

КОДОН-ОПТИМИЗИРОВАННАЯ НУКЛЕИНОВАЯ КИСЛОТА, КОДИРУЮЩАЯ БЕЛОК SMN1

ОБЛАСТЬ ТЕХНИКИ

Настоящая заявка относится к области генетики, генной терапии и молекулярной биологии. Более конкретно, настоящее изобретение относится к выделенной кодон-оптимизированной нуклеиновой кислоте, которая кодирует белок SMN1 (белок выживаемости моторных (двигательных) нейронов), экспрессионной кассете и вектору на ее основе, а также к рекомбинантному вирусу на основе AAV9 (аденоассоциированный вирус 9 серотипа) для увеличения экспрессии гена SMN1 в целевых клетках, и их применению.

УРОВЕНЬ ТЕХНИКИ

Спинальная мышечная атрофия (SMA) представляет собой аутомное рецессивное нервно-мышечное нарушение, вызванное мутациями в гене выживаемости моторных (двигательных) нейронов 1 (SMN1) и утратой кодируемого SMN белка (Lefebvre et al., Cell (1995) 80:155-165). Отсутствие SMN ведет к дегенерации двигательных нейронов в брюшном (переднем) роге спинного мозга, что ведет к слабости проксимальных мышц, отвечающих за ползание, ходьбу, движение шеи и глотание, и непроизвольно сокращающихся мышц, которые управляют дыханием и кашлем (Sumner C.J., NeuroRx (2006) 3:235-245). Таким образом, пациенты с SMA предрасположены к пневмониям и другим пульмональным проблемам, таким как рестриктивное легочное заболевание.

Генная терапия представляет собой перспективный способ лечения спинальной мышечной атрофии (SMA).

Аденоассоциированные вирусные (AAV) векторы считаются эффективными для генной терапии ЦНС, поскольку они обладают подходящим профилем токсичности и иммуногенности, их можно использовать в трансдукции нервных клеток, и они способны опосредовать длительную экспрессию в ЦНС.

Аденоассоциированный вирус (AAV) представляет собой небольшой (20 нм), неспособный к самостоятельной репликации, безоболочечный вирус. У человека и приматов описано множество различных серотипов AAV. Геном аденоассоциированного вируса содержит (+ или -) одноцепочечную ДНК (ssDNA) длиной около 4,7 тысяч нуклеотидов. На концах молекулы геномной ДНК располагаются инвертированные концевые повторы (англ. inverted terminal repeats, ITRs). Геном содержит две открытые рамки считывания (англ. ORF): Rep и Cap, содержащие в себе несколько альтернативных

рамок считывания, кодирующих различные белковые продукты. Продукты Rep имеют важное значение для репликации AAV, при этом ген Cap, помимо других альтернативных продуктов, кодирует 3 капсидных белка (VP1, VP2 и VP3). Белки VP1, VP2 и VP3 находятся в соотношении 1:1:10, образуя икосаэдрический капсид (Xie Q. et al. The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. Proc Natl Acad Sci USA, 2002; 99:10405-10410). При образовании рекомбинантного вектора AAV (rAAV) кассета экспрессии, фланкированная ITR, упаковывается в капсид AAV. Гены, необходимые для репликации AAV, не входят в кассету. Рекомбинантный AAV считается самым безопасным и одним из наиболее широко используемых вирусных векторов для переноса генов *in vivo*. Векторы могут инфицировать клетки из тканей множества типов, обеспечивая мощную и устойчивую трансгенную экспрессию. Они также являются непатогенными и имеют низкий профиль иммуногенности (High KA et al., «rAAV human trial experience» Methods Mol Biol. 2011; 807:429-57).

Одной из насущных целей исследований в области разработки эффективной генотерапии является кодон-оптимизация генов интереса в составе векторов для получения максимального уровня экспрессии генов интереса, что, в свою очередь, позволит использовать для достижения значимого эффекта более низкие дозы вектора.

Одним из свойств генетического кода является вырожденность - способность разных кодонов (тринуклеотидов) кодировать одну и ту же аминокислоту. Такие кодоны, которые дают одну и ту же аминокислоту в процессе трансляции, называются синонимичными. В природных последовательностях выбор одного из синонимичных кодонов осуществляется случайным образом в процессе эволюции, однако частоты использования синонимичных кодонов отличаются: для каждой аминокислоты есть более и менее предпочтительные. Кодон-оптимизация - это широко используемая в мире техника, направленная на повышение продуктивности наработки белковых молекул, которая заключается в рациональном сопоставлении каждой аминокислоты в белковой последовательности одного из подходящих синонимичных кодонов. Один из распространенных принципов кодон-оптимизации подразумевает использование наиболее частых кодонов, впоследствии были предложены и другие подходы, такие как гармонизация (воспроизведение распределения частот используемых кодонов), но и они не всегда дают увеличение продуктивности. Помимо частот кодонов на эффективность наработки может влиять GC-состав последовательности (отношение количества гуанинов и цитозинов к суммарной длине последовательности), в частности, было показано, что завышенный GC-состав ассоциирован с повышением количества мРНК в клетках млекопитающих Grzegorz Kudla ET AL., High Guanine and Cytosine Content Increases mRNA Levels in Mammalian Cells, June

2006, Volume 4, Issue 6, e180, pp. 933-942). Также стоит отметить, что устойчивые элементы вторичной структуры мРНК, т.е. имеющие низкую свободную энергию фолдинга, могут снижать эффективность.

Различные варианты кодон-оптимизации последовательности гена интереса могут приводить к следующему (в сравнение с геном дикого типа):

- а) уровень экспрессии генов интереса будет незначительно увеличен;
- б) уровень экспрессии генов интереса будет значительно увеличен;
- в) уровень экспрессии генов интереса останется приблизительно на том же уровне;
- г) уровень экспрессии генов интереса будет понижен.

Таким образом, есть потребность в получении кодон-оптимизированной последовательности гена SMN1 для увеличения экспрессии гена SMN1 в целевых клетках.

Было установлено, что кодон-оптимизированная последовательность SMN1 (SMN1-GeneBeam (или сокращенно SMN1-GB)), которая имеет нуклеотидную последовательность SEQ ID NO: 2, неожиданно увеличивает транскрипцию гена SMN1 более чем в 3 раза, то есть неожиданно увеличивает количество копий мРНК SMN1-GeneBeam в более чем 3 раза по сравнению с SMN1-WT (дикого типа), что в свою очередь приводит к значительному увеличению экспрессии гена SMN1 и, соответственно SMN белка.

Краткое описание изобретения

В одном из аспектов настоящее изобретение относится к выделенной кодон-оптимизированной нуклеиновой кислоте, которая кодирует белок SMN1 (белок выживаемости моторных нейронов) с SEQ ID NO: 1, и включает последовательность нуклеиновой кислоты SEQ ID NO: 2.

В одном из аспектов настоящее изобретение относится к экспрессионной кассете, которая включает вышеуказанную кодон-оптимизированную нуклеиновую кислоту.

В некоторых вариантах экспрессионная кассета включает следующие элементы в направлении от 5'-конца к 3'-концу:

- левый (первый) ITR (инвертированные концевые повторы);
- CMV (цитомегаловирусный) энхансер;
- CMV (цитомегаловирусный) промотер;
- интрон гена hBG1 (ген субъединицы гемоглобина гамма-1);
- вышеуказанную кодон-оптимизированную нуклеиновую кислоту гена SMN1;
- сигнал полиаденилирования hGH1 (сигнал полиаденилирования гена гормона роста человека);
- правый (второй) ITR.

В некоторых вариантах экспрессионная кассета включает нуклеиновую кислоту с последовательностью SEQ ID NO: 4.

В одном из аспектов настоящее изобретение относится к экспрессионному вектору, который включает вышеуказанную кодон-оптимизированную нуклеиновую кислоту или вышеуказанную кассету.

В одном из аспектов настоящее изобретение относится к рекомбинантному вирусу на основе AAV9 (аденоассоциированный вирус 9 серотипа) для увеличения экспрессии гена SMN1 в целевых клетках, который включает капсид и вышеуказанную экспрессионную кассету.

В некоторых вариантах рекомбинантный вирус на основе AAV9 имеет капсид, который включает белок VP1 AAV9.

В некоторых вариантах рекомбинантный вирус на основе AAV9 имеет капсид, который включает белок VP1 AAV9, имеющий аминокислотную последовательность SEQ ID NO: 5.

В некоторых вариантах рекомбинантный вирус на основе AAV9 имеет капсид, который включает белок VP1 AAV9, имеющий аминокислотную последовательность SEQ ID NO: 5 с одной или несколькими точечными мутациями.

В некоторых вариантах рекомбинантный вирус на основе AAV9 имеет капсид, который включает белок VP1 AAV9, имеющий аминокислотную последовательность SEQ ID NO: 5 или аминокислотную последовательность SEQ ID NO: 5 с одной или несколькими точечными мутациями, а экспрессионная кассета включает следующие элементы в направлении от 5'-конца к 3'-концу:

CMV энхансер;

CMV промотер;

интрон гена hBG1;

вышеуказанную кодон-оптимизированную нуклеиновую кислоту гена SMN1;

сигнал полиаденилирования hGH1;

правый ITR.

В некоторых вариантах рекомбинантный вирус на основе AAV9 имеет капсид, который включает белок VP1 AAV9, имеющий аминокислотную последовательность SEQ ID NO: 5 или аминокислотную последовательность SEQ ID NO: 5 с одной или несколькими точечными мутациями, а экспрессионная кассета включает нуклеиновую кислоту с SEQ ID NO: 4.

В одном из аспектов настоящее изобретение относится к фармацевтической композиции для доставки гена SMN1 в целевые клетки, которая включает вышеуказанный

рекомбинантный вирус на основе AAV9 в сочетании с одним или несколькими фармацевтически приемлемыми эксципиентами.

В одном из аспектов настоящее изобретение относится к применению вышеуказанного рекомбинантного вируса на основе AAV9 или вышеуказанной композиции для доставки гена SMN1 в целевые клетки.

Краткое описание чертежей

Фигура 1. Экспрессия SMN1 на уровне мРНК после трансфекции. Клетки HEK293 и HSMC были трансфецированы 5 мкг плазмид pAAV-SMN1-WT и pAAV-SMN1-GB (кодирующих ген SMN1 без кодон-оптимизации и с кодон-оптимизацией по алгоритму GeneBeam). Через 72 ч количество копий гена SMN1 в каждом образце было определено с помощью количественной ПЦР ($n = 3$). Также было определено количество копий гена домашнего хозяйства GAPDH. Все полученные количества для SMN1 были нормализованы на 10000 копий гена GAPDH в каждом образце. Представлены данные по нормализованному среднему количеству копий SMN1-WT, SMN-GB для обеих клеточных линий, с указанием стандартного отклонения. Также представлено соотношение нормализованного количества копий SMN1-GB и SMN1-WT в каждой линии.

Фигура 2. Экспрессия SMN1 на уровне белка после трансфекции. Клетки HSMC были трансфецированы 5 мкг плазмид pAAV-SMN1-WT и pAAV-SMN1-GB (кодирующих ген SMN1 без кодон-оптимизации и с кодон-оптимизацией по алгоритму GeneBeam). Через 72 ч клетки были покрашены первичными антителами к белку SMN1 и вторичными антителами, мечеными Alexa Fluor 488, в каждом образце ($n = 3$). Представлена средняя интенсивность флуоресцентного сигнала для живых клеток в образцах за вычетом фонового сигнала, полученного на клетках, окрашенных вторичными антителами без первичных антител, с указанием стандартного отклонения.

Фигура 3. Соотношение экспрессии SMN1 на уровне мРНК и белка после трансдукции. Клетки HSMC были трансдуцированы вирусами AAV9-SMN1-WT и AAV9-SMN1-GB в 3 независимых экспериментах, в каждом из которых эффективность трансдукции составила не менее 50% по контрольному GFP-содержащему вирусу. Экспрессия SMN1 была определена на уровне мРНК и белка (см. выше), после чего посчитано соотношение экспрессии SMN1-GB и SMN1-WT. Средние соотношения вместе со стандартными отклонениями представлены в фигуре.

Определения и общие методы

Если иное не определено в настоящем документе, научные и технические термины, используемые в связи с настоящим изобретением, будут иметь значения, которые обычно понятны специалистам в данной области.

Кроме того, если по контексту не требуется иное, термины в единственном числе включают в себя термины во множественном числе, и термины во множественном числе включают в себя термины в единственном числе. Как правило, используемая классификация и методы культивирования клеток, молекулярной биологии, иммунологии, микробиологии, генетики, аналитической химии, химии органического синтеза, медицинской и фармацевтической химии, а также гибридизации и химии белка и нуклеиновых кислот, описанные в настоящем документе, хорошо известны специалистам и широко применяются в данной области. Ферментативные реакции и способы очистки осуществляют в соответствии с инструкциями производителя, как это обычно осуществляется в данной области, или как описано в настоящем документе.

«Выделенный» означает измененный или удаленный из природного состояния. Например, нуклеиновая кислота или пептид, в природе присутствующие в животном, не являются «выделенными», но те же нуклеиновая кислота или пептид, частично или полностью отделенные от материалов, сопутствующих им в их природном состоянии, являются «выделенными». Выделенная нуклеиновая кислота или белок могут существовать, по существу, в очищенной форме или могут существовать в неприродном окружении, таком как, например, генетически модифицированной клетке.

Определения «встречающийся в природе», «нативный» или «дикого типа» используют для описания объекта, который можно обнаружить в природе как отличающийся от получаемого искусственно. Например, белок или нуклеотидная последовательность, присутствующие в организме (включая вирус), которые можно изолировать из источника в природе, и которые не модифицированы умышленно специалистом в лаборатории, являются встречающимися в природе.

Термин «геном» относится к полному генетическому материалу организма.

В настоящем описании и в последующей формуле изобретения, если контекстом не предусмотрено иное, слова «включать» и «содержать» или их вариации, такие как «имеющий», «включает», «включающий», «содержит» или «содержащий», следует понимать как включение указанного целого или группы целых, но не исключение любого другого целого или группы целых.

Белок (Пептид)

В настоящем описании, термины «пептид», «полипептид» и «белок» используют взаимозаменяемо, и они относятся к соединению, состоящему из аминокислотных остатков, ковалентно связанных пептидными связями. Белок или пептид должен содержать по меньшей мере две аминокислоты, и не существует ограничений по максимальному количеству аминокислот, которые может содержать последовательность белка или пептида. Полипептиды включают любой пептид или белок, содержащий две или более аминокислот, соединенных друг с другом пептидными связями. Как применяют в настоящем описании, термин относится и к коротким цепям, также общепринято обозначаемым в этой области, например, как пептиды, олигопептиды и олигомеры, и к более длинным цепям, как правило, обозначаемым в этой области как белки, множество типов которых существует. «Полипептиды» включают, помимо прочего, например, биологически активные фрагменты, по существу, гомологичные полипептиды, олигопептиды, гомодимеры, гетеродимеры, варианты полипептидов, модифицированные полипептиды, производные, аналоги, слитные белки. Полипептиды включают природные пептиды, рекомбинантные пептиды, синтетические пептиды или их комбинацию.

Молекулы нуклеиновых кислот

Термины «нуклеиновая кислота», «нуклеиновая последовательность» или «нуклеиновокислотная последовательность», «полинуклеотид», «олигонуклеотид», «полинуклеотидная последовательность» и «нуклеотидная последовательность», которые используются равнозначно в данном описании, обозначают четкую последовательность нуклеотидов, модифицированных или не модифицированных, определяющую фрагмент или участок нуклеиновой кислоты, содержащую или не содержащую неприродные нуклеотиды и являющуюся либо двухцепочечной ДНК или РНК, либо одноцепочечной ДНК или РНК, либо продуктами транскрипции указанных ДНК.

Специалист в этой области имеет общие знания о том, что нуклеиновые кислоты являются полинуклеотидами, которые можно гидролизовать до мономерных «нуклеотидов». Мономерные нуклеотиды можно гидролизовать в нуклеозиды. Как применяют в настоящем описании, полинуклеотиды включают, в качестве неограничивающих примеров, все последовательности нуклеиновой кислоты, получаемые любыми способами, доступными в этой области, включая, в качестве неограничивающих примеров, рекомбинантные способы, т.е. клонирование последовательностей нуклеиновой кислоты из рекомбинантной библиотеки или генома клетки, использование обычной технологии клонирования и ПЦР и т.п., и способами синтеза.

Здесь также следует упомянуть, что данное изобретение не относится к нуклеотидным последовательностям в их природной хромосомной среде, т.е. в природном состоянии. Последовательности данного изобретения были выделены и/или очищены, т.е. были взяты прямо или косвенно, например, путем копирования, при этом их среда была по меньшей мере частично модифицирована. Таким образом, также здесь следует подразумевать изолированные нуклеиновые кислоты, полученные путем генетической рекомбинации, например, с помощью принимающих клеток (клеток-хозяев), или полученные путем химического синтеза.

«Выделенная» молекула нуклеиновой кислоты представляет собой молекулу нуклеиновой кислоты, которая идентифицирована и отделена от по меньшей мере одной молекулы нуклеиновой кислоты-примеси, с которой она обычно связана в естественном источнике нуклеиновой кислоты нуклеазы. Выделенная молекула нуклеиновой кислоты отличается от той формы или набора, в которых она находится в естественных условиях. Таким образом, выделенная молекула нуклеиновой кислоты отличается от молекулы нуклеиновой кислоты, существующей в клетках в естественных условиях. Однако выделенная молекула нуклеиновой кислоты включает молекулу нуклеиновой кислоты, находящуюся в клетках, в которых в норме происходит экспрессия нуклеазы, например, в случае, если молекула нуклеиновой кислоты имеет локализацию в хромосоме, отличную от ее локализации в клетках в естественных условиях.

Термин нуклеотидная последовательность охватывает его комплемент, если не указано иное. Таким образом, нуклеиновую кислоту, имеющую определенную последовательность следует понимать как охватывающие ее комплементарную цепь с ее комплементарной последовательностью.

Термины «трансформация», «трансфекция», «трансдукция» относятся к любому способу или средствам, с помощью которых нуклеиновая кислота вводится в клетку или организм-хозяин, и могут быть использованы взаимозаменяемо для передачи аналогичного значения. Такие способы включают в себя без ограничения трансфекцию, электропорацию, микроинъекции, инфицирование, ПЭГ-сплавление и тому подобное.

Аденоассоциированный вирус (AAV)

Вирусы семейства Parvoviridae представляют собой небольшие ДНК-содержащие вирусы животных. Семейство Parvoviridae может быть разделено на два подсемейства: Parvovirinae, представители которого инфицируют позвоночных животных, и Densovirinae, представители которого инфицируют насекомых. К 2006 году были описаны 11 серотипов аденоассоциированного вируса (Mori, S. ET AL., 2004, «Two novel adeno-associated viruses

from cynomolgus monkey: pseudotyping characterization of capsid protein», *Virology*, Т. 330 (2): 375-83). Все известные серотипы могут инфицировать клетки многих видов тканей. Тканевая специфичность определяется серотипом белков капсида, поэтому векторы на основе аденоассоциированного вируса конструируют, задавая необходимый серотип. Дополнительная информация по парвовирусам и другим представителям Parvoviridae описана в литературе (Kenneth I. Berns, «Parvoviridae: The Viruses and Their Replication», Chapter 69 in *Fields Virology* (3d Ed. 1996)).

Геномная организация всех известных серотипов AAV очень сходна. Геном AAV представляет собой линейную одноцепочечную молекулу ДНК, которая содержит менее чем примерно 5000 нуклеотидов (нт) в длину. Инвертированные концевые повторы (ITR) фланкируют уникальные кодирующие нуклеотидные последовательности репликации неструктурных белков (Rep) и структурных белков (Cap). Ген Cap кодирует белки VP (VP1, VP2 и VP3), которые образуют капсид. Концевые 145 нуклеотидов являются самокомплементарными и организованы таким образом, что может быть сформирован энергетически стабильный внутримолекулярный дуплекс, образующий Т-образную шпильчатую структуру. Такие шпильчатые структуры функционируют как точки начала репликации ДНК вируса, являясь праймерами для клеточного ДНК-полимеразного комплекса. После инфекции клеток млекопитающих AAV дикого типа (wtAAV) гены Rep (например, Rep78 и Rep52) экспрессируются с помощью P5 промотора и P19 промотора, соответственно, и оба белка Rep выполняют определенную функцию в репликации генома вируса. Сплайсинг в открытой рамке считывания Rep (Rep ORF) приводит к экспрессии фактически четырех белков Rep (например, Rep78, Rep68, Rep52 и Rep40). Однако было показано, что несплайсированная мРНК, кодирующая белки Rep78 и Rep52, является достаточной для продукции вектора AAV в клетках млекопитающих.

Вектор

Термин «вектор» при использовании в настоящем документе означает молекулу нуклеиновой кислоты, способную транспортировать другую нуклеиновую кислоту, с которой она соединена.

Термины «инфекционная единица» (ие), «инфекционная частица» или «репликационная единица», как используется в отношении вирусного титра, относятся к числу инфекционных частиц рекомбинантного вектора AAV, которое измеряют посредством анализа инфекционных центров, также известного как анализ репликационных центров, описанный, например, в публикации McLaughlin et al., *J. Virol.* (1988) 62:1963-1973.

Термин «гетерологичный», когда он относится к последовательностям нуклеиновых кислот, таким как кодирующие последовательности и последовательности регуляции, обозначает последовательности, которые обычно не соединены вместе и/или обычно не связаны с конкретной клеткой. Таким образом, «гетерологичная» область конструкции нуклеиновой кислоты или вектора представляет собой фрагмент нуклеиновой кислоты, расположенный внутри или присоединенный к другой молекуле нуклеиновой кислоты, которая в природе не найдена совместно с другой молекулой. Например, гетерологичная область конструкции нуклеиновой кислоты может содержать кодирующую последовательность, фланкированную последовательностями, которые в природе не найдены совместно с кодирующей последовательностью. Другой пример гетерологичной кодирующей последовательности представляет собой конструкцию, где сама кодирующая последовательность не найдена в природе (например, синтетические последовательности, которые содержат кодоны, отличные от нативного гена).

Как применяют в настоящем описании, термин «экспрессия» определяют как транскрипцию и/или трансляцию конкретной нуклеотидной последовательности, запускаемую ее промотором.

Применение

«Генная терапия» представляет собой вставку генов в клетки и/или тканей субъекта для лечения заболевания, обычно, наследственных заболеваний, при этом дефектный мутантный аллель заменяется или дополняется функциональным аллелем.

«Лечить», «лечение» и «терапия» относятся к методу смягчения или устранения биологического расстройства и/или по меньшей мере одного из сопутствующих ему симптомов. Используемый в данном документе, термин «облегчить» болезнь, заболевание или состояние, означает уменьшение тяжести и/или частоты возникновения симптомов заболевания, расстройства или состояния. Кроме того, содержащиеся в данном документе ссылки на «лечение» включают ссылки на лечебную, паллиативную и профилактическую терапию.

В одном аспекте субъект лечения или пациент является млекопитающим, предпочтительно человеческим субъектом. Вышеупомянутый субъект может быть мужского или женского пола любого возраста.

Термин «нарушение» означает любое состояние, которое можно улучшить в результате лечения по настоящему изобретению. В определение данного термина входят хронические и острые нарушения или заболевания, включающие в себя патологические

состояния, которые вызывают предрасположенность млекопитающего к возникновению данного нарушения.

«Заболевание» является состоянием здоровья животного, где животное не может поддерживать гомеостаз, и где, если заболевание не облегчают, то здоровье животного продолжает ухудшаться.

Термин «субъект», «пациент», «индивидуум» и т.п. используют в настоящем описании взаимозаменяемо, и они относятся к любому животному, поддающемуся воздействию способами, представленными в настоящем описании. В конкретных неограничивающих вариантах осуществления субъект, пациент или индивидуум является человеком.

«Терапевтически эффективным количеством» считается количество вводимого в процессе лечения терапевтического агента, которое избавит в определенной степени от одного или нескольких симптомов заболевания, по поводу которого проводится лечение.

Подробное описание изобретения

Кодон-оптимизированная нуклеиновой кислоте

В одном из аспектов настоящее изобретение относится к выделенной кодон-оптимизированной нуклеиновой кислоте, которая кодирует белок SMN1 (белок выживаемости моторных нейронов) с SEQ ID NO: 1, и включает последовательность нуклеиновой кислоты SEQ ID NO: 2.

Для получения кодон-оптимизированного гена SMN1 за основу была взята соответствующая аминокислотная последовательность белка SMN_HUMAN: MAMSSGGSGGGVPEQEDSVLFRRGTGQSDSDIWDDTALIKAYDKAVASFKHALKNG DICETSGPKKTTPKRKPAKKNKSQKKNTAASLQQWKVGDKCSAIWSEDGCIYPATIASI DFKRETCVVVYTYGNGNREEQNLSDLLSPICEVANNIEQNAQENENESQVSTDESENSRSP GNKSDNIKPKSAPWNSFLPPPPMPGPRLGPGKPKLKFNGPPPPPPPPPHLLSCWLPPFPS GPPIPPPPPICPDSLDDADALGSMLISWMSGYHTGYMGFRQNQKEGRCSHSLN (SEQ ID NO:1)

Данная аминокислотная последовательность SEQ ID NO:1 переводилась в нуклеотидную последовательность путем последовательного сопоставления каждой аминокислоте начиная с N-конца одного из кодирующих её синонимичных кодонов.

Подробно по кодон-оптимизированного гена SMN1 и отбор финальной последовательности описано в примере 1.

Финальная кодон-оптимизированная последовательность SMN1 (SMN1-GeneBeam)

имеет следующую нуклеотидную последовательность:

ATGGCCATGAGCAGCGGCGGCAGCGGCGGCGGCGTGCCTGAGCAAGAGGAC
AGCGTGCTGTTTCAGAAGAGGCACCGGCCAGAGCGACGACAGCGACATCTGGGACGA
CACCGCCCTGATCAAGGCCTACGACAAGGCCGTGGCCAGCTTCAAGCACGCCCTGA
AGAACGGCGACATCTGCGAGACCAGCGGCAAGCCCAAGACCACCCCCAAGAGAAA
GCCCCGCAAGAAGAACAAGAGCCAGAAGAAGAACACCGCCGCCAGCCTGCAGCAG
TGGAAGGTGGGCGACAAGTGCAGCGCCATCTGGAGCGAGGACGGCTGCATCTACCC
CGCCACCATCGCCAGCATCGACTTCAAGAGAGAGACCTGCGTGGTGGTGTACACCG
GCTACGGCAACAGAGAGGAGCAGAACCTGAGCGACCTGCTGAGCCCCATCTGCGAG
GTGGCCAACAACATCGAGCAGAACGCCCAAGAGAACGAGAACGAGAGCCAAGTGA
GCACCGACGAGAGCGAGAACAGCAGAAGCCCCGGCAACAAGAGCGACAACATCAA
GCCCAAGAGCGCCCCCTGGAACAGCTTCCTGCCCCCTCCCCCCCCCTATGCCCGGCCC
TAGACTGGGCCCTGGCAAGCCTGGCCTGAAGTTCAACGGCCCCCCCCCCCCCTCCTCC
TCCTCCTCCTCCTCACCTGCTGAGCTGCTGGCTGCCCCCTTCCCCAGCGGCCCTCCT
ATCATCCCTCCTCCCCCCCCCATCTGCCCCGACAGCCTGGACGACGCCGACGCCCTG
GGCAGCATGCTGATCAGCTGGTACATGAGCGGCTACCACACCGGCTACTACATGGG
CTTCAGACAGAACCAGAAGGAGGGCCGGTGCAGCCACAGCCTGAACTAG (SEQ ID
NO:2).

Данная финальная кодон-оптимизированная нуклеотидная последовательность SMN1 (SMN1-GeneBeam) характеризуется повышенным индексом адаптации кодонов (стандартная метрика для оценки последовательности на предмет частот использованных кодонов) по сравнению с кодирующей последовательностью гена SMN дикого типа (SMN1-WT):

ATGGCGATGAGCAGCGGCGGCAGTGGTGGCGGCGTCCCGGAGCAGGAGGATTCCGT
GCTGTTCCGGCGCGGCACAGGCCAGAGCGATGATTCTGACATTTGGGATGATACAG
CACTGATAAAAGCATATGATAAAGCTGTGGCTTCATTTAAGCATGCTCTAAAGAATG
GTGACATTTGTGAAACTTCGGGTAAACCAAAAACCACACCTAAAAGAAAACCTGCT
AAGAAGAATAAAAGCCAAAAGAAGAATACTGCAGCTTCCTTACAACAGTGGAAAGT
TGGGGACAAATGTTCTGCCATTTGGTCAGAAGACGGTTGCATTTACCCAGCTACCAT
TGCTTCAATTGATTTTAAGAGAGAAACCTGTGTTGTGGTTTACACTGGATATGGAAA
TAGAGAGGAGCAAAATCTGTCCGATCTACTTTCCCCAATCTGTGAAGTAGCTAATAA
TATAGAACAAAATGCTCAAGAGAATGAAAATGAAAGCCAAGTTTCAACAGATGAAA
GTGAGAACTCCAGGTCTCCTGGAATAAATCAGATAACATCAAGCCCAAATCTGCT
CCATGGAACTCTTTTCTCCCTCCACCACCCCCCATGCCAGGGCCAAGACTGGGACCA

GGAAAGCCAGGTCTAAAATTCAATGGCCCCACCCGCCACCGCCACCACCACCACC
CCACTTACTATCATGCTGGCTGCCTCCATTTCTTCTGGACCACCAATAATTCCCCCA
CCACCTCCCATATGTCCAGATTCTCTTGATGATGCTGATGCTTTGGGAAGTATGTAA
TTTCATGGTACATGAGTGGCTATCATACTGGCTATTATATGGGTTTCAGACAAAATC
AAAAAGAAGGAAGGTGCTCACATTCCTTAAATTAA (SEQ ID NO:3).

Индекс адаптации кодонов для финальной кодон-оптимизированной нуклеотидной последовательности гена SMN1 (SEQ ID NO:2) нашей последовательности равен 98%, для последовательности дикого типа - 75%.

GC-состав последовательности дикого типа равен 45%, то есть отличается от целевого значения на 15%, GC-состав финальной кодон-оптимизированной нуклеотидной последовательности гена SMN1 (SEQ ID NO:2) оптимизированной последовательности равен 64%, т.е. отличается от целевого значения на 4%.

Финальная кодон-оптимизированная нуклеотидная последовательность гена SMN1 (SEQ ID NO:2) и нуклеотидная последовательность гена SMN1 дикого типа (SEQ ID NO:3) Последовательности идентичны на 71%.

Экспрессионная кассета. Экспрессионный вектор.

В одном из аспектов настоящее изобретение относится к экспрессионной кассете, которая включает вышеуказанную кодон-оптимизированную нуклеиновую кислоту.

Термин «экспрессионная кассета» при использовании в данном документе, в частности, относится к фрагменту ДНК, который способен в соответствующей обстановке запускать экспрессию полинуклеотида, кодирующего представляющий интерес полипептид, который включен в указанную экспрессионную кассету. При введении в клетку-хозяина экспрессионная кассета помимо прочего способна задействовать клеточные механизмы для транскрипции полинуклеотида, кодирующего представляющий интерес полипептид, в РНК, которая затем обычно дополнительно процессируется и, наконец, транслируется в представляющий интерес полипептид. Экспрессионная кассета может содержаться в экспрессионном векторе.

Экспрессионная кассета по настоящему изобретению содержит в качестве элемента промотор. Термин «промотор», используемый в настоящем документе, в частности, относится к элементу ДНК, который способствует транскрипции полинуклеотида, с которым функционально связан промотор. Промотор может также составлять часть элемента «промотор/энхансер». Хотя физические границы между элементами «промотор» и «энхансер» не всегда ясны, термин «промотор» обычно относится к месту на молекуле нуклеиновой кислоты, с которым связывается РНК-полимераза и/или связанные с ней

факторы, и с которого иницируется транскрипция. Эnhансеры усиливают активность промотора во времени, а также пространственно. В данной области известно множество промоторов, которые транскрипционно активны в широком диапазоне типов клеток. Промоторы могут быть разделены на два класса: на тех, которые функционируют конститутивно, и тех, которые регулируются индукцией или снятием репрессии. Для экспрессии белка пригодны оба класса. Промоторы, которые используются для продукции высокого уровня полипептидов в эукариотических клетках и, в частности, в клетках млекопитающих, должны быть сильными и, предпочтительно, должны быть активными в широком диапазоне типов клеток. Сильные конститутивные промоторы, которые способны запускать экспрессию во многих типах клеток, хорошо известны в данной области и, поэтому, нет необходимости в их подробном описании в данном документе. В соответствии с идеей настоящего изобретения - предпочтительно использовать промотор цитомегаловируса (CMV). Промотор или промотор/энхансер, полученные из немедленной ранней (IE) области цитомегаловируса (hCMV) человека, в особенности подходят в качестве промотора в экспрессионной кассете по настоящему изобретению. Немедленная ранняя (IE) область цитомегаловируса (hCMV) человека и полученные из нее функциональные запускающие экспрессию фрагменты и/или функциональные усиливающие экспрессию фрагменты, например, описаны в EP0173177 и EP0323997, а также хорошо известны в данной области. Таким образом, несколько фрагментов немедленной ранней (IE) области hCMV могут использоваться в качестве промотора и/или промотора/энхансера. Согласно одному варианту осуществления изобретения промотор CMV человека используется в экспрессионной кассете по настоящему изобретению.

В некоторых вариантах экспрессионная кассета включает следующие элементы в направлении от 5'-конца к 3'-концу:

- левый (первый) ITR (инвертированные концевые повторы);
- CMV (цитомегаловирусный) энхансер;
- CMV (цитомегаловирусный) промотор;
- интрон гена hBG1 (ген субъединицы гемоглобина гамма-1);
- вышеуказанную кодон-оптимизированную нуклеиновую кислоту гена SMN1;
- сигнал полиаденилирования hGH1 (сигнал полиаденилирования гена гормона роста человека);
- правый (второй) ITR.

В некоторых вариантах левый (первый) ITR (инвертированные концевые повторы) имеет следующую последовательность нуклеиновой кислоты:

Cctgcaggcagctgcgcctcgtcgtcactgaggccgcccggcgctggggcgacctttggtcgcccggcctcagtgagcgagcgagcgcgagagagggagtgccaaactccatcactaggggttct (SEQ ID NO: 8).

В некоторых вариантах CMV (цитомегаловирусный) энхансер имеет следующую последовательность нуклеиновой кислоты:

cgttacataacttacggtaaatggcccgctggctgaccgccaacgacccccgccattgacgtcaataatgacgtatgttcccatagtaacgCcaatagggactttccattgacgtcaatgggtggagtatttacggtaaaactgccacttggcagttacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgctggcattatgccagttacatgaccttatgggactttcctacttggcagttacatctacgtattagtcacgtattaccatg (SEQ ID NO: 9).

В некоторых вариантах CMV (цитомегаловирусный) промотор имеет следующую последовательность нуклеиновой кислоты:

gtgatgcggttttggcagttacatcaatggcggtggatagcgggttgactcacggggatttccaagtctccacccattgacgtcaatgggagttgttttGcaccaaaatcaacgggactttccaaaatgtcgtgaacaactccgccccattgacgcaaatgggcggtaggcgtgtacgggtgggaggtctatataagcagagct (SEQ ID NO: 10).

В некоторых вариантах интрон гена hBG1 (ген субъединицы гемоглобина гамма-1) имеет следующую последовательность нуклеиновой кислоты:

cgaatccccggccgggaacgggtgcattggaacgcggattccccgtgccaagagtacgtaagtaccgcctatagagtctataggccacaaaatgctttcttttaataatactttttgtttatctatttctaatactttccctaactctttttcagggaataatgatacaatgtatcatgcctctttgaccattctaaagaataacagtgataatttctgggttaaggcaatagcaatatttctcatataaatatttctcatataaattgtaactgatgtaaagggtttcatattgctaatagcagctacaatccagctaccattctgctttatfttatggttgggataaggctggattattctgagccaagctaggccctttgctaatactgttcatactcttctctccacagctcctgggcaacgtgctggtctgtgtgctggcccatcactttggcaagaattgggat (SEQ ID NO: 11).

В некоторых вариантах сигнал полиаденилирования hGH1 (сигнал полиаденилирования гена гормона роста человека) имеет следующую последовательность нуклеиновой кислоты:

Acgggtggcatccctgtgacccctccccagtgccctcctggccctggaagttgccactccagtgcccaccagcctgtcctaataaaataagttgcatcattttgtctgactaggtgtccttctataatattatgggtggaggggggtgtatggagcaaggggcaagttgggaagacaacctgtagggcctgcgggggtctattgggaaccaagctggagtgagtgacagtcacaaacttggctcactgcaatctccgcctcctgggttcaggcattctcctcagcctcccagttgttgggattccaggcatgaccaggctcagctaattttgtttttgtagagacggggttcaccatattggccaggctggtctccaactcctaactcaggtgatctaccaccttggcctccaaattgctgggattacaggcgtgaaccactgtcccttccctgtcctt (SEQ ID NO: 12).

В некоторых вариантах правый (второй) ITR имеет следующую последовательность нуклеиновой кислоты:

aggaaccctagtgatggagttggccactccctctctgcgcctcgtcgtcactgaggccgggacgacaaaggtgcccagcggcggtttgcccggggcctcagtgagcgagcgagcgcgagctgcctgcagg (SEQ ID NO: 13).

В некоторых вариантах экспрессионная кассета имеет следующую последовательность нуклеиновой кислоты:

ctgcaggcagctgcgcgctcgtcgtcactgaggccgcccggcgctggcgaccttggtcgcccggcctcagtgagcgagcgag
cgcgagagagggagtgcccaactccatcactaggggttctgcggccgcacgcgtctagttattaatagtaatcaattacggggcattag
ttcatagcccatafatggagttccggttacataacttacggtaaatggcccgcctggctgaccgccaacgacccccgcccattgacgtca
ataatgacgtatgttcccatagtaacgCcaatagggactttccattgacgtcaatgggtggagtatttacggtaaaactcccacttggcagtac
atcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgccagttacatgaccttatggg
actttctacttggcagttacatctacgtattagtcacgtattaccatgggtgatgcgggttttggcagttacatcaatgggcgtggatagcggttg
actcacggggatttccaagtctcacccttattgacgtcaatgggagttgttttGcaccaaaatcaacgggactttccaaaatgtcgaaca
actccgccccattgacgcaaatggcggttaggcgtgtacgggtggaggtctatataagcagagctcgtttagtgaaccgtcagatcgctg
gagacgccatccacgtgtttgacctccatagaagacaccgggaccgatccagcctccgcggttcgaatcccgccgggaacgggtgc
attggaacgcggattccccgtgccaagagtacgtaagtaccgcctatagagtctataggccccacaaaaatgtttctcttttaataatacttt
ttgtttatcttatttctaatactttccctaactcttctttcagggaataatgatacaatgtatcatgcctctttgcaccattctaaagaataacagt
ataatttctgggttaaggcaatagcaatatttctgcatataaatatttctgcatataaattgtaactgatgtaagaggtttcatattgctaatagcac
tacaatccagctaccattctgctttattttatgggtgggataaggctggattattctgagtcgaagctaggcccttttctaatacatgttcatactc
ttatcttctccccacagctcctgggcaacgtgctgtgtgtgtggtggccatcactttggcaagaattgggattcgaacatCGATTGT
AATTCATGAGCCACCATGGCCATGAGCAGCGGCGGCAGCGGCGGCGGCGTGCCTGA
GCAAGAGGACAGCGTGCTGTTTCTCAGAAGAGGACACCGGCCAGAGCGACGACAGCGAC
ATCTGGGACGACACCGCCCTGATCAAGGCCTACGACAAGGCCGTGGCCAGCTTCAA
GCACGCCCTGAAGAACGGCGACATCTGCGAGACCAGCGGCAAGCCCAAGACCACCC
CCAAGAGAAAGCCCGCCAAGAAGAACAAGAGCCAGAAGAAGAACACCGCCGCCAG
CCTGCAGCAGTGGAAGGTGGGCGACAAGTGCAGCGCCATCTGGAGCGAGGACGGCT
GCATCTACCCCGCCACCATCGCCAGCATCGACTTCAAGAGAGAGACCTGCGTGGTG
GTGTACACCGGCTACGGCAACAGAGAGGAGCAGAACCTGAGCGACCTGCTGAGCCC
CATCTGCGAGGTGGCCAACAACATCGAGCAGAACGCCCAAGAGAACGAGAACGAG
AGCCAAGTGAGACCGACGAGAGCGAGAACAGCAGAACGCCCCGGAACAAGAGCG
ACAACATCAAGCCCAAGAGCGCCCCCTGGAACAGCTTCCTGCCCCCTCCCCCCCCTA
TGCCCCGGCCCTAGACTGGGCCCTGGCAAGCCTGGCCTGAAGTTCAACGGCCCCCCCC
CCCCCTCCTCCTCCTCCTCCTCCTCACCTGCTGAGCTGCTGGCTGCCCCCCTTCCCCAG
CGGCCCTCCTATCATCCCTCCTCCCCCCCCCATCTGCCCCGACAGCCTGGACGACGC
CGACGCCCTGGGCAGCATGCTGATCAGCTGGTACATGAGCGGCTACCACACCGGCT
ACTACATGGGCTTCAGACAGAACCAGAAGGAGGGCCGGTGCAGCCACAGCCTGAAC
TGATctagagtcgacctgcagaagcttgctcgcagcgcgtgctcgagagatctacgggtggcatccctgtgaccttccccagtc
ctctcctggccctggaagttgccactccagtgcccaccagcctgtcctaataaaatgaagttgcatcattttgtctgactaggtgtccttataa
tattatggggtggaggggggtggtatggagcaaggggcaagttgggaagacaacctgtagggcctcggggtctattgggaaccaagct

ggagtgcagtggcacaatcttggctcactgcaatctccgcctcctgggttcaagcgattctcctgcctcagcctcccgagttgttgggattcca
ggcatgcatgaccaggctcagctaattttgttttttggtagagacggggtttaccatattggccaggctggctccaactcctaattcaggt
gatctaccacaccttggcctccaaattgctgggattacaggcgtgaaccactgctcccttccctgtccttctgatttttaggtaaccacgtgcg
gaccgagcgcccgaggaacccctagtgatggagtggccactccctctctgcgcgctcgctcgctcactgaggccgggcgaccaaaagg
tcggccgacgccccgggctttgccccgggcgccctcagtgcgcgagcgagcgcgagctgcctgcagg (SEQ ID NO: 4).

В одном из аспектов настоящее изобретение относится к экспрессионному вектору, который включает вышеуказанную кодон-оптимизированную нуклеиновую кислоту или вышеуказанную экспрессионную кассету.

В некоторых вариантах осуществления изобретения вектор представляет собой плазмиду, т.е. кольцевую двухцепочечную часть ДНК, в которую могут быть лигированы дополнительные сегменты ДНК.

В некоторых вариантах осуществления изобретения вектор представляет собой вирусный вектор, в котором дополнительные сегменты ДНК могут быть лигированы в вирусный геном.

В некоторых вариантах осуществления изобретения векторы способны к автономной репликации в клетке-хозяине, в которую они введены (например, бактериальные векторы, имеющие бактериальный сайт инициации репликации и эписомные векторы млекопитающих). В других вариантах осуществления изобретения векторы (например, неэписомальные векторы млекопитающих) могут быть интегрированы в геном клетки-хозяина при введении в клетку-хозяина, и таким образом реплицируются вместе с геном хозяина. Более того, некоторые векторы способны направлять экспрессию генов, с которыми они функционально соединены. Такие векторы упоминаются в данном документе как «рекомбинантные экспрессирующие векторы» (или просто «экспрессирующие векторы» («вектор экспрессии» или «экспрессионный вектор»)).

Экспрессионные векторы включают плазмиды, ретровирусы, аденовирусы, аденоассоциированные вирусы (AAV), вирусы растений, такие как вирус мозаики цветной капусты, вирусы табачной мозаики, космиды, YAC, EBV полученные эписомы и тому подобное. Молекулы ДНК могут быть лигированы в вектор таким образом, что последовательности, контролирующие транскрипцию и трансляцию в векторе, выполняют предусмотренную функцию регуляции транскрипции и трансляции ДНК. Экспрессионный вектор и последовательности контроля экспрессии могут быть выбраны таким образом, чтобы быть совместимыми с используемой экспрессирующей клеткой-хозяином. Молекулы ДНК могут быть введены в экспрессионный вектор стандартными способами (например, лигированием комплементарных сайтов рестрикции или лигированием тупых концов, если сайты рестрикции отсутствуют).

Рекомбинантный экспрессионный вектор также может кодировать сигнальный пептид, который облегчает выработку белка-интереса клеткой-хозяином. Ген белка-интереса может быть клонирован в вектор таким образом, что сигнальный пептид соединен с рамкой считывания аминоконца белка-интереса. Сигнальным пептидом может быть сигнальный пептид иммуноглобулина или гетерологичный сигнальный пептид (то есть, сигнальный пептид белка не иммуноглобулиновой природы).

Помимо гена SMN1-GB по данному изобретению, рекомбинантная экспрессия векторов по данному изобретению может нести регулирующие последовательности, которые контролируют экспрессию гена SMN1-GB в клетке-хозяине. Специалистам в этой области будет понятно, что дизайн экспрессионного вектора, включая выбор регулирующих последовательностей, может зависеть от таких факторов, как селекция клетки-хозяина для трансформации, уровень экспрессии желаемого белка, и т.д. Предпочтительные регулирующие последовательности для экспрессирующей клетки-хозяина млекопитающих включают вирусные элементы обеспечивающие высокий уровень экспрессии белков в клетках млекопитающих, таких как промоторы и/или энхансеры, полученные из ретровирусной LTR, цитомегаловируса (CMV) (например, CMV промотора/энхансера), обезьяньего вируса 40 (SV40) (например, SV40 промотора/энхансера), аденовируса, (например, большого позднего промотора аденовируса (AdMLP)), вирус полиомы, а также сильных промоторов млекопитающих, таких как промотор нативных иммуноглобулинов или промотор актина.

Выражение «контролирующие последовательности» относится к последовательностям ДНК, необходимым для экспрессии функционально связанной кодирующей последовательности в определенном организме-хозяине. Пригодные для прокариот контролирующие последовательности представляют собой, например, промотор, необязательно оператор и сайт связывания рибосомы. Как известно, в эукариотических клетках присутствуют промоторы, сигналы полиаденилирования и энхансеры.

В контексте настоящего описания термин «промотор» или «регуляторная последовательность транскрипции» или «регуляторная последовательность» относится к фрагменту нуклеиновой кислоты, который контролирует транскрипцию одной или нескольких кодирующих последовательностей, и который расположен против направления считывания информации относительно направления транскрипции от сайта инициации транскрипции кодирующей последовательности, а также который структурно идентифицируется по наличию сайта связывания для ДНК-зависимой РНК-полимеразы, сайтов инициации транскрипции и других последовательностей ДНК, включающих, без

ограничения, сайты связывания фактора транскрипции, сайты связывания репрессора и активатора белка, а также любые другие последовательности нуклеотидов, известные специалистам в данной области, которые непосредственно или опосредованно регулируют уровень транскрипции с данным промотором. «Конститутивный» промотор представляет собой такой промотор, который активен в большинстве тканей в обычных физиологических условиях и условиях развития. «Индукцибельный» промотор представляет собой промотор, который подвергается физиологической регуляции или регуляции в ходе развития, например, при воздействии химического индуктора. «Тканеспецифичный» промотор активен только в конкретных типах тканей или клеток.

Термины «энхансеры» или «энхансер», используемые в изобретении, могут относиться к последовательности ДНК, которая расположена как смежная с последовательностью ДНК, кодирующей рекомбинантный продукт. Энхансерные элементы обычно расположены в 5'-направлении от промоторного элемента или могут быть расположены ниже или в пределах кодирующей последовательности ДНК (например, последовательности ДНК, транскрибированной или транслированной в рекомбинантный продукт или продукты). Таким образом, энхансерный элемент может быть расположен на расстоянии 100 пар оснований, 200 пар оснований или 300 или больше пар оснований перед последовательностью ДНК, которая кодирует рекомбинантный продукт, или после этой последовательности. Энхансерные элементы могут увеличивать количество экспрессируемого рекомбинантного продукта от последовательности ДНК, превышая экспрессию, обусловленную одиночным промоторным элементом. Специалистам в данной области техники доступно множество энхансерных элементов.

В дополнение к вышеуказанным генам и регуливающим последовательностям, рекомбинантные векторы экспрессии изобретения могут нести дополнительные последовательности, такие как последовательности, которые регулируют репликацию вектора в клетках-хозяевах (например, точки начала репликации) и гены селективируемого маркера. Ген селективируемого маркера облегчает селекцию клеток-хозяев, в которые был введен вектор (см., например, патенты США 4,399,216, 4,634,665 и 5,179,017). Например, обычно ген селективируемого маркера придает устойчивость к лекарственным средствам, таким как G418, гигромицин или метотрексат, клетке-хозяину, в которую вектор введен. Например, гены селективируемого маркера включают ген дигидрофолат редуктазы (DHFR) (для использования в dhfr-клетках-хозяевах при селекции/амплификации метотрексата), ген нео (для селекции G418) и ген синтетазы глутамата.

Термин «последовательность контроля экспрессии», используемый в данном описании, означает полинуклеотидные последовательности, которые необходимы для

воздействия на экспрессию и процессинг кодирующих последовательностей, к которым они лигированы. Контролирующие экспрессию последовательности включают соответствующие последовательности инициации транскрипции, терминации, промотора и энхансера; эффективные сигналы процессинга РНК, такие как сплайсинг и сигналы полиаденилирования; последовательности, которые стабилизируют цитоплазматическую мРНК; последовательности, которые повышают эффективность трансляции (т.е. консенсусная последовательность Козака); последовательности, которые повышают стабильность белка; и, при желании, последовательности, которые усиливают секрецию белка. Характер таких контролирующих последовательностей различается в зависимости от организма-хозяина; в прокариотах такие контролирующие последовательности, как правило, включают промотор, сайт связывания рибосомы, а также последовательности терминации транскрипции; в эукариотах, как правило, такие контролирующие последовательности включают промоторы и последовательности терминации транскрипции. Термин «контролирующие последовательности» включает, как минимум, все компоненты, наличие которых имеет важное значение для экспрессии и процессинга, и может также включать дополнительные компоненты, чье присутствие является полезным, например, лидирующие последовательности и последовательности слившихся клеток.

В контексте настоящего описания термин «функционально связанный» относится к связи полинуклеотидных (или полипептидных) элементов в функциональную связь. Нуклеиновая кислота является «функционально связанной», если она находится в условиях функциональной связи с другой последовательностью нуклеиновой кислоты. Например, регуляторная последовательность транскрипции функционально связана с кодирующей последовательностью, если она влияет на транскрипцию указанной кодирующей последовательности. Термин «функционально связанный» означает, что связанные последовательности ДНК являются, как правило, непрерывными, и при необходимости соединения двух участков, кодирующих белок, являются также непрерывными и находятся в рамке считывания.

В одном из вариантов настоящего изобретения «экспрессионный вектор» относится к вектору, содержащему одну или несколько интересующих полинуклеотидных последовательностей, интересующих генов или «трансгенов», которые фланкированы парвовирусными или инвертированными концевыми повторяющимися последовательностями (ITR).

Ни кассета, ни вектор по изобретению не содержит нуклеотидные последовательности генов, кодирующих неструктурные белки (Rep) и структурные белки (Cap) аденоассоциированного вируса.

Рекомбинантный вирус на основе AAV9 (аденоассоциированный вирус 9 серотипа)

В одном из аспектов настоящее изобретение относится к рекомбинантному вирусу на основе AAV9 (аденоассоциированный вирус 9 серотипа) для увеличения экспрессии гена SMN1 в целевых клетках, который включает капсид и вышеуказанную экспрессионную кассету.

Термин «рекомбинантный вирус на основе AAV» (или «вирусоподобная частица на основе AAV», или «рекомбинантный вирусный штамм AAV», или «рекомбинантный вектор AAV», или «вектор rAAV») в контексте настоящего описания относится к вышеуказанной экспрессионной кассете (или вышеуказанному экспрессионному вектору), которая заключена внутри капсида AAV.

Ген Сар, помимо других альтернативных продуктов, кодирует 3 капсидных белка (VP1, VP2 и VP3). Белки VP1, VP2 и VP3 находятся в соотношении 1:1:10, образуя икосаэдрический капсид (Xie Q. et al. The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. Proc Natl Acad Sci USA, 2002; 99:10405-10410). Транскрипция этих генов начинается с одного промотора, р40. Молекулярная масса соответствующих белков (VP1, VP2 и VP3) составляет 87, 72 и 62 кДа, соответственно. Все три белка транслируются с одной мРНК. После транскрипции пре-мРНК может подвергаться сплайсингу двумя разными способами, при этом вырезается более длинный или более короткий интрон и образуются мРНК различной нуклеотидной длины.

При образовании рекомбинантного вируса на основе AAV (rAAV) кассета экспрессии, фланкированная ИКП (ITR), упаковывается в капсид AAV. Гены, необходимые для репликации AAV, как было указано выше, не входят в кассету.

ДНК экспрессионной кассеты упакована в вирусный капсид в виде одноцепочечной молекулы ДНК (оцДНК) длиной приблизительно 3000 нуклеотидов. После инфицирования клетки вирусом, одноцепочечную ДНК конвертируют в форму двухцепочечной ДНК (дцДНК). Только дцДНК могут использовать белки клетки, которые транскрибируют содержащийся ген или гены в РНК.

В некоторых вариантах рекомбинантный вирус на основе AAV9 имеет капсид, который включает белок VP1 AAV9.

В некоторых вариантах рекомбинантный вирус на основе AAV9 имеет капсид, который включает белок VP1 AAV9, имеющий следующую аминокислотную последовательность

MAADGYLPDWLEDNLSEGIREWWALKPGAPQPKANQQHQDNARGLVLPGYKYLPGP
NGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLYKNHADADEFQERLKEDTSFGG

NLGRAVFAQKKRLLEPLGLVEEAAKTAPGKKRPVEQSPQEPDSSAGIGKSGAQAQPAKKRL
NFGQTGDTEVPDPQPIGEPPAAPSGVGS LTMASGGGAPVADNNEGADGVGSSSGNWH
CDSQWLGD RVITTSTRTWALPTYNNHLYKQISNSTSGGSSNDNAYFGYSTPWGYFDNF
RFHCHFSPRDWQRLINNNWGFRPKRLNFKLFNIQVKEVTDNNGVKTIANNL TSTVQVFT
DSDYQLPYVLGSAHEGCLPPFPADVFMIPQYGYLTLNDGSQAVGRSSFYCLEYFPSQML
RTGNNFQFSYEFENVPFHSSYAHSQSLDRLMNPLIDQYLYLSKTINGSGQNQQTLKFSV
AGPSNMAVQGRNYIPGPSYRQQRVSTTVTQNNNSEFAWPGASSWALNGRNSLMNPGP
AMASHKEGEDRFFPLSGSLIFGKQGTGRDNVDADKVMITNEEEIKTTNPVATESYGQVA
TNHQSAQAQAQTGWVQNQGILPGMVWQDRDVYLQGPWAKIPHTDGNFHPSPMLMGGF
GMKHPPPQILIKNTVPADPPTAFNKDKLNSFITQYSTGQVSVEIEWELQKENS KRWNPE
IQYTSNYYKSNNVEFAVNTEGVYSEPRPIGTRYLTRNL (SEQ ID NO: 5).

В некоторых вариантах рекомбинантный вирус на основе AAV9 имеет капсид, который включает белок VP1 AAV9, имеющий аминокислотную последовательность SEQ ID NO: 5 с одной или несколькими точечными мутациями.

В некоторых вариантах рекомбинантный вирус на основе AAV9 имеет капсид, который включает белок VP2 AAV9.

В некоторых вариантах рекомбинантный вирус на основе AAV9 имеет капсид, который включает белок VP2 AAV9, имеющий следующую аминокислотную последовательность:

TAPGKKRPVEQSPQEPDSSAGIGKSGAQAQPAKKRLNFGQTGDTEVPDPQPIGEPPAAPSG
VGSLTMASGGGAPVADNNEGADGVGSSSGNWHCDSQWLGD RVITTSTRTWALPTYNN
HLYKQISNSTSGGSSNDNAYFGYSTPWGYFDNF RFHCHFSPRDWQRLINNNWGFRPKRL
NFKLFNIQVKEVTDNNGVKTIANNL TSTVQVFTDSDYQLPYVLGSAHEGCLPPFPADV
MIPQYGYLTLNDGSQAVGRSSFYCLEYFPSQMLRTGNNFQFSYEFENVPFHSSYAHSQS
LDRLMNPLIDQYLYLSKTINGSGQNQQTLKFSVAGPSNMAVQGRNYIPGPSYRQQRVS
TTVTQNNNSEFAWPGASSWALNGRNSLMNPGPAMASHKEGEDRFFPLSGSLIFGKQGT
GRDNVDADKVMITNEEEIKTTNPVATESYGQVATNHQSAQAQAQTGWVQNQGILPGM
VWQDRDVYLQGPWAKIPHTDGNFHPSPMLMGGFGMKHPPPQILIKNTVPADPPTAFNK
DKLNSFITQYSTGQVSVEIEWELQKENS KRWNPEIQYTSNYYKSNNVEFAVNTEGVYSE
PRPIGTRYLTRNL (SEQ ID NO: 6).

В некоторых вариантах рекомбинантный вирус на основе AAV9 имеет капсид, который включает белок VP2 AAV9, имеющий аминокислотную последовательность SEQ ID NO: 6 с одной или несколькими точечными мутациями.

В некоторых вариантах рекомбинантный вирус на основе AAV9 имеет капсид, который включает белок VP3 AAV9.

В некоторых вариантах рекомбинантный вирус на основе AAV9 имеет капсид, который включает белок VP3 AAV9, имеющий следующую аминокислотную последовательность

MASGGGAPVADNNEGADGVGSSSGNWHCDSQWLGD RVITTSTRTWALPTYNNHLYK
QISNSTSGGSSNDNAYFGYSTPWGYFDNRFHCHFS PRDWQRLINNNWGF RPKRLNFKL
FNIQVKEVTDNNGVKTIANNLTSTVQVFTDS DYQLPYVLGSAHEGCLPPFPADVFMIPQ
YGYLTLNDGSQAVGRSSFYCLEYFPSQMLRTGN NFQFSYEFENVFPFHSSYAHQS LDR L
MNPLIDQYLYLSKTINGSGQNQQTLKFSVAGPSNMAVQGRNYIPGPSYRQQRVSTTVT
QNNNSEFAWPGASSWALNGRNSLMNPGPAMASHKEGEDRFFPLSGSLIFGKQGTGRDN
VDADKVMITNEEEIKTTNPVATESYGQVATNHQSAQAQAQTGWVQNQGILPGMVWQD
RDVYLQGPIWAKIPHTDGNFHPSPLMGGFGMKHPP PQILIKNTPVPADPPTAFNKDKLNS
FITQYSTGQVSVEIEWELQKENS KRWNPEIQYTSNYYKSNNVEFAVNTEGVYSEPRPIGT
RYLTRNL (SEQ ID NO: 7).

В некоторых вариантах рекомбинантный вирус на основе AAV9 имеет капсид, который включает белок VP3 AAV9, имеющий аминокислотную последовательность SEQ ID NO: 7 с одной или несколькими точечными мутациями.

В некоторых вариантах рекомбинантный вирус на основе AAV9 имеет капсид, который включает белки VP1, VP2 и VP3 AAV9.

В некоторых вариантах рекомбинантный вирус на основе AAV9 имеет капсид, который включает белки VP1 с аминокислотной последовательность SEQ ID NO: 5, VP2 с аминокислотной последовательность SEQ ID NO: 6 и VP3 с аминокислотной последовательность SEQ ID NO: 7.

В некоторых вариантах рекомбинантный вирус на основе AAV9 имеет капсид, который включает белки VP1 с аминокислотной последовательность SEQ ID NO: 5 с одной или несколькими точечными мутациями, VP2 с аминокислотной последовательность SEQ ID NO: 6 с одной или несколькими точечными мутациями и VP3 с аминокислотной последовательность SEQ ID NO: 7 с одной или несколькими точечными мутациями.

Под «несколькими точечными мутациями» подразумеваются две, три, четыре, пять, шесть, семь, восемь, девять или десять точечных замен.

Особенно предпочтительные варианты включают замены (мутации), которые являются консервативными по природе, т.е. те замены, которые имеют место в семействе аминокислот, которые объединены по их боковым цепям. В частности, аминокислоты обычно делят на четыре семейства: (1) кислые - аспартат и глутамат; (2) основные - лизин, аргинин, гистидин; (3) неполярные - аланин, валин, лейцин, изолейцин, пролин, фенилаланин, метионин, триптофан; и (4) незаряженные полярные - глицин, аспарагин,

глутамин, цистеин, серин треонин, тирозин. Фенилаланин, триптофан и тирозин иногда классифицируют как ароматические аминокислоты. Например, достаточно обосновано предсказание о том, что выделенная замена лейцина на изолейцин или валин, аспартата на глутамат, треонина на серин или схожая консервативная замена аминокислоты на структурно родственную аминокислоту не окажет важного влияния на биологическую активность. Например, полипептид, представляющий интерес, может включать вплоть до приблизительно 5-10 консервативных или неконсервативных аминокислотных замен, при условии, что желаемая функция молекулы остается незатронутой.

Вариант точечных мутаций в последовательностях белков VP1, VP2 или VP3 AAV9 с помощью аминокислотных замен представляет собой замену, по меньшей мере, одного аминокислотного остатка в белке VP1, VP2 или VP3 AAV9 на другой аминокислотный остаток.

Консервативные замены показаны в таблице А под заголовком «предпочтительные замены».

Таблица А		
Исходный остаток	Примеры замены	Предпочтительные замены
Ala (A)	Val; Leu; Ile	Val
Arg(R)	Lys; Gin; Asn	Lys
Asn(N)	Gin; His; Asp, Lys; Arg	Gin
Asp (D)	Glu; Asn	Glu
Cys (C)	Ser; Ala	Ser
Gln(Q)	Asn; Glu	Asn
Glu (E)	Asp; Gin	Asp
Gly(G)	Ala	Ala
His (H)	Asn; Gin; Lys; Arg	Arg
Ile (I)	Leu; Val; Met; Ala; Phe; Норлейцин	Leu
Leu (L)	Норлейцин; Ile; Val; Met; Ala; Phe	Ile
Lys (K)	Arg; Gin; Asn	Arg

Met (M)	Leu; Phe; Ile	Leu
Phe(F)	Trp; Leu; Val; Ile; Ala; Tyr	Tyr
Pro (P)	Ala	Ala
Ser(S)	Thr	Thr
Thr (T)	Val; Ser	Ser
Trp(W)	Tyr; Phe	Tyr
Tyr(Y)	Trp; Phe; Thr; Ser	Phe
Val (V)	Ile; Leu; Met; Phe; Ala; Норлейцин	Leu

В некоторых вариантах рекомбинантный вирус на основе AAV9 имеет капсид, который включает белок VP1 AAV9, имеющий аминокислотную последовательность SEQ ID NO: 5 или аминокислотную последовательность SEQ ID NO: 5 с одной или несколькими точечными мутациями, а экспрессионная кассета включает следующие элементы в направлении от 5'-конца к 3'-концу:

CMV энхансер;

CMV промотер;

интрон гена hBG1;

вышеуказанную кодон-оптимизированную нуклеиновую кислоту гена SMN1;

сигнал полиаденилирования hGH1;

правый ITR.

В некоторых вариантах рекомбинантный вирус на основе AAV9 имеет капсид, который включает белки VP1 с аминокислотной последовательностью SEQ ID NO: 5, VP2 с аминокислотной последовательностью SEQ ID NO: 6 и VP3 с аминокислотной последовательностью SEQ ID NO: 7, а экспрессионная кассета включает следующие элементы в направлении от 5'-конца к 3'-концу:

CMV энхансер;

CMV промотер;

интрон гена hBG1;

вышеуказанную кодон-оптимизированную нуклеиновую кислоту гена SMN1;

сигнал полиаденилирования hGH1;

правый ITR.

В некоторых вариантах рекомбинантный вирус на основе AAV9 имеет капсид, который включает белки VP1 с аминокислотной последовательность SEQ ID NO: 5 с одной или несколькими точечными мутациями, VP2 с аминокислотной последовательность SEQ ID NO: 6 с одной или несколькими точечными мутациями и VP3 с аминокислотной последовательность SEQ ID NO: 7 с одной или несколькими точечными мутациями, а экспрессионная кассета включает следующие элементы в направлении от 5'-конца к 3'-концу:

CMV энхансер;

CMV промотер;

интрон гена hBG1;

вышеуказанную кодон-оптимизированную нуклеиновую кислоту гена SMN1;

сигнал полиаденилирования hGH1;

правый ITR.

В некоторых вариантах рекомбинантный вирус на основе AAV9 имеет капсид, который включает белок VP1 AAV9, имеющий аминокислотную последовательность SEQ ID NO: 5 или аминокислотную последовательность SEQ ID NO: 5 с одной или несколькими точечными мутациями, а экспрессионная кассета включает нуклеиновую кислоту с SEQ ID NO: 4.

В некоторых вариантах рекомбинантный вирус на основе AAV9 имеет капсид, который включает белки VP1 с аминокислотной последовательность SEQ ID NO: 5, VP2 с аминокислотной последовательность SEQ ID NO: 6 и VP3 с аминокислотной последовательность SEQ ID NO: 7, а экспрессионная кассета включает нуклеиновую кислоту с SEQ ID NO: 4.

В некоторых вариантах рекомбинантный вирус на основе AAV9 имеет капсид, который включает белки VP1 с аминокислотной последовательность SEQ ID NO: 5 с одной или несколькими точечными мутациями, VP2 с аминокислотной последовательность SEQ ID NO: 6 с одной или несколькими точечными мутациями и VP3 с аминокислотной последовательность SEQ ID NO: 7 с одной или несколькими точечными мутациями, а экспрессионная кассета включает нуклеиновую кислоту с SEQ ID NO: 4.

Фармацевтическая композиция

В одном из аспектов настоящее изобретение относится к фармацевтической композиции для доставки гена SMN1 в целевые клетки, которая включает вышеуказанный

рекомбинантный вирус на основе AAV9 в сочетании с одним или несколькими фармацевтически приемлемыми эксципиентами.

В конкретных вариантах осуществления настоящее изобретение относится к фармацевтической композиции, содержащей рекомбинантный вирус на основе AAV9 по изобретению в фармацевтически приемлемом носителе или в других фармацевтических агентах, адъювантах, разбавителях и т.д. Носитель для инъекций обычно является жидким. Носитель для других способов введения может быть или твердым, или жидким, таким как стерильная апиrogenная вода или стерильный апиrogenный фосфатно-солевой буферный раствор. Для введения путем ингаляции носитель является вдыхаемым и предпочтительно находится в твердой или жидкой дисперсной форме. В качестве инъекционной среды предпочтительно использовать воду, содержащую добавки, общепринятые для инъекционных растворов, такие как стабилизирующие агенты, соли или солевые растворы и/или буферы.

«Фармацевтическая композиция» обозначает композицию, включающую в себя вышеуказанный рекомбинантный вирус на основе AAV9 по изобретению и, по крайней мере, один из компонентов, выбранных из группы, состоящей из фармацевтически приемлемых и фармакологически совместимых эксципиентов, таких как наполнители, растворители, разбавители, носители, вспомогательные, распределяющие, средства доставки, консерванты, стабилизаторы, эмульгаторы, суспендирующие агенты, загустители, регуляторы пролонгированной доставки, выбор и соотношение которых зависит от природы и способа назначения и дозировки. Фармацевтические композиции по настоящему изобретению и способы их изготовления будут бесспорно очевидными для специалистов в этой области. Производство фармацевтических композиций предпочтительно должно соответствовать требованиям GMP (надлежащей производственной практики). Композиция может включать буферную композицию, тонические агенты, стабилизаторы и солюбилизаторы.

«Фармацевтически приемлемым» считается материал, который не имеет биологических или других противопоказаний, например, материал можно вводить субъекту без каких-либо нежелательных биологических эффектов. Таким образом, такие фармацевтические композиции можно использовать, например, для трансфекции клетки *ex vivo* или для введения *in vivo* рекомбинантного вируса на основе AAV9 по изобретению непосредственно субъекту.

Термин «эксципиент» или «вспомогательное вещество» используется в данном документе для описания любого компонента, отличающегося от ранее описанных по данному изобретению. Это вещества неорганического или органического происхождения,

используемые в процессе производства, изготовления лекарственных препаратов для придания им необходимых физико-химических свойств.

Под «стабилизатором» понимается вспомогательное вещество или смесь двух и более вспомогательных веществ, которые обеспечивают физическую и/или химическую стабильность активного агента.

Под термином «буфер», «буферная композиция», «буферный агент» понимается раствор, способный сохранять значение pH, благодаря взаимодействию кислотных и щелочных компонентов, входящих в его состав, который дает возможность препарату вектора на основе rAAV5, проявлять устойчивость к изменениям pH. В общем случае, преимущественными являются значения pH фармацевтической композиции от 4,0 до 8,0. В качестве буферных агентов могут быть использованы, например, ацетатный, фосфатный, цитратный, гистидиновый, сукцинатный и т.п. буферные растворы, но, не ограничиваясь ими.

Фармацевтическая композиция является «стабильной», если активный агент сохраняет свою физическую стабильность и/или химическую стабильность и/или биологическую активность в течение заявленного срока годности при температуре хранения, например, при 2-8 °C. Предпочтительно, чтобы активный агент сохранял и физическую, и химическую стабильность, а также биологическую активность. Период хранения выбирается на основании результатов исследования стабильности при ускоренном и естественном хранении.

Фармацевтическая композиция по данному изобретению может изготавливаться, упаковываться или широко продаваться в виде единичной стандартной дозы или множества единичных стандартных доз в виде готовой лекарственной формы. Используемый в данном документе термин «единичная стандартная доза» означает дискретное количество фармацевтической композиции, содержащей заранее определенное количество активного ингредиента. Количество активного ингредиента обычно равно дозировке активного ингредиента, который будет вводиться субъекту, или удобной части такой дозировки, например, половине или трети такой дозировки.

Применение

В одном из аспектов настоящее изобретение относится к применению вышеуказанного рекомбинантного вируса на основе AAV9 или вышеуказанной композиции для доставки гена SMN1 в целевые клетки.

Любой способ введения рекомбинантного вируса на основе AAV9, принятый в данной области, может соответствующим образом использоваться для вышеуказанного рекомбинантного вируса на основе AAV9 по данному изобретению.

Рекомбинантный вирус на основе AAV9 предпочтительно вводят в клетку в биологически эффективном количестве. «Биологически эффективное» количество рекомбинантного вируса представляет собой количество, которое достаточно, чтобы вызвать инфекцию (или трансдукцию) и экспрессию гетерологичной последовательности нуклеиновой кислоты в клетке. Если вирус вводят в клетку *in vivo* (например, вирус вводят субъекту, как описано ниже), «биологически эффективное» количество вирусного вектора представляет собой количество, которое достаточно, чтобы вызвать трансдукцию и экспрессию гетерологичной последовательности нуклеиновой кислоты в клетке-мишени.

Клетка для введения вышеуказанного рекомбинантного вируса на основе AAV9 по изобретению может быть клеткой любого типа, включая в себя без ограничения нервные клетки (включающие в себя клетки периферической и центральной нервной системы, в частности, клетки головного мозга), легочные клетки, эпителиальные клетки (например, эпителиальные клетки кишечника и дыхательных путей), мышечные клетки, клетки поджелудочной железы (в том числе островковые клетки), печеночные клетки, клетки миокарда, костные клетки (например, стволовые клетки костного мозга), гемопоэтические стволовые клетки, клетки селезенки, кератиноциты, фибробласты, эндотелиальные клетки, клетки предстательной железы, половые клетки и тому подобное. Альтернативно, клетка для введения вышеуказанного рекомбинантного вируса на основе AAV9 может быть любой клеткой-предшественником. В качестве дополнительной альтернативы, клетки могут представлять собой стволовые клетки (например, нервные стволовые клетки, стволовые клетки печени). Кроме того, клетки могут происходить от любых видов, как указано выше.

Вышеуказанный рекомбинантный вирус на основе AAV9 не используется для модификации генетической целостности клеток зародышевой линии человека.

Примеры

Для наилучшего понимания изобретения приводятся следующие примеры. Эти примеры приведены только в иллюстративных целях и не должны толковаться как ограничивающие сферу применения изобретения в любой форме.

Все публикации, патенты и патентные заявки, указанные в этой спецификации включены в данный документ путем отсылки. Хотя вышеупомянутое изобретение было довольно подробно описано путем иллюстрации и примера в целях исключения

двусмысленного толкования, специалистам в данной области на основе идей, раскрытых в данном изобретении, будет вполне понятно, что могут быть внесены определенные изменения и модификации без отклонения от сущности и объема прилагаемых вариантов осуществления изобретения.

Материалы и общие методы

Методы рекомбинантной ДНК

Для манипуляций с ДНК использовали стандартные методы, описанные у Sambrook J. и др., *Molecular cloning: A laboratory manual*; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989. Реагенты для молекулярной биологии использовали согласно инструкциям производителей. Вкратце, плазмидную ДНК нарабатывали для дальнейших манипуляций в клетках *E. coli*, выращиваемых под селективным давлением с антибиотиками для того, чтобы плазмиды не терялись в клеточной популяции. Плазмидную ДНК выделяли из клеток коммерческими наборами, измеряли концентрацию и использовали для клонирования с помощью обработки эндонуклеазами рестрикции или методами ПЦР-амплификации. Фрагменты ДНК лигировали между собой с помощью лигаз и трансформировали в бактериальные клетки для отбора клонов и дальнейших работ. Все полученные генетические конструкции подтверждали по паттернам рестрикции и полным секвенированием по Сэнгеру.

Синтез генов

Требуемые сегменты генов получали из олигонуклеотидов, созданных путем химического синтеза. Генные сегменты длиной от 300 до 1000 п. н., которые фланкированы уникальными сайтами рестрикции, собирали путем ренатурации олигонуклеотидов друг на друге с последующей ПЦР-амплификацией с крайних праймеров. В результате получали смесь фрагментов, включая нужный. Фрагменты клонировали по сайтам рестрикции в промежуточные векторы, после чего последовательности ДНК субклонированных фрагментов подтверждали путем секвенирования ДНК.

Определение последовательностей ДНК

Последовательности ДНК определяли путем секвенирования по Сэнгеру. Анализ последовательностей ДНК и белков и обработку данных о последовательностях осуществляли в программе SnapGene 4.2 и выше для создания, картирования, анализа, аннотирования и иллюстрации последовательностей.

Культивирование клеточных культур

В экспериментах были использованы клеточные линии HEK293 (Human Embryonic Kidney clone 293) и HSMC (Human Skeletal Muscle Cells). Клетки культивировались в стандартных условиях при 37°C и 5%CO₂, на полной питательной среде DMEM с добавлением 10% FBS и антибиотика. Для культивирования HSMC культуральный пластик предварительно покрывался коллагеном (Gibco). Пересев клеток осуществлялся при достижении 80-90% конfluence. Жизнеспособность клеток оценивалась с помощью окраски Trypan Blue и камеры Горяева либо с помощью окраски PI и проточной цитометрии.

Трансфекция клеток

Клеточные линии засевали накануне трансфекции в 6-луночные планшеты таким образом, чтобы они достигали 70-80% конfluence к моменту трансфекции. Трансфекцию осуществляли с помощью коммерческих наборов для липофекции по протоколу производителя. Через 72 ч клетки обрабатывали растворами трипсина или аналогичными, снимали с подложки, отмывали в фосфатном буфере и собирали для дальнейшего анализа экспрессии целевых генов и белков. При постановке каждой трансфекции использовали контрольную плазмиду, экспрессирующую GFP, для контроля эффективности трансфекции (доля GFP-позитивных клеток в процентах). Дальнейший анализ проводили только в том случае, если эффективность трансфекции составляла не менее 50%.

Все работы проводились в 3 независимых экспериментах.

Анализ генной экспрессии

Экспрессию SMN1 на уровне мРНК проверяли с помощью количественной ПЦР. Вкратце, использовали праймеры и пробу, специфичные к последовательности SMN1 дикого типа или GeneBeam. Для контроля количества исходной РНК использовали праймеры и пробу, специфичные к гену домашнего хозяйства GAPDH. Для каждого набора праймеров и проб строили калибровочные кривые с применением известного количества копий линеаризованной плазмидной ДНК, содержащей амплифицируемую последовательность соответствующего гена. Анализ экспрессии осуществляли, определяя по калибровочным кривым количество копий SMN1-GeneBeam, SMN1-WT и GAPDH в каждом образце, после чего нормализовали количество копий SMN1 на 10000 копий

GAPDH. Полученные значения сравнивали между собой для различных образцов в рамках одного эксперимента.

Определение экспрессии белка SMN1 с помощью проточной цитометрии

Оценку содержания белка SMN1 в клетках проводили посредством внутриклеточного окрашивания, с последующим анализом с помощью проточной цитометрии. Вкратце, клетки снимали с культуральных планшетов с помощью TrypLE, отмывали в PBS, фиксировали в растворе 4% параформальдегида, пермеабилizировали с помощью 0,5% раствора Triton X-100 в PBS, инкубировали в блокирующем буфере с 1-5% BSA и окрашивали двухэтапно с использованием первичных антител к SMN1 и вторичных антител, меченных Alexa Fluor 488. После окрашивания клетки однократно промывали в PBS и анализировали на проточном цитометре. Оценивали среднюю интенсивность сигнала за вычетом сигнала, окрашенного вторичными антителами без добавления первичных антител.

Сборка и очистка вирусных частиц рекомбинантных векторов AAV

Для сборки вирусных частиц AAV, содержащих ген SMN1 или контрольный ген GFP, использовали упаковочные клетки HEK293, в которые трансфецировали 3 плазмиды:

Плазмида, содержащая геном AAV с кассетой для экспрессии трансгена (SMN1 или GFP);

Плазмида для экспрессии гена Cap серотипа AAV9 и гена Rep серотипа AAV2. Каждый ген с помощью альтернативных рамок считывания кодирует несколько белковых продуктов;

Плазмида для экспрессии генов аденовируса Ad5, необходимых для сборки и упаковки капсидов AAV.

Через 72 часа клетки лизировали и проводили очистку и концентрирование вирусных частиц с помощью методов фильтрации и хроматографии. Титр вирусных частиц определяли с помощью количественной ПЦР с праймерами и пробой, специфичными к участку рекомбинантного вирусного генома и выражали в виде количества копий вирусных геномов на 1 мл.

Трансдукция клеточных культур

Клеточные линии засеивали аналогично экспериментам для трансфекции, после чего добавляли препарат с вирусными частицами и через 72 ч клетки анализировали. Эффективность трансдукции считали, анализируя процент GFP+ клеток.

Для используемых культур предварительно были поставлены эксперименты с проверкой эффективности трансдукции. Кратко, препарат вируса AAV9-GFP трансдуцировали в клеточные линии в различных соотношениях клеток и вирусных частиц. Отношение количества вирусных частиц и клеток называют multiplicity of infection (MOI). MOI вируса AAV9-GFP варьировали от 50000 до 1000000. В результате для каждой линии были определены диапазоны MOI, в пределах которых эффективность трансдукции менялась линейно в зависимости от MOI. Дальнейшие работы с трансдукцией клеточных линий проводили в их линейных диапазонах.

После трансдукции анализ экспрессии генов и белков осуществляли как описано выше.

Все работы проводились в 3 независимых экспериментах.

Пример 1. Способ получения кодон -оптимизированного гена SMN1

Для получения кодон-оптимизированного гена SMN1 за основу была взята соответствующая аминокислотная последовательность белка SMN_HUMAN (SEQ ID NO:1).

Данная аминокислотная последовательность SEQ ID NO:1 переводилась в нуклеотидную последовательность путем последовательного сопоставления каждой аминокислоте начиная с N-конца одного из кодирующих её синонимичных кодонов с учетом одного или совокупности следующих признаков:

1) частоты использования данного кодона (Yasukazu Nakamura ET AL., Codon usage tabulated from the international DNA sequence databases; its status 1999, Nucleic Acids Research, 1999, Vol. 27, No. 1, doi: [10.1093/nar/27.1.292](https://doi.org/10.1093/nar/27.1.292));

2) GC-состава концевой участка получаемой нуклеотидной последовательности (целевым значением GC-состава считалось 60% исходя из статьи Grzegorz Kudla ET AL., High Guanine and Cytosine Content Increases mRNA Levels in Mammalian Cells, PLoS Biol, June 2006, Volume 4, Issue 6, e180, doi: [10.1371/journal.pbio.0040180](https://doi.org/10.1371/journal.pbio.0040180), поэтому кодон был тем предпочтительнее, чем меньше получалась разность текущего GC-состава с целевым);

3) свободной энергии фолдинга концевой участка получаемой нуклеотидной последовательности (вторичные структуры определялись при помощи алгоритма Зукера, Michael Zuker ET AL., Optimal computer folding of large RNA sequences using thermodynamics

and auxiliary information, *Nucleic Acids Research*, Volume 9, Issue 1, 10 January 1981, Pages 133–148, doi: [10.1093/nar/9.1.133](https://doi.org/10.1093/nar/9.1.133)).

В процессе построения также избегалась генерация смысловых нуклеотидных последовательностей, таких как сайты рестрикции, участки внутренней посадки рибосомы, сайты сплайсинга т.д.

В результате перевода аминокислотной последовательности SEQ ID NO:1 в нуклеотидную последовательность был получен ряд кодон-оптимизированных нуклеотидных последовательностей гена SMN1.

Из вышеуказанного ряда кодон-оптимизированных нуклеотидных последовательностей SMN1 несколько последовательностей не показывали увеличения транскрипции гена SMN1 в дальнейших исследованиях, то есть не было достоверного увеличения количества копий мРНК SMN1-opt по сравнению с SMN1-WT на какой-либо из использованных клеточных линий, или данное увеличение было незначительным.

Большинство кодон-оптимизированных нуклеотидных последовательностей гена SMN1 показал увеличение транскрипции гена SMN1 в 1,5-2 раза в дальнейших исследованиях, то есть достоверно увеличивая количество копий мРНК SMN1-opt по сравнению с SMN1-WT на всех использованных клеточных линиях.

Одна из вышеуказанного ряда кодон-оптимизированных нуклеотидных последовательностей гена SMN1 неожиданно показала увеличение транскрипции гена SMN1 более чем в 3 раза в дальнейших исследованиях, то есть неожиданно увеличивая количество копий мРНК SMN1-opt в более чем 3 раза по сравнению с SMN1-WT на всех использованных клеточных линиях (см. Примеры 3-4). Данная финальная кодон-оптимизированная нуклеотидная последовательность гена SMN1 получила условное название SMN1-GeneBeam (или сокращенно SMN1-GB).

Финальная кодон-оптимизированная последовательность SMN1 (SMN1-GeneBeam) имеет нуклеотидную последовательность представленную SEQ ID NO: 2.

Данная финальная кодон-оптимизированная нуклеотидная последовательность SMN1 (SMN1-GeneBeam) характеризуется повышенным индексом адаптации кодонов (Paul M. Sharp ET AL., The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications, *Nucleic Acids Research*, Volume 15, Issue 3, 11 February 1987, Pages 1281–1295, doi: [10.1093/nar/15.3.1281](https://doi.org/10.1093/nar/15.3.1281) - стандартная метрика для оценки последовательности на предмет частот использованных кодонов) по сравнению с кодирующей последовательностью гена SMN дикого типа (SMN1-WT с SEQ ID NO: 3).

Индекс адаптации кодонов для финальной кодон-оптимизированной нуклеотидной последовательности гена SMN1 (SEQ ID NO: 2) равен 98%, для последовательности дикого типа - 75%.

GC-состав последовательности дикого типа равен 45%, то есть отличается от целевого значения на 15%, GC-состав финальной кодон-оптимизированной нуклеотидной последовательности гена SMN1 (SEQ ID NO: 2) равен 64%, т.е. отличается от целевого значения на 4%.

Финальная кодон-оптимизированная нуклеотидная последовательность гена SMN1 (SEQ ID NO:2) и нуклеотидная последовательность гена SMN1 дикого типа (SEQ ID NO:3) идентичны на 71%.

Пример 2. Сборка генетических конструкций, несущих рекомбинантный геном AAV и кодирующих ген SMN1.

Последовательность гена SMN1 дикого типа была получена путём амплификации со специфическими праймерами с кДНК, синтезированной на основе тотальной РНК клеток HEK293. В процессе амплификации с 5'-конца гена была добавлена последовательность Kozak и сайт рестрикции ClaI, а с 3'-конца – сайт рестрикции XbaI. После этого последовательность гена SMN1 была клонирована рестриктазно-лигазным методом по сайтам ClaI и XbaI в коммерческую конструкцию pAAV-GFP Control plasmid (VPK-402) от CellBiolab (США), с заменой гена GFP на SMN1, получив конструкцию pAAV-SMN1-WT.

Последовательность SMN1-GeneBeam была собрана как описано выше. В силу сложности последовательности, несмотря на её относительно небольшой размер, провели серию субклонирований фрагментов гена в промежуточных векторах pGEMT с верификацией сиквенсов для каждого вектора. Далее из нескольких промежуточных векторов с помощью ПЦР собрали полноразмерную версию гена и заклонировали в промежуточный вектор pGEMT. В качестве конечной генетической конструкции использовали конструкцию pAAV-SMN1-WT, заменив SMN1 дикого типа на SMN1-GeneBeam по сайтам ClaI и XbaI, добавленным на концы последовательности SMN1-GeneBeam с помощью ПЦР.

Конечный вектор содержит все необходимые элементы для экспрессии гена и сборки в составе генома рекомбинантного AAV:

- 1) Терминальные повторы ITR на концах последовательности, которая инкапсулируется в вирусный капсид;
- 2) Элементы для экспрессии целевого гена (промотор, энхансер, интрон, последовательность Kozak, трансген, сайт полиаденилирования);

3) Ориджин бактериальной репликации и ген устойчивости к антибиотикам для наработки плазмидной ДНК в бактериальных клетках.

Важно отметить, что генетические конструкции, содержащие гены SMN1-WT и SMN1-GeneBeam, отличаются только последовательностями генов SMN1, а в остальном полностью идентичны.

Пример 3. Проверка экспрессии SMN1 с генетических конструкций.

Генетические конструкции pAAV-GFP, pAAV-SMN1-WT и pAAV-SMN1-GB трансфецировали в клетки HEK293 и HSMC как описано выше. Использовали 5 мкг ДНК на 1 лунку. Через 72 ч клетки собрали и проанализировали экспрессию SMN1 (с нормализацией на GAPDH) как описано выше.

Было обнаружено, что кодон-оптимизация гена SMN1 влияет на транскрипцию SMN1, достоверно увеличивая количество копий мРНК SMN1-GB в несколько раз по сравнению с SMN1-WT на обеих использованных клеточных линиях (фиг. 1). В частности, для клеток HEK293 отношение нормализованной экспрессии SMN1-GB к SMN1-WT составило 3,9, а для клеток HSMC – 12,8.

Данное свойство SMN1-GeneBeam, как показывают полученные данные, не является клеточно-специфичным, при этом обеспечивает увеличение экспрессии целевого гена в клетках в несколько раз, что может быть важным преимуществом при разработке генотерапевтических препаратов. При этом данное свойство не обусловлено никакими отличиями в кассете генной экспрессии и свойствами возможных вирусных капсидов, несущих геном с генов SMN1-GeneBeam, поскольку данный анализ проводили на генетических конструкциях, которые отличаются только кодон-оптимизацией генов SMN1, а в остальном полностью идентичны.

Клетки HSMC были выбраны для проверки экспрессии SMN1 на уровне белка с помощью проточной цитометрии, как описано выше. Было показано, что сигнал от SMN1-специфичных антител в клетках, трансфецированных pAAV-SMN1-GB, выше, чем в клетках, трансфецированных pAAV-SMN1-WT, в 12,2 раз при использованном количестве ДНК 5 мкг на 1 лунку (фиг. 2). Данный результат означает, что SMN1-GB не имеет преимуществ в трансляции, однако в силу увеличенной транскрипции количество конечного белка в клетках также увеличивается.

Пример 4. Создание вирусных препаратов, экспрессирующих SMN1

Плазмиды pAAV-SMN1-WT и pAAV-SMN1-GB вместе с остальными плазмидами, необходимыми для получения вирусных частиц рекомбинантного AAV (см. выше), были

использованы для биопроцесса получения AAV. В качестве серотипа был выбран серотип AAV9 дикого типа или с одной или несколькими точечными мутациями.

Во всех случаях сравнение свойств SMN1 дикого типа и SMN1-GeneBeam проводили только в случае идентичности используемого серотипа и мутаций капсида, если они присутствовали. Все серотипы на базе AAV9, дикого типа или с мутациями, в дальнейшем обозначаются как AAV9 без указания мутаций.

В результате биопроцесса были получены рекомбинантные вирусные частицы, обозначенные как AAV9-SMN1-WT и AAV9-SMN1-GB, а также контрольные частицы AAV9-GFP. После определения титров вирусных частиц, все 3 препарата с одинаковыми MOI (значения MOI варьировали между экспериментами от 50 до 200 тысяч) были использованы для трансдукции пермиссивных клеток – первичных миоцитов человека HSMC. Дальнейший анализ проводили только в том случае, если эффективность трансдукции составляла не менее 50%.

После успешной трансдукции клетки снимали с подложки, отмывали в фосфатном буфере и анализировали экспрессию SMN1 на уровне мРНК и белка как описано выше. Было показано, что увеличение транскрипционной активности SMN1-GeneBeam сохраняется, таким образом, мРНК SMN1-GeneBeam было детектировано в 7,3 раз больше, чем для SMN1 дикого типа. Аналогичное увеличение наблюдалось и на уровне белка (в 6,8 раз) (фиг. 3), что показывает отсутствие преимуществ SMN1-GB на уровне трансляции, однако детектируемое увеличение эффективности транскрипции позволяет с помощью препарата AAV9-SMN1-GB обеспечивать более высокий уровень экспрессии SMN1 в целевых клетках, что является важным преимуществом при лечении, например, спинальной мышечной атрофии, где уровень экспрессии белка SMN1 определяет тип заболевания от 0 (эмбриональная леталь) до 4 (не требует специального лечения).