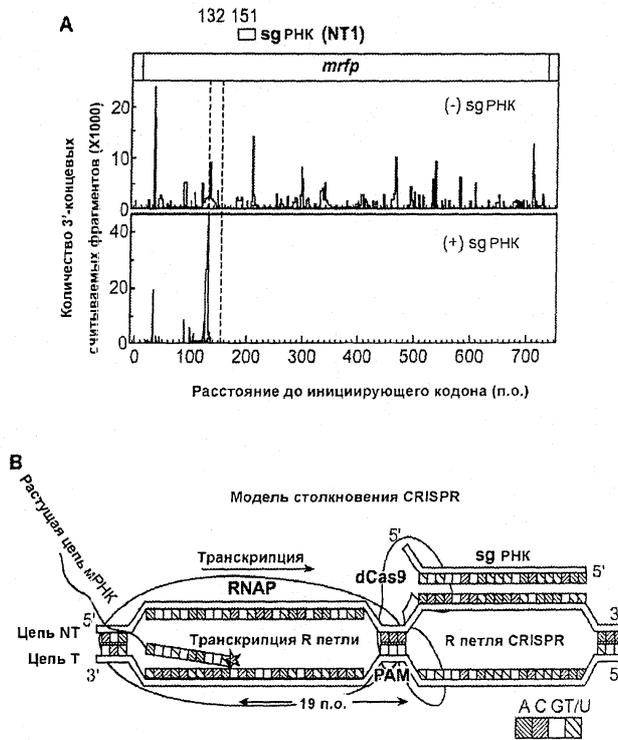


Фиг. 39

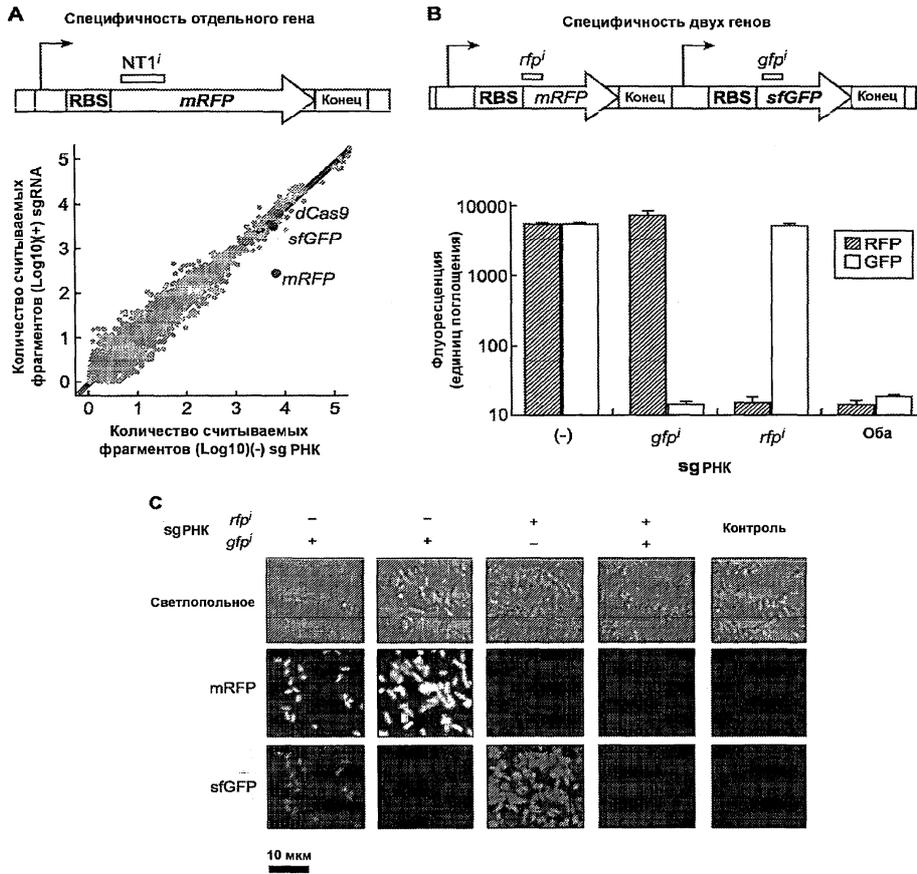




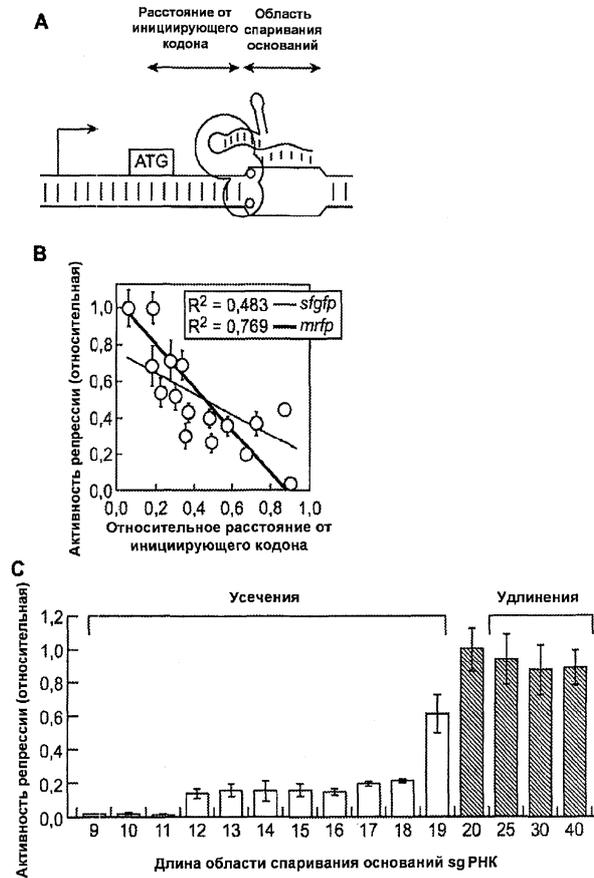
Фиг. 40

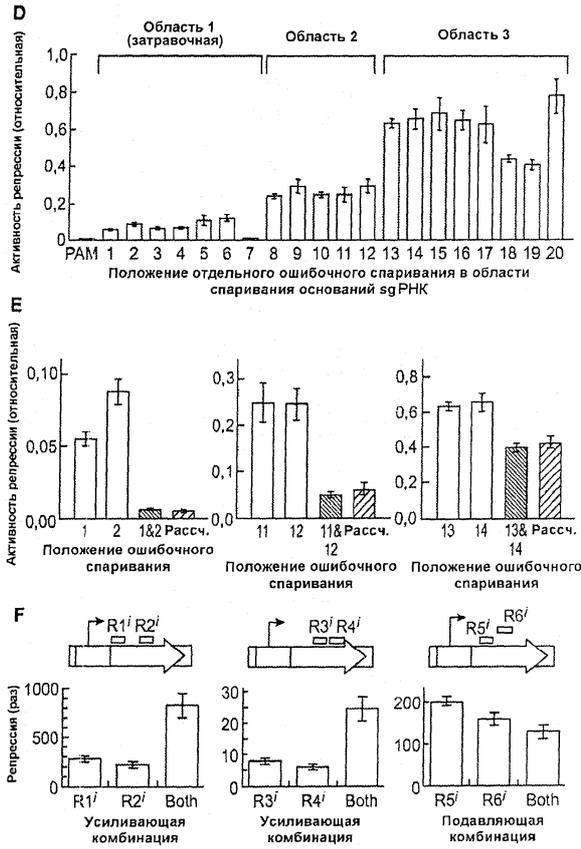


Фиг. 41

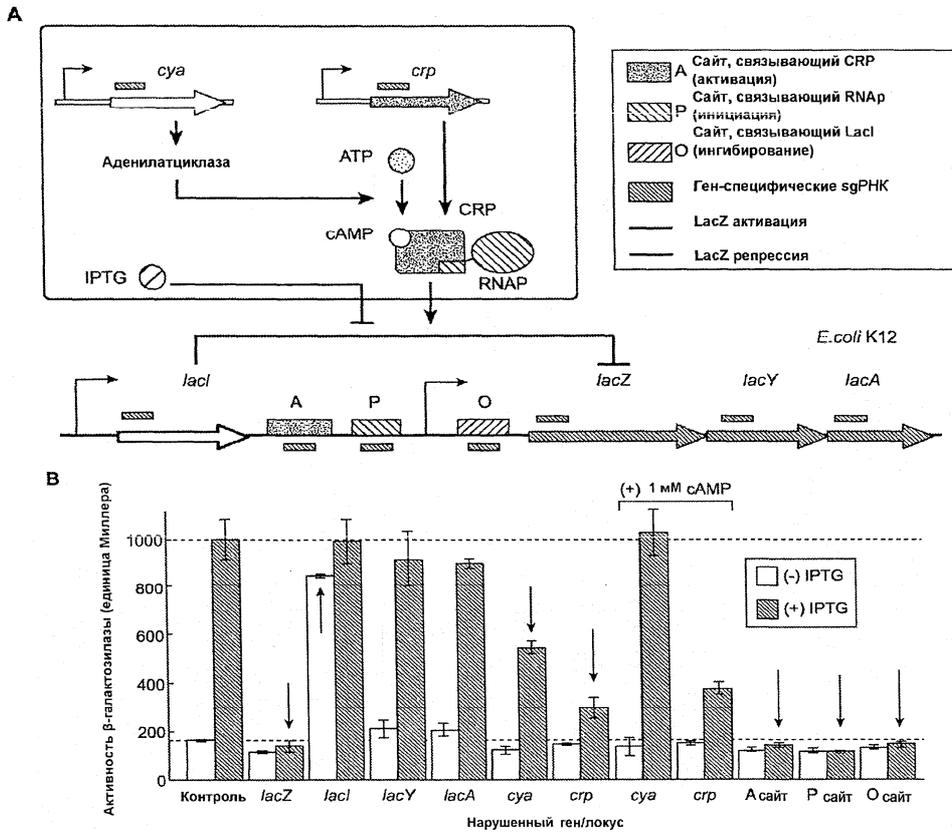


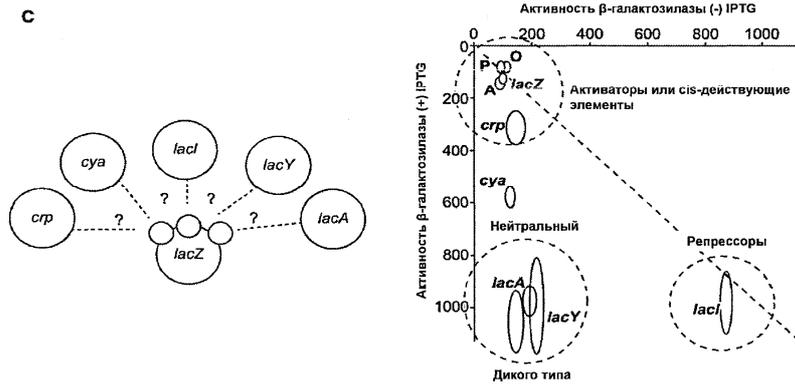
Фиг. 42



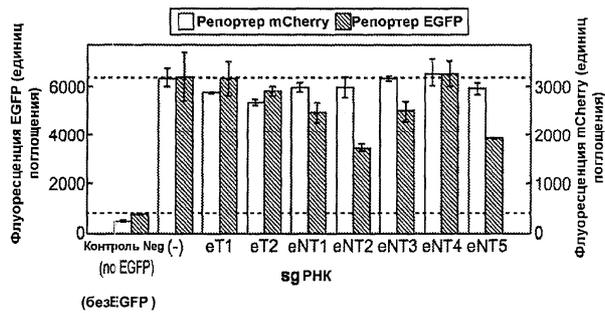
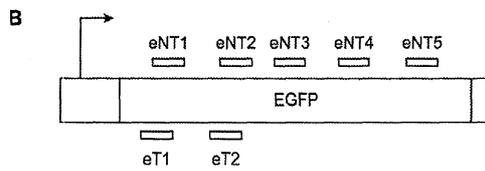
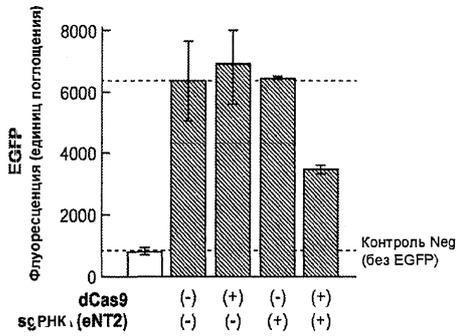
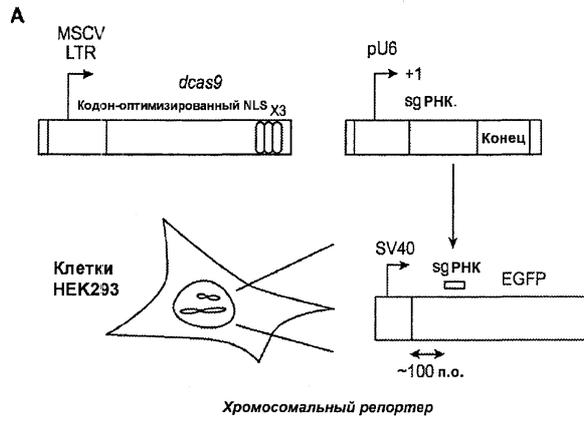


Фиг. 43

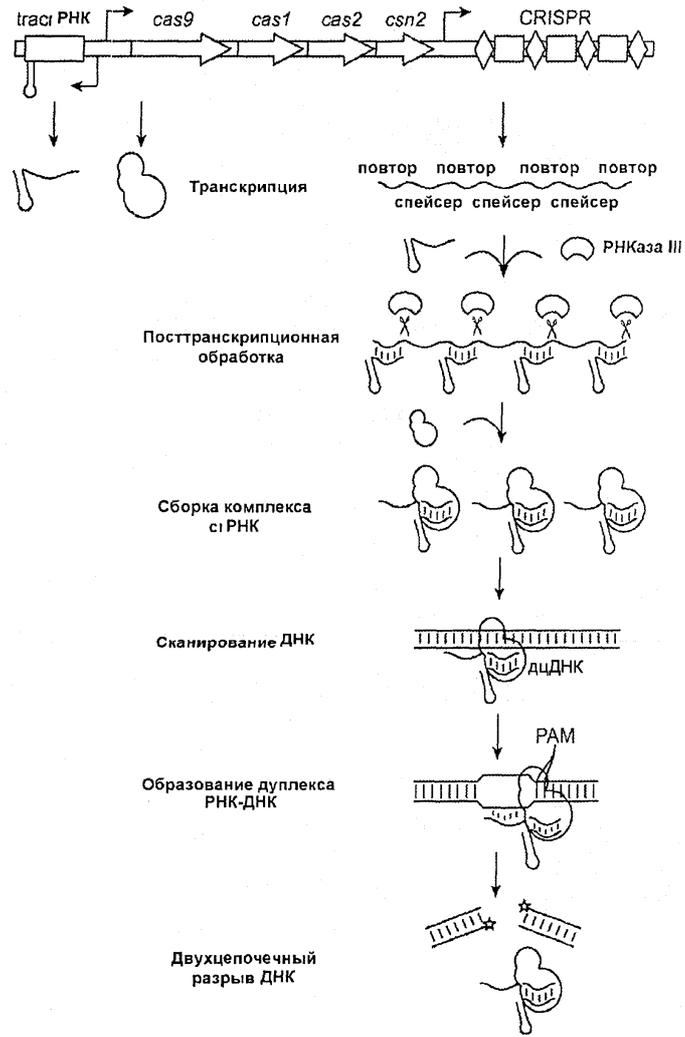




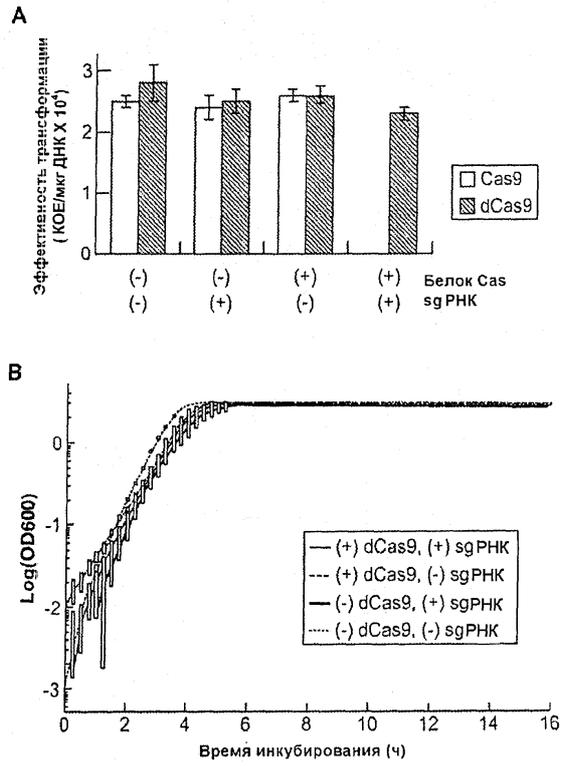
Фиг. 44



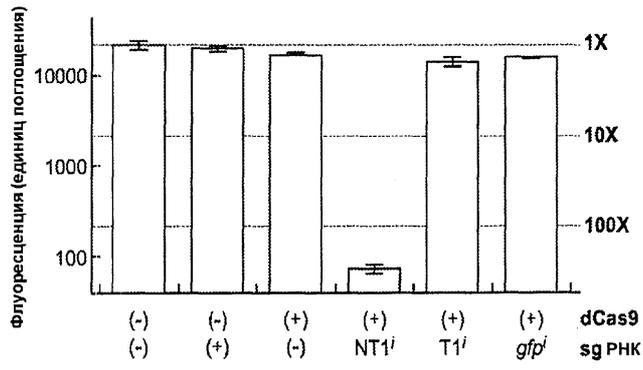
Фиг. 45



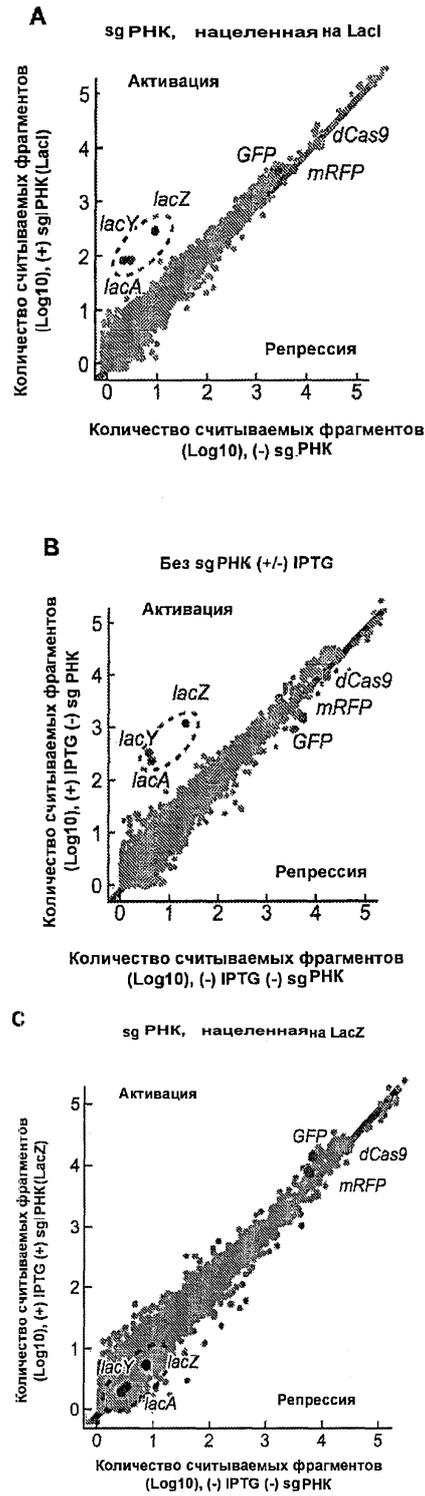
Фиг. 46



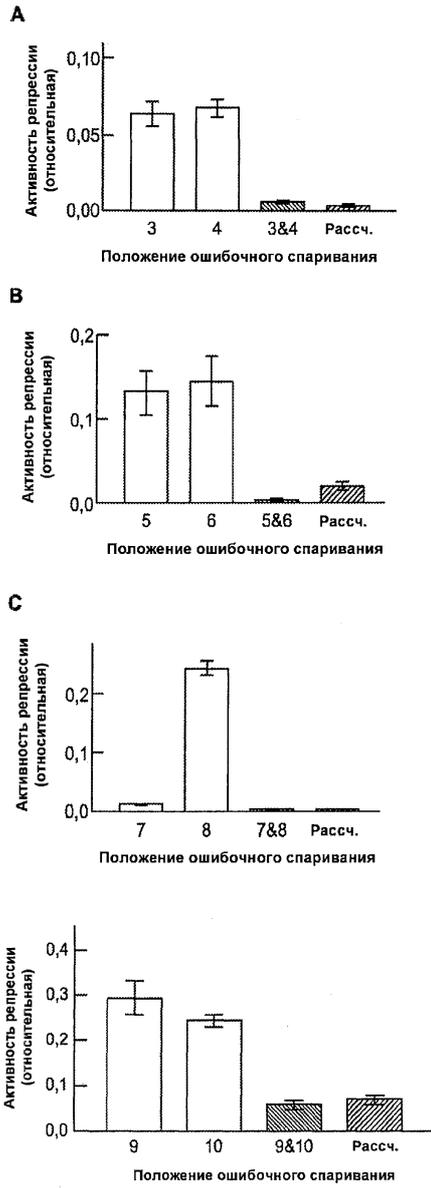
Фиг. 47



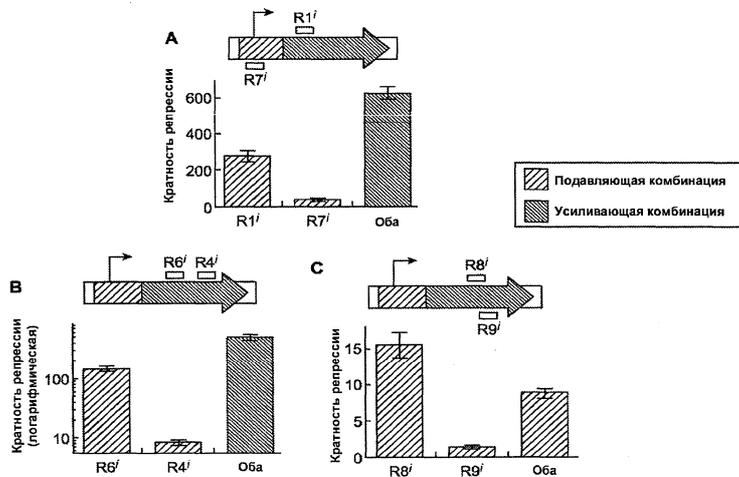
Фиг. 48



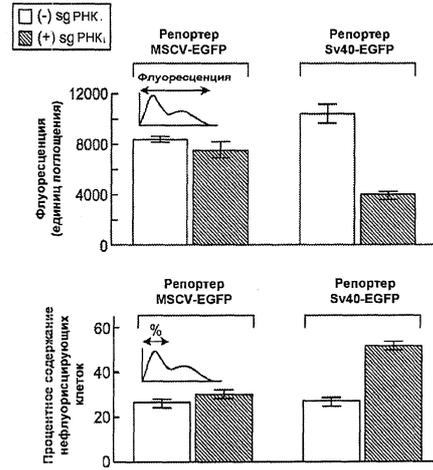
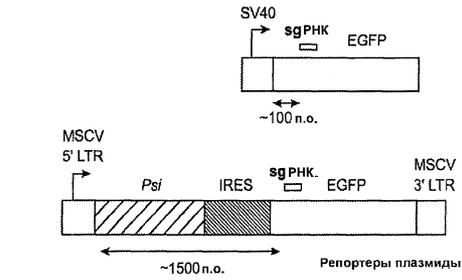
Фиг. 49



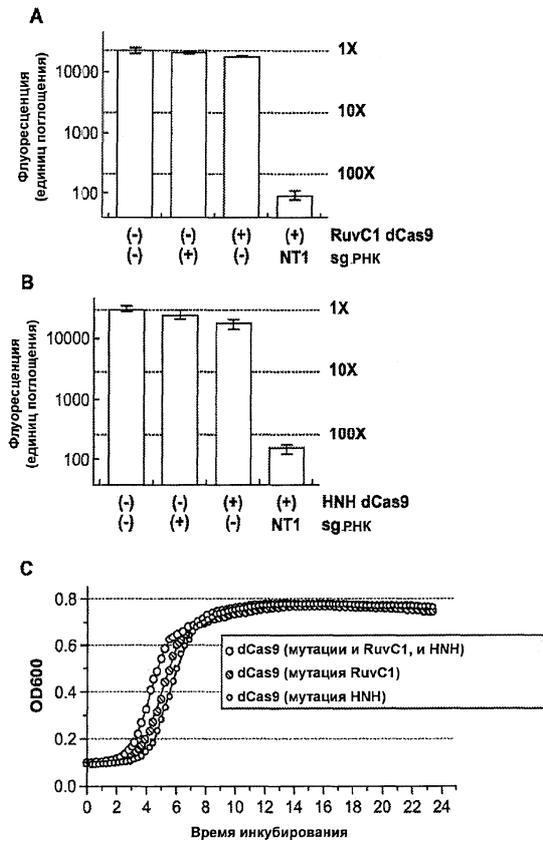
Фиг. 50



Фиг. 51



Фиг. 52



Фиг. 53

## А

Название белка	Функция
<b>Активаторы транскрипции</b>	
GAL4	Активация транскрипции
VP16	Активация транскрипции
VP64	Активация транскрипции
Субдомен р65 (NFKB)	Активация транскрипции
<b>Репрессоры транскрипции</b>	
KRAB	Репрессия транскрипции
Домен взаимодействия Mad mSin3 (SID)	Репрессия транскрипции
Домен репрессора ERF (ERD)	Репрессия транскрипции
<b>Лизинметилтрансферазы гистона (KMT)</b>	
Семейство KMT1: SUV39H1, SUV39H2, G9A, ESET/SETDB1 и гомологи (Ctcf4, Su(var) 3-9)	Образование гетерохроматина/ репрессия транскрипции
Семейство KMT2: hSET1A, hSET1B, MLL1-5, ASH 1, и гомологи (Trx, Ttr, Ash1)	Активация транскрипции
Семейство KMT3: SYMD2, NSD1	Активация транскрипции
Семейство KMT4: DOT1L и гомологи	Активация транскрипции
Семейство KMT5: Pr-SET7/8, SUV4-20H1 и гомологи (PR-set7, Suv4-20, Set9)	Ответ на повреждение DNA, репрессия транскрипции
KMT6: EZH2	Подавление Polycomb
KMT8: RIZ1	Репрессия транскрипции
<b>Лизиндеметилаты гистона (KDM)</b>	
KDM1: LSD1/BHC110 и гомологи (SpLsd1/Swm1/Saf1 10, Su(var)3-3)	Активация и репрессия транскрипции, образование гетерохроматина
Семейство KDM3: JHDM2a/b	Активация гена андрогенового рецептора, сперматогенез
Семейство KDM4: JMJD2A/JHDM3A, JMJD2B, JMJD2C/GASC1, JMJD2D и гомологи (Rph1)	Удлинение транскрипции, репрессия транскрипции, образование гетерохроматина, целостность генома
Семейство KDM5: JARID1A/RBP2, JARID1B/PLU-1, JARID1C/SMCX, JARID1D/SMCY и гомологи (Lid, Jhn2, Jmj2)	Репрессия транскрипции
Семейство KDM6: UTX, JMJD3	Активация транскрипции

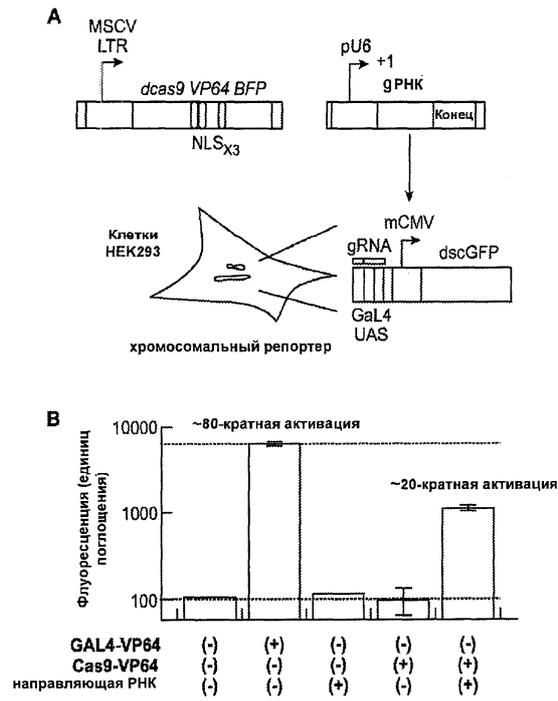
## В

Название белка	Функция
<b>Лизинацетилазы гистона (KAT)</b>	
Семейство KAT2: hGCN5, PCAF и гомологи (dGCN5/PCAF, Gcn5)	Активация транскрипции, репарация DNA
Семейство KAT3: CBP, p300 и гомологи (dCBP/NEJ)	Активация транскрипции, репарация DNA
KAT4: TAF1 и гомологи (dTAF1)	Активация транскрипции
KAT5: TIP60/PLIP и гомологи	Активация транскрипции, репарация DNA
KAT6: MOZ/MYST3, MORF/MYST4 и гомологи (Mst2, Sas3, CG1894)	Активация и удлинение транскрипции, репликация DNA
KAT7: HBO1/MYST2 и гомологи (CHM, Mst2)	Транскрипция, DNA replication
KAT8: HMOF/MYST1 и гомологи (dMOF, CG1894, Sas2, Mst2)	Границы хроматина, компенсация дозы, репарация DNA
Семейство KAT13: SRC1, ACTR, P160, CLOCK и гомологи	Активация транскрипции
<b>Лизиндеацетилазы гистона</b>	
Класс I: HDAC1, HDAC2, HDAC3, HDAC8 и его гомологи (Rpd3, Hos1, Ctr6)	Репрессия транскрипции, образование гетерохроматина
Класс IIa: HDAC4, HDAC5, HDAC7, HDAC9 и его гомологи (Hda1, Ctr3 etc.)	Репрессия транскрипции, образование гетерохроматина
Класс III: SIRT1, SIRT2 и его гомологи (Sir2, Hst1, Hst2, Hst3, Hst4)	Репрессия транскрипции, образование гетерохроматина
Класс IV: HDAC11	Репрессия транскрипции
<b>Метилазы ДНК (модификация аденозина или цитозина)</b>	
Dam (E. coli)	Система рестрикции
Dcm (E. coli)	Система рестрикции
M. SssI (Spiroplasma sp)	Система рестрикции
DNMT1	Репрессия транскрипции, импринтинг, образование гетерохроматина
DNMT3a/DNMT3b, MET1, DRM3 (растения) и гомологи	Репрессия транскрипции, импринтинг, образование гетерохроматина
Хромометилазы, напр. ZMET2, CMT1, CMT2 (растения)	Репрессия транскрипции, импринтинг, образование гетерохроматина
<b>Деметилазы ДНК</b>	
Семейство дезаминаз AID/Aobec: AID	Активация транскрипции, целостность генома
Семейство дезоксигеназ TET: TET1	Активация транскрипции, целостность генома
Семейство гликозилаз DEMETER: DME, DML1, DML2, ROS1	Активация транскрипции, целостность генома

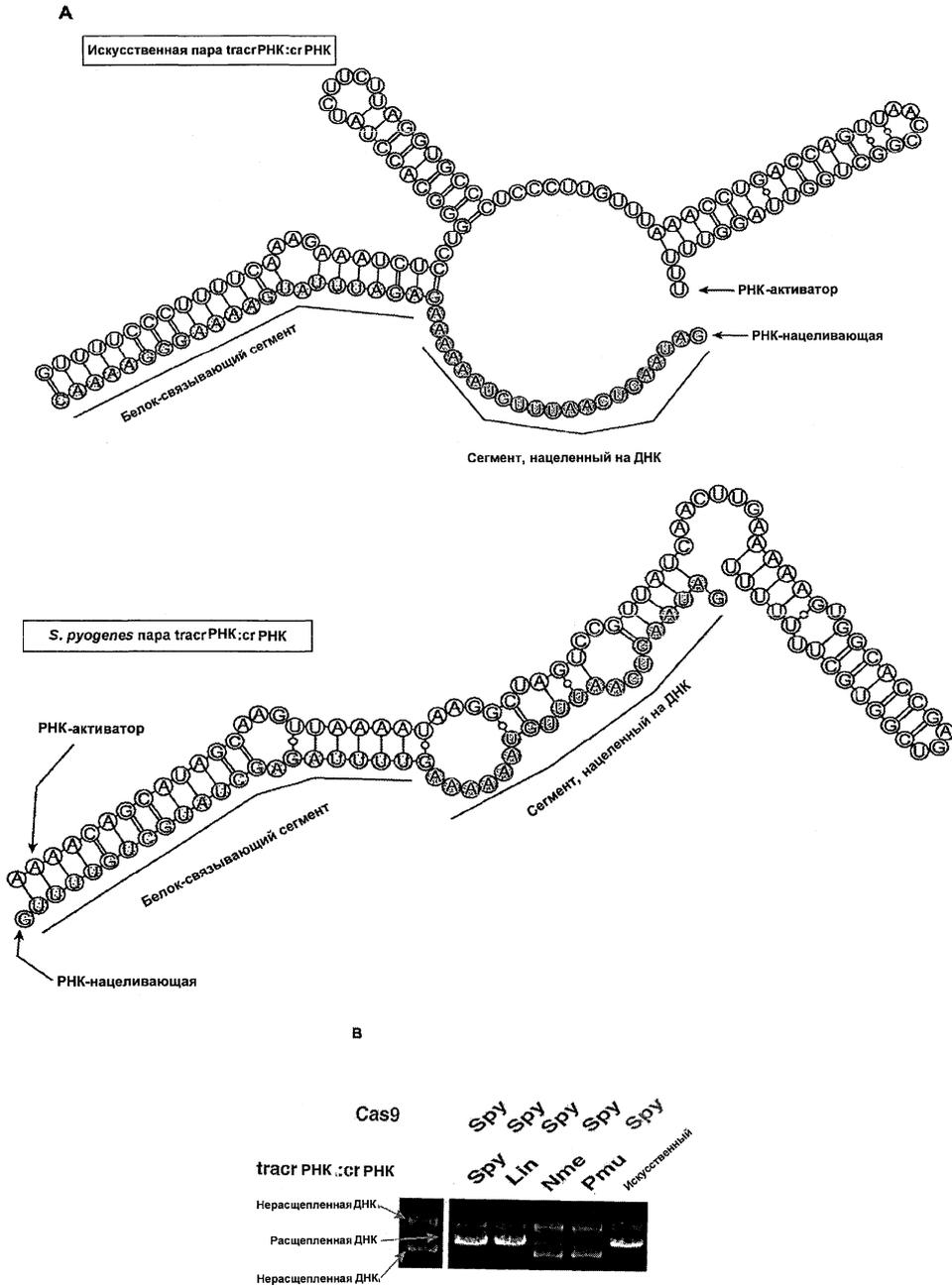
С

Название белка	Функция
<b>Связующие элементы</b>	
CTCF	Изолирование хроматина, подавление распространения гетерохроматина
<b>Периферические элементы набора</b>	
Ламин А	Репрессия транскрипции
Ламин В	Репрессия транскрипции
<b>Элементы докинга белка</b>	
FKBP/FRB ( <i>S. pombe</i> )	Рекрутирование на основе рапамицина
Рi1/Aby1 ( <i>E. coli</i> )	Рекрутирование на основе АВА

Фиг. 54







Фиг. 57

