



(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ

(43) Дата публикации заявки
2021.11.18

(51) Int. Cl. C12N 15/10 (2006.01)
C12N 15/63 (2006.01)

(22) Дата подачи заявки
2019.10.14

(54) СПОСОБЫ НАПРАВЛЕННОЙ ВСТАВКИ ДНК В ГЕНЫ

(31) 62/746,497; 62/830,654; 62/864,432

(72) Изобретатель:

(32) 2018.10.16; 2019.04.08; 2019.06.20

Балтес Николас (US)

(33) US

(86) PCT/US2019/056083

(74) Представитель:

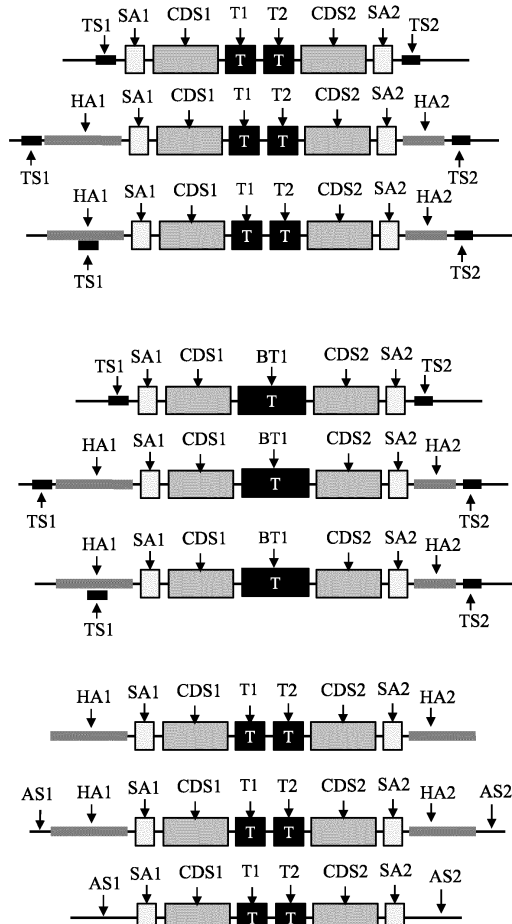
(87) WO 2020/081438 2020.04.23

Нилова М.И. (RU)

(71) Заявитель:

БЛЮЭЛЛЕЛЕ, ЛЛС (US)

(57) Способы и композиции для модификации кодирующей последовательности эндогенных генов с помощью редкощелящих эндонуклеаз и транспозаз. Описанные в настоящей заявке способы и композиции могут применяться для модификации кодирующей последовательности эндогенных генов.



СПОСОБЫ НАПРАВЛЕННОЙ ВСТАВКИ ДНК В ГЕНЫ

ССЫЛКА НА РОДСТВЕННЫЕ ЗАЯВКИ

5 Данная заявка испрашивает приоритет относительно ранее поданных и одновременно рассматриваемых заявок: USSN 62/746497, поданной 16 октября 2018 г., USSN 62/830654, поданной 8 апреля 2019 г., и USSN 62/864432, поданной 20 июня 2019 г., полное содержание которых включено в настоящий документ посредством ссылки.

СПИСОК ПОСЛЕДОВАТЕЛЬНОСТЕЙ

10 Настоящая заявка содержит Список последовательностей, представленный в формате ASCII через EFS-Web, полное содержание которого включено в данный документ посредством ссылки. Вышеуказанная копия в формате ASCII, созданная 14 октября 2019 г., называется SEQUENCE_LISTING_BA2018-4WO_P12987WO00.txt и имеет размер 517036 байт.

15 ОБЛАСТЬ ТЕХНИКИ

Настоящая заявка относится к области редактирования генома. Более конкретно, настоящая заявка относится к направленной модификации эндогенных генов с применением редкощепящих эндонуклеаз или транспозаз.

20 УРОВЕНЬ ТЕХНИКИ

Моногенные нарушения вызываются одной или более мутациями в одном гене, примеры которых включают серповидно-клеточную анемию (ген бета-гемоглобина), муковисцидоз (ген трансмембранного регулятора проводимости муковисцидоза) и болезнь Тея-Сакса (ген бета-гексозаминидазы А). Моногенные расстройства представляют интерес для генной терапии, поскольку замена дефектного гена функциональной копией может обеспечить терапевтические преимущества. Однако одним из узких мест для создания эффективных способов лечения является размер функциональной копии гена. Многие способы доставки, включая способы с применением вирусов, имеют ограничения по размеру, которые препятствуют доставке больших трансгенов. Кроме того, многие гены имеют альтернативные паттерны сплайсинга, приводящие к тому, что один ген кодирует несколько белков. Способы коррекции неполных участков дефектного гена могут стать альтернативным средством лечения моногенных заболеваний.

35 РАСКРЫТИЕ ИЗОБРЕТЕНИЯ

Редактирование генов обещает исправить мутации, обнаруженные при генетических

нарушениях; однако остается еще много проблем для создания эффективных способов лечения отдельных расстройств, в том числе тех, которые вызваны мутациями с приобретением функции, или там, где требуется точная репарация. Эти проблемы наблюдаются при таких расстройствах, как спиноцеребеллярная атаксия 3 и спиноцеребеллярная атаксия 6, при которых нарушение вызвано мутациями с приобретением функции (увеличенный тринуклеотидный повтор) на 3'-конце генов.

Представленные в настоящей заявке способы обеспечивают новые подходы к коррекции мутаций, обнаруженных на 3'-конце генов. Раскрытие в настоящей заявке основано, по меньшей мере частично, на конструкции бимодульных трансгенов, совместимых с интеграцией через несколько путей репарации. Описанные в настоящей заявке трансгены могут быть интегрированы в гены посредством пути гомологичной рекомбинации, пути негомологичного соединения концов или пути гомологичной рекомбинации и пути негомологичного соединения концов или посредством транспозиции. Кроме того, результат интеграции в любом случае (HR (гомологичная рекомбинация), прямое NHEJ (негомологичное соединение концов), обратное NHEJ; прямая транспозиция или обратная транспозиция) может привести к точной коррекции/изменению белкового продукта гена. Описанные в настоящей заявке трансгены можно применять для исправления или введения мутаций в 3'-область интересующих генов. Эти способы особенно полезны в тех случаях, когда необходимо точное редактирование генов или когда целевой мутированный эндогенный ген не может быть «заменен» синтетической копией, поскольку он превышает размер стандартных векторов или вирусных векторов. Представленные в настоящей заявке способы могут применяться для прикладных исследований (например, генная терапия) или фундаментальных исследований (например, создания моделей на животных или понимания функции генов).

Представленные в настоящей заявке способы совместимы с существующими носителями для доставки *in vivo* (например, с помощью аденоассоциированных вирусных векторов и липидных наночастиц), и они решают несколько проблем, связанных с достижением точного изменения продуктов гена.

В одном из вариантов воплощения настоящего изобретения, представлен способ интеграции трансгена в эндогенный ген. Способ может включать доставку трансгена, когда трансген несет первую и вторую акцепторные последовательности сплайсинга, первую и вторую неполные кодирующие последовательности, а также первый и второй терминатор. В некоторых вариантах воплощения изобретения, первый и второй терминаторы могут быть заменены одним двунаправленным терминатором. Способ дополнительно включает введение одной или более редкощепящих эндонуклеаз, нацеленных на участок внутри

эндогенного гена, где трансген затем интегрируется в эндогенный ген. Трансген может быть нацелен на сайт внутри интрона или на стык интрон-экзон. Первая и вторая неполные кодирующие последовательности могут быть ориентированы в направлении «хвост к хвосту», так что интеграция трансгена в любом направлении (то есть прямом или обратном) с помощью NHEJ может привести к точному изменению белкового продукта гена. В других вариантах воплощения изобретения, трансген может включать левое и правое плечо гомологии для обеспечения интеграции посредством HR. Эти трансгены могут содержаться в векторе аденоассоциированного вируса (AAV), где трансген может быть интегрирован через HR (через плечи гомологии) или посредством прямого направленного NHEJ или обратного направленного NHEJ (посредством прямой интеграции вектора AAV в целевом двухцепочечном разрыве). В одном из вариантов воплощения изобретения, векторы с первой и второй кодирующей последовательностью и левым и правым плечом гомологии могут дополнительно включать первый и второй сайты для расщепления одной или более редкощепляющимися эндонуклеазами. Расщепление одной или более редкощепляющимися эндонуклеазами может привести к высвобождению линейного трансгена с плечами гомологии, способного интегрироваться в геном через HR или NHEJ. В другом варианте воплощения изобретения, векторы с первой и второй кодирующей последовательностью могут фланкироваться первым и вторым сайтами для расщепления одной или более редкощепляющимися эндонуклеазами. Расщепление одной или более редкощепляющимися эндонуклеазами может привести к высвобождению линейного трансгена, способного интегрироваться в геном через NHEJ. В другом варианте воплощения изобретения, векторы с первой и второй кодирующей последовательностью могут фланкироваться левым и правым концами транспозона. Доставка CRISPR-ассоциированной транспозазы (например, Cas6/7/8 вместе с TniQ, TnsA, TnsB и TnsC) может привести к интеграции трансгена посредством транспозиции.

Эти способы можно применять для изменения С-конца белков, продуцируемых эндогенными генами. В некоторых вариантах воплощения изобретения, эндогенный ген может включать ген ATXN3 или ген CACNA1A. ATXN3 является геном, кодирующим фермент атаксин-3. Атаксин-3 является членом убиквитин-протеасомной системы, которая способствует разрушению избыточных или поврежденных белков. Спинаocerebellарная атаксия 3 типа представляет собой генетическое заболевание, вызванное размножением тринуклеотидного повтора на 3'-конце гена ATXN3. CACNA1A является геном, кодирующим белки, участвующие в формировании кальциевых каналов. Спинаocerebellарная атаксия 6 типа представляет собой генетическое заболевание, вызванное мутациями в гене CACNA1A. Мутации, вызывающие SCA6, включают

экспансию тринуклеотидного повтора на 3'-конце гена *SACNA1A*. В некоторых вариантах воплощения изобретения, способы, представленные в настоящей заявке, можно применять для изменения 3'-конца эндогенного гена *ATXN3* или гена *SACNA1A*. В определенных вариантах воплощения изобретения, мишенью для интеграции трансгенов, описанных в настоящей заявке, может быть интрон 9 гена *ATXN3* или интрон 46 гена *SACNA1A*.

Если не указано иное, все технические и научные термины, используемые в настоящей заявке, имеют значение, которое общепринято специалистами в области техники, к которой относится настоящее изобретение. Хотя способы и материалы, подобные или эквивалентные способам и материалам, описанным в настоящей заявке, могут быть использованы для практической реализации изобретения, подходящие способы и материалы описаны ниже. Все публикации, заявки на патенты, патенты и другие ссылки, упомянутые в настоящей заявке, включены посредством ссылки в полном объеме для любых целей. В случае возникновения расхождений настоящая заявка, включая определения, имеет преимущественную силу. Кроме того, материалы, способы и примеры носят только иллюстративный характер и не предназначены для ограничения.

Подробности одного или более вариантов воплощения изобретения изложены в описании ниже. Другие особенности, цели и преимущества изобретения станут очевидны из описания и формулы изобретения.

КРАТКОЕ ОПИСАНИЕ ЧЕРТЕЖЕЙ

На ФИГ. 1 представлены трансгены для нацеленной вставки в эндогенные гены. TS1, целевой сайт 1; SA1, акцепторный сайт сплайсинга 1, CDS1, кодирующая последовательность 1; T1, терминатор 1, TS2, целевой сайт 2; SA2, акцепторный сайт сплайсинга 2, CDS2, кодирующая последовательность 2; T2, терминатор 2; HA1, плечо гомологии 1; HA2, плечо гомологии 2; BT1, двунаправленный терминатор 1; AS1, дополнительная последовательность 1; AS2, дополнительная последовательность 2.

На ФИГ. 2 представлена интеграция трансгена в иллюстративный ген. Трансген содержит два сайта-мишени для одной или более редкощепящих эндонуклеаз, две акцепторные последовательности сплайсинга, две кодирующие последовательности (3.1 и 3.2) и два терминатора (Т). Интеграция происходит через негомологичное соединение концов (NHEJ).

На ФИГ. 3 представлена интеграция трансгена в иллюстративный ген. Трансген включает два плеча гомологии, два сайта-мишени для одной или более редкощепящих эндонуклеаз, две акцепторные последовательности сплайсинга, две кодирующие последовательности (3.1 и 3.2) и два терминатора. Интеграция происходит либо

посредством гомологичной рекомбинации (HR), либо посредством негомологичного соединения концов (NHEJ).

На ФИГ. 4 представлены экзон 46, интрон 46 и интрон 47 гена *CACNA1A*. Также показан трансген pB1011-D1 для интеграции в ген *CACNA1A*.

5 На ФИГ. 5 представлены результаты интеграции трансгена pB1011-D1 в ген *CACNA1A*.

На ФИГ. 6 представлены экзон 9, интрон 9, экзон 10, интрон 10 и экзон 11 гена *ATXN3*. Также показан трансген pB1012-D1 для интеграции в ген *ATXN3*.

На ФИГ. 7 представлены результаты интеграции трансгена pB1012-D1 в ген *ATXN3*.

10 На ФИГ. 8 представлены изображения гелей, обнаруживающих интеграцию трансгенов в гене *ATXN3*. 1, Лэддер 100 п.н. с верхней полосой 1517 п.н. 2, соединение pBA1135 5'; 3, соединение pBA1136 5'; 4, соединение pBA1137 5'; 5, соединение pBA1135 3'; 6, соединение pBA1136 3'; 7, соединение pBA1 137 3'; 8, лэддер 1 т.п.н. с более темными полосами на уровне 500 п.н., 1000 п.н. и 3000 п.н.; 9 - лэддер 1 т.п.н. с более темными
15 полосами на уровне 500 п.н., 1000 п.н. и 3000 п.н.; 10, инвертированное соединение pBA1 135 5'; 11, лэддер 1 т.п.н. с более темными полосами на уровне 500 п.н., 1000 п.н. и 3000 п.н.; 12, инвертированное соединение pBA1 136 5'; 13, лэддер 1 т.п.н. с более темными полосами на уровне 500 п.н., 1000 п.н. и 3000 п.н.; 14, пара праймеров oNJB156 + oNJB1 13;
20 oNJB1 14 + oNJB170; 15, пара праймеров 114 + 162; 16, пара праймеров oNJB1 16 + oNJB1 13; 17, пара праймеров oNJB1 14 + oNJB170; 18, пара праймеров oNJB167 + oNJB 170; 19, лэддер 100 п.н. с темной полосой на уровне 500 п.н.; 20, геномная ДНК от трансфекции pBA1 135 и нуклеазой; 21 - геномная ДНК от трансфекции pBA1 136 и нуклеазой; 22 - геномная ДНК от трансфекции pBA1 137 и нуклеазой; 23 - геномная ДНК от трансфекции водой; 24, без контроля ДНК.

25 ПОДРОБНОЕ ОПИСАНИЕ ИЗОБРЕТЕНИЯ

В настоящей заявке раскрыты способы и композиции для модификации кодирующей последовательности эндогенных генов. В некоторых вариантах воплощения изобретения, способы включают вставку трансгена в эндогенный ген, где трансген обеспечивает неполную кодирующую последовательность, которая заменяет кодирующую
30 последовательность эндогенного гена.

В одном из вариантов воплощения настоящего изобретения, описан способ интеграции трансгена в эндогенный ген; способ включает введение трансгена, при этом трансген содержит первую и вторую акцепторные последовательности сплайсинга, первую и вторую неполные кодирующие последовательности и один двунаправленный терминатор
35 или первый и второй терминатор, и введение одной или более редкощепящих эндонуклеаз,

нацеленных на сайт внутри эндогенного гена, при этом трансген интегрирован в эндогенный ген. Способ может включать конструирование трансгена так, чтобы первый акцептор сплайсинга был функционально связан с первой неполной кодирующей последовательностью, а второй акцептор сплайсинга был функционально связан со второй

5 неполной кодирующей последовательностью. Компоновка также может включать в себя наличие первой неполной кодирующей последовательности, функционально связанной с первым терминатором, и второй неполной кодирующей последовательности, функционально связанной со вторым терминатором. В варианте воплощения изобретения, два терминатора могут быть заменены одним двунаправленным терминатором. В варианте

10 воплощения изобретения, трансгены с первым и вторым акцепторами сплайсинга, первыми и вторыми неполными кодирующими последовательностями и первым и вторым терминаторами могут быть ориентированы "хвост к хвосту". Трансгены с ориентацией последовательностей «хвост к хвосту» могут дополнительно содержать первый и второй сайты-мишени для одной или более редкощепящих эндонуклеаз, при этом сайты-мишени

15 фланкируют первый и второй акцепторы сплайсинга. В другом варианте воплощения изобретения, трансгены могут содержать левое и правое плечо гомологии, которые фланкируют первый и второй акцепторы сплайсинга. В этом варианте воплощения изобретения, трансген может находиться в аденоассоциированном вирусном векторе. В другом варианте воплощения изобретения, трансген может дополнительно содержать

20 первый и второй сайты-мишени для одной или более редкощепящих эндонуклеаз, при этом сайты-мишени фланкируют первый и второй акцепторы сплайсинга. Первый и второй сайты-мишени могут фланкировать первое и второе плечо гомологии. В вариантах воплощения настоящего изобретения, трансгены могут быть интегрированы в интрон эндогенного гена или в соединение интрон-экзон. Трансгены могут быть интегрированы в

25 интрон или в соединение интрон-экзон гена ATXN3 или гена CACNA1A. Трансген может содержать первую и вторую неполные кодирующие последовательности, кодирующие пептид, продуцируемый экзоном 10 непатогенного гена ATXN3, и может быть нацелен на интрон 9 или соединение интрона 9 с экзоном 10 патогенного гена ATXN3. Трансген может содержать первую и вторую неполные кодирующие последовательности, кодирующие

30 пептид, продуцируемый экзоном 47 непатогенного гена CACNA1A, и может быть нацелен на интрон 46 или соединение интрона 46 с экзоном 47 патогенного гена CACNA1A. В некоторых вариантах воплощения изобретения, редкощепящая эндонуклеаза может быть нуклеазой CRISPR/Cas12a или нуклеазой CRISPR Cas9. Первая и вторая неполные кодирующие последовательности кодируют одни и те же аминокислоты. В варианте

35 воплощения изобретения, первая и вторая кодирующие последовательности могут

отличаться последовательностью нуклеиновой кислоты, но кодируют одни и те же аминокислоты. Трансген может находиться в векторе, где формат вектора выбран из двухцепочечной линейной ДНК, двухцепочечной кольцевой ДНК или вирусного вектора. Вирусный вектор может включать аденовирусный вектор, аденоассоциированный вирусный вектор или лентивирусный вектор. Представленные в настоящей заявке способы можно применять с трансгеном равным или менее 4,7 т.п.н. Трансген может содержать первую и вторую неполные кодирующие последовательности, которые кодируют неполный пептид из функционального белка, продуцируемого целевым эндогенным геном. Целевой эндогенный ген может быть абберрантным.

10 В другом варианте воплощения настоящего изобретения, представлены полинуклеотиды ДНК с первой и второй акцепторной последовательностью сплайсинга, первой и второй неполными кодирующими последовательностями, одним двунаправленным терминатором или первым и вторым терминатором, необязательно, первым и вторым плечом гомологии, и, необязательно, первым и вторым сайтом-мишенью

15 редкощепящей эндонуклеазы. Полинуклеотиды ДНК могут включать конструкцию, в которой первый акцептор сплайсинга функционально связан с первой неполной кодирующей последовательностью, а второй акцептор сплайсинга функционально связан со второй кодирующей последовательностью. Компоновка также может включать в себя наличие первой неполной кодирующей последовательности, функционально связанной с

20 первым терминатором, и второй неполной кодирующей последовательности, функционально связанной со вторым терминатором. В варианте воплощения изобретения, два терминатора могут быть заменены одним двунаправленным терминатором. В одном из вариантов воплощения изобретения, полинуклеотиды ДНК с первым и вторым акцепторами сплайсинга, первой и второй кодирующими последовательностями и первым

25 и вторым терминаторами могут быть ориентированы в направлении «хвост к хвосту». Полинуклеотиды ДНК с ориентацией последовательностей «хвост к хвосту» могут дополнительно содержать первый и второй сайт-мишень для одной или более редкощепящих эндонуклеаз, причем сайты-мишени фланкируют первый и второй акцепторы сплайсинга. В другом варианте воплощения изобретения, полинуклеотиды ДНК

30 могут содержать левое и правое плечо гомологии, которые фланкируют первый и второй акцепторы сплайсинга. В этом варианте воплощения изобретения, полинуклеотид ДНК может содержаться в аденоассоциированном вирусном векторе. В другом варианте воплощения изобретения, полинуклеотиды ДНК могут дополнительно содержать первый и второй сайты-мишени для одной или более редкощепящих эндонуклеаз, при этом сайты-

35 мишени фланкируют первый и второй акцепторы сплайсинга. Первый и второй сайты-

мишени могут фланкировать первое и второе плечо гомологии. В вариантах воплощения изобретения, полинуклеотиды ДНК, описанные в настоящей заявке, могут быть интегрированы в интрон эндогенного гена или в соединение интрон-экзон. Полинуклеотиды ДНК могут быть интегрированы в интрон или в соединение интрон-экзон гена ATXN3 или гена SACSNA1A. Полинуклеотид ДНК может содержать первую и вторую 5 неполные кодирующие последовательности, кодирующие пептид, продуцируемый экзоном 10 непатогенного гена ATXN3. Полинуклеотид ДНК может содержать первую и вторую неполные кодирующие последовательности, кодирующие пептид, продуцируемый экзоном 47 непатогенного гена SACSNA1A. Первая и вторая неполные кодирующие 10 последовательности кодируют одни и те же аминокислоты. В одном из вариантов воплощения изобретения, первая и вторая кодирующие последовательности могут отличаться последовательностью нуклеиновой кислоты, но кодируют одни и те же аминокислоты. Полинуклеотиды ДНК могут содержаться в векторе, где формат вектора выбран из двухцепочечной линейной ДНК, двухцепочечной кольцевой ДНК или вирусного 15 вектора. Вирусный вектор может быть выбран из аденовирусного вектора, аденоассоциированного вирусного вектора или лентивирусного вектора. Описанные в настоящей заявке полинуклеотиды ДНК могут иметь размер равный или менее 4,7 т.п.н.

В одном из вариантов воплощения настоящего изобретения, описан способ интеграции трансгена в эндогенный ген, способ включает введение трансгена, при этом 20 трансген содержит левый и правый конец транспозона, первую и вторую акцепторную последовательность сплайсинга, первую и вторую неполные кодирующие последовательности и один двунаправленный терминатор или первый и второй терминатор, и введение транспозазы, нацеленной на эндогенный ген, где трансген интегрирован в эндогенный ген. Способ может включать конструирование трансгена так, чтобы первый 25 акцептор сплайсинга был функционально связан с первой неполной кодирующей последовательностью, а второй акцептор сплайсинга был функционально связан со второй кодирующей последовательностью. Компоновка также может включать в себя наличие первой неполной кодирующей последовательности, функционально связанной с первым терминатором, и второй неполной кодирующей последовательности, функционально 30 связанной со вторым терминатором. В варианте воплощения изобретения, два терминатора могут быть заменены одним двунаправленным терминатором. В варианте воплощения изобретения, трансгены с первым и вторым акцепторами сплайсинга, первой и второй кодирующими последовательностями и первым и вторым терминаторами могут быть ориентированы в направлении «хвост к хвосту». Трансгены с ориентацией 35 последовательностей "хвост к хвосту" могут дополнительно содержать левый и правый

конец транспозона, фланкирующие первый и второй акцепторы сплайсинга. В вариантах воплощения изобретения, описанные в настоящей заявке трансгены могут быть интегрированы в интрон эндогенного гена или в соединение интрон-экзон. Трансгены могут быть интегрированы в интрон или в соединение интрон-экзон гена ATXN3 или гена

5 CACNA1A. Трансген может содержать первую и вторую неполные кодирующие последовательности, кодирующие пептид, продуцируемый экзоном 10 непатогенного гена ATXN3, и может быть нацелен на интрон 9 или соединение интрона 9 с экзоном 10 патогенного гена ATXN3. Трансген может содержать первую и вторую неполные кодирующие последовательности, кодирующие пептид, продуцируемый экзоном 47

10 непатогенного гена CACNA1A, и может быть нацелен на интрон 46 или соединение интрона 46 с экзоном 47 патогенного гена CACNA1A. Транспозаза может быть транспозазой CRISPR, где транспозаза CRISPR включает белок Cas12k или Cas6. Первая и вторая неполные кодирующие последовательности кодируют одни и те же аминокислоты. В одном из вариантов воплощения изобретения, первая и вторая кодирующие

15 последовательности могут отличаться последовательностью нуклеиновой кислоты, но кодируют одни и те же аминокислоты. Трансген может находиться в векторе, где формат вектора выбран из двухцепочечной линейной ДНК, двухцепочечной кольцевой ДНК или вирусного вектора. Вирусный вектор может включать аденовирусный вектор, аденоассоциированный вирусный вектор или лентивирусный вектор. Представленные в

20 настоящей заявке способы можно применять с трансгеном равным или менее 4,7 т.п.н. Левый конец может содержать последовательность, показанную в SEQ ID NO: 41, а правый конец может содержать последовательность, показанную в SEQ ID NO: 13.

В другом варианте воплощения настоящего изобретения, представлены полинуклеотиды ДНК с первой и второй акцепторными последовательностями сплайсинга,

25 первой и второй неполными кодирующими последовательностями, одним двунаправленным терминатором или первым и вторым терминатором и левым и правым концом транспозона. Полинуклеотиды ДНК могут включать конструкцию, в которой первый акцептор сплайсинга функционально связан с первой неполной кодирующей последовательностью, а второй акцептор сплайсинга функционально связан со второй

30 кодирующей последовательностью. Компоновка также может включать в себя наличие первой неполной кодирующей последовательности, функционально связанной с первым терминатором, и второй неполной кодирующей последовательности, функционально связанной со вторым терминатором. В варианте воплощения изобретения, два терминатора могут быть заменены одним двунаправленным терминатором. В варианте воплощения

35 изобретения, полинуклеотиды ДНК с первым и вторым акцепторами сплайсинга, первой и

второй кодирующими последовательностями и первым и вторым терминаторами могут быть ориентированы в направлении «хвост к хвосту». Полинуклеотиды ДНК с ориентацией последовательностей "хвост к хвосту" могут дополнительно содержать левый и правый конец транспозона, фланкирующие первый и второй акцепторы сплайсинга. В вариантах воплощения изобретения, полинуклеотиды ДНК, описанные в настоящей заявке, могут быть интегрированы в интрон эндогенного гена или в соединение интрон-экзон. Полинуклеотиды ДНК могут быть интегрированы в интрон или в соединение интрон-экзон гена ATXN3 или гена SACSNA1A. Полинуклеотид ДНК может содержать первую и вторую неполные кодирующие последовательности, кодирующие пептид, продуцируемый экзоном 10 непатогенного гена ATXN3. Полинуклеотид ДНК может содержать первую и вторую неполные кодирующие последовательности, кодирующие пептид, продуцируемый экзоном 47 непатогенного гена SACSNA1A. Первая и вторая неполные кодирующие последовательности кодируют одни и те же аминокислоты. В одном из вариантов воплощения изобретения, первая и вторая кодирующие последовательности могут отличаться последовательностью нуклеиновой кислоты, но кодируют одни и те же аминокислоты. Полинуклеотиды ДНК могут содержаться в векторе, где формат вектора выбран из двухцепочечной линейной ДНК, двухцепочечной кольцевой ДНК или вирусного вектора. Вирусный вектор может быть выбран из аденовирусного вектора, аденоассоциированного вирусного вектора или лентивирусного вектора. Описанные в настоящей заявке полинуклеотиды ДНК могут иметь размер равный или менее 4,7 т.п.н. Левый конец может содержать последовательность, показанную в SEQ ID NO: 41, а правый конец может содержать последовательность, показанную в SEQ ID NO: 13.

В одном из вариантов воплощения настоящего изобретения, представлен способ интеграции трансгена в эндогенный ген, способ, включающий введение трансгена, при этом трансген содержит первую и вторую акцепторные последовательности сплайсинга, первую и вторую кодирующие последовательности, один двунаправленный терминатор или первый и второй терминатор, и первого и второго плеча гомологии, при этом трансген интегрирован в эндогенный ген. Способ может включать конструирование трансгена так, чтобы первый акцептор сплайсинга был функционально связан с первой неполной кодирующей последовательностью, а второй акцептор сплайсинга был функционально связан со второй кодирующей последовательностью. Компоновка также может включать в себя наличие первой неполной кодирующей последовательности, функционально связанной с первым терминатором, и второй неполной кодирующей последовательности, функционально связанной со вторым терминатором. В варианте воплощения изобретения, два терминатора могут быть заменены одним двунаправленным терминатором. Плечи

гомологии могут фланкировать первую и вторую акцепторную последовательность сплайсинга, первую и вторую кодирующие последовательности, один двунаправленный терминатор или первый и второй терминатор. Кодирующая последовательность может кодировать полную кодирующую последовательность или неполную кодирующую последовательность. В варианте воплощения изобретения, трансгены с первым и вторым акцепторами сплайсинга, первой и второй кодирующими последовательностями и первым и вторым терминаторами могут быть ориентированы «хвост к хвосту». Трансгены с ориентацией последовательностей «хвост к хвосту» могут дополнительно содержать первый и второй сайты-мишени для одной или более редкощепящих эндонуклеаз, при этом сайты-мишени фланкируют первый и второй акцепторы сплайсинга. В другом варианте воплощения изобретения, трансгены могут содержать левое и правое плечо гомологии, которые фланкируют первый и второй акцепторы сплайсинга. В этом варианте воплощения изобретения, трансген может находиться в аденоассоциированном вирусном векторе. В другом варианте воплощения изобретения, трансген может дополнительно содержать первый и второй сайт-мишень для одной или более редкощепящих эндонуклеаз, при этом сайты-мишени фланкируют первый и второй акцепторы сплайсинга. Первый и второй сайты-мишени могут фланкировать первое и второе плечо гомологии. В вариантах воплощения настоящего изобретения, трансгены могут быть интегрированы в интрон эндогенного гена или в соединение интрон-экзон.

При практической реализации способов, а также получении и применении композиций, раскрытых в настоящей заявке, используются, если не указано иное, обычные методики в молекулярной биологии, биохимии, структуре и анализе хроматина, вычислительной химии, культуре клеток, рекомбинантной ДНК и связанных областях, которые находятся в пределах области техники. Эти методики полностью описаны в научной литературе. См., например, Sambrook et al. MOLECULAR CLONING: A LABORATORY MANUAL, Second edition, Cold Spring Harbor Laboratory Press, 1989 и Third edition, 2001; Ausubel et al., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, 1987 и периодические обновления; серию METHODS IN ENZYMOLOGY, Academic Press, San Diego; Wolffe, CHROMATIN STRUCTURE AND FUNCTION, Third edition, Academic Press, San Diego, 1998; METHODS IN ENZYMOLOGY, Vol. 304, "Chromatin" (P. M. Wassarman and A. P. Wolffe, eds.), Academic Press, San Diego, 1999; и METHODS IN MOLECULAR BIOLOGY, Vol. 119, "Chromatin Protocols" (P. B. Becker, ed.) Humana Press, Totowa, 1999.

В контексте данного документа, термины «нуклеиновая кислота» и «полинуклеотид» могут быть взаимозаменяемыми. Нуклеиновая кислота и полинуклеотид

могут относиться к дезоксирибонуклеотидному или рибонуклеотидному полимеру в линейной или кольцевой конформации и в одноцепочечной или двухцепочечной форме. Эти термины не следует рассматривать как ограничивающие длину полимера. Термины могут охватывать известные аналоги природных нуклеотидов, а также нуклеотиды, модифицированные по основанию, сахару и/или фосфатному фрагменту.

Термины «полипептид», «пептид» и «белок» могут использоваться взаимозаменяемо для обозначения аминокислотных остатков, ковалентно связанных вместе. Этот термин также применяется к белкам, в которых одна или более аминокислот являются химическими аналогами или модифицированными производными соответствующих встречающихся в природе аминокислот.

Термины «функционально связанный» или «оперативно связанный» используются взаимозаменяемо и относятся к сопоставлению двух или более компонентов (таких как элементы последовательности), в котором компоненты расположены так, что оба компонента функционируют нормально и допускают возможность того, что по меньшей мере один из компонентов может выполнять функцию, выполняемую по меньшей мере одним из других компонентов. В качестве примера, транскрипционная регуляторная последовательность, такая как промотор, функционально связана с кодирующей последовательностью, если транскрипционная регуляторная последовательность контролирует уровень транскрипции кодирующей последовательности в ответ на присутствие или отсутствие одного или более факторов регуляции транскрипции. Последовательность, регулирующая транскрипцию, обычно функционально связана в цис-кодирующей последовательности, но не обязательно непосредственно примыкает к ней. Например, энхансер является регуляторной последовательностью транскрипции, которая функционально связана с кодирующей последовательностью, даже если они не являются смежными. Кроме того, в качестве примера, акцептор сплайсинга может быть функционально связан с неполной кодирующей последовательностью, если акцептор сплайсинга позволяет очертить 3'-границу интрона и если трансляция полученной зрелой мРНК приводит к включению пептидной последовательности, кодируемой неполной кодирующей последовательностью с образованием конечного белкового продукта.

В контексте данного документа, термин «расщепление» относится к разрыву ковалентного остова молекулы нуклеиновой кислоты. Расщепление можно инициировать множеством способов, включая без ограничений ферментативный или химический гидролиз фосфодиэфирной связи. Расщепление может относиться как к одноцепочечному разрыву, так и к двухцепочечному разрыву. Двухцепочечный разрыв может произойти в результате двух отдельных одноцепочечных разрывов. Расщепление нуклеиновой кислоты

может привести к образованию либо тупых концов, либо смещенных концов. В некоторых вариантах воплощения изобретения, для целенаправленного расщепления двухцепочечной или одноцепочечной ДНК, применяются редкощепящие эндонуклеазы.

5 Термин «экзогенная» молекула может относиться к небольшой молекуле (например, сахара, липиды, аминокислоты, жирные кислоты, фенольные соединения, алкалоиды) или макромолекуле (например, белок, нуклеиновая кислота, углевод, липид, гликопротеин, липопротеин, полисахарид), или любому модифицированному производному вышеуказанных молекул, или любому комплексу, содержащему одну или более вышеуказанных молекул, генерируемых или присутствующих вне клетки, или обычно не присутствующих в клетке. Экзогенные молекулы можно вводить в клетки. Способы введения или «впрыскивания» экзогенных молекул в клетки могут включать липид-опосредованный перенос, электропорацию, прямую инъекцию, слияние клеток, бомбардировку частицами, совместное осаждение фосфатом кальция, перенос, опосредованный DEAE-декстраном и перенос, опосредованный вирусным вектором.

10 Указанный в данном документе термин «впрыскивание» может относиться к доставке, обеспечению или введению экзогенных молекул в клетку. Если трансген или редкощепящая эндонуклеаза вводится в клетку, то трансген или редкощепящая эндонуклеаза доставляется, предоставляется или вводится в клетку. Редкощепящая эндонуклеаза может вводиться в виде очищенного белка, нуклеиновой кислоты или смеси очищенного белка и нуклеиновой кислоты. Нуклеиновая кислота (т.е. РНК или ДНК) может кодировать редкощепящую эндонуклеазу или часть редкощепящей эндонуклеазы (например, гРНК). Введение может осуществляться с помощью таких способов, как липид-опосредованный перенос, электропорация, прямая инъекция, слияние клеток, бомбардировка частицами, совместное осаждение фосфатом кальция, перенос, опосредованный DEAE-декстраном, перенос, опосредованный вирусным вектором, или любыми подходящими средствами доставки очищенного белка или нуклеиновых кислот, или смеси очищенного белка и нуклеиновых кислот.

20

25

«Эндогенная» молекула является молекулой, которая присутствует в конкретной клетке на определенной стадии развития в определенных условиях окружающей среды.

30 Эндогенная молекула может быть нуклеиновой кислотой, хромосомой, геномом митохондрии, хлоропласта или другой органеллы или встречающейся в природе эписомальной нуклеиновой кислотой. Дополнительные эндогенные молекулы могут включать белки, например, факторы транскрипции и ферменты.

В контексте данного документа, термин «ген» относится к области ДНК, кодирующей продукт гена, включая все области ДНК, которые регулируют

35

5 продуцирование продукта гена. Соответственно, ген включает без ограничений промоторные последовательности, терминаторы, регуляторные последовательности трансляции, такие как сайты связывания рибосом и внутренние сайты входа в рибосомы, энхансеры, сайленсеры, инсуляторы, граничные элементы, источники репликации, сайты

10 прикрепления матрикса и регионы контроля локуса. В контексте данного документа, термин «ген дикого типа» относится к форме гена, которая присутствует с наибольшей частотой в конкретной популяции.

10 Термин «эндогенный ген» относится к области ДНК, обычно присутствующей в конкретной клетке, которая кодирует продукт гена, а также ко всем областям ДНК, которые регулируют выработку продукта гена.

15 Термин «экспрессия гена» относится к преобразованию информации, содержащейся в гене, в продукт гена. Продукт гена может быть продуктом прямой транскрипции гена. Например, генный продукт может представлять собой без ограничения мРНК, тРНК, рРНК, антисмысловую РНК, рибозим, структурную РНК или белок, полученный путем

20 транскрипции мРНК. Генные продукты также включают РНК, модифицированные такими процессами, как кэппинг, полиаденилирование, метилирование и редактирование, и белки, модифицированные, например, метилированием, ацетилированием, фосфорилированием, убиквитинированием, АДФ-рибозилированием, миристилированием и гликозилированием.

25 Термин «кодирование» относится к преобразованию информации, содержащейся в нуклеиновой кислоте, в продукт, при этом такой продукт может быть результатом продукта прямой транскрипции последовательности нуклеиновой кислоты. Например, продукт может представлять собой без ограничения мРНК, тРНК, рРНК, антисмысловую РНК, рибозим, структурную РНК или белок, полученный путем транскрипции мРНК. Генные продукты также включают РНК, модифицированные такими процессами, как кэппинг,

30 полиаденилирование, метилирование и редактирование, и белки, модифицированные, например, метилированием, ацетилированием, фосфорилированием, убиквитинированием, АДФ-рибозилированием, миристилированием и гликозилированием.

35 Термин «сайт-мишень» или «последовательность-мишень» относится к части нуклеиновой кислоты, с которой будет связываться редкощепящая эндонуклеаза или транспозаза, ассоциированная с CRISPR, при условии, что существуют достаточные условия для связывания.

В контексте данного документа, термин «рекомбинация» относится к процессу обмена генетической информацией между двумя полинуклеотидами. Термин «гомологичная рекомбинация (HR)» относится к специальной форме рекомбинации, которая может иметь место, например, во время репарации двухцепочечных разрывов.

Гомологичная рекомбинация требует гомологии нуклеотидной последовательности, присутствующей на «донорной» молекуле. Молекула-донор может использоваться клеткой в качестве матрицы для восстановления двухцепочечного разрыва. Информация внутри донорной молекулы, которая отличается от геномной последовательности на двухцепочечном разрыве или рядом с ним, может быть стабильно включена в геномную ДНК клетки.

В контексте данного документа, термин «интеграция» относится к процессу добавления ДНК к целевой области ДНК. В соответствии с настоящей заявкой, интеграции можно способствовать несколькими различными способами, включая негомологичное соединение концов, гомологичную рекомбинацию или направленную транспозицию. Например, интеграцию поставляемой пользователем молекулы ДНК в целевой ген можно облегчить за счет негомологичного соединения концов. В соответствии с настоящей заявкой, в целевом гене делается направленный двухцепочечный разрыв и вводится предъявляемая пользователем молекула ДНК. Предъявляемая пользователем молекула ДНК может содержать открытые концы ДНК для облегчения захвата во время репарации целевого гена за счет негомологичного соединения концов. Открытые концы могут присутствовать на молекуле ДНК при введении (т. е. введении линейной молекулы ДНК) или образовываться при введении в клетку (т. е. редкощеплящая эндонуклеаза расщепляет предоставленную пользователем молекулу ДНК внутри клетки, чтобы обнажить концы). Кроме того, предъявляемая пользователем молекула ДНК может содержаться в вирусном векторе, включая аденоассоциированный вирусный вектор. В другом примере интеграция происходит посредством гомологичной рекомбинации. В соответствии с настоящей заявкой, предъявляемая пользователем ДНК может содержать левое и правое плечо гомологии. В другом примере интеграция происходит через транспозицию. В соответствии с настоящей заявкой, предъявляемая пользователем ДНК содержит левый и правый конец транспозона.

В контексте данного документа, термин «трансген» относится к последовательности нуклеиновых кислот, которая может быть перенесена в организм или клетку. Трансген может содержать ген или последовательность нуклеиновых кислот, обычно не присутствующих в организме или клетке-мишени. Кроме того, трансген может содержать копию гена или последовательности нуклеиновых кислот, которые обычно присутствуют в организме или клетке-мишени. Трансген может быть экзогенной последовательностью ДНК, введенной в цитоплазму или ядро клетки-мишени. В одном из вариантов воплощения настоящего изобретения, трансгены содержат неполные кодирующие последовательности, при этом неполные кодирующие последовательности кодируют часть белка,

продуцируемого геном в клетке-хозяине.

В контексте данного документа, термин «патогенный» относится ко всему, что может вызвать заболевание. Патогенная мутация может относиться к модификации гена, вызывающей заболевание. Патогенный ген относится к гену, содержащему модификацию, вызывающую заболевание. Например, патогенный ген ATXN3 у пациентов со спиноцеребеллярной атаксией 3 относится к гену ATXN3 с расширенным тринуклеотидным повтором CAG, при этом расширенный тринуклеотидный повтор CAG вызывает заболевание.

В контексте данного документа, термин «хвост к хвосту» относится к ориентации двух единиц в противоположном и обратном направлениях. Две единицы могут быть двумя последовательностями в одной молекуле нуклеиновой кислоты, где 3'-конец каждой последовательности расположен рядом друг с другом. Например, первая нуклеиновая кислота, имеющая элементы в направлении от 5' к 3', [акцептор сплайсинга 1] - [неполная кодирующая последовательность 1] - [терминатор 1], и вторая нуклеиновая кислота, имеющая элементы [акцептор сплайсинга 2] - [неполная кодирующая последовательность 2] - [терминатор 2] может быть размещен в ориентации «хвост к хвосту», что приводит к [акцептор сплайсинга 1] - [неполная кодирующая последовательность 1] - [терминатор 1] - [терминатор 2 RC] - [неполная кодирующая последовательность 2 RC] - [акцептор сплайсинга 2 RC], где RC означает обратный комплемент.

Термин «соединение интрон-экзон» относится к определенному участку внутри гена. Конкретное местоположение находится между последним нуклеотидом в интроне и первым нуклеотидом следующего экзона. При интеграции трансгена, описанного в настоящей заявке, трансген может быть интегрирован в «соединение интрон-экзон». Если трансген включает карго, карго будет интегрирован сразу после последнего нуклеотида в интроне. В некоторых случаях интеграция трансгена в соединение интрон-экзон может привести к удалению последовательности внутри экзона (например, интеграция посредством HR и замена последовательности внутри экзона карго в трансгене).

В контексте данного документа, термин «гомологичный» относится к последовательности нуклеиновых кислот или аминокислот, имеющей сходство со второй последовательностью нуклеиновых кислот или аминокислот. В некоторых вариантах воплощения изобретения, гомологичные последовательности могут иметь по меньшей мере 80% идентичности последовательностей (например, 81, 85, 90, 95, 96, 97, 98 или 99% идентичности последовательностей) друг с другом.

В контексте данного документа, термин «неполная кодирующая последовательность» относится к последовательности нуклеиновых кислот, которая

кодирует неполный белок. Неполная кодирующая последовательность может кодировать белок, который содержит одну или менее аминокислот по сравнению с белком дикого типа или функциональным белком. Неполная кодирующая последовательность может кодировать неполный белок, гомологичный белку дикого типа или функциональному белку. Термин «неполная кодирующая последовательность» применительно к ATXN3 относится к последовательности нуклеиновых кислот, которая кодирует неполный белок ATXN3. Неполный белок ATXN3 имеет одну или менее аминокислот по сравнению с белком ATXN3 дикого типа. При модификации 3'-конца гена одна или менее аминокислот могут происходить от N-конца белка. Если ген ATXN3 имеет 11 экзонов, то неполная кодирующая последовательность может включать последовательность, кодирующую пептид, продуцируемый экзонами 2-11 или 3-11, или 4-11, или 5-11, или 6-11, или 7-11, или 8-11, или 9-11, или 10-11, или 11.

В способах и композициях, представленных в настоящей заявке, могут применяться трансгены, имеющие последовательность карго. Термин «карго» может относиться к таким элементам, как полная или неполная кодирующая последовательность гена, неполная последовательность гена, несущего однонуклеотидный полиморфизм относительно WT (дикого типа) или измененной мишени, акцептор сплайсинга, терминатор, регулятор транскрипции, метки очистки (например, глутатион-S-трансфераза, поли (His), мальтозосвязывающий белок, Strep-tag, Myc-tag, AviTag, HA-tag или хитинсвязывающий белок) или репортерные гены (например, GFP, RFP, lacZ, cat, люцифераза, пуρο, неомицин). Указанный в данном документе термин «карго» может относиться к последовательности трансгена, интегрированной в целевой сайт. Например, «карго» может относиться к последовательности трансгена между двумя плечами гомологии, двумя сайтами-мишенями редкощепящей эндонуклеазы или левым и правым концом транспозона.

Термин «гомологическая последовательность» относится к последовательности нуклеиновых кислот, которая гомологична второй нуклеиновой кислоте. Гомологичная последовательность, например, может присутствовать на молекуле-доноре как «плечо гомологии» или «гомологичное плечо». Гомологичное плечо может быть последовательностью нуклеиновых кислот внутри донорной молекулы, которая способствует гомологичной рекомбинации со второй нуклеиновой кислотой. Указанное в данном документе гомологичное плечо также может называться «плечом». В молекуле-доноре с двумя гомологичными плечами плечи гомологии можно обозначать как «плечо 1» и «плечо 2». В одном аспекте последовательность карго может быть фланкирована первым и вторым гомологичными плечами.

Термин «двунаправленный терминатор» относится к терминатору, который может

прекращать транскрипцию полимеразы РНК либо в смысловом, либо в антисмысловом направлении. В отличие от двух однонаправленных терминаторов в ориентации хвост к хвосту, двунаправленный терминатор может содержать нехимическую последовательность ДНК. Примеры двунаправленных терминаторов включают терминатор ARO4, TRP1, TRP4, ADH1, CYC1, GAL1, GAL7 и GAL10.

5' или 3'-конец молекулы нуклеиновой кислоты указывает направление и химическую ориентацию нуклеиновой кислоты. Указанный в данном документе «5'-конец гена» может включать экзон со стартовым кодоном, но не экзон со стоп-кодоном. Указанный в данном документе «3'-конец гена» может включать экзон со стоп-кодоном, но не экзон со стартовым кодоном.

Термин ген «ATXN3» относится к гену, который кодирует фермент атаксин-3. Репрезентативная последовательность гена ATXN3 может быть обнаружена с помощью эталонной последовательности NCBI: NG 008198.2 и соответствующей SEQ ID NO: 42. Границы экзона и интрона можно определить с помощью последовательности, представленной в SEQ ID NO: 42. В частности, экзон 1 включает последовательность от 1 до 54. Экзон 2 включает последовательность от 9745 до 9909. Экзон 3 включает последовательность от 10446 до 10490. Экзон 4 включает последовательность от 12752 до 12837. Экзон 5 включает последовательность от 13265 до 13331. Экзон 6 включает последовательность от 17766 до 17853. Экзон 7 включает последовательность от 23325 до 23457. Экзон 8 включает последовательность от 24117 до 24283. Экзон 9 включает последовательность от 25522 до 25618. Экзон 10 включает последовательность от 35530 до 35648. Экзон 11 включает последовательность от 42169 до 48031. Интрон 1 включает последовательность от 55 до 9744. Интрон 2 включает последовательность от 9910 до 10445. Интрон 3 включает последовательность от 10491 до 12751. Интрон 4 включает последовательность от 12838 до 13264. Интрон 5 включает последовательность от 13332 до 17765. Интрон 6 включает последовательность от 17854 до 23324. Интрон 7 включает последовательность от 23458 до 24116. Интрон 8 включает последовательность от 24284 до 25521. Интрон 9 включает последовательность от 25619 до 35529. Интрон 10 включает последовательность от 35649 до 42168.

Термин ген «CACNA1 A» относится к гену, который кодирует белок альфа-A субъединицы кальциевого потенциалзависимого канала. Репрезентативная последовательность гена CACNA1A может быть обнаружена с помощью эталонной последовательности NCBI: NG 011569.1 и соответствующей SEQ ID NO: 43. Границы экзона и интрона можно определить с помощью последовательности, представленной в SEQ ID NO: 43. В частности, экзон 1 включает последовательность от 1 до 529. Экзон 2

включает последовательность от 51249 до 51354. Экзон 3 включает последовательность от 53446 до 53585. Экзон 4 включает последовательность от 134682 до 134773. Экзон 5 включает последовательность от 140992 до 141144. Экзон 6 включает последовательность от 146662 до 146855. Экзон 7 включает последовательность от 170552 до 170655. Экзон 8 включает последовательность от 171968 до 172083. Экзон 9 включает последовательность от 173536 до 173592. Экзон 10 включает последовательность от 176125 до 176217. Экзон 11 включает последовательность от 189140 до 189349. Экзон 12 включает последовательность от 193680 до 193792. Экзон 13 включает последовательность от 197933 до 198045. Экзон 14 включает последовательность от 198210 до 198341. Экзон 15 включает последовательность от 198607 до 198679. Экзон 16 включает последовательность от 202577 до 202694. Экзон 17 включает последовательность от 202848 до 202915. Экзон 18 включает последовательность от 205805 до 205911. Экзон 19 включает последовательность от 207108 до 207917. Экзон 20 включает последовательность от 219495 до 219958. Экзон 21 включает последовательность от 221255 до 221393. Экзон 22 включает последовательность от 223065 до 223194. Экзон 23 включает последовательность от 229333 до 229392. Экзон 24 включает последовательность от 230505 до 230611. Экзон 25 включает последовательность от 243628 до 243727. Экзон 26 включает последовательность от 244851 до 245011. Экзон 27 включает последовательность от 246760 до 246897. Экзон 28 включает последовательность от 248910 до 249111. Экзон 29 включает последовательность от 251202 до 251366. Экзон 30 включает последовательность от 253360 до 253470. Экзон 31 включает последовательность от 261196 до 261279. Экзон 32 включает последовательность от 270731 до 270847. Экзон 33 включает последовательность от 271187 до 271252. Экзон 34 включает последовательность от 271425 до 271540. Экзон 35 включает последовательность от 274601 до 274751. Экзон 36 включает последовательность от 276252 до 276379. Экзон 37 включает последовательность от 277666 до 277762. Экзон 38 включает последовательность от 281689 до 281794. Экзон 39 включает последовательность от 291853 до 291960. Экзон 40 включает последовательность от 292128 до 292228. Экзон 41 включает последовательность от 293721 до 293830. Экзон 42 включает последовательность от 293939 до 294077. Экзон 43 включает последовательность от 294245 до 294358. Экзон 44 включает последовательность от 295809 до 295844. Экзон 45 включает последовательность от 296963 до 297149. Экзон 46 включает последовательность от 297452 до 297705. Экзон 47 включает последовательность от 298413 до 300019. Интрон 1 включает последовательность от 530 до 51248. Интрон 2 включает последовательность от 51355 до 53445. Интрон 3 включает последовательность от 53586 до 134681. Интрон 4 включает последовательность от 134774 до 140991. Интрон 5 включает последовательность от 141145 до 146661. Интрон 6 включает последовательность от 146856 до 170551. Интрон 7 включает

последовательность от 170656 до 171967. Интрон 8 включает последовательность от 172084 до 173535. Интрон 9 включает последовательность от 173593 до 176124. Интрон 10 включает последовательность от 176218 до 189139. Интрон 11 включает последовательность от 189350 до 193679. Интрон 12 включает последовательность от 193793 до 197932. Интрон 13 включает последовательность от 198046 до 198209. Интрон 14 включает последовательность от 198342 до 198606. Интрон 15 включает последовательность от 198680 до 202576. Интрон 16 включает последовательность от 202695 до 202847. Интрон 17 включает последовательность от 202916 до 205804. Интрон 18 включает последовательность от 205912 до 207107. Интрон 19 включает последовательность от 207918 до 219494. Интрон 20 включает последовательность от 219959 до 221254. Интрон 21 включает последовательность от 221394 до 223064. Интрон 22 включает последовательность от 223195 до 229332. Интрон 23 включает последовательность от 229393 до 230504. Интрон 24 включает последовательность от 230612 до 243627. Интрон 25 включает последовательность от 243728 до 244850. Интрон 26 включает последовательность от 245012 до 246759. Интрон 27 включает последовательность от 246898 до 248909. Интрон 28 включает последовательность от 249112 до 251201. Интрон 29 включает последовательность от 251367 до 253359. Интрон 30 включает последовательность от 253471 до 261195. Интрон 31 включает последовательность от 261280 до 270730. Интрон 32 включает последовательность от 270848 до 271186. Интрон 33 включает последовательность от 271253 до 271424. Интрон 34 включает последовательность от 271541 до 274600. Интрон 35 включает последовательность от 274752 до 276251. Интрон 36 включает последовательность от 276380 до 277665. Интрон 37 включает последовательность от 277763 до 281688. Интрон 38 включает последовательность от 281795 до 291852. Интрон 39 включает последовательность от 291961 до 292127. Интрон 40 включает последовательность от 292229 до 293720. Интрон 41 включает последовательность от 293831 до 293938. Интрон 42 включает последовательность от 294078 до 294244. Интрон 43 включает последовательность от 294359 до 295808. Интрон 44 включает последовательность от 295845 до 296962. Интрон 45 включает последовательность от 297150 до 297451. Интрон 46 включает последовательность от 297706 до 298412.

Процент идентичности конкретной последовательности нуклеиновой кислоты или аминокислоты и последовательности, на которую ссылается конкретный идентификационный номер последовательности, определяют следующим образом. Сначала последовательность нуклеиновой кислоты или аминокислоты сравнивается с последовательностью, указанной в конкретном идентификационном номере

последовательности, с помощью программы BLAST 2 Sequences (B12seq) из автономной версии BLASTZ, содержащей BLASTN версии 2.0.14 и BLASTP версии 2.0.14. Эту автономную версию BLASTZ можно скачать в Интернете по ссылке: fig.com/blast или ncbi.nlm.nih.gov. Инструкции, объясняющие как использовать программу B12seq, можно найти в файле readme, прилагаемом к BLASTZ. B12seq выполняет сравнение двух последовательностей с помощью алгоритма BLASTN или BLASTP. BLASTN применяется для сравнения последовательностей нуклеиновых кислот, а BLASTP – для сравнения аминокислотных последовательностей. Для сравнения двух последовательностей нуклеиновых кислот параметры устанавливаются следующим образом: -i устанавливается в файл, содержащий первую сравниваемую последовательность нуклеиновой кислоты (например, C:\seq1.txt); -j устанавливается в файл, содержащий вторую сравниваемую последовательность нуклеиновой кислоты (например, C:\seq2.txt); -p устанавливается в blastn; -o устанавливается в любое желаемое имя файла (например, C:\output.txt); -q устанавливается в -1; -r устанавливается в 2; а для всех остальных параметров установлены значения по умолчанию. Например, следующую команду можно использовать для создания выходного файла, содержащего сравнение двух последовательностей нуклеиновой кислоты: C:\B12seq -ic seq1.txt -jc:\seq2.txt -p blastn -oc:\output.txt -q -1 -r 2. Для сравнения двух аминокислотных последовательностей параметры B12seq устанавливаются следующим образом: -i устанавливается в файл, содержащий первую сравниваемую аминокислотную последовательность (например, C:\seq1.txt); -j устанавливается в файл, содержащий вторую сравниваемую аминокислотную последовательность (например, C:\seq2.txt); -p устанавливается в blastp; -o устанавливается в любое желаемое имя файла (например, C:\output.txt); а для всех остальных параметров установлены значения по умолчанию. Например, следующую команду можно использовать для создания выходного файла, содержащего сравнение двух аминокислотных последовательностей: C:\B12seq -ic:\seq1.txt -jc:\seq2.txt -p blastp -oc:\output.txt. Если две сравниваемые последовательности имеют общую гомологию, то указанный выходной файл представит эти области гомологии как выровненные последовательности. Если две сравниваемые последовательности не имеют одинаковой гомологии, то указанный выходной файл не будет представлять выровненные последовательности.

После выравнивания, количество совпадений определяется путем подсчета количества положений, в которых идентичный нуклеотидный или аминокислотный остаток присутствует в обеих последовательностях. Процент идентичности последовательностей определяется путем деления количества совпадений либо на длину последовательности, указанной в идентифицированной последовательности, либо на сочлененную длину

(например, 100 последовательных нуклеотидов или аминокислотных остатков из последовательности, указанной в идентифицированной последовательности) с последующим умножением полученного значения на 100. Значение идентичности последовательности в процентах округляется до ближайшей десятой.

5 В одном из вариантов воплощения настоящего изобретения, представлены способы модификации 3'-конца эндогенных генов, где эндогенные гены имеют по меньшей мере один интрон между двумя кодирующими экзонами. Интрон может быть любым интроном, который удаляется из предшественника матричной РНК с помощью обычного механизма обработки информационной РНК. Интрон может иметь размер от 20 п.н. до более 500 т.п.н.
10 и содержать элементы, включая донорный сайт сплайсинга, последовательность ответвления и акцепторный сайт. Раскрытые в настоящей заявке трансгены для модификации 3'-конца эндогенных генов могут содержать множество функциональных элементов, включая сайты-мишени для редкощепящих эндонуклеаз, гомологичные плечи, акцепторные последовательности сплайсинга, кодирующие последовательности и терминаторы транскрипции (ФИГ. 1).
15

В одном из вариантов воплощения изобретения, трансген содержит два сайта-мишени для одной или более редкощепящих эндонуклеаз. Сайты-мишени могут иметь подходящую последовательность и длину для расщепления редкощепящей эндонуклеазой. Целевой сайт может быть расщеплен системами CRISPR, эффекторными нуклеазами TAL,
20 нуклеазами типа «цинковые пальцы» или мегануклеазами, или комбинацией систем CRISPR, нуклеаз TALE, нуклеаз или мегануклеаз «цинковые пальцы», или любой другой сайт-специфической нуклеазой. Сайты-мишени могут быть расположены так, что расщепление редкощепящей эндонуклеазой приводит к высвобождению трансгена из вектора. Вектор может включать вирусные векторы (например, аденоассоциированные
25 векторы) или невирусные векторы (например, плазмиды, миникольцевые векторы). Если трансген содержит два сайта-мишени, сайты-мишени могут быть одной и той же последовательности (т.е. нацелены одной и той же редкощепящей эндонуклеазой) или они могут быть разными последовательностями (т.е. нацелены двумя или более разными редкощепящими эндонуклеазами).

30 В одном из вариантов воплощения изобретения, трансген содержит первый и второй сайты-мишени для одной или более редкощепящих эндонуклеаз, а также первое и второе плечо гомологии. Первое и второе плечо гомологии могут включать последовательность, которая гомологична геномной последовательности в желаемом сайте интеграции или рядом с ним. Плечи гомологии могут иметь длину, подходящую для участия в
35 гомологичной рекомбинации с последовательностью в желаемом сайте интеграции или

рядом с ним. Длина каждого плеча гомологии может составлять от 20 нт (нуклеотидов) до 10000 нт (например, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000 нт). В одном из вариантов воплощения изобретения, плечи гомологии могут содержать функциональные элементы, включая сайт-мишень для редкощепящей эндонуклеазы и/или акцепторную последовательность сплайсинга. В одном из вариантов воплощения изобретения, первое плечо гомологии (например, левое плечо гомологии) может содержать последовательность, гомологичную интрон-мишени, которая включает акцепторный сайт сплайсинга целевого интрона. В другом варианте воплощения изобретения, второе плечо гомологии может содержать последовательность, гомологичную геномной последовательности ниже целевого интрона, (например, последовательность экзона, последовательность 3' UTR). Однако второе плечо гомологии не должно обладать акцепторными функциями сплайсинга в обратном направлении комплемента. Чтобы определить, включает ли последовательность акцепторные функции сплайсинга, можно предпринять несколько шагов, включая анализ *in silico* и экспериментальные исследования. Чтобы определить, есть ли потенциал для акцепторных функций сплайсинга, последовательность, необходимая для второго плеча гомологии, может быть найдена среди последовательностей консенсусных плеч (например, YTRAC) и акцепторных сайтов сплайсинга (например, Y-богатого NCAGG). Если присутствуют акцепторные последовательности ответвлений или сплайсинга, можно ввести однонуклеотидные полиморфизмы для разрушения функции или можно выбрать другую, но смежную последовательность, не содержащую такие последовательности. Предпочтительно, окно последовательности, которое можно использовать для второго плеча гомологии, простирается от 1 п.н. до 10 т.п.н. ниже интрона, на который нацелена интеграция. Чтобы экспериментально определить, обладает ли вторая гомология акцепторной функцией сплайсинга, можно сконструировать синтетическую конструкцию, содержащую второе плечо гомологии внутри интрона в репортерном гене. Затем конструкцию можно вводить в клетки соответствующего типа и контролировать функцию сплайсинга.

В одном из вариантов воплощения изобретения, трансген содержит две акцепторные последовательности сплайсинга, называемые в настоящей заявке первой и второй акцепторными последовательностями сплайсинга. Первая и вторая акцепторные последовательности сплайсинга расположены внутри трансгена в противоположных направлениях (т.е. в ориентациях «хвост к хвосту») и фланкируют внутренние последовательности (т.е. кодирующие последовательности и терминаторы). Если трансген интегрируется в интрон в прямом или обратном направлении, акцепторные

последовательности сплайсинга облегчают удаление соседней/вышележащей последовательности интрона во время процессинга мРНК. Первая и вторая акцепторные последовательности сплайсинга могут быть одинаковыми или разными последовательностями. Одна или обе акцепторные последовательности сплайсинга могут
5 быть акцепторной последовательностью сплайсинга интрона, в который должен быть интегрирован трансген. Одна или обе акцепторные последовательности сплайсинга могут быть синтетической акцепторной последовательностью сплайсинга или акцепторной последовательностью сплайсинга от интрона из другого гена.

В одном из вариантов воплощения изобретения, трансген содержит первую и вторую
10 кодирующие последовательности, функционально связанные с первой и второй акцепторными последовательностями сплайсинга. Первая и вторая кодирующие последовательности расположены внутри трансгена в противоположных направлениях (т.е. в ориентации «хвост к хвосту»). Если трансген интегрируется в эндогенный ген в прямом или обратном направлении, первая или вторая кодирующая последовательность
15 транскрибируется в мРНК промотором эндогенного гена. Для исправления дефектных кодирующих последовательностей, введения мутаций или введения новых пептидных последовательностей могут быть разработаны кодирующие последовательности. Первая и вторая кодирующие последовательности могут быть одной и той же последовательностью нуклеиновой кислоты и кодировать один и тот же белок. Альтернативно, первая и вторая
20 кодирующие последовательности могут быть разными последовательностями нуклеиновой кислоты и кодировать один и тот же белок (т.е. с применением вырожденности кодонов). Кодирующая последовательность может кодировать метки очистки (например, глутатион-S-трансферазу, поли (His), мальтозо-связывающий белок, Strep-tag, Myc-tag, AviTag, HA-tag или хитин-связывающий белок) или репортерные белки (например, GFP, RFP, lacZ, cat,
25 люцифераза, пуго, неомицин). В одном из вариантов воплощения изобретения, трансген содержит первую и вторую неполные кодирующие последовательности, функционально связанные с первой и второй акцепторными последовательностями сплайсинга, и трансген не содержит промотора.

В одном из вариантов воплощения изобретения, трансген может содержать
30 двунаправленный терминатор или первый и второй терминаторы, функционально связанные с первой и второй кодирующей последовательностью. Двунаправленный терминатор или первый и второй терминаторы расположены внутри трансгена в противоположных направлениях (то есть в ориентациях «хвост к хвосту»). Если трансген интегрируется в эндогенный ген в прямом или обратном направлении, двунаправленный
35 терминатор или первый и второй терминаторы прекращают транскрипцию с промотора

эндогенного гена. Первый и второй терминаторы могут быть одинаковыми или разными терминаторами.

В одном из вариантов воплощения настоящего изобретения, представлен трансген, содержащий первый и второй сайты-мишени редкощепящей эндонуклеазы, первую и вторую акцепторные последовательности сплайсинга, первую и вторую кодирующие последовательности и один двунаправленный терминатор или первый и второй терминаторы. Трансген может быть интегрирован в эндогенные гены с помощью способов, независимых от гомологии, включая негомологичное соединение концов и альтернативное негомологичное соединение концов, или соединение концов, опосредованное микрогомологией. В одном аспекте трансген интегрирован в интрон внутри эндогенного гена (ФИГ. 2).

В другом варианте воплощения настоящего изобретения, представлен трансген, содержащий первое и второе плечи гомологии, первый и второй сайты-мишени редкощепящей эндонуклеазы, первую и вторую акцепторные последовательности сплайсинга, первую и вторую кодирующие последовательности и один двунаправленный терминатор или первый и второй терминаторы. Трансген может быть интегрирован в эндогенные гены как с помощью способов, зависимых от гомологии (например, зависимо от синтеза отжига цепей и опосредованного микрогомологией соединения концов), так и способов, независимых от гомологии (например, негомологичного соединения концов и альтернативного негомологичного соединения концов). В одном аспекте трансген интегрирован в интрон внутри эндогенного гена (ФИГ. 3). В другом аспекте трансген интегрирован в конец интрона или в начало расположенного ниже экзона (ФИГ. 3).

В другом варианте воплощения настоящего изобретения, представлен трансген, содержащий первое и второе плечи гомологии, первую и вторую кодирующие последовательности, первую и вторую акцепторные последовательности сплайсинга и один двунаправленный терминатор или первый и второй терминаторы (ФИГ. 1). В другом варианте воплощения настоящего изобретения, представлен трансген, содержащий первую и вторую кодирующие последовательности, первую и вторую акцепторные последовательности сплайсинга и один двунаправленный терминатор или первый и второй терминаторы.

В другом варианте воплощения настоящего изобретения, представлен трансген, содержащий первое и второе плечи гомологии, первую и вторую кодирующие последовательности, первую и вторую акцепторные последовательности сплайсинга, один двунаправленный терминатор или первый и второй терминаторы, а также первую и вторую дополнительные последовательности (ФИГ. 1). В некоторых вариантах воплощения

изобретения, дополнительная последовательность может представлять собой любую дополнительную последовательность, которая присутствует в трансгене на 5'- и 3'-концах, однако дополнительная последовательность не должна содержать какой-либо элемент, который функционирует как акцептор сплайсинга. Дополнительная последовательность может быть, например, инвертированными концевыми повторами вирусного генома. Дополнительная последовательность может присутствовать в трансгене, имеющем линейный формат. Линейный формат позволяет интегрировать NHEJ. Например, трансген, содержащийся в векторе аденоассоциированного вируса, где дополнительная последовательность представляет собой инвертированные концевые повторы, может быть непосредственно интегрирован NHEJ в целевой сайт после расщепления редкощепящей эндонуклеазой (т. е. нет необходимости в процессинге трансгена). В другом примере дополнительная последовательность представляет собой левый и правый концы транспозона.

В другом варианте воплощения настоящего изобретения, представлены трансгены в вирусных векторах, включая аденоассоциированные вирусы и аденовирусы, при этом трансген содержит первую и вторую акцепторные последовательности сплайсинга, первую и вторую кодирующие последовательности и один двунаправленный терминатор или первый и второй терминаторы. Из-за инвертированных концевых повторов вирусных векторов трансгены также содержат первую и вторую дополнительные последовательности.

В другом варианте воплощения настоящего изобретения, представлены трансгены в вирусных векторах, включая аденоассоциированные вирусы и аденовирусы, при этом трансген содержит первое и второе плечо гомологии, первую и вторую акцепторные последовательности сплайсинга, первую и вторую кодирующие последовательности и один двунаправленный терминатор или первый и второй терминаторы. Из-за инвертированных концевых повторов вирусных векторов трансгены также содержат первую и вторую дополнительные последовательности.

В некоторых вариантах воплощения изобретения, трансгены, представленные в настоящей заявке, могут быть интегрированы с транспозазами. Транспозазы могут включать транспозазы CRISPR (Strecker et al., *Science* 10.1126/science.aax9181, 2019; Klompe et al., *Nature*, 10.1038/s41586-019-1323-z, 2019). Транспозазы можно применять в комбинации с трансгеном, содержащим первую и вторую акцепторные последовательности сплайсинга, первую и вторую кодирующие последовательности, один двунаправленный терминатор или первый и второй терминаторы (ФИГ.1), а также левый и правый концы транспозона. Транспозазы CRISPR могут включать белок CRISPR TypeV-U5, C2C5, Cas12k,

а также белки *tnsB*, *tnsC* и *tniQ*. В некоторых вариантах воплощения изобретения, Cas12k может происходить из *Scytonema hofmanni* (SEQ ID NO: 30) или *Anabaena cylindrica* (SEQ ID NO: 31). В одном из вариантов воплощения настоящего изобретения, трансгены, содержащие левый (SEQ ID NO: 32) и правый концы транспозона (SEQ ID NO: 33), могут быть доставлены в клетки вместе с ShCas12k, *tnsB*, *tnsC*, *TniQ* и гРНК (SEQ ID NO: 14). Альтернативно, транспозаза CRISPR может включать белок Cas6 вместе со вспомогательными белками, включая Cas7, Cas8 и *TniQ*. В одном из вариантов воплощения настоящего изобретения, трансгены, содержащие левый (SEQ ID NO: 41) и правый концы транспозона (SEQ ID NO: 13), могут быть доставлены в эукариотические клетки вместе с Cas6 (SEQ ID NO: 37), Cas7 (SEQ ID NO: 37), Cas8 (SEQ ID NO: 37), *TniQ* (SEQ ID NO: 37), *TnsA* (SEQ ID NO: 37), *TnsB* (SEQ ID NO: 37), *TnsC* (SEQ ID NO: 37) и гРНК (SEQ ID NO: 12). Белки можно вводить в клетки непосредственно в виде очищенного белка или кодировать на РНК или ДНК. Если последовательность кодируется на РНК или ДНК, последовательность может быть оптимизирована по кодонам для экспрессии в эукариотических клетках. гРНК (SEQ ID NO: 12) может быть размещена ниже промотора РНК polIII и завершена терминатором поли (Т).

В некоторых вариантах воплощения изобретения, трансгены, описанные в настоящей заявке, могут иметь комбинацию из элементов, включая акцепторы сплайсинга, неполные кодирующие последовательности, терминаторы, плечи гомологии, левый и правый концы транспозазы и сайты для расщепления редкощепящими эндонуклеазами. В одном из вариантов воплощения изобретения, комбинация может быть от 5' до 3', [акцептор сплайсинга 1] - [неполная кодирующая последовательность 1] - [терминатор 1] - [терминатор 2 RC] - [неполная кодирующая последовательность 2 RC] - [акцептор сплайсинга 2 RC], где RC означает обратный комплемент. Эта комбинация может содержаться в линейной молекуле ДНК или молекуле AAV и может быть интегрирована NHEJ посредством целевого разрыва в гене-мишени. В другом варианте воплощения изобретения, комбинация может представлять собой от 5' до 3' [сайт 1 расщепления редкощепящей эндонуклеазой] - [акцептор сплайсинга 1] - [неполная кодирующая последовательность 1] - [терминатор 1] - [терминатор 2 RC] - [неполная кодирующая последовательность 2 RC] - [акцептор сплайсинга 2 RC] - [сайт 1 расщепления редкощепящей эндонуклеазой]. В другом варианте воплощения изобретения, комбинация может представлять собой от 5' до 3' [сайт 1 расщепления редкощепящей эндонуклеазой] - [плечо гомологии 1] - [акцептор сплайсинга 1] - [неполная кодирующая последовательность 1] - [терминатор 1] - [терминатор 2 RC] - [неполная кодирующая последовательность 2 RC] - [акцептор сплайсинга 2 RC] - [плечо гомологии 2] - [сайт 2 расщепления редкощепящей

эндонуклеазой]. В этой комбинации для содействия HR и NHEJ можно использовать одну или более редкощепящих эндонуклеаз. Например, одна редкощепящая нуклеаза может расщеплять целевой ген (т.е. желаемый интрон), а сайты расщепления, фланкирующие плечи гомологии, могут быть сконструированы таким образом, чтобы представлять собой одну и ту же целевую последовательность внутри интрона. В другом варианте воплощения изобретения, комбинация может представлять собой от 5' до 3' [плечо гомологии 1 + сайт расщепления редкощепящей эндонуклеазой 1] - [акцептор сплайсинга 1] - [неполная кодирующая последовательность 1] - [терминатор 1] - [терминатор 2 RC] - [неполная кодирующая последовательность 2 RC] - [акцептор сплайсинга 2 RC] - [плечо гомологии 2] - [сайт 1 расщепления редкощепящей эндонуклеазой]. В этой комбинации одна или более редкощепящих эндонуклеаз могут способствовать HR и NHEJ. Например, единичная редкощепящая нуклеаза может расщеплять в пределах плеча гомологии 1, ниже плеча гомологии 2, и в геномном сайте-мишени (т.е. в сайте с гомологией с последовательностью в плече гомологии 1). В другом варианте комбинация может быть от 5' до 3' [левый конец для транспозазы] - [акцептор сплайсинга 1] - [неполная кодирующая последовательность 1] - [терминатор 1] - [терминатор 2 RC] - [неполная кодирующая последовательность 2 RC] - [акцептор сплайсинга 2 RC] - [правый конец транспозазы]. Во всех вариантах воплощения изобретения, акцептор сплайсинга 1 и акцептор сплайсинга 2 могут быть одинаковыми или разными последовательностями; неполная кодирующая последовательность 1 и неполная кодирующая последовательность 2 могут быть одинаковыми или разными последовательностями; терминатор 1 и терминатор 2 могут быть одинаковыми или разными последовательностями.

В вариантах воплощения изобретения, трансген, содержащий структуру [сайт 1 расщепления редкощепящей эндонуклеазой] - [плечо гомологии 1] - [акцептор сплайсинга 1] - [неполная кодирующая последовательность 1] - [терминатор 1] - [терминатор 2 RC] - [неполная кодирующая последовательность 2 RC] - [акцептор сплайсинга 2 RC] - [плечо гомологии 2] - [сайт 2 расщепления редкощепящей эндонуклеазой], может быть интегрирован в ДНК посредством доставки одной или более редкощепящих эндонуклеаз. Если доставляется одна редкощепящая эндонуклеаза, редкощепящая эндонуклеаза может высвободить трансген путем расщепления в сайтах расщепления редкощепящей эндонуклеазой 1 и 2. Кроме того, та же самая редкощепящая эндонуклеаза может создать разрыв в гене-мишени, имитируя вставку через HR или NHEJ.

В других вариантах воплощения изобретения, трансген, содержащий структуру [плечо гомологии 1 + сайт расщепления редкощепящей эндонуклеазой 1] - [акцептор сплайсинга 1] - [неполная кодирующая последовательность 1] - [терминатор 1] -

[терминатор 2 RC] - [неполная кодирующая последовательность 2 RC] - [акцептор сплайсинга 2 RC] - [плечо гомологии 2] - [сайт расщепления редкощепящей эндонуклеазой 1], может быть интегрирован в ДНК с полной доставкой одной или более редкощепящих эндонуклеаз. Если доставляется одна редкощепящая эндонуклеаза, редкощепящая эндонуклеаза может высвободить транскрипт путем расщепления в сайтах расщепления редкощепящей эндонуклеазы 1 и 2. Кроме того, та же редкощепящая эндонуклеаза может создать разрыв в гене-мишени, имитируя вставку через HR или NHEJ. Интеграция с помощью HR может происходить, если расщепление происходит выше сайта интеграции (то есть в пределах плеча гомологии).

10 В вариантах воплощения изобретения, местом интеграции транскриптов может быть интрон или соединение интрон-экзон. При нацеливании на интрон неполная кодирующая последовательность может включать последовательность, кодирующую пептид, продуцируемый следующими экзонами внутри эндогенного гена. Например, если транскрипт предназначен для интеграции в интрон 9 эндогенного гена с 11 экзонами, то неполная кодирующая последовательность может включать последовательность, кодирующую пептид, продуцируемый экзонами 10 и 11 эндогенного гена. При нацеливании на соединение интрон-экзон транскрипт может быть сконструирован таким образом, чтобы он содержал плечи гомологии с последовательностью, гомологичной 3' указанного интрона.

15 В некоторых вариантах воплощения изобретения, неполные кодирующие последовательности могут быть полными кодирующими последовательностями. Полная кодирующая последовательность может кодировать эндогенный ген (например, фактор VIII, фактор IX или INS) или репортерные гены (например, RFP, GFP, cat, lacZ, люциферазу). Полные кодирующие последовательности могут быть функционально связаны с акцепторами и терминаторами сплайсинга и помещены в транскрипт в ориентации «хвост к хвосту».

20 Способы и композиции, представленные в настоящей заявке, можно применять для модификации эндогенных генов в клетках. Эндогенные гены могут включать фибриноген, протромбин, тканевой фактор, фактор V, фактор VII, фактор VIII, фактор IX, фактор X, фактор XI, фактор XII (фактор Хагемана), фактор XIII (фактор стабилизации фибрина), фактор фон Виллебранда, прекалликреин, высокомолекулярный кининоген (фактор Фитцджеральда), фибронектин, антитромбин III, кофактор гепарина II, протеин C, протеин S, протеин Z, ингибитор протеина Z, связанный с протеином Z, плазминоген, альфа-2-антиплазмин, тканевый активатор плазминогена, урокиназу, ингибитор активатора плазминогена-1, ингибитор активатора плазминогена-2, глюкоцереброзидазу (GBA), α -галактозидазу A (GLA), идуронатсульфатазу (IDS), идуронидазу (IDUA), кислую

сфингомиелиназу (SMPD1), MMAA, MMAB, MMACHC, MMADHC (C2orf25), MTRR, LMBRD1, MTR, пропионил-СоА карбоксилазу (PCC) (субъединицы PCCA и/или PCCA), белок-переносчик глюкозо-6-фосфата (G6PT) или глюкозо-6-фосфатазу (G6Pase), рецептор LDL (LDLR), ApoB, LDLRAP-1, PCSK9, митохондриальный белок, такой как NAGS (N-ацетилглутаматсинтетаза), CPS1 (карбамоилфосфатсинтетаза I) и OTC (орнитинтранскарбамилаза), ASS (синтетаза аргининоянтарной кислоты), ASL (лиаза аргининосукциназойной кислоты) и/или ARG1 (аргиназа), и/или семейство растворенных носителей (SLC25A13, аспартат/носитель глутамата), полипептид глюкуронилтрансферазы AUGT1A1 или UDP, фумарилацетоацетатгидролазу (FAH), белок аланин-глиоксилатаминотрансферазы (AGXT), белок глиоксилатредуктазы/гидроксипируватредуктазы (GRHPR), белок генов TRHPR, белок ATP7B, белок фенилаланингидроксилазы (PAH), белок USH2A, ATXN белок и белок липопротеинлиазы (LPL).

Трансген может включать последовательность для модификации последовательности, кодирующей полипептид, который отсутствует, нефункционален или имеет мутацию увеличения функции у субъекта, имеющего генетическое заболевание, включая без ограничений следующие генетические заболевания: ахондроплазию, ахроматопсию, дефицит кислой мальтазы, дефицит аденозиндезаминазы, аденолейкодистрофию, синдром Айкарди, дефицит альфа-1-антитрипсина, альфа-талассемию, синдром нечувствительности к андрогенам, синдром перта, аритмогенную дисплазию правого желудочка, атаксию телеанги-тазию, синдром Барта, бета-талассемию, синдром голубого резинового пузырчатого невуза, синдром Канавана, хронические гранулематозные заболевания (ХГБ), синдром кошачьего крика, муковисцидоз, болезнь деркума, эктодермальную дисплазию, анемию фанкони, прогрессирующую оссифицирующую фибродисплазию, синдром ломкой X-хромосомы, галактоземию, болезнь Гоше, генерализованные ганглиозидозы (например, GM1), гемохроматоз, мутацию гемоглобина С в 6-м кодоне бета-глобина (HbC), гемофилию, болезнь Хантингтона, синдром Гурлера, гипофосфатазию, синдром клейфентера, болезнь Краббса, синдром Лангера-Гидиона, дефицит адгезии лейкоцитов, лейкодистрофию, синдром удлиненного интервала QT, синдром Марфана, синдром Мебиуса, мукополисахаридоз (МПС), синдром ногтей-надколенника, несхарный нефрогенный диабет, , нейрофиброматоз, болезнь Неймана-Пика, несовершенный остеогенез, порфирию, синдром Прадера-Вилли, прогерия, синдром Протея, ретинобластому, синдром Ретта, синдром Рубинштейна-Тайби, синдром Санфилиппо, тяжелый комбинированный иммунодефицит (ТКИД), синдром Швахмана, серповидноклеточную болезнь (серповидноклеточная анемия), синдром Смита-

Магениса, синдром Стиклера, болезнь Тея-Сакса, синдром тромбоцитопении с отсутствием лучевой кости (TAR), синдром Тричера Коллинза, трисомию, туберозный склероз, синдром Тернера, нарушение цикла мочевины, болезнь фон Гиппеля-Ландау, синдром, Синдром Вильямса, болезнь Вильсона, синдром Вискотта-Олдрича, X-связанный лимфопролиферативный синдром, лизосомные болезни накопления (например, болезнь Гоше, GM1, болезнь Фабри и болезнь Тея-Сакса), мукополисахидоз (например, болезнь Хантера, болезнь Гурлера), гемоглобинопатии (например, серповидноклеточные заболевания, HbC, α -талассемия, β -талассемия) и гемофилии.

Дополнительные заболевания, которые можно лечить с помощью направленной интеграции, включают болезнь фон Виллебранда, синдром Ушера, поликистоз почек, спиноцеребеллярную атаксию 3 типа и спиноцеребеллярную атаксию 6 типа.

В одном из вариантов воплощения изобретения, геномная модификация представляет собой вставку трансгена в эндогенную геномную последовательность CACNA1A. Трансген может включать синтетическую и неполную кодирующую последовательность для белка CACNA1A. Неполная кодирующая последовательность может быть гомологичной кодирующей последовательности в гене CACNA1A дикого типа или функциональном варианте гена CACNA1A дикого типа, или мутанте гена CACNA1A дикого типа. В одном из вариантов воплощения изобретения, трансген, кодирующий неполный белок CACNA1A, вставлен в интрон 46 или начало экзона 47.

В другом варианте воплощения изобретения, геномная модификация представляет собой вставку трансгена в эндогенную геномную последовательность ATXN3. Трансген может включать синтетическую и неполную кодирующую последовательность для белка ATXN3. Неполная кодирующая последовательность может быть гомологичной кодирующей последовательности в гене ATXN3 дикого типа или функциональном варианте гена ATXN3 дикого типа, или мутанте гена ATXN3 дикого типа. В одном из вариантов воплощения изобретения, трансген, кодирующий неполный белок ATXN3, вставлен в интрон 9 или начало экзона 10.

В одном из вариантов воплощения изобретения, способы и композиции, описанные в настоящей заявке, можно применять для модификации 3'-конца эндогенного гена, тем самым приводя к модификации C-конца белка, кодируемого эндогенным геном. Модификация 3'-конца кодирующей последовательности эндогенного гена может включать замену последнего кодирующего экзона (т.е. экзона, содержащего стоп-кодон), вплоть до экзона, который находится между экзоном с начальным кодированием и последним экзоном. Указанный в настоящем документе термин «замена» относится к вставке ДНК в ген, при этом вставленная ДНК обеспечивает информацию для получения мРНК и белка 1

или более экзонов. Замена может происходить путем интеграции трансгена в эндогенный ген, при этом трансген содержит одну или более кодирующих последовательностей, функционально связанных с акцептором сплайсинга. Вставка может приводить или не приводить к удалению последовательности внутри эндогенного гена (например, к удалению интронов и экзонов). Например, если ген содержит 72 экзона, а стартовый кодон находится внутри экзона 1, модификация может включать замену экзонов 2-72, 3-72, 4-72, 5-72, 6-72, 7-72, 8-72, 9-72, 10-72, 11-72, 12-72, 13-72, 14-72, 15-72, 16-72, 17-72, 18-72, 19-72, 20-72, 21-72, 22-72 или 23-72, или 24-72, или 25-72, или 26-72, или 27-72, или 28-72, или 29-72, или 30-72, или 31-72, или 32-72, или 33-72, или 34-72, или 35-72, или 36-72, или 37-72, или 38-72, или 39-72, или 40-72, или 41-72, или 42-72, или 43-72, или 44-72, или 45-72, или 46-72, или 47-72, или 48-72, или 49-72, или 50-72, или 51-72, или 52-72, или 53-72, или 54-72, или 55-72, или 56-72, или 57-72, или 58-72, или 59-72, или 60-72, или 61-72, или 62-72, или 63-72, или 64-72, или 65-72, или 66-72, или 67-72, или 68-72, или 69-72, или 70-72, или 71-72, или 72. В одном из вариантов воплощения изобретения, экзоны эндогенного гена могут быть заменены путем интеграции трансгена в эндогенный ген, при этом трансген содержит первую и вторую неполные кодирующие последовательности, где первая и вторая неполные кодирующие последовательности кодируют пептид, продуцируемый экзонами эндогенных генов. Например, первая и вторая кодирующие последовательности трансгена могут кодировать пептид, который продуцируется экзонами эндогенного гена 2-72, 3-72, 4-72, 5-72, 6-72, 7-72, 8-72, 9-72, 10-72, 11-72, 12-72, 13-72, 14-72, 15-72, 16-72, 17-72, 18-72, 19-72, 20-72, 21-72, 22-72 или 23-72, или 24-72, или 25-72, или 26-72, или 27-72, или 28-72, или 29-72, или 30-72, или 31-72, или 32-72, или 33-72, или 34-72, или 35-72, или 36-72, или 37-72, или 38-72, или 39-72, или 40-72, или 41-72, или 42-72, или 43-72, или 44-72, или 45-72, или 46-72, или 47-72, или 48-72, или 49-72, или 50-72, или 51-72, или 52-72, или 53-72, или 54-72, или 55-72, или 56-72, или 57-72, или 58-72, или 59-72, или 60-72, или 61-72, или 62-72, или 63-72, или 64-72, или 65-72, или 66-72, или 67-72, или 68-72, или 69-72, или 70-72, или 71-72, или 72. Трансген может быть интегрирован в эндогенный ген в вышестоящем интроне или в начале экзона, соответствующего первому экзону в неполной кодирующей последовательности трансгена (ФИГ. 2). Трансген может быть сконструирован размером равным или менее 4,7 т.п.н. и включен в вектор и частицу AAV, и доставлен *in vivo* в клетки-мишени.

В варианте воплощения изобретения, трансген представляет собой последовательность ДНК, которая несет первую и вторую неполные кодирующие последовательности, при этом неполные кодирующие последовательности кодируют неполный белок, при этом неполный белок гомологичен соответствующей области в

функциональном белке, полученном из гена дикого типа. Ген-хозяин или эндогенный ген представляет собой ген, в котором экспрессия белка является aberrантной, другими словами, не экспрессируется, экспрессируется на низких уровнях или экспрессируется, но мРНК или белковый продукт, или его часть нефункциональны, имеют сниженные функции или имеют мутацию с приобретением функции, что приводит к расстройству в организме хозяина.

В соответствии с настоящей заявкой, донорная молекула может находиться в вирусном или невирусном векторе. Векторы могут быть в форме кольцевой или линейной двухцепочечной или одноцепочечной ДНК. Молекула-донор может быть конъюгирована или связана с реагентом, который способствует стабильности или обновлению клеток. Реагентом могут быть липиды, фосфат кальция, катионные полимеры, DEAE-декстран, дендримеры, пептиды, проникающие в клетки с помощью полиэтиленгликоля (PEG), инкапсулированные в газовой капсуле микропузырьки или магнитные шарики. Молекула-донор может быть включена в вирусную частицу. Вирус может быть ретровирусным, аденовирусным, аденоассоциированными векторами (AAV), вирусом простого герпеса, вирусом оспы, гибридным аденовирусным вектором, вирусом Эпштейна-Барр, лентивирусом или вирусом простого герпеса.

В некоторых вариантах воплощения настоящего изобретения, векторы AAV могут быть получены из любого AAV. В некоторых вариантах воплощения изобретения, вектор AAV получен из дефектного и непатогенного парвовирусного аденоассоциированного вируса 2 типа. Все такие векторы происходят из плазмиды, которая сохраняет только инвертированные концевые повторы AAV длиной 145 п.н., фланкирующие кассету экспрессии трансгена. Эффективный перенос генов и стабильная доставка трансгена за счет интеграции в геномы трансдуцированной клетки являются ключевыми особенностями этой векторной системы. (Wagner et al., *Lancet* 351:9117 1702-3, 1998; Kearns et al., *Gene Ther.* 9:748-55, 1996). Другие серотипы AAV, включая AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9 и AAVrh.10 и любой новый серотип AAV, также можно применять в соответствии с настоящим изобретением. В некоторых вариантах воплощения изобретения, химерный AAV применяется в том случае, если вирусное происхождение последовательностей длинного концевого повтора (LTR) вирусной нуклеиновой кислоты гетерологично вирусному происхождению последовательностей капсида. Неограничивающие примеры включают химерный вирус с LTR, полученным из AAV2, и капсидами, полученными из AAV5, AAV6, AAV8 или AAV9 (т.е. AAV2/5, AAV2/6, AAV2/8 и AAV2/9 соответственно).

Описанные в настоящей заявке конструкции также могут быть включены в систему

аденовирусных векторов. Векторы на основе аденовирусов обладают очень высокой эффективностью трансдукции во многих типах клеток и не требуют деления клеток. С такими векторами можно получить высокий титр и высокие уровни экспрессии.

5 Описанные в настоящей заявке способы и композиции могут применяться к любому эукариотическому организму, в котором желательно изменить организм посредством геномной модификации. К эукариотическим организмам относятся растения, водоросли, животные, грибы и простейшие. Эукариотические организмы также могут включать клетки растений, клетки водорослей, клетки животных, клетки грибов и клетки простейших.

10 Примеры клеток млекопитающих включают без ограничений ооциты, клетки K562, клетки CHO (яичника китайского хомячка), клетки HEP-G2, клетки BaF-3, клетки Шнайдера, клетки COS (клетки почек обезьяны, экспрессирующие Т-антиген SV40), клетки CV-1, клетки HuTu80, клетки NTERA2, клетки NB4, клетки HL-60 и клетки HeLa, клетки 293 (см., например, Graham et al. (1977) *J. Gen. Virol.* 36:59) и клетки миеломы, такие как SP2 или NS0 (см., например, Galfre and Milstein (1981) *Meth. Enzymol.* 73(B):3 46). Также
15 можно применять мононуклеарные клетки периферической крови (МКПК) или Т-клетки, а также эмбриональные и стволовые клетки взрослой особи. Например, стволовые клетки, которые можно применять, включают эмбриональные стволовые клетки (ES), индуцированные плюрипотентные стволовые клетки (iPSC), мезенхимальные стволовые клетки, гемопоэтические стволовые клетки, стволовые клетки печени, стволовые клетки
20 кожи и нейрональные стволовые клетки.

 Способы и композиции по настоящему изобретению можно применять в производстве модифицированных организмов. Модифицированными организмами могут быть мелкие млекопитающие, домашние животные, домашний скот и приматы. Неограничивающие примеры грызунов могут включать мышей, крыс, хомяков, песчанок и
25 морских свинок. Неограничивающие примеры домашних животных могут включать кошек, собак, кроликов, ежей и хорьков. Неограничивающие примеры домашнего скота могут включать лошадей, коз, овец, свиней, лам, альпак и крупный рогатый скот. Неограничивающие примеры приматов могут включать обезьян-капуцинов, шимпанзе, лемуров, макак, мартышек, тамаринов, паукообразных обезьян, беличьих обезьян и зеленых
30 мартышек. Способы и композиции по настоящему изобретению можно применять на людях.

 Примеры растений и растительных клеток, которые можно модифицировать с применением описанных в настоящей заявке способов, включают без ограничений однодольные растения (например, пшеница, кукуруза, рис, просо, ячмень, сахарный
35 тростник), двудольные растения (например, соя, картофель, томат, люцерна), плодовые

культуры (например, помидоры, яблоки, груши, клубника, апельсин), кормовые культуры (например, люцерна), корнеплоды (например, морковь, картофель, сахарная свекла, ямс), листовые овощные культуры (например, салат, шпинат); вегетативные культуры для потребления (например, соя и другие бобовые, кабачки, перец, баклажаны, сельдерей и т. д.), цветущие растения (например, петуния, роза, хризантема), хвойные деревья и сосны (например, сосна, пихта, ель); тополя (например, гибрид осины и тополя белого); волокнистые культуры (хлопок, джут, лен, бамбук), растения, применяемые в фиторемедиации (например, растения, накапливающие тяжелые металлы); масличные культуры (например, подсолнечник, семена рапса) и растения, применяемые в экспериментальных целях (например, резуховидка). Раскрытые в настоящей заявке способы можно применять в пределах родов *Asparagus*, *Avena*, *Brassica*, *Citrus*, *Citrullus*, *Capsicum*, *Cucurbita*, *Daucus*, *Erigeron*, *Glycine*, *Gossypium*, *Hordeum*, *Lactuca*, *Lolium*, *Lycopersicon*, *Malus*, *Manihot*, *Nicotiana*, *Orychophragmus*, *Oryza*, *Persea*, *Phaseolus*, *Pisum*, *Pyrus*, *Prunus*, *Raphanus*, *Secale*, *Solanum*, *Sorghum*, *Triticum*, *Vitis*, *Vigna* и *Zea*. Термин «растительные клетки» включает изолированные растительные клетки, а также целые растения или части целых растений, такие как семена, каллус, листья и корни. Настоящее изобретение также включает семена вышеописанных растений, в которых семена модифицированы с помощью описанных в настоящей заявке композиций и/или способов. Кроме того, настоящее изобретение включает потомство, клоны, клеточные линии или клетки трансгенных растений, описанных выше, при этом указанное потомство, клон, клеточная линия или клетка имеет трансген или генную конструкцию. Примеры видов водорослей включают микроводоросли, диатомовые водоросли, *Botryococcus braunii*, *Chlorella*, *Dunaliella tertiolecta*, *Gracilaria*, *Pleurochrysis carterae*, *Sorgassum* и *Ulva*.

Способы, представленные в настоящей заявке, могут включать применение редкощепящих эндонуклеаз для стимуляции гомологичной рекомбинации или негомологичной интеграции молекулы трансгена в эндогенный ген. Редкощепящая эндонуклеаза может включать CRISPR, TALEN или нуклеазы цинкового пальца (ZFN). Система CRISPR может включать CRISPR/Cas9 или CRISPR/Cas12a (Cpf1). Система CRISPR может включать варианты, демонстрирующие широкие возможности PAM (мотив, смежный с протоспейсером) (Hu et al., *Nature* 556, 57-63, 2018; Nishimasu et al., *Science* DOI: 10.1126, 2018) или более высокую активность точного связывания или расщепления (Kleinstiver et al., *Nature* 529:490-495, 2016). Реагент для редактирования гена может быть в формате нуклеазы (Mali et al., *Science* 339:823-826, 2013; Christian et al., *Genetics* 186:757-761, 2010), никазы (Cong et al., *Science* 339:819-823, 2013; Wu et al., *Biochemical and Biophysical Research Communications* 1:261-266, 2014), CRISPR-*FokI* dimers (Tsai et al.,

Nature Biotechnology 32:569-576, 2014) или спаренных нуклеаз CRISPR (Ran et al., *Cell* 154:1380-1389, 2013).

Способы и композиции, представленные в настоящей заявке, можно применять в
5 обстоятельствах, когда желательно модифицировать 3'-конец кодирующей
последовательности эндогенного гена. Например, пациенты с SCA3 или SCA6 имеют
увеличенные повторы CAG в экзонах 10 (предпоследний экзон) и экзоне 47 (последний
экзон), соответственно. Пациентам с SCA3 или SCA6 может быть полезна замена экзонов
10-11 и экзона 47, соответственно. В других примерах пациенты с генетическими
10 нарушениями из-за мутаций потери функции на 3'-конце эндогенного гена могут получить
пользу от замены последних экзонов указанного гена.

Далее изобретение будет описано в следующих примерах, которые не ограничивают
объем изобретения, описанный в формуле изобретения.

ПРИМЕРЫ

Пример 1: Направленная интеграция ДНК в ген ATXN3

15 Были сконструированы три плазмиды с трансгенами, предназначенными для
интеграции в ген ATXN3 в клетках человека. Все трансгены были разработаны для вставки
в интрон 9 или в место соединения интрона 9 и экзона 10 гена ATXN3, и все трансгены
были разработаны для вставки по меньшей мере одного акцептора сплайсинга и по меньшей
мере одной функциональной кодирующей последовательности для экзонов 10 и 11 гена
20 ATXN3. Первая плазида, обозначенная pBA1135, содержала левое и правое плечо
гомологии с последовательностями, гомологичными 3'-концу интрона 9 и 5'-концу интрона
10 (т. е. успешное нацеливание гена привело бы к удалению экзона 10 и замене его карго
последовательности в пределах pBA1135). Между плечами гомологии, от 5' до 3', находился
акцептор сплайсинга (акцептор сплайсинга из интрона 9 ATXN3), кодирующая
25 последовательность для экзонов 10 и 11 ATXN3, терминатор SV40, обратный терминатор
BGH, обратная кодирующая последовательность для экзонов 10 и 11 (кодон
скорректирован) и акцептор обратного сплайсинга. Последовательность трансгена
pBA1135 показана в SEQ ID NO: 17. Была сконструирована соответствующая нуклеаза Cas9
для расщепления i) внутри интрона 9 гена ATXN3, ii) внутри левого плеча гомологии
30 pBA1135 и iii) на участке 3'-конец правого плеча гомологии pBA1135. Ожидается, что
успешное расщепление плазмиды высвободит трансген, что позволит использовать
последовательность в качестве матрицы для HR или для интеграции через NHEJ. Сайт-
мишень пРНК Cas9 показан в SEQ ID NO: 18. Отдельные элементы в pBA1135 показаны в
SEQ ID NO: 44-51. SEQ ID NO: 44 включает левое плечо гомологии, сайт-мишень нуклеазы
35 и акцептор сплайсинга. SEQ ID NO: 45 содержит неполную кодирующую

последовательность (экзоны 10 и 11) непатогенного гена ATXN3. SEQ ID NO: 46 содержит терминаторную последовательность р(A) SV40. SEQ ID NO: 47 содержит терминатор BGH в обратном комплементе. SEQ ID NO: 48 содержит обратный комплемент, неполную кодирующую последовательность со скорректированными кодонами (экзоны 10 и 11) непатогенного гена ATXN3. SEQ ID NO: 49 содержит последовательность акцептора сплайсинга. SEQ ID NO: 50 включает последовательность для правого плеча гомологии. SEQ ID NO: 51 содержит последовательность сайта-мишени для нуклеазы. Вторая плаزمиды, обозначенная рВА1136, содержала то же карго, что и рВА1135, однако плечи гомологии были удалены. Сайты-мишени нуклеаз сохраняли для облегчения высвобождения трансгена из плазмиды. Ожидается, что успешное расщепление плазмиды приведет к высвобождению трансгена, что позволит использовать последовательность для интеграции NHEJ в ген ATXN3. Последовательность рВА1136 показана в SEQ ID NO: 19. Третья плаزمиды, обозначенная рВА1137, содержала ту же последовательность, что и рВА1135, за исключением обратных последовательностей и сайта-мишени нуклеазы (т.е. обратного терминатора, обратной кодирующей последовательности и обратной последовательности акцептора сплайсинга). Плазмиду рВА1137 использовали в качестве контроля для традиционных способов, основанных на HR. Последовательность рВА1137 представлена в SEQ ID NO: 20.

Трансфекцию проводили с использованием клеток НЕК293Т. Клетки НЕК293Т поддерживали при 37°C и 5% CO₂ в среде DMEM с высоким содержанием глюкозы с добавлением 10% фетальной бычьей сыворотки (FBS). Клетки НЕК293Т трансфицировали 2 мкг донора, 2 мкг гидовой РНК (формат РНК) и 2 мкг Cas9 (формат РНК). Трансфекцию проводили посредством электропорации. Геномную ДНК выделяли через 72 часа после трансфекции и оценивали на предмет интеграции. Список праймеров, используемых для обнаружения интеграции или геномной ДНК, показан в Таблице 1.

Таблица 1: Праймеры для обнаружения интеграции трансгенов в ATXN3.

Название праймера	Последовательность (от 5' до 3')	SEQ ID NO:
ONJB043	CAAAGGTGCCCTTGAGGTT	21
ONJB044	AGGAGAAGTCTGCCGTTACT	22
oNJB113	GGACAAACCACAAGTACAATGC	23
oNJB114	TAGGAAAGGACAGTGGGAGT	24
ONJB116	CCATTATGTCTCAGTTGTTTCAGTG	25
oNJB156	CCAGACCATCTCAGACACC	26
oNJB162	GGCTGGGCTTCCACTTAC	27

oNJB167	GTGGTTTGTCCAAACTCATCAA	28
oNJB170	AGTAACTCTGCACTTCCCATTG	29

Для обнаружения интеграции рВА1135, рВА1136 и рВА1137 проводили ПЦР на геномной ДНК. Что касается рВА1137, трансген был разработан для точной интеграции с помощью HR. Соответственно, полосы были обнаружены в ПЦР с 5'- и 3'-соединениями, что указывают на точную вставку в экзон 10 (ФИГ. 8, дорожки 4 и 7). Ожидаемые размеры

5 полосы составили 1520 п.н. для 5'-соединения и 786 п.н. для 3'-соединения. Праймеры oNJB113 и oNJB116 использовали для ПЦР с 5'-соединениями. Праймеры oNJB167 и oNJB170 использовали для ПЦР с 3'-соединением. Что касается рВА1136, поскольку плечи гомологии отсутствовали, предполагалось, что трансген будет вставлен посредством вставки NHEJ. Полосы подходящего размера наблюдались для трансгена,

10 интегрирующегося в прямом и обратном направлениях. Интеграцию в прямом направлении можно увидеть на ФИГ. 8, дорожки 3 (ожидаемый размер приблизительно 1520 п.н.) и 6 (ожидаемый размер приблизительно 1519 п.н.). Интегрирование в обратном направлении можно увидеть на ФИГ. 8, дорожка 12 (ожидаемый размер приблизительно 1520 п.н.). Праймеры oNJB113 и oNJB116 использовали для ПЦР с 5'-соединениями. Праймеры

15 oNJB114 и oNJB170 использовали для ПЦР с 3'-соединением. Праймеры oNJB116 и oNJB114 использовали для ПЦР с обратным 5'-соединением. Что касается ррВА1135, на трансгене присутствовали как плечи гомологии, так и сайты расщепления нуклеазами. Интеграцию с помощью HR наблюдали путем обнаружения полос в ПЦР с 5'- и 3'-соединениями (ФИГ. 8, дорожки 2 и 5). Кроме того, интеграцию посредством NHEJ

20 наблюдали путем обнаружения полос в ПЦР с обратным 5'-соединением (ФИГ. 8, дорожка 10). Ожидаемый размер для ПЦР с 5'-соединениями составил 1520 п.н. Ожидаемый размер для ПЦР с 3'-соединением составил 1157 п.н. Ожидаемый размер для ПЦР с обратным 5'-соединением составил приблизительно 1520 п.н. Праймеры oNJB113 и oNJB116 использовали для ПЦР с 5'-соединениями. Праймеры oNJB114 и oNJB170 использовали для

25 ПЦР с 3'-соединением. Праймеры oNJB116 и oNJB114 использовали для ПЦР с обратным 5'-соединением.

Эти результаты демонстрируют, что описанные трансгены, содержащие двунаправленные неполные кодирующие последовательности, могут быть интегрированы в геномную ДНК посредством множества различных путей репарации.

30 Пример 2: Направленная интеграция ДНК в ген CACNA1A

Трансген, нацеленный на CACNA1A, предназначен для замены 3'-конца кодирующей последовательности CACNA1A. Конструируют плазмиду с трансгеном, предназначенным для интеграции кодирующей последовательности WT в интрон 46 или

5 начало экзона 47 (ФИГ. 4). Трансген содержит первое плечо гомологии, которое гомологично последовательности, следующей сразу за донорным сайтом сплайсинга в интроне 46. Первое плечо гомологии также включает целевой сайт для нуклеазы (SEQ ID NO: 9) и акцепторную последовательность сплайсинга. За первым плечом гомологии следует первая кодирующая последовательность, содержащая экзон 47 CACNA1A и последовательность нерасширенного повтора CAG (SEQ ID NO: 3). После первой кодирующей последовательности следует последовательность терминации поли(А) SV40 (SEQ ID NO: 4). При ориентации «хвост к хвосту» присутствует второй набор функциональных элементов. Начало второго набора элементов включает целевой сайт для нуклеазы (SEQ ID NO: 9), за которым следует второе плечо гомологии. Второе плечо гомологии содержит 446 п.н., которые гомологичны последовательности, следующей сразу за стоп-кодированием (SEQ ID NO: 8). Путем анализа *in silico* эта последовательность была определена как несодержащая консенсусную ветвь или акцепторные последовательности сплайсинга. За вторым плечом гомологии следует второй акцептор сплайсинга от интрона 15 1 бета-актина карпа (SEQ ID NO: 7). Следом за акцептором сплайсинга находится оптимизированная по кодонам версия экзона 47 CACNA1A (SEQ ID NO: 6) и терминатор поли(А) bGH (SEQ ID NO: 5).

Соответствующая нуклеаза Cas12a предназначена для создания трех двухцепочечных разрывов после трансфекции плазмиды: 1) внутри интрона 46 эндогенного гена CACNA1A, 2) внутри первого плеча гомологии в трансгене pBA1011-D1 и 3) после второго плеча гомологии в трансгене pBA10111-D1. Целевая последовательность для нуклеазы Cas12a представлена в SEQ ID NO: 9.

20 Подтверждение функции трансгена и векторов CRISPR достигается путем трансфекции клеток HEK293. Клетки HEK293 поддерживают при 37°C и 5% CO₂ в среде DMEM с высоким содержанием глюкозы без L-глутамин без пирувата натрия, дополненной 10% фетальной бычьей сывороткой (FBS) и 1% раствором пенициллин-стрептомицина (PS) 100X. Клетки HEK293 трансфицируют каждой из плазмидных конструкций и их комбинаций с применением Lipofectamine 3000. Через два дня после трансфекции ДНК экстрагируется и оценивается на предмет мутаций и целевых вставок в гене CACNA1A. Нуклеазную активность анализируют с помощью анализа Cel-I или путем глубокого секвенирования ампликонов, содержащих последовательность-мишень CRISPR/Cas12a. Успешная интеграция трансгена анализируется с помощью ПЦР (ФИГ. 5).

Пример 3: Направленная интеграция ДНК в ген ATXN3

35 Трансген, нацеленный на ATXN3, предназначен для замены 3'-конца последовательности, кодирующей ATXN (экзоны 10 и 11). Конструируют плазмиду с

трансгеном, предназначенным для интеграции кодирующей последовательности WT в интрон 9 или начало экзона 10 (ФИГ. 5). Трансген включает первое плечо гомологии, которое гомологично интрону последовательности 9 (SEQ ID NO: 10). Первое плечо гомологии также включает целевой сайт для нуклеазы Cas12a и акцепторную последовательность сплайсинга. За первым плечом гомологии следует первая кодирующая последовательность, содержащая экзон 10 и 11 ATXN3 и последовательность нерасширенного повтора CAG. После первой кодирующей последовательности следует последовательность терминации поли(A) SV40. При ориентации «хвост к хвосту» присутствует второй набор функциональных элементов. Начало второго набора элементов включает целевой сайт для нуклеазы Cas12a, за которым следует второе плечо гомологии. Второе плечо гомологии содержит 379 п.н., что гомологично последовательности, следующей сразу за концом экзона 10 (то есть началом интрона 10). Эта последовательность была определена с помощью анализа *in silico* с тем, чтобы иметь ограниченное количество потенциальных последовательностей ответвлений или акцепторов сплайсинга. За вторым плечом гомологии идет второй акцептор сплайсинга от интрона 1 бета-актина карпа. За акцептором сплайсинга идет оптимизированная по кодонам версия экзонов 10 и 11 ATXN3 и терминатор поли(A) bGH.

Соответствующая нуклеаза Cas12a предназначена для создания трех двухцепочечных разрывов после трансфекции плазмиды: 1) внутри интрона 9 эндогенного гена ATXN3, 2) внутри первого плеча гомологии в трансгене pBA1012-D1 и 3) после второго плеча гомологии в трансгене pBA1012-D1. Целевая последовательность для нуклеазы Cas12a представлена в SEQ ID NO: 11.

Подтверждение функции трансгена и векторов CRISPR достигается путем трансфекции клеток HEK293. Клетки HEK293 поддерживают при 37°C и 5% CO₂ в среде DMEM с высоким содержанием глюкозы без L-глутамин без пирувата натрия, дополненной 10% фетальной телячьей сывороткой (FBS) и 1% раствором пенициллин-стрептомицина (PS) 100X. Клетки HEK293 трансфицируют каждой из плазмидных конструкций и их комбинаций с применением Lipofectamine 3000. Через два дня после трансфекции ДНК экстрагируют и оценивают на наличие мутаций и целевых вставок в гене ATXN3. Нуклеазную активность анализируют с помощью анализа Cel-I или путем глубокого секвенирования ампликонов, содержащих последовательность-мишень CRISPR/Cas12a. Успешная интеграция трансгена анализируется с помощью ПЦР (ФИГ. 7).

Пример 4: Направленная интеграция ДНК в ген ATXN3 путем применения транспозаз Cas12k

Трансген, нацеленный на ATXN3, предназначен для замены 3'-конца кодирующей

последовательности ATXN (экзоны 10 и 11). Плазида конструируется с трансгеном, предназначенным для интеграции кодирующей последовательности WT в интрон 9 или начало экзона 10. Трансген включает правый и левый концы транспозона, первый и второй акцепторы сплайсинга, первую и вторую кодирующие последовательности (кодирующие аминокислоты из экзонов 10 и 11), а также первый и второй терминаторы. Последовательность между правым и левым концами транспозона показана в SEQ ID NO: 17.

Плазмиды сконструированы для экспрессии *tnsB*, *tnsC*, *tniQ* и *Cas12k Scytonema hofmanni* (SEQ ID NO: 30) с использованием эукариотических промоторов. Вторая плазида сконструирована для экспрессии соответствующей гидовой РНК *Cas12k* (SEQ ID NO: 14). Целевая последовательность гидовой РНК CCGCCGACCTTTCACTTTC (SEQ ID NO: 15). Плазмиды транспозонов *Cas12k* котрансформируются в клетках НЕК293 плазмидой, несущей трансген, нацеленный на ATXN3. Клетки НЕК293 поддерживают при 37°C и 5% CO₂ в среде DMEM с высоким содержанием глюкозы без L-глутамина и без пирувата натрия, дополненной 10% фетальной телячьей сывороткой (FBS) и 1% раствором пенициллин-стрептомицина (PS) 100X. Клетки НЕК293 трансфицируют каждой из плазмидных конструкций и их комбинаций с применением Lipofectamine 3000. Через два дня после трансфекции ДНК экстрагируют и оценивают на предмет целевых вставок в ген ATXN3. Интеграцию трансгена анализируют с помощью ПЦР.

20 Пример 5: Направленная интеграция ДНК в ген CACNA1A

Трансген, нацеленный на CACNA1A, предназначен для замены 3'-конца кодирующей последовательности CACNA1A. Плазида конструируется с трансгеном, предназначенным для интеграции кодирующей последовательности WT в интрон 46 или начало экзона 47. Трансген включает правый и левый концы транспозона, первый и второй акцепторы сплайсинга, первую и вторую кодирующие последовательности (кодирующие аминокислоты из экзона 47), а также первый и второй терминаторы.

Плазмиды сконструированы для экспрессии *Scytonema hofmanni tnsB*, *tnsC*, *tniQ* и *Cas12k* (SEQ ID NO: 30) с использованием эукариотических промоторов. Вторую плазмиду конструируют для экспрессии соответствующей гидовой РНК *Cas12k* (SEQ ID NO: 14). Гидовая РНК предназначена для целевой последовательности CCGGATCCCGGCTGTGACC (SEQ ID NO: 16). Плазмиды транспозонов *Cas12k* котрансформируются в клетках НЕК293 плазмидой, несущей трансген, нацеленный на ATXN3. Клетки НЕК293 поддерживают при 37°C и 5% CO₂ в среде DMEM с высоким содержанием глюкозы без L-глутамина и без пирувата натрия, дополненной 10% фетальной телячьей сывороткой (FBS) и 1% раствором пенициллин-стрептомицина (PS) 100X. Клетки

НЕК293 трансфицируют каждой из плазмидных конструкций и их комбинаций с применением Lipofectamine 3000. Через два дня после трансфекции ДНК экстрагируют и оценивают на предмет целевых вставок в ген АТХN3. Интеграцию трансгена анализируют с помощью ПЦР.

5

ДРУГИЕ ВАРИАНТЫ ВОПЛОЩЕНИЯ ИЗОБРЕТЕНИЯ

Следует понимать, что, хотя настоящее изобретение описано в разделе «Подробное описание изобретения», данное описание приведено в качестве примера, а не для ограничения объема изобретения, описанного в прилагаемой формуле изобретения, объем которой включает и другие аспекты, преимущества и модификации настоящего изобретения.

10

ФОРМУЛА ИЗОБРЕТЕНИЯ:

1. Способ интеграции трансгена в эндогенный ген, включающий:
 - а. введение трансгена, при этом трансген содержит
 - i. первую и вторую акцепторные последовательности сплайсинга,
 - ii. первую и вторую неполные кодирующие последовательности, и
 - iii. один двунаправленный терминатор или первый и второй терминаторы;
 - б. введение по меньшей мере одной редкощепящей эндонуклеазы, нацеленной на сайт эндогенного гена,
при этом трансген интегрирован в эндогенный ген.
- 5 2. Способ по п.1, отличающийся тем, что первый акцептор сплайсинга функционально связан с первой неполной кодирующей последовательностью, а второй акцептор сплайсинга функционально связан со второй неполной кодирующей последовательностью.
3. Способ по п.2, отличающийся тем, что первая неполная кодирующая последовательность функционально связана с первым терминатором, а вторая неполная
15 кодирующая последовательность функционально связана со вторым терминатором.
4. Способ по п.2, отличающийся тем, что первая и вторая неполные кодирующие последовательности функционально связаны с двунаправленным терминатором.
5. Способ по п.3, отличающийся тем, что первый и второй акцепторы сплайсинга, первая и вторая кодирующие последовательности, а также первый и второй терминаторы
20 направлены в ориентации «хвост к хвосту».
6. Способ по п.5, отличающийся тем, что трансген дополнительно содержит первый и второй сайты-мишени для одной или более редкощепящих эндонуклеаз, при этом сайты-мишени фланкируют первый и второй акцепторы сплайсинга.
7. Способ по п.5, отличающийся тем, что трансген дополнительно содержит левое и
25 правое плечи гомологии, которые фланкируют первый и второй акцепторы сплайсинга.
8. Способ по п.7, отличающийся тем, что трансген содержится в аденоассоциированном вирусном векторе.
9. Способ по п.7, отличающийся тем, что трансген дополнительно содержит первый и второй сайты-мишени для одной или более редкощепящих эндонуклеаз, при этом сайты-мишени фланкируют первый и второй акцепторы сплайсинга.
30
10. Способ по п.9, отличающийся тем, что первый и второй сайты-мишени фланкируют первое и второе плечи гомологии.
11. Способ по п.1, отличающийся тем, что трансген интегрирован в интрон эндогенного гена или в место соединения интрон-экзон.
- 35 12. Способ по п.1, отличающийся тем, что трансген интегрирован в интрон или в место

соединения интрон-экзон гена ATXN3 или гена CACNA1A.

13. Способ по п. 12, отличающийся тем, что трансген содержит первую и вторую неполные кодирующие последовательности, кодирующие пептид, продуцируемый экзоном 10 непатогенного гена ATXN3, и нацелен на интрон 9 или место соединения интрона 9 с экзоном 10 патогенного гена ATXN3.
14. Способ по п. 12, отличающийся тем, что трансген содержит первую и вторую неполные кодирующие последовательности, кодирующие пептид, продуцируемый экзоном 47 непатогенного гена CACNA1A, и нацелен на интрон 46 или место соединения интрона 46 с экзоном 47 патогенного гена CACNA1A.
15. Способ по п.1, отличающийся тем, что нуклеаза представляет собой нуклеазу CRISPR/Cas12a или нуклеазу CRISPR/Cas9.
16. Способ по п.1, отличающийся тем, что первая и вторая неполные кодирующие последовательности кодируют одни и те же аминокислоты.
17. Способ по п.1, отличающийся тем, что первая и вторая кодирующие последовательности отличаются последовательностью нуклеиновой кислоты, но кодируют одни и те же аминокислоты.
18. Способ по п.1, отличающийся тем, что трансген содержится в векторе, где формат вектора выбран из двухцепочечной линейной ДНК, двухцепочечной кольцевой ДНК или вирусного вектора.
19. Способ по п.18, отличающийся тем, что вирусный вектор выбран из аденовирусного вектора, аденоассоциированного вирусного вектора или лентивирусного вектора.
20. Способ по п.19, отличающийся тем, что трансген равен или менее 4,7 т.п.н.
21. Способ по п.1, отличающийся тем, что указанный эндогенный ген представляет собой ген дикого типа, указанных неполных кодирующих последовательностей.
22. Способ по п.21, отличающийся тем, что указанный эндогенный ген является aberrantным, а неполные кодирующие последовательности кодируют неполный белок из функциональной версии указанного эндогенного гена.
23. Полинуклеотид ДНК, содержащий:
- первую и вторую акцепторные последовательности сплайсинга,
 - первую и вторую неполные кодирующие последовательности,
 - один двунаправленный терминатор или первый и второй терминаторы,
 - необязательно, первое и второе плечи гомологии, и
 - необязательно, первый и второй сайты-мишени редкощепящей эндонуклеазы.
24. Способ интеграции трансгена в эндогенный ген, включающий:

- 5
- a. введение трансгена, при этом трансген содержит
 - i. левый и правый концы транспозона,
 - ii. первую и вторую акцепторные последовательности сплайсинга,
 - iii. первую и вторую неполные кодирующие последовательности, и
 - iv. один двунаправленный терминатор или первый и второй терминаторы;
 - б. введение транспозазы

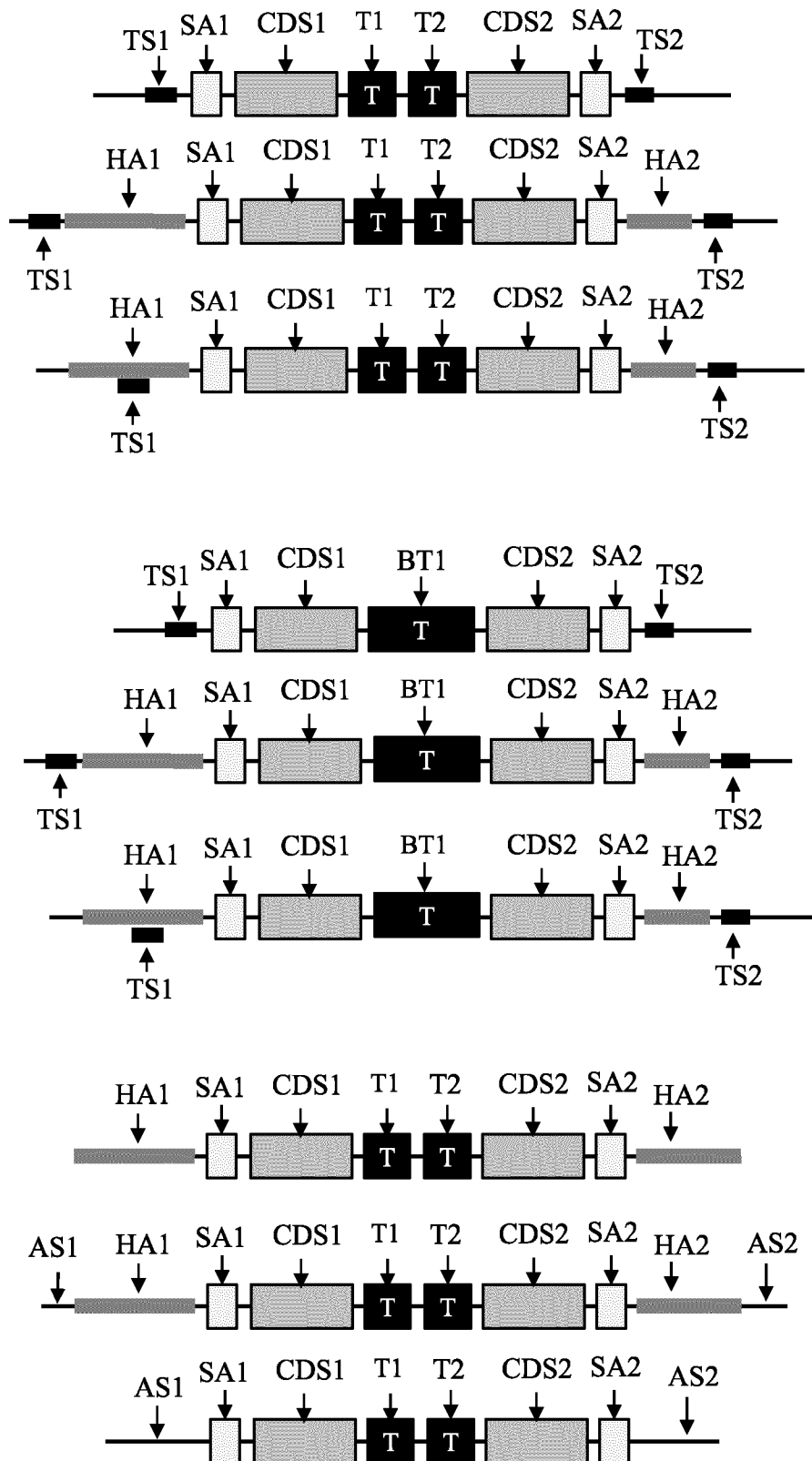
при этом трансген интегрирован в эндогенный ген.

25. Способ интеграции трансгена в эндогенный ген, включающий:
- 10
- a. введение трансгена, при этом трансген содержит
 - i. первую и вторую акцепторные последовательности сплайсинга,
 - ii. первую и вторую кодирующие последовательности, и
 - iii. один двунаправленный терминатор или первый и второй терминаторы;
 - б. введение по меньшей мере одной редкощепящей эндонуклеазы, нацеленной на сайт эндогенного гена,

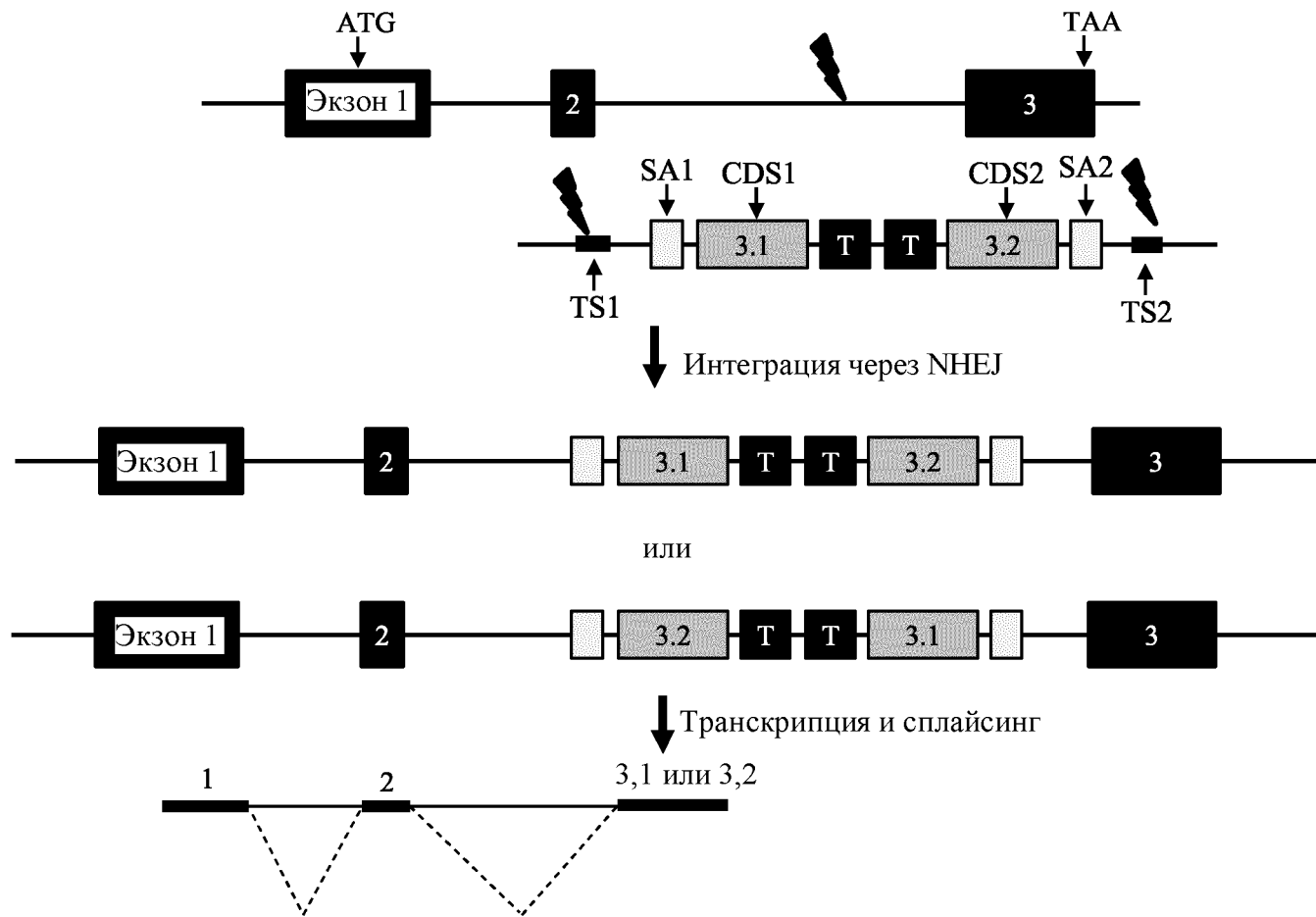
15 при этом трансген интегрирован в эндогенный ген.

26. Способ интеграции трансгена в эндогенный ген, включающий:
- 20
- a. введение трансгена, при этом трансген содержит
 - i. первую и вторую акцепторные последовательности сплайсинга,
 - ii. первую и вторую кодирующие последовательности,
 - iii. один двунаправленный терминатор или первый и второй терминаторы, и
 - iv. первое и второе плечи гомологии
- при этом трансген интегрирован в эндогенный ген.

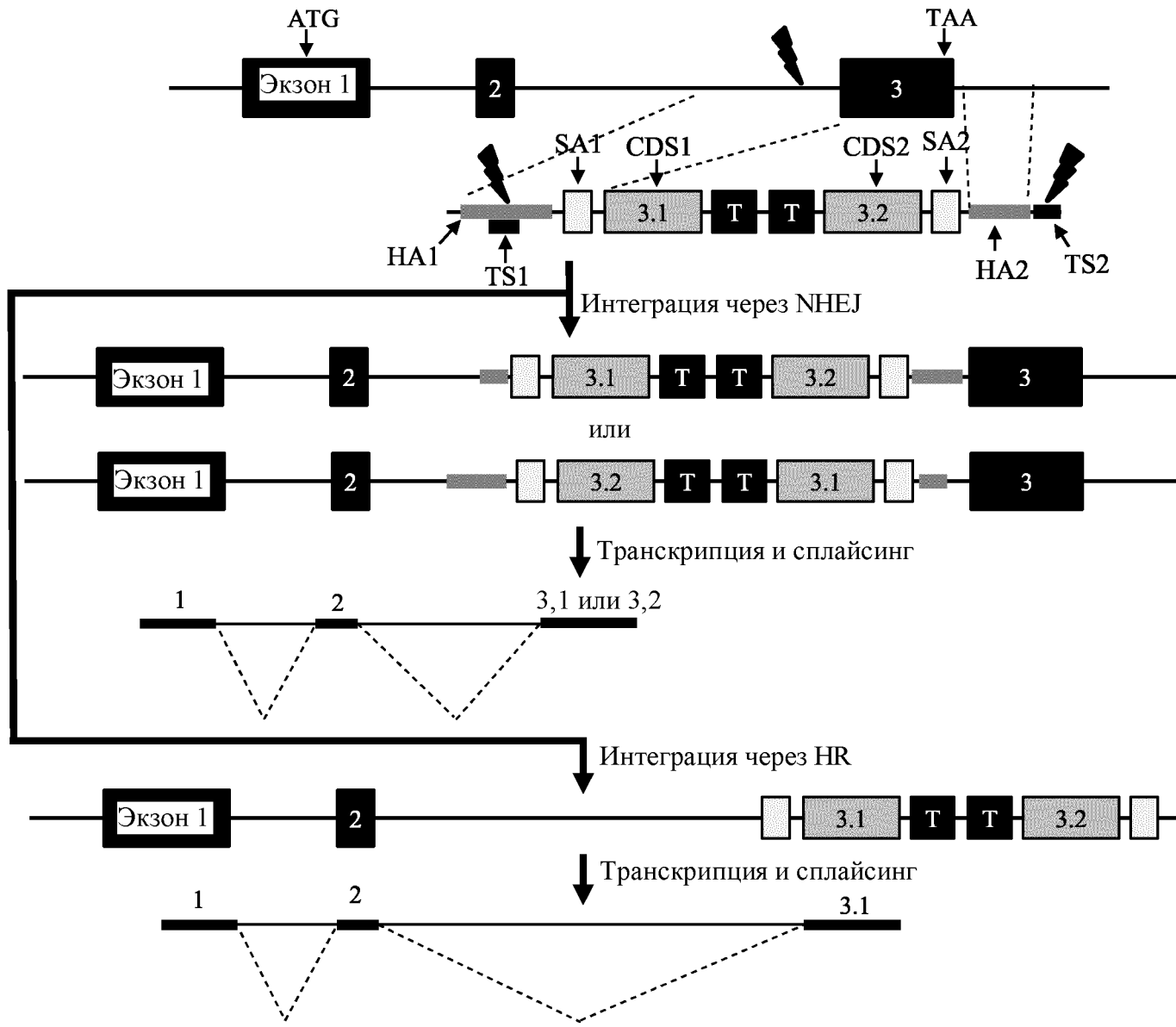
1/8



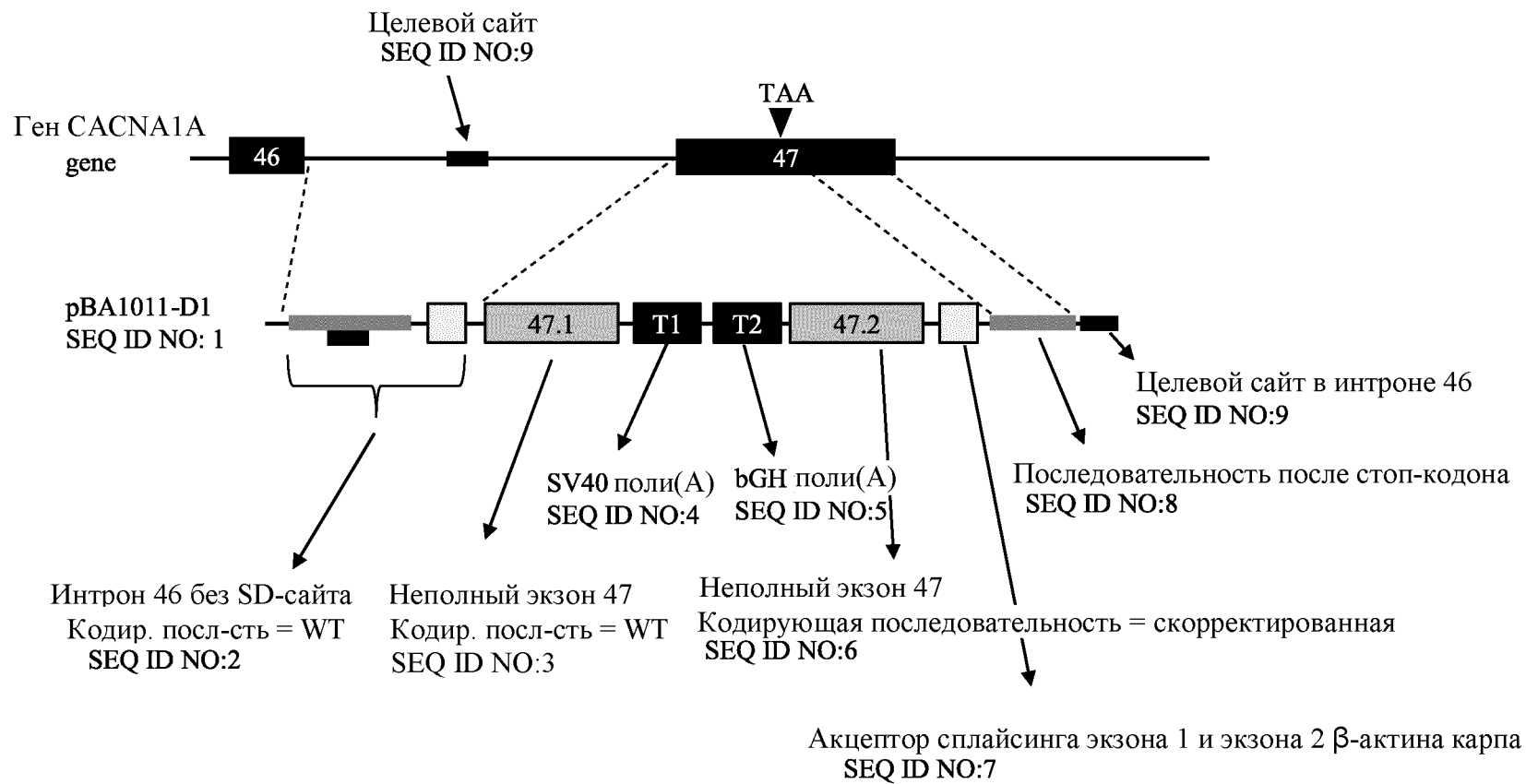
ФИГ. 1



ФИГ. 2



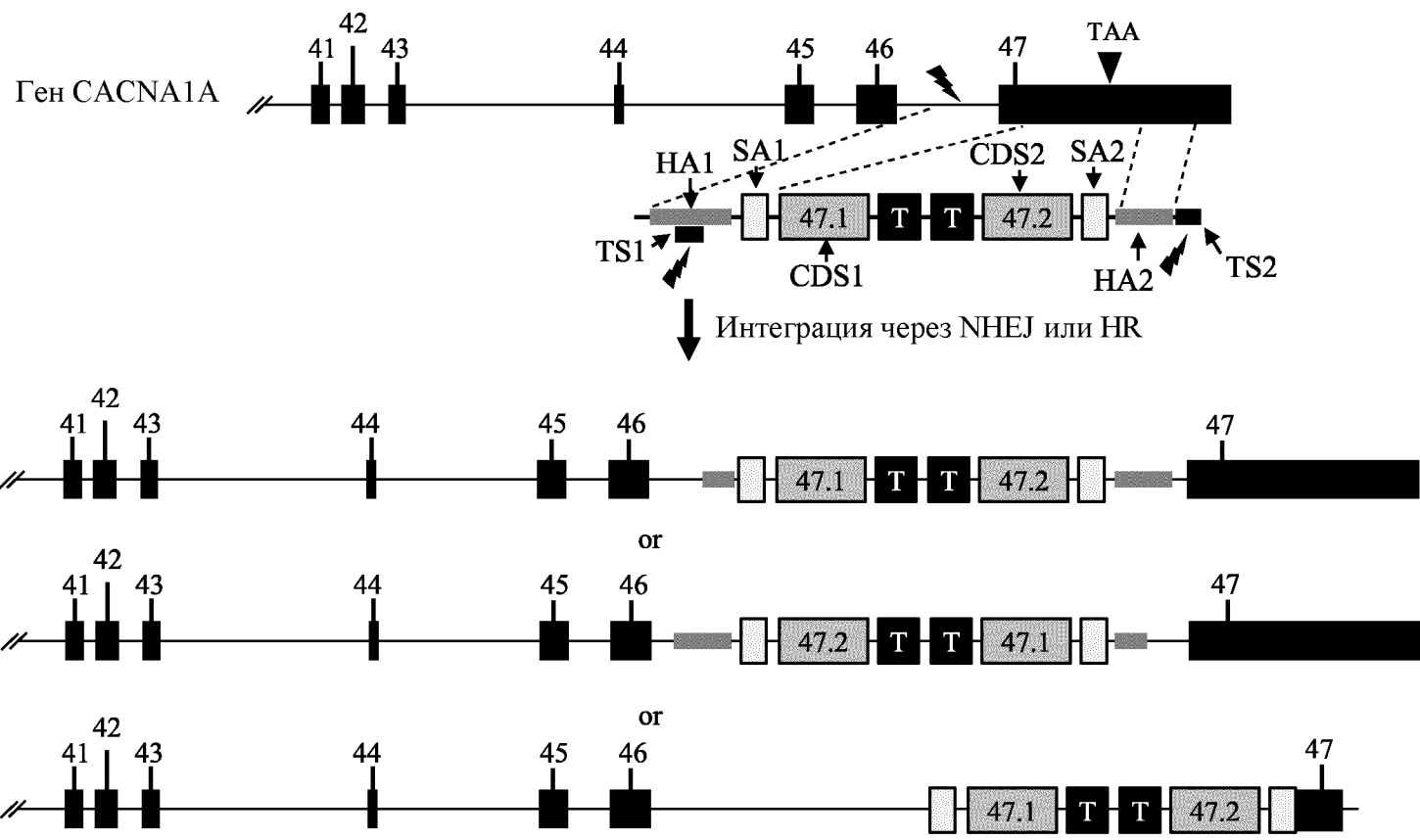
ФИГ. 3



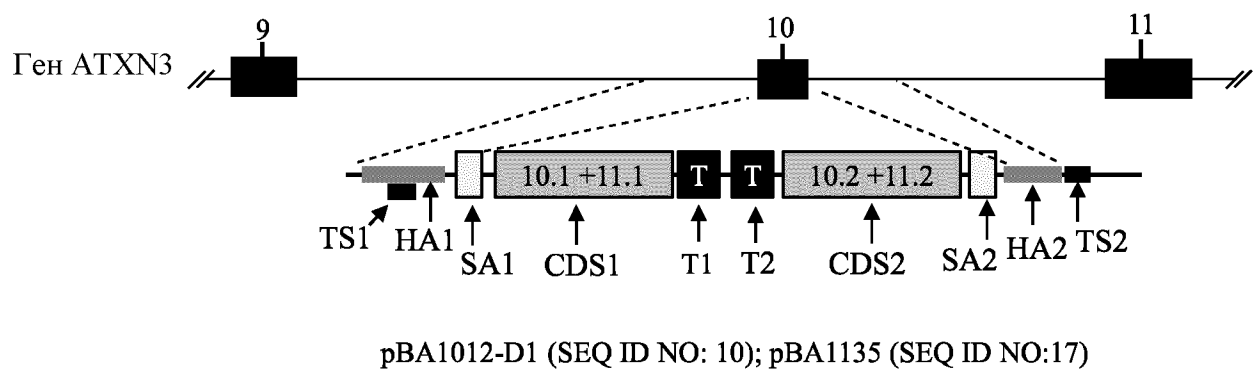
4/8

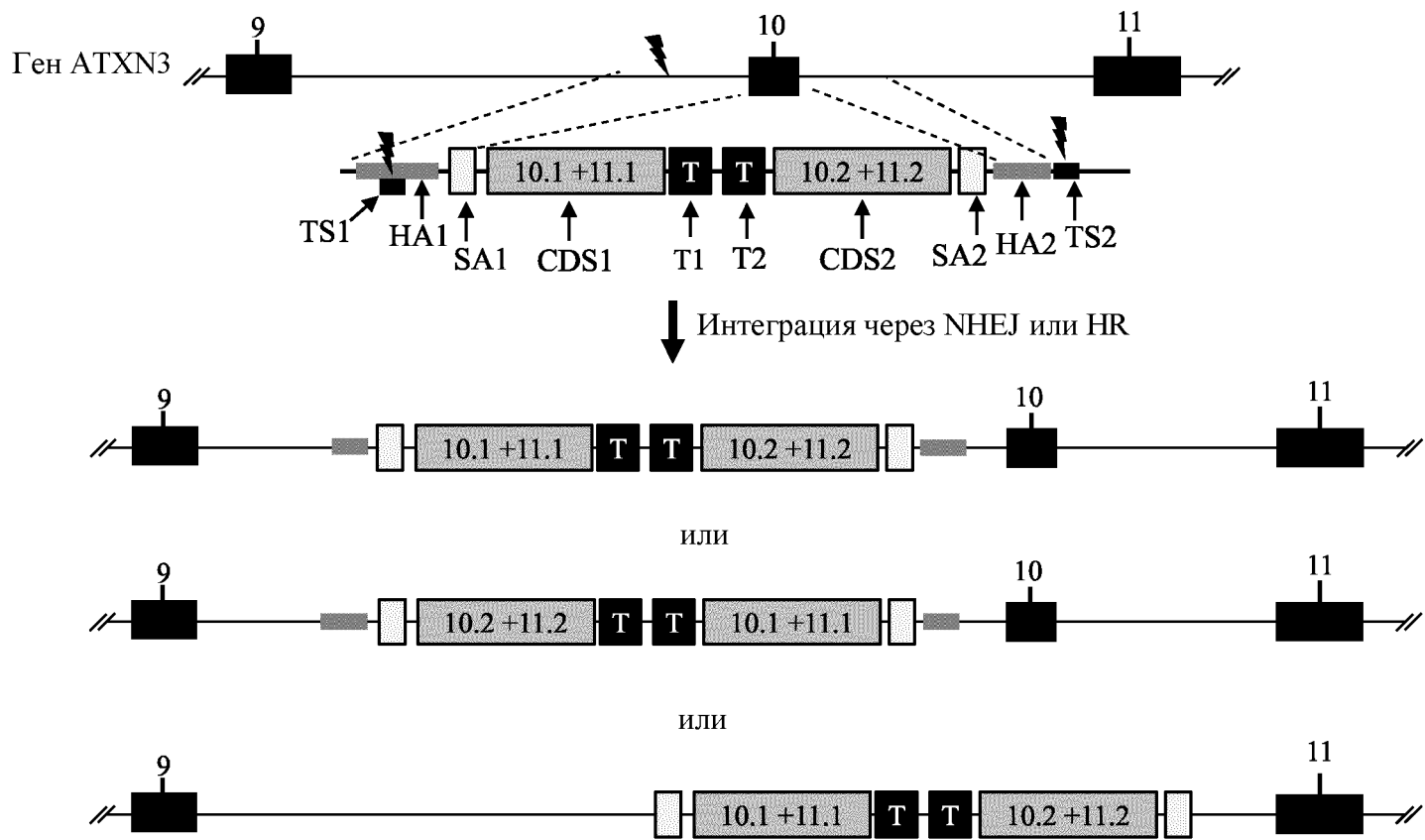
4/8

ФИГ. 4

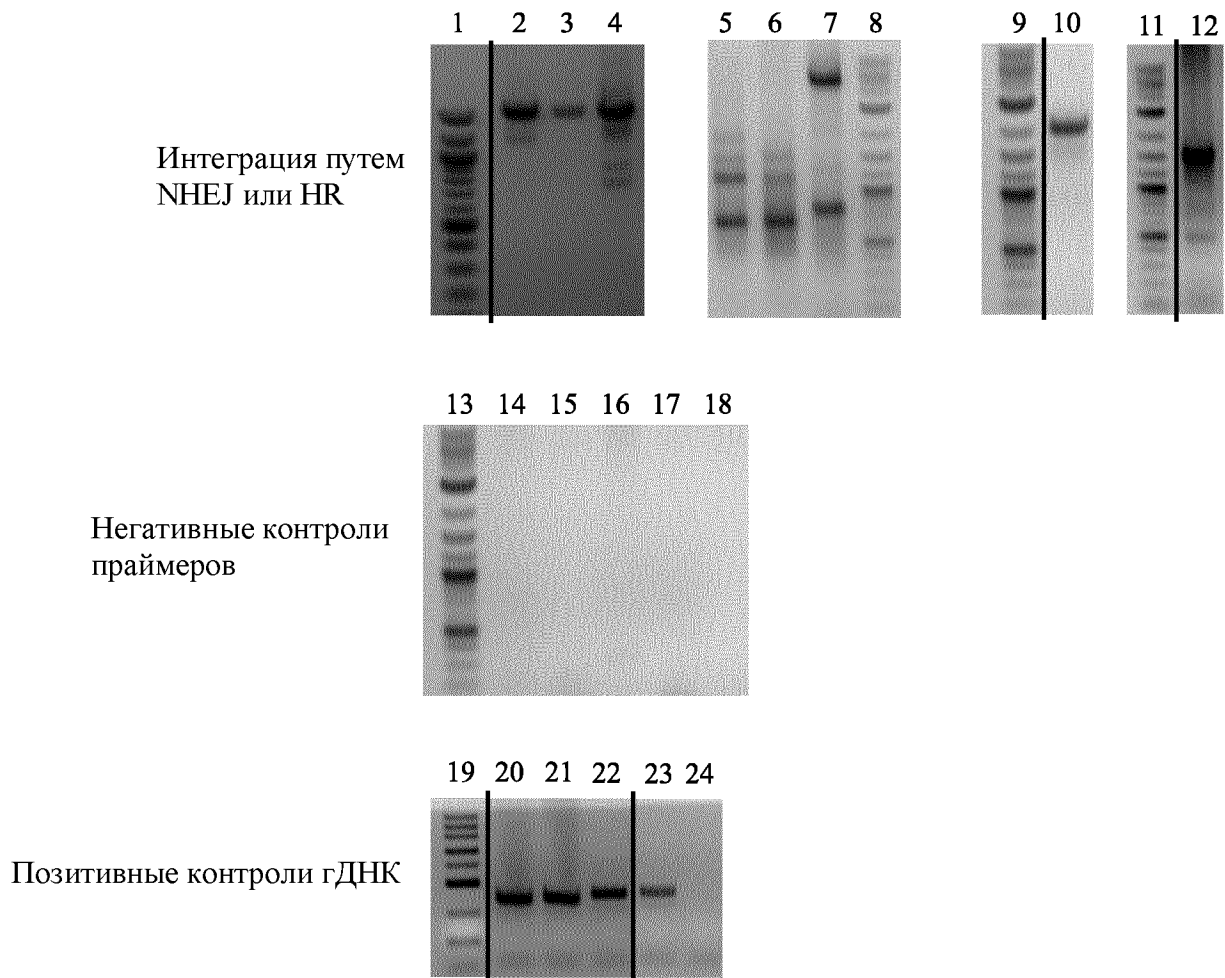


ФИГ. 5





ФИГ. 7

8/8**ФИГ. 8**