

(19)



**Евразийское
патентное
ведомство**

(21)

202390350

(13)

A1

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ

(43) Дата публикации заявки
2023.05.22

(22) Дата подачи заявки
2021.07.15

(51) Int. Cl. *C12N 9/22* (2006.01)
A61K 38/43 (2006.01)
C07K 14/31 (2006.01)
C12N 9/78 (2006.01)
C12N 15/62 (2006.01)

(54) СТАБИЛИЗИРУЮЩИЕ УРАЦИЛ БЕЛКИ И ИХ АКТИВНЫЕ ФРАГМЕНТЫ И ВАРИАНТЫ И СПОСОБЫ ПРИМЕНЕНИЯ

(31) 63/052,175

(32) 2020.07.15

(33) US

(86) PCT/US2021/041809

(87) WO 2022/015969 2022.01.20

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

(72) Изобретатель:

**Боуэн Тайсон Д., Кроули Александра
Брайнер, Элич Тэдд Д. (US)**

(74) Представитель:

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

(57) В заявке описаны композиции и способы, включающие использование стабилизирующих урацил полипептидов для целенаправленного редактирования нуклеиновых кислот. Композиции содержат стабилизирующие урацил полипептиды. Описаны также слитые белки, содержащие I) ДНК-связывающий полипептид; II) дезаминазу; и III) стабилизирующий урацил полипептид (USP). Слитые белки включают РНК-управляемые нуклеазы, слитые с дезаминазами и дополнительно слитые с USP, необязательно в комплексе с гидовыми РНК. Композиции включают также молекулы нуклеиновых кислот, кодирующие USP или слитые белки. Описаны также векторы и клетки-хозяева, содержащие молекулы нуклеиновых кислот, кодирующие USP или слитые белки.

202390350

A1

A1

202390350

СТАБИЛИЗИРУЮЩИЕ УРАЦИЛ БЕЛКИ И ИХ АКТИВНЫЕ ФРАГМЕНТЫ И ВАРИАНТЫ И СПОСОБЫ ПРИМЕНЕНИЯ

5

Перекрестная ссылка на родственные заявки

Настоящая заявка претендует на приоритет предварительной заявки на патент США № 63/052175, поданной 15 июля 2020 г., которая в полном объеме включена в настоящее описание в качестве ссылки.

10

Сообщение, касающееся перечня последовательностей

Настоящая заявка содержит перечень последовательностей, поданный в ASCII-формате вместо бумажной копии, и он включен в спецификацию в качестве ссылки. ASCII-копия, обозначенная как L103438_1220WO_0106_5_SL.txt и имеющая размер 658586 байтов, создана 14 июля 2021 года, и представлена в электронном виде через EFS-Web.

15

Область техники, к которой относится изобретение

Настоящее изобретение относится к области молекулярной биологии и редактирования генов.

Предпосылки создания изобретения

20

Целенаправленное редактирование или модификация генома быстро становится важным инструментом фундаментальных и прикладных исследований. Первоначальные методы включали сконструированные нуклеазы, такие как мегануклеазы, слитые белки с доменом цинковых пальцев или TALEN, требующие создания химерных нуклеаз со сконструированными, программируемыми, специфическими для последовательности ДНК-связывающими доменами, которые обладают специфичностью в отношении каждой конкретной последовательности-мишени. РНК-управляемые нуклеазы (RGN), такие как ассоциированные с короткими палиндромными повторами, регулярно расположенными группами (Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) белки (Cas) бактериальной системы CRISPR-Cas, позволяют таргетировать специфические последовательности путем создания комплексов нуклеаз с гидовой РНК, которая специфически гибридизуется с конкретной последовательность-мишенью. Получение специфических для

25

30

мишени гидовых РНК является менее дорогостоящим и более эффективным, чем создание химерных нуклеаз для каждой последовательности-мишени. Указанные РНК-управляемые нуклеазы можно применять для редактирования геномов путем интродукции специфического для последовательности двухцепочечного разрыва, который репарируется с помощью подверженного ошибкам негомологичного соединения концов (NHEJ), для введения мутации в определенное местоположение в геноме.

Кроме того, RGN можно применять в подходах целенаправленного редактирования ДНК. Целенаправленное редактирование последовательностей нуклеиновых кислот, например, путем целенаправленного расщепления, позволяющего интродуцировать специфическую модификацию в геномную ДНК, обеспечивает очень тонкий подход к изучению функции генов и экспрессии генов. Такое целенаправленное редактирование можно применять также для таргетинга генетических заболеваний у людей или для введения агрономически ценных мутаций в геномы культурных растений. Создание инструментов редактирования генома обеспечивает новые подходы на основе редактирования генов к терапии млекопитающих и к агробιοтехнологиям.

Краткое изложение сущности изобретения

Предложены композиции и способы модификации молекулы ДНК-мишени. Композиции находят применение в модификации представляющей интерес молекулы ДНК-мишени. Предложенные композиции содержат стабилизирующие урацил полипептиды. Предложены также слитые белки, содержащие ДНК-связывающий полипептид, полипептид дезаминазы и стабилизирующий урацил полипептид. Предложенные композиции включают также молекулы нуклеиновых кислот, которые кодируют стабилизирующие урацил полипептиды или слитые белки, и векторы и клетки-хозяева, содержащие молекулы нуклеиновых кислот. Представленные в настоящем описании способы предназначены для связывания представляющей интерес последовательности-мишени в представляющей интерес молекуле ДНК-мишени и модификации представляющей интерес молекулы ДНК-мишени.

Подробное описание изобретения

Специалисту в области, к которой относится это изобретение, должны быть очевидны многочисленные модификации и другие варианты осуществления

изобретения, представленные в настоящем описании, которые обладают полезными свойствами, указанными в предыдущих описаниях. Таким образом, должно быть очевидно, что настоящее изобретение не ограничено описанными конкретными вариантами осуществления и подразумевается, что другие модификации и другие варианты осуществления изобретения подпадают под объем прилагаемой формулы изобретения. Хотя в настоящем описании применяют специфические понятия, их используют только в общем и описательном смысле, а не в целях ограничения.

I. Краткий обзор

В настоящем описании предложены стабилизирующие урацил полипептиды (USP), которые стабилизируют остатки урацила в молекуле ДНК и кодирующие их молекулы нуклеиновых кислот.

Целенаправленное редактирование нуклеотидных оснований, которое обозначают также как редактирование оснований, разработано Комог с соавторами в 2016 г. с использованием цитозиндезаминазы (гАРОВЕС1), функционально связанной с модифицированной РНК-управляемой нуклеазой (SpCas9) (Nature 533, ss. 420-424). В системе, описанной Комог с соавторами, гидовая РНК направляет слитый белок гАРОВЕС1-Cas9 к последовательности ДНК-мишени, в которой гАРОВЕС1 дезаминирует целевой цитозин (С) с образованием урацила (U), который обладает свойствами тимина (Т) при спаривании оснований. С использованием этой системы можно интродуцировать целенаправленные мутации С>Т в молекулу ДНК.

Основным недостатком редактирования оснований с использованием слияния гАРОВЕС1-Cas9 *in vivo* является то, что клеточная урацил-ДНК-гликозилаза (UDG) распознает гетеродуплекс U:G ДНК и катализирует удаление урацила из ДНК, образуя абасический (не содержащий азотистого основания) сайт, инициируя эксцизионную репарацию оснований с реверсией пары U:G в пару С:G в качестве наиболее распространенного результата, хотя обнаружено также образование инсерционно-делеционных мутаций (инделов). При встраивании ингибитора урацил-ДНК-гликозилазы (UGI) в слитый белок гАРОВЕС1-Cas9 урацил сохранялся в течение промежутка времени, достаточно продолжительного для того, чтобы произошла репликация, приводящая к требуемой мутации С>Т.

При создании настоящего изобретения было установлено, что путем стабилизации урацила, образованного из дезаминированного цитозина, можно предупреждать образование абасического сайта и с большей долей вероятности интродуцировать требуемую мутацию С>Т. Это достигалось путем

5 идентификации стабилизирующих урацил белков (которые обозначают также как стабилизирующие урацил полипептиды или USP).

В некоторых вариантах осуществления USP предложен в качестве компонента слитого белка, который содержит ДНК-связывающий полипептид, полипептид дезаминазы и стабилизирующий урацил полипептид. В некоторых

10 вариантах осуществления изобретения ДНК-связывающий полипептид представляет собой или имеет происхождение из мегануклеазы, слитого белка с доменом цинковых пальцев или TALEN. В некоторых вариантах осуществления изобретения ДНК-связывающий полипептид представляет собой РНК-

15 управляемую нуклеазу, такую как полипептид Cas9, которая связывается с гидовой РНК (которую обозначают также как гРНК), которая, в свою очередь, связывается с нуклеотидной последовательностью-мишенью посредством гибридизации цепей. В других вариантах осуществления изобретения USP применяют индивидуально.

В некоторых вариантах осуществления изобретения полипептид

20 дезаминазы может представлять собой домен дезаминазы, который может дазаминировать нуклеотидное основание, такое, например, как цитидин. Дезаминирование нуклеотидного основания с помощью дезаминазы может приводить к точечной мутации соответствующего остатка, которую в настоящем описании обозначают «редактирование нуклеиновой кислоты» или

25 «редактирование основания». Таким образом, слитые белки, содержащие полипептид РНК-управляемой нуклеазы (RGN) и полипептид дезаминазы, можно применять для целенаправленного редактирования нуклеотидных последовательностей.

Указанные слитые белки можно применять для целенаправленного

30 редактирования ДНК *in vitro*, например, для создания мутантных клеток. Указанные мутантные клетки могут присутствовать в растениях или животных. Указанные слитые белки можно применять также для интродукции целенаправленных мутаций, например, для коррекции генетических дефектов в

клетках млекопитающих *ex vivo*, например, в клетках, полученных из организма субъекта, которые затем реинтродуцируют в организм этого же или другого субъекта; и для интродукции целенаправленных мутаций *in vivo*, например, для коррекции генетических дефектов или интродукции дезактивирующих мутаций в ассоциированные с заболеванием гены у млекопитающего. Указанные слитые белки можно применять также для интродукции целенаправленных мутаций в растительные клетки, например, для интродукции полезных или агрономически ценных признаков или аллелей.

В контексте настоящего описания понятия «белок», «пептид» и «полипептид» используют взаимозаменяемо, и они относятся к полимеру аминокислотных остатков, сцепленных друг с другом пептидными (амидными) связями. Понятия относятся к белку, пептиду или полипептиду любого размера, структуры или функции. Как правило, белок, пептид или полипептид должен состоять по меньшей мере из трех аминокислот. Белок, пептид или полипептид может относиться к индивидуальному белку или к совокупности белков. Одну или несколько аминокислот в белке, пептиде или полипептиде можно модифицировать, например, путем добавления химической субстанции, такой как углеводная группа, гидроксильная группа, фосфатная группа, фарнезильная группа, изофарнезильная группа, группа жирных кислот, линкер для конъюгации, функционализации или другой модификации и т.д. Белок, пептид или полипептид может представлять собой также единичную молекулу или может представлять собой многомолекулярный комплекс. Белок, пептид или полипептид может представлять собой просто фрагмент встречающегося в естественных условиях белка или пептида. Белок, пептид или полипептид может представлять собой встречающийся в естественных условиях, рекомбинантный или синтетический белок, пептид или полипептид, или любую их комбинацию.

Любой из представленных в настоящем описании белков можно получать любым методом, известным в данной области. Например, белки, представленные в настоящем описании, можно получать путем рекомбинантной экспрессии и очистки белка, который является особенно пригодным для слитых белков, содержащих пептидный линкер. Методы рекомбинантной экспрессии и очистки белков хорошо известны и включают методы, описанные у Green и Sambrook, в: *Molecular Cloning: A Laboratory Manual* (4-ое изд., изд-во Cold Spring Harbor

Laboratory Press, Cold Spring Harbor, N.Y. 2012), содержание публикации в полном объеме включено в настоящее описание в качестве ссылки

II. Стабилизирующие урацил белки

В настоящем описании предложены новые стабилизирующие урацил полипептиды (USP), и они представлены в виде SEQ ID NO: 1-16. USP, указанные в настоящем описании, могут найти применение для стабилизации при необходимости урацила в молекуле ДНК.

В контексте настоящего описания понятия «стабилизирующий урацил белок», «стабилизирующий урацил полипептид» и «USP» относятся к полипептиду, обладающему стабилизирующей урацил активностью. В контексте настоящего описания понятие «стабилизирующая урацил активность» относится к способности молекулы (например, полипептида) повышать частоту мутаций по меньшей мере одного из таких оснований как цитидин, дезоксицитидин или цитозин, в тимидин, дезокситимидин или тимин в молекуле нуклеиновой кислоты с помощью дезаминазы по сравнению с частотой мутаций с помощью дезаминазы в отсутствии указанной молекулы (например, стабилизирующего урацил полипептида). Не вдаваясь в теорию или механизм действия, можно предположить, что предложенные в настоящем описании USP могут функционировать путем поддержания присутствия урацила в одноцепочечной ДНК, который образовался в результате дезаминирования цитидинового, дезоксицитидинового или цитозинового основания, в течение периода времени, достаточного для того, чтобы имели место репликация и интродукция требуемой мутации C>T. Стабилизирующая урацил активность может проявляться за счет ингибирования урацил-ДНК-гликозилазы, пути эксцизионной репарации оснований или механизмов репарации ошибочного спаривания.

В некоторых вариантах осуществления изобретения предложенные в настоящем описании USP или их активные варианты или фрагменты, которые сохраняют стабилизирующую урацил активность, повышает частоту мутаций по меньшей мере одного из таких оснований как цитидин, дезоксицитидин или цитозин, в тимидин, дезокситимидин или тимин в молекуле нуклеиновой кислоты с помощью дезаминазы по меньшей мере на 10%, по меньшей мере на 15%, по меньшей мере на 20%, по меньшей мере на 25%, по меньшей мере на 30%, по меньшей мере на 35%, по меньшей мере на 40%, по меньшей мере на

45%, по меньшей мере на 50%, по меньшей мере на 55%, по меньшей мере на 60%, по меньшей мере на 70%, по меньшей мере на 80%, по меньшей мере на 90%, по меньшей мере на 95%, по меньшей мере на 96%, по меньшей мере на 97%, по меньшей мере на 98%, по меньшей мере на 99%, по меньшей мере на 100%, по меньшей мере на 150%, по меньшей мере на 200% или более по сравнению с частотой мутаций с помощью дезаминазы в отсутствие USP. И, наоборот, частота мутаций с помощью дезаминазы по меньшей мере одного из таких оснований, как цитидин, дезоксицитидин или цитозин, в любое из нуклеотидных оснований, отличных от тимидина, дезокситимидина или тимина (т.е. гуанозина, дезоксигуанозина, гуанина, аденозина, дезоксиаденозина, аденина) в молекуле нуклеиновой кислоты снижается с помощью предложенных в настоящем описании USP или их активных вариантов или фрагментов по меньшей мере на 10%, по меньшей мере на 15%, по меньшей мере на 20%, по меньшей мере на 25%, по меньшей мере на 30%, по меньшей мере на 35%, по меньшей мере на 40%, по меньшей мере на 45%, по меньшей мере на 50%, по меньшей мере на 55%, по меньшей мере на 60%, по меньшей мере на 70%, по меньшей мере на 80%, по меньшей мере на 90%, по меньшей мере на 95%, по меньшей мере на 96%, по меньшей мере на 97%, по меньшей мере на 98%, по меньшей мере на 99% или более по сравнению с частотой мутаций с помощью дезаминазы в отсутствие USP. Повышение или снижение частоты мутаций цитидина, дезоксицитидина или цитозина в другое нуклеотидное основание можно измерять путем сравнения частоты мутаций с помощью конкретной дезаминазы в конкретное нуклеотидное основание в присутствии USP или без него. В тех вариантах осуществления изобретения, в которых дезаминазу направляют к конкретной области молекулы нуклеиновой кислоты посредством слияния с ДНК-связывающим полипептидом, частоту мутаций цитидиновых, дезоксицитидиновых или цитозиновых оснований в последовательности-мишени или по соседству с последовательностью-мишенью, с которой связывается ДНК-связывающий полипептид, можно измерять с помощью метода, известного в данной области, включая полимеразную цепную реакцию (ПЦР), полиморфизм длин рестрикционных фрагментов (RFLP) или секвенирование ДНК.

Предложенные в настоящем описании новые USP или их активные варианты или фрагменты, которые сохраняют стабилизирующую урацил

активность, можно интродуцировать в клетку в виде части слияния дезаминаза-ДНК-связывающий полипептид и/или можно совместно экспрессировать со слиянием ДНК-связывающий полипептид-дезаминаза или слиянием ДНК-связывающий полипептид-дезаминаза-USP для повышения эффективности интродукции требуемой мутации C>T в молекулу ДНК-мишени. Предложенные в настоящем описании новые USP, сохраняющие стабилизирующую урацил активность, имеют аминокислотную последовательность, представленную в любой из SEQ ID NO: 1-16, или ее вариант или фрагмент. В некоторых вариантах осуществления изобретения USP имеет аминокислотную последовательность, идентичную по меньшей мере на 50%, по меньшей мере на 55%, по меньшей мере на 60%, по меньшей мере на 65%, по меньшей мере на 70%, по меньшей мере на 75%, по меньшей мере на 80%, по меньшей мере на 81%, по меньшей мере на 82%, по меньшей мере на 83%, по меньшей мере на 84%, по меньшей мере на 85%, по меньшей мере на 86%, по меньшей мере на 87%, по меньшей мере на 88%, по меньшей мере на 89%, по меньшей мере на 90%, по меньшей мере на 91%, по меньшей мере на 92%, по меньшей мере на 93%, по меньшей мере на 94%, по меньшей мере на 95%, по меньшей мере на 96%, по меньшей мере на 97%, по меньшей мере на 98% или по меньшей мере на 99% аминокислотной последовательности, представленной в любой из SEQ ID NO: 1-16. В конкретных вариантах осуществления изобретения USP содержит аминокислотную последовательность, обладающую по меньшей мере 80%-ной идентичностью с одной из последовательностей SEQ ID NO: 1, 2, 4, 5 и 7-15. В других вариантах осуществления изобретения USP содержит аминокислотную последовательность, обладающую по меньшей мере 81%-ной идентичностью с SEQ ID NO: 3 или 16. В следующих вариантах осуществления изобретения USP содержит аминокислотную последовательность, обладающую по меньшей мере 82%-ной идентичностью с SEQ ID NO: 6.

III. Слитые белки

В некоторых объектах настоящего изобретения предложены слитые белки, которые содержат ДНК-связывающий полипептид и полипептид дезаминазы, и в некоторых вариантах осуществления изобретения полипептид USP. Указанные слитые белки можно применять для целенаправленного редактирования ДНК *in vitro*, *ex vivo* или *in vivo*.

В контексте настоящего описания понятие «слитый белок» относится к гибриднему полипептиду, который содержит домены по меньшей мере из двух различных белков. Слитый белок может содержать различные домены, например, ДНК-связывающий домен и дезаминазу. В некоторых вариантах осуществления изобретения слитый белок представляет собой комплекс или ассоциацию с нуклеиновой кислотой, такой, например, как РНК.

Полипептид дезаминазы содержит домен дезаминазы, который обладает способностью дезаминировать нуклеотидное основание, такое, например, как цитидин. Дезаминирование нуклеотидного основания с помощью дезаминазы может приводить к точечной мутации в соответствующем остатке, что в контексте настоящего описания обозначают как «редактирование нуклеиновой кислоты» или «редактирование оснований». Таким образом, слитые белки, содержащие вариант или домен полипептида RGN и домен дезаминазы, можно применять для целенаправленного редактирования нуклеотидных

последовательностей. В некоторых вариантах осуществления изобретения дезаминаза содержит аминокислотную последовательность, идентичную по меньшей мере на 80%, по меньшей мере на 81%, по меньшей мере на 82%, по меньшей мере на 83%, по меньшей мере на 84%, по меньшей мере на 85%, по меньшей мере на 86%, по меньшей мере на 87%, по меньшей мере на 88%, по меньшей мере на 89%, по меньшей мере на 90%, по меньшей мере на 91%, по меньшей мере на 92%, по меньшей мере на 93%, по меньшей мере на 94%, по меньшей мере на 95%, по меньшей мере на 96%, по меньшей мере на 97%, по меньшей мере на 98%, по меньшей мере на 99% или более аминокислотной последовательности, представленной в одной из SEQ ID NO 47, 48 и 76-94. В некоторых вариантах осуществления изобретения дезаминаза содержит аминокислотную последовательность, обладающую по меньшей мере 80%-ной идентичностью с одной из последовательностей SEQ ID NO: 47, 48 и 76-94. В некоторых вариантах осуществления изобретения дезаминаза содержит аминокислотную последовательность, обладающую по меньшей мере 85%-ной идентичностью с одной из последовательностей SEQ ID NO: 47, 48 и 76-94. В некоторых вариантах осуществления изобретения дезаминаза содержит аминокислотную последовательность, обладающую по меньшей мере 90%-ной идентичностью с одной из последовательностей SEQ ID NO: 47, 48 и 76-94. В

некоторых вариантах осуществления изобретения дезаминаза содержит аминокислотную последовательность, обладающую по меньшей мере 95%-ной идентичностью с одной из последовательностей SEQ ID NO: 47, 48 и 76-94. В других вариантах осуществления изобретения дезаминаза содержит

5 аминокислотную последовательность, обладающую по меньшей мере 99%-ной идентичностью с одной из последовательностей SEQ ID NO: 47, 48 и 76-94. В конкретных вариантах осуществления изобретения дезаминаза содержит аминокислотную последовательность, представленную в одной из SEQ ID NO: 47, 48 и 76-94.

10 Предложенные в настоящем описании слитые белки содержат ДНК-связывающий полипептид. В контексте настоящего описания понятие «ДНК-связывающий полипептид» относится к любому полипептиду, который обладает способностью связываться с ДНК. В конкретных вариантах осуществления изобретения компонент предложенных в настоящем описании слитых белков,

15 представляющий собой ДНК-связывающий полипептид, связывается с двухцепочечной ДНК. В конкретных вариантах осуществления изобретения ДНК-связывающий полипептид связывается с ДНК специфическим для последовательности образом. В контексте настоящего описания понятия «специфическим для последовательности» или «специфическим для

20 последовательности образом» относятся к избирательному взаимодействию со специфической нуклеотидной последовательностью.

Две полинуклеотидные последовательности можно считать практически комплементарными, если две последовательности гибридизуются друг с другом в строгих условиях. Аналогично этому, считается, что ДНК-связывающий

25 полипептид связывается с конкретной последовательностью-мишенью специфическим для последовательности образом, если ДНК-связывающий полипептид связывается с этой последовательностью в строгих условиях. Под «строгими условиями» или «строгими условиями гибридизации» подразумевают условия, в которых две полинуклеотидные последовательности (или полипептид,

30 который связывается со специфической последовательностью-мишенью) будут связываться друг с другом в заметно большей степени, чем с другими последовательностями (например, по меньшей мере в 2 раза по сравнению с фоном). Строгие условия зависят от последовательности и могут отличаться в

различных обстоятельствах. Как правило, строгие условия должны представлять собой условия, в которых концентрация соли составляет менее чем примерно 1,5М в пересчете на ион Na, как правило, концентрация составляет примерно от 0,01 до 1,0М в пересчете на ион Na (или других солей) при pH от 7,0 до 8,3, и температура составляет по меньшей мере примерно 30°C для коротких последовательностей (например, из 10-50 нуклеотидов) и по меньшей мере примерно 60°C для длинных последовательностей (например, состоящих более чем из 50 нуклеотидов). Для достижения строгих условий можно также добавлять дестабилизирующие агенты, такие как формамид. Пример гибридизации в расслабленных условиях (в условиях пониженной жесткости) включает гибридизацию в буферном растворе, содержащем 30-35% формамида, 1М NaCl, 1% ДСН (додecilсульфат натрия) при 37°C, и промывку от 1× до 2× SSC (20× SSC = 3,0М NaCl/0,3М тринатрий цитрат) при 50-55°C. Пример гибридизации в условиях умеренной жесткости включает гибридизацию в 40-45% формамида, 1,0М NaCl, 1% ДСН при 37°C и промывку в от 0,5× до 1× SSC при 55-60°C. Пример гибридизации в очень строгих условиях включает гибридизацию в 50% формамида, 1М NaCl, 1% ДСН при 37°C и промывку в 0,1× SSC при 60-65°C. Необязательно промывочные буферы могут содержать от примерно 0,1% до примерно 1% ДСН. Продолжительность гибридизации, как правило, составляет менее чем примерно 24 ч, обычно от примерно 4 до примерно 12 ч. Продолжительность периода промывки должна по меньшей мере быть достаточной для достижения равновесия.

T_m представляет температуру (при определенной ионной силе и величине pH), при которой 50% комплементарных последовательностей-мишеней гибридизуются с идеально соответствующей последовательностью. Для гибридов ДНК-ДНК T_m можно аппроксимировать из уравнения Meinkoth и Wahl, Anal. Biochem. 138, 1984, сс. 267-284: $T_m = 81,5^\circ\text{C} + 16,6 (\log M) + 0,41 (\% \text{GC}) - 0,61 (\% \text{form.}) - 500/L$; в котором M обозначает молярность одновалентных катионов, %GC обозначает процент гуанозиновых и цитозиновых нуклеотидов в ДНК, % form обозначает процент формамида в растворе для гибридизации, и L обозначает длину гибрида в парах оснований. Как правило, строгие условия выбирают такими, чтобы они были примерно на 5°C ниже температуры плавления (T_m) для конкретной последовательности и ее

комплемента при определенной ионной силе и pH. Однако в случае очень жестких условий можно использовать гибридизацию и/или промывку при температуре на 1, 2, 3 или 4°C ниже, чем температура плавления (T_m); в случае умеренных условий можно использовать гибридизацию и/или промывку при температуре на 6, 7, 8, 9 или 10°C ниже, чем температура плавления (T_m); в случае ослабленных условий можно использовать гибридизацию и/или промывку при температуре на 11, 12, 13, 14, 15 или 20°C ниже, чем температура плавления (T_m). С учетом уравнения, композиций для гибридизации и промывки и требуемой T_m , обычным специалистам в данной области должны быть очевидны возможные вариации в строгости гибридизации и/или промывочных растворов. Обширное руководство по гибридизации нуклеиновых кислот представлено у Tijssen, *Laboratory Techniques in Biochemistry and Molecular Biology—Hybridization with Nucleic Acid Probes*, часть I, глава 2, 1995 (изд-во Elsevier, New York); и в: *Current Protocols in Molecular Biology*, глава 2, 1995, под ред. Ausubel и др. (изд-во Greene Publishing and Wiley-Interscience, New York) (см. Sambrook и др. *Molecular Cloning: A Laboratory Manual*, 1989 (2-ое изд., изд-во Cold Spring Harbor Laboratory Press, Plainview, New York)).

В конкретных вариантах осуществления изобретения специфический для последовательности ДНК-связывающий полипептид представляет собой РНК-управляемый ДНК-связывающий полипептид (RGDBP). В контексте настоящего описания понятия «РНК-управляемый ДНК-связывающий полипептид» и «RGDBP» относятся к полипептидам, которые обладают способностью связываться с ДНК посредством гибридизации ассоциированной молекулы РНК с последовательностью ДНК-мишени.

В некоторых вариантах осуществления изобретения ДНК-связывающий полипептид слитого белка представляет собой нуклеазу, такую как специфическая для последовательности нуклеаза. В контексте настоящего описания понятие «нуклеаза» относится к ферменту, который катализирует расщепление фосфодиэфирных связей между нуклеотидами в молекуле нуклеиновой кислоты. В некоторых вариантах осуществления изобретения ДНК-связывающий полипептид представляет собой эндонуклеазу, которая обладает способностью расщеплять фосфодиэфирные связи между нуклеотидами внутри молекулы нуклеиновой кислоты, в то время как в других вариантах

осуществления изобретения ДНК-связывающий полипептид представляет собой экзонуклеазу, которая обладает способностью расщеплять нуклеотиды на любом конце (5' или 3') молекулы нуклеиновой кислоты. В некоторых вариантах осуществления изобретения специфическую для последовательности нуклеазу выбирают из группы, которая состоит из мегануклеазы, нуклеазы с цинковыми пальцами, слитого белка, содержащего эффекторный ДНК-связывающий домен TAL - нуклеазу (TALEN), и РНК-управляемую нуклеазу (RGN) или их вариантов, нуклеазная активность которых снижена или заингибирована.

В контексте настоящего описания понятие «мегануклеаза» или «хоминг-эндонуклеаза» относится к эндонуклеазам, которые связываются с распознаваемым сайтом в двухцепочечной ДНК, имеющим длину от 12 до 40 пар оснований (bp). Примерами мегануклеаз являются (но не ограничиваясь только ими) ферменты, принадлежащие к семейству LAGLIDADG, которое содержит консервативный аминокислотный мотив LAGLIDADG (SEQ ID NO: 75). Понятие «мегануклеаза» может относиться к димерной или одноцепочечной мегануклеазе.

В контексте настоящего описания понятие «нуклеаза с цинковыми пальцами» или «ZFN» относится к химерному белку, содержащему ДНК-связывающий домен с цинковыми пальцами и нуклеазный домен.

В контексте настоящего описания понятие «слитый белок, содержащий эффекторный ДНК-связывающий домен TAL - нуклеазу» или «TALEN» относится к химерному белку, который содержит эффекторный ДНК-связывающий домен TAL и домен нуклеазы.

В контексте настоящего описания понятие «РНК-управляемая нуклеаза» или «RGN» относится к РНК-управляемому ДНК-связывающему полипептиду, который обладает нуклеазной активностью. RGN рассматриваются как «РНК-управляемые», поскольку гидовая РНК образует комплекс с РНК-управляемой нуклеазой для того, чтобы «принуждать» РНК-управляемые нуклеазы связываться с последовательностью-мишенью, и в некоторых вариантах осуществления изобретения интродуцировать одноцепочечный или двухцепочечный разрыв в последовательности-мишени.

Примеры RGN, которые можно применять в предложенных в настоящем описании композициях и способах, включают (но не ограничиваясь только ими)

примеры, описанные в следующих публикациях: WO 2020/139783, WO 2019/236566, WO 2021/030344, WO/2021/138247 и в заявках на патент PCT/US2021/028843 и PCT/US2021/031794, поданных 23 апреля 2021 года и 11 мая 2021 года соответственно, каждая из которых в полном объеме включена в

5 настоящее описание в качестве ссылки. В некоторых вариантах осуществления изобретения предлагаемый в настоящем описании слитый белок содержит RGN, содержащую аминокислотную последовательность, которая идентична по меньшей мере на 80%, по меньшей мере на 81%, по меньшей мере на 82%, по меньшей мере на 83%, по меньшей мере на 84%, по меньшей мере на 85%, по

10 меньшей мере на 86%, по меньшей мере на 87%, по меньшей мере на 88%, по меньшей мере на 89%, по меньшей мере на 90%, по меньшей мере на 91%, по меньшей мере на 92%, по меньшей мере на 93%, по меньшей мере на 94%, по меньшей мере на 95%, по меньшей мере на 96%, по меньшей мере на 97%, по меньшей мере на 98%, по меньшей мере на 99% или более одной из SEQ ID NO:

15 40 и 95-142. В некоторых вариантах осуществления изобретения предлагаемый в настоящем описании слитый белок содержит RGN, содержащую аминокислотную последовательность, которая идентична по меньшей мере на 80% одной из SEQ ID NO: 40 и 95-142. В некоторых вариантах осуществления изобретения предлагаемый в настоящем описании слитый белок содержит RGN,

20 содержащую аминокислотную последовательность, которая идентична по меньшей мере на 85% одной из SEQ ID NO: 40 и 95-142. В некоторых вариантах осуществления изобретения предлагаемый в настоящем описании слитый белок содержит RGN, содержащую аминокислотную последовательность, которая идентична по меньшей мере на 90% одной из SEQ ID NO: 40 и 95-142. В

25 некоторых вариантах осуществления изобретения предлагаемый в настоящем описании слитый белок содержит RGN, содержащую аминокислотную последовательность, которая идентична по меньшей мере на 95% одной из SEQ ID NO: 40 и 95-142. В других вариантах осуществления изобретения предлагаемый в настоящем описании слитый белок содержит RGN, содержащую

30 аминокислотную последовательность, которая идентична по меньшей мере на 99% одной из SEQ ID NO: 40 и 95-142. В некоторых вариантах осуществления изобретения предлагаемый в настоящем описании слитый белок содержит RGN,

содержащую аминокислотную последовательность, которая представлена в одной из SEQ ID NO: 40 и 95-142.

Согласно настоящему изобретению белок RGN, в котором в результате мутации нуклеаза стала неактивной или «убитой», как, например, в случае dCas9, в настоящем описании обозначают как РНК-управляемый ДНК-связывающий полипептид. Одним из примеров пригодного Cas9-домена с инактивированной нуклеазной активностью является мутантный домен D10A/H840A Cas9 (см., например, Qi и др., Cell. 152(5), 2013, сс. 1173-1183, содержание публикации в полном объеме включено в настоящее описание в качестве ссылки). Кроме того, можно определять пригодные Cas9-домены с инактивированной нуклеазной активностью других известных РНК-управляемых нуклеаз (RGN) (например, вариант RGN APG08290.1 с инактивированной нуклеазной активностью, который описан в публикации патента США № 2019/0367949, содержание которого в полном объеме включено в настоящее описание в качестве ссылки).

Понятие «полипептид RGN» включает полипептиды RGN, расщепляющие только одну цепь нуклеотидной последовательности-мишени, которые обозначают в настоящем описании как никазы. Указанные RGN имеют единственный функционирующий нуклеазный домен. RGN, представляющие собой никазы (RGN-никазы) могут представлять собой встречающиеся в естественных условиях никазы или могут представлять собой белки RGN, которые в естественных условиях расщепляют обе цепи двухцепочечной молекулы нуклеиновой кислоты, с измененными в результате мутации в дополнительных нуклеазных доменах, в результате чего нуклеазная активность этих мутантных доменов уменьшается или элиминируется, что превращает эти белки в никазу. В некоторых вариантах осуществления изобретения RGN-никаза в слитом белке содержит мутацию D10A (например, nAPG07433.1 (SEQ ID NO: 41)), что придает RGN способность расщеплять только цепь-мишень с неотредактированными основаниями (цепь, которая содержит PAM (примыкающий к протоспейсеру мотив) и соединена путем спаривания оснований с гРНК) дуплекса нуклеиновой кислоты. В некоторых вариантах осуществления изобретения RGN-никаза в слитом белке содержит мутацию D10A или эквивалентную ей мутацию в одной из SEQ ID NO: 40 и 95-142. В

некоторых вариантах осуществления изобретения RGN-никаза в слитом белке содержит мутацию H840A, которая придает RGN способность расщеплять только нецелевую цепь с отредактированными основаниями (цепь, которая не содержит РАМ и не соединена путем спаривания оснований с гРНК) дуплекса нуклеиновой кислоты. RGN-никаза, содержащая мутацию H840A или эквивалентную мутацию, имеет инактивированный HNH-домен. RGN-никаза, содержащая мутацию D10A или эквивалентную мутацию, имеет инактивированный RuvC-домен. Дезаминаза действует на нецелевую цепь. Никаза, содержащая мутацию D10A или эквивалентную мутацию, имеет инактивированный нуклеазный RuvC-домен и не обладает способностью расщеплять нецелевую цепь ДНК, т.е. цепь, в которой желательно редактирование оснований.

Кроме того, дополнительные примеры приемлемых Cas9-доменов с инактивированной нуклеазной активностью включают (но не ограничиваясь только ими) мутантные домены D10A/D839A/H840A и D10A/D839A/H840A/N863A (см., например, Mali и др., Nature Biotechnology, 31(9), 2013, сс. 833-838, содержание публикации в полном объеме включено в настоящее описание в качестве ссылки). Другие белки RGN, приемлемые для превращения в результате мутации в никазы, должны стать очевидны специалистам в данной области на основе настоящего описания и знаний в данной области (такие, например, как RGN, описанные в публикации PCT WO 2019/236566), и они подпадают под объем настоящего изобретения.

В некоторых вариантах осуществления изобретения RGN-никаза, сохраняющая никазную активность, содержит аминокислотную последовательность, которая по меньшей мере на 60%, по меньшей мере на 65%, по меньшей мере на 70%, по меньшей мере на 75%, по меньшей мере на 80%, по меньшей мере на 85%, по меньшей мере на 90%, по меньшей мере на 95%, по меньшей мере на 96%, по меньшей мере на 97%, по меньшей мере на 98%, по меньшей мере на 99% или по меньшей мере на 99,5% идентична nAPG07433.1 (SEQ ID NO: 41).

Любой известный в данной области метод интродукции мутаций в аминокислотную последовательность, такой как ПЦР-опосредуемый мутагенез или сайтнаправленный мутагенез, можно применять для создания никаза или

RGN с «убитой, (лишенной ферментативной активности)» нуклеазной активностью (см., например, публикацию патента США № 2014/0068797 и патент США № 9790490; каждый из которых в полном объеме включен в настоящее описание в качестве ссылки). РНК-управляемые нуклеазы (RGN) позволяют целенаправленно манипулировать с одним участком в геноме и являются ценными в контексте таргетинга генов для терапевтических и исследовательских применений. В различных организмах, включая млекопитающих, РНК-управляемые нуклеазы нашли применение в генной инженерии путем стимулирования либо негомологичного соединения концов, либо гомологичной рекомбинации. RGN включают белки CRISPR-Cas, которые представляют собой РНК-управляемые нуклеазы, направляемые к последовательности-мишени с помощью гидовой РНК (гРНК), в качестве компонента системы ассоциированной с короткими палиндромными повторами, регулярно расположенными группами (CRISPR) РНК-управляемой нуклеазы, или их варианты или фрагменты.

Некоторыми объектами настоящего описания являются слитые белки, которые содержат РНК-управляемый ДНК-связывающий полипептид, полипептид дезаминазы и USP. В некоторых вариантах осуществления изобретения РНК-управляемый ДНК-связывающий полипептид представляет собой РНК-управляемую нуклеазу. В других вариантах осуществления изобретения РНК-управляемая нуклеаза представляет собой встречающийся в естественных условиях белок CRISPR-Cas или его активные варианты или фрагменты. Системы CRISPR-Cas подразделяют на системы класса I или класса II. Системы класса II содержат единую эффекторную нуклеазу и включают типы II, V и VI. Каждый класс подразделяют на типы (типы I, II, III, IV, V, VI), а некоторые типы дополнительно подразделяют на подтипы (например, тип II-A, тип II-B, тип II-C, тип V-A, тип V-B).

В некоторых вариантах осуществления изобретения белок CRISPR-Cas представляет собой встречающийся в естественных условиях белок CRISPR-Cas типа II или его активный вариант или фрагмент. В контексте настоящего описания понятие «белок CRISPR-Cas типа II», «эффекторный белок CRISPR-Cas типа II» или «Cas9» относится к эффекторному белку CRISPR-Cas, для которого требуется транс-активирующая РНК (tracrRNA), и он содержит два

нуклеазных домена (RuvC и HNH), каждый из которых ответствен за расщепление одной цепи двухцепочечной молекулы ДНК.

В других вариантах осуществления изобретения белок CRISPR-Cas представляет собой встречающийся в естественных условиях белок CRISPR-Cas типа V или его активный вариант или фрагмент. В контексте настоящего описания понятие «белок CRISPR-Cas типа V», «эффекторный белок CRISPR-Cas типа V» или «Cas12» относится к эффекторному белку CRISPR-Cas, который расщепляет dsДНК и содержит один нуклеазный домен RuvC или расщепленный нуклеазный домен RuvC, и в нем отсутствует HNH-домен (Zetsche и др., 2015, Cell doi:10.1016/j.cell.2015.09.038; Shmakov и др., 2017, Nat Rev Microbiol doi:10.1038/nrmicro.2016.184; Yan и др., 2018, Science doi:10.1126/science.aav7271; Harrington и др., 2018, Science doi:10.1126/science.aav4294). Следует отметить, что Cas12a обозначают также как Cpf1, и для него не требуется tracrPHK, хотя для других белков CRISPR-Cas типа V, таких как Cas12b, tracrPHK требуется. Для большинства эффекторных белков типа V мишенью может являться также ssДНК (одноцепочечная ДНК), при этом, часто не требуется PAM (Zetsche и др., 2015; Yan и др., 2018; Harrington и др., 2018). Под понятие «белок CRISPR-Cas типа V» подпадают уникальные RGN, которые содержат расщепленные нуклеазные домены RuvC, например, описанные в предварительной заявке на патент США № 62/955014, поданной 30 декабря 2019 г., содержание которой в полном объеме включено в настоящее описание в качестве ссылки.

В следующих вариантах осуществления изобретения белок CRISPR-Cas представляет собой встречающийся в естественных условиях белок CRISPR-Cas типа VI или его активный вариант или фрагмент. В контексте настоящего описания понятие «белок CRISPR-Cas типа VI», «эффекторный белок CRISPR-Cas типа VI» или «Cas13» относится к эффекторным белкам CRISPR-Cas, для которых не требуется tracrPHK и которые содержат два HEPN-домена, расщепляющие РНК. Понятие «гидовая РНК» относится к нуклеотидной последовательности, обладающей комплементарностью с нуклеотидной последовательностью-мишенью, достаточной для того, чтобы гибридизоваться с последовательностью-мишенью и направлять специфическое для последовательности связывание ассоциированной RGN с нуклеотидной

последовательностью-мишенью. Для CRISPR-Cas RGN соответствующая гидовая РНК представляет собой одну или несколько молекул РНК (как правило, одну или две), которые могут связываться с RGN и управляют связыванием RGN с конкретной нуклеотидной последовательностью-мишенью, и в тех случаях, когда RGN обладает нуклеазной активностью, расщеплять также нуклеотидную последовательность-мишень. Гидовая РНК содержит CRISPR РНК (crРНК) и в некоторых вариантах осуществления изобретения транс-активирующую CRISPR РНК (tracrРНК).

CRISPR РНК содержит спейсерную последовательность и последовательность повторов CRISPR. «Спейсерная последовательность» представляет собой нуклеотидную последовательность, которая непосредственно гибридизуется с представляющей интерес нуклеотидной последовательностью-мишенью. Спейсерную последовательность конструируют так, чтобы она была полностью или частично комплементарной представляющей интерес последовательности-мишени. В различных вариантах осуществления изобретения спейсерная последовательность может содержать от примерно 8 нуклеотидов до примерно 30 нуклеотидов или более. Например, спейсерная последовательность может содержать примерно 8, примерно 9, примерно 10, примерно 11, примерно 12, примерно 13, примерно 14, примерно 15, примерно 16, примерно 17, примерно 18, примерно 19, примерно 20, примерно 21, примерно 22, примерно 23, примерно 24, примерно 25, примерно 26, примерно 27, примерно 28, примерно 29, примерно 30 или более нуклеотидов. В некоторых вариантах осуществления изобретения спейсерная последовательность содержит от примерно 10 до примерно 26 нуклеотидов или от примерно 12 до примерно 30 нуклеотидов. В конкретных вариантах осуществления изобретения спейсерная последовательность содержит примерно 30 нуклеотидов. В некоторых вариантах осуществления изобретения степень комплементарности между спейсерной последовательностью и соответствующей ей последовательностью-мишенью при оптимальном выравнивании с использованием приемлемого алгоритма сравнительного анализа первичной структуры составляет примерно или более чем примерно 50%, примерно 60%, примерно 70%, примерно 75%, примерно 80%, примерно 81%, примерно 82%, примерно 83%, примерно 84%, примерно 85%, примерно 86%, примерно 87%,

примерно 88%, примерно 89%, примерно 90%, примерно 91%, примерно 92%,
примерно 93%, примерно 94%, примерно 95%, примерно 96%, примерно 97%,
примерно 98%, примерно 99% или более. В конкретных вариантах
5 осуществления изобретения спейсерная последовательность не имеет вторичной
структуры, что можно предсказывать с помощью любого приемлемого
алгоритма укладки полинуклеотидов, включая (но не ограничиваясь только ими)
mFold (см., например, Zuker и Stiegler *Nucleic Acids Res.* 9, 1981, сс. 133-148) и
RNAfold (см., например, Gruber и др., *Cell* 106(1), 2008, сс. 23-24).

Последовательность повторов CRISPR РНК содержит нуклеотидную
10 последовательность, которая сама по себе или совместно с гибридизованной с
ней tracrРНК образует структуру, которая распознается молекулой RGN. В
различных вариантах осуществления изобретения последовательность повторов
CRISPR РНК может содержать от примерно 8 нуклеотидов до примерно 30
нуклеотидов или более. Например, последовательность повторов CRISPR РНК
15 может содержать примерно 8, примерно 9, примерно 10, примерно 11, примерно
12, примерно 13, примерно 14, примерно 15, примерно 16, примерно 17,
примерно 18, примерно 19, примерно 20, примерно 21, примерно 22, примерно
23, примерно 24, примерно 25, примерно 26, примерно 27, примерно 28,
примерно 29, примерно 30 или более нуклеотидов. В некоторых вариантах
20 осуществления изобретения степень комплементарности между
последовательностью повторов CRISPR и соответствующей
последовательностью tracrРНК при оптимальном выравнивании с
использованием приемлемого алгоритма сравнительного анализа первичной
структуры составляет примерно или более чем примерно 50%, примерно 60%,
25 примерно 70%, примерно 75%, примерно 80%, примерно 81%, примерно 82%,
примерно 83%, примерно 84%, примерно 85%, примерно 86%, примерно 87%,
примерно 88%, примерно 89%, примерно 90%, примерно 91%, примерно 92%,
примерно 93%, примерно 94%, примерно 95%, примерно 96%, примерно 97%,
примерно 98%, примерно 99% или более.

30 В некоторых вариантах осуществления изобретения гидовая РНК
дополнительно содержит молекулу tracrРНК. Молекула транс-активирующей
РНК CRISPR или tracrРНК содержит нуклеотидную последовательность,
содержащую область, которая обладает комплементарностью, достаточной для

гибридизации с последовательностью повторов CRISPR crPHK, которую в контексте настоящего описания обозначают также как область «анти-повтора». В некоторых вариантах осуществления изобретения молекула tracrPHK содержит также область, имеющую вторичную структуру (например, типа стебель-петля),
5 или образует вторичную структуру при гибридизации с соответствующей ей crPHK. В конкретных вариантах осуществления изобретения область tracrPHK, которая полностью или частично комплементарна последовательности повторов CRISPR, находится на 5'-конце молекулы, и 3'-конец tracrPHK содержит вторичную структуру. Указанная область, имеющая вторичную структуру, как
10 правило, содержит несколько структур типа шпильки, включая центральную (nexus) шпильку, примыкающую к последовательности «анти-повтора». Часто концевые шпильки, которые присутствуют на 3'-конце tracrPHK, могут варьироваться по структуре и количеству, но они часто содержат богатую GC шпильку Rho-независимого терминатора транскрипции, за которой следует
15 цепочка остатков U на 3'-конце (см., например, Briner и др., Molecular Cell 56, 2014, сс. 333-339, Briner и Barrangou, 2016 Cold Spring Harb Protoc; doi: 10.1101/pdb.top090902 и публикацию патента США № 2017/0275648, каждый из документов в полном объеме включен в настоящее описание в качестве ссылки).

В различных вариантах осуществления изобретения область «анти-повтора» tracrPHK, которая полностью или частично комплементарна
20 последовательности повтора CRISPR, содержит от примерно 6 нуклеотидов до примерно 30 нуклеотидов или более. Например, область спаривания оснований между последовательностью «анти-повтора» tracrPHK и последовательностью повтора CRISPR может иметь примерно 6, примерно 7, примерно 8, примерно 9,
25 примерно 10, примерно 11, примерно 12, примерно 13, примерно 14, примерно 15, примерно 16, примерно 17, примерно 18, примерно 19, примерно 20, примерно 21, примерно 22, примерно 23, примерно 24, примерно 25, примерно 26, примерно 27, примерно 28, примерно 29, примерно 30 или более нуклеотидов. В конкретных вариантах осуществления изобретения область
30 «анти-повтора» tracrPHK, которая полностью или частично комплементарна последовательности повтора CRISPR, имеет примерно 10 нуклеотидов в длину. В некоторых вариантах осуществления изобретения степень комплементарности между последовательностью повтора CRISPR и соответствующей

последовательностью «анти-повтора» tracrRNA при оптимальном выравнивании с использованием приемлемого алгоритма сравнительного анализа первичной структуры составляет примерно или более чем примерно 50%, примерно 60%, примерно 70%, примерно 75%, примерно 80%, примерно 81%, примерно 82%,
5 примерно 83%, примерно 84%, примерно 85%, примерно 86%, примерно 87%, примерно 88%, примерно 89%, примерно 90%, примерно 91%, примерно 92%, примерно 93%, примерно 94%, примерно 95%, примерно 96%, примерно 97%, примерно 98%, примерно 99% или более.

В различных вариантах осуществления изобретения полная tracrRNA может
10 содержать от примерно 60 нуклеотидов до более чем примерно 210 нуклеотидов. Например, tracrRNA может иметь в длину примерно 60, примерно 65, примерно 70, примерно 75, примерно 80, примерно 85, примерно 90, примерно 95, примерно 100, примерно 105, примерно 110, примерно 115, примерно 120, примерно 125, примерно 130, примерно 135, примерно 140, примерно 150,
15 примерно 160, примерно 170, примерно 180, примерно 190, примерно 200, примерно 210 нуклеотидов или более. В конкретных вариантах осуществления изобретения tracrRNA имеет в длину от примерно 100 до примерно 201 нуклеотида, включая примерно 95, примерно 96, примерно 97, примерно 98, примерно 99, примерно 100, примерно 105, примерно 106, примерно 107,
20 примерно 108, примерно 109 и примерно 100 нуклеотидов в длину.

Гидовые РНК образуют комплекс с РНК-управляемыми нуклеазами для «принуждения» к связыванию с последовательностью-мишенью и интродукции одноцепочечного или двухцепочечного разрыва последовательности-мишени. После расщепления последовательности-мишени разрыв можно репарировать
25 так, чтобы модифицировать в процессе репарации последовательность ДНК последовательности-мишени. В настоящем описании представлены способы применения мутантных вариантов РНК-управляемых нуклеаз, которые представляют собой либо инактивированную нуклеазу, либо нуклеазы, которые связываются с дезаминазами для модификации последовательности-мишени в
30 ДНК клетки-хозяина. Мутантные варианты РНК-управляемых нуклеаз, в которых нуклеазная активность является инактивированной или значительно пониженной, можно обозначать как РНК-управляемые ДНК-связывающие полипептиды, которые представляют собой полипептиды, обладающие

способностью связываться, но не обязательно расщеплять последовательность-мишень. РНК-управляемые нуклеазы, обладающие способностью расщеплять только одну цепь двухцепочечной молекулы нуклеиновой кислоты, в настоящем описании обозначают как никазы.

5 Нуклеотидная последовательность-мишень связывается с RGN и гибридизуется с гидовой РНК, ассоциированной с RGN. Затем последовательность-мишень может расщепляться RGN, если полипептид обладает нуклеазной активностью, включая активность в качестве никазы.

10 Гидовая РНК может представлять собой единую гидовую РНК или двойную систему гидовых РНК. Единая гидовая РНК содержит sgРНК и необязательно tracrРНК на одной молекуле РНК, в то время как двойная система гидовых РНК содержит sgРНК и tracrРНК в двух различных молекулах РНК, которые гибридизуются друг с другом через по меньшей мере часть последовательности повторов CRISPR sgРНК и по меньшей мере часть tracrРНК, которая может быть
15 полностью или частично комплементарна последовательности повторов CRISPR sgРНК. В некоторых из указанных вариантов осуществления изобретения, в которых гидовая РНК представляет собой единую гидовую РНК, sgРНК и необязательно tracrРНК разделенных линкерной нуклеотидной последовательностью.

20 В целом, линкерная нуклеотидная последовательность представляет собой последовательность, которая не включает комплементарные основания во избежание образования вторичной структуры внутри последовательности или среди нуклеотидов линкерной нуклеотидной последовательности. В некоторых вариантах осуществления изобретения линкерная нуклеотидная
25 последовательность между sgРНК и tracrРНК имеет в длину по меньшей мере 3, по меньшей мере 4, по меньшей мере 5, по меньшей мере 6, по меньшей мере 7, по меньшей мере 8, по меньшей мере 9, по меньшей мере 10, по меньшей мере 11, по меньшей мере 12 или более нуклеотидов. В конкретных вариантах осуществления изобретения линкерная нуклеотидная последовательность единой
30 гидовой РНК имеет в длину по меньшей мере 4 нуклеотида.

В конкретных вариантах осуществления изобретения гидовую РНК можно интродуцировать в клетку-, органеллу- или эмбрион-мишень в виде молекулы РНК. Гидовую РНК можно транскрибировать *in vitro* или химически

синтезировать. В других вариантах осуществления изобретения нуклеотидную последовательность, кодирующую гидовую РНК, интродуцируют в клетку, органеллу или эмбрион. В некоторых из указанных вариантов осуществления изобретения нуклеотидную последовательность, кодирующую гидовую РНК, функционально связывают с промотором (например, промотором РНК-полимеразы III). Промотор может представлять собой нативный промотор или может быть гетерологичным относительно кодирующей гидовую РНК нуклеотидной последовательности.

В различных вариантах осуществления изобретения гидовую РНК можно интродуцировать в клетку-, органеллу- или эмбрион-мишень в виде рибонуклеопротеинового комплекса, указанного в настоящем описании, в котором гидовая РНК связана с полипептидом РНК-управляемой нуклеазы.

Гидовая РНК направляет ассоциированную РНК-управляемую нуклеазу к представляющей интерес конкретной нуклеотидной последовательности-мишени посредством гибридизации гидовой РНК с нуклеотидной последовательностью-мишенью. Нуклеотидная последовательность-мишень может содержать ДНК, РНК или комбинацию их обеих и может быть одноцепочечной или двухцепочечной. Нуклеотидная последовательность-мишень может представлять собой молекулу геномной ДНК (т.е. хромосомной ДНК), плазмидной ДНК или РНК (например, матричной РНК, рибосомной РНК, транспортной РНК, микроРНК, малой интерферирующей РНК). Нуклеотидная последовательность-мишень может быть связана (и в некоторых вариантах осуществления изобретения расщеплена) с использованием РНК-управляемой нуклеазы *in vitro* или в клетке. Хромосомная последовательность, таргетируемая RGN, может представлять собой ядерную, пластидную или митохондриальную хромосомную последовательность. В некоторых вариантах осуществления изобретения нуклеотидная последовательность-мишень является уникальной в геноме-мишени.

В некоторых вариантах осуществления изобретения нуклеотидная последовательность-мишень находится по соседству с примыкающим к протоспейсеру мотивом (РАМ). РАМ, как правило, находится на расстоянии от примерно 1 до примерно 10 нуклеотидов от нуклеотидной последовательности-мишени, в том числе на расстоянии примерно 1, примерно 2, примерно 3,

примерно 4, примерно 5, примерно 6, примерно 7, примерно 8, примерно 9 или примерно 10 нуклеотидов от нуклеотидной последовательности-мишени. РАМ может примыкать к 5'- или 3'-концу последовательности-мишени. В некоторых вариантах осуществления изобретения РАМ примыкает к 3'-концу

последовательности-мишени. Как правило, РАМ представляет собой консенсусную последовательность, состоящую примерно из 2-6 нуклеотидов, но в конкретных вариантах осуществления изобретения может иметь в длину 1, 2, 3, 4, 5, 6, 7, 8, 9 или более нуклеотидов.

После распознавания соответствующей последовательности РАМ RGN может расщеплять нуклеотидную последовательность-мишень в специфическом сайте расщепления. В контексте настоящего описания сайт расщепления состоит из двух конкретных нуклеотидов в нуклеотидной последовательности-мишени, между которыми нуклеотидная последовательность расщепляется RGN. Сайт расщепления может содержать 1-ый и 2-ой, 2-ой и 3-ий, 3-ий и 4-ый, 4-ый и 5-ый, 5-ый и 6-ой, 7-ой и 8-ой или 8-ой и 9-ый нуклеотиды от РАМ в любом 5'- или 3'-направлении. Поскольку RGN могут расщеплять нуклеотидную последовательность-мишень, приводя к образованию чередующихся (расположенных в шахматном порядке) концов, то в некоторых вариантах осуществления изобретения сайт расщепления определяют на основе расстояния двух нуклеотидов от РАМ на позитивной (+) цепи полинуклеотида и расстоянии двух нуклеотидов от РАМ на негативной (-) цепи полинуклеотида.

RGN можно применять для доставки слитого полипептида, полинуклеотида или малой молекулы в качестве «полезного груза» в конкретное положение в геноме. В некоторых вариантах осуществления изобретения RGN, представляющая собой неактивную нуклеазу или нуклеазу, функционально связана с дезаминазой, а также с USP, предлагаемым в изобретении.

В контексте настоящего описания понятие «дезаминаза» или «полипептид дезаминазы» относится к полипептиду, который катализирует реакцию дезаминирования. Дезаминаза может представлять собой встречающийся в естественных условиях фермент дезаминазу или его активный фрагмент или вариант. В некоторых вариантах осуществления изобретения дезаминаза представляет собой цитидиндезаминазу, катализирующую гидролитическое дезаминирование цитидина или дезоксицитидина с образованием урацила или

дезоксигуанилат соответственно. Цитидиндезаминазы могут воздействовать как на ДНК, так и на РНК, и, как правило, воздействуют на одноцепочечные молекулы нуклеиновых кислот. В предпочтительных вариантах осуществления изобретения RGN, обладающая высокой активностью в отношении цепи-мишени «разрезает» цепь-мишень, в то время как комплементарную нецелевую цепь модифицирует дезаминаза. Клеточный механизм репарации ДНК может восстанавливать разрезанную цепь-мишень, используя в качестве матрицы модифицированную нецелевую цепь, интродуцируя тем самым мутацию в ДНК.

В некоторых вариантах осуществления изобретения цитидиндезаминаза представляет собой дезаминазу семейства комплекса редактирования мРНК апополипротеина В (APOBEC). В некоторых из указанных вариантов осуществления изобретения дезаминаза представляет собой дезаминазу семейства APOBEC1. В некоторых вариантах осуществления изобретения цитидиндезаминаза представляет собой индуцируемую активацией цитидиндезаминазу (AID). В некоторых вариантах осуществления изобретения дезаминаза представляет собой ACF1/ASE-дезаминазу. В конкретных вариантах осуществления изобретения дезаминаза представляет собой аденозиндезаминазу. В некоторых из указанных вариантов осуществления изобретения дезаминаза представляет собой дезаминазу семейства ADAT. Дополнительные приемлемые ферменты дезаминазы или их домены должны стать очевидны специалистам в данной области после ознакомления с настоящим описанием.

Одним из примеров приемлемого типа полипептидов дезаминаз являются цитидиндезаминазы, например, семейства APOBE. Семейство комплекса редактирования мРНК апополипротеина В (APOBEC) ферментов цитозиндезаминаз включает 11 белков, которые служат для инициации мутагенеза контролируемым и выгодным образом (Conticello и др., *Genome Biology*, 9(6), 2008, с. 229). Только один член семейства, а именно, индуцируемая активацией цитидиндезаминаза (AID), ответственна за созревание антител путем превращения цитозина в ssДНК в урацилы зависящим от транскрипции и настроенным на цепь образом (Reynaud и др., *Nature Immunology*, 4(7), 2003, сс. 631-638). Фермент комплекса 3 редактирования апополипротеина В (APOBEC3) обеспечивает защиту человеческих клеток от определенного штамма ВИЧ-1 путем дезаминирования цитозина полученной

путем обратной транскрипции вирусной ssДНК (Bhagwat и др., DNA Repair (Amst), 3(1), 2004, сс. 85-89). Для проявления каталитической активности всех этих белков требуется Zn^{2+} -координирующий мотив (HisX- Glu-X₂₃₋₂₆-Pro-Cys-X₂₋₄-Cys) и связанная молекула воды. Действие остатка Glu состоит в активации молекулы воды в отношении гидроксида цинка для нуклеофильной атаки в реакции дезаминирования. Каждый член семейства преимущественно дезаминирует свою конкретную «горячую точку», варьирующую от WRC (где W обозначает А или Т и R обозначает А или G) в случае hAID, до TTC в случае hAPOBEC3F (Navaratnam и др., Intl J Hematol 83(3), 2006, сс. 195-200). Недавнее исследование кристаллической структуры каталитического домена APOBEC3G выявило вторичную структуру, состоящую из сердцевины в виде пятицепочечной β -складки, фланкированной шестью α -спиралями, которая, вероятно, является консервативной для всего семейства (Holden и др., Nature 456(7218), 2008, сс. 121-124). Установлено, что активные центральные петли ответственны и за связывание ssДНК, и за определение идентичности «горячей точки» (Chelico и др., J Biol Chem 284(41), 2009, сс. 27761-27765).

Сверхэкспрессия указанных ферментов связана с геномной нестабильностью и раком, что подчеркивает важность специфического для последовательности таргетинга (Pham и др., Biochem 44(8), 2005, сс. 2703-2715). В некоторых вариантах осуществления изобретения полипептид дезаминазы может представлять собой полипептид дезаминазы, который может дезаминировать цитидин с образованием урацила. Дезаминирование нуклеотидного основания дезаминазой может приводить к точечной мутации соответствующего остатка, модифицируя тем самым молекулу ДНК. Указанные действия модификации в настоящем описании обозначают также как редактирование нуклеиновых кислот или редактирование оснований. Таким образом, слитые белки, содержащие вариант или домен Cas9, домен дезаминазы и домен USP, можно применять для целенаправленного редактирования нуклеотидных последовательностей.

В некоторых вариантах осуществления изобретения RGN, представляющая собой неактивную нуклеазу, или RGN, представляющая собой нуклеазу, слитую с дезаминазой и USP, предлагаемым в изобретении, можно направлять к конкретному местоположению в молекуле нуклеиновой кислоты (т.е. в молекуле нуклеиновой кислоты-мишени), которое в некоторых вариантах осуществления

изобретения представляет собой конкретный локус в геноме, для изменения экспрессии требуемой последовательности. В некоторых вариантах осуществления изобретения связывание слитого белка с последовательностью-мишенью приводит к дезаминированию нуклеотидного основания, результатом которого является превращение одного нуклеотидного основания в другое. В некоторых вариантах осуществления изобретения связывание указанного слитого белка с последовательностью-мишенью приводит к дезаминированию нуклеотидного основания, смежного с последовательностью-мишенью.

Нуклеотидное основание, смежное с последовательностью-мишенью, которое подверглось дезаминированию и мутации с использованием представленных в описании композиций и способов, может находиться на расстоянии 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 или 100 пар оснований от 5'- или 3'-конца последовательности-мишени (связанной с помощью гРНК) в молекуле нуклеиновой кислоты-мишени. Некоторыми объектами настоящего изобретения являются слитые белки, которые содержат (I) полипептид RGN, представляющий собой неактивную нуклеазу или никазу; (II) полипептид дезаминазы; и (III) стабилизирующий урацил полипептид.

В настоящем описании предложены слитые белки различных конфигураций. В некоторых вариантах осуществления изобретения полипептид дезаминазы слит с N-концом полипептида RGN. В некоторых вариантах осуществления изобретения полипептид дезаминазы слит с C-концом полипептида RGN.

В некоторых вариантах осуществления изобретения домен USP, домен дезаминазы и РНК-управляемый ДНК-связывающий полипептид слиты друг с другом через линкер. Между тремя функциональными доменами слитого белка можно использовать линкеры различной длины и гибкости (например, варьирующие от очень гибких линкеров типа $(GGGGS)_n$ и $(G)_n$ до более жестких линкеров типа $(EAAAK)_n$ и $(XP)_n$) для достижения оптимальной длины для проявления дезаминазной активности в конкретных вариантах применения. В контексте настоящего описания понятие «линкер» относится к химической группе или молекуле, связывающей две молекулы или два фрагмента, например, связывающий домен и расщепляющий домен нуклеазы. В некоторых вариантах

осуществления изобретения линкер соединяет РНК-управляемую нуклеазу и дезаминазу. В некоторых вариантах осуществления изобретения линкер соединяет dCas9 и дезаминазу. Как правило, линкер расположен между двумя группами, молекулами или другими фрагментами или фланкирован ими и
5 соединен с каждой/каждым из них ковалентной связью, соединяя таким образом два элемента. В некоторых вариантах осуществления изобретения линкер представляет собой аминокислоту или несколько аминокислот (например, пептид или белок). В некоторых вариантах осуществления изобретения линкер представляет собой органическую молекулу, группу, органический полимер или
10 химический фрагмент. В некоторых вариантах осуществления изобретения линкер имеет в длину 5-100 аминокислот, например, имеет в длину 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150 или 150-200 аминокислот. Можно рассматривать также более длинные или более короткие
15 линкеры.

В некоторых вариантах осуществления изобретения линкер содержит мотив (GGGGS)_n, (G)_n, (EAAAK)_n, или (XP)_n или комбинацию любых указанных мотивов, в которых «n» независимо обозначает целое число от 1 до 30. В некоторых вариантах осуществления изобретения «n» независимо обозначает 1,
20 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 или 30 или, если присутствует более одного линкера или более одного линкерного мотива, то любую их комбинацию. Дополнительные приемлемые линкерные мотивы и конфигурации линкеров должны быть очевидны
специалистам в данной области. В некоторых вариантах осуществления изобретения приемлемые мотивы и конфигурации линкеров включают
25 описанные у Chen и др., Adv Drug Deliv Rev. 65(10), 2013, сс. 1357-1369 (содержание публикации в полном объеме включено в настоящее описание в качестве ссылки). Дополнительные приемлемые последовательности линкеров должны быть очевидны специалистам в данной области.

В некоторых вариантах осуществления изобретения общая архитектура приведенных в качестве примеров слитых белков, представленных в настоящем описании, содержит следующую структуру: [NH₂]-[дезаминаза]-[полипептид RGN]-[USP]-[COOH]; [NH₂]-[USP]-[дезаминаза]-[полипептид RGN]- [COOH];
30

[NH₂]-[USP]-[полипептид RGN]-[дезминаза]-[COOH]; [NH₂]-[полипептид RGN]-[дезминаза]-[USP]-[COOH]; [NH₂]-[полипептид RGN]-[полипептид USP]-[полипептид дезминазы]-[COOH]; или [NH₂]-[полипептид дезминазы]-[полипептид USP]-[полипептид RGN]-[COOH], в которой NH₂ обозначает N-конец слитого белка, а COOH обозначает C-конец слитого белка.

Некоторыми объектами настоящего изобретения являются слитые белки дезминаза-RGN-USP, слитые белки дезминаза-не обладающая нуклеазной активностью RGN-USP и слитые белки дезминаза-представляющая собой никазу RGN-USP, с повышенной эффективностью C→T-редактирования нуклеотидных оснований по сравнению со сходными слитым белком, который не содержит домен USP.

В некоторых вариантах осуществления изобретения слитый белок содержит структуру: [NH₂]-[дезминаза]-[не обладающая нуклеазной активностью RGN]-[USP]-[COOH]; [NH₂]-[полипептид дезминазы]-[USP]-[не обладающая нуклеазной активностью RGN]-[COOH]; [NH₂]-[USP]-[дезминаза]-[не обладающая нуклеазной активностью RGN]-[COOH]; [NH₂]-[USP]-[не обладающая нуклеазной активностью RGN]-[дезминаза]-[COOH]; [NH₂]-[не обладающая нуклеазной активностью RGN]-[дезминаза]-[USP]-[COOH]; или [NH₂]-[не обладающая нуклеазной активностью RGN]-[USP]-[дезминаза]-[COOH]. Должно быть очевидно, что понятие «не обладающая нуклеазной активностью RGN» обозначает любую RGN, включающую любой белок CRISPR-Cas, которая в результате мутации лишена нуклеазной активности. Также должно быть очевидно, что «USP» обозначает один или несколько полипептидов USP.

В других вариантах осуществления изобретения слитый белок содержит структуру: [NH₂]-[дезминаза]-[RGN-никаза]-[USP]-[COOH]; [NH₂]-[дезминаза]-[USP]-[RGN-никаза]-[COOH]; [NH₂]-[USP]-[дезминаза]-[RGN-никаза]-[COOH]; [NH₂]-[USP]-[RGN-никаза]-[дезминаза]-[COOH]; [NH₂]-[RGN-никаза]-[дезминаза]-[USP]-[COOH]; или [NH₂]-[RGN-никаза]-[USP]-[дезминаза]-[COOH]. Должно быть очевидно, что понятие «RGN-никаза» обозначает любую RGN, включающую любой белок CRISPR-Cas, которая в результате мутации приобрела никазную активность. Также должно быть очевидно, что «USP» обозначает один или несколько полипептидов USP.

В некоторых вариантах осуществления изобретения слитый белок содержит цитидиндезаминазу, обладающую по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94, RGN (или ее никазную форму), обладающую по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 40, 41, 95-142, и USP, обладающий по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

В некоторых вариантах осуществления изобретения слитый белок содержит цитидиндезаминазу, обладающую по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94, RGN (или ее никазную форму), обладающую по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 40, 41, 95-142, и USP, обладающий по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

В некоторых вариантах осуществления изобретения слитый белок содержит цитидиндезаминазу, обладающую по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94, RGN (или ее никазную форму), обладающую по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 40, 41, 95-142, и USP, обладающий по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

В некоторых вариантах осуществления изобретения слитый белок содержит цитидиндезаминазу, обладающую по меньшей мере 95%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94, RGN (или ее никазную форму), обладающую по меньшей мере 95%-ной идентичностью последовательности с одной из SEQ ID NO: 40, 41, 95-142, и USP, обладающий по меньшей мере 95%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

В некоторых вариантах осуществления изобретения слитый блок содержит цитидиндезаминазу, которая имеет аминокислотную последовательность, представленную в одной из SEQ ID NO: 47, 48 и 76-94, RGN (или ее никазу), которая имеет аминокислотную последовательность, представленную в одной из

SEQ ID NO: 40, 41 и 95-142, и USP, который имеет аминокислотную последовательность, представленную в одной из SEQ ID NO: 1-16.

В некоторых вариантах осуществления изобретения символ «-», используемый в общей архитектуре выше, указывает на присутствие
5 необязательной линкерной последовательности. В некоторых вариантах осуществления изобретения слитые белки, представленные в настоящем описании, не содержат линкерную последовательность. В некоторых вариантах осуществления изобретения присутствует по меньшей мере одна из необязательных линкерных последовательностей.

10 Другими примерами особенностей, которые могут иметь место, являются локализация последовательностей, например, ядерная локализация последовательностей, цитоплазматическая локализация последовательностей, экспорт последовательностей, например, ядерный экспорт последовательностей, или другая локализация последовательностей, а также метки
15 последовательностей, которая являются ценными для солюбилизации, очистки или детекции слитых белков. Приемлемые последовательности сигналов локализации и последовательности белковых меток представлены в настоящем описании и включают (но ограничиваясь только ими) метки на основе белка-носителя биотинкарбоксилазы (БССР), тус-метки, метки на основе
20 кальмодулина, FLAG-метки, метки на основе гемагглютенина (НА), полигистидиновые метки, которые обозначают также как гистидиновые метки или His-метки, метки на основе мальтозосвязывающего белка (MBP), pus-метки, метки на основе глутатин-S-трансферазы (GST), на основе зеленого
флуоресцентного белка (GFP), метки на основе тиоредоксина, S-метки, Sof-метки (Softag) (например, Softag 1, Softag 3), стреп-метки (streptag), метки на
25 основе биотинлигазы, FlAsH-метки, V5-метки и SBP-метки. Дополнительные приемлемые последовательности должны быть очевидны специалистам в данной области.

В конкретных вариантах осуществления изобретения предложенные в
30 настоящем описании слитые белки содержат по меньшей мере один проникающий в клетку домен, который облегчает поглощение клеткой слитого белка. Проникающие в клетку домены известны в данной области и, как правило, содержат участки положительно заряженных аминокислотных остатков

(т.е. поликатионные проникающие в клетку домены), чередующихся полярных аминокислотных остатков и неполярных аминокислотных остатков (т.е. амфипатические проникающие в клетку домены) или гидрофобных аминокислотных остатков (т.е. гидрофобные проникающие в клетку домены) (см., например, Milletti F. Drug Discov Today 17, 2012, сс. 850-860). Примером проникающего в клетку домена является (но не ограничиваясь только им) транс-активирующий активатор транскрипции (TAT) из вируса иммунодефицита человека 1.

В некоторых вариантах осуществления изобретения USP или слитые белки, предложенные в настоящем описании, содержат дополнительно последовательность ядерной локализации (NLS). Сигнал ядерной локализации, сигнал пластидной локализации, сигнал митохондриальной локализации, сигнал локализации с двойным таргетингом и/или проникающий в клетку домен могут быть локализованы на аминоконце (N-конец), карбоксильном конце (C-конец) или локализованы внутри слитого белка.

В некоторых вариантах осуществления изобретения NLS сливают с N-концом слитого белка или USP. В некоторых вариантах осуществления изобретения NLS сливают с C-концом слитого белка или USP. В некоторых вариантах осуществления изобретения NLS сливают с N-концом USP в слитом белке. В некоторых вариантах осуществления изобретения NLS сливают с C-концом USP в слитом белке. В некоторых вариантах осуществления изобретения NLS сливают с N-концом полипептида RGN в слитом белке. В некоторых вариантах осуществления изобретения NLS сливают с C-концом полипептида RGN в слитом белке. В некоторых вариантах осуществления изобретения NLS сливают с N-концом полипептида дезаминазы в слитом белке. В некоторых вариантах осуществления изобретения NLS сливают с C-концом полипептида дезаминазы в слитом белке. В некоторых вариантах осуществления изобретения NLS сливают со слитым белком или UPS через один или несколько линкеров. В некоторых вариантах осуществления изобретения NLS сливают со слитым белком или UPS без линкера. В некоторых вариантах осуществления изобретения NLS содержит аминокислотную последовательность, представляющую собой одну из последовательностей NLS, предлагаемых в настоящем описании, или на которую сделана ссылка в настоящем описании. В

некоторых вариантах осуществления изобретения NLS содержит аминокислотную последовательность, представленную в SEQ ID NO: 42 или SEQ ID NO: 45.

5 В некоторых вариантах осуществления изобретения предложенные в настоящем описании слитые белки содержат полноразмерную последовательность стабилизирующего урацил белка, например, одну из SEQ ID NO: 1-16. Однако в других вариантах осуществления изобретения предложенные в настоящем описании слитые белки не содержат полноразмерную последовательность USP, а содержат лишь ее фрагмент. Например, в некоторых
10 вариантах осуществления изобретения предложенный в настоящем описании слитый белок дополнительно содержит РНК-управляемый ДНК-связывающий домен, домен дезаминазы и активный фрагмент USP.

В некоторых вариантах осуществления изобретения предлагаемый в изобретении слитый белок содержит RGN, дезаминазу и USP, при этом USP
15 имеет аминокислотную последовательность, которая по меньшей мере на 50%, по меньшей мере на 55%, по меньшей мере на 60%, по меньшей мере на 65%, по меньшей мере на 70%, по меньшей мере на 75%, по меньшей мере на 80%, по меньшей мере на 85%, по меньшей мере на 90%, по меньшей мере на 95%, по меньшей мере на 96%, по меньшей мере на 97%, по меньшей мере на 98%, по
20 меньшей мере на 99% или на 100% идентична любой из SEQ ID NO: 1-16. Примеры указанных слитых белков представлены ниже в разделе «Примеры».

В некоторых вариантах осуществления изобретения слитый белок содержит один полипептид USP. В некоторых вариантах осуществления изобретения слитый белок содержит по меньшей мере два полипептида USP, которые
25 функционально связаны либо непосредственно, либо через линкер. В некоторых вариантах осуществления изобретения слитый белок содержит один полипептид USP, а второй полипептид USP совместно экспрессируют со слитым белком.

Другим вариантом осуществления изобретения является рибонуклеопротеиновый комплекс, который содержит слитый белок и гидовую
30 РНК, либо в виде единой гидовой РНК, либо в виде двойной гидовой РНК (которые совместно обозначают как гРНК).

IV. Нуклеотиды, кодирующие стабилизирующие урацил полипептиды, слитые белки и/или гРНК

В настоящем описании представлены полинуклеотиды, которые кодируют предложенные в настоящем изобретении стабилизирующие урацил полипептиды (SEQ ID NO: 17-32). В настоящем описании представлены также полинуклеотиды, которые кодируют слитые белки, содержащие дезаминазу и ДНК-связывающий полипептид, например, например, мегануклеазу, слитый белок, содержащий домен цинковых пальцев, или TALEN. В настоящем описании представлены также полинуклеотиды, которые кодируют слитые белки, содержащие USP, домен дезаминазы и РНК-управляемый ДНК-связывающий полипептид. Указанный РНК-управляемый ДНК-связывающий полипептид может представлять собой RGN или вариант RGN. Вариант белка может представлять собой неактивную нуклеазу или никазу. RGN может представлять собой белок CRISPR-Cas или его активный вариант или фрагмент. SEQ ID NO: 40 и 41 представляют собой примеры RGN и никазного варианта RGN соответственно (но не ограничиваясь только ими). Примеры нуклеаз CRISPR-Cas хорошо известны в данной области и с помощью сходных соответствующих мутаций можно создавать мутантные варианты, которые представляют собой также никазы или неактивные нуклеазы.

Вариантом осуществления изобретения является полинуклеотид, кодирующий слитый белок, который содержит RGN, дезаминазу и USP, представленный в настоящем описании (SEQ ID NO: 1-16 или его вариант). В некоторых вариантах осуществления изобретения второй полинуклеотид кодирует гидовую РНК, которая необходима RGN для таргетинга представляющей интерес нуклеотидной последовательности. В других вариантах осуществления изобретения гидовая РНК и слитый белок кодируются одним и тем же полинуклеотидом.

Подразумевается, что в настоящем описании применение понятия «полинуклеотид» не ограничено полинуклеотидами, содержащими ДНК. Специалистам в данной области должно быть очевидно, что полинуклеотиды могут содержать рибонуклеотиды (РНК) и комбинации рибонуклеотидов и дезоксирибонуклеотидов. Указанные дезоксирибонуклеотиды и рибонуклеотиды включают как встречающиеся в естественных условиях молекулы, так и

синтетические аналоги. Представленные в настоящем описании полинуклеотиды включают также все формы последовательностей, включая (но не ограничиваясь только ими) одноцепочечные формы, двухцепочечные формы, структуры типа стебель-петля и т.п.

5 Вариантом осуществления изобретения является молекула нуклеиновой кислоты, содержащая последовательность, которая по меньшей мере на 50%, по меньшей мере на 55%, по меньшей мере на 60%, по меньшей мере на 65%, по меньшей мере на 70%, по меньшей мере на 75%, по меньшей мере на 80%, по меньшей мере на 85%, по меньшей мере на 90%, по меньшей мере на 95%, по
10 меньшей мере на 96%, по меньшей мере на 97%, по меньшей мере на 98%, по меньшей мере на 99% или на 100% идентична любой из SEQ ID NO: 17-32, где молекула нуклеиновой кислоты кодирует USP, обладающий стабилизирующей урацил активностью. Молекула нуклеиновой кислоты может содержать также гетерологичный промотор или терминатор. Молекула нуклеиновой кислоты
15 может кодировать слитый белок, в котором кодируемый USP функционально связан с ДНК-связывающим полипептидом и/или дезаминазой. В некоторых вариантах осуществления изобретения молекула нуклеиновой кислоты кодирует слитый белок, в котором кодируемый USP функционально связан с RGN и/или дезаминазой.

20 Молекулы нуклеиновых кислот, содержащие полинуклеотид, который кодирует USP, предлагаемый в изобретении, могут иметь кодоны, оптимизированные для экспрессии в представляющем интерес организме. Кодирующая последовательность с «оптимизированными кодонами» представляет собой полинуклеотидную кодирующую последовательность, в
25 которой создана частота встречаемости предпочтительных кодонов, имитирующая частоту встречаемости предпочтительных кодонов или условий транскрипции в конкретной клетке-хозяине. Экспрессия в конкретной клетке-хозяине или конкретном организме-хозяине повышается в результате изменения одного или несколько кодонов на уровне нуклеиновой кислоты, при этом
30 транслируемая аминокислотная последовательность не изменяется. В молекулах нуклеиновых кислот оптимизация кодонов может быть полной или частичной. В данной области доступны таблицы кодонов и другие ссылки, предоставляющие информацию о предпочтениях для широкого спектра организмов (см., например,

у Campbell и Gowri Plant Physiol. 92, 1990, сс. 1-11 обсуждение предпочтительных для растений часто встречающихся кодонов). В данной области известны методы синтеза предпочтительных для растений генов (см., например, патенты США №№ 5380831 и 5436391, и Murray и др., Nucleic Acids Res. 17, 1989, сс. 477-498, которые включены в настоящее описание в качестве ссылки).

Полинуклеотиды, кодирующие USP, слитые белки и/или гРНК, представленные в настоящем описании, могут находиться в кассетах экспрессии для экспрессии *in vitro* или экспрессии в представляющей/представляющим интерес клетке, органелле, эмбрионе или организме. Кассета должна включать 5'- и 3'- регуляторные последовательности, функционально связанные с полинуклеотидом, который кодирует USP и/или слитый белок, содержащий USP, РНК-управляемый ДНК-связывающий полипептид и дезаминазу и/или гРНК, представленные в настоящем описании, что обеспечивает экспрессию полинуклеотида. Кассета может содержать дополнительно по меньшей мере один дополнительный ген или генетический элемент, подлежащий котрансформации в организме. Если включают дополнительные гены или элементы, то компоненты функционально связывают. Подразумевается, что понятие «функционально связанные» означает функциональную связь между двумя или большим количеством элементов. Например, функциональная связь между промотором и представляющей интерес кодирующей последовательностью (например, область, которая кодирует USP, дезаминазу, РНК-управляемый ДНК-связывающий полипептид и/или гРНК) представляет собой функциональную связь, которая обеспечивает экспрессию представляющей интерес кодирующей области. Функционально связанные элементы могут быть смежными или несмежными. При использовании для обозначения соединения двух кодирующих белок областей, под функционально связанными подразумевается, что кодирующие области находятся в одной и той же рамке считывания. Альтернативно этому, дополнительный(ые) ген(ы) или элемент(ы) могут находиться в нескольких кассетах экспрессии. Например, нуклеотидная последовательность, кодирующая предложенный в настоящем описании стабилизирующий урацил полипептид, либо индивидуально, либо в виде компонента слитого белка, может находиться в одной кассете экспрессии, в

то время как нуклеотидная последовательность, кодирующая гРНК, может присутствовать в другой кассете экспрессии. В другом примере нуклеотидная последовательность, кодирующая предложенный в настоящем описании USP, может присутствовать индивидуально в первой кассете экспрессии, вторая кассета экспрессии может включать слитый белок, содержащий USP, а нуклеотидная последовательность, кодирующая гРНК, может находиться в третьей кассете экспрессии. Предложенная указанная кассета экспрессии содержит несколько сайтов рестрикции и/или сайтов рекомбинации для встраивания полинуклеотидов, регуляция транскрипции которых находится под контролем регуляторных областей. Можно применять также кассеты экспрессии, которые содержат ген селектируемого маркера.

Кассета экспрессии должна включать в 5'→3' направлении транскрипции область инициации транскрипции (и в некоторых вариантах осуществления изобретения инициации трансляции) (т.е. промотор), кодирующий USP полинуклеотид, предлагаемый в изобретении, и область терминирования транскрипции (и в некоторых вариантах осуществления изобретения терминирования трансляции) (т.е. область терминирования), которые функционируют в представляющем интерес организме. Промоторы, предлагаемые в изобретении, обладают способностью направлять экспрессию или управлять экспрессией кодирующей последовательности в клетке-хозяине. Регуляторные области (например, промоторы, регулирующие транскрипцию области и области терминирования трансляции) могут быть эндогенными или гетерологичными для клетки-хозяина или друг для друга. В контексте настоящего описания понятие «гетерологичный» касательно последовательности представляет собой последовательность, имеющую происхождение из чужеродного вида, или, если из этого же вида, то она в значительной степени изменена по сравнению с ее нативной формой в композиции и/или в геномном локусе путем преднамеренного вмешательства человека. В контексте настоящего описания химерный ген содержит кодирующую последовательность, функционально связанную с областью инициации транскрипции, которая является гетерологичной относительно кодирующей последовательности.

Из Ti-плазмиды *A. tumefaciens* можно получать приемлемые области терминирования, такие как области терминирования октопинсинтазы и нопалинсинтазы

(см., также Guerineau и др., Mol. Gen. Genet. 262, 1991, сс. 141-144; Proudfoot Cell 64, 1991, сс. 671-674; Sanfacon и др., Genes Dev. 5, 1991, сс. 141-149; Mogen и др., Plant Cell 2, 1990, сс. 1261-1272; Munroe и др., Gene 91, 1990, сс. 151-158; Ballas и др., Nucleic Acids Res. 17, 1989, сс. 7891-7903; и Joshi и др., Nucleic Acids Res. 15, 1987, сс. 9627-9639).

Дополнительные регуляторные сигналы включают (но не ограничиваясь только ими) сайты инициации начала транскрипции, операторы, активаторы, энхансеры, другие регуляторные элементы, сайты связывания рибосом, иницирующий кодон, сигналы терминации и т.п. (см., например, патенты США №№. 5039523 и 4853331; ЕРО 0480762А2; Sambrook и др. Molecular Cloning: A Laboratory Manual, под ред. Maniatis и др. (изд-во Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.), 1992 г. (обозначено ниже как «Sambrook 11»; Advanced Bacterial Genetics под ред. Davis и др., (изд-во Cold Spring Harbor Laboratory Press), Cold Spring Harbor, N.Y., 1980 г. и процитированные в указанных публикациях ссылки).

При создании кассеты экспрессии можно манипулировать с различными ДНК-фрагментами так, чтобы получать последовательности ДНК в правильной ориентации и при необходимости в правильной рамке считывания. Для достижения этого можно применять адаптеры или линкеры для соединения ДНК-фрагментов или можно осуществлять другие манипуляции для создания приемлемых сайтов рестрикции, удаления излишних ДНК, удаления сайтов рестрикции или т.п. Для этой цели можно осуществлять мутагенез *in vitro*, репарацию с помощью праймеров, рестрикцию, отжиг, повторные замены, например, транцизии и трансверсии.

Для воплощения изобретения на практике можно применять целый ряд промоторов. Промоторы можно выбирать с учетом требуемого результата. Нуклеиновые кислоты можно объединять с конститутивными, индуцибельными, специфическими для стадии роста, специфическими для типа клеток, предпочтительными для ткани, тканеспецифическими или другими промоторами для экспрессии в представляющем интерес организме (см., например, промоторы, описанные в ВО 99/43838 и патентах США №№ 8575425; 7790846; 8147856; 8586832; 7772369; 7534939; 6072050; 5659026; 5608149; 5608144;

5604121; 5569597; 5466785; 5399680; 5268463; 5608142; и 6177611; которые включены в настоящее описание в качестве ссылки).

Для экспрессии в растениях конститутивные промоторы включают также промотор CaMV 35S (Odell и др., Nature 313, 1985, сс. 810-812); актина риса (McElroy и др., Plant Cell 2, 1990, сс. 163-171); убиквитина (Christensen и др., Plant Mol. Biol. 12, 1989, сс. 619-632 и Christensen и др., Plant Mol. Biol. 18, 1992, сс. 675-689); pEMU (Last и др., Theor. Appl. Genet. 81, 1991, сс. 581-588); и MAS (Velten и др., EMBO J. 3, 1984, сс. 2723-2730).

Примерами индуцибельных промоторов являются промотор Adh1, который индуцируется гипоксией или холодовым стрессом, промотор Hsp70, который индуцируется тепловым стрессом, промотор PPDK и промотор пепкарбоксилазы, которые оба индуцируются светом. Также ценными являются промоторы, которые индуцируются химическими стимулами, такие как промотор In2-2, представляющий собой защитное средство (патент США № 5364780), промотор Axig1, который индуцируется ауксином и является специфическим для тапетума, но также обладающий активностью в каллусе (PCT US01/22169), чувствительные к стероидам промоторы (см., например, описание промотора ERE, который индуцируется эстрогеном, и индуцируемого глюкокортикостероидом промотора у Schena и др., Proc. Natl. Acad. Sci. USA 88, 1991, сс. 10421-10425 и McNellis и др., Plant J. 14(2), 1998, сс.247-257) и индуцируемые тетрациклином и подавляемые тетрациклином промоторы (см., например, Gatz и др., Mol. Gen. Genet. 227, 1991, сс. 229-237 и патенты США №№ 5814618 и 5789156), которые включены в настоящее описание в качестве ссылки).

Тканеспецифические или тканепредпочтительные промоторы можно применять для целевой экспрессии экспрессируемой конструкции в конкретной ткани. В конкретных вариантах осуществления изобретения тканеспецифические или тканепредпочтительные промоторы обладают активностью в ткани растения. Примеры промоторов, находящихся под контролем развития в растениях, включают промоторы, которые иницируют транскрипцию прежде всего в определенных тканях, таких как листья, корни, плоды, семена или цветки. «Тканеспецифический» промотор представляет собой промотор, который иницирует транскрипцию только в определенных тканях. В отличие от конститутивной экспрессии генов тканеспецифическая экспрессия является

результатом нескольких взаимодействующих уровней регуляции генов.

Фактически промоторы из гомологичных или близкородственных видов растений можно предпочтительно использовать для достижения эффективной и надежной экспрессии трансгенов в конкретных тканях. В некоторых вариантах осуществления изобретения экспрессия включает применение

5 тканепредпочтительного промотора. «Тканепредпочтительный» промотор представляет собой промотор, который инициирует транскрипцию предпочтительно, но не обязательно исключительно или только, в определенных тканях.

10 В некоторых вариантах осуществления изобретения молекулы нуклеиновых кислот, которые кодируют USP, представленный в настоящем описании, содержат специфический для типа клеток промотор. «Специфический для типа клеток» промотор представляет собой промотор, который главным образом контролирует экспрессию в конкретных типах клеток в одном или нескольких

15 органах. Некоторые примеры растительных клеток, в которых функционирующие в растениях специфические для типа клеток промоторы, могут главным образом проявлять активность, включают, например, BETL-клетки, сосудистые клетки в корнях, листьях, клетки стеблей и клетки стволов. Молекулы нуклеиновых кислот могут включать также предпочтительные для

20 типа клеток промоторы. «Предпочтительный для типа клеток» промотор представляет собой промотор, который предпочтительно контролирует экспрессию в основном, но не обязательно полностью или исключительно, в определенных типах клеток в одном или нескольких органах. Некоторые примеры растительных клеток, в которых функционирующие в растениях

25 предпочтительные для типа клеток промоторы, предпочтительно могут проявлять активность, включают, например, BETL-клетки, сосудистые клетки в корнях, листьях, клетки стеблей и клетки стволов.

Нуклеотидные последовательности, кодирующие USP, слитые белки и/или гРНК, могут быть функционально связаны с промоторной последовательностью,

30 которая распознается, например, фаговой РНК-полимеразой, для синтеза мРНК *in vitro*. В указанных вариантах осуществления изобретения транскрибируемую *in vitro* РНК можно очищать для применения в способах, представленных в настоящем описании. Например, промоторная последовательность может

представлять собой последовательность промотора T7, T3 или SP6 или вариант последовательности промотора T7, T3 или SP6. В указанных вариантах осуществления изобретения экспрессируемый белок и/или РНК можно очищать для применения в способах модификации генома, представленных в настоящем описании.

В конкретных вариантах осуществления изобретения полинуклеотиды, кодирующие USP, слитый белок и/или гРНК, можно также связывать с сигналом полиаденилирования (например, полиА-сигналом SV40 и другими сигналами, функционирующими в растениях) и/или по меньшей с одной последовательностью терминации транскрипции. Кроме того, последовательность, кодирующую дезаминазу или слитый белок, можно связывать также с последовательностью(ями), которая(ые) кодирует(ют) по меньшей мере один сигнал ядерной локализации, по меньшей мере один проникающий в клетку домен и/или по меньшей мере один сигнальный пептид, который обеспечивает транспорт белков к определенным субклеточным местоположениям, что еще будет описано ниже.

Полинуклеотид, кодирующий USP, слитый белок и/или гРНК, может находиться в векторе или в нескольких векторах. Понятие «вектор» относится к полинуклеотидной композиции, предназначенной для переноса, доставки или интродукции нуклеиновой кислоты в клетку-хозяина. Приемлемые векторы включают плазмидные векторы, фагмиды, космиды, искусственные хромосомы/минихромсомы, транспозоны и вирусные векторы (например, лентивирусные векторы, аденоассоциированные вирусные векторы, бакуловирусный вектор). Вектор может содержать дополнительные контролирующие экспрессию последовательности (например, энхансерные последовательности, последовательности Козак, последовательности полиаденилирования, последовательности терминации транскрипции), последовательности селективируемых маркеров (например, гены, обуславливающие устойчивость к антибиотикам), сайт инициации репликации и т.п. Дополнительную информацию можно почерпнуть в «Current Protocols in Molecular Biology», под ред. Ausubel и др., изд-во John Wiley & Sons, New York, 2003 г., или в «Molecular Cloning: A Laboratory Manual», Sambrook и Russell, изд-во Cold Spring Harbor Press, Cold Spring Harbor, N.Y., 3-е изд., 2001 г.

Вектор может включать также ген селективируемого маркера для отбора трансформированных клеток. Гены селективируемых маркеров применяют для отбора трансформированных клеток или тканей. Маркерные гены включают гены, кодирующие фактор устойчивости к антибиотикам, например, гены, кодирующие неомицинфосфотрансферазу II (NEO) и

5 гигромицинфосфотрансферазу (HPT), в также гены, обуславливающие устойчивость к гербицидам, таким как глюфосинат аммония, бромоксинил, имидазолиноны и 2,4-дихлорфеноксиацетат (2,4-D).

В некоторых вариантах осуществления изобретения кассета экспрессии или вектор, содержащая/содержащий последовательность, которая кодирует слитый

10 белок, содержащий РНК-управляемый ДНК-связывающий полипептид, такой как RGN, может дополнительно содержать последовательность, кодирующую гРНК. Последовательность(и), кодирующая(ие) гРНК, может(гут) быть функционально связаны(ы) по меньшей мере с одной контролирующей транскрипцию

15 последовательностью для экспрессии гРНК в представляющем интерес организме-хозяине или представляющей интерес клетке-хозяине. Например, полинуклеотид, кодирующий гРНК, может быть функционально связан с промоторной последовательностью, которая распознается РНК-полимеразой III (Pol III). Примеры приемлемых промоторов Pol III включают (но не

20 ограничиваясь только ими) промоторы млекопитающих U6, U3, H1 и 7SL РНК и промоторы риса U6 и U3.

Как указано, предназначенные для экспрессии конструкции, содержащие нуклеотидные последовательности, которые кодируют USP, слитые белки и/или гРНК, можно применять для трансформации представляющих интерес

25 организмов. Методы трансформации включают интродукцию нуклеотидной конструкции в представляющий интерес организм. Под «интродукцией» подразумевается встраивание нуклеотидной конструкции в клетку-хозяина таким образом, чтобы конструкция получала доступ к внутренней области клетки-хозяина. При осуществлении способов, предлагаемых в изобретении, не

30 требуется применение конкретного метода интродукции нуклеотидной конструкции в организм-хозяин, а требуется только, чтобы нуклеотидная конструкция получала доступ к внутренней области по меньшей мере одной клетке организма-хозяина. Клетка-хозяин может представлять собой

эукариотическую или прокариотическую клетку. В конкретных вариантах осуществления изобретении эукариотическая клетка-хозяин представляет собой клетку растения, клетку млекопитающего или клетку насекомого. Методы интродукции нуклеотидных конструкций в растения и другие клетки-хозяева известны в данной области, и они включают (но не ограничиваясь только ими) методы стабильной трансформации, методы кратковременной трансформации и опосредуемые вирусами методы.

Методы приводят к получению трансформированного организма, такого как растение, включая целое растение, а также органы растения (например, листья, стебли, корни и т.д.), семена, растительные клетки, побеги, зародыши и их потомство. Растительные клетки могут быть дифференцированными или недифференцированными (например, каллус, клетки суспензионной культуры, протопласты, клетки листьев, клетки корней, клетки флоэмы, пыльца).

Понятие «трансгенные организмы» или «трансформированные организмы» или «стабильно трансформированные» организмы или клетки, или ткани относится к организмам, в которые включен или интегрирован полинуклеотид, кодирующий дезаминазу, предлагаемую в изобретении. Очевидно, что другие экзогенные или эндогенные нуклеотидные последовательности или ДНК-фрагменты могут также быть включены в клетку-хозяина. Опосредуемая *Agrobacterium* и биобаллистическая трансформация остаются двумя преимущественно используемыми подходами к трансформации растительных клеток. Однако трансформацию клетки-хозяина можно осуществлять путем заражения, трансфекции, микроинъекции, электропорации, бомбардировки микроснарядами, биобаллистической бомбардировки или бомбардировки частицами, с помощью карбидокремниевых волокон, опосредованного ультразвуком, опосредуемого ПЭГ, опосредуемого фосфатом кальция совместного осаждения, с помощью технологии на основе поликатионного ДМСО, процедуры на основе DEAE-декстрана и опосредуемой вирусами и опосредуемой липосомами трансформации и т.п. Опосредуемая вирусами интродукция полинуклеотида, кодирующего дезаминазу, слитый белок и/или гРНК, включает опосредуемую ретровирусом, лентивирусом, аденовирусом и аденоассоциированным вирусом интродукцию и экспрессию, а также применение *Caulimoviruses* (например, вируса мозаики цветной капусты),

Geminiviruses (например, вируса золотисто-желтой мозаики фасоли или вируса полосатости кукурузы) и РНК-вирусов растений (например, вируса мозаики табака).

Протоколы трансформации, а также протоколы интродукции
5 полипептидных или полинуклеотидных последовательностей в растения могут
варьироваться в зависимости от типа клетки-хозяина (например, клетка
однодольного или двудольного растения), являющейся мишенью для
трансформации. Методы трансформации известны в данной области и включают
методы, изложенные в патентах США №№ 8575425; 7692068; 8802934; 7541517;
10 каждый из которых включен в настоящее описание в качестве ссылки (см., также
Rakoczy-Trojanowska M. Cell Mol Biol Lett. 7, 2002, сс. 849-858; Jones и др., Plant
Methods 1, 2005, с. 5; Rivera и др., Physics of Life Reviews 9, 2012, сс. 308-345;
Bartlett и др., Plant Methods 4, 2008, сс. 1-12; Bates G.W. Methods in Molecular
Biology 111, 1999, сс. 359-366; Binns и Thomashow Annual Reviews in
15 Microbiology 42, 1988, сс. 575-606; Christou P. The Plant Journal 2, 1992, сс. 275-
281; Christou P. Euphytica 85, 1995, сс. 13-27; Tzfira и др., TRENDS in Genetics
20, 2004, сс. 375-383; Yao и др., Journal of Experimental Botany 57, 2006, сс. 3737-
3746; Zupan и Zambryski Plant Physiology 107, 1995, сс. 1041-1047; Jones и др.,
Plant Methods 1, 2005, с. 5).

20 Трансформация может приводить к стабильному или кратковременному
встраиванию нуклеиновой кислоты в клетку. Подразумевается, что понятие
«стабильная трансформация» означает, что нуклеотидная конструкция,
интродуцированная в клетку-хозяина, интегрирована в геном клетки-хозяина и
может наследоваться ее потомством. Подразумевается, что понятие
25 «кратковременная трансформация» означает, что полинуклеотид
интродуцирован в клетку-хозяина, но не интегрирован в геном клетки-хозяина.

Методы трансформации хлоропластов известны в данной области (см.,
например Svab и др., Proc. Natl. Acad. Sci. USA 87, 1990, сс. 8526-8530; Svab и
Maliga Proc. Natl. Acad. Sci. USA 90, 1993, сс. 913-917; Svab и Maliga EMBO J.
30 12, 1993, сс. 601-606). Метод основан на доставке с помощью генной пушки
частиц ДНК, содержащих селективируемый маркер, и таргетирующих ДНК в геном
пластиды посредством гомологичной рекомбинации. Кроме того,
трансформацию пластид можно осуществлять путем трансаактивации

«молчащего» переносимого пластидой трансгена с помощью тканепредпочтительной экспрессии кодируемой в ядре и направленной к пластиде РНК-полимеразы. Такая система описана у McBride и др., Proc. Natl. Acad. Sci. USA 91, 1994, сс. 7301-7305.

5 Трансформированные клетки можно выращивать в трансгенном организме, таком как растение, общепринятыми методами (см., например, McCormick и др., Plant Cell Reports 5, 1986, сс. 81-84). Указанные растения можно затем выращивать и либо опылять такой же трансформированной линией, либо
10 другими линиями, и идентифицировать полученный в результате гибрид, конститутивно экспрессирующий требуемый фенотипический признак. Можно выращивать два или большее количество поколений для гарантии того, что экспрессия требуемого фенотипического признака стабильно поддерживается и наследуется, а затем собирать семена для гарантии достижения экспрессии
15 требуемого фенотипического признака. Таким образом, в настоящем изобретении предложен трансформированный семенной материал (обозначенный также как «трансгенный семенной материал»), содержащий нуклеотидную конструкцию, предлагаемую в изобретении, например, кассету экспрессии, предлагаемую в изобретении, стабильно включенную в его геном.

Альтернативно этому, трансформированные клетки можно
20 интродуцировать в организм. Указанные клетки можно получать из организма, в котором клетки трансформировали с помощью подхода *ex vivo*. Представленные в настоящем описании последовательности можно применять для трансформации любого вида растения, включая (но не ограничиваясь только ими) однодольные и двудольные растения. Примеры представляющих интерес
25 растений включают (но не ограничиваясь только ими) кукурузу (маис), сорго, пшеницу, подсолнечник, томаты, крестоцветные, перец, картофель, хлопчатник, рис, сою, сахарную свеклу, сахарный тростник, табак, ячмень и рапс масличный, *Brassica* sp., люцерну, рожь, просо, сафлор, арахис, сладкий картофель, кассайю, кофе, кокос, ананас, цитрусовые деревья, какао, чай, бананы, авокадо, инжир, гуаву,
30 манго, оливки, папайю, кешью, макадамию, миндаль, овес, овощные культуры, декоративные растения и хвойные растения.

Овощные культуры включают (но не ограничиваясь только ими) томаты, листовой салат, зеленую фасоль, лимскую фасоль, горох и представителей рода

Curcumis, таких как огурец, дыня и мускусная дыня. Декоративные растения включают (но не ограничиваясь только ими) азалию, гортензию, гибискус, розы, тюльпаны, нарциссы, петунии, гвоздику, пуансеттию и хризантему. Предпочтительно согласно настоящему изобретению растения представляют собой культурные растения (например, кукуруза, сорго, пшеница, подсолнечник, томаты крестоцветные, перец, картофель, хлопчатник, рис, соя, сахарная свекла, сахарный тростник, табак, ячмень, рапс и т.д.).

В контексте настоящего описания понятие «растение» включает клетки растений, протопласты растений, культуры тканей из клеток растений, из которых можно регенерировать растения, каллус растений, маточный корень растений и растительные клетки, которые остаются неповрежденными в растениях или частях растений, таких как зародыши, пыльца, семяпочки, семена, листья, цветы, ветви, плоды, ядра, початки, шелуха, стебли, корни, кончики корней, пыльники и т.п. Подразумевается, что понятие «зерно» означает зрелые семена, произведенные коммерческими производителями для целей, отличных от выращивания или воспроизводства вида. Потомство, варианты и мутанты регенерированных растений также включены в объем изобретения при условии, что эти части содержат интродуцированные полинуклеотиды. Кроме того, предложен обработанный растительный продукт или побочный продукт, который сохраняет указанные в настоящем описании последовательности, включая, например, соевую муку.

Полинуклеотиды, кодирующие USP, слитые белки и/или гРНК, можно применять для трансформации любых эукариотических видов, включая (но не ограничиваясь только ими) животных (например, млекопитающих, насекомых, рыб, птиц и рептилий), грибы, амёбы, водоросли и дрожжи. Полинуклеотиды, кодирующие USP, слитые белки и/или гРНК, можно применять для трансформации любых прокариотических видов, включая (но не ограничиваясь только ими) археи и бактерии (например, *Bacillus* spp., *Klebsiella* spp., *Streptomyces* spp., *Rhizobium* spp., *Escherichia* spp., *Pseudomonas* spp., *Salmonella* spp., *Shigella* spp., *Vibrio* spp., *Yersinia* spp., *Mycoplasma* spp., *Agrobacterium* spp. и *Lactobacillus* spp.).

Общепринятые методы переноса генов на основе вирусов и на другой основе можно применять для интродукции нуклеиновых кислот в клетки- или

ткани-мишени млекопитающих. Указанные методы можно применять для введения нуклеиновых кислот, кодирующих слитый белок, предлагаемый в изобретении, и необязательно гРНК в клетки в культуре или в организм-хозяин. Примеры невирусных систем доставки включают ДНК-плазмиды, РНК

- 5 (например, транскрипт вектора, указанного в настоящем описании), «голую» нуклеиновую кислоту и нуклеиновую кислоту в комплексе со средством доставки, таким как липосома. Системы доставки на основе вирусного вектора включают ДНКовые и РНКовые вирусы, которые имеют либо эписомные, либо интегрированные геномы после доставки в клетку. Примеры включают (но не
- 10 ограничиваясь только ими) векторы на основе *Caulimoviruses* (например, вируса мозаики цветной капусты), *Geminiviruses* (например, вируса золотисто-желтой мозаики фасоли или вируса полосатости кукурузы) и РНКовых вирусов растений (например, вируса мозаики табака). Обзор технологий генной терапии см. у
- 15 Anderson, Science 256, 1992, cc. 808-813; Nabel и Feigner, TIBTECH 11, 1993, cc. 211-217; Mitani и Caskey, TIBTECH 11, 1993, cc. 162-166; Dillon, TIBTECH 11, 1993, cc. 167-175; Miller, Nature 357, 1992, cc. 455-460; Van Brunt, Biotechnology 6(10), 1988, cc. 1149-1154; Vigne, Restorative Neurology and Neuroscience 8, 1995, cc. 35-36; Kremer и Perricaudet, British Medical Bulletin 51(1), 1995, cc. 31-44; Haddada и др., в: Current Topics in Microbiology and Immunology, под ред.
- 20 Doerfler и Bohm, 1995; и Yu и др., Gene Therapy 1, 1994, cc. 13-26.

Методы невирусной доставки нуклеиновых кислот включают липофекцию, опосредуемую *Agrobacterium* трансформацию, нуклеофекцию, микроинъекцию, биобаллистическую доставку, применение виросом, липосом, иммунолипосом, поликатиона или липида, конъюгатов нуклеиновых кислот, «голой» ДНК,

25 искусственных вирионов и агента, усиливающего поглощение ДНК. Липофекция описана, например, в патентах США №№ 5049386, 4946787 и 4897355, и реагенты для липофекции поступают в продажу (например, Transfectam™ и Lipofectin™). Катионные и нейтральные липиды, которые можно применять для эффективной липофекции полинуклеотидов на основе распознавания

30 рецепторов, включают липиды, описанные у Feigner, WO 91/17424; WO 91/16024. Доставку можно осуществлять в клетки-мишени (например, введение *in vitro* или *ex vivo*) или в ткани-мишени (например, введение *in vivo*).

Приготовление комплексов липид:нуклеиновая кислота, включающих

таргетирующие липосомы, такие как иммунолипидные комплексы, хорошо известно специалисту в данной области (см., например, Crystal, Science 270, 1995, сс. 404-410; Blaese и др., Cancer Gene Ther. 2, 1995, сс. 291- 297; Behr и др., Bioconjugate Chem. 5, 1994, сс. 382-389; Remy и др., Bioconjugate Chem. 5, 1994, сс. 647-654; Gao и др., Gene Therapy 2, 1995, сс. 710-722; Ahmad и др., Cancer Res. 52, 1992, сс. 4817-4820; патенты США №№ 4186183, 4217344, 4235871, 4261975, 4485054, 4501728, 4774085, 4837028 и 4946787).

При применении для доставки нуклеиновых кислот систем на основе РНКовых или ДНКовых вирусов используется преимущество сложных технологических процессов нацеливания вируса на конкретные клетки организма и доставки вирусного «полезного груза» в ядро. Вирусные векторы можно вводить пациентам непосредственно (*in vivo*) или их можно использовать для обработки клеток *in vitro* и модифицированные клетки необязательно можно вводить пациентам (*ex vivo*). Общепринятые системы для переноса генов на основе вирусов могут включать вирусные векторы на основе ретровирусов, лентивирусов, аденовирусов и аденоассоциированных вирусов и вируса герпеса простого. Интеграция в геном хозяина возможна при использовании методов переноса генов на основе ретровирусов, лентивирусов и аденоассоциированных вирусов, что часто приводит к длительной экспрессии встроенного трансгена. Кроме того, обнаружена высокая эффективность трансдукции во многих различных типах клеток и тканей-мишеней.

Тропизм ретровируса можно изменять путем включения чужеродных оболочечных белков, расширяя потенциальную целевую популяцию клеток-мишеней. Лентивирусные векторы представляют собой ретровирусные векторы, которые обладают способностью трансдуцировать или заражать неделящиеся клетки и, как правило, продуцируют высокие вирусные титры. Таким образом, выбор ретровирусной системы переноса генов должен зависеть от ткани-мишени. Ретровирусные векторы состоят из цис-активных длинных концевых повторов с емкостью упаковки вплоть до 6-10 т.п.н. чужеродной последовательности. Для репликации и упаковки векторов, которые затем применяют для интеграции терапевтического гена в клетку-мишень для обеспечения постоянной экспрессии трансгена, достаточно минимальное количество цис-активных LTR. Широко применяемые ретровирусные векторы

включают векторы на основе вируса лейкоза мыши (MuLV), вируса лейкоза гиббонов (GaLV), вируса иммунодефицита обезьян (SIV), вируса иммунодефицита человека (ВИЧ) и их комбинации (см., например, Buchscher и др., J. Virol. 66, 1992, сс. 2731-2739; Johann и др., J. Virol. 66, 1992, сс. 1635-1640; Somnerfelt и др., Virol. 176, 1990, сс. 58-59; Wilson и др., J. Virol. 63, 1989, сс. 2374-2378; Miller и др., J. Virol. 65, 1991, сс. 2220-2224; PCT/US94/05700).

В вариантах применения, для которых предпочтительной является кратковременная экспрессия, можно использовать системы на основе аденовирусов. Векторы на основе аденовирусов обладают очень высокой эффективностью трансдукции во многих типах клеток, и для них не требуется деление клеток. При применении указанных векторов можно получать высокие титры и уровни экспрессии. Указанный вектор можно получать в больших количествах в относительно простой системе. Векторы на основе аденоассоциированного вируса («AAV») можно применять также для трансдукции клеток нуклеиновыми кислотами-мишенями, например, для получения *in vitro* нуклеиновых кислот и пептидов и для процедур генной терапии *in vivo* и *ex vivo* (см., например, West и др., Virology 160, 1987, сс. 38-47; патент США № 4797368; WO 93/24641; Katin, Human Gene Therapy 5, 1994, сс. 793-801; Muzyczka, J. Clin. Invest. 94, 1994, с. 1351). Конструирование рекомбинантных AAV-векторов описано в многочисленных публикациях, включая патент США № 5173414; Tratschin и др., Mol. Cell. Biol. 5, 1985, сс. 3251-3260; Tratschin и др., Mol. Cell. Biol. 4, 1984, сс. 2072-2081; Hermonat и Muzyczka, PNAS 81, 1984, сс. 6466-6470 и Samulski и др., J. Virol. 63, 1989, сс. 3822-3828). Упаковочные клетки, как правило, используют для формирования вирусных частиц, способных инфицировать клетку-хозяина. Указанные клетки включают 293-клетки, которые применяют для упаковки аденовирусов, и ψ J2-клетки или PA317-клетки, которые применяют для упаковки ретровирусов.

Вирусные векторы, применяемые в генной терапии, как правило, создают путем получения клеточной линии, в которой содержащий нуклеиновые кислоты вектор упакован в вирусную частицу. Векторы, как правило, содержат минимальные вирусные последовательности, необходимые для упаковки и последующей интеграции в хозяина, другие вирусные последовательности,

заменяют кассетой экспрессии, содержащей подлежащий(е) экспрессии полинуклеотид(ы). Недостающие вирусные функции обычно поставляются в транс-упаковочной клеточной линии. Например, AAV –векторы, применяемые в генной терапии, как правило, несут только ITR-последовательности из генома AAV, которые требуются для упаковки и интеграции в геном хозяина. Вирусную ДНК упаковывают в клеточную линию, которая содержит хелперную плазмиду, кодирующую другие гены AAV, а именно, *rep* и *cap*, но лишенные ITR-последовательностей.

Клеточную линию можно инфицировать также аденовирусом в качестве хелпера. Хелперный вирус усиливает репликацию AAV-вектора и экспрессию генов AAV с хелперной плазмиды. Хелперная плазида не упаковывается в значительных количествах из-за отсутствия ITR-последовательностей.

Загрязнение аденовирусом можно снижать, например, путем тепловой обработки, к которой аденовирус является более чувствительным, чем AAV.

Дополнительные методы доставки нуклеиновых кислот в клетки известны специалистам в данной области (см., например, US 20030087817, включенный в настоящее описание в качестве ссылки).

В некоторых вариантах осуществления изобретения клетку-хозяина кратковременно или не кратковременно трансфектируют одним или несколькими векторами, указанными в описании. В некоторых вариантах осуществления изобретения клетку трансфектируют в том виде, в котором она встречается в естественных условиях у субъекта. В некоторых вариантах осуществления изобретения клетку, которую трансфектируют, изымают из организма субъекта. В некоторых вариантах осуществления изобретения клетку получают из клеток из организма субъекта, например, в виде клеточной линии. В некоторых вариантах осуществления изобретения клетка или клеточная линия является прокариотической. В других вариантах осуществления изобретения клетка или клеточная линия является эукариотической. В следующих вариантах осуществления изобретения клетку или клеточную линию получают из видов насекомых, птиц, растений или грибов. В некоторых вариантах осуществления изобретения клетка или клеточная линия может представлять собой клетку или клеточную линию млекопитающего, такого, например, как человек, обезьяна, мышь, корова, свинья, коза, хомяк, крыса, кошка или собака. В данной области

известно широкое разнообразие клеточных линий для культур тканей. Примеры клеточных линий включают (но не ограничиваясь только ими) C8161, CCRF-CEM, MOLT, mIMCD-3, NHDF, HeLaS3, Huh1, Huh4, Huh7, HUVEC, HASMC, HEK_n, HEK_a, MiaPaCell, Panel, PC-3, TF1, CTLL-2, CIR, Rat6, CVI, RPTE, A10, T24, 182, A375, ARH-77, Calul, SW480, SW620, SKOV3, SK-UT, CaCo2, P388D1, SEM-K2, WEHI- 231, HB56, TIB55, lurkat, 145.01, LRMB, Bcl-1, BC-3, IC21, DLD2, Raw264.7, NRK, NRK-52E, MRC5, MEF, Hep G2, HeLa B, HeLa T4, COS, COS-1, COS-6, COS-M6A, линию эпителия почки обезьян BS-C-1, линию фибробластов мышиноного эмбриона BALB/3T3, 3T3 Swiss, 3T3-L1, линию фибробластов эмбриона человека 132-d5; линию фибробластов мыши 10.1, 293-T, 3T3, 721, 9L, A2780, A2780ADR, A2780цис, A172, A20, A253, A431, A-549, ALC, B16, B35, BCP-I-клетки, BEAS-2B, bEnd.3, BHK-21, BR 293, BxPC3, C3H-10T1/2, C6/36, Cal-27, CHO, CHO-7, CHO-IR, CHO-K1, CHO-K2, CHO-T, CHO Dhfr^{-/-}, COR-L23, COR-L23/CPR, COR-L235010, CORL23/ R23, COS-7, COV-434, CML T1, CMT, CT26, D17, DH82, DU145, DuCaP, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, HEK-293, HeLa, Hepalclc7, HL-60, HMEC, HT-29, lurkat, 1Y-клетки, K562-клетки, Ku812, KCL22, KGI, KYOI, LNCap, Ma-Mel 1-48, MC-38, MCF-7, MCF-10A, MDA-MB-231, MDA-MB-468, MDA-MB-435, MDCKII, MDCKII, MOR/0.2R, MONO-MAC 6, MTD-1A, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NALM-1, NW-145, клеточные линии OPCN/OPCT, Peer, PNT-1A/PNT 2, RenCa, RIN-5F, RMA/RMAS, Saos-2 cells, Sf-9, SkBr3, T2, T-47D, T84, клеточную линию THP1, U373, U87, U937, VCaP, клетки Vero, WM39, WT-49, X63, YAC-1, YAR и их трансгенные варианты. Клеточные линии можно получать из различных источников, известных специалистам в данной области (см., например, Американскую коллекцию типовых (ATCC) (Манассас, шт. Вирджиния)).

В некоторых вариантах осуществления изобретения клетку, трансфектированную одним или несколькими векторами, указанными в настоящем описании, используют для создания новой клеточной линии, содержащей одну или несколько имеющих происхождение из вектора последовательностей. В некоторых вариантах осуществления изобретения клетку, кратковременно трансфектированную слитым белком, предлагаемым в изобретении, и необязательно гРНК или рибонуклеопротеиновым комплексом,

предлагаемым в изобретении, и модифицированную посредством активности слитого белка или рибонуклеопротеинового комплекса, применяют для создания новой клеточной линии, которая содержит клетки, содержащие модификацию, но лишённые любой другой экзогенной последовательности. В некоторых вариантах осуществления изобретения клетки, кратковременно или не кратковременно трансфектированные одним или несколькими векторами, указанными в настоящем описании, или клеточные линии, полученные из указанных клеток, применяют для оценки одного или нескольких тестируемых соединений.

В некоторых вариантах осуществления изобретения один или несколько векторов, указанных в настоящем описании, применяют для получения трансгенного животного кроме человека или трансгенного растения. В некоторых вариантах осуществления изобретения трансгенное животное представляет собой насекомое. В других вариантах осуществления изобретения насекомое относится к насекомым-паразитам, таким как комар или клещ. В других вариантах осуществления изобретения насекомое относится к насекомым – вредителям растений, таким как кукурузный корневой червь или осенний шелкопряд. В некоторых вариантах осуществления изобретения трансгенное животное представляет собой птицу, такую как курица, индейка, гусь или утка. В некоторых вариантах осуществления изобретения трансгенное животное представляет собой млекопитающее, такое как человек, мышь, крыса, хомяк, обезьяна, человекообразная обезьяна, кролик, свинья, корова, лошадь, коза, овца, кошка или собака.

V. Варианты и фрагменты полипептидов и полинуклеотидов

В настоящем описании предложены активные варианты и фрагменты встречающихся в естественных условиях (т.е. дикого типа) стабилизирующих урацил полипептидов, аминокислотные последовательности которых представлены в SEQ ID NO: 1-16, и кодирующих их полинуклеотидов.

Хотя активность варианта или фрагмента может быть изменена по сравнению с представляющим интерес полинуклеотидом или полипептидом, вариант и фрагмент должны сохранять функциональность представляющего интерес полинуклеотида или полипептида. Например, вариант или фрагмент может обладать повышенной активностью, пониженной активностью, другим

спектром активности или любым другим изменением активности по сравнению с представляющим интерес полинуклеотидом или полипептидом.

Фрагменты и варианты встречающихся в естественных условиях USP, которые представлены в настоящем описании, должны сохранять такую
5 активность, чтобы, в том случае, когда они являются частью слитого белка, дополнительно содержащего дезаминазу или ее фрагмент и/или ДНК-связывающий полипептид или его фрагмент, то указанный слитый белок должен обладать повышенной эффективностью редактирования нуклеотидных оснований (C→T) по сравнению со сходным слитым белком, который не
10 содержит USP-домен.

Понятие «фрагмент» относится к участку полинуклеотидной или полипептидной последовательности, предлагаемой в изобретении. «Фрагменты» или «биологически активные участки» включают полинуклеотиды, содержащие количество смежных нуклеотидов, достаточное для сохранения биологической
15 активности (т.е. дезаминазной активности в отношении нуклеиновых кислот). «Фрагменты» или «биологически активные участки» включают полипептиды, содержащие количество смежных аминокислотных остатков, достаточное для сохранения биологической активности. Фрагменты USP включают фрагменты, более короткие, чем полноразмерные последовательности из-за использования
20 альтернативного расположенного в прямом направлении стартового сайта. Биологически активный участок USP может представлять собой полипептид, который содержит, например, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110 или большее количество смежных аминокислотных остатков любой из SEQ ID NO: 1-16, или его вариант. Указанные биологически активные участки можно
25 получать с помощью методов рекомбинации и оценивать их активность.

Как правило, подразумевается, что понятие «варианты» относится к практически сходным последовательностям. В случае полинуклеотидов вариант содержит делецию и/или добавление одного или нескольких нуклеотидов в одном или нескольких внутренних сайтах в нативном полинуклеотиде и/или
30 замену одного или нескольких нуклеотидов в одном или нескольких сайтах в нативном полинуклеотиде. В контексте настоящего описания «нативный» или «дикого типа» полинуклеотид или полипептид содержит встречающуюся в естественных условиях нуклеотидную последовательность или аминокислотную

последовательность соответственно. В случае полинуклеотидов консервативные варианты включают такие последовательности, которые в результате вырожденности генетического кода кодируют нативную аминокислотную последовательность представляющего интерес гена. Встречающиеся в естественных условиях указанные аллельные варианты можно идентифицировать с помощью хорошо известных методов молекулярной биологии, таких, например, как полимерная цепная реакция (ПЦР), и методов гибридизации, которые изложены ниже. Варианты полинуклеотидов включают также полученные методами синтеза полинуклеотиды, которые можно создавать, например, с помощью сайтнаправленного мутагенеза, но которые все еще кодируют представляющий интерес полипептид или полинуклеотид. Как правило, варианты конкретных полинуклеотидов, указанных в настоящем описании, должны быть по меньшей мере примерно на 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или более идентичны последовательности конкретного полинуклеотида при определении с помощью программ сравнительного анализа первичной структуры последовательности и параметров, указанных в другом месте в настоящем описании.

Варианты конкретного полинуклеотида, указанного в настоящем описании (т.е. референс-полинуклеотида) можно оценивать также путем сравнения процента идентичности последовательностей между полипептидом, кодируемым вариантом полинуклеотида, и полипептидом, кодируемым референс-полинуклеотидом. Процент идентичности последовательностей между любыми двумя полипептидами можно рассчитывать с использованием программ сравнительного анализа первичной структуры последовательности и параметров, указанных в другом месте в настоящем описании. Когда любую заданную пару полинуклеотидов, указанных в настоящем описании, оценивают путем сравнения процента идентичности последовательностей, присущего двум кодируемым ими полипептидам, то процент идентичности последовательностей между двумя кодируемыми полипептидами, составляет по меньшей мере примерно 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или более.

В конкретных вариантах осуществления изобретения представленные в настоящем описании полинуклеотиды кодируют USP, содержащий аминокислотную последовательность, которая по меньшей мере на 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или более идентична аминокислотной последовательности, представленной в любой из SEQ ID NO: 1-16.

Биологически активный вариант стабилизирующего урацил полипептида, предлагаемого в изобретении, может отличаться всего лишь примерно 1-15 аминокислотными остатками, всего лишь примерно 1-10, например, примерно 6-10, всего лишь 5, всего лишь 4, всего лишь 3, всего лишь 2 или всего лишь 1 аминокислотным остатком. В конкретных вариантах осуществления изобретения полипептиды могут содержать N-концевое или C-концевое усечение, полученное путем делеции по меньшей 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60 аминокислотных остатков или более либо из N-, либо из C-конца полипептида.

Очевидно, что можно осуществлять модификации в USP, представленных в настоящем описании, создавая вариант белков и полинуклеотидов. Созданные человеком изменения можно интродуцировать, используя методы сайтнаправленного мутагенеза. Альтернативно этому, можно идентифицировать нативные, пока неизвестные или пока неидентифицированные полинуклеотиды и/или полипептиды, структурно и/или функционально родственные последовательностям, представленным в настоящем описании, которые подпадают под объем настоящего изобретения. Консервативные аминокислотные замены можно осуществлять в неконсервативных областях, которые не могут изменять функцию стабилизирующего урацил полипептида. Альтернативно этому, можно осуществлять модификации, повышающие активность стабилизирующего урацил полипептида.

Варианты полинуклеотидов и белков включают также последовательности и белки, полученные с помощью мутагенной и рекомбиногенной процедуры, такой как перетасовка ДНК. С помощью указанной процедуры манипулируют с одним или несколькими различными USP, указанными в настоящем описании (например, SEQ ID NO: 1-16), для создания нового USP, обладающего требуемыми свойствами. Таким путем создают библиотеки рекомбинантных

полинуклеотидов из популяции родственных последовательностей полинуклеотидов, содержащих последовательность областей, которые обладают существенной идентичностью последовательностей и которые можно подвергать гомологичной рекомбинации *in vitro* или *in vivo*. Например, с помощью такого

5 подхода последовательность мотивов, кодирующих представляющий интерес домен, можно перетасовывать между последовательностями USP, представленными в настоящем описании, и другими идентифицированными в дальнейшем генами USP с получением нового гена, кодирующего представляющий интерес белок с улучшенными свойствами, такими как

10 повышенная величина K_m в случае фермента. Стратегии перетасовки ДНК известны в данной области (см., например, Stemmer, Proc. Natl. Acad. Sci. USA 91, 1994, сс. 10747-10751; Stemmer, Nature 370, 1994, сс. 389-391; Cramer и др., Nature Biotech. 15, 1997, сс. 436-438; Moore и др., J. Mol. Biol. 272, 1997, сс. 336-347; Zhang и др., Proc. Natl. Acad. Sci. USA 94, 1997, сс. 4504-4509; Cramer и др.,

15 Nature 391, 1998, сс. 288-291; и патенты США №№ 5605793 и 5837458). «Перетасованная» нуклеиновая кислота представляет собой нуклеиновую кислоту, полученную с помощью процедуры перетасовки, такой как любая процедура перетасовки, изложенная в настоящем описании. Перетасованные нуклеиновые кислоты получают путем рекомбинации (физической или

20 виртуальной) двух или большего количества нуклеиновых кислот (или характерных сегментов), например, с использованием искусственного и необязательно рекурсивного подхода. Как правило, одну или несколько стадий скрининга применяют в процессах перетасовки для идентификации представляющих интерес нуклеиновых кислот; указанную стадию скрининга

25 можно осуществлять до или после любой стадии рекомбинации. В некоторых (но не во всех) вариантах перетасовки требуется осуществлять несколько раундов рекомбинации до селекции для повышения разнообразия пула, подлежащего скринингу. Общий процесс рекомбинации и селекции необязательно рекурсивно повторяют. В зависимости от контекста понятие «перетасовка» может

30 относиться к полному процессу рекомбинации и селекции или, альтернативно этому, может относиться просто к связанным с рекомбинацией частям полного процесса.

В контексте настоящего описания понятие «идентичность последовательности» или «идентичность» в контексте двух полинуклеотидных или полипептидных последовательностей относится к остаткам в двух последовательностях, которые являются одинаковыми при выравнивании для максимального соответствия в заданном окне сравнения. Когда процент идентичности последовательности используют применительно к белкам, то известно, что положения остатков, которые не являются идентичными, часто отличаются консервативными аминокислотными заменами, при которых аминокислотные остатки заменены другими аминокислотными остатками с аналогичными химическими свойствами (например, зарядом или гидрофобностью) и поэтому не изменяют функциональные свойства молекулы. Когда последовательности отличаются друг от друга консервативными заменами, то процент идентичности последовательности может быть скорректирован в сторону увеличения с учетом консервативного характера замены. Последовательности, которые отличаются друг от друга указанными консервативными заменами, рассматриваются как «сходные последовательности» или как «сходство (последовательностей)». Средства осуществления указанной корректировки известны специалистам в данной области. Как правило, это включает оценку консервативной замены как частичное, а не полное несоответствие, увеличивая тем самым процент идентичности последовательности. Так, например, если идентичной аминокислоте присваивается оценка 1, а неконсервативной замене присваивается оценка ноль, то консервативной замене присваивается оценка от нуля до 1. Оценка консервативных замен определяют, например, как предложено в программе PC/GENE (фирма Intelligenetics, Маунтин-Вью, шт. Калифорния).

В контексте настоящего описания понятие «процент идентичности последовательности» означает величину, определенную путем сравнения двух оптимально выровненных последовательностей в окне сравнения, при этом, часть полинуклеотидной последовательности в окне сравнения может содержать добавления или делеции (например, бреши) по сравнению с референс-последовательностью (которая не содержит добавления или делеции) для оптимального сравнительного анализа первичной структуры двух последовательностей. Процент рассчитывают путем определения количества

положений, в которых идентичное нуклеотидное основание или аминокислотный остаток присутствует в обеих последовательностях, получая количество совпадающих положений, деления количества совпадающих положений на общее количество положений в окне сравнения и умножение результата на 100, получая в результате процент идентичности последовательности.

Если не указано иное, то понятие «величины идентичности/сходства последовательностей», согласно настоящему описанию относится к величине, полученной с помощью версии 10 программы GAP, с использованием следующих параметров: % идентичности и % сходства для нуклеотидной последовательности, полученных с использованием веса брешы 50 и веса длины брешы 3, и матрицы баллов nwsgapdna; % идентичности и % сходства для аминокислотной последовательности, полученных с использованием веса брешы 8 и веса длины брешы 2, и матрицы баллов BLOSUM62; или с помощью любой эквивалентной программы. Под «эквивалентной программой» подразумевается любая программа для сравнения последовательностей, которая для любой из двух рассматриваемых последовательностей производит выравнивание, имеющее идентичные совпадения нуклеотидов или аминокислотных остатков и идентичный процент идентичности последовательности при сравнении с соответствующим выравниванием, произведенным с помощью версии 10 программы GAP.

Две последовательности являются «оптимально выровненными», когда они выровнены для оценки сходства с использованием определенной матрицы аминокислотных замен (например, BLOSUM62), штрафа за наличие брешы и штрафа за расширение брешы для получения максимально возможного балла для этой пары последовательностей. Матрицы аминокислотных замен и их использование для количественной оценки сходства между двумя последовательностями хорошо известны в данной области и описаны, например, у Dayhoff и др. «A model of evolutionary change in proteins» в: «Atlas of Protein Sequence and Structure», т. 5, Suppl. 3 (под ред. М. О. Dayhoff), изд-во Natl. Biomed. Res. Found., Washington, D.C. 1978, сс. 345-352, и у Henikoff и др., Proc. Natl. Acad. Sci. USA 89, 1992, сс.10915-10919. Матрицу BLOSUM62 часто применяют в качестве матрицы установки баллов по умолчанию в протоколах сравнительного анализа первичной структуры последовательности. Штраф за

наличие бреши вводится за интродукцию единственной аминокислотной бреши в одной из выровненных последовательностей, а штраф за расширение бреши вводится за каждое дополнительное пустое аминокислотное положение, встроенное в уже открытую брешь. Выравнивание определяется положениями аминокислот в каждой последовательности, на которых начинается и заканчивается выравнивание, и, необязательно, вставкой бреши или нескольких брешей в одну или обе последовательности, так образом, чтобы получать максимально возможный балл. Хотя оптимальное выравнивание и подсчет баллов можно выполнять вручную, процесс облегчается за счет использования компьютерного алгоритма выравнивания, например, программы gapped BLAST 2.0, описанной у Altschul и др., *Nucleic Acids Res.* 25, 1997, сс. 3389-3402, и доступной для общественности на веб-сайте Национального центра биотехнологической информации (www.ncbi.nlm.nih.gov). Оптимальные выравнивания, включая множественные выравнивания, можно осуществлять с использованием, например, программы PSI-BLAST, доступной через www.ncbi.nlm.nih.gov и описанной у Altschul и др., *Nucleic Acids Res.* 25, 1997, сс. 3389-3402.

Касательно аминокислотной последовательности, оптимально выровненной с референс-последовательностью, аминокислотный остаток «соответствует» положению в референс-последовательности, с которым остаток сравнивают при выравнивании. «Положение» обозначают номером, который последовательно идентифицирует каждую аминокислоту в референс-последовательности на основе ее положения относительно N-конца. Из-за делеций, инсерцией, усечений, слияний и т.д., которые следует принимать во внимание при определении оптимального выравнивания, в целом, номер аминокислотного остатка в тестируемой последовательности, определенный путем простого подсчета, начиная с N-конца, не обязательно должен совпадать с номером соответствующего положения в референс-последовательности. Например, в случае, когда в выровненной тестируемой последовательности есть делеция, то в месте делеции не будет аминокислоты, соответствующей положению в референс-последовательности. Если в выровненной референс-последовательности есть инсерция, то эта инсерция не будет соответствовать ни одному аминокислотному положению в референс-последовательности. В случае

усечений или слияний могут быть участки аминокислот либо в референс-, либо в выравненной последовательности, которые не соответствуют ни одной аминокислоте в соответствующей последовательности.

VI. Антитела

5 Предложены также антитела к USP, слитым белкам или к рибонуклеопротеинам, содержащим USP, предлагаемые в настоящем изобретении, которые содержат аминокислотную последовательность, представленную в SEQ ID NO: 1-16, или ее активные варианты или фрагменты. Методы получения антител хорошо известны в данной области (см., например, 10 Harlow и Lane, Antibodies: A Laboratory Manual, изд-во Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. 1988; и патент США № 4196265). Эти антитела можно применять в наборах для детекции и выделения USP или слитых белков или рибонуклеопротеинов, содержащих USP, которые указаны в настоящем описании. Таким образом, в настоящем описании предложены 15 наборы, содержащие антитела, которые специфически связываются с полипептидами или рибонуклеопротеинами, указанными в настоящем описании, включая, например, полипептиды, содержащие последовательность, идентичную по меньшей мере на 85% любой из SEQ ID NO: 1-16.

VII. Системы и рибонуклеопротеиновые комплексы для связывания

20 представляющий интерес последовательности-мишени и способы их получения

В настоящем описании предложена система, которая таргетирует нуклеотидную последовательность и модифицирует нуклеотидную последовательность-мишень. В некоторых вариантах осуществления изобретения РНК-управляемый ДНК-связывающий полипептид, такой как RGN, 25 и гРНК ответственны для направленного переноса рибонуклеопротеинового комплекса к представляющей интерес нуклеотидной последовательности; полипептид дезаминазы ответствен за C>U-модификацию таргетированной нуклеотидной последовательности; стабилизирующий урацил полипептид обеспечивает сохранение урацила в молекуле ДНК, в результате чего 30 происходит требуемая репарация ДНК с интродукцией мутации C>T. Гидовая РНК гибридизуется с представляющей интерес последовательностью-мишенью и образует также комплекс с РНК-управляемым ДНК-связывающим полипептидом, «принуждая» тем самым к связыванию РНК-управляемого ДНК-

связывающего полипептида с последовательностью-мишенью. РНК-управляемый ДНК-связывающий полипептид представляет собой один домен трехдоменного слитого белка; второй домен представляет собой дезаминазу, а третий домен представляет собой USP, указанный в настоящем описании. В

5 некоторых вариантах осуществления изобретения РНК-управляемый ДНК-связывающий полипептид представляет собой RGN, такую как Cas9. Другие примеры РНК-управляемых ДНК-связывающих полипептидов включают RGN, описанные в публикации заявки на патент США № 2019/0367949 (которая в полном объеме включена в настоящее описание в качестве ссылки). В некоторых

10 вариантах осуществления изобретения РНК-управляемый ДНК-связывающий полипептид представляет собой полипептид CRISPR-Cas типа II или его активный вариант или фрагмент. В некоторых вариантах осуществления изобретения РНК-управляемый ДНК-связывающий полипептид представляет собой полипептид CRISPR-Cas типа V или его активный вариант или фрагмент.

15 В других вариантах осуществления изобретения РНК-управляемый ДНК-связывающий полипептид представляет собой полипептид CRISPR-Cas типа VI. В других вариантах осуществления изобретения для ДНК-связывающего домена слитого белка не требуется «управление» с помощью РНК, как в случае полипептида нуклеазы цинковых пальцев, TALEN или мегануклеазы. В

20 некоторых из указанных вариантов осуществления изобретения нуклеазная активность каждого из ферментов инактивирована. В следующих вариантах осуществления изобретения РНК-управляемый ДНК-связывающий полипептид содержит аминокислотную последовательность RGN, такую как аминокислотная последовательность, которая идентична по меньшей мере на 80%

25 последовательности APG07433.1 (SEQ ID NO: 40), или ее активный вариант или фрагмент, такой как никаза APG07433.1 (SEQ ID NO: 41). В следующих вариантах осуществления изобретения РНК-управляемый ДНК-связывающий полипептид содержит аминокислотную последовательность RGN, такую как аминокислотная последовательность, которая идентична по меньшей мере на

30 85% последовательности APG07433.1 (SEQ ID NO: 40), или ее активный вариант или фрагмент, такой как никаза APG07433.1 (SEQ ID NO: 41). В следующих вариантах осуществления изобретения РНК-управляемый ДНК-связывающий полипептид содержит аминокислотную последовательность RGN, такую как

аминокислотная последовательность, которая идентична по меньшей мере на 90% последовательности APG07433.1 (SEQ ID NO: 40), или ее активный вариант или фрагмент, такой как никаза APG07433.1 (SEQ ID NO: 41). В следующих вариантах осуществления изобретения РНК-управляемый ДНК-связывающий полипептид содержит аминокислотную последовательность RGN, такую как аминокислотная последовательность, которая идентична по меньшей мере на 95% последовательности APG07433.1 (SEQ ID NO: 40), или ее активный вариант или фрагмент, такой как никаза APG07433.1 (SEQ ID NO: 41). В следующих вариантах осуществления изобретения РНК-управляемый ДНК-связывающий полипептид содержит аминокислотную последовательность RGN, такую как APG07433.1 (SEQ ID NO: 40), или ее активный вариант или фрагмент, такой как никаза APG07433.1 (SEQ ID NO: 41).

Указанная система для связывания представляющей интерес последовательности-мишени, предложенная в настоящем описании, может представлять собой рибонуклеопротеиновый комплекс, в котором по меньшей мере одна молекула РНК связана по меньшей мере с одним белком. Рибонуклеопротеиновые комплексы, представленные в настоящем описании, содержат по меньшей мере одну гидовую РНК в качестве РНК-компонента и слитый белок, содержащий дезаминазу, USP, предлагаемый в изобретении, и РНК-управляемый ДНК-связывающий полипептид в качестве белкового компонента. Рибонуклеопротеиновый комплекс можно очищать из клетки или организма, трансформированной/трансформированного полинуклеотидами, которые кодируют слитый белок и гидовую РНК, и культивировать в условиях, обеспечивающих экспрессию слитого белка и гидовой РНК. Таким образом, предложен способ создания USP, слитого белка или слитого белка рибонуклеопротеинового комплекса. Указанные способы включают культивирование клетки, содержащей нуклеотидную последовательность, которая кодирует USP, слитый белок и в некоторых вариантах осуществления изобретения нуклеотидную последовательность, которая кодирует гидовую РНК, в условиях, в которых экспрессируется USP или слитый белок (и в некоторых вариантах осуществления изобретения гидовая РНК). USP, слитый белок или слитый рибонуклеопротеин затем можно очищать из лизата культивированных клеток.

Методы очистки USP, слитого белка или слитого рибонуклеопротеинового комплекса из лизата биологического образца известны в данной области (например, гель-фильтрация и/или аффинная хроматография, 2D-ПААГ, ЖХВР, хроматография с обращенной фазой, иммунопреципитация). В конкретных методах USP или слитый белок получают методом рекомбинации, и они содержат метку для очистки, способствующую их очистке, такую как (но не ограничиваясь только ими) глутатион-S-трансфераза (GST), хитинсвязывающий белок (CBP), мальтозосвязывающий белок, тиоредоксин (TRX), поли(NANP), метка для тандемной аффинной очистки (TAP) мус, AcV5, AU1, AU5, E, ECS, E2, FLAG, HA, nus, Softag 1, Softag 3, Strep, SBP, Glu-Glu, HSV, KT3, S, S1, T7, V5, VSV-G, 6×His, биотинкарбоксильный белок-носитель (BCCP) и кальмодулин. Как правило, меченные USP, слитый белок или слитый рибонуклеопротеиновый комплекс очищают, используя иммунопреципитацию или другие аналогичные методы, известные в данной области.

«Выделенный» или «очищенный» полипептид или его биологически активный участок является в значительной степени или практически свободным от компонентов, которые обычно сопровождают полипептид или взаимодействуют с ним в естественных условиях. Таким образом, выделенный или очищенный полипептид практически свободен от другого клеточного материала или клеточной среды при его получении методами рекомбинации или практически свободен от химических предшественников или других химических веществ при его получении путем химического синтеза. Белок, который практически свободен от клеточного материала, включает препараты белка, имеющие менее чем примерно 30%, 20%, 10%, 5% или 1% (в пересчете на сухую массу) загрязняющего белка. Когда белок, предлагаемый в изобретении, или его биологически активный участок получают путем рекомбинации, то оптимальная культуральная среда содержит менее чем примерно 30%, 20%, 10%, 5% или 1% (в пересчете на сухую массу) химических предшественников или химических соединений, которые не являются представляющим интерес белком.

VIII. Способы модификации последовательности-мишени

В настоящем описании предложены способы модификации представляющей интерес молекулы нуклеиновой кислоты-мишени (например, молекулы ДНК-мишени). Способы включают доставку системы, содержащей по

меньшей мере одну гидовую РНК или полинуклеотид, который ее кодирует, и по меньшей мере один слитый белок, который содержит USP, предлагаемый в изобретении, дезаминазу и РНК-управляемый ДНК-связывающий полипептид или кодирующий его полинуклеотид, к последовательности-мишени или клетке, органелле или эмбриону, содержащей/содержащему последовательность-мишень. В некоторых из указанных вариантов осуществления изобретения слитый белок содержит одну из аминокислотных последовательностей SEQ ID NO: 1-16 или ее активный вариант или фрагмент.

В некоторых вариантах осуществления изобретения способы включают приведение в контакт молекулы ДНК с (а) слитым белком, содержащим USP, дезаминазу и РНК-управляемый ДНК-связывающий полипептид, такой, например, как домен неактивной нуклеазы или нуклеазы Cas9; и (б) гРНК, которая обеспечивает направленный перенос слитого белка, указанного в подпункте (а), к нуклеотидной последовательности-мишени молекулы ДНК; в которых молекулу ДНК приводят в контакт со слитым белком и гРНК в эффективном количестве и в условиях, пригодных для дезаминирования нуклеотидного основания. В некоторых вариантах осуществления изобретения молекула ДНК-мишени содержит последовательность, ассоциированную с заболеванием или нарушением, и в которых дезаминирование нуклеотидного основания приводит к образованию последовательности, не ассоциированной с заболеванием или нарушением. В некоторых вариантах осуществления изобретения заболевание или нарушение поражает животных. В следующих вариантах осуществления изобретения заболевание или нарушение поражает млекопитающих, таких как люди, коровы, лошади, собаки, кошки, козы, овцы, свиньи, обезьяны, крысы, мыши или хомяки. В некоторых вариантах осуществления изобретения последовательность ДНК-мишени находится в аллеле культурного растения, где конкретный аллель представляющего интерес признака приводит к меньшей агрономической ценности растения. Дезаминирование нуклеотидного основания приводит к образованию аллеля с улучшенным признаком и повышает агрономическую ценность растения.

В некоторых вариантах осуществления изобретения требуемая мутация представляет собой точечную мутацию Т→С, ассоциированную с заболеванием или нарушением, и при этом дезаминирование мутантного С-основания

позволяет получать последовательность, не ассоциированную с заболеванием или нарушением. В некоторых вариантах осуществления изобретения дезаминирование корректирует точечную мутацию в последовательности, ассоциированной с заболеванием или нарушением.

5 В некоторых вариантах осуществления изобретения последовательность, ассоциированная с заболеванием или нарушением, кодирует белок, и дезаминирование интродуцирует стоп-кодон в последовательность, ассоциированную с заболеванием или нарушением, что приводит к усечению кодируемого белка. В некоторых вариантах осуществления изобретения
10 приведение в контакт осуществляют *in vivo* в организме субъекта, чувствительного к возникновению заболевания, имеющему заболевание или нарушение, или у которого они диагностированы. В некоторых вариантах осуществления изобретения заболевание или нарушение представляет собой заболевание, ассоциированное с точечной мутацией или мутацией одного
15 основания в геноме. В некоторых вариантах осуществления изобретения заболевание представляет собой генетическое заболевание, рак, метаболическое заболевание или лизосомную болезнь накопления.

Так, предложенные в настоящем описании композиции и способы можно применять для лечения заболевания или нарушения, ассоциированного с
20 последовательностью (т.е. последовательность является причинным фактором заболевания или нарушения или причинным фактором симптомов, ассоциированных с заболеванием или нарушением), которую изменяют путем мутации для лечения заболевания или нарушения или снижения симптомов, ассоциированных с заболеванием или нарушением. В контексте настоящего
25 описания понятие «лечить» или «лечение» относится к введению фармацевтической композиции, указанной в настоящем описании, которая содержит USP или слитый белок, субъекту, который имеет заболевание или нарушение. Лечение может быть профилактическим, предупреждающим
30 появление симптомов, ассоциированных с заболеванием или нарушением у субъекта, чувствительного к заболеванию или нарушению (например, имеющему генетическую предрасположенность). Требуемые действия лечения включают (но не ограничиваясь только ими) предупреждение возникновения или рецидива заболевания, облегчение симптомов, уменьшение любого прямых или косвенных

патологических последствий заболевания, снижение скорости прогрессирования заболевания, улучшение или ослабление состояния заболевания, а также ремиссию или улучшение прогноза.

5 Фармацевтическая композиция представляет собой композицию, которую применяют для предупреждения, уменьшения интенсивности, излечения или другого варианта лечения заболевания или нарушения, которая содержит действующее вещество (т.е. USP или слитый белок или молекулу нуклеиновой кислоты, кодирующую его) и фармацевтически приемлемый носитель. Понятие «фармацевтически приемлемый носитель» относится к одному или нескольким 10 совместимым твердым или жидким наполнителям, разбавителям или капсулирующим веществам, которые пригодны для введения человеку или другому позвоночному животному. В некоторых вариантах осуществления изобретения фармацевтическая композиция содержит фармацевтически приемлемый носитель, который не встречается в природе.

15 Фармацевтические композиции, которые применяют в способах, указанных в настоящем описании, приготавливают в сочетании с приемлемыми носителями, эксципиентами и другими агентами, которые обеспечивают приемлемый перенос, доставку, переносимость и т.п. Множество приемлемых форм известно специалистам в данной области. Их примерами являются (но не 20 ограничиваясь только ими) стерильный разбавитель, такой как вода для инъекций, физиологический раствор, нелетучие масла, полиэтиленгликоли, глицерин, пропиленгликоль или другие синтетические растворители; антибактериальные средства, такие как бензиловый спирт или метилпарабены; антиоксиданты, такие как аскорбиновая кислота или бисульфит натрия; 25 хелатообразующие агенты, такие как этилендиаминтетрауксусная кислота; буферы, такие как ацетаты, цитраты или фосфаты, и средства для регулирования тоничности, такие как хлорид натрия или декстроза. Конкретными носителями для внутривенного введения являются физиологический раствор или забуренный фосфатом физиологический раствор (ЗФР). Фармацевтические композиции для 30 орального или парентерального применения можно приготавливать в виде стандартной дозы лекарственной формы, пригодной для подбора дозы действующих веществ. Такие лекарственные формы в виде стандартных доз включают, например, таблетки, драже, капсулы, инъекции (ампулы),

суппозитории и т.д. Указанные композиции могут содержать также адъюванты, включая консерванты, смачивающие агенты, эмульгирующие агенты и диспергирующие агенты. Предупреждение активности микроорганизмов можно гарантировать с помощью различных антибактериальных и противогрибковых средств, например, парабенов, хлорбутанола, фенола, сорбиновой кислоты и т.п. Может требоваться включение придающих изотоничность агентов, таких, например, как сахара, хлорид натрия и т.п. Для пролонгирования абсорбции инъектируемой фармацевтической формы можно использовать агенты, замедляющие абсорбцию, например, моностеарат алюминия и желатин.

Фармацевтические композиции, содержащие USP или слитый белок, или молекулы нуклеиновых кислот, которые их кодируют, или клетки, которые их содержат, можно вводить субъекту любым путем, например, орально, трансбуккально, парентерально, местно, путем вдыхания или инсuffляции (т.е. через рот или через нос) или ректально. Введение можно осуществлять, например, один раз, несколько раз и/или в течение одного или более продолжительных периодов.

Эффективное количество фармацевтической композиции, предлагаемой в изобретении, представляет собой любое количество, которое является эффективным для достижения определенной цели (например, предотвращение или восстановление после заболевания, включая частичное восстановление, или предотвращение или замедление нарушения или заболевание, вызываемого специфической последовательностью). Эффективное количество, которое обычно выражают в мг/кг, обычные специалисты в данной области могут определять общепринятыми методами.

В тех вариантах осуществления изобретения, в который способ включает доставку полинуклеотида, кодирующего геновую РНК и/или слитый белок, клетку или эмбрион можно затем культивировать в условиях, в которых происходит экспрессия геновой РНК и/или слитого белка. В различных вариантах осуществления изобретения способ включает приведение в контакт последовательности-мишени с рибонуклеопротеиновым комплексом, содержащим гРНК и слитый белок (который содержит USP, предлагаемый в изобретении, дезаминазу и РНК-управляемый ДНК-связывающий полипептид). В конкретных вариантах осуществления изобретения способ включает

интродукцию в клетку, органеллу или эмбрион, которые содержат последовательность-мишень, рибонуклеопротеинового комплекса, предлагаемого в изобретении. Рибонуклеопротеиновый комплекс, предлагаемый в изобретении, может представлять собой комплекс, очищенный от

5 биологического образца, полученный методом рекомбинации и в последующем очищенный или собранный *in vitro* комплекс, указанный в настоящем описании. В тех вариантах осуществления изобретения, в которых рибонуклеопротеиновый комплекс, который приводят в контакт с последовательностью-мишенью или клеткой, органеллой или эмбрионом, был собран *in vitro*, способ может
10 дополнительно включать сборку *in vitro* комплекса до приведения его в контакт с последовательностью-мишенью, клеткой, органеллой или эмбрионом.

Очищенный или собранный *in vitro* рибонуклеопротеиновый комплекс, предлагаемый в изобретении, можно интродуцировать в клетку, органеллу или эмбрион с использованием любого метода, известного в данной области,
15 включая (но не ограничиваясь только ею) электропорацию. Альтернативно этому, слитый белок, содержащий USP, предлагаемый в изобретении, дезаминазу и РНК-управляемый ДНК-связывающий полипептид, и полинуклеотид, который кодирует или содержит гидовую РНК, можно интродуцировать в клетку, органеллу или эмбрион с помощью любого метода,
20 известного в данной области (например, электропорации).

После доставки или приведения в контакт с последовательностью-мишенью или клеткой, органеллой или эмбрионом, которые содержат последовательность-мишень, гидовая РНК «принуждает» слитый белок к связыванию с последовательностью-мишенью специфическим для последовательности
25 образом. Затем последовательность-мишень можно модифицировать с помощью домена дезаминазы и USP-домена слитого белка. В некоторых вариантах осуществления изобретения связывание этого слитого белка с последовательностью-мишенью приводит к модификации нуклеотида, примыкающего к последовательности-мишени. Нуклеотидное основание,
30 примыкающее к последовательности-мишени, может находиться на расстоянии 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 или 100 пар оснований от 5'- или 3'-конца последовательности-мишени. Слитый белок, содержащий USP, предлагаемый в

изобретении, дезаминазу и РНК-управляемый ДНК-связывающий полипептид, может интродуцировать целевые мутации С>Т с более высокой эффективностью по сравнению со слитым белком, который содержит только дезаминазу и РНК-управляемый ДНК-связывающий полипептид.

5 Методы оценки связывания слитого белка с последовательностью-мишенью известны в данной области и включают анализы иммунопреципитации хроматина, анализы сдвига подвижности в геле, анализы извлечения ДНК, репортерные анализы, анализ захвата на микропланшете и детекции. Аналогично
10 этому, методы оценки расщепления или модификации последовательности-мишени известны в данной области и включают анализы расщепления *in vitro* или *in vivo*, в которых расщепление подтверждают с использованием ПЦР, секвенирования или гель-электрофореза без метки или с присоединением соответствующей метки (например, радиоизотопа, флуоресцентной субстанции) к последовательности-мишени для облегчения детекции продуктов расщепления.
15 Альтернативно этому, можно применять анализ экспоненциальной реакции амплификации, опосредуемой никтирующим ферментом (NTEXPAR) (см., например, Zhang и др., Chem. Sci. 7, 2016, сс. 4951-4957). *In vivo* расщепление можно оценивать с помощью Surveyor-анализа (анализа на нуклеазу Surveyor) (Guschin и др., Methods Mol Biol 649, 2010, сс. 247-256).

20 В некоторых вариантах осуществления изобретения методы включают применение РНК-связывающего ДНК-управляемого домена в качестве части слитого белка, образующего комплекс с одной или несколькими гидовыми РНК. Для несколько гидовых РНК мишенями могут являться различные области в одном гене или мишенями могут являться несколько генов. Указанное
25 множественное таргетирование позволяет дезаминазному домену слитого белка модифицировать нуклеиновые кислоты, интродуцируя тем самым несколько мутаций в представляющую интерес молекулу нуклеиновой кислоты-мишени (например, в геном). USP-домен слитого белка повышает эффективность интродукции требуемых мутаций.

30 В тех вариантах осуществления изобретения, в которых способ включает применение РНК-управляемой нуклеазы (RGN), такой как RGN-никаза (т.е. фермента, который обладает способностью расщеплять только одну цепь двухцепочечного полинуклеотида, например, nAPG07433.1 (SEQ ID NO: 41)),

способ может включать интродукцию двух различных RGN или вариантов RGN, мишенью которых являются идентичные или перекрывающиеся последовательности-мишени, и расщепление различных цепей полинуклеотида.

Например, RGN-никазу, которая расщепляет только позитивную (+) цепь двухцепочечного полинуклеотида, можно интродуцировать наряду со второй RGN-никазой, которая расщепляет только негативную (-) цепь двухцепочечного полинуклеотида. Альтернативно этому, можно получать два различных слитых белка, при этом каждый слитый белок содержит различные RGN с различной РАМ-распознающей последовательностью, в результате чего большее разнообразие нуклеотидных последовательностей могут являться целями для мутации.

Обычному специалисту в данной области должно быть очевидно, что любой из представленных в настоящем описании способов можно применять для таргетирования одной последовательности-мишени или нескольких последовательностей-мишеней. Таким образом, способы включают применение слитого белка, содержащего один РНК-управляемый ДНК-связывающий полипептид в комбинации с несколькими различными гидовыми РНК, мишенью которых могут являться несколько различных последовательностей в одном гене и/или нескольких генах. Дезаминазный домен слитого белка затем может интродуцировать мутации в каждую из последовательностей-мишеней. USP-домен слитого белка повышает эффективность интродукции требуемых мутаций. Под объем настоящего изобретения подпадают также способы, в которых несколько различных гидовых РНК интродуцируют в комбинации с несколькими РНК-управляемыми ДНК-связывающими полипептидами. Указанные РНК-управляемые ДНК-связывающие полипептиды могут представлять собой несколько RGN или вариантов RGN. Указанные гидовые РНК и системы гидовая РНК/слитый белок могут таргетировать несколько различных последовательностей в одном гене и/или нескольких генах.

IX. Клетки, содержащие полинуклеотид с генетической модификацией

Некоторыми вариантами осуществления изобретения являются способы применения слитых белков, представленных в настоящем описании. В некоторых вариантах осуществления изобретения слитый белок применяют для интродукции точечной мутации в молекулу нуклеиновой кислоты-мишени путем

дезаминирования нуклеотидного основания-мишени, например, остатка С. В некоторых вариантах осуществления изобретения дезаминирование нуклеотидного основания-мишени приводит к коррекции генетического дефекта, например, к коррекции точечной мутации, которая приводит к утрате функции

5 генного продукта. В некоторых вариантах осуществления изобретения генетический дефект ассоциирован с заболеванием или нарушением, например, с лизосомной болезнью накопления или метаболическим заболеванием таким, например, диабет типа I. В некоторых вариантах осуществления изобретения способы, представленные в настоящем описании, применяют для интродукции

10 дезактивирующей точечной мутации в ген или аллель, который кодирует генный продукт, ассоциированный с заболеванием или нарушением. Например, в некоторых вариантах осуществления изобретения в настоящем описании предложены способы, в которых применяют слитый белок для интродукции дезактивирующей точечной мутации в онкоген (например, для лечения

15 пролиферативного заболевания). В некоторых вариантах осуществления изобретения дезактивирующая мутация приводит к образованию преждевременного стоп-кодона в кодирующей последовательности, что приводит к экспрессии усеченного генного продукта, например, усеченного белка, лишенного функции полноразмерного белка. В некоторых вариантах

20 осуществления изобретения целью способов, представленных в настоящем описании, является восстановление функции гена с нарушенной функцией путем редактирования генома. Слитые белки, представленные в настоящем описании, можно апробировать для терапий на основе редактирования генома *in vitro* в культуре человеческих клеток, предназначенных для лечения людей путем

25 коррекции заболевания, ассоциированного с мутацией. Специалистам в данной области должно быть очевидно, что слитые белки, представленные в настоящем описании, например, слитые белки, содержащие РНК-управляемый ДНК-связывающий домен, домен дезаминазы и USP, предлагаемый в изобретении, можно применять для коррекции любой единичной точечной мутации С>Т.

30 Дезаминирование мутантного С с превращением в U приводит к коррекции мутации.

В некоторых вариантах осуществления изобретения слитый белок, содержащий РНК-управляемый ДНК-связывающий домен, домен дезаминазы и

USP, предлагаемый в изобретении, можно применять для создания мутаций в гене-мишени или области-мишени представляющего интерес гена. В некоторых вариантах осуществления изобретения слитый белок, предлагаемый в изобретении, можно применять для насыщающего мутагенеза гена-мишени или области-мишени представляющего интерес гена-мишени, с последующим высокопроизводительным прямым генетическим скринингом для выявления новых мутаций и/или фенотипов. В других вариантах осуществления изобретения слитый белок, представленный в настоящем описании, можно применять для создания мутации в целевом геномном положении, которое может содержать или может не содержать последовательность ДНК. Библиотеки клеточных линий, созданных с помощью направленного мутагенеза, описанного выше, можно применять также для изучения функции генов или экспрессии генов.

Слитые белки, предлагаемые в изобретении, можно применять также для эффективного создания линий с нокаутом (КО), включая полные библиотеки КО-линий путем целевого встраивания нонсенс-мутаций. Слитые белки, содержащие РНК-управляемый ДНК-связывающий домен, домен дезаминазы и USP, предлагаемый в изобретении, могут преобразовывать три кодона (CAA, CAG и CGA) в стоп-кодона (TAG, TAA или TGA), если таргетируют кодирующую цепь ДНК, и могут преобразовывать TGG в стоп-кодон, если таргетируют некодирующую цепь ДНК У Billon и др., Mol Cell 67, 2017, сс. 1068-1079 (публикация включена в настоящее описание в качестве ссылки) представлена база данных из более чем 3,4 миллиона гидовых РНК восьми видов эукариот, которую можно применять для создания стоп-кодонов. Кроме того, Kuscu и др., Nature Methods 14(7), 2017, сс. 710-714 (публикация включена в настоящее описание в качестве ссылки) идентифицировали ~260000 уникальных i-стоп-sgРНК в геноме человека, которые могут таргетировать около 17000 генов, интродуцируя ранние (преждевременные) стоп-кодона. В некоторых вариантах осуществления изобретения КО-линии представляют собой линии эукариотических клеток. В других вариантах осуществления изобретения КО-линии представляют собой линии прокариотических клеток. В некоторых вариантах осуществления изобретения КО-линии, созданные с использованием слитого белка, предлагаемого в изобретении, представляют собой линии

человеческих клеток. В других вариантах осуществления изобретения КО-линии представляют собой линии клеток млекопитающих, например, линии клеток мышей, крыс, обезьян, кошек, собак, коров, свиней, овец или лошадей. В других вариантах осуществления изобретения КО-линии представляют собой клетки птиц. В других вариантах осуществления изобретения КО-линии представляют собой клетки насекомых. В других вариантах осуществления изобретения КО-линии представляют собой клетки микробов. В следующих вариантах осуществления изобретения КО-линии представляют собой клетки растений. В других вариантах осуществления изобретения КО-линии представляют собой клетки *Arabidopsis*, сои, кукурузы, хлопчатника, томатов, картофеля или фасоли. В следующих вариантах осуществления изобретения клеточные линии представляют собой семена растений.

В некоторых вариантах осуществления изобретения слитый белок, представленный в настоящем описании, можно применять для терапевтического редактирования генома. Например, слитый белок, содержащий РНК-управляемый ДНК-связывающий домен, домен дезаминазы и USP, предлагаемый в изобретении, можно применять для создания целенаправленных нонсенс-мутаций в PCSK9 (пропротеинконвертаза субтилизин/кексин типа 9). PCSK9 участвует в гомеостазе липопротеинов, и агенты, которые блокируют PCSK9, могут снижать концентрацию частиц липопротеинов низкой плотности (ЛПНП) в крови. Встречающиеся в естественных условиях у индивидуумов варианты PCSK9 с нонсенс-мутациями приводят к значительному снижению уровней холестерина в крови и снижают риски сердечно-сосудистых заболеваний (Cohen и др., N Engl J Med 354, 2006, сс. 1264-1272). У Chadwick и др., Arterioscler Thromb Vasc Biol 37, 2017, сс. 1741-1747; (публикация включена в настоящее описание в качестве ссылки) описана успешная интродукция целенаправленных мутаций C>T в ген PCSK9, что приводило к снижению уровней белка PCSK9 и уровней холестерина в плазме у мышей. В некоторых вариантах осуществления изобретения аналогичный подход можно осуществлять с использованием слитого белка, предлагаемого в изобретении. Кроме того, специалистам в данной области должно быть очевидно, что предложенные слитые белки можно применять для коррекции других точечных мутаций и мутаций,

ассоциированных с другими видами рака и с заболеваниями, отличными от рака, включая другие пролиферативные заболевания.

В настоящем описании представлены клетки и организмы, содержащие представляющую интерес молекулу нуклеиновой кислоты-мишени, модифицированную с использованием процесса, который опосредован слитым белком, необязательно гРНК, указанной в настоящем описании. В некоторых из указанных вариантов осуществления изобретения слитый белок содержит USP, который содержит аминокислотную последовательность, представленную в любой из SEQ ID NO: 1-16, или ее активный вариант или фрагмент. В некоторых вариантах осуществления изобретения слитый белок содержит USP, который содержит аминокислотную последовательность, идентичную по меньшей мере на 50%, по меньшей мере на 55%, по меньшей мере на 60%, по меньшей мере на 65%, по меньшей мере на 70%, по меньшей мере на 75%, по меньшей мере на 80%, по меньшей мере на 85%, по меньшей мере на 90%, по меньшей мере на 95%, по меньшей мере на 96%, по меньшей мере на 97%, по меньшей мере на 98% или по меньшей мере на 99% любой из SEQ ID NO: 1-16. В некоторых вариантах осуществления изобретения слитый белок дополнительно содержит дезаминазу и РНК-управляемый ДНК-связывающий полипептид. В других вариантах осуществления изобретения слитый белок содержит дезаминазу и RGN или ее вариант, например, имеющий аминокислотную последовательность, которая по меньшей мере на 80% идентична последовательности APG07433.1 (SEQ ID NO: 40), или ее обладающий никазной активностью вариант nAPG07433.1 (SEQ ID NO: 41). В следующих вариантах осуществления изобретения слитый белок содержит дезаминазу и RGN или ее вариант, например, имеющий аминокислотную последовательность, которая по меньшей мере на 85% идентична последовательности APG07433.1 (SEQ ID NO: 40), или ее обладающий никазной активностью вариант nAPG07433.1 (SEQ ID NO: 41). В следующих вариантах осуществления изобретения слитый белок содержит дезаминазу и RGN или ее вариант, например, имеющий аминокислотную последовательность, которая по меньшей мере на 90% идентична последовательности APG07433.1 (SEQ ID NO: 40), или ее обладающий никазной активностью вариант nAPG07433.1 (SEQ ID NO: 41). В следующих вариантах осуществления изобретения слитый белок содержит дезаминазу и RGN или ее

вариант, например, имеющий аминокислотную последовательность, которая по меньшей мере на 95% идентична последовательности APG07433.1 (SEQ ID NO: 40), или ее обладающий никазной активностью вариант nAPG07433.1 (SEQ ID NO: 41). В следующих вариантах осуществления изобретения слитый белок
5 содержит дезаминазу и RGN или ее вариант, такой как APG07433.1 (SEQ ID NO: 40), или ее обладающий никажной активностью вариант nAPG07433.1 (SEQ ID NO: 41). В других вариантах осуществления изобретения слитый белок содержит дезаминазу и Cas9 или ее вариант, такой, например, как dCas9 или никаза Cas9. В некоторых вариантах осуществления изобретения слитый белок содержит
10 вариант с инактивированной нуклеазой или вариант с никажной активностью CRISPR-Cas полипептида типа II. В других вариантах осуществления изобретения слитый белок содержит вариант с инактивированной нуклеазой или вариант с никажной активностью CRISPR-Cas полипептида типа V. В следующих вариантах осуществления изобретения слитый белок содержит вариант с
15 инактивированной нуклеазой или вариант с никажной активностью CRISPR-Cas полипептида типа VI.

Модифицированные клетки могут представлять собой эукариотические клетки (например, клетку млекопитающего, растения, насекомого, птицы) или прокариотические клетки. Предложены также органеллы и эмбрионы, которые
20 содержат по меньшей мере одну нуклеотидную последовательность, модифицированную с помощью процесса, в котором используют слитый белок, представленный в настоящем описании. Генетически модифицированные клетки, организмы, органеллы и эмбрионы могут быть гетерозиготными или гомозиготными по модифицированной нуклеотидной последовательности.

Мутация(и), интродуцированная(ые) с помощью домена дезаминазы слитого белка, может(гут) приводить к измененной экспрессии (повышающая регуляция или понижающая регуляция), к инактивации или экспрессии
25 измененного белкового продукта или интегрированной последовательности. В те случаях, когда мутация(и) приводит(ят) либо к инактивации гена, либо к экспрессии нефункционального белкового продукта, генетически
30 модифицированные клетку, организм, органеллу или эмбрион обозначают как имеющие «нокаут». Нокаут-фенотип может являться результатом делеционной мутации (т.е. делеции по меньшей мере одного нуклеотида), инсерционной

мутации (т.е. инсерции по меньшей мере одного нуклеотида) или нонсенс-мутации (т.е. замены по меньшей мере одного нуклеотида, такой как интродукция стоп-кодона).

В других вариантах осуществления изобретения мутация(и),
5 интродуцированная(ые) с помощью домена дезаминазы слитого белка, может(гут) приводить к производству варианта белкового продукта. Экспрессированный вариант белкового продукт может иметь по меньшей мере одну аминокислотную замену и/или добавление или делецию по меньшей мере одной аминокислоты. Вариант белкового продукта может обладать
10 модифицированными характеристиками или видами активности по сравнению с белком дикого типа, включая (но не ограничиваясь только ими) измененную ферментативную активность или субстратную специфичность.

В следующих других вариантах осуществления изобретения мутация(и), интродуцированная(ые) с помощью домена дезаминазы слитого белка,
15 может(гут) приводить к измененной схеме экспрессии белка. В качестве примера (но не ограничиваясь только указанным) мутация(и) в регуляторных областях, контролирующей экспрессию белкового продукта, может(гут) приводить к сверхэкспрессии или понижающей регуляции белкового продукта или измененной тканевой или временной схеме экспрессии.

20 Некоторыми объектами настоящего изобретения являются наборы, содержащие слитый белок, который содержит РНК-управляемый ДНК-связывающий полипептид, такой как полипептид RGN, например, домен Cas9 с инактивированной нуклеазой, и дезаминазу, предлагаемые в изобретении, и необязательно линкер, расположенный между Cas9-доменом и дезаминазой.
25 Кроме того, в некоторых вариантах осуществления изобретения набор содержит пригодные реагенты, буферы и/или инструкции по применению слитого белка, например, для редактирования *in vitro* или *in vivo* ДНК или РНК. В некоторых вариантах осуществления изобретения набор содержит инструкции, касающиеся конструирования и использования приемлемых гРНК для целенаправленного
30 редактирования нуклеотидной последовательности.

Х: Дополнительные варианты применения USP

USP, представленные в настоящем описании, могут также найти применение, отличное от геномного редактирования оснований. В целом, USP

могут найти применения, в которых требуется стабилизация урациловых нуклеотидных оснований в молекуле ДНК. Естественным путем или с помощью человека урацил может быть интродуцирован в геномную ДНК в результате повреждения ДНК, вызванного активными формами кислорода, ионизирующим излучением и/или алкилирующими агентами. В исследованиях механизмов репарации ДНК, таких как путь репарации, путем удаления основания (BER), или в исследованиях, в которых оценивают способность к репарации ДНК, можно использовать USP, предлагаемый в изобретении, для ингибирования репарации урацила.

Кроме того, USP, представленные в настоящем описании, могут оказаться ценными для лечения различных видов рака. Например, фторпиримидины, включая 5-фторурацил (5-FU) и его дезоксирибонуклеозидный метаболит 5-фтордезоксиуридин (5-FdU), нашли широкое применение для лечения различных солидных опухолей, включая колоректальных раков. 5-FdU обладает активностью в отношении ингибирования тимидилатсинтазы, что в результате приводит к интродукции урацила и включению 5-FU в геном клетки. Как описано выше, репарирующие основания ферменты, такие как UDG, распознают урациловые нуклеотидные основания в геномной ДНК и удаляют их. У Yan и др., Oncotarget 7 (37), 2016, сс. 59299-59313) описано, что истощенные по UDG клетки сохранялись и для них продемонстрировано устойчивое повреждение ДНК после обработки 5-FdU, что указывает на то, что действия UDG, приводящие к удалению урацила, и 5-FU играют роль в эффективности лечения опухолей с помощью 5-FdU. Доставка USP, предлагаемого в изобретении, в комбинации с фторпиримидинами, могут повышать эффективность указанного пути лечения опухолей. Таким образом, предложены фармацевтические композиции, которые содержат USP, представленный в настоящем описании, и фторпиримидин, наряду со способами лечения рака с помощью применяемых в эффективном количестве указанных фармацевтических композиций. Фторпиримидины представляют собой класс противораковых антиметаболитов, которые включают капецитабин, кармофур, доксифлуридин, фторурацил, 5-фтордезоксиуридин и тегафур.

В настоящем описании, если объект указан в единственном числе, то это включает также более чем один (т.е. по меньшей мере один) указанный объект. Например, понятие «полипептид» означает один или несколько полипептидов.

Все публикации и заявки на патент, упомянутые в спецификации, являются показателем существующего уровня техники в области, к которой относится данное изобретение. Все публикации и заявки на патент включены в настоящее описание в качестве ссылки в той же степени, как если бы каждая отдельная публикация или заявка на патент была конкретно и индивидуально включена в качестве ссылки.

Хотя представленное выше изобретение было описано в некоторых деталях в качестве иллюстрации и примера для целей ясности понимания, должно быть очевидно, что в рамках прилагаемой формулы изобретения могут быть осуществлены определенные изменения и модификации.

Вариантами осуществления изобретения, не ограничивающими его объем, являются следующие варианты:

1. Выделенный полипептид, содержащий аминокислотную последовательность, которая обладает:

а) по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1, 2, 4, 5 и 7-15;

б) по меньшей мере 81%-ной идентичностью последовательности с SEQ ID NO: 3 или 16; или

в) по меньшей мере 82%-ной идентичностью последовательности с SEQ ID NO: 6;

где указанный полипептид обладает стабилизирующей урацил активностью и где указанный полипептид дополнительно содержит гетерологичную аминокислотную последовательность.

2. Выделенный полипептид по варианту осуществления изобретения 1, где аминокислотная последовательность обладает по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

3. Выделенный полипептид по варианту осуществления изобретения 1, где аминокислотная последовательность обладает по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

4. Выделенный полипептид по варианту осуществления изобретения 1, где аминокислотная последовательность обладает по меньшей мере 95%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

5 5. Выделенный полипептид по варианту осуществления изобретения 1, где аминокислотная последовательность обладает 100%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

6. Выделенный полипептид по варианту осуществления изобретения 1, где полипептид имеет одну из последовательностей SEQ ID NO: 33-39.

10 7. Фармацевтическая композиция, содержащая не встречающийся в естественных условиях фармацевтически приемлемый носитель и полипептид, содержащий аминокислотную последовательность, которая обладает:

а) по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1, 2, 4, 5 и 7-15;

15 б) по меньшей мере 81%-ной идентичностью последовательности с SEQ ID NO: 3 или 16; или

в) по меньшей мере 82%-ной идентичностью последовательности с SEQ ID NO: 6;

в которой указанный полипептид обладает стабилизирующей урацил активностью.

20 8. Фармацевтическая композиция по варианту осуществления изобретения 7, в которой полипептид содержит аминокислотную последовательность, обладающую по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

25 9. Фармацевтическая композиция по варианту осуществления изобретения 7, в которой полипептид содержит аминокислотную последовательность, обладающую по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

30 10. Фармацевтическая композиция по варианту осуществления изобретения 7, в которой полипептид содержит аминокислотную последовательность, обладающую по меньшей мере 95%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

11. Фармацевтическая композиция по варианту осуществления изобретения 7, в которой полипептид содержит аминокислотную последовательность,

обладающую 100%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

12. Фармацевтическая композиция, содержащая не встречающийся в естественных условиях фармацевтически приемлемый носитель и молекулу нуклеиновой кислоты, которая содержит полинуклеотид, кодирующий аминокислотную последовательность, которая обладает:

а) по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1, 2, 4, 5 и 7-15;

б) по меньшей мере 81%-ной идентичностью последовательности с SEQ ID NO: 3 или 16; или

в) по меньшей мере 82%-ной идентичностью последовательности с SEQ ID NO: 6;

где указанный полипептид обладает стабилизирующей урацил активностью.

13. Фармацевтическая композиция по варианту осуществления изобретения 12, в которой полипептид содержит аминокислотную последовательность, обладающую по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

14. Фармацевтическая композиция по варианту осуществления изобретения 12, в которой полипептид содержит аминокислотную последовательность, обладающую по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

15. Фармацевтическая композиция по варианту осуществления изобретения 12, в которой полипептид содержит аминокислотную последовательность, обладающую по меньшей мере 95%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16..

16. Фармацевтическая композиция по варианту осуществления изобретения 12, в которой полипептид содержит аминокислотную последовательность, обладающую 100%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16..

17. Фармацевтическая композиция по варианту осуществления изобретения 7 или 12, в которой полипептид имеет последовательность, представленной в одной из SEQ ID NO: 33-39.

18. Фармацевтическая композиция по одному из вариантов осуществления изобретения 7-17, дополнительно содержащая фторпиримидин.

19. Молекула нуклеиновой кислоты, содержащая полинуклеотид, который кодирует полипептид, содержащий аминокислотную последовательность,

5 которая обладает:

а) по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1, 2, 4, 5 и 7-15;

б) по меньшей мере 81%-ной идентичностью последовательности с SEQ ID NO: 3 или 16; или

10 в) по меньшей мере 82%-ной идентичностью последовательности с SEQ ID NO: 6;

где указанный полипептид обладает стабилизирующей урацил активностью; и

15 где указанная молекула нуклеиновой кислоты дополнительно содержит гетерологичный промотор, функционально связанный с указанным полинуклеотидом.

20 20. Молекула нуклеиновой кислоты по варианту осуществления изобретения 19, где полипептид содержит аминокислотную последовательность, обладающую по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

21. Молекула нуклеиновой кислоты по варианту осуществления изобретения 19, где полипептид содержит аминокислотную последовательность, обладающую по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

25 22. Молекула нуклеиновой кислоты по варианту осуществления изобретения 19, где полипептид содержит аминокислотную последовательность, обладающую по меньшей мере 95%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

30 23. Молекула нуклеиновой кислоты по варианту осуществления изобретения 19, где полипептид содержит аминокислотную последовательность, обладающую 100%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

24. Молекула нуклеиновой кислоты по варианту осуществления изобретения 19, где полипептид имеет последовательность, представленную в одной из SEQ ID NO: 33-39.

25. Композиция, содержащая фторпиримидин и полипептид, который
5 содержит аминокислотную последовательность, обладающую:

а) по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1, 2, 4, 5 и 7-15;

б) по меньшей мере 81%-ной идентичностью последовательности с SEQ ID NO: 3 или 16; или

10 в) по меньшей мере 82%-ной идентичностью последовательности с SEQ ID NO: 6;

где указанный полипептид обладает стабилизирующей урацил активностью.

26. Композиция по варианту осуществления изобретения 25, в которой
15 полипептид содержит аминокислотную последовательность, обладающую по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

27. Композиция по варианту осуществления изобретения 25, в которой
20 полипептид содержит аминокислотную последовательность, обладающую по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

28. Композиция по варианту осуществления изобретения 25, в которой
25 полипептид содержит аминокислотную последовательность, обладающую по меньшей мере 95%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

29. Композиция по варианту осуществления изобретения 25, в которой полипептид содержит аминокислотную последовательность, обладающую 100%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

30. Композиция, содержащая фторпиримидин и молекулу нуклеиновой кислоты, которая кодирует полипептид, содержащий аминокислотную последовательность, которая обладает:

а) по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1, 2, 4, 5 и 7-15;

б) по меньшей мере 81%-ной идентичностью последовательности с SEQ ID NO: 3 или 16; или

в) по меньшей мере 82%-ной идентичностью последовательности с SEQ ID NO: 6;

5 где указанный полипептид обладает стабилизирующей урацил активностью.

31. Композиция по варианту осуществления изобретения 30, в которой полипептид содержит аминокислотную последовательность, обладающую по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

32. Композиция по варианту осуществления изобретения 30, в которой полипептид содержит аминокислотную последовательность, обладающую по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

15 33. Композиция по варианту осуществления изобретения 30, в которой полипептид содержит аминокислотную последовательность, обладающую по меньшей мере 95%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

20 34. Композиция по варианту осуществления изобретения 30, в которой полипептид содержит аминокислотную последовательность, обладающую 100%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

35. Композиция по варианту осуществления изобретения 25 или 30, в которой полипептид имеет последовательность, представленную в одной из SEQ ID NO: 33-39.

25 36. Слитый белок, содержащий: (I) ДНК-связывающий полипептид; (II) дезаминазу; и (III) по меньшей мере один стабилизирующий урацил полипептид (USP), обладающий по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

30 37. Слитый белок по варианту осуществления изобретения 36, в котором по меньшей мере один USP обладает по меньшей мере 85%-ной идентичностью последовательности одной из SEQ ID NO: 1-16.

38. Слитый белок по варианту осуществления изобретения 36, в котором по меньшей мере один USP обладает по меньшей мере 90%-ной идентичностью последовательности одной из SEQ ID NO: 1-16.

5 39. Слитый белок по варианту осуществления изобретения 36, в котором по меньшей мере один USP обладает по меньшей мере 95%-ной идентичностью последовательности одной из SEQ ID NO: 1-16.

40. Слитый белок по варианту осуществления изобретения 36, в котором по меньшей мере один USP обладает 100%-ной идентичностью последовательности одной из SEQ ID NO: 1-16.

10 41. Слитый белок по варианту осуществления изобретения 36, в котором USP имеет последовательность, представленную в одной из SEQ ID NO: 33-39.

42. Слитый белок по варианту осуществления изобретения 36 или 41, в котором дезаминаза представляет собой цитидиндезаминазу.

15 43. Слитый белок по варианту осуществления изобретения 42, в котором цитидиндезаминаза представляет собой индуцируемую активацией цитидиндезаминазу (AID) или член семейства дезаминаз комплекса редактирования мРНК аполипопротеина В (APOBEC).

20 44. Слитый белок по варианту осуществления изобретения 43, в котором цитидиндезаминаза содержит аминокислотную последовательность, обладающую по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94.

25 45. Слитый белок по варианту осуществления изобретения 43, в котором цитидиндезаминаза содержит аминокислотную последовательность, обладающую по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94.

46. Слитый белок по варианту осуществления изобретения 43, в котором цитидиндезаминаза содержит аминокислотную последовательность, обладающую по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94.

30 47. Слитый белок по варианту осуществления изобретения 43, в котором цитидиндезаминаза содержит аминокислотную последовательность, обладающую по меньшей мере 95%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94.

48. Слитый белок по варианту осуществления изобретения 43, в котором цитидиндезаминаза содержит аминокислотную последовательность, обладающую 100%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94.

5 49. Слитый белок по одному из вариантов осуществления изобретения 36-44, в котором ДНК-связывающий полипептид представляет собой мегануклеазу, слитый белок с доменом цинковых пальцев или TALEN.

10 50. Слитый белок по одному из вариантов осуществления изобретения 36-44, в котором ДНК-связывающий полипептид представляет собой РНК-управляемый ДНК-связывающий полипептид.

51. Слитый белок по варианту осуществления изобретения 50, в котором РНК-управляемый ДНК-связывающий полипептид представляет собой полипептид РНК-управляемой нуклеазы (RGN).

15 52. Слитый белок по варианту осуществления изобретения 51, в котором RGN представляет собой полипептид CRISPR-Cas типа II.

53. Слитый белок по варианту осуществления изобретения 51, в котором RGN представляет собой полипептид CRISPR-Cas типа в котором RGN представляет собой полипептид CRISPR-Cas типа V.

20 54. Слитый белок по варианту осуществления изобретения 51, в котором RGN содержит аминокислотную последовательность, обладающую по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 40 и 95-142.

25 55. Слитый белок по варианту осуществления изобретения 51, в котором RGN содержит аминокислотную последовательность, обладающую по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 40 и 95-142.

30 56. Слитый белок по варианту осуществления изобретения 51, в котором RGN содержит аминокислотную последовательность, обладающую по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 40 и 95-142.

57. Слитый белок по варианту осуществления изобретения 51, в котором RGN содержит аминокислотную последовательность, обладающую по меньшей

мере 95%-ной идентичностью последовательности с одной из SEQ ID NO: 40 и 95-142.

58. Слитый белок по варианту осуществления изобретения 51, в котором RGN содержит аминокислотную последовательность, обладающую 100%-ной идентичностью последовательности с одной из SEQ ID NO: 40 и 95-142.

59. Слитый белок по одному из вариантов осуществления изобретения 51-58, в котором RGN представляет собой RGN-никазу.

60. Слитый белок по одному из вариантов осуществления изобретения 51-59, в котором слитый белок содержит RGN, цитидиндезаминазу и USP.

61. Слитый белок по варианту осуществления изобретения 60, в котором слитый белок содержит RGN, обладающую по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 40, 41 и 95-142, цитидиндезаминазу, обладающую по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94, и USP, обладающий по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

62. Слитый белок по варианту осуществления изобретения 60, в котором слитый белок содержит RGN, обладающую по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 40, 41 и 95-142, цитидиндезаминазу, обладающую по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94, и USP, обладающий по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

63. Слитый белок по варианту осуществления изобретения 60, в котором слитый белок содержит RGN, обладающую по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 40, 41 и 95-142, цитидиндезаминазу, обладающую по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94, и USP, обладающий по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

64. Слитый белок по варианту осуществления изобретения 60, в котором слитый белок содержит RGN, обладающую по меньшей мере 95%-ной идентичностью последовательности с одной из SEQ ID NO: 40, 41 и 95-142,

цитидиндезаминазу, обладающую по меньшей мере 95%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94, и USP, обладающий по меньшей мере 95%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

5 65. Слитый белок по варианту осуществления изобретения 60, в котором слитый белок содержит RGN, обладающую 100%-ной идентичностью последовательности с одной из SEQ ID NO: 40, 41 и 95-142, цитидиндезаминазу, обладающую 100%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94, и USP, обладающий 100%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

10 66. Слитый белок по одному из вариантов осуществления изобретения 36-66, где слитый белок дополнительно содержит по меньшей мере один сигнал ядерной локализации (NLS).

15 67. Молекула нуклеиновой кислоты, содержащая полинуклеотид, который кодирует слитый белок, содержащий: (I) ДНК-связывающий полипептид; (II) деаминазу; и (III) по меньшей мере один стабилизирующий урацил полипептид (USP), где USP кодируется нуклеотидной последовательностью, которая:

а) обладает по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 17-32,

20 б) представлена в одной из SEQ ID NO: 17-32,

в) кодирует аминокислотную последовательность, идентичную по меньшей мере на 80% одной из SEQ ID NO: 1-16, и дополнительно включает последовательность, представленную в одной из SEQ ID NO: 33-39,

25 г) кодирует аминокислотную последовательность, идентичную по меньшей мере на 80% одной из SEQ ID NO: 1-16, или

д) кодирует аминокислотную последовательность, представленную в одной из SEQ ID NO: 1-16.

30 68. Молекула нуклеиновой кислоты по варианту осуществления изобретения 67, где USP кодируется нуклеотидной последовательностью, которая:

а) обладает по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 17-32,

б) представлена в одной из SEQ ID NO: 17-32,

в) кодирует аминокислотную последовательность, идентичную по меньшей мере на 85% одной из SEQ ID NO: 1-16, и дополнительно включает последовательность, представленную в одной из SEQ ID NO: 33-39,

5 г) кодирует аминокислотную последовательность, идентичную по меньшей мере на 85% одной из SEQ ID NO: 1-16, или

д) кодирует аминокислотную последовательность, представленную в одной из SEQ ID NO: 1-16.

10 69. Молекула нуклеиновой кислоты по варианту осуществления изобретения 67, где USP кодируется нуклеотидной последовательностью, которая:

а) обладает по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 17-32,

б) представлена в одной из SEQ ID NO: 17-32,

15 в) кодирует аминокислотную последовательность, идентичную по меньшей мере на 90% одной из SEQ ID NO: 1-16, и дополнительно включает последовательность, представленную в одной из SEQ ID NO: 33-39,

г) кодирует аминокислотную последовательность, идентичную по меньшей мере на 90% одной из SEQ ID NO: 1-16, или

20 д) кодирует аминокислотную последовательность, представленную в одной из SEQ ID NO: 1-16.

70. Молекула нуклеиновой кислоты по варианту осуществления изобретения 67, где USP кодируется нуклеотидной последовательностью, которая:

25 а) обладает по меньшей мере 95%-ной идентичностью последовательности с одной из SEQ ID NO: 17-32,

б) представлена в одной из SEQ ID NO: 17-32,

в) кодирует аминокислотную последовательность, идентичную по меньшей мере на 95% одной из SEQ ID NO: 1-16, и дополнительно включает последовательность, представленную в одной из SEQ ID NO: 33-39,

30 г) кодирует аминокислотную последовательность, идентичную по меньшей мере на 95% одной из SEQ ID NO: 1-16, или

д) кодирует аминокислотную последовательность, представленную в одной из SEQ ID NO: 1-16.

71. Молекула нуклеиновой кислоты по варианту осуществления изобретения 67, где USP кодируется нуклеотидной последовательностью, которая:

а) представлена в одной из SEQ ID NO: 17-32,

5 б) кодирует аминокислотную последовательность, идентичную на 100% одной из SEQ ID NO: 1-16, и дополнительно включает последовательность, представленную в одной из SEQ ID NO: 33-39, или

 в) кодирует аминокислотную последовательность, представленную в одной из SEQ ID NO: 1-16.

10 72. Молекула нуклеиновой кислоты по варианту осуществления изобретения 67, где дезаминаза представляет собой цитидиндезаминазу.

 73. Молекула нуклеиновой кислоты по варианту осуществления изобретения 72, где цитидиндезаминаза представляет собой индуцируемую активацией цитидиндезаминазу (AID) или член семейства дезаминаз комплекса редактирования мРНК аполипопротеина В (APOBEC)..

15 74. Молекула нуклеиновой кислоты по варианту осуществления изобретения 73, где цитидиндезаминаза содержит аминокислотную последовательность, обладающую по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94.

20 75. Молекула нуклеиновой кислоты по варианту осуществления изобретения 73, где цитидиндезаминаза содержит аминокислотную последовательность, обладающую по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94.

 76. Молекула нуклеиновой кислоты по варианту осуществления изобретения 73, где цитидиндезаминаза содержит аминокислотную последовательность, обладающую по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94.

25 77. Молекула нуклеиновой кислоты по варианту осуществления изобретения 73, где цитидиндезаминаза содержит аминокислотную последовательность, обладающую по меньшей мере 95%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94.

30 78. Молекула нуклеиновой кислоты по варианту осуществления изобретения 73, где цитидиндезаминаза содержит аминокислотную

последовательность, обладающую 100%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94.

79. Молекула нуклеиновой кислоты по одному из вариантов осуществления изобретения 67-78, где ДНК-связывающий полипептид представляет собой мегануклеазу, слитый белок с доменом цинковых пальцев или TALEN.

80. Молекула нуклеиновой кислоты по одному из вариантов осуществления изобретения 67-78, где ДНК-связывающий полипептид представляет собой РНК-управляемый ДНК-связывающий полипептид.

81. Молекула нуклеиновой кислоты по варианту осуществления изобретения 80, в где РНК-управляемый ДНК-связывающий полипептид представляет собой полипептид РНК-управляемой нуклеазы (RGN).

82. Молекула нуклеиновой кислоты по варианту осуществления изобретения 81, где RGN представляет собой полипептид CRISPR-Cas типа II.

83. Молекула нуклеиновой кислоты по варианту осуществления изобретения 81, где RGN представляет собой полипептид CRISPR-Cas типа V.

84. Молекула нуклеиновой кислоты по варианту осуществления изобретения 81, где RGN содержит аминокислотную последовательность, обладающую по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 40 и 95-142.

85. Молекула нуклеиновой кислоты по варианту осуществления изобретения 81, где RGN содержит аминокислотную последовательность, обладающую по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 40 и 95-142.

86. Молекула нуклеиновой кислоты по варианту осуществления изобретения 81, где RGN содержит аминокислотную последовательность, обладающую по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 40 и 95-142.

87. Молекула нуклеиновой кислоты по варианту осуществления изобретения 81, где RGN содержит аминокислотную последовательность, обладающую по меньшей мере 95%-ной идентичностью последовательности с одной из SEQ ID NO: 40 и 95-142.

88. Молекула нуклеиновой кислоты по варианту осуществления изобретения 81, где RGN содержит аминокислотную последовательность,

обладающую 100%-ной идентичностью последовательности с одной из SEQ ID NO: 40 и 95-142.

89. Молекула нуклеиновой кислоты по одному из вариантов осуществления изобретения 81-88, где RGN представляет собой RGN-никазу.

5 90. Молекула нуклеиновой кислоты по одному из вариантов осуществления изобретения 81-88, где слитый белок содержит RGN, цитидиндезаминазу и USP.

91. Молекула нуклеиновой кислоты по варианту осуществления изобретения 90, где слитый белок содержит RGN, обладающую по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 40, 41 и 95-
10 142, цитидиндезаминазу, обладающую по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94, и USP, обладающий по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

92. Молекула нуклеиновой кислоты по варианту осуществления изобретения 90, где слитый белок содержит RGN, обладающую по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 40, 41 и 95-
15 142, цитидиндезаминазу, обладающую по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94, и USP, обладающий по меньшей мере 85%-ной идентичностью
20 последовательности с одной из SEQ ID NO: 1-16.

93. Молекула нуклеиновой кислоты по варианту осуществления изобретения 90, где слитый белок содержит RGN, обладающую по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 40, 41 и 95-
25 142, цитидиндезаминазу, обладающую по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94, и USP, обладающий по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

94. Молекула нуклеиновой кислоты по варианту осуществления изобретения 90, где слитый белок содержит RGN, обладающую по меньшей мере
30 95%-ной идентичностью последовательности с одной из SEQ ID NO: 40, 41 и 95-142, цитидиндезаминазу, обладающую по меньшей мере 95%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94, и

USP, обладающий по меньшей мере 95%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

95. Молекула нуклеиновой кислоты по варианту осуществления изобретения 90, где слитый белок содержит RGN, обладающую 100%-ной идентичностью последовательности с одной из SEQ ID NO: 40, 41 и 95-142, цитидиндезаминазу, обладающую 100%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94, и USP, обладающий 100%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

96. Молекула нуклеиновой кислоты по одному из вариантов осуществления изобретения 67-95, в которой полинуклеотид, кодирующий слитый белок, функционально связан на его 5'-конце с гетерологичным промотором.

97. Молекула нуклеиновой кислоты по одному из вариантов осуществления изобретения 67-95, в которой полинуклеотид, кодирующий слитый белок, функционально связан на его 3'-конце с гетерологичным терминатором.

98. Молекула нуклеиновой кислоты по одному из вариантов осуществления изобретения 67-97, где слитый белок содержит один ли несколько сигналов ядерной локализации.

99. Молекула нуклеиновой кислоты по одному из вариантов осуществления изобретения 67-98, где слитый белок имеет кодоны, оптимизированные для экспрессии в эукариотической клетке.

100. Молекула нуклеиновой кислоты по одному из вариантов осуществления изобретения 67-99, в которой слитый белок имеет кодоны, оптимизированные для экспрессии в прокариотической клетке.

101. Вектор, содержащий молекулу нуклеиновой кислоты по одному из вариантов осуществления изобретения 67-100.

102. Вектор, содержащий молекулу нуклеиновой кислоты по одному из вариантов осуществления изобретения 81-95, дополнительно содержащий по меньшей мере одну нуклеотидную последовательность, которая кодирует гидовую РНК (гРНК), обладающую способностью к гибридизации с последовательностью-мишенью.

103. Вектор по варианту осуществления изобретения 102, в котором гРНК представляет собой единую гидовую РНК.

104. Вектор по варианту осуществления изобретения 102, в котором гРНК представляет собой двойную гидовую РНК.

105. Клетка, содержащая слитый белок по одному из вариантов осуществления изобретения 36-66.

5 106. Клетка, содержащая слитый белок по одному из вариантов осуществления изобретения 51-66, где клетка дополнительно содержит гидовую РНК.

107. Клетка, содержащая молекулу нуклеиновой кислоты по одному из вариантов осуществления изобретения 67-100.

10 108. Клетка, содержащая вектор по одному из вариантов осуществления изобретения 101-104.

109. Клетка по одному из вариантов осуществления изобретения 105-108, где клетка представляет собой прокариотическую клетку.

15 110. Клетка по одному из вариантов осуществления изобретения 105-108, где клетка представляет собой эукариотическую клетку.

111. Клетка по варианту осуществления изобретения 110, где клетка представляет собой клетку насекомого, птицы или млекопитающего.

112. Клетка по варианту осуществления изобретения 110, где клетка представляет собой клетку растения или грибную клетку.

20 113. Фармацевтическая композиция, содержащая фармацевтически приемлемый носитель и молекулу нуклеиновой кислоты по одному из вариантов осуществления изобретения 19-24 и 67-100, композицию по одному из вариантов осуществления изобретения 25-35, слитый белок по одному из вариантов осуществления изобретения 36-66, вектор по одному из вариантов осуществления изобретения 101-104 или клетку по одному из вариантов осуществления изобретения 105-111.

25

114. Способ получения слитого белка, включающий культивирование клетки по одному из вариантов осуществления изобретения 105-112 в условиях, в которых происходит экспрессия слитого белка.

30 115. Способ получения слитого белка, включающий интродукцию в клетку молекулы нуклеиновой кислоты по одному из вариантов осуществления изобретения 67-100 или вектора по одному из вариантов осуществления

изобретения 101-104 и культивирование клетки в условиях, в которых происходит экспрессия слитого белка.

116. Способ по варианту осуществления изобретения 114 или 115, дополнительно включающий очистку указанного слитого белка.

5 117. Способ получения рибонуклеопротеинового комплекса, слитого с RGN, включающий интродукцию в клетку молекулы нуклеиновой кислоты по одному из вариантов осуществления изобретения 81-95 и молекулы нуклеиновой кислоты, которая содержит кассету экспрессии, кодирующую гидовую РНК, или вектора по одному из вариантов осуществления изобретения 102-104, и
10 культивирование клетки в условиях, в которых происходит экспрессия слитого белка и гРНК и образование рибонуклеопротеинового комплекса, слитого с RGN.

118. Способ по варианту осуществления изобретения 117, дополнительно включающий очистку указанного рибонуклеопротеинового комплекса, слитого с
15 RGN.

119. Система для модификации молекулы ДНК-мишени, содержащей последовательность ДНК-мишени, где указанная система содержит:

а) слитый белок, содержащий полипептид РНК-управляемой нуклеазы (RGN), цитидиндезаминазу и по меньшей мере один стабилизирующий урацил
20 полипептид (USP), где USP по меньшей мере на 80% идентичен одной из SEQ ID NO: 1-16, или нуклеотидную последовательность, кодирующую указанный слитый белок; и

б) одну или несколько гидовых РНК, обладающих способностью к гибридизации с указанной последовательностью ДНК-мишени, или одну или
25 несколько нуклеотидных последовательностей, кодирующих одну или несколько гидовых РНК (гРНК);

где указанные нуклеотидные последовательности, кодирующих одну или несколько гидовых РНК и кодирующих слитый белок, каждая функционально связаны с промотором, гетерологичным относительно указанной нуклеотидной
30 последовательности;

и

где одна или несколько гидовых РНК обладают способностью образовывать комплекс со слитым белком для того, чтобы «принуждать» указанный слитый

белок к связыванию с указанной последовательностью ДНК-мишени и модифицировать молекулу ДНК-мишени.

120. Система по варианту осуществления изобретения 119, в которой USP по меньшей мере на 85% идентичен одной из SEQ ID NO: 1-16.

5 121. Система по варианту осуществления изобретения 119, в которой USP по меньшей мере на 90% идентичен одной из SEQ ID NO: 1-16.

122. Система по варианту осуществления изобретения 119, в которой USP по меньшей мере на 95% идентичен одной из SEQ ID NO: 1-16.

10 123. Система по варианту осуществления изобретения 119, в которой USP на 100% идентичен одной из SEQ ID NO: 1-16.

124. Система по варианту осуществления изобретения 119, в которой USP содержит последовательность, представленную в одной из SEQ ID NO: 33-39.

15 125. Система по одному из вариантов осуществления изобретения 119-124, в которой последовательность ДНК-мишени локализована по соседству с примыкающим к протоспейсеру мотивом (PAM), который распознается RGN.

126. Система по одному из вариантов осуществления изобретения 119-125, в которой молекула ДНК-мишени находится внутри клетки.

127. Система по варианту осуществления изобретения 126, в которой клетка представляет собой эукариотическую клетку.

20 128. Система по варианту осуществления изобретения 127, в которой эукариотическая клетка представляет собой клетку растения.

129. Система по варианту осуществления изобретения 127, в которой эукариотическая клетка представляет собой клетку млекопитающего.

25 130. Система по варианту осуществления изобретения 127, в которой эукариотическая клетка представляет собой клетку насекомого.

131. Система по варианту осуществления изобретения 126, в которой клетка представляет собой прокариотическую клетку.

30 132. Система по одному из вариантов осуществления изобретения 119-131, в которой RGN в слитом белке представляет собой полипептид CRISPR-Cas типа II.

133. Система по одному из вариантов осуществления изобретения 119-131, в которой RGN в слитом белке представляет собой полипептид CRISPR-Cas типа V.

134. Система по одному из вариантов осуществления изобретения 119-131, в которой RGN в слитом белке по меньшей мере на 80% идентична одной из SEQ ID NO: 40 и 95-142.

5 135. Система по одному из вариантов осуществления изобретения 119-131, в которой RGN в слитом белке по меньшей мере на 85% идентична одной из SEQ ID NO: 40 и 95-142.

136. Система по одному из вариантов осуществления изобретения 119-131, в которой RGN в слитом белке по меньшей мере на 90% идентична одной из SEQ ID NO: 40 и 95-142.

10 137. Система по одному из вариантов осуществления изобретения 119-131, в которой RGN в слитом белке по меньшей мере на 95% идентична одной из SEQ ID NO: 40 и 95-142.

138. Система по одному из вариантов осуществления изобретения 119-131, в которой RGN в слитом белке на 100% идентична одной из SEQ ID NO: 40 и 95-142.

139. Система по одному из вариантов осуществления изобретения 119-138, в которой RGN в слитом белке представляет собой RGN-никазу.

20 140. Система по одному из вариантов осуществления изобретения 119-139, в которой цитидиндезаминаза по меньшей мере на 80% идентична одной из SEQ ID NO: 47, 48 и 76-94.

141. Система по одному из вариантов осуществления изобретения 119-139, в которой цитидиндезаминаза по меньшей мере на 85% идентична одной из SEQ ID NO: 47, 48 и 76-94.

25 142. Система по одному из вариантов осуществления изобретения 119-139, в которой цитидиндезаминаза по меньшей мере на 90% идентична одной из SEQ ID NO: 47, 48 и 76-94.

143. Система по одному из вариантов осуществления изобретения 119-139, в которой цитидиндезаминаза по меньшей мере на 95% идентична одной из SEQ ID NO: 47, 48 и 76-94.

30 144. Система по одному из вариантов осуществления изобретения 119-139, в которой цитидиндезаминаза на 100% идентична одной из SEQ ID NO: 47, 48 и 76-94.

145. Система по одному из вариантов осуществления изобретения 119-144, в которой слитый белок содержит RGN, обладающую по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 40, 41, 95 и 142, цитидиндезаминазу, обладающую по меньшей мере 80%-ной идентичностью последовательности с одной из EQ ID NO: 47, 48 и 76-94, и USP, обладающий по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

146. Система по одному из вариантов осуществления изобретения 119-144, в которой слитый белок содержит RGN, обладающую по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 40, 41, 95 и 142, цитидиндезаминазу, обладающую по меньшей мере 85%-ной идентичностью последовательности с одной из EQ ID NO: 47, 48 и 76-94, и USP, обладающий по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

147. Система по одному из вариантов осуществления изобретения 119-144, в которой слитый белок содержит RGN, обладающую по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 40, 41, 95 и 142, цитидиндезаминазу, обладающую по меньшей мере 90%-ной идентичностью последовательности с одной из EQ ID NO: 47, 48 и 76-94, и USP, обладающий по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

148. Система по одному из вариантов осуществления изобретения 119-144, в которой слитый белок содержит RGN, обладающую по меньшей мере 95%-ной идентичностью последовательности с одной из SEQ ID NO: 40, 41, 95 и 142, цитидиндезаминазу, обладающую по меньшей мере 95%-ной идентичностью последовательности с одной из EQ ID NO: 47, 48 и 76-94, и USP, обладающий по меньшей мере 95%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

149. Система по одному из вариантов осуществления изобретения 119-144, в которой слитый белок содержит RGN, обладающую 100%-ной идентичностью последовательности с одной из SEQ ID NO: 40, 41, 95 и 142, цитидиндезаминазу, обладающую 100%-ной идентичностью последовательности с одной из EQ ID

NO: 47, 48 и 76-94, и USP, обладающий 100%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

150. Система по одному из вариантов осуществления изобретения 119-149, в которой слитый белок содержит один или несколько сигналов ядерной
5 локализации.

151. Система по одному из вариантов осуществления изобретения 119-150, в которой слитый белок имеет кодоны, оптимизированные для экспрессии в эукариотической клетке.

152. Система по одному из вариантов осуществления изобретения 119-151,
10 в которой нуклеотидные последовательности, кодирующие одну или несколько гидовых РНК и нуклеотидная последовательность, кодирующая слитый белок, локализованы в одном векторе.

153. Способ модификации молекулы ДНК-мишени, содержащей последовательность ДНК-мишени, где указанный способ включает доставку
15 системы по одному из вариантов осуществления изобретения 119-152 к указанной молекуле ДНК-мишени или клетке, содержащей молекулу ДНК-мишень.

154. Способ по варианту осуществления изобретения 153, в котором указанная модифицированная молекула ДНК-мишени содержит мутацию С>Т по
20 меньшей мере одного нуклеотида в молекуле ДНК-мишени.

155. Способ по варианту осуществления изобретения 153, в котором указанная модифицированная молекула ДНК-мишени содержит мутацию С>Т по меньшей мере одного нуклеотида в последовательности ДНК-мишени.

156. Способ модификации молекулы ДНК-мишени, содержащей
25 последовательность-мишень, включающий:

а) сборку рибонуклеотидного комплекса RGN-дезаминаза-USP путем объединения *in vitro*:

I) одной или нескольких гидовых РНК, обладающих способностью к гибридизации с последовательностью ДНК-мишени; и

30 II) слитого белка, содержащего полипептид РНК-управляемой нуклеазы (RGN), цитидиндезаминазу и по меньшей мере один стабилизирующий урацил полипептид (USP), где USP по меньшей мере на 80% идентичен одной из SEQ ID NO: 1-16;

в условиях, пригодных для образования рибонуклеотидного комплекса RGN-дезаминаза-USP; и

б) приведение в контакт указанной молекулы ДНК-мишени или клетки, содержащей указанную молекулу ДНК-мишени, с собранным *in vitro*

5 рибонуклеотидным комплексом RGN-дезаминаза-USP;

в котором одна или несколько гидовых РНК гибридизуются с последовательностью ДНК-мишени, «принуждая» тем самым указанный слитый белок к связыванию с указанной последовательностью ДНК-мишени, в результате чего происходит модификация молекулы ДНК-мишени.

10 157. Способ по варианту осуществления изобретения 156, в котором USP по меньшей мере на 85% идентичен одной из SEQ ID NO: 1-16.

158. Способ по варианту осуществления изобретения 156, в котором USP по меньшей мере на 90% идентичен одной из SEQ ID NO: 1-16.

15 159. Способ по варианту осуществления изобретения 156, в котором USP по меньшей мере на 95% идентичен одной из SEQ ID NO: 1-16.

160. Способ по варианту осуществления изобретения 156, в котором USP на 100% идентичен одной из SEQ ID NO: 1-16.

161. Способ по варианту осуществления изобретения 156, в котором USP содержит последовательность, представленную в одной из SEQ ID NO: 33-39.

20 162. Способ по одному из вариантов осуществления изобретения 156-161, в котором указанная модифицированная молекула ДНК-мишени содержит мутацию С>Т по меньшей мере одного нуклеотида в молекуле ДНК-мишени.

25 163. Способ по одному из вариантов осуществления изобретения 156-161, в котором указанная модифицированная молекула ДНК-мишени содержит мутацию С>Т по меньшей мере одного нуклеотида в последовательности ДНК-мишени.

164. Способ по одному из вариантов осуществления изобретения 156-163, в котором RGN в слитом белке представляет собой полипептид CRISPR-Cas типа II.

30 165. Способ по одному из вариантов осуществления изобретения 156-163, в котором RGN в слитом белке представляет собой полипептид CRISPR-Cas типа V.

166. Способ по одному из вариантов осуществления изобретения 156-163, в котором RGN в слитом белке по меньшей мере на 80% идентична одной из SEQ ID NO: 40 и 95-142.

5 167. Способ по одному из вариантов осуществления изобретения 156-163, в котором RGN в слитом белке по меньшей мере на 85% идентична одной из SEQ ID NO: 40 и 95-142.

168. Способ по одному из вариантов осуществления изобретения 156-163, в котором RGN в слитом белке по меньшей мере на 90% идентична одной из SEQ ID NO: 40 и 95-142.

10 169. Способ по одному из вариантов осуществления изобретения 156-163, в котором RGN в слитом белке по меньшей мере на 95% идентична одной из SEQ ID NO: 40 и 95-142.

15 170. Способ по одному из вариантов осуществления изобретения 156-163, в котором RGN в слитом белке на 100% идентична одной из SEQ ID NO: 40 и 95-142.

171. Способ по одному из вариантов осуществления изобретения 156-170, в котором RGN в слитом белке представляет собой RGN-никазу.

20 172. Способ по одному из вариантов осуществления изобретения 156-171, в котором цитидиндезаминаза по меньшей мере на 80% идентична одной из SEQ ID NO: 47, 48 и 76-94.

173. Способ по одному из вариантов осуществления изобретения 156-171, в котором цитидиндезаминаза по меньшей мере на 85% идентична одной из SEQ ID NO: 47, 48 и 76-94.

25 174. Способ по одному из вариантов осуществления изобретения 156-171, в котором цитидиндезаминаза по меньшей мере на 90% идентична одной из SEQ ID NO: 47, 48 и 76-94.

175. Способ по одному из вариантов осуществления изобретения 156-171, в котором цитидиндезаминаза по меньшей мере на 95% идентична одной из SEQ ID NO: 47, 48 и 76-94.

30 176. Способ по одному из вариантов осуществления изобретения 156-171, в котором цитидиндезаминаза на 100% идентична одной из SEQ ID NO: 47, 48 и 76-94.

177. Способ по одному из вариантов осуществления изобретения 156-176, в котором слитый белок содержит RGN, обладающую по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 40, 41 и 95-142, цитидиндезаминазу, обладающую по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94, и USP, обладающий по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

178. Способ по одному из вариантов осуществления изобретения 156-176, в котором слитый белок содержит RGN, обладающую по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 40, 41 и 95-142, цитидиндезаминазу, обладающую по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94, и USP, обладающий по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

179. Способ по одному из вариантов осуществления изобретения 156-176, в котором слитый белок содержит RGN, обладающую по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 40, 41 и 95-142, цитидиндезаминазу, обладающую по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94, и USP, обладающий по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

180. Способ по одному из вариантов осуществления изобретения 156-176, в котором слитый белок содержит RGN, обладающую по меньшей мере 95%-ной идентичностью последовательности с одной из SEQ ID NO: 40, 41 и 95-142, цитидиндезаминазу, обладающую по меньшей мере 95%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94, и USP, обладающий по меньшей мере 895%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

181. Способ по одному из вариантов осуществления изобретения 156-176, в котором слитый белок содержит RGN, обладающую 100%-ной идентичностью последовательности с одной из SEQ ID NO: 40, 41 и 95-142, цитидиндезаминазу, обладающую 100%-ной идентичностью последовательности с одной из SEQ ID

NO: 47, 48 и 76-94, и USP, обладающий 100%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

182. Способ по одному из вариантов осуществления изобретения 156-181, в котором слитый белок содержит один или несколько сигналов ядерной
5 локализации.

183. Способ по одному из вариантов осуществления изобретения 156-182, в котором слитый белок имеет кодоны, оптимизированные для экспрессии в эукариотической клетке.

184. Способ по одному из вариантов осуществления изобретения 156-183, в
10 котором указанная последовательность ДНК-мишени расположена по соседству с примыкающим к протоспейсеру мотивом (PAM).

185. Способ по одному из вариантов осуществления изобретения 156-184, в котором молекула ДНК-мишени находится внутри клетки.

186. Способ по варианту осуществления изобретения 185, в котором клетка
15 представляет собой эукариотическую клетку.

187. Способ по варианту осуществления изобретения 186, в котором эукариотическая клетка представляет собой клетку растения.

188. Способ по варианту осуществления изобретения 186, в котором эукариотическая клетка представляет собой клетку млекопитающего.

20 189. Способ по варианту осуществления изобретения 186, в котором эукариотическая клетка представляет собой клетку насекомого.

190. Способ по варианту осуществления изобретения 185, клетка представляет собой прокариотическую клетку.

25 191. Способ по одному из вариантов осуществления изобретения 185-190, который дополнительно включает отбор клетки, содержащей указанную модифицированную молекулу ДНК.

192. Клетка, содержащая последовательность ДНК-мишени, модифицированную согласно способу по варианту осуществления изобретения 191.

30 193. Клетка по варианту осуществления изобретения 192, где клетка представляет собой эукариотическую клетку.

194. Клетка по варианту осуществления изобретения 193, где эукариотическая клетка представляет собой клетку растения.

195. Растение, содержащее клетку по варианту осуществления изобретения 194.

196. Семя, содержащее клетку по варианту осуществления изобретения 194.

197. Клетка по варианту осуществления изобретения 193, где
5 эукариотическая клетка представляет собой клетку млекопитающего.

198. Клетка по варианту осуществления изобретения 193, где
эукариотическая клетка представляет собой клетку насекомого.

199. Клетка по варианту осуществления изобретения 192, где клетка
представляет собой прокариотическую клетку.

10 200. Способ получения генетически модифицированной клетки с
коррекцией мутации, являющейся причиной генетически наследуемого
заболевания (каузальная мутация), где способ включает интродукцию в клетку:

а) слитого белка, содержащего полипептид РНК-управляемой нуклеазы
(RGN), цитидиндезаминазу и по меньшей мере один стабилизирующий урацил
15 полипептид (USP), где USP по меньшей мере на 80% идентичен одной из SEQ ID
NO: 1-16, или полинуклеотида, кодирующего указанный слитый белок, где
указанный полинуклеотид, кодирующий слитый белок, функционально связан с
промотором, обеспечивающим экспрессию слитого белка в клетке; и

б) одной или нескольких гидовых РНК (гРНК), обладающих способностью
20 к гибридизации с последовательностью ДНК-мишени, или полинуклеотида,
кодирующего указанную гРНК, где указанный полинуклеотид, кодирующий
гРНК, функционально связан с промотором, обеспечивающим экспрессию гРНК
в клетке;

в результате чего слитый белок и гРНК таргетируют локализацию
25 каузальной мутации в геноме и модифицируют геномную последовательность,
устраняя каузальную мутацию.

201. Способ по варианту осуществления изобретения 200, в котором USP по
меньшей мере на 85% идентичен одной из SEQ ID NO: 1-16.

202. Способ по варианту осуществления изобретения 200, в котором USP по
30 меньшей мере на 90% идентичен одной из SEQ ID NO: 1-16.

203. Способ по варианту осуществления изобретения 200, в котором USP по
меньшей мере на 95% идентичен одной из SEQ ID NO: 1-16.

204. Способ по варианту осуществления изобретения 200, в котором USP на 100% идентичен одной из SEQ ID NO: 1-16.

205. Способ по варианту осуществления изобретения 200, в котором указанная RGN в слитом белке представляет собой нуклеотид.

5 206. Способ по варианту осуществления изобретения 200 или 205, в котором USP содержит последовательность, представленную в одной из SEQ ID NO: 33-39.

10 207. Способ по одному из вариантов осуществления изобретения 200-206, в котором геномная модификация включает мутацию C>T по меньшей мере одного нуклеотида в последовательности ДНК-мишени.

208. Способ по одному из вариантов осуществления изобретения 200-207, в котором клетка представляет собой клетку животного.

209. Способ по варианту осуществления изобретения 208, в котором клетка животного представляет собой клетку млекопитающего.

15 210. Способ по варианту осуществления изобретения 209, в котором клетка имеет происхождение из собаки, кошки, мыши, крысы, кролика, лошади, овцы, козы, коровы, свиньи или человека.

211. Способ по одному из вариантов осуществления изобретения 200-210, в котором коррекция каузальной мутации включает интродукцию стоп-кодона.

20 212. Композиция, которая содержит:

а) слитый белок, содержащий: (I) ДНК-связывающий полипептид; и (II) дезаминазу; или молекулу нуклеиновой кислоты, кодирующую слитый белок; и

25 б) стабилизирующий урацил полипептид (USP), обладающий по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16; или молекулу нуклеиновой кислоты, кодирующую USP.

213. Композиция по варианту осуществления изобретения 212, в которой USP обладает по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

30 214. Композиция по варианту осуществления изобретения 212, в которой USP обладает по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

215. Композиция по варианту осуществления изобретения 212, в которой USP обладает по меньшей мере 95%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

5 216. Композиция по варианту осуществления изобретения 212, в которой USP обладает 100%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

10 217. Композиция по варианту осуществления изобретения 212, в которой слитый белок дополнительно содержит стабилизирующий урацил полипептид (USP), обладающий по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

218. Композиция по варианту осуществления изобретения 212, в которой слитый белок дополнительно содержит стабилизирующий урацил полипептид (USP), обладающий по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

15 219. Композиция по варианту осуществления изобретения 212, в которой слитый белок дополнительно содержит стабилизирующий урацил полипептид (USP), обладающий по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

20 220. Композиция по варианту осуществления изобретения 212, в которой слитый белок дополнительно содержит стабилизирующий урацил полипептид (USP), обладающий по меньшей мере 95%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

25 221. Композиция по варианту осуществления изобретения 212, в которой слитый белок дополнительно содержит стабилизирующий урацил полипептид (USP), обладающий 100%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

222. Композиция по варианту осуществления изобретения 212, в которой ДНК-связывающий полипептид представляет собой мегануклеазу, слитый белок с доменом цинковых пальцев или TALEN.

30 223. Композиция по варианту осуществления изобретения 212, в которой ДНК-связывающий полипептид представляет собой РНК-управляемый ДНК-связывающий полипептид.

224. Композиция по варианту осуществления изобретения 223, в которой РНК-управляемый ДНК-связывающий полипептид представляет собой полипептид РНК-управляемой нуклеазы (RGN).

5 225. Композиция по варианту осуществления изобретения 224, в которой RGN представляет собой RGN-никазу.

226. Вектор, содержащий молекулу нуклеиновой кислоты, кодирующую слитый белок, и молекулу нуклеиновой кислоты, кодирующую стабилизирующий урацил полипептид (USP), в котором указанный слитый белок содержит ДНК-связывающий полипептид и дезаминазу, и в котором указанный
10 USP обладает по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

227. Вектор по варианту осуществления изобретения 226, в котором указанный USP обладает по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

15 228. Вектор по варианту осуществления изобретения 226, в котором указанный USP обладает по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

229. Вектор по варианту осуществления изобретения 226, в котором указанный USP обладает по меньшей мере 95%-ной идентичностью
20 последовательности с одной из SEQ ID NO: 1-16.

230. Вектор по варианту осуществления изобретения 226, в котором указанный USP обладает 100%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

231. Вектор по варианту осуществления изобретения 226, в котором слитый
25 белок дополнительно содержит стабилизирующий урацил полипептид (USP), обладающий по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

232. Вектор по варианту осуществления изобретения 226, в котором слитый белок дополнительно содержит стабилизирующий урацил полипептид (USP),
30 обладающий по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

233. Вектор по варианту осуществления изобретения 226, в котором слитый белок дополнительно содержит стабилизирующий урацил полипептид (USP),

обладающий по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

234. Вектор по варианту осуществления изобретения 226, в котором слитый белок дополнительно содержит стабилизирующий урацил полипептид (USP),
5 обладающий по меньшей мере 95%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

235. Вектор по варианту осуществления изобретения 226, в котором слитый белок дополнительно содержит стабилизирующий урацил полипептид (USP),
10 обладающий 100%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

236. Вектор по варианту осуществления изобретения 226, в котором ДНК-связывающий полипептид представляет собой мегануклеазу, слитый белок с доменом цинковых пальцев или TALEN.

237. Вектор по варианту осуществления изобретения 226, в котором ДНК-связывающий полипептид представляет собой РНК-управляемый ДНК-связывающий полипептид.
15

238. Вектор по варианту осуществления изобретения 237, в котором РНК-управляемый ДНК-связывающий полипептид представляет собой полипептид РНК-управляемой нуклеазы (RGN).

20 239. Вектор по варианту осуществления изобретения 238, в котором RGN представляет собой RGN-никазу.

240. Клетка, содержащая вектор по одному из вариантов осуществления изобретения 226-239.

241. Клетка, которая содержит:

25 а) слитый белок, содержащий: (I) ДНК-связывающий полипептид; и (II) дезаминазу; или молекулу нуклеиновой кислоты, кодирующую слитый белок;
и

б) стабилизирующий урацил полипептид (USP), обладающий по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16;
30 или молекулу нуклеиновой кислоты, кодирующую USP.

242. Клетка по варианту осуществления изобретения 241, в которой USP обладает по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

243. Клетка по варианту осуществления изобретения 241, в которой USP обладает по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

5 244. Клетка по варианту осуществления изобретения 241, в которой USP обладает по меньшей мере 95%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

245. Клетка по варианту осуществления изобретения 241, в которой USP обладает 100%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

10 246. Клетка по варианту осуществления изобретения 241, в которой слитый белок дополнительно содержит стабилизирующий урацил полипептид (USP), обладающий по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

15 247. Клетка по варианту осуществления изобретения 241, в которой слитый белок дополнительно содержит стабилизирующий урацил полипептид (USP), обладающий по меньшей мере 85%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

20 248. Клетка по варианту осуществления изобретения 241, в которой слитый белок дополнительно содержит стабилизирующий урацил полипептид (USP), обладающий по меньшей мере 90%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

25 249. Клетка по варианту осуществления изобретения 241, в которой слитый белок дополнительно содержит стабилизирующий урацил полипептид (USP), обладающий по меньшей мере 95%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

250. Клетка по варианту осуществления изобретения 241, в которой слитый белок дополнительно содержит стабилизирующий урацил полипептид (USP), обладающий 100%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

30 251 Клетка по варианту осуществления изобретения 241, в которой ДНК-связывающий полипептид представляет собой мегануклеазу, слитый белок с доменом цинковых пальцев или TALEN.

252. Клетка по варианту осуществления изобретения 241, в которой ДНК-связывающий полипептид представляет собой РНК-управляемый ДНК-связывающий полипептид.

253. Клетка по варианту осуществления изобретения 252, в которой РНК-управляемый ДНК-связывающий полипептид представляет собой полипептид РНК-управляемой нуклеазы (RGN).

254. Клетка по варианту осуществления изобретения 253, в которой RGN представляет собой RGN-никазу.

Представленные ниже примеры даны для иллюстрации и не направлены на ограничение объема изобретения.

Экспериментальный раздел

Пример 1: Стабилизирующие урацил полипептиды

Аминокислотные последовательности стабилизирующих урацил полипептидов (USP), предлагаемых в изобретении, представлены в SEQ ID NO: 1-16, которые приведены в таблице 1. Все заявляемые USP имеют происхождение из *Staphylococcus* spp. и состоят из 112-116 аминокислот. Полипептиды USP, представленные в настоящем описании, несут от 16,4% до 20,7% отрицательных зарядов, для них ожидаемая величина pI составляет 3,86-4,55, исключением является APG05963, для которого ожидаемая величина pI составляет 5,26.

USP APG06351, APG03399, APG04638, APG09242, APG02463, APG04080, APG01791, APG04001 и APG03327 обладают уникальной консенсусной С-концевой последовательностью «KEGGNDHE» (SEQ ID NO: 33). USP APG05198 и APG05756 обладают уникальной С-концевой последовательностью «EKENYNNE» (SEQ ID NO: 34). APG05963 обладает уникальной С-концевой последовательностью «EKEKHKNE» (SEQ ID NO: 35); APG06702 обладает уникальной С-концевой последовательностью «DKGDDNHD» (SEQ ID NO: 36); APG05316 обладает уникальной С-концевой последовательностью «QKGGQ» (SEQ ID NO: 37); APG09230 обладает уникальной С-концевой последовательностью «KGENKYE» (SEQ ID NO: 38); и APG04100 обладает уникальной С-концевой последовательностью «KQGENNHE» (SEQ ID NO: 39).

Таблица 1: Стабилизирующие урацил полипептиды

USP ID	SEQ ID NO
APG03399	1
APG06702	2
APG05198	3
APG04638	4
APG06351	5
APG05963	6
APG09230	7
APG04100	8
APG04001	9
APG02463	10
APG09242	11
APG04080	12
APG05316	13
APG03327	14
APG01791	15
APG05756	16

Пример 2: Слитые белки, содержащие USP, обладают повышенной активностью в отношении редактирования оснований в клетках млекопитающих

- 5 Идентифицировали остатки, которые предположительно дезактивируют домен RuvC RGN APG07433.1 (SEQ ID NO: 40; публикация PCT WO 2019/236566, включена в настоящее описание в качестве ссылки), и RGN модифицировали с получением никазного варианта (nAPG07433.1; SEQ ID NO: 41). Получали слитые белки, содержащие цитидиндезаминазу, а именно
- 10 APG09980 (SEQ ID NO: 47; см. PCT/US2019/068079, включенную в настоящее описание в качестве ссылки) или APG07386CTD (SEQ ID NO: 48; см. PCT/US2019/068079). Из числа USP, представленных в таблице 1, выбирали три USP для анализа активности в слитых белках, а именно APG03399, APG06702 и APG05198. Синтезировали нуклеотидные последовательности дезаминазы, USP
- 15 и nRGN с оптимизированными для экспрессии кодонами в виде слитых белков с N-концевой меткой ядерной локализации и клонировали в экспрессионной плазмиде pTwist CMV (фирма Twist Biosciences). Слитый белок, не содержащий USP, предлагаемый в изобретении, содержал, начиная с аминоконца, NLS SV40 (SEQ ID NO: 42), функционально связанный на C-конце с меткой 3× FLAG (SEQ
- 20 ID NO: 43), функционально связанной на C-конце с дезаминазой, функционально связанной на C-конце с пептидным линкером (SEQ ID NO: 44), функционально

связанным на С-конце с nRGN (например, с nAPG07433.1, имеющей SEQ ID NO: 41), и, наконец, функционально связанной на С-конце с NLS нуклеоплазмина (SEQ ID NO: 45). Слитый белок, содержащий USP, предлагаемый в изобретении, содержал, начиная с аминоконца, NLS SV40 (SEQ ID NO: 42), функционально

5 связанный на С-конце с меткой 3× FLAG (SEQ ID NO: 43), функционально связанной на С-конце с дезаминазой, функционально связанной на С-конце с пептидным линкером (SEQ ID NO: 44), функционально связанным на С-конце с nRGN (например, SEQ ID NO: 41), функционально связанной на С-конце со

10 второй линкерной последовательностью (SEQ ID NO: 46), функционально связанной на С-конце с USP, предлагаемым в изобретении, и, наконец, функционально связанной на С-конце с NLS нуклеоплазмина (SEQ ID NO: 45). В

таблице 2 представлены полученные и протестированные в отношении активности слитые белки. Все слитые белки содержали по меньшей мере один NLS и метку 3× FLAG, как указано выше.

15 Таблица 2: Слитые белки, проанализированные в отношении редактирования C>T

Слитый белок	SEQ ID NO
APG09980-nAPG07433.1	49
APG09980-nAPG07433.1-APG03399	50
APG09980-nAPG07433.1-APG06702	51
APG09980-nAPG07433.1-APG05198	52
APG07386-CTD-nAPG07433.1	53
APG07386-CTD-nAPG07433.1-APG03399	54
APG07386-CTD-nAPG07433.1-APG06702	55
APG07386-CTD-nAPG07433.1-APG05198	56

Получали также экспрессионные плазмиды, содержащие кассету экспрессии, кодирующую sgРНК. Человеческие геномные последовательности-мишени и последовательности sgРНК для направления слитых белков к

20 геномным мишеням указаны в таблице 3. Указаны также геномные локусы для каждой последовательности-мишени.

Таблица 3: Последовательности гидовых РНК

sgРНК ID	Последовательность-мишень	Последовательность sgРНК	Геномный локус
SGN000143	57	63	69
SGN000169	58	64	70

sgPHK ID	Последовательность-мишень	Последовательность sgPHK	Геномный локус
SGN000173	59	65	71
SGN000929	60	66	72
SGN000930	61	67	73
SGN001101	62	68	74

Клетки НЕК293FT в 24-луночных планшетах при уровне конфлюэнтности 75-90% совместно трансфектировали 500 нг плазмиды, содержащей кассету экспрессии, которая содержала последовательность, кодирующую слитый белок, указанный в таблице 2, и 500 нг плазмиды, содержащей кассету экспрессии, которая кодировала sgPHK, указанную в таблице 3, с использованием реагента Lipofectamine 2000 (фирма Life Technologies). Затем клетки инкубировали при 37°C в течение 72 ч. После инкубации экстрагировали геномную ДНК с использованием NucleoSpin 96 Tissue (фирма Macherey-Nagel) согласно протоколу производителя. Амплифицировали с помощью ПЦР геномную область, фланкирующую геномный сайт-мишень, и полученные продукты очищали с использованием набора для очистки и концентрирования ДНК ZR-96 DNA Clean and Concentrator (фирма Zymo Research) согласно протоколу производителя. Затем очищенные ПЦР-продукты передавали для секвенирования нового поколения с использованием системы Illumina MiSeq (2×250). Результаты анализировали в отношении образования инделов или специфической мутации цитозина в области до нуклеотида +30, где последний нуклеотид на 3'-конце последовательности-мишени, указанной в таблице 3, представляет собой нуклеотид +1, и где нуклеотид +30 находится на расстоянии 29 нуклеотидов в обратном направлении или в 5'-направлении от нуклеотида +1 в последовательности-мишени, представленной в SEQ ID NO: 57-62 (последовательности-мишени SEQ ID NO: 57-62 указаны строчным текстом в последовательностях геномного локуса, представленных в SEQ ID NO: 69-74 соответственно).

В таблицах с 4 по 15 представлены результаты редактирования основания цитидина для каждой комбинации слитого белка из таблицы 2 и гидовой РНК из таблицы 3. Для обозначения слитых белков использованы их SEQ ID NO. Нумерация цитидинов (C) в таблицах 4-15 соответствует нумерации, использованной в предыдущем параграфе, где последний нуклеотид на 3'-конце

- последовательности-мишени обозначен как нуклеотид +1 и нумерация идет в направлении 3' → 5' в последовательности-мишени и соответствующей последовательности геномного локуса. Важно отметить, что, если сравнивать активность слитого белка дезаминаза-nRGN с активностью соответствующего белка дезаминаза-nRGN плюс USP при использовании одной и той же гидовой РНК, то превращение цитидина в требуемый тимидин является более высоким при менее частом превращении в аденозин или гуанозин.

Таблица 4: Частота редактирования C>N при использовании дезаминазы APG09980 и гидовой конструкции SGN000143

Слитый белок (SEQ ID NO)		C10	C15	C20	C21	C22	C25
49	A	2,8	0,1	0	0,2	0,6	0,1
	G	1,1	0,1	0,1	0,1	0,2	0
	T	11	0,4	0,2	0,3	2	0,7
50	A	0	0	0	0	0	0
	G	0	0	0	0	0	0
	T	23,5	1,4	1,7	2	4,6	0,7
51	A	0,1	0	0	0	0	0
	G	0	0	0	0	0	0
	T	16,3	1,2	1,5	1,6	3,1	0,6
52	A	0,2	0	0	0	0	0
	G	0	0	0	0	0	0
	T	21,4	1,6	1,1	1,4	2,5	0,7

10

Результаты, представленные в таблице 4, демонстрируют, что частота редактирования C>T в положении C10 возрастала в образцах с USP.

Таблица 5: Частота редактирования C>N при использовании дезаминазы APG09986 и гидовой конструкции SGN000143

Слитый белок (SEQ ID NO.)		C10	C15	C20	C21	C22	C25
53	A	0,1	0,4	0	0,5	1	0
	G	0	0,8	0	0	0,1	0
	T	0,5	4	0,3	0,7	3,8	0
54	A	0	0	0	0	0,1	0
	G	0	0	0	0	0	0
	T	1	4,5	1	1,7	3,1	0,1
55	A	0	0,1	0	0	0,1	0
	G	0	0	0	0	0	0
	T	1,6	9,7	2,5	3,2	5,8	0,2

Слитый белок (SEQ ID NO.)		C10	C15	C20	C21	C22	C25
56	A	0	0,1	0	0	0	0
	G	0	0	0	0	0	0
	T	1,9	10,2	2,6	3,1	7	0,2

Результаты, представленные в таблице 5, демонстрируют, что частота редактирования C>T в положении C15 возросла в образцах с USP.

Таблица 6: Частота редактирования C>N при использовании дезаминазы

5 APG09980 и гидовой конструкции SGN000169

Слитый белок (SEQ ID NO.)		C9	C13	C15	C18	C20	C23
49	A	0,5	0,8	3,6	0,3	1,3	1,3
	G	1,5	1,8	8,3	0	0,2	0,1
	T	1,5	23,9	16,3	0,3	8,6	4,9
50	A	0	0	0,6	0	0,1	0,2
	G	0,2	0	2,5	0	0	0
	T	7	38,1	35,8	0,5	16,6	13,4
51	A	0,2	0,3	1,3	0,2	0,3	0,4
	G	0,6	0,6	2,1	0,1	0,1	0
	T	9,2	42,4	40,8	0,8	20,4	16,6
52	A	0,8	1	2,4	0,7	0,6	1,5
	G	0,5	1,7	3,7	0,5	0,5	0,2
	T	8,5	37,8	35,7	1,2	17,1	14,9

Результаты, представленные в таблице 6, демонстрируют, что частота образования C>T в положениях C13 и C15 возросла при добавлении USP.

Таблица 7: Частота редактирования C>N при использовании дезаминазы

10 APG07386 и гидовой конструкции SGN000169

Слитый белок (SEQ ID NO.)		C9	C13	C15	C18	C20	C23
53	A	0,8	0,7	2	0	0,5	0,3
	G	5,4	0,8	10,2	0,1	0,1	0
	T	3,5	5,1	11,2	0,3	2,6	2,4
54	A	0,1	0,5	0,8	0	0,3	0,2
	G	0,7	0,2	2,5	0,2	0,1	0
	T	8,5	15,9	24,8	1,5	6,4	7,1
55	A	0,3	0,5	1,3	0,4	0,2	0,4
	G	2	0,5	3,5	0,1	0,2	0,1

Слитый белок (SEQ ID NO.)		C9	C13	C15	C18	C20	C23
	T	14,3	22,7	37,1	2	9,2	7,9
56	A	0,1	0,1	0,7	0	0,2	0
	G	0,7	0	2,9	0	0	0
	T	14,7	24	36,4	1,3	8,4	8,9

Результаты, представленные в таблице 7, демонстрируют, что частота образования C>T в положениях C13 и C15 возрастала при добавлении USP.

Таблица 8: Частота редактирования C>N при использовании дезаминазы

5 APG09980 и гидовой конструкции SGN000930

Слитый белок (SEQ ID NO.)		C17	C19	C22
49	A	0,7	0,4	1,4
	G	34,5	0,4	0
	T	2,6	2,4	2,2
50	A	0,3	0,1	0,3
	G	8,6	0,3	0
	T	33	2,8	3,2
51	A	0,2	0,1	0,2
	G	5,1	0,1	0,1
	T	35,7	3,6	3,3
52	A	0,4	0,3	0,9
	G	11,2	0	0
	T	23,7	2,5	2,5

Результаты, представленные в таблице 8, демонстрируют, что частота редактирования C>G в положении C17 снижалась во всех образцах с USP в пользу изменений C>T.

10 Таблица 9: Частота редактирования C>N при использовании дезаминазы
APG07386 и гидовой конструкции SGN000930

Слитый белок (SEQ ID NO.)		C17	C19	C22
53	A	0,2	0	0,4
	G	14,7	0	0,4
	T	1,1	0,2	2,2
54	A	0	0	0
	G	1,4	0	0
	T	9,1	0,4	3,7
55	A	0,4	0,1	0,3

Слитый белок (SEQ ID NO.)		C17	C19	C22
	G	2,6	0,2	0,2
	T	11,8	0,9	4,5
56	A	0,1	0,2	0
	G	2,7	0	0,1
	T	17,4	0,8	6,7

Результаты, представленные в таблице 9, демонстрируют, что частота редактирования C>G в положении C17 снижалась во всех образцах с USP в пользу изменений C>T.

5 Таблица 10: Частота редактирования C>N при использовании дезаминазы APG09980 и гидовой конструкции SGN000173

Слитый белок (SEQ ID NO.)		C1	C2	C3	C4	C7	C8	C10	C11	C17	C20
49	A	0	0	0,1	0,3	3,1	6,9	0,3	0	3,4	0,1
	G	0	0,1	0	0,2	0,3	1	0,1	0,4	3,4	0,1
	T	0,1	0	0	2,5	11,1	12,3	1,7	0,2	12,1	1,7
50	A	0	0,1	0	0,1	0,9	2,9	0	0,4	1,5	0
	G	0,1	0,1	0	0	0,5	0,8	0,4	0,3	1,9	0,2
	T	0,1	0,1	0,3	14	39	39,3	28,7	13	31,1	12
51	A	0	0	0	0	1,2	1,6	0	0	0,9	0
	G	0	0	0	0	0	1,4	0,1	0	2,2	0,1
	T	0	0	0,3	10,4	33,5	33,4	23,8	10,9	26,8	10,3
52	A	0	0	0	0	1,8	1,2	0,3	0	1,5	0
	G	0	0	0	0	0	0,8	0	0,5	2,7	0
	T	0	0	0,6	10,5	33,1	36,8	26,7	13,2	28,6	9,6

Результаты, представленные в таблице 10, демонстрируют, что частота образования C>T в положениях C4, C7, C8, C10, C11, C17 и C20 возрастала при добавлении USP.

10

Таблица 11: Частота редактирования C>N при использовании дезаминазы APG07386 и гидовой конструкции SGN000173

Слитый белок (SEQ ID NO.)		C1	C2	C3	C4	C7	C8	C10	C11	C17	C20
53	A	0	0	0	0,1	0,4	1,1	0	2,5	2,8	0
	G	0	0	0	0	0,1	0,3	0	0,5	2	0
	T	0	0	0	0,5	1,4	2,5	0	7	11,3	0,7

Слитый белок (SEQ ID NO.)		C1	C2	C3	C4	C7	C8	C10	C11	C17	C20
54	A	0	0	0	0	0	0,1	0	0,7	0,8	0
	G	0	0	0	0	0	0	0	0	0,6	0
	T	0	0	0	1,6	6,1	8,6	2,2	13,3	23	1,8
55	A	0	0	0	0	0,2	0	0	1	0,9	0
	G	0	0	0	0	0	0,3	0,1	0,1	0,5	0
	T	0	0	0,2	2,2	9,1	11,2	3,2	13	21,3	2
56	A	0	0	0	0	0	0	0	0,9	0,8	0
	G	0	0	0	0	0	0,1	0	0,2	0,5	0
	T	0	0	0,1	2,7	12,3	15	4,3	18,3	27,1	1,9

Результаты, представленные в таблице 11, демонстрируют, что частота образования C>T в положениях C7, C8, C11 и C17 возрастала при добавлении USP.

5 Таблица 12: Частота редактирования C>N при использовании дезаминазы APG09980 и гидовой конструкции SGN000929

Слитый белок (SEQ ID NO.)		C6	C9	C13	C23
49	A	0,2	1,1	0,2	5,2
	G	0,2	1,5	1,4	20,1
	T	1,2	5,2	1	7,4
50	A	0	0,1	0	1,3
	G	0	0,1	0	3,3
	T	2,2	9,4	3,2	30,8
51	A	0	0,2	0	0,8
	G	0	0,1	0	1,6
	T	1,5	9,2	3,5	34
52	A	0,5	0,6	0	1,8
	G	0,1	0,5	0	6,5
	T	1,6	8,6	2,7	22,3

Результаты, представленные в таблице 12, демонстрируют, что частота образования C>T в положении C23 возрастала при добавлении USP.

10 Таблица 13: Частота редактирования C>N при использовании дезаминазы APG07386 и гидовой конструкции SGN000929

Слитый белок (SEQ ID NO.)		C6	C9	C13	C23
53	A	1,5	0	0	1,5

Слитый белок (SEQ ID NO.)		C6	C9	C13	C23
	G	0,6	0	0,1	5,2
	T	5	0,3	0,1	2,1
54	A	0	0	0	0,1
	G	0	0	0	0,1
	T	4,2	0,3	0	4,6
55	A	0,2	0	0,1	0
	G	0	0	0	1,4
	T	10,4	0,9	0,3	8,2
56	A	0,1	0	0	0,2
	G	0,2	0	0	0,5
	T	10,8	0,6	0,4	8,5

Результаты, представленные в таблице 13, демонстрируют, что частота образования C>T в положениях C6 и C23 возрастала при добавлении USP.

Таблица 14: Частота редактирования C>N при использовании дезаминазы

5 APG09980 и гидовой конструкции SGN001101

Слитый белок (SEQ ID NO.)		C2	C6	C7	C18
49	A	0,1	1,8	0,4	3,3
	G	0	2,8	0,1	18,5
	T	0,1	6,5	1,8	5,3
50	A	0,2	0,3	0,2	0,6
	G	0,1	0,1	0,2	2,1
	T	0,1	9,1	0,7	25,4
51	A	0,2	0,1	0,1	0,3
	G	0,1	0,1	0	2,4
	T	0,1	7,9	0,7	21,8
52	A	0,2	0,9	0,2	1,2
	G	0	0,8	0	4,8
	T	0,2	7,7	1,3	19,4

Результаты, представленные в таблице 14, демонстрируют, что частота образования C>T в положении C18 возрастала при добавлении USP.

Таблица 15: Частота редактирования C>N при использовании дезаминазы

10 APG03786 и гидовой конструкции SGN001101

Слитый белок (SEQ ID NO.)		C2	C6	C7	C18
53	A	0,1	0	4	2,2

Слитый белок (SEQ ID NO.)		C2	C6	C7	C18
	G	0	0	2,1	11
	T	0	0	10,4	3,9
54	A	0,1	0	0,1	0
	G	0	0	0,1	0,6
	T	0	1,7	8,5	6,7
55	A	0,2	0,1	0,6	0,3
	G	0	0	0,5	1,4
	T	0	1,7	10,7	9,2
56	A	0,1	0	0,3	0,2
	G	0	0	0,4	1,4
	T	0	1,9	12,1	9,8

Результаты, представленные в таблице 15, демонстрируют, что частота образования C>T в положении C18 возрастала при добавлении USP. Частота конверсии C>G в положении C18 снижалась при добавлении USP.

- 5 В таблицах 16 и 17 представлены частоты образования инделов для каждой протестированной комбинации слитый белок/гидовая последовательность. Слитые белки обозначены их SEQ ID NO. Эти данные демонстрируют, что слитые белки, содержащие USP, представленные в настоящем описании, снижают частоту образования инделов во всех
- 10 протестированных геномных положениях-мишенях.

Таблица 16: Частота инсерции и делеции при использовании APG09980 и USP

ID sgГНК	SEQ ID NO: 49	SEQ ID NO: 50	SEQ ID NO: 51	SEQ ID NO: 52
SGN000143	2,78	0	0	0,62
SGN000169	18,02	1,62	3,89	9,1
SGN000173	27,46	7,86	7,17	7,48
SGN000929	3,49	0,2	0,65	2,36
SGN000930	6,03	0,81	0,65	1,65
SGN001101	2,5	1,66	0,92	1,86

- 15 Таблица 17: Частота инсерции и делеции при использовании APG03796 и USP

sgRNA ID	SEQ ID NO: 53	SEQ ID NO: 54	SEQ ID NO: 55	SEQ ID NO: 56
SGN000143	0,13	0,05	0	0,06

sgRNA ID	SEQ ID NO: 53	SEQ ID NO: 54	SEQ ID NO: 55	SEQ ID NO: 56
SGN000169	5,6	1,38	3,11	0,32
SGN000173	13,8	0,99	3,49	2,57
SGN000929	0,5	0	0,3	0,16
SGN000930	1,7	0	0,23	0,31
SGN001101	1,78	0	0,48	0,39

Пример 3: Тестирование различных форматов доставки

Для определения того, можно ли редакторы оснований доставлять в различных форматах, тестировали доставку мРНК с использованием первичных

5 Т-клеток. Очищенные CD3+ Т-клетки или PBMC подвергали оттаиванию, активировали с использованием покрытых CD3/CD28 гранул (фирма ThermoFisher) в течение 3 дней, после чего осуществляли нуклеофекцию с использованием модуля Lonza 4D-Nucleofector X и полосок Nucleocuvette. Для доставки как мРНК, так и РНП использовали набор для первичных клеток P3

10 Primary Cell. Клетки трансфектировали с использованием программ ЕО-115 и ЕН-115 для доставки мРНК и РНП соответственно. Клетки культивировали в среде для размножения клеток CTS OpTimizer T (фирма ThermoFisher), содержащей IL-2, IL-7 и IL-15 (фирма Miltenyi Biotec) в течение 4 дней после осуществления нуклеофекции, после чего собирали с использованием набора для

15 выделения из ткани геномной ДНК (Nucleospin Tissue genomic DNA isolation kit, фирма Machery Nagel).

Ампликоны, окружающие сайты редактирования, создавали с помощью ПЦР и подвергали секвенированию нового поколения (NGS) с использованием платформы Illumina Nexterra, осуществляя секвенирование спаренных концов

20 длиной 2×250 пар оснований. Оценку частоты редактирования оснований осуществляли путем расчета общей частоты замен для каждого образца. Среднее значение и количество образцов для каждой протестированной гидовой конструкции приведены в представленных ниже в таблицах 18 и 19.

Для конструкций APG09980-nAPG07433.1-APG03399 и APG05840-nAPG07433.1-APG03399 при доставке с помощью мРНК установлены высокие частоты редактирования оснований в отношении нескольких мишеней.

25 Благодаря встраиванию USP2 в конструкцию для редактирования основания

имели место очень низкие частоты образования инделов, несмотря на высокую частоту замен.

Таблица 18: Средняя частота редактирования оснований при использовании APG09980-nAPG07433.1-APG03399

Слитая конструкция	Обозначение гена	SGN	Средний % замен	Средний % инделов	N
APG09980-nAPG07433.1-APG03399	Ген 1	SGN000754	23,32917436	1,128931	6
APG09980-nAPG07433.1-APG03399	Ген 1	SGN000755	59,37254849	7,1037823	4
APG09980-nAPG07433.1-APG03399	Ген 2	SGN001061	13,60100568	0,4214674	3
APG09980-nAPG07433.1-APG03399	Ген 2	SGN001062	26,9304354	4,3225871	4
APG09980-nAPG07433.1-APG03399	Ген 2	SGN001063	75,27761104	0,8163273	4
APG09980-nAPG07433.1-APG03399	Ген 2	SGN001064	72,94658862	1,0468487	3

5

Таблица 19: Средняя частота редактирования оснований при использовании APG05840-nAPG07433.1-APG03399

Слитая конструкция	Обозначение гена	SGN	Средний % замен	Средний % инделов	N
APG05840-nAPG07433.1-APG03399	Ген 1	SGN000754	57,7775198	5,14624384	4
APG05840-nAPG07433.1-APG03399	Ген 1	SGN000755	68,352455	4,98538891	3
APG05840-nAPG07433.1-APG03399	Ген 2	SGN001061	14,6830209	0	2
APG05840-nAPG07433.1-APG03399	Ген 2	SGN001062	39,7312597	2,9887885	4
APG05840-nAPG07433.1-APG03399	Ген 2	SGN001063	70,4564399	0,25727852	4
APG05840-nAPG07433.1-APG03399	Ген 2	SGN001064	53,2112842	1,98008536	3

ПЕРВОНАЧАЛЬНАЯ ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Выделенный полипептид, содержащий аминокислотную последовательность, которая обладает:

5 а) по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1, 2, 4, 5 и 7-15;

 б) по меньшей мере 81%-ной идентичностью последовательности с SEQ ID NO: 3 или 16; или

10 в) по меньшей мере 82%-ной идентичностью последовательности с SEQ ID NO: 6;

 где указанный полипептид обладает стабилизирующей урацил активностью и где указанный полипептид дополнительно содержит гетерологичную аминокислотную последовательность.

15 2. Выделенный полипептид по п. 1, где полипептид имеет последовательность, представленную в одной из SEQ ID NO: 33-39.

20 3. Фармацевтическая композиция, содержащая не встречающийся в естественных условиях фармацевтически приемлемый носитель и полипептид, содержащий аминокислотную последовательность, которая обладает:

 а) по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1, 2, 4, 5 и 7-15;

 б) по меньшей мере 81%-ной идентичностью последовательности с SEQ ID NO: 3 или 16; или

25 в) по меньшей мере 82%-ной идентичностью последовательности с SEQ ID NO: 6;

 в которой указанный полипептид обладает стабилизирующей урацил активностью.

30 4. Фармацевтическая композиция, содержащая не встречающийся в естественных условиях фармацевтически приемлемый носитель и молекулу нуклеиновой кислоты, которая содержит полинуклеотид, кодирующий

полипептид, который содержит аминокислотную последовательность, обладающую:

а) по меньшей мере 80%-ной идентичностью последовательности с любой из SEQ ID NO: 1, 2, 4, 5 и 7-15;

5 б) по меньшей мере 81%-ной идентичностью последовательности с SEQ ID NO: 3 или 16; или

в) по меньшей мере 82%-ной идентичностью последовательности с SEQ ID NO: 6;

10 где указанный полипептид обладает стабилизирующей урацил активностью.

5. Фармацевтическая композиция по п. 3 или п. 4, в которой полипептид имеет последовательность, представленную в одной из SEQ ID NO: 33-39.

15 6. Фармацевтическая композиция по одному из п.п. 3-5, дополнительно содержащая фторпиримидин.

20 7. Молекула нуклеиновой кислоты, содержащая полинуклеотид, который кодирует полипептид, содержащий аминокислотную последовательность, которая обладает:

а) по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1, 2, 4, 5 и 7-15;

б) по меньшей мере 81%-ной идентичностью последовательности с SEQ ID NO: 3 или 16; или

25 в) по меньшей мере 82%-ной идентичностью последовательности с SEQ ID NO: 6;

где указанный полипептид обладает стабилизирующей урацил активностью; и

30 где указанная молекула нуклеиновой кислоты дополнительно содержит гетерологичный промотор, функционально связанный с указанным полинуклеотидом.

8. Молекула нуклеиновой кислоты по п. 7, где полипептид имеет последовательность, представленную в одной из SEQ ID NO: 33-39.

9. Композиция, содержащая фторпиримидин и полипептид, который

а) по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1, 2, 4, 5 и 7-15;

б) по меньшей мере 81%-ной идентичностью последовательности с SEQ ID NO: 3 или 16; или

в) по меньшей мере 82%-ной идентичностью последовательности с SEQ ID NO: 6;

где указанный полипептид обладает стабилизирующей урацил активностью.

10. Композиция, содержащая фторпиримидин и молекулу нуклеиновой кислоты, которая кодирует полипептид, содержащий аминокислотную последовательность, которая обладает:

а) по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1, 2, 4, 5 и 7-15;

б) по меньшей мере 81%-ной идентичностью последовательности с SEQ ID NO: 3 или 16; или

в) по меньшей мере 82%-ной идентичностью последовательности с SEQ ID NO: 6;

где указанный полипептид обладает стабилизирующей урацил активностью.

11. Композиция по п. 9 или п. 10, в которой полипептид имеет последовательность, представленную в одной из SEQ ID NO: 33-39.

12. Слитый белок, содержащий: (I) ДНК-связывающий полипептид; (II) дезаминазу; и (III) по меньшей мере один стабилизирующий урацил полипептид (USP), обладающий по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

13. Слитый белок по п. 12, в котором USP имеет последовательность, представленную в одной из SEQ ID NO: 33-39.

5 14. Слитый белок по п. 12 или 13, в котором дезаминаза представляет собой цитидиндезаминазу.

10 15. Слитый белок по п. 14, в котором цитидиндезаминаза представляет собой индуцируемую активацией цитидиндезаминазу (AID) или член семейства дезаминаз комплекса редактирования мРНК аполипопротеина В (APOBEC).

15 16. Слитый белок по п. 15, в котором цитидиндезаминаза содержит аминокислотную последовательность, обладающую по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94.

17. Слитый белок по одному из п.п. 12-16, в котором ДНК-связывающий полипептид представляет собой мегануклеазу, слитый белок с доменом цинковых пальцев или TALEN.

20 18. Слитый белок по одному из п.п. 12-16, в котором ДНК-связывающий полипептид представляет собой РНК-управляемый ДНК-связывающий полипептид.

25 19. Слитый белок по п. 18, в котором РНК-управляемый ДНК-связывающий полипептид представляет собой полипептид РНК-управляемой нуклеазы (RGN).

20. Слитый белок по п. 19, в котором RGN представляет собой полипептид CRISPR-Cas типа II.

30 21. Слитый белок по п. 19, в котором RGN представляет собой полипептид CRISPR-Cas типа в котором RGN представляет собой полипептид CRISPR-Cas типа V.

22. Слитый белок по п. 19, в котором RGN содержит аминокислотную последовательность, обладающую по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 40 и 95-142.

5 23. Слитый белок по одному из п.п. 19-22, в котором RGN представляет собой RGN-никазу.

24. Слитый белок по одному из п.п. 19-23, в котором слитый белок содержит RGN, цитидиндезаминазу и USP.

10

25. Слитый белок по п. 24, в котором указанная RGN обладает по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 40, 41 и 95-142, указанная цитидиндезаминаза обладает по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94, и
15 указанный USP обладает по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

26. Слитый белок по одному из п.п. 12-25, где слитый белок дополнительно содержит по меньшей мере один сигнал ядерной локализации (NLS).

20

27. Молекула нуклеиновой кислоты, содержащая полинуклеотид, кодирующий слитый белок, содержащий: (I) ДНК-связывающий полипептид; (II) деаминазу; и (III) по меньшей мере один стабилизирующий урацил полипептид (USP), где USP кодируется нуклеотидной последовательностью, которая:

25 а) обладает по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 17-32,

 б) представлена в одной из SEQ ID NO: 17-32,

 в) кодирует аминокислотную последовательность, идентичную по меньшей мере на 80% одной из SEQ ID NO: 1-16, и дополнительно включает
30 последовательность, представленную в одной из SEQ ID NO: 33-39,

 г) кодирует аминокислотную последовательность, идентичную по меньшей мере на 80% одной из SEQ ID NO: 1-16, или

д) кодирует аминокислотную последовательность, представленную в одной из SEQ ID NO: 1-16.

28. Молекула нуклеиновой кислоты по п. 27, где дезаминаза представляет собой цитидиндезаминазу.

29. Молекула нуклеиновой кислоты по п. 28, где цитидиндезаминаза представляет собой индуцируемую активацией цитидиндезаминазу (AID) или член семейства дезаминаз комплекса редактирования мРНК аполипопротеина В (APOBEC).

30. Молекула нуклеиновой кислоты по п. 29, где цитидиндезаминаза содержит аминокислотную последовательность, обладающую по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94.

31. Молекула нуклеиновой кислоты по одному из п.п. 27-30, где ДНК-связывающий полипептид представляет собой мегануклеазу, слитый белок с доменом цинковых пальцев или TALEN.

32. Молекула нуклеиновой кислоты по одному из п.п. 27-30, где ДНК-связывающий полипептид представляет собой РНК-управляемый ДНК-связывающий полипептид.

33. Молекула нуклеиновой кислоты по п. 32, в которой РНК-управляемый ДНК-связывающий полипептид представляет собой полипептид РНК-управляемой нуклеазы (RGN).

34. Молекула нуклеиновой кислоты по п. 33, где RGN представляет собой полипептид CRISPR-Cas типа II.

35. Молекула нуклеиновой кислоты по п. 33, где RGN представляет собой полипептид CRISPR-Cas типа V.

36. Молекула нуклеиновой кислоты по п. 33, где RGN содержит аминокислотную последовательность, обладающую по меньшей мере 80%-ной идентичностью с одной из SEQ ID NO: 40 и 95-142.

5

37. Молекула нуклеиновой кислоты по одному из п.п. 33-36, где RGN представляет собой RGN-никазу.

10 38. Молекула нуклеиновой кислоты по одному из п.п. 33-36, где слитый белок содержит RGN, цитидиндезаминазу и USP.

15 39. Молекула нуклеиновой кислоты по п. 38, где указанная RGN обладает по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 40, 41 и 95-142, указанная цитидиндезаминаза обладает по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94, и указанный USP обладает по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

20 40. Молекула нуклеиновой кислоты по одному из п.п. 27-39, в которой полинуклеотид, кодирующий слитый белок, функционально связан на его 5'-конце с гетерологичным промотором.

25 41. Молекула нуклеиновой кислоты по одному из п.п. 27-39, в которой полинуклеотид, кодирующий слитый белок, функционально связан на его 3'-конце с гетерологичным терминатором.

42. Молекула нуклеиновой кислоты по одному из п.п. 29-41, где слитый белок содержит один или несколько сигналов ядерной локализации.

30 43. Молекула нуклеиновой кислоты по одному из п.п. 27-42, где слитый белок имеет кодоны, оптимизированные для экспрессии в эукариотической клетке.

44. Молекула нуклеиновой кислоты по одному из п.п. 27-43, где слитый белок имеет кодоны, оптимизированные для экспрессии в прокариотической клетке.

5 45. Вектор, содержащий молекулу нуклеиновой кислоты по одному п.п. 27-44.

10 46. Вектор, содержащий молекулу нуклеиновой кислоты по одному из п.п. 33-39, который дополнительно содержит по меньшей мере одну нуклеотидную последовательность, кодирующую гидовую РНК (гРНК), которая обладает способностью к гибридизации с последовательностью-мишенью.

15 47. Вектор по п. 46, в котором гРНК представляет собой единую гидовую РНК.

48. Вектор по п. 46, в котором гРНК представляет собой двойную гидовую РНК.

20 49. Клетка, содержащая слитый белок по одному из п.п. 12-26.

50. Клетка, содержащая слитый белок по одному из п.п. 18-25, где клетка дополнительно содержит гидовую РНК.

25 51. Клетка, содержащая молекулу нуклеиновой кислоты по одному п.п. 27-44.

52. Клетка, содержащая вектор по одному из п.п. 45-48.

30 53. Клетка по одному из п.п. 49-52, где клетка представляет собой прокариотическую клетку.

54. Клетка по одному из п.п. 49-52, где клетка представляет собой эукариотическую клетку.

55. Клетка по п. 54, где клетка представляет собой клетку насекомого, птицы или млекопитающего.

5 56. Клетка по п. 54, где клетка представляет собой клетку растения или грибную клетку.

57. Фармацевтическая композиция, содержащая фармацевтически приемлемый носитель и молекулу нуклеиновой кислоты по одному из п.п. 7, 8, 10 27-44, композицию по одному из п.п. 9-11, слитый белок по одному из п.п. 12-26, вектор по одному из п.п. 45-48 или клетку по одному из п.п. 49-55.

58. Способ получения слитого белка, включающий культивирование клетки по одному из п.п. 49-56 в условиях, в которых происходит экспрессия слитого 15 белка.

59. Способ получения слитого белка, включающий интродукцию в клетку молекулы нуклеиновой кислоты по одному из п.п. 27-44 или вектора по одному из п.п. 45-48 и культивирование клетки в условиях, в которых происходит 20 экспрессия слитого белка.

60. Способ по варианту п. 58 или п. 59, дополнительно включающий очистку указанного слитого белка.

25 61. Способ получения рибонуклеопротеинового комплекса, слитого с RGN, включающий интродукцию в клетку молекулы нуклеиновой кислоты по одному из п.п. 33-39 и молекулы нуклеиновой кислоты, которая содержит кассету экспрессии, кодирующую гидовую РНК, или вектора по одному из п.п. 46-48, и культивирование клетки в условиях, в которых происходит экспрессия слитого 30 белка и гРНК и образование рибонуклеопротеинового комплекса, слитого с RGN.

62. Способ по п. 61, дополнительно включающий очистку указанного рибонуклеопротеинового комплекса, слитого с RGN.

63. Система для модификации молекулы ДНК-мишени, содержащей последовательность ДНК-мишени, где указанная система содержит:

а) слитый белок, содержащий полипептид РНК-управляемой нуклеазы (RGN), цитидиндезаминазу и по меньшей мере один стабилизирующий урацил полипептид (USP), где USP по меньшей мере на 80% идентичен одной из SEQ ID NO: 1-16, или нуклеотидную последовательность, кодирующую указанный слитый белок; и

б) одну или несколько гидовых РНК, обладающих способностью к гибридизации с указанной последовательностью ДНК-мишени, или одну или несколько нуклеотидных последовательностей, кодирующих одну или несколько гидовых РНК (гРНК);

где указанные нуклеотидные последовательности, кодирующие одну или несколько гидовых РНК и кодирующие слитый белок, каждая функционально связаны с промотором, гетерологичным относительно указанной нуклеотидной последовательности;

и

где одна или несколько гидовых РНК обладают способностью образовывать комплекс со слитым белком для того, чтобы «принуждать» указанный слитый белок к связыванию с указанной последовательностью ДНК-мишени и модификации молекулы ДНК-мишени.

64. Система по п. 63, в которой последовательность ДНК-мишени локализована по соседству с примыкающим к протоспейсеру мотивом (PAM), который распознается RGN.

65. Система по п. 63 или п. 64, в которой молекула ДНК-мишени находится внутри клетки.

66. Система по п. 65, в которой клетка представляет собой эукариотическую клетку.

67. Система по п. 66, в которой эукариотическая клетка представляет собой клетку растения.

5 68. Система по п. 66, в которой эукариотическая клетка представляет собой клетку млекопитающего.

69. Система по п. 66, в которой эукариотическая клетка представляет собой клетку насекомого.

10

70. Система по п. 65, в которой клетка представляет собой прокариотическую клетку.

15 71. Система по одному из п.п. 63-70, в которой RGN в слитом белке представляет собой полипептид CRISPR-Cas типа II.

72. Система по одному из п.п. 63-70, в которой RGN в слитом белке представляет собой полипептид CRISPR-Cas типа V.

20 73. Система по одному из п.п. 63-70, в которой RGN в слитом белке по меньшей мере на 80% идентична одной из SEQ ID NO: 40 и 95-142.

74. Система по одному из п.п. 63-73, в которой RGN в слитом белке представляет собой RGN-никазу.

25

75. Система по одному из п.п. 63-74, в которой цитидиндезаминаза по меньшей мере на 80% идентична одной из SEQ ID NO: 47, 48 и 76-94.

30 76. Система по одному из п.п. 63-75, в которой USP содержит последовательность, представленную в одной из SEQ ID NO: 33-39.

77. Система по одному из п.п. 63-76, в которой RGN обладает по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 40, 41,

95 и 142, цитидиндезаминаза обладает по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94, и USP обладает по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

5

78. Система по одному из п.п. 63-77, в которой слитый белок содержит один или несколько сигналов ядерной локализации.

10

79. Система по одному из п.п. 63-78, в которой слитый белок имеет кодоны, оптимизированные для экспрессии в эукариотической клетке.

15

80. Система по одному из п.п. 63-79, в которой нуклеотидные последовательности, кодирующие одну или несколько гидовых РНК и нуклеотидная последовательность, кодирующая слитый белок, локализованы в одном векторе.

20

81. Способ модификации молекулы ДНК-мишени, содержащей последовательность ДНК-мишени, где указанный способ включает доставку системы по одному из п.п. 63-80 к указанной молекуле ДНК-мишени или клетке, содержащей молекулу ДНК-мишень.

25

82. Способ по п. 81, в котором указанная модифицированная молекула ДНК-мишени содержит мутацию С>Т по меньшей мере одного нуклеотида в молекуле ДНК-мишени.

30

83. Способ по п. 81, в котором указанная модифицированная молекула ДНК-мишени содержит мутацию С>Т по меньшей мере одного нуклеотида в последовательности ДНК-мишени.

84. Способ модификации молекулы ДНК-мишени, содержащей последовательность-мишень, включающий:

а) сборку рибонуклеотидного комплекса RGN-дезаминаза-USP *in vitro* путем объединения:

I) одной или нескольких гидовых РНК, обладающих способностью к гибридизации с последовательностью ДНК-мишени; и

II) слитого белка, содержащего полипептид РНК-управляемой нуклеазы (RGN), цитидиндезаминазу и по меньшей мере один стабилизирующий урацил полипептид (USP), где USP по меньшей мере на 80% идентичен одной из SEQ ID NO: 1-16;

в условиях, пригодных для образования рибонуклеотидного комплекса RGN-дезаминаза-USP; и

б) приведение в контакт указанной молекулы ДНК-мишени или клетки, содержащей указанную молекулу ДНК-мишени, с собранным *in vitro* рибонуклеотидным комплексом RGN-дезаминаза-USP;

в котором одна или несколько гидовых РНК гибридизуются с последовательностью ДНК-мишени, «принуждая» тем самым указанный слитый белок к связыванию с указанной последовательностью ДНК-мишени, и в результате происходит модификация молекулы ДНК-мишени.

85. Способ по п. 84, в котором указанная модифицированная молекула ДНК-мишени содержит мутацию С>Т по меньшей мере одного нуклеотида в молекуле ДНК-мишени.

86. Способ по п. 84, в котором указанная модифицированная молекула ДНК-мишени содержит мутацию С>Т по меньшей мере одного нуклеотида в последовательности ДНК-мишени.

87. Способ по одному из п.п. 84-86, в котором RGN в слитом белке представляет собой полипептид CRISPR-Cas типа II.

88. Способ по одному из п.п. 84-86, в котором RGN в слитом белке представляет собой полипептид CRISPR-Cas типа V.

89. Способ по одному из п.п. 84-86, в котором RGN в слитом белке по меньшей мере на 80% идентична одной из SEQ ID NO: 40 и 95-142.

90. Способ по одному из п.п. 84-89, в котором RGN в слитом белке представляет собой RGN-никазу.

5 91. Способ по одному из п.п. 84-90, в котором цитидиндезаминаза по меньшей мере на 80% идентична одной из SEQ ID NO: 47, 48 и 76-94.

92. Способ по одному из п.п. 84-91, в котором USP содержит последовательность, представленную в одной из SEQ ID NO: 33-39.

10 93. Способ по одному из п.п. 84-92, в котором RGN обладает по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 40, 41 и 95-142, цитидиндезаминаза обладает по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 47, 48 и 76-94, и USP обладает по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

15

94. Способ по одному из п.п. 84-93, в котором слитый белок содержит один или несколько сигналов ядерной локализации.

20 95. Способ по одному из п.п. 84-94, в котором слитый белок имеет кодоны, оптимизированные для экспрессии в эукариотической клетке.

96. Способ по одному из п.п. 84-95, в котором указанная последовательность ДНК-мишени локализована по соседству с примыкающим к протоспейсеру мотивом (PAM).

25

97. Способ по одному из п.п. 84-96, в котором молекула ДНК-мишени находится внутри клетки.

30 98. Способ по п. 97, в котором клетка представляет собой эукариотическую клетку.

99. Способ по 98, в котором эукариотическая клетка представляет собой клетку растения.

5 100. Способ по п. 98, в котором эукариотическая клетка представляет собой клетку млекопитающего.

101. Способ по п. 98, в котором эукариотическая клетка представляет собой клетку насекомого.

10 102. Способ по п. 97, клетка представляет собой прокариотическую клетку.

103. Способ по одному из п.п. 97-102, дополнительно включающий отбор клетки, которая содержит указанную модифицированную молекулу ДНК.

15 104. Клетка, содержащая последовательность ДНК-мишени, модифицированную согласно способу по п. 103.

105. Клетка по п. 104, где клетка представляет собой эукариотическую клетку.

20 106. Клетка по п. 105, где эукариотическая клетка представляет собой клетку растения.

107. Растение, содержащее клетку по п. 106.

25 108. Семя, содержащее клетку по п. 106.

109. Клетка по п. 105, где эукариотическая клетка представляет собой клетку млекопитающего.

30 110. Клетка по п. 105, где эукариотическая клетка представляет собой клетку насекомого.

111. Клетка по п. 104, где клетка представляет собой прокариотическую клетку.

112. Способ получения генетически модифицированной клетки с
5 коррекцией мутации, являющейся причиной генетически наследуемого заболевания, где способ включает интродукцию в клетку:

а) слитого белка, содержащего полипептид РНК-управляемой нуклеазы (RGN), цитидиндезаминазу и по меньшей мере один стабилизирующий урацил полипептид (USP), где USP по меньшей мере на 80% идентичен одной из SEQ ID
10 NO: 1-16, или полинуклеотида, кодирующего указанный слитый белок, где указанный полинуклеотид, кодирующий слитый белок, функционально связан с промотором, обеспечивающим экспрессию слитого белка в клетке; и

б) одной или нескольких гидовых РНК (гРНК), обладающих способностью к гибридизации с последовательностью ДНК-мишени, или полинуклеотида,
15 кодирующего указанную гРНК, где указанный полинуклеотид, кодирующий гРНК, функционально связан с промотором, обеспечивающим экспрессию гРНК в клетке;

в котором слитый белок и гРНК таргетируют положение каузальной мутации в геноме и модифицируют геномную последовательность, устраняя
20 каузальную мутацию.

113. Способ по п. 112, в котором указанная RGN в слитом белке представляет собой нуклеазу.

25 114. Способ по п. 112 или п. 113, в котором USP содержит последовательность, представленную в одной из SEQ ID NO: 33-39.

115. Способ по одному из п.п. 112-114, в котором геномная модификация включает интродукцию мутации С>Т по меньшей мере одного нуклеотида в
30 последовательности ДНК-мишени.

116. Способ по одному из п.п. 112-115, в котором клетка представляет собой клетку животного.

117. Способ по п 116, в котором клетка животного представляет собой клетку млекопитающего.

5 118. Способ по п. 117, в котором клетка имеет происхождение из собаки, кошки, мыши, крысы, кролика, лошади, овцы, козы, коровы, свиньи или человека.

10 119. Способ по одному из п.п. 112-118, в котором коррекция каузальной мутации включает интродукцию стоп-кодона.

120. Композиция, содержащая:

а) слитый белок, содержащий: (I) ДНК-связывающий полипептид; и (II) дезаминазу; или молекулу нуклеиновой кислоты, кодирующую слитый белок; и

15 б) стабилизирующий урацил полипептид (USP), обладающий по меньшей мере 80%-ной идентичностью с одной из SEQ ID NO: 1-16; или молекулу нуклеиновой кислоты, кодирующую USP.

20 121. Композиция по п. 120, в которой слитый белок дополнительно содержит стабилизирующий урацил полипептид (USP), обладающий по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

25 122. Композиция по п. 120 или п. 121, в которой ДНК-связывающий полипептид представляет собой мегануклеазу, слитый белок с доменом цинковых пальцев или TALEN.

30 123. Композиция по п. 120 или п. 121, в которой ДНК-связывающий полипептид представляет собой РНК-управляемый ДНК-связывающий полипептид.

124. Композиция по п. 123, в которой РНК-управляемый ДНК-связывающий полипептид представляет собой полипептид РНК-управляемой нуклеазы (RGN).

125. Композиция по п. 124, в которой RGN представляет собой RGN-никазу.

5 126. Вектор, содержащий молекулу нуклеиновой кислоты, кодирующую
слитый белок, и молекулу нуклеиновой кислоты, кодирующую
стабилизирующий урацил полипептид (USP), где указанный слитый белок
содержит ДНК-связывающий полипептид и дезаминазу, и где указанный USP
обладает по меньшей мере 80%-ной идентичностью последовательности с одной
10 из SEQ ID NO: 1-16.

127. Вектор по п. 126, в котором слитый белок дополнительно содержит
стабилизирующий урацил полипептид (USP), обладающий по меньшей мере
80%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

15 128. Вектор по п. 126 или п. 127, в котором ДНК-связывающий полипептид
представляет собой мегануклеазу, слитый белок с доменом цинковых пальцев
или TALEN.

20 129. Вектор по п. 126 или 127, в котором ДНК-связывающий полипептид
представляет собой РНК-управляемый ДНК-связывающий полипептид.

130. Вектор по п. 129, в котором РНК-управляемый ДНК-связывающий
полипептид представляет собой полипептид РНК-управляемой нуклеазы (RGN).

25 131. Вектор по п. 130, в котором RGN представляет собой RGN-никазу.

132. Клетка, содержащая вектор по одному из п.п. 126-131.

30 133. Клетка, содержащая:

а) слитый белок, содержащий: (I) ДНК-связывающий полипептид; и (II)
дезаминазу; или молекулу нуклеиновой кислоты, кодирующую слитый белок;

и

б) стабилизирующий урацил полипептид (USP), обладающий по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16; или молекулу нуклеиновой кислоты, кодирующую USP.

5 134. Клетка по п. 133, в которой слитый белок дополнительно содержит стабилизирующий урацил полипептид (USP), обладающий по меньшей мере 80%-ной идентичностью последовательности с одной из SEQ ID NO: 1-16.

10 135. Клетка по п. 134, в которой ДНК-связывающий полипептид представляет собой мегануклеазу, слитый белок с доменом цинковых пальцев или TALEN.

15 136. Клетка по п. 134, в которой ДНК-связывающий полипептид представляет собой РНК-управляемый ДНК-связывающий полипептид.

137. Клетка по п. 136, в которой РНК-управляемый ДНК-связывающий полипептид представляет собой полипептид РНК-управляемой нуклеазы (RGN).

138. Клетка по п. 137, в которой RGN представляет собой RGN-никазу.