(19) Евразийское патентное ведомство

(21) 201990213 (13) A1

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ

(43) Дата публикации заявки
2019.11.29

(22) Дата подачи заявки
2017.07.05

(51) Int. Cl. ***C12N 15/113*** (2010.01)
C12N 9/16 (2006.01)
C12N 15/63 (2006.01)
C12N 15/85 (2006.01)
A61K 38/46 (2006.01)
A61K 48/00 (2006.01)
A61P 35/00 (2006.01)

(54) КОМПОЗИЦИИ И СПОСОБЫ, ВКЛЮЧАЮЩИЕ УСОВЕРШЕНСТВОВАНИЯ ГИДОВЫХ РНК CRISPR С ИСПОЛЬЗОВАНИЕМ ПРОМОТОРА H1

(31) **62/358,335**
(32) **2016.07.05**
(33) **US**
(86) **PCT/US2017/040707**
(87) **WO 2018/009534 2018.01.11**

(71) Заявитель:
ДЗЕ ДЖОНС ХОПКИНС ЮНИВЕРСИТИ (US)

(72) Изобретатель:
Джаскула-Ранга Винод, Зэк Дональд, Велсби Дерек С. (US)

(74) Представитель:
Медведев В.Н. (RU)

(57) Настоящее изобретение относится к композициям и способам, включающим усовершенствование системы CRISPR (например, CRISPR-ассоциированная (Cas) 9 (системы CRISPR-Cas9, не Cas9 CRISPR)). Такие композиции могут содержать модификации промоторной области H1, добавление модификаций 5'UTR, различные ортологичные последовательности промотора H1, новые компактные последовательности двунаправленного промотора с активностью в отношении как pol II, так и pol III, вставку консенсусных последовательностей Козак, последовательностей терминации, вставку двунаправленного промотора pol II/pol III с зависимой от условий экспрессией, вставку донорной матричной последовательности для коррекции мутаций или их комбинации. Другие аспекты изобретения относятся к модификациям Cas9 посредством слитых конструкций для посттрансляционной регуляции клеточного цикла, сконструированных частичных участков-мишеней, чтобы нуклеаза могла связываться без расщепления, участков ауторегуляции и N-концевых модификаций для модулирования времени полужизни.

201990213 A1

201990213 A1

ОПИСАНИЕ ИЗОБРЕТЕНИЯ

2420-554578EA/032

КОМПОЗИЦИИ И СПОСОБЫ, ВКЛЮЧАЮЩИЕ УСОВЕРШЕНСТВОАНИЯ ГИДОВЫХ РНК CRISPR С ИСПОЛЬЗОВАНИЕМ ПРОМОТОРА H1

ПЕРЕКРЕСТНАЯ ССЫЛКА

По настоящей заявке испрашивается приоритет предварительной заявке США № 62/358335, поданной 5 июля 2016 года, полое содержание которой включено в настоящее описание в качестве ссылки.

УРОВЕНЬ ТЕХНИКИ

Короткие палиндромные повторы, регулярно расположенные кластерами (CRISPR), вместе с генами cas (CRISPR-ассоциированные), составляют адаптивную иммунную систему, которая обеспечивает приобретенную резистентность против вторгающихся чужеродных нуклеиновых кислот в бактериях и археях (Barrangou et al. (2007) *Science* 315:1709-12). CRISPR состоит из наборов коротких консервативных повторяющихся последовательностей, между которыми располагаются уникальные вариабельные последовательности ДНК сходного размера, называемые спейсерами, которые часто происходят из фаговой или плазмидной ДНК (Barrangou et al. (2007) *Science* 315:1709-12; Bolotin et al. (2005) *Microbiology* 151: 2551-61; Mojica et al. (2005) *J. Mol. Evol.* 60:174-82). Система CRISPR-Cas функционируют путем приобретения коротких фрагментов чужеродной ДНК (спейсеров), которые встраиваются в область CRISPR и обеспечивают иммунитет против последующих воздействий фагов и плазмид, которые имеют соответствующие последовательности (Barrangou et al. (2007) *Science* 315:1709-12; Brouns et al. (2008) *Science* 321:960-64). Именно эта интерференция/иммунитет CRISPR-Cas обеспечивает опосредуемое cr-РНК подавление чужеродных нуклеиновых кислот (Horvath & Barrangou (2010) *Science* 327:167-70; Deveau et al. (2010) *Annu. Rev. Microbiol.* 64:475-93; Marraffini & Sontheimer (2010) *Nat. Rev. Genet.* 11:181-90; Bhaya et al. (2011) *Annu. Rev. Genet.* 45:273-97; Wiedenheft et al. (2012) *Nature* 482:331-338).

Применение конструкций CRISPR, в основе которых лежит нуклеазная активность белка Cas9 (Makarova et al. (2011) *Nat. Rev. Microbiol.* 9:467-77), в сочетании с синтетической гидовой РНК (гРНК), недавно произвело переворот в генной инженерии, позволив беспрецедентное манипулирование последовательностями ДНК. Конструкции CRISPR/Cas9 просто и быстро синтезировать, и они могут быть мультиплексными. Однако, несмотря на относительную простоту их синтеза, CRISPR имеют технологические ограничения, связанные с их доступом к области генома, являющейся мишенью, который зависит как от свойств самого белка Cas9, так и от его гРНК.

Расщепление системой CRISPR требует комплементарного спаривания оснований гРНК с последовательностью ДНК из 20 нуклеотидов и соответствующего примыкающего к протоспейсеру мотива (PAM) - короткого нуклеотидного мотива, находящегося с 3'-стороны от области-мишени (Jinek et al. (2012) *Science* 337: 816-821). Теоретически, с использованием технологии CRISPR можно осуществлять нацеливание на любую уникальную последовательность N_{20}-PAM в геноме. Специфичность связывания последовательности PAM с ДНК, которая варьируется в зависимости от вида, из которого происходит конкретная используемая Cas9, имеет одно ограничение. В настоящее время имеющим наименьшие ограничения и наиболее часто используемым белком Cas9 является белок из *S. pyogenes*, который распознает последовательность NGG, и, таким образом, можно осуществлять нацеливание на любую уникальную последовательность из 21 нуклеотида в геноме, после которой следуют два гуанозиновых нуклеотида (N_{20}NGG). Расширение доступных областей для нацеливания, осуществляемого белковым компонентом, ограничивается открытием и применением новых белков Cas9 с измененными требованиями к PAM (Cong et al. (2013) *Science* 339: 819-823; Hou et al. (2013) *Proc. Natl. Acad. Sci. U.S.A.* 110(39):15644-9), или ожиданием получения новых вариантов Cas9 посредством мутагенеза или направленной эволюции.

Второе технологическое ограничение системы CRISPR обусловлено тем, что экспрессия гРНК начинается с 5'-гуанозинового нуклеотида. Применение класса промоторов РНК

полимеразы III типа III является особенно пригодным для экспрессии гРНК, поскольку эти короткие некодирующие транскрипты имеют строго определенные концы, и все необходимые для транскрипции элементы, за исключением нуклеотида 1+, содержатся в вышележащей промоторной области. Однако, поскольку обычно используемый промотор U6 требует гуанозинового нуклеотида для инициации транскрипции, использование промотора U6 далее ограничивает геномные области, на которые можно осуществлять нацеливание, до $GN_{19}NGG$ (Mali et al. (2013) *Science* 339:823-826; Ding et al. (2013) *Cell Stem Cell* 12:393-394 Ranganathan, V *et al. Nature communications* 5, 4516 (2014)). Альтернативные подходы, такие как транскрипция *in vitro* с помощью промоторов T7, T3 или SP6, также могут потребовать инициирующего гуанозинового нуклеотида(ов) (Adhya et al. (1981) *Proc. Natl. Acad. Sci. U.S.A.* 78:147-151; Melton et al. (1984) *Nucleic Acids Res.* 12:7035-7056; Pleiss et al. (1998) *RNA* 4:1313-1317).

СУЩНОСТЬ ИЗОБРЕТЕНИЯ

Для применения настоящего изобретения на практике, как правило, используются, если нет иных указаний, общепринятые способы клеточной биологии, культивирования клеток, молекулярной биологии, трансгенной биологии, микробиологии, технологии рекомбинантных нуклеиновых кислот (например, ДНК), иммунологии и РНК-интерференции (РНК-i), которые входят в пределы компетентности в данной области техники. Неограничивающее описание некоторых из этих способов приведено в следующих публикациях: Ausubel, F., et al., (eds.), *Current Protocols in Molecular Biology, Current Protocols in Immunology, Current Protocols in Protein Science* и *Current Protocols in Cell Biology*, все John Wiley & Sons, N.Y., издание от декабря 2008 года; Sambrook, Russell, and Sambrook, *Molecular Cloning. A Laboratory Manual,* 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2001; Harlow, E. and Lane, D., Antibodies - A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1988; Freshney, R. I., "Culture of Animal Cells, A Manual of Basic Technique", 5th ed., John Wiley & Sons, Hoboken, N.J., 2005. Неограничивающая информация, касающаяся

терапевтических средств и заболеваний человека, представлена в Goodman and Gilman's The Pharmacological Basis of Therapeutics, 11th Ed., McGraw Hill, 2005, Katzung, B. (ed.) Basic and Clinical Pharmacology, McGraw-Hill/Appleton & Lange 10th ed. (2006) или 11th edition (July 2009). Неограничивающая информация, касающаяся генов и генетических нарушений, представлена в McKusick, V. A.: Mendelian Inheritance in Man. A Catalog of Human Genes and Genetic Disorders. Baltimore: Johns Hopkins University Press, 1998 (12th edition) или более последней базе данных в интернете: Online Mendelian Inheritance in Man, OMIM™. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, Md.) и National Center for Biotechnology Information, National Library of Medicine (Bethesda, Md.), от 1 мая 2010 года, доступных в Интернете: http://www.ncbi.nlm.nih.gov/omim/ и в Online Mendelian Inheritance in Animals (OMIA), базе данных генов, наследственных заболеваний и признаков у видов животных (отличных от человека и мыши), доступной в интернете: http://omia.angis.org.au/contact.shtml. Все патенты, патентные заявки и другие публикации (например, научные статьи, книги, интернет-сайты и базы данных), упоминаемые в настоящем описании, включены в настоящее описание в качестве ссылки в полном объеме. В случае противоречия между описанием и включенными ссылками (включая любые изменения в них, которые могут быть основаны на включенной ссылке), настоящее описание имеет преимущество. Если нет иных указаний, в настоящем описании используются стандартные принятые в данной области термины. В настоящем описании используются стандартные сокращенные обозначения различных терминов.

В настоящем описании предусматриваются композиции и способы, содержащие усовершенствования системы CRISPR/Cas9 (т.е. гидовые РНК CRISPR с использованием промотора H1). В некоторых вариантах осуществления усовершенствования включают модификации промоторной области H1. В некоторых вариантах осуществления композиции включают усиление двунаправленной экспрессии pol II

под контролем H1 с использованием модификаций 5'UTR. В некоторых вариантах осуществления композиции включают модулирование двунаправленной экспрессии с использованием различных ортологичных последовательностей промотора H1. В некоторых вариантах осуществления композиции содержат новые компактные двунаправленные промоторные последовательности с активностью в отношении как pol II, так и pol III (например, 7sk, 5'UTR, консенсусные последовательности Козак или их комбинации). В некоторых вариантах осуществления способ включает скрининг экспрессии с двунаправленных промоторов с активностью в отношении как РНК-pol II, так и РНК-pol III. В некоторых вариантах осуществления композиции включают зависимую от условий экспрессию pol II/pol III с двунаправленного промотора (например, участки TetR и TetO), который может регулировать ферментативную активность рибонуклеопротеинов или РНК-направленных нуклеаз. В некоторых вариантах осуществления усовершенствование включает добавление донорной матричной последовательности для коррекции мутаций (например, направляемая гомологией репарация (HDR)).

В других вариантах осуществления настоящего изобретения усовершенствование включает модификации компонента системы CRISPR/Cas9. В некоторых вариантах осуществления модификации внесены в нуклеазы (например, РНК-направленные нуклеазы). В некоторых вариантах осуществления нуклеазу (например, Cas9) модифицируют путем посттранскрипционной регуляции клеточного цикла (например, слитые белки, содержащие геминин (Gem) или Cdt1). В некоторых вариантах осуществления нуклеазу (например, Cas9) модифицируют путем конструирования частичных участков-мишеней, так чтобы нуклеаза могла связываться без расщепления ДНК. В некоторых вариантах осуществления нуклеазу (например, Cas9) модифицируют путем модулирования ее времени полужизни с использованием типа N-концевых аминокислот.

Один аспект изобретения относится к не встречающейся в природе нуклеазной системе, содержащей один или несколько векторов, содержащих: a) промотор, функционально связанный по

меньшей мере с одной нуклеотидной последовательностью, кодирующей гидовую РНК (гРНК) нуклеазной системы, где гРНК гибридизуется с последовательностью-мишенью молекулы или молекулы РНК в клетке и где молекула ДНК или молекула РНК кодирует один или несколько продуктов генов, экспрессируемых в клетке; и b) регуляторный элемент, функционирующий в клетке, функционально связанный с нуклеотидной последовательностью, кодирующей нацеленную на геном нуклеазу,

где компоненты (a) и (b) расположены на одном и том же или на различных векторах системы, где гРНК осуществляет нацеливание на последовательность-мишень и гибридизуется с ней, и нуклеаза расщепляет молекулу ДНК или РНК для изменения экспрессии одного или нескольких продуктов генов.

В некоторых вариантах осуществления система представляет собой CRISPR (например, CRISPR-ассоциированная (Cas) 9 (CRISPR-Cas9, не-Cas9 системы CRISPR, система CRISPR-Cpf-1, и т.п.)).

В некоторых вариантах осуществления система упакована в одну частицу аденоассоциированного вируса (AAV).

В некоторых вариантах осуществления упаковывающий аденоассоциированный вирус выбран из аденовируса серотипа 2, аденовируса серотипа 5 или аденовируса серотипа 35.

В некоторых вариантах осуществления аденоассоциированный упаковывающий вирус представляет собой аденовирус серотипа 5.

В некоторых вариантах осуществления система инактивирует один или несколько продуктов генов.

В некоторых вариантах осуществления нуклеазная система вырезает по меньшей мере одну мутацию гена.

В некоторых вариантах осуществления промотор выбран из группы, состоящей из промотора H1, 7sk, RPPH1-PARP2 человека, SRP-RPS29, 7sk1-GSTA4, SNAR-G-1-CGB1, SNAR- CGB2, RMRP-CCDC107, tRNA(Lys)-ALOXE3, RNU6-9-MED16: tRNA (Gly)-DPP9, RNU6-2-THEM259, SNORD13-C8orf41, RPPH1-PARP2 мыши и RPPH1-PARP2 крысы.

В некоторых вариантах осуществления промотор представляет собой промотор H1.

В некоторых вариантах осуществления промотор H1 содержит нуклеотидную последовательность, обладающую по меньшей мере 80%

идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 12 или 32.

В некоторых вариантах осуществления промотор H1 содержит нуклеотидную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 12.

В некоторых вариантах осуществления промотор H1 содержит нуклеотидную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 32.

В некоторых вариантах осуществления промотор является ортологом промотора H1.

В некоторых вариантах осуществления ортологичный промотор H1 происходит из плацентарного млекопитающего.

В некоторых вариантах осуществления ортологичный промотор H1 происходит из *ailuropoda melanoleuca, bos taurus, callithrix jacchus, canis familiaris, cavia porcellus, chlorocebus sabaeus, choloepus hoffmanni, dasypus novemcinctus, dipodomys ordii, equus caballus, erinaceus europaeus, felis catus, gorilla gorilla, homo sapiens, ictidomys tridecemlineatus, loxodonta africana, macaca mulatta, mus musculus, mustela putorius furo, myotis lucifugus, nomascus leucogenys, ochotona princeps, oryctolagus cuniculus, otolemur garnettii, ovis aries, pan troglodytes, papio anubis, pongo abelii, procavia capensis, pteropus vampyrus, rattus norvegicus, sus scrofa, tarsius syrichta, tupaia belangeri, tursiops truncatus, vicugna pacos*.

В некоторых вариантах осуществления ортологичный промотор H1 происходит из мыши или крысы.

В некоторых вариантах осуществления ортологичный промотор H1 содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 84-119.

В некоторых вариантах осуществления ортологичный промотор H1 содержит нуклеотидные последовательности, указанные в группе последовательностей, состоящей из SEQ ID NO: 84-119.

В некоторых вариантах осуществления промотор H1 является двунаправленным. Промотор H1 представляет собой промотор как pol

II, так и pol III

В некоторых вариантах осуществления промотор H1 содержит: a) элементы контроля, которые обеспечивают транскрипцию в одном направлении по меньшей мере одной нуклеотидной последовательности, кодирующей гРНК; и b) элементы контроля, которые обеспечивают транскрипцию в противоположном направлении нуклеотидной последовательности, кодирующей нацеленную на геном нуклеазу.

В некоторых вариантах осуществления промотор представляет собой промотор 7sk.

В некоторых вариантах осуществления промотор 7sk происходит из человека, крысы или мыши.

В некоторых вариантах осуществления промотор 7sk выбран из группы, состоящей из 7sk1, 7sk2 и 7sk3.

В некоторых вариантах осуществления промотор 7sk представляет собой 7sk1.

В некоторых вариантах осуществления 7sk1 происходит из человека.

В некоторых вариантах осуществления 7sk1 человека содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 3.

В некоторых вариантах осуществления 7sk1 человека содержит нуклеотидную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 3.

В некоторых вариантах осуществления 7sk1 происходит из мыши.

В некоторых вариантах осуществления 7sk1 мыши содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с SEQ ID NO: 6.

В некоторых вариантах осуществления 7sk1 мыши содержит нуклеотидную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 6.

В некоторых вариантах осуществления промотор 7sk представляет собой 7sk2.

В некоторых вариантах осуществления 7sk2 происходит из

человека.

В некоторых вариантах осуществления 7sk2 человека содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с SEQ ID NO: 4.

В некоторых вариантах осуществления 7sk2 человека содержит нуклеотидную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 4.

В некоторых вариантах осуществления 7sk2 происходит из мыши.

В некоторых вариантах осуществления 7sk2 мыши содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичность с SEQ ID NO: 7.

В некоторых вариантах осуществления 7sk2 мыши содержит нуклеотидную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 7.

В некоторых вариантах осуществления промотор 7sk представляет собой 7sk3.

В некоторых вариантах осуществления 7sk3 происходит из человека.

В некоторых вариантах осуществления 7sk3 человека содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с SEQ ID NO: 5.

В некоторых вариантах осуществления 7sk3 человека содержит нуклеотидную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 5.

В некоторых вариантах осуществления 7sk3 происходит из мыши.

В некоторых вариантах осуществления 7sk3 мыши содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с SEQ ID NO: 8.

В некоторых вариантах осуществления 7sk3 мыши содержит нуклеотидную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 8.

В некоторых вариантах осуществления промотор имеет по меньшей мере одну модификацию.

В некоторых вариантах осуществления по меньшей мере одна

модификация промотора включает элемент, который позволяет зависимую от условий регуляцию.

В некоторых вариантах осуществления элемент представляет собой отвечающий на tet промотор.

В некоторых вариантах осуществления отвечающий на tet промотор содержит Tet-репрессор (TetR) и Tet-оператор (TetO), встроенные в промотор H1.

В некоторых вариантах осуществления TetR содержит аминокислотную последовательность, обладающую по меньшей мере 80% идентичностью с SEQ ID NO: 14.

В некоторых вариантах осуществления TetR содержит аминокислотную последовательность, обладающую аминокислотной последовательностью, указанной в SEQ ID NO: 14.

В некоторых вариантах осуществления TetO содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 11.

В некоторых вариантах осуществления TetO содержит нуклеотидную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 11.

В некоторых вариантах осуществления H1-TetO содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 13.

В некоторых вариантах осуществления H1-TetO содержит нуклеотидную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 13.

В некоторых вариантах осуществления по меньшей мере одна модификация промотора включает область, которая позволяет ауторегуляцию.

В некоторых вариантах осуществления область ауторегуляции содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 33 или 34.

В некоторых вариантах осуществления область ауторегуляции содержит нуклеотидную последовательность, указанную в SEQ ID NO:

33.

В некоторых вариантах осуществления область ауторегуляции содержит нуклеотидную последовательность, указанную в SEQ ID NO: 34.

В некоторых вариантах осуществления способ дополнительно включает по меньшей мере одну последовательность терминатора.

В некоторых вариантах осуществления по меньшей мере одна последовательность терминатора выбрана из группы, состоящей из последовательности SV40 или синтетических поли A (SPA) последовательностей.

В некоторых вариантах осуществления последовательность терминатора представляет собой последовательность SV40 из 120 (SEQ ID NO: 123) или 240 (SEQ ID NO: 122) пар оснований.

В некоторых вариантах осуществления последовательность терминатора представляет собой последовательность SPA из 49 пар оснований.

В некоторых вариантах осуществления способ дополнительно включает последовательности 5′-нетранслируемой области (5′UTR).

В некоторых вариантах осуществления способ дополнительно включает последовательность Козак.

В некоторых вариантах осуществления последовательность Козак содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с SEQ ID NO: 1 или 2.

В некоторых вариантах осуществления последовательность Козак содержит нуклеотидную последовательность, указанную в SEQ ID NO: 1.

В некоторых вариантах осуществления последовательность Козак содержит нуклеотидную последовательность, указанную в SEQ ID NO: 2.

В некоторых вариантах осуществления способ дополнительно включает последовательность РНК, которая опосредует кэп-независимую инициацию трансляции.

В некоторых вариантах осуществления последовательность РНК выбрана из группы, состоящей из 6.947 или 6.967.

В некоторых вариантах осуществления нацеленная на геном

нуклеаза представляет собой белок Cas9.

В некоторых вариантах осуществления Cas9 содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 61.

В некоторых вариантах осуществления Cas9 содержит нуклеотидную последовательность, указанную в SEQ ID NO: 61.

В некоторых вариантах осуществления Cas9 содержит по меньшей мере одну модификацию.

В некоторых вариантах осуществления по меньшей мере одна модификация в Cas9 содержит изменение расщепляющей последовательности.

В некоторых вариантах осуществления изменение расщепляющей последовательности выбрано из группы, состоящей из T2A, P2A, E2A и F2A.

В некоторых вариантах осуществления T2A содержит аминокислотную последовательность, обладающую по меньшей мере 80% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 36.

В некоторых вариантах осуществления T2A содержит аминокислотную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 36.

В некоторых вариантах осуществления P2A содержит аминокислотную последовательность, обладающую по меньшей мере 80% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 35.

В некоторых вариантах осуществления P2A содержит аминокислотную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 35.

В некоторых вариантах осуществления E2A содержит аминокислотную последовательность, обладающую по меньшей мере 80% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 37.

В некоторых вариантах осуществления E2A содержит аминокислотную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 37.

В некоторых вариантах осуществления F2A содержит аминокислотную последовательность, обладающую по меньшей мере 80% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 38.

В некоторых вариантах осуществления F2A содержит аминокислотную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 38.

В некоторых вариантах осуществления по меньшей мере одна модификация в Cas9 является кодон-оптимизированной для экспрессии в клетке.

В некоторых вариантах осуществления Cas9 дополнительно содержит линкерную последовательность, функционально слитую в рамке считывания с зависимым от клеточного цикла белком (слитая конструкция Cas9).

В некоторых вариантах осуществления зависимый от клеточного цикла белок выбран из группы, состоящей из APE2, ATR, BRCA1, Chk1, Cdc5, Cdc6, Cdc7, Cdc45, Cdt1, CSA, CSB, Ctf18, DDB1, DDB2, DNA2, DUT, Elg1, EndoV, Esp1, экзонуклеазы 1, FBH1, FEN1, геминина, Hus1, KNTC2 (NDC80), Ku80, лигазы 1, Mad2, MBD4, Mcm3, Mcm4, Mcm5, Mcm6, Mcm7, Mcm8, Mcm10, MGMT, MLH3, Mms4, MPG, MSH2, Mus81, NBS1, NEIL2, NEIL3, NTH1, Orc1, Orc3, PARP1, PCNA, Pif1, PMS1, PMS2, PNK, Pola p180, Pola p70, Pola Spp1 (Prim2a), Polb, Pold p125, Pole Dpb3, Pole Dpb4, Pole Pol2, Poli, Poll, Polm, Psf1, Psf2, Psf3, Rad1, Rad18, Rad23A, Rad23B, Rad51, Rad51D, Rad54, Rad6A, RPA34, RPA70, Scc1, Scc3, Sir2, SIRT1 (сиртуин), TDG, TDP1, TIMELESS, Tin2, топоизомеразы I, топоизомеразы IIIa, топоизомеразы IIIb, Ubc13, UNG, XAB2, XPC, XPF, XPG, Xrcc2 и XRCC4.

В некоторых вариантах осуществления зависимый от клеточного цикла белок представляет собой геминин.

В некоторых вариантах осуществления зависимый от клеточного цикла белок представляет собой геминин человека.

В некоторых вариантах осуществления геминин человека содержит аминокислоты из положений 1-110 (hGem(1-110)).

В некоторых вариантах осуществления hGem(1-110) содержит аминокислотную последовательность, обладающую по меньшей мере

80% идентичностью с аминокислотной последовательностью, указанной в SEQ ID NO: 19.

В некоторых вариантах осуществления hGem(1-110) содержит аминокислотную последовательность, указанную в SEQ ID NO: 19.

В некоторых вариантах осуществления зависимый от клеточного цикла белок представляет собой Cdt1.

В некоторых вариантах осуществления зависимый от клеточного цикла белок представляет собой Cdt1 человека.

В некоторых вариантах осуществления Cdt1 человека содержит аминокислоты из положений 30-120 (hCdt1(30-120)).

В некоторых вариантах осуществления hCdt1(30-120) содержит аминокислотную последовательность, обладающую по меньшей мере 80% идентичностью с аминокислотной последовательностью, указанной в SEQ ID NO: 18.

В некоторых вариантах осуществления hCdt1(30-120) содержит аминокислотную последовательность, указанную в SEQ ID NO: 18.

В некоторых вариантах осуществления слитая конструкция Cas9 содержит аминокислотную последовательность, обладающую по меньшей мере 80% идентичностью с аминокислотной последовательностью, выбранной из группы, состоящей из SEQ ID NO: 20-27.

В некоторых вариантах осуществления слитая конструкция Cas9 содержит аминокислотную последовательность, выбранную из группы, состоящей из SEQ ID NO: 20-27.

В некоторых вариантах осуществления Cas9 функционально слита в рамке считывания с белком убиквитином (Ub-Cas9).

В некоторых вариантах осуществления Ub-Cas9 содержит по меньшей мере одну N-концевую модификацию.

В некоторых вариантах осуществления модифицированная на N-конце конструкция Ub-Cas9 содержит аминокислотную последовательность, обладающую по меньшей мере 80% идентичностью с аминокислотной последовательностью, выбранной из группы, состоящей из SEQ ID NO: 39-58.

В некоторых вариантах осуществления модифицированная на N-конце конструкция Ub-Cas9 содержит аминокислотную последовательность, выбранную из группы, состоящей из SEQ ID NO:

39-58.

В некоторых вариантах осуществления белок убиквитин содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 60.

В некоторых вариантах осуществления белок убиквитин содержит нуклеотидную последовательность, указанную в SEQ ID NO: 60.

В некоторых вариантах осуществления способ дополнительно включает никазу SaCas9.

В некоторых вариантах осуществления способ дополнительно включаета донорную матричную последовательность.

В некоторых вариантах осуществления по меньшей мере один вектор содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 62 или 67.

В некоторых вариантах осуществления по меньшей мере один вектор содержит нуклеотидную последовательность, указанную в SEQ ID NO: 62.

В некоторых вариантах осуществления по меньшей мере один вектор содержит нуклеотидную последовательность, указанную в SEQ ID NO: 67.

В некоторых вариантах осуществления донорная матричная последовательность корректирует по меньшей мере одну мутацию гена.

В некоторых вариантах осуществления по меньшей мере одна мутация гена представляет собой *rd10* или *rd12*.

В некоторых вариантах осуществления по меньшей мере один вектор содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 63 или 64.

В некоторых вариантах осуществления по меньшей мере один вектор содержит нуклеотидную последовательность, указанную в SEQ ID NO: 63.

В некоторых вариантах осуществления по меньшей мере один вектор содержит нуклеотидную последовательность, указанную в SEQ

ID NO: 64.

В некоторых вариантах осуществления по меньшей мере один вектор содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 65 или 66.

В некоторых вариантах осуществления по меньшей мере один вектор содержит нуклеотидную последовательность, указанную в SEQ ID NO: 65.

В некоторых вариантах осуществления по меньшей мере один вектор содержит нуклеотидную последовательность, указанную в SEQ ID NO: 66.

В некоторых вариантах осуществления промотор функционально связан по меньшей мере с одной, двумя, тремя, четырьмя, пятью, шестью, семью, восемью, девятью или десятью гРНК.

В некоторых вариантах осуществления последовательность-мишень содержит нуклеотидную последовательность $AN_{19}NGG$, $GN_{19}NGG$, $CN_{19}NGG$ или $TN_{19}NGG$.

В некоторых вариантах осуществления клетка представляет собой эукариотическую или неэукариотическую клетку.

В некоторых вариантах осуществления эукариотическая клетка представляет собой клетку млекопитающего или человека.

В некоторых вариантах осуществления эукариотическая клетка представляет собой фоторецепторную клетку сетчатки.

В некоторых вариантах осуществления один или несколько продуктов генов представляют собой родопсин.

В некоторых вариантах осуществления экспрессия одного или нескольких продуктов генов снижается.

Другой аспект изобретения относится к способу изменения экспрессии одного или нескольких продуктов генов в клетке, где клетка содержит молекулу ДНК, кодирующую один или несколько продуктов генов, причем способ включает введение в клетку не встречающейся в природе нуклеазной системы, как указано выше.

В некоторых вариантах осуществления система инактивирует один или несколько продуктов генов.

В некоторых вариантах осуществления нуклеазная система вырезает по меньшей мере одну мутацию гена.

В некоторых вариантах осуществления экспрессия одного или нескольких продуктов генов снижается.

В некоторых вариантах осуществления клетка представляет собой эукариотическую или неэукариотическую клетку.

В некоторых вариантах осуществления эукариотическая клетка представляет собой клетку млекопитающего или человека.

В некоторых вариантах осуществления клетка представляет собой фоторецепторную клетку сетчатки.

В некоторых вариантах осуществления клетка представляет собой ганглионарную клетку сетчатки.

В некоторых вариантах осуществления эукариотическая клетка представляет собой злокачественную клетку.

В некоторых вариантах осуществления пролиферация злокачественной клетки ингибируется или снижается.

В некоторых вариантах осуществления апоптоз злокачественной клетки усиливается или повышается.

Другой аспект изобретения относится к способу лечения заболевания, выбранного из группы, состоящей из дистрофии сетчатки, дистрофии роговицы, болезни экспансии микросателлитов и злокачественной опухоли у индивидуума, нуждающегося в этом, причем способ включает: (a) предоставление не встречающейся в природе нуклеазной системы, как указано выше; и (b) введение индивидууму эффективного количества системы.

В некоторых вариантах осуществления заболевание представляет собой дистрофию роговицы.

В некоторых вариантах осуществления заболевание представляет собой дистрофию сетчатки.

В некоторых вариантах осуществления дистрофия сетчатки выбрана из группы, состоящей из врожденного амавроза Лебера (LCA), пигментного ретинита (RP) и глаукомы.

В некоторых вариантах осуществления дистрофия роговицы выбрана из группы, состоящей из дистрофии эпителия базальной мембраны роговицы, эпителиальной рецидивирующей эрозивной дистрофии, субэпителиальной муцинозной дистрофии роговицы, дистрофии роговицы Месманна, эпителиальной дистрофии роговицы Лиша, желатинозной каплевидной дистрофии роговицы, дистрофии

роговицы Рейса-Бюклерса, дистрофии роговицы Тиля-Бенке, решетчатой дистрофии роговицы 1 типа (классическая), решетчатой дистрофии роговицы 2 типа, решетчатой дистрофии роговицы типа III, решетчатой дистрофии роговицы типа IIIA, решетчатой дистрофии роговицы типа I/IIIA, решетчатой дистрофии роговицы типа IV, полиморфного амилоидоза (роговицы), гранулярной дистрофии роговицы 1 типа, гранулярной дистрофии роговицы 2 типа, макулярной дистрофии роговицы, дистрофии роговицы Шнайдера, врожденной стромальной дистрофии роговицы, дистрофии роговицы Флека, задней аморфной дистрофии роговицы, центральной облаковидной дистрофии Франсуа, предесцеметовой дистрофии роговицы, эндотелиальной дистрофии роговицы Фукса, задней полиморфной дистрофии роговицы, врожденной наследственной эндотелиальной дистрофии и X-сцепленной эндотелиальной дистрофии роговицы.

В некоторых вариантах осуществления болезни экспансии микросателлитов выбраны из группы, состоящей из блефарофимоза, птоза и эпикантуса-инверсной синдактилии, ключично-черепной дисплазии, центрального врожденного гиповентиляционного синдрома, синдрома DM (миотоническая дистрофия) Хаддада, FRAXA (синдром ломкой X-хромосомы), FRAXE (задержка умственного развития, обусловленная ломкой XE), FRDA (атаксия Фридрейха), эндотелиальной дистрофии роговицы Фукса, FXTAS (ассоциированный с ломкой X синдром тремора/атаксии), ладонно-подошвенно-генитального синдрома, HD (болезнь Гентингтона), голопрозэнцефалии, задержки умственного развития с дефицитом гормонов роста, задержки умственного развития, эпилепсии, синдрома веста, синдрома Партингтона, окулофарингеальной мышечной дистрофии, SBMA (спинальная и бульбарная мышечная атрофия), SCA1 (спинально-церебеллярная атаксия 1 типа), SCA12 (спинально-церебеллярная атаксия 12 типа), SCA17 (спинально-церебеллярная атаксия 17 типа), SCA2 (спинально-церебеллярная атаксия 2 типа), SCA3 (спинально-церебеллярная атаксия 3 типа или болезнь Мачадо-Джозефа), SCA6 (спинально-церебеллярная атаксия 6 типа), SCA7 (спинально-церебеллярная атаксия 7 типа), SCA8 (спинально-церебеллярная атаксия 8 типа) и синполидактилии.

В некоторых вариантах осуществления введение индивидууму происходит посредством имплантации, инъекции или вирусным путем.

В некоторых вариантах осуществления введение индивидууму осуществляется посредством субретинальной инъекции.

В некоторых вариантах осуществления введение индивидууму осуществляется местным путем, внутрисосудистым путем, внутрикожным путем, трансдермально, парентерально, внутривенно, внутримышечно, интраназально, подкожно, регионарным путем, подкожно, интратрахеальным путем, внутрибрюшинно, интраартериально, внутрь мочевого пузыря, внутрь опухоли, перитуморальным путем, ингаляционным путем, системно, путем перфузии, посредством промывания, непосредственно посредством инъекции или пероральным введением состава.

В некоторых вариантах осуществления введение индивидууму осуществляется местным путем на поверхность глаза.

В некоторых вариантах осуществления введение индивидууму осуществляется на и снаружи роговицы, склеры, во внутриглазное, субконъюнктивальное, субтеноново или ретробульбарное пространство, или в или вокруг век.

В некоторых вариантах осуществления индивидуума лечат по меньшей мере одним дополнительным средством против злокачественной опухоли.

В некоторых вариантах осуществления средство против злокачественной опухоли выбрано из группы, состоящей из паклитаксела, цисплатина, топотекана, гемцитабина, блеомицина, этопозида, карбоплатина, доцетаксела, доксорубицина, топотекана, циклофосфамида, трабектедина, олапариба, тамоксифена, лектрозола и бевацизумаба.

В некоторых вариантах осуществления индивидуума лечат по меньшей мере одной дополнительной терапией против злокачественной опухоли.

В некоторых вариантах осуществления терапия против злокачественной опухоли представляет собой лучевую терапию, химиотерапию или хирургическую операцию.

В некоторых вариантах осуществления злокачественная опухоль представляет собой солидную опухоль.

В некоторых вариантах осуществления злокачественная опухоль выбрана из группы, состоящей из злокачественной опухоли головного мозга, желудочно-кишечной злокачественной опухоли, рака полости рта, рака молочной железы, рака яичника, рака предстательной железы, рака поджелудочной железы, рака легкого, рака печени, рака горла, рака желудка и рака почки.

В некоторых вариантах осуществления злокачественная опухоль представляет собой злокачественную опухоль головного мозга.

В некоторых вариантах осуществления системное введение выбрано из группы, состоящей из перорального, внутривенного, внутрикожного, внутрибрюшинного, подкожного и внутримышечного введения.

В некоторых вариантах осуществления индивидуумом является млекопитающее.

В некоторых вариантах осуществления млекопитающим является человек.

Выше в настоящем описании приведены определенные аспекты описанного в настоящем описании изобретения, которые полностью или частично направлены на описанное в настоящем описании изобретение, другие аспекты станут понятными из продолжения описания совместно с прилагаемыми примерами и чертежами, как описано в настоящем описании ниже.

КРАТКОЕ ОПИСАНИЕ ЧЕРТЕЖЕЙ

После описания описанных настоящего изобретения в общих терминах, ниже упоминаются прилагаемые чертежи, которые не обязательно изображены в масштабе, и где:

Фиг.1 содержит шесть панелей, A-F, демонстрирующих эффекты последовательностей 5'UTR на экспрессию Pol II под контролем H1. На **фиг.1A-1C** представлены эффекты 5'UTR и энхансера трансляции на экспрессию Pol II под контролем H1. На **фиг.1D** и **1E** представлены эффекты последовательностей Козак 5'UTR на экспрессию Pol II под контролем H1. На фиг.1F представлены последовательности терминаторов. Функциональные терминаторы были примерно эквивалентными. Наиболее коротким эффективным терминатором является терминатор SPA из 49 п.н.

Фиг.2 содержит три панели, A, B и C, демонстрирующих модулирование двунаправленной экспрессии с использованием различных ортологичных последовательностей.

Фиг.3 содержит две панели, A и B, демонстрирующих выравнивание различных ортологичных последовательностей H1.

Фиг.4 содержит две панели, A и B, демонстрирующих скрининг мутантов Pol II и Pol III под контролем двунаправленных промоторов.

Фиг.5 демонстрирует дальнейшее выравнивание различных ортологичных последовательностей H1.

Фиг.6 содержит две панели, A и B, демонстрирующих зависимую от условий экспрессию Pol III под контролем H1 с использованием TetO, регулирующую ферментативную активность рибонуклеопротеина. На **фиг.6A** представлена схема сконструированных областей TetO в промоторе H1. На **фиг.6B** показано, что экспрессия Pol II под контролем сконструированного H1 не изменялась в значительной степени в результате присутствия последовательностей оператора Tet.

Фиг.7 содержит две панели, A и B, демонстрирующих зависимую от условий экспрессию Pol III под контролем H1, регулирующую ферментативную активность рибонуклеопротеина. На **фиг.7A** показано, что экспрессия Pol III подавляется в присутствии TetR и в отсутствие доксициклина. На **фиг.7B** показано, что экспрессия Pol III дерепрессируется в присутствии TetR и в отсутствие доксициклина.

Фиг.8 содержит две панели, A и B, демонстрирующих зависимую от условий экспрессию Pol III под контролем H1, регулирующую ферментативную активность рибонуклеопротеина. На **фиг.8A** представлен активный ферментный комплекс в присутствии доксициклина. На **фиг.8B** представлен неактивный ферментный комплекс в отсутствии доксициклина.

Фиг.9 содержит три панели, A, B и C, демонстрирующих схему клеточного ответа на разрывы ДНК. На **фиг.9A** показано, что клеточный ответ на разрывы ДНК модулируется клеточным циклом. На **фиг.9B** показано, что фаза клеточного цикла в значительной степени определяет выбор каскада ДНК. На **фиг.9C** показано, что

РНК-направляемые нуклеазы могут быть слиты либо с доменом hGEM1, либо с доменом hCDT1, что приводит к зависимой от клеточного цикла регуляции этих белков.

Фиг.10 содержит три панели, A, B и C. На **фиг.10A** представлена регулируемая клеточным циклом конструкция, которая является активной в ходе фазы S, G2 и M (сверху), и регулируемая клеточная циклом конструкция, которая является активной в ходе фазы G1 (снизу). На **фиг.10B** показана олигонуклеотидная последовательность, использованная для нокина участка EcoRI в ген родопсина и положение мишени CRISPR и участка разрезания. На **фиг.1°C** показана последовательность после успешного встраивания посредством HDR в ген родопсина и включения участка EcoRI в ген.

Фиг.11 содержит три панели, A, B и C. На **фиг.11A** представлено количественное определение NHEJ с использованием анализа T7 EndoI для различных регулируемых клеточным циклом конструкций и с различными количествами донорной олигонуклеотидной последовательности. На **фиг.11B** представлено количественное определение HDR с использованием анализа EcoRI для различных регулируемых клеточным циклом конструкций и с различными количествами донорной олигонуклеотидной последовательности. На **фиг.11C** представлено количественное определение данных NHEJ и HDR. Конструкции Gem предпочтительно индуцируют HDR, в то время как конструкции Cdt предпочтительно индуцируют NHEJ. Смесь обоих из них (gem/cdt) является сходной с Cas9 без слияния, и общая скорость NHEJ и HDR является приблизительно равной среди всех условий, как и ожидалось.

Фиг.12 содержит две панели, A и B, демонстрирующих ауторегуляцию РНК-направляемых нуклеаз с использованием частичных областей-мишеней. На **фиг.12A** представлено, что Cas9 с полной комплементарностью с участком-мишенью приводит к расщеплению ДНК (белые стрелки). На **фиг.12B** показано, что Cas9 с высоким спариванием оснований, но без комплементарности, все еще связывается, однако отсутствует активность расщепления ДНК.

Фиг.13 содержит три панели, A, B и C, демонстрирующих ауторегуляцию РНК-направляемых нуклеаз с использованием

частичных областей-мишеней. На **фиг.13A** показано связывание полимеразы с последовательностью-мишенью из 20 оснований. На **фиг.12B** показано, что Cas9 (версия Cas9 с убитой нуклеазой) связывается с ДНК, но не расщепляет ее. На **фиг.13C** показано, что Cas9 с высоким спариванием оснований, но без комплементарности, все еще связывается, однако отсутствует активность расщепления ДНК.

На **фиг.14** представлена ауторегуляция РНК-наплавленных нуклеаз с использованием частичных участков-мишеней. Полная комплементарность в желаемой области приводит к расщеплению ДНК (белые стрелки) (левое нижнее изображение). Неполная комплементарность в сконструированной промоторной области приводит к отсутствию расщепления ДНК и подавлению ее собственной экспрессии (правое нижнее изображение).

На **фиг.15** представлена ауторегуляция РНК-направленных нуклеаз с использованием частичных участков-мишеней: GFP-репортер. На столбиковой диаграмме показано, что присутствие частичных последовательностей-мишеней не влияет в значительной степени на экспрессию Pol II со сконструированного H1.

На **фиг.16** представлена ауторегуляция РНК-направленных нуклеаз с использованием частичных участков-мишеней: Cas9 и гРНК.

Фиг.17 содержит две панели, A и B. На **фиг.17A** показано, что метионин (Met), за которым следуют глицин (Gly), пролин (Pro), серин (Ser), треонин (Thr), аланин (Ala), валин (Val) или цистеин (Cys), процессируются метионинаминопептидазами, которые отщепляют N-концевой метионин. На **фиг.17B** показано, что правило N-конца позволяет оценить время полужизни белков в клетке.

На **фиг.18** показано, что линейный полиубиквитин распознается в клетке деубиквитинизирующими ферментами, которые отщепляют индивидуальные пептиды Ub. Этот процесс может быть использован для получения определенных N-концевых остатков путем слияния Ub с N-концом любого белка.

Фиг.19 содержит две панели, A и B, демонстрирующих, что простые изменения N-концевых аминокислот, посредством либо Met-аминопептидаз, либо деубиквитинилирования, или альтернативными

способами, могут регулировать время полужизни РНК-направленной нуклеазы. На **фиг.19A** и **19B** показано, что уровни Cas9, РНК-направленной нуклеазы, могут модулироваться типом N-концевой аминокислоты, вплоть до 8-10 раз, что является значительным диапазоном для экспрессии нуклеазы.

На **фиг.20** показано, как широко используемые пептиды 2A также могут быть использованы для модулирования уровней белка в клетке.

Фиг.21 содержит три панели, A-C, иллюстрирующих способность AAV к упаковыванию. AAV дикого типа имеет размер ~4,7 т.п.н. и рекомбинантный AAV может вместить вплоть до 5,2 т.п.н. На **фиг.21B** приведена иллюстрация размера вируса AAV с использованием двунаправленного промотора H1 для экспрессии SpCas9 и одной гРНК. На **фиг. 21C** представлена конструкция с использованием более сильного промотора мыши. Первый набор экспериментов проводили с использованием промотора человека. В конструкциях на **фиг.21B** терминатор SV40 был заменен меньшим терминатором SPA.

Фиг.22 содержит девять панелей, A-I. На **фиг.22A** представлена проблема доставки AAV. Размер ДНК, требуемой для экспрессии Cas9 и гРНК общепринятыми способами, превышает 5,2 т.п.н.: промотор Pol II (~0,5 т.п.н.), SpCas9 (~4,1 т.п.н.), терминатор Pol II (~0,2 т.п.н.), промотор U6 (~0,3 т.п.н.) и гРНК (~0,1 т.п.н.). На фиг **фиг.22B** показано, что Cas9 *S. aureus* меньше на ~1 т.п.н., чем Cas9 *S. pyogenes*, что позволяет упаковать ее в AAV. На **фиг.22C** представлен геномный локус H1, показывающий H1RNA (транскрипт pol III), экспрессируемый в прямой ориентации, и ген PARP-2 (транскрипт pol II), транскрибируемый в противоположном направлении. Эта последовательность размером ~200 п.н. представляет собой чрезвычайно компактный двунаправленный промотор. На **фиг.22D** представлено сокращение "инструкций" по доставке SpCas9, наиболее широко используемого и наиболее широко изученного белка Cas9. Важно, что белок SpCas9 может быть нацелен на значительно большее количество геномных участков, чем белок SaCas9. Это означает, что может быть осуществлено нацеливание на большее

количество мутаций/заболеваний. На **фиг.22E** представлено сокращение "инструкций" по доставке ряда других белков Cas9, для которых было показано, что они являются эффективными в отношении редактирования генома в эукариотических клетках. Также можно доставлять SaCas9 в значительно более компактном векторе, чем с использованием стандартных промоторов. На **фиг.22F** показано, что настоящее изобретение обладает потенциалом к значительно большему количеству геномных мишеней и, таким образом, к значительно большему количеству потенциальных мутаций и заболеваний. На **фиг.22G** приведена иллюстрация размера вируса AAV с использованием двунаправленного промотора H1 для экспрессии SaCas9 и одной гРНК. Приблизительно 1 т.п.н. пространства доступно для матрицы HDR. На **фиг.22H** представлена иллюстрация размера вируса AAV с использованием двунаправленного промотора H1 для экспрессии SaCas9, и одной гРНК, и матрицы. На **фиг.22I** показано следующее: 1. аденоассоциированный вирус (AAV) является наиболее безопасным и наиболее часто используемым вектором в генной терапии. Вирус имеет один недостаток: он очень маленький. 2. По своей природе он состоит из одноцепочечного ДНК-генома из 4700 нуклеотидов. Концы генома представляют собой инвертированные повторяющиеся последовательности, известные как ITR (оранжевый). 3. Эти последовательности (по ~150 п.н. каждая) являются единственными элементами, требуемыми для упаковывания ДНК в AAV, так что вирусные последовательности (*Rep* и *Cap*) могут быть устранены. 4. Любой груз (синий) может быть встроен между ITR и в вирус может быть встроено всего вплоть до 5200 нуклеотидов, что означает, что вирус может вместить груз вплоть до 4900 нуклеотидов. 5. Рекомбинантный аденоассоциированный вирус (rAAV) используют в качестве высокоэффективного носителя для доставки груза в клетку. После вхождения в клетку одноцепочечный геном становится двухцепочечным и остается внутри на протяжении жизни клетки.

На **фиг.23** представлен механизм сайт-специфической рекомбинации из одного вируса AAV, содержащего Cas9, одну гРНК и матрицу.

Фиг.24 содержит две панели, A и B. На **фиг.24A** представлены

потенциальные конфигурации вектора AAV для обеспечения HDR. На **фиг.24B** представлены потенциальные конфигурации вектора AAV для обеспечения HDR в интронной области РНК-направленной полимеразы.

Фиг.25 содержит две панели, A и B. На **фиг.25A** проиллюстрирован подход с никазой Cas9 с использованием SaCas9, и одной гРНК, и матрицы. На **фиг.25B** приведено изображение нацеливающей последовательности rd12 для коррекции рецессивной мутации RPE65.

Фиг.26 содержит две панели, A и B. На **фиг.26A** представлена иллюстрация размера вируса AAV с использованием двунаправленного промотора H1 для экспрессии SpCas9 и одной гРНК. Для матрицы HDR доступно приблизительно 0,7 т.п.н. пространства. На **фиг.26B** представлен общий подход для доставки РНК-направленных нуклеаз, гидовой РНК и матрицы в одном AAV.

На **фиг.27** представлена нацеливающая последовательность rd10 для коррекции рецессивной мутации PDE6b.

На **фиг.28** представлен клонирующий вектор, который является легко адаптируемым путем встраивания матриц HDR.

Фиг.29 содержит две панели, A и B. На **фиг.29A** представлена нарисованная схема и таблица компактных двунаправленных промоторных последовательностей с активностью как pol II, так и pol III. На **фиг.29B** представлена столбиковая диаграмма, демонстрирующая экспрессию GFP посредством конструкций, содержащий компактные двунаправленные промоторные последовательности, представленные на **фиг.29A.** На **фиг.29B** представлены, начиная с оси y справа и двигаясь влево вдоль оси x, следующие обозначения столбиков, 1) отрицательный контроль, 2) hH1Koz (SEQ ID NO: 71 (или SEQ ID NO: 12 или 32)), 3) hSk1 (SEQ ID NO: 73), 4) biSRP (SEQ ID NO: 72), 5) biCGB1 (SEQ ID NO: 74), 6) biAloxE3 (SEQ ID NO: 77), 7) biDPP9 (SEQ ID NO: 79), 8) biTHEM (SEQ ID NO: 83) и 9) biORF (SEQ ID NO: 80).

Фиг.30 содержит две панели, A и B. На **фиг.30A** представлена нарисованная схема и таблица ортологичных двунаправленных промоторных последовательностей с активностью как pol II, так и pol III. На **фиг.30B** представлена столбиковая диаграмма, демонстрирующая экспрессию GFP с конструкций, содержащих

ортологичные двунаправленный промоторные последовательности, представленные на **фиг. 30A.** На **фиг.30B** представлены, начиная с оси y справа и двигаясь влево вдоль оси x, следующие обозначения столбиков, 1) отрицательный контроль, 2) hH1+ (соответствующий RPPH1-PARP2 человека), 3) mH1+ (соответствующий RPPH1-PARP2 мыши) и 4) rH1+ (соответствующий RPPH1-PARP2 крысы).

ПОДРОБНОЕ ОПИСАНИЕ

Описанное в настоящем описании изобретение описано более подробно ниже с отсылкой на прилагаемые чертежи, на которых показаны некоторые, но не все, варианты осуществления описанного в настоящем описании изобретения. Везде подобные цифры относятся к подобным элементам. Описанное в настоящем описании изобретение может быть осуществлено в множестве различных форм и его не следует считать ограниченным вариантами осуществления, описанными в настоящем описании; скорее, эти варианты осуществления предоставлены для того, чтобы настоящее описание удовлетворяло требованиям законодательства. Действительно, специалисту в области, к которой относится настоящее изобретение, станут очевидными многие модификации и другие варианты осуществления описанного в настоящем описании изобретения с использованием идей, представленных в описании ниже и на прилагаемых чертежах. Таким образом, следует понимать, что настоящее изобретение не ограничивается конкретными описанными вариантами осуществленияs и что подразумевается, что в объем прилагаемой формулы изобретения включены модификации и другие варианты осуществления.

Технологии редактирования генома, такие как нуклеазы с цинковыми пальцами (ZFN) (Porteus, and Baltimore (2003) *Science* 300: 763; Miller et al. (2007) *Nat. Biotechnol.* 25:778-785; Sander et al. (2011) *Nature Methods* 8:67-69; Wood et al. (2011) *Science* 333:307) and transcription activator-like effectors nucleases (TALEN) (Wood et al. (2011) *Science* 333:307; Boch et al. (2009) *Science* 326:1509-1512; Moscou and Bogdanove (2009) *Science* 326:1501; Christian et al. (2010) *Genetics* 186:757-761; Miller et al. (2011) *Nat. Biotechnol.* 29:143-148; Zhang et al. (2011) *Nat. Biotechnol.* 29:149-153; Reyon et al. (2012) *Nat.*

Biotechnol. 30:460-465) обеспечили возможность проведения направленных модификаций генома и имеют потенциал к точной коррекции обуславливающих заболевание мутаций. Являясь эффективными, эти технологии имеют практические ограничения, поскольку пары ZFN и TALEN требуют синтеза больших и уникальных распознающих белков для данного участка ДНК, являющегося мишенью. Недавно несколько групп сообщили о высокоэффективном редактировании генома с использованием сконструированной системы CRISPR/Cas9 типа II, которая преодолевает эти ключевые ограничения (Cong et al. (2013) *Science* 339:819-823; Jinek et al. (2013) *eLife* 2:e00471; Mali et al. (2013) *Science* 339:823-826; Cho et al. (2013) *Nat. Biotechnol.* 31:230-232; Hwang et al. (2013) *Nat. Biotechnol.* 31:227-229). В отличие от ZFN и TALEN, получение которых является относительно времязатратным и трудоемким, синтез конструкций CRISPR, которые основаны на нуклеазной активности белка Cas9, сопряженного с синтетической гидовой РНК (гРНК), является простым и быстрым, и они могут быть мультиплексными. Однако, несмотря на относительную простоту синтеза, CRISPR имеют технологические ограничения, связанные с их доступом к подлежащему нацеливанию геномному пространству, который зависит как от свойств самой Cas9, так и от синтеза ее гРНК.

Расщепление системой CRISPR требует комплементарного спаривания оснований гРНК с последовательностью ДНК из 20 нуклеотидов и соответствующего примыкающего к протоспейсеру мотива (PAM) - короткого нуклеотидного мотива, находящегося с 3'-стороны от области-мишени (Jinek et al. (2012) *Science* 337: 816-821). Теоретически, с использованием технологи CRISPR можно осуществлять нацеливание на любую уникальную последовательность N_{20}-PAM в геноме. Специфичность связывания последовательности PAM с ДНК, которая варьируется в зависимости от вида, из которого происходит конкретная используемая Cas9, имеет одно ограничение. В настоящее время имеющим наименьшие ограничения и наиболее часто используемым белком Cas9 является белок из *S. pyogenes*, который распознает последовательность NGG, и, таким образом, можно осуществлять нацеливание на любую уникальную

последовательность их 21 нуклеотида в геном, после которой следуют два гуанозиновых нуклеотида (N_{20}NGG). Расширение доступных областей для нацеливания, осуществляемого белковым компонентом, ограничивается открытием и применением новых белков Cas9 с измененными требованиями к PAM (Cong et al. (2013) *Science* 339: 819-823; Hou et al. (2013) *Proc. Natl. Acad. Sci. U.S.A.* 110(39):15644-9), или ожиданием получения новых вариантов Cas9 посредством мутагенеза или направленной эволюции. Второе технологическое ограничение системы CRISPR является результатом экспрессии гРНК, начинающейся с 5'-гуанозинового нуклеотида. Применение класса промоторов РНК полимеразы III типа III является особенно пригодным для экспрессии гРНК, поскольку эти короткие некодирующие транскрипты имеют строго определенные концы, и все необходимые для транскрипции элементы, за исключением нуклеотида 1+, содержатся в вышележащей промоторной области. Однако, поскольку обычно используемый промотор U6 требует гуанозинового нуклеотида для инициации транскрипции, использование промотора U6 далее ограничивает геномные области, подлежащие нацеливанию, до GN_{19}NGG (Mali et al. (2013) *Science* 339:823-826; Ding et al. (2013) *Cell Stem Cell* 12:393-394 Ranganathan, V *et al. Nature communications* 5, 4516 (2014)). Альтернативные подходы, такие как транскрипция *in vitro* с помощью промоторов T7, T3 или SP6, также могут потребовать инициирующего гуанозинового нуклеотида(ов) (Adhya et al. (1981) *Proc. Natl. Acad. Sci. U.S.A.* 78:147-151; Melton et al. (1984) *Nucleic Acids Res.* 12:7035-7056; Pleiss et al. (1998) *RNA* 4:1313-1317).

Описанное в настоящем описании изобретение относится к композициям и способам, включающим усовершенствования системы CRISPR/Cas9 (т.е. гидовой РНК CRISPR с использованием промотора H1; WO2015/19561, включенная в настоящее описание в качестве ссылки в полном объеме). Такая модифицированная система CRISPR/Cas9 может содержать модификации промоторной области H1.

В некоторых вариантах осуществления модифицированная система CRISPR/Cas9 включает усиление двунаправленной экспрессии

pol II под контролем H1 с использованием модификаций 5'UTR. В некоторых вариантах осуществления модифицированная система CRISPR/Cas9 включает модулирование двунаправленной экспрессии с использованием различных ортологичных последовательностей промотора H1. В некоторых вариантах осуществления модифицированная система CRISPR/Cas9 содержит новые компактные двунаправленные промоторные последовательности с активностью как pol II, так и pol III (например, 7sk, 5'UTR, консенсусные последовательности Козак или их комбинации).

В некоторых вариантах осуществления модифицированная система CRISPR/Cas9 включает зависимую от условий экспрессию pol II/pol III под контролем двунаправленного промотора (в том числе компактного и ортологичного промотора) (например, участки TetR и TetO), который может регулировать ферментативную активность рибонуклеопротеинов или РНК-направленных нуклеаз. В некоторых вариантах осуществления усовершенствование включает добавление донорной матричной последовательности для коррекции мутаций (например, направляемая гомологией репарация (HDR)).

В некоторых вариантах осуществления компактные двунаправленные промоторы включают, но не ограничиваются ими, RPPH1-PARP2 (человек), SRP-RPS29, 7sk1-GSTA4, SNAR-G-1-CGB1, SNAR-CGB2, RMRP-CCDC107, tRNA(Lys)-ALOXE3, RNU6-9-MED16: tRNA (Gly)-DPP9, RNU6-2-THEM259 или SNORD13-C8orf41.

В некоторых вариантах осуществления ортологичные двунаправленные промоторы включают, но не ограничиваются ими, RPPH1-PARP2 (мышь) или RPPH1-PARP2 (крыса), или промоторы, происходящие из *ailuropoda melanoleuca, bos taurus, callithrix jacchus, canis familiaris, cavia porcellus, chlorocebus sabaeus, choloepus hoffmanni, dasypus novemcinctus, dipodomys ordii, equus caballus, erinaceus europaeus, felis catus, gorilla gorilla, homo sapiens, ictidomys tridecemlineatus, loxodonta africana, macaca mulatta, mus musculus, mustela putorius furo, myotis lucifugus, nomascus leucogenys, ochotona princeps, oryctolagus cuniculus, otolemur garnettii, ovis aries, pan troglodytes, papio anubis, pongo abelii, procavia capensis,*

pteropus vampyrus, rattus norvegicus, sus scrofa, tarsius syrichta, tupaia belangeri, tursiops truncatus, vicugna pacos.

Таблица 3: Примеры компактных двунаправленных промоторов

Ген нкРНК	Ген белка	Расстояние
RPPH1	PARP2	230 п.н.
SRP	RPS29	233 п.н.
7sk1	GSTA4	239 п.н.
SNAR-G1	CGB1	308 п.н.
SNAR	CGB2	308 п.н.
RMRP	CCDC107	361 п.н.
тРНК(Lys)	ALOXE3	376 п.н.
RNU6-9	MED16	412 п.н.
тРНК (Gly)	DPP9	484 п.н.
SNORD13	C8orf41	847 п.н.
RNU6-2	THEM259	525 п.н.

RPPH1-PARP2 (человек):

GGAATTCGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCCCAGTGTCACTA GGCGGGAACACCCAGCGCGCGTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAGGGGAGTGGC GCCCTGCAATATTTGCATGTCGCTATGTGTTCTGGGAAATCACCATAAACGTGAAATGTCTTTG GATTTGGGAATCTTATAAGTTCTGTATGAGACCACTTTTTCCC (SEQ ID NO: 71)

SRP-RPS29:

CTTGCTCTCAGCAGTGCAACGAGGTAAAAGGAAGAAGCTGGCCCACGCATGCGCTCTTC AAATTTTTGAGACAGTTTACCCAGAATGCAGTGCTCAAAGGAAACGCGTGCGCAGTGTGGTCAG GTTGTTTCGCTGGGTGAGTAAAATGAAATCTTAGAGGCGTTGTGGGCTGGCCCAGTTGATGACG TCACCATACCACAGCTTCTAGTGCTATTCTGCGCCGGTATCCGACC (SEQ ID NO: 72)

7sk1-GSTA4:

AGTATTTAGCATGCCCCACCCATCTGCAAGGCATTCTGGATAGTGTCAAAACAGCCGGA AATCAAGTCCGTTTATCTCAAACTTTAGCATTTTGGGAATAAATGATATTTGCTATGCTGGTTA AATTAGATTTTAGTTAAATTTCCTGCTGAAGCTCTAGTACGATAAGCAACTTGACCTAAGTGTA AAGTTGAGACTTCCTTCAGGTTTATATAGCTTGTGCGCCGCTTGGGTACCTC (SEQ ID NO: 73)

SNAR-G-1-CGB1:

GTCTCTCTCTTAGCGGGATATCTTCCGCAAGCACTGGGAATGTGGACATGGAAAGTAAA TTGAGTCTCCGTGGGGGAGTGAGACAGGGAGTGAGGGGTGTTGGACGCGGCACGGGAACCTGGC CAGAGTCAGCGGACCCAATTGGCTGCTCTCTCTCAGATGCAGTTCCCCTTCCTCCCTCCAGGGG

GCGCCACGGAACGCAGGGCCCTCACTGGCCCTGGGGACTGGGTGACGTCAGGGATGAGCCTCTT
GTGATTGGCTCCATCACCCTGCGTAAGATCAAAGGGAAGAAAGGATGGGCCCGACAA (SEQ
ID NO: 74)

SNAR-CGB2:

GTCTCTCTCTTAGCGGGATATCTTCCGCAAGCACTGGGGATGTGGACATGGAAAGTAAA
TTGAGTCTCCGTGGGGGAGTGAGACAGGGAGTGAGGGGTGTTGGACGCGGCACGGGAACCCGGC
CGGAGTCAGCGGACCCAATTGGCTGCTCTCTCTCAGATACAGTTCCCCTTCCTCCCTCCAGGGG
GCGCCACGGAACGCAGGGCCCTCACTGGCCCTGGGGACTGGGTGACGTCAGGGGTGAGCCTCTC
CTGATTGGCTCCATCACCCTGCGTAAGGTCAAAAGGAAGAAAGGAGATCCCCGACAC (SEQ
ID NO: 75)

RMRP-CCDC107:

TGCCGGCCCACGGGTGGAGGGATCGGGCGGGCGGTGCCGAAGCGGTCCGGCATTGGCCG
GCCGCCCCAACGCGCACGCGCACGCGAGCAGGCCGGCCGGCTCCGGGGAGGCCACGCCCACTCC
CCGTAGGGCGGGGCCAGACCATATTTGCATAAGATAGTGTCATTCTAGCTTTCCTGTATTTGTT
CATTTCGTGTCTATTAGCTATTCTGCTAGCCACAATGCCTCTGAAAGCCTATAGTCTTAGAAAG
TTATGCCCGAAAACGGTTTTTTTAATCTCACGCCACCAACTTTCTCACCCTAATCATAAAACAC
AATTTCTTTAGGGCTATAAAATACTACTCTGTGAAGCTGAGGACGT (SEQ ID NO: 76)

tRNA(Lys)-ALOXE3:

TCTTTCCGCTCCAGGACCGCCCTGGGCCTGCAGGATCCTGGGCGGGAGCCCAGGTGTCC
GGGATCTGGGCCACTAGGGACTGGGGAGGAACCTCTCAGAGAAGCCCATAGCCCGCAGCGGCCC
CGCGCGGCCGGTTCCGGCGCCGCACTGTTCCAGCCTCTACTATGGTACAGTCCCTGCGTCGCAG
CCTCGGCGGGGGCTCTAAGAACGGGAGGCAGAAAAAGCTCAATCAGCAGCAGGCGAGCTTCACC
CGCTGCTTCCAAATCTGTGCCAAAATATTCTATGCTGCACAGATAAAATCCTCTGTCGGTTCTA
CAAGCCTGGCTTTTCCTATAGAGAACCCTCTTATAAGCAAAAAGTAAAGCTCTCGTGAAGA
(SEQ ID NO: 77)

RNU6-9-MED16:

GAGGGCAGTCACCAGCTCCTGGCCCGTGCGCCAAGCTCAGCGGGCGTCCGCGGTGCGAT
CTTCCCTAGCGCCTCGGGTCTGGCGCCGCCATCTTCCTCGGTAACAACCAGTCGCCTGAGGCGT
GGGGCCGCCTCCCAAAGACTTCTGGGAGGGCGGTGCGGCTCAGGCTCTGCCCCGCCTCCGGGGC
TATTTGCATACGACCATTTCCAGTAATTCCCAGCAGCCACCGTAGCTATATTTGGTAGAACAAC
GAGCACTTTCTCAACTCCAGTCAATAACTACGTTAGTTGCATTACACATTGGGCTAATATAAAT
AGAGGTTAAATCTCTAGGTCATTTAAGAGAAGTCGGCCTATGTGTACAGACATTTGTTCCAGGG
GCTTTAAATAGCTGGTGGTGGAACTCAATATTC (SEQ ID NO: 78)

tRNA(Gly)-DPP9:

TAACCGCTCAGCTGACCTCAGGAGGGCAGGGGTGCCTTCTAAAGGGTCCAGAGAGCCTC
CATTCCAGCTGCAGGCGTGGGACACAGACCGGGACGTGGGGCGGCGGCCGGACTGGGCAGGTCG

TCCCGGGTCCAGCGGCGCCTCACGGTCGCGGCTCCATGCCCGGGACTGCGACCCCGGAAGTGGC
GGGAGCGGGGGACGACAGCCGCGGCGGACACAGGGGACCCGCCGGCTCAGGCACCTTTGACCCG
GAAGTTGAGCGACCCAGGCGGCGGCCTGGGATTGGACACCACCAGGCACGTACCAAGGCGTCCG
CGGCGCTTGGGGGGGAGCCCGCGGCGCGGCGGCCTAAGGTGCGTAACGCCCCATGAACGACATC
TTCCGGTGGGTTAGGGAGAGACACCCCCCTGTGACTTGGTATCACTCAGTCAAACCCATGATCC
CCCACTATTAAGGATATCCGGAGAGGATGCTACCTATCAGG (SEQ ID NO: 79)

SNORD13-C8orf41:

TCCTGACTGCAGCACCAGAAGGCTGGTCTCTCCCACAGAACGAGGATGGAGGCGGGGAG
GGATCCGTTGAAGAGGGAAGGAGCGATCACCCAAAGAGAACTAAAATCAAATAAAATAAAACAG
AGAGATGTCTTGGAGGAGGGGGCGAGTCTGACCGGGATAAGAATAAAGAGAAAGGGTGAACCCG
GGAGGCGGAGTTTGCAGTGAGCCGAGATCGCGCCACTGCACTCCAGCCTGGGCGACAGAGTGAG
ACTCCGTCTCAGTAAAAAAAAAAAAAAAAAAAAGAATAAAGAGGAAAGGACGCAAGAAAGGGAA
AGGGGACTCTCAGGGAGTAAAAGAGTCTTACACTTTTAACAGTGACGTTAAAAGACTACTGTTG
CCTTTCTGAAGACTAAAAAGAAAAAAAACTTAAAAATTTAAAGAAATAAACTTCTGAGCCATGT
CACCAACTTAACCACCCCCAGGTACCTGCAACGGCTCGCGCCCGCCGGTGTCTAACAGGATCCG
GACCTAGCTCATATTGCTGCCGCAAAACGCAAGGCTAGCTTCCGCCAGTACTGCCGCAACACCT
TCTTATTTCACGACGTATGGTCGTAAAGCAATAAAGATCCAGGCTCGGGAAAATGACGGAGAGG
TGGAACTATAGAGAATAAATTTGCATATATAATAATCCGCTCGCTAATTGTGTTTCTGTTTTCC
TTTGCTAAGGTAGAAACAAAAGAATAATCACAGAATCTCAGTGGGACTTTGAAAATATCCAGGA
TTTTATACGTGAAGAATGGATGTATCGCATTACGGTAGTCACCCTATGTGTAAATTAGTGGCAC
ATACTTGGCACTCCTTAATGTCAACTATAAGATG (SEQ ID NO: 80)

RNU6-2-THEM259;

GCCTCCCAGCGTCGCGCCCTAACGACCCGCAAGTGTCCGAGGGCGCCTCCCGGCCGCCA
TCGGCCGCCCTCGCAGCCGCCGCTCTCCTCACGGCCTCCCGGCCGCCGCCGCCATCTTCCGCTT
TCTCGTCCGGCTGCGGCGCTGCTGACGCTAGCGAGTCGCCACGCCGGGCAAGAGCGGCCCCCCT
GCGCCCGCAGAGAACGCTGGGATGCCAGCGGCGCCCGCGGAGGCCTCACCCCCTACCTCGGCCG
CTCCAGGGGGCGGGCCTGCATCTGGGCCACCTCTTTTGCATATTGGCACCCACAATCCACCGCG
GCTATGAGGCCAGTATAAGGCGGTAAAATTACGATAAGATATGGGATTTTACGTGATCGAAGAC
ATCAAAGTAAGCGTAAGCACGAAAGTTGTTCTGCAACATACCACTGTAGGAAATTATGCTAAAT
ATGAAACCGACCATAAGTTATCCTAACCAAAAGATGATTTGATTGAAGGGCTTAAAATAGGTGT
GACAGTAACCCTTGAGTC (SEQ ID NO: 83)

Таблица 4: Примеры ортологичных двунаправленных промоторов

Ген нкРНК	**Ген белка**	**Расстояние**	**Организм**
RPPH1	PARP2	230 п.н.	Человек
RPPH1	PARP2	172 п.н.	Мышь

RPPH1	PARP2	201 п.н.	Крыса

RPPH1-PARP2 (мышь):

CGCTCTTGAAGGACGACGTCATCATCCCTTGCCCGGATGCGCGGGCTTCTTGTCTAGCA CAGGAGCCTGGGGTAGAGCGCATGCAAATTACGCGCTGTGCTTTGTGGGAAATCACCCTAAACG AAAAATTTATTCCTCTTTCGAGCCTTATAGTGGCGGCCGGTCTACATCC (SEQ ID NO: 81)

RPPH1-PARP2 (крыса):

GGCTGATGAGCTTCCCCCGCCCACTAGGAGTGTGAAGACCTGCCGCCATAATAAGACTC CAAAAGACAGTGAATTTAACACTTACGGTGACTTCCCACAAAGCACAGCGTGTAATTTGCATGC GCTCTAGCCCAGGCTCCAGCTCCGGACCAGAAGCCCGCGCATCCCGGCAAAGGGTGATGACGTC GTCCTTCAAGCGCT (SEQ ID NO: 82)

В некоторых вариантах осуществления проводят модификации нуклеаз (например, РНК-направленных нуклеаз). В некоторых вариантах осуществления нуклеазу (например, Cas9) модифицируют посредством посттранскрипционной регуляции клеточного цикла (например, слитые белки, содержащие геминин (Gem) или Cdt1). В некоторых вариантах осуществления нуклеазу (например, Cas9) модифицируют путем конструирования частичных участков-мишеней, чтобы нуклеаза могла связываться без расщепления ДНК. В некоторых вариантах осуществления нуклеазу (например, Cas9) модифицируют путем модулирования ее времени полужизни с использованием типа N-концевой аминокислоты.

Такая модифицированная система CRISPR/Cas9 может с точностью нацеливаться на геномные области или способствовать репарации дефектных геномных областей с большей эффективностью, безопасностью и точностью. Более того, эта модификация обеспечивает компактную систему CRISPR/Cas9, которая позволяет более высокоточное нацеливание на онкогены относительно существующих технологий CRISPR, TALEN или цинковых пальцев.

Другие аспекты изобретения относятся к способам, включающим скрининг экспрессии для дополнительных двунаправленных промоторов с активностью в отношении как РНК-pol II, так и РНК-pol III.

I. УСОВЕРШЕНСТВОВАНИЯ/МОДИФИКАЦИИ СИСТЕМЫ CRISPR/CAS9.

A. Композиции

В некоторых вариантах осуществления описанное в настоящем

описании изобретение относится к композициям, содержащим усовершенствования системы CRISPR/Cas9, описанной ранее в WO2015/195621 (включенной в настоящее описание в качестве ссылки в полном объеме). Такие усовершенствования включают не встречающуюся в природе нуклеазную систему (например, CRISPR-Cas9), содержащую один или несколько векторов, содержащих: a) промотор (например, ортологичный промотор H1 или 7sk), функционально связанный по меньшей мере с одной нуклеотидной последовательностью, кодирующей гидовую РНК (гРНК) нуклеазной системы, где гРНК гибридизуется с последовательностью молекулы ДНК, являющейся мишенью, в клетке, и где молекула ДНК кодирует один или несколько продуктов генов, экспрессируемых в клетке; и b) регуляторный элемент, функционирующий в клетке, функционально связанный с нуклеотидной последовательностью, кодирующей нацеленную на геном нуклеазу (например, модифицированный белок Cas9), где компоненты (a) и (b) расположены на одном и том же или на различных векторах системы, где гРНК осуществляет нацеливание на последовательность-мишень и гибридизуется с ней, и нуклеаза расщепляет молекулу ДНК, изменяя экспрессию одного или нескольких продуктов генов. В некоторых вариантах осуществления система упакована в одну частицу аденоассоциированного вируса (AAV). В некоторых вариантах осуществления система инактивирует один или несколько продуктов генов. В некоторых вариантах осуществления система вырезает по меньшей мере одну мутацию гена. В некоторых вариантах осуществления промотор H1 содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 12 или 32. В некоторых вариантах осуществления ортологичный промотор H1 происходит из мыши или крысы. В некоторых вариантах осуществления ортологичный промотор H1 содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 84-119. В некоторых вариантах осуществления ортологичный промотор H1 содержит

нуклеотидные последовательности из группы, состоящей из SEQ ID NO: 84-119. В некоторых вариантах осуществления промотор H1 содержит: a) элементы контроля, которые обеспечивают транскрипцию в одном направлении по меньшей мере одной нуклеотидной последовательности, кодирующей гРНК; и b) элементы контроля, которые обеспечивают транскрипцию в противоположном направлении нуклеотидной последовательности, кодирующей нацеленную на геном нуклеазу. В некоторых вариантах осуществления промотор представляет собой промотор 7sk человека, мыши или крысы. В некоторых вариантах осуществления 7sk выбран из группы, состоящей из 7sk1 (вариант 1 7sk, в котором используется эндогенная 5'UTR GSTA4), 7sk2 (вариант 2 7sk, в котором используется только последовательность Козак в 5'UTR) и 7sk3 (вариант 7sk, в котором используется 5'UTR бета-глобина). В некоторых вариантах осуществления 7sk1 человека содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 3. В некоторых вариантах осуществления 7sk1 мыши содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с SEQ ID NO: 6. В некоторых вариантах осуществления 7sk2 человека содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с SEQ ID NO: 4. В некоторых вариантах осуществления 7sk2 мыши содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с SEQ ID NO: 7. В некоторых вариантах осуществления 7sk3 человека содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с SEQ ID NO: 5. В некоторых вариантах осуществления 7sk3 мыши содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с SEQ ID NO: 8. В некоторых вариантах осуществления ортологичный промотор H1 или промотор 7sk имеет по меньшей мере одну модификацию. В некоторых

вариантах осуществления по меньшей мере одна модификация промотора содержит элемент, который позволяет зависимую от условий регуляцию. В некоторых вариантах осуществления элемент представляет собой отвечающий на tet промотор. В некоторых вариантах осуществления отвечающий на tet промотор содержит Tet-репрессор (TetR) и Tet-оператор (TetO) встроенный в промотор H1. В некоторых вариантах осуществления TetR содержит аминокислотную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с SEQ ID NO: 14. В некоторых вариантах осуществления TetO содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 11. В некоторых вариантах осуществления H1-TetO содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 13. В некоторых вариантах осуществления по меньшей мере одна модификация промотора включает участок, который позволяет ауторегуляцию. В некоторых вариантах осуществления участок ауторегуляции содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 33 или 34. В некоторых вариантах осуществления нуклеазная система дополнительно содержит по меньшей мере одну последовательность терминатора. В некоторых вариантах осуществления последовательность терминатора представляет собой любую последовательность нуклеиновой кислоты, выбранную из группы, состоящей из SEQ ID NO: 120-130. В некоторых вариантах осуществления по меньшей мере одна последовательность терминатора выбрана из группы, состоящей из последовательности SV40 или синтетических поли A (SPA) последовательностей. В некоторых вариантах осуществления последовательность терминатора представляет собой последовательность SV40 из 120 (SEQ ID NO: 123) или 240 (SEQ ID NO: 122) пар оснований. В некоторых вариантах осуществления последовательность терминатора

представляет собой последовательность SPA из 49 пар оснований. В некоторых вариантах осуществления нуклеазная система дополнительно содержит последовательности 5′-нетранслируемой области (5′UTR). В некоторых вариантах осуществления нуклеазная система дополнительно содержит последовательность Козак. В некоторых вариантах осуществления последовательность Козак содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с SEQ ID NO: 1 или 2. В некоторых вариантах осуществления нуклеазная система дополнительно содержит последовательность РНК, которая опосредует кэп-независимую инициацию трансляции. В некоторых вариантах осуществления последовательность РНК выбрана из группы, состоящей из 6.947 или 6.967 (Wellensiek I. (2013) *Nature Methods*, 10:747-750). В некоторых вариантах осуществления Cas9 содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 61. В некоторых вариантах осуществления Cas9 содержит по меньшей мере одну модификацию. В некоторых вариантах осуществления по меньшей мере одна модификация в Cas9 включает изменение расщепляющей последовательности. В некоторых вариантах осуществления нуклеаза выбрана из группы, состоящей из Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (также известной как Csn1 и Csx12), Cas10, Cas13a, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, Cpf1, C2c1, C2c2 и C2c3. В некоторых вариантах осуществления изменение в расщепляющей последовательности выбрано из группы, состоящей из T2A, P2A, E2A и F2A. В некоторых вариантах осуществления T2A содержит аминокислотную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 36. В некоторых вариантах осуществления P2A содержит аминокислотную последовательность, обладающую по меньшей мере 80%, 85%, 90%,

95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 35. В некоторых вариантах осуществления P2A содержит аминокислотную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 35. В некоторых вариантах осуществления E2A содержит аминокислотную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 37. В некоторых вариантах осуществления F2A содержит аминокислотную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 38. В некоторых вариантах осуществления по меньшей мере одна модификация в Cas9 является кодон-оптимизированной для экспрессии в клетке. В некоторых вариантах осуществления Cas9 дополнительно содержит линкерную последовательность, функционально слитую в рамке считывания зависимым от клеточного цикла белком (слитая конструкция Cas9). В некоторых вариантах осуществления зависимый от клеточного цикла белок выбран из группы, состоящей из APE2, ATR, 53BP1, BRCA1, Chk1, Cdc5, Cdc6, Cdc7, Cdc45, Cdt1, CSA, CSB, Ctf18, CtIP, DDB1, DDB2, DNA2, DUT, Elg1, EndoV, Esp1, экзонуклеазы 1, FBH1, FEN1, геминина, Hus1, KNTC2 (NDC80), Ku80, лигазы 1, Mad2, MBD4, Mcm3, Mcm4, Mcm5, Mcm6, Mcm7, Mcm8, Mcm10, MGMT, MLH3, Mms4, MPG, MSH2, Mus81, NBS1, NEIL2, NEIL3, NTH1, Orc1, Orc3, PARP1, PCNA, Pif1, Pin1, PMS1, PMS2, PNK, Pola p180, Pola p70, Pola Spp1 (Prim2a), Polb, Pold p125, Pole Dpb3, Pole Dpb4, Pole Pol2, Poli, Poll, Polm, Psf1, Psf2, Psf3, Rad1, Rad18, Rad23A, Rad23B, Rad51, Rad51D, Rad54, Rad6A, Rag1, Rag2, RPA34, RPA70, Scc1, Scc3, Sir2, SIRT1 (сиртуин), TDG, TDP1, TIMELESS, Tin2, топоизомеразы I, топоизомеразы IIIa, топоизомеразы IIIb, Ubc13, UNG, XAB2, XPC, XPF, XPG, Xrcc2 и XRCC4. В некоторых вариантах осуществления зависимый от клеточного цикла белок представляет собой геминин. В некоторых вариантах осуществления зависимый от клеточного цикла белок представляет собой геминин человека. В некоторых вариантах осуществления геминин человека содержит аминокислоты из

положений 1-110 (hGem(1-110)). В некоторых вариантах осуществления hGem(1-110) содержит аминокислотную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с аминокислотной последовательностью, указанной в SEQ ID NO: 19. В некоторых вариантах осуществления зависимый от клеточного цикла белок представляет собой Cdt1. В некоторых вариантах осуществления зависимый от клеточного цикла белок представляет собой Cdt1 человека. В некоторых вариантах осуществления Cdt1 человека содержит аминокислоты из положений 30-120 (hCdt1(30-120)). В некоторых вариантах осуществления hCdt1(30-120) содержит аминокислотную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с аминокислотной последовательностью, указанной в SEQ ID NO: 18. В некоторых вариантах осуществления слитая конструкция Cas9 содержит аминокислотную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с аминокислотной последовательностью, выбранной из группы, состоящей из SEQ ID NO: 20-27. В некоторых вариантах осуществления Cas9 функционально слита в рамке считывания с убиквитином (Ub-Cas9). В некоторых вариантах осуществления Ub-Cas9 содержит по меньшей мере одну N-концевую модификацию. В некоторых вариантах осуществления Ub-Cas9 с N-концевой модификацией содержит аминокислотную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с аминокислотной последовательностью, выбранной из группы, состоящей из SEQ ID NO: 39-58. В некоторых вариантах осуществления белок убиквитин содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 60. В некоторых вариантах осуществления нуклеазная система дополнительно содержит никазу SaCas9. В некоторых вариантах осуществления нуклеазная система дополнительно содержит донорную матричную последовательность. В некоторых вариантах осуществления по меньшей мере один вектор содержит нуклеотидную

последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 62 или 67. В некоторых вариантах осуществления донорная матричная последовательность корректирует по меньшей мере одну мутацию гена. В некоторых вариантах осуществления по меньшей мере одна мутация гена представляет собой *rd10* или *rd12*. В некоторых вариантах осуществления по меньшей мере один вектор содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 63 или 64. В некоторых вариантах осуществления по меньшей мере один вектор содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 65 или 66. В некоторых вариантах осуществления промотор функционально связан по меньшей мере с одной, двумя, тремя, четырьмя, пятью, шестью, семью, восемью, девятью или десятью гРНК. В некоторых вариантах осуществления последовательность-мишень содержит нуклеотидную последовательность $AN_{19}NGG$, $GN_{19}NGG$, $CN_{19}NGG$ или $TN_{19}NGG$. В некоторых вариантах осуществления клетка представляет собой эукариотическую или неэукариотическую клетку. В некоторых вариантах осуществления эукариотическая клетка представляет собой клетку млекопитающего или человека. В некоторых вариантах осуществления эукариотическая клетка представляет собой фоторецепторную клетку сетчатки. В некоторых вариантах осуществления один или несколько продуктов генов представляют собой родопсин. В некоторых вариантах осуществления экспрессия одного или нескольких продуктов генов снижается.

Таблица 5: Примеры ортологичных последовательностей H1

mus_musculus

TTCAGGATGTAGACCGGCCGCCACTATAAGGCTCGAAAGAGGAATAAATTTTTCGTTTAGGGTGATTTCCCACAAAGCACAGCGCGTAATTTGCATGCGCTCTACCCCAGGCTCCTGTGCTAGACAAGAAGCCCGCGCATCCGGGCAAGGGATGATGACGTCGTCCTTCAAGAGCG (SEQ ID NO: 84)

rattus_norvegicus

AGGAGTGTGAAGACCTGCCGCCATAATAAGACTCCAAAAGACAGTGAATTTAACACTTA
CGGTGACTTCCCACAAAGCACAGCGTGTAATTTGCATGCGCTCTAGCCCAGGCTCCAGCTCCGG
ACCAGAAGCCCGCGCATCCCGGCAAAGGGTGATGACGTCGTCCTTCAAGCGCT (SEQ ID
NO: 85)

dipodomys_ordii

AGGAAAGACTTCGCTGAGGCAGACTTTATAAGGCTCCCGCGCAGAAAGAAACTTTATAG
TTATGGTGATTTCCCACAAGCCACTGCGTCATGCAAATAAAGCAGGGTACGGCTTCCATGTACC
TTAAGGTTTTTTTCTAGGCCGCGTACGCTCTGCGTATTCAGCCACGTGACCCTGAGCCAGTGGT
TGTTGGGAGCACGTTGTGGACCTCTGCGTTTGGATTCC (SEQ ID NO: 86)

ictidomys_tridecemlineatus

GAAAGGGACTCCGCACAAGCAGAGTTTATAAGGCTCCCATCTGTACAGCCATTTCTCGG
TCATGGTAACTACCCACAACACACAGCGATATGCAAATATAGCAGAGCGTGTCTTCCCGCGCGC
GCCTGGTCGTCTCGGCGCCGGCGCGCTGCGTGGGGCGGAACTGTGACAGAGACCCTGCGATTCC
TGGGAGCTGGCTGATGACATCAGTGTCTAACCTCC (SEQ ID NO: 87)

cavia_porcellus

GAGAAAGAAAGGCTCAAACCTAGCCTTATAAGGCTCCCAAATGTCGGTATATTTTTTGG
TTATGGTGACTTCCCACAATGCATAGCGATATGTAGATATTGCCAGGAGTACCTCCCACTTCTG
GTCCTGTCAGCTCTTTTCTAGGACGCGCGCGCTGCAGGTTTCCAGCCTGTGATTGGGCCAGCAA
TTCCGGGAATGAATTGATGACGTCAGCGTTTGAATTCC (SEQ ID NO: 88)

ochotona_princeps

GGGGGAAGCTGGGCTCGATCAGCCTTTATAAAGCTCCAAAAACTCAAGACATTTTTCTG
TTACGGTGGCTTCCCACAGTACACAGCGACATGCAAATAGCTTGCCAATGAATTCGCGGACCGC
TTCCCGCCCCGGCGCAGGCGCGCGGACGCTGTCTCCCCTGGACGCGCGCTCGCGGTTCCCGGGA
GCTGGCTGATGACGTTCGGTCTCC (SEQ ID NO: 89)

oryctolagus_cuniculus

GGGGAGAGGTGGATCCGAACAGACTTTATAAAGCTCCGAAAGCCCAAGGCATCTTTCCC
TTACGGTAGCTTCCCACAAGACATAGCGACATGCAAATTTCTTGAAGTATGCTTCAGACGCGCT
TCTCGCCACAGCGCAAGCGCGCTGTGTGCTGACGCGGGAACGGGCCAGGGCGCGGTTCCCGGGA
GCGGGTTGATGACGTTAGATCTCC (SEQ ID NO: 90)

callithrix_jacchus

GAGGAAAAGTAGTCCCACAGACAACTTATAAGATTCCCATACCCTAAGACATTTCACGA
TTATGGTGACTTCCCAGAAGACACAGCGACATGCAAATATTGCAGGTCGTGTTTCGCCTGTCCC
TCACAGTCGTCTTCCTGCCAGGGCGCACGCGCGCTGGGTTTCCCGCCAACTGACGCTGGGCTCG
CGATTCCTTGGAGCGGGTTGATGACGTCAGCGTTTGAATTCC (SEQ ID NO: 91)

хлорcebus_sabaeus

GGGGAAGGGTGGTCCCTTACAGAACTTATAAGATTCCCAAACTCAAAGACATTTCACGT

TTATGGTGACTTCCCAGAAGACATAGCGACATGCAAATATTGCAGGGCGTCACACCCCTCTCCC
TCACAGTCATCTTCCTGCCAGGGCGCACGCGCGCTGGGTGTTCTCGCGTAGTGACACTGGGCCC
GCGATTCCTTGGAGCGGGTTGATGACGTCAGCGTTCGAATTCC (SEQ ID NO: 92)

macaca_mulatta

GGGGAAGGGTGGTCCCACACAGAACTTATAAGATTCCCATACTCAAAGACATTTCTCGT
TTATGGTGACTTCCCAGAAGACACAGCGACATGCAAATATTGTAGGGCGTCACACCCCTGTCCC
TCACAGTCATCTTCCTGCCAGGGCGCACGCGCGCTGGGTGTTCCCGCGTAGTGACACTGGGCCC
GCGATTCCTTGGAGCGGGTTGATGACGTCAGCGTTCGAATTCC (SEQ ID NO: 93)

papio_anubis

GGGGAAAGGTGGTACCATACAGAACTTATAAGATTCCCATACTCAAAGACATTTCACGA
TTATGGTGACTTCCCAGAAGACACAGCGACATGCAAATATTGTAGGGCGTCACACCCCCTGTCC
CTCACAGTCATCTTCCTGCCAGGGCGCACGCGCGCTGGGTGTTCCCGCGTAGTGACACTGGGCC
CGCGATTCCTTGGAGCGGGTTGATGACGTCAGCGTTCGAATTCC (SEQ ID NO: 94)

gorilla_gorilla

GGGAAAGGGTGGTCCCACACAGAACTTATAAGACTCCCATATCCAAAGACATTTCACGG
TTATGGTGATTTCCCAGAACACATAGCGACATGTAAATATTGCAGGGCGCCACTCCCCAGTCCC
TCACAGCCATCTTCCTGCCAGGGCGCACGCGCGCTGGGTGTTCCCGCCTAGTGACACTGGGCCC
GCGATTCCTTGGAGCGGGTTGATGACGTCAGCGTTCGAATTCC (SEQ ID NO: 95)

homo_sapiens

GGGAAAAAGTGGTCTCATACAGAACTTATAAGATTCCCAAATCCAAAGACATTTCACGT
TTATGGTGATTTCCCAGAACACATAGCGACATGCAAATATTGCAGGGCGCCACTCCCCTGTCCC
TCACAGCCATCTTCCTGCCAGGGCGCACGCGCGCTGGGTGTTCCCGCCTAGTGACACTGGGCCC
GCGATTCCTTGGAGCGGGTTGATGACGTCAGCGTTCGAATTCC (SEQ ID NO: 96)

pan_troglodytes

GGGAAAGGGTGGTGCCACACAGAACTTATAAGATTCCCATATGCAAAGACATTTCACGT
TTATGGTGATTTCCCAGAACACATAGCGACATGCAAATATTGCAGGGCGCCACTCCCCTGTCCC
TCACTGCCATCTTCCTGCCAGGGCGCACGCGCGCTGGGTGTTCCCGCCTAGTGACACTGGGCCC
GCGATTCCTTGGAGCGGGTTGATGACGTCAGCGTTCGAATTCC (SEQ ID NO: 97)

pongo_abelii

GAGAAAGGGTGGTCCCGTCCAGAACTTATAAGATTCCCATACCCAAAGACATTTCACGT
TTATGGTGACTTCCCAGAATGCATAGCGACATGCAAATATTGCAGGGCGTCACTCCCCTGTCCC
TCACAGCCATCTTCCTGCCAGGGCGCCCGCGCGCTGGTGTTCCCGCCTAGTGACACTGGGCCCA
CGATTCCTTGGAGCGGGTTGATGACGTCAGCGCTCGTATTCC (SEQ ID NO: 98)

nomascus_leucogenys

GGGGAAAAGTAGTAGACCTTATAAGATTCCCAAACCCAAAGACATTTCTCGTTTATGGT
GACTTCCCAGAAGACATAGCGACATGCAAATATTGCAGGGCGCCACTCCCCTGTCCCTCACAGC

CATCTTCCTGCCAGGGCGCACGCGCGCTGGGTGTTCCCGCCTAGTGACACTCGGCCCGCGATTC
CTTGGAGCGGGTTGATGACGTCAGCGTTCGAATTCC (SEQ ID NO: 99)

tarsius_syrichta

GCGAGAGGGTGGGTCCACACAGAGCTTATAAGGTTCACAAGTAAAGATATTTCACGGTG
ACGGTGACTTCCCACAATACACTGCGACATGCAAATATAGCCGGGCGTGCCTCCCCGATCCCGG
AAGAGCGACTCCTAGCCAGTGCGCACGCGCGCTGCGTGTTCGCGTCCTAGGTCGCTGGGCCCGC
GGTTCCTGGGAGCGGGTGGTGACGTCAGCGGCCCAGCTTC (SEQ ID NO: 100)

otolemur_garnettii

GCCTAAAAGGGCGCTTGCACAGAATTTATAAGGTTCCCAAACAGAGACACATTTCATTA
TTATGGTGACTTCCCACAATGCACAGCGCCATGCAAATATGCTAGGACGCCTCCCCCCGCTACC
TTAAGGTCGTCAACTAACCAGTGCGCGCGCGCACTGCGCGTTTCCCGCCGGTGACTCAATGCCC
GCGTTTGGTGGGAGCTAGTTGGTGACCTCAGTTCTGGAGGCTC (SEQ ID NO: 101)

tupaia_belangeri

GGGGGAAGCTGGGTCCACTGAGTTCTTATAAGGTTTCCAGTCCTAGAGCGATTTTACCA
TTGCGGTGATTTCCCAGCATCCGTAGCTACATGCAAATAGCGCGGGGCGCGTCTCTCAGGTCCC
TCCCCGCCCTCTCACTGTACGTACCCGCGTCCTAGGGACGCCGCGCCCGGGGTTCCCGGACGTC
AGCGTTCCGACGCA (SEQ ID NO: 102)

ailuropoda_melanoleuca

AGGGAAAGCCGCGCCTGGGGCGGATTTATAAGGCTTCCATATCTAAAGGCATTTCACAG
TCATGGTGACTTCCCACAATACATAGCAACATGCAAATATCGCGGGGAGAACCTCCCCTGTCCC
TTGTACGCGGCTTCTAAAGACGCACGCGCGCTCTGTGTTCCCGCCCTGTGACTCTAGGCGGGCA
ATTCCTGGGACAGTGTTCTGACGGGAACGTTCAGGCTCC (SEQ ID NO: 103)

mustela_putorius_furo

GGGAAAGGGTGGACCCACCGAGCATTTATAAGGCTCCCGCATCTAAAGACATTTTACAG
TTATGGTGACTTCCCACAACGCGTAGCAACATGCAAATATCGTGGAGAGTACCGCCCCTGTCCC
ATGCACGCGTCTTCTCAGCAGCACGCACGCGCGCTGTGTTCCCGCCCTGTGACTCCAGGCGGGT
ATTTCCAGGGGCGGGTTGCTGACAGGAACGTTCAGGCTTC (SEQ ID NO: 104)

canis_familiaris

GCAGCGCAGCCCTCTCGCCGCTTATAAAGTGCCGCCCGCACGGCCCTTCTCGCTCACGG
CGACTTCCCATAACACACAGCAGCATGCAAATACCGCGGGGAGCCCCGCCCCGCCCCGGCCCCC
GCACCGCCTCGGGACGCATGCGCCGGCTCTCCGTTCCCGCCTTGGGCCGGCGGCGGGCGGGCGG
GCGAGCGGGCGGGAGCGGCTCCGGCGGGGACGAGCGGGCGCC (SEQ ID NO: 105)

felis_catus

GGGAAAGGGTGGCCCCGCCGAGCATTTATAAGACTCCCATACCTAAAGACATTTCTCAG
TTATGGTGATTTCCCACAACACACAGCAACATGCAAATATCGAGGGGTGTACCGCCCCTGTCCT
TTGTAGACGTCTTCTCTCCAGGACGCACGCGCGCTGTATTCCCGCCTTGTGACTCTAGGCGGGC

GATTCCTGGGAGAGGGTTGATGACGTCCAAGTTCTGGCTTC (SEQ ID NO: 106)

equus_caballus

GGGGGAAAACAGCCCATGGCTGCATTTATAAGACTCACAGATCTAAAGCCATTTCACGA
ATAGGGTGACTTCCCACAATACACAGCGACATGCAAACATAGCGGGGCGTGCCTTTCCTGTACC
CTGTGGGCATCTCTCCTGGACGCACGCGCGCCGGGTGTTCCCGCGCTGTGACTCTAGGCAAGCG
CTTCCTGGGAGAGAGTTGATGACGGCAGCATTCGGGCTCC (SEQ ID NO: 107)

myotis_lucifugus

GGGAGAAGGAGGCGTAGAGGATATATAAGGCCCCCTTATGTGTAGTCCTTTTACGGTTA
GGGTGACTTCCCACAACGCATAGCGACATGCAAATTTGACGGGCGTGCCTCCTCTGTCCCTGCG
GGCAACTTCTCTCCTGGACGCGCGCGCGCTGCGTGTTCCCGCCTTTTGACTCCAGCCGAGCGAA
TCCTGGGAGAGGGCAGGTGACGTCAACAGTCAGGCTCG (SEQ ID NO: 108)

pteropus_vampyrus

GCGAGAAAAATTCTTCACGCAGAATATATAAGGATCCCATATCTGAAGACATTTTACGA
TTACGGCGATTTCCCACAACACATAGCGACATGTAAATGTAGTGGGGCATGCCTCCCCTGTCCC
TTGTGGGCAGCTTCTCGCCAGAACGCACGCGCGGTGCGTGTTCCCGCCTTGTGACTAAGTTGGC
GAGTCAGGGAGGAGATTGATGACGTCAGCTCACCCGCTCC (SEQ ID NO: 109)

bos_taurus

GGCAAACACCGCACGCAAATAGCACTTATAATGTGCTCATACCTAGAGCCACTTTTCGG
TTACGGTGACTTCTCAAAAAGACAGTGGAACATGCAAATATTACAGTGCGTCCCGCCCCTGGTA
GGTCTACGCTAGGACGCACGCGCACTACGGTTCCCGCCTATAGACTGCGCTGGCGATTCCTGGG
AGCGGACTGATGACGTCAGCGTTCGGGATCC (SEQ ID NO: 110)

ovis_aries

GGCGAACAATGCGCGCAAACAGCATTTATAATGAGCTCATACCTAAAGCCACTTTACGG
TTACGGTGACTTCCCACAAGACATTGCGGCATGCAAATATTTTAGTGCGTCCCGCCCCTGGTAG
TTCCACGCTAGGACGCACACGCACTACGGTTCCCGCCTTTAGACTGCGCTGGCGATTCCAGGAG
CGGACTGATGACGTCAGCGTTGGGGCTCC (SEQ ID NO: 111)

tursiops_truncatus

GCCGAAAACCAGGCTCAAGCCACATTTATAAGGCTCCCAAATCTAAGTACATTTGTCGG
TTATGGTGACTTCCCGCACCACATTGCGACATGCAAATACTGCGGAGCGTCCCTCCCCTGGCAA
CTCCTCGCTGGGACGCACGCGCGCTACGTGCTCCCGCCTTTTGACTGCGCCGGCGATACTTGGG
AGAGGGTTGATGACGTCAGCGTTCTGGCTCC (SEQ ID NO: 112)

vicugna_pacos

GGGAAAGGGTGGGCTCACGCAGCCTTTATAAGACTCCCAAACTTAAAGACATTTCTCGG
TTATGGCGACTTCCCACAAGACATAGCGACATGCAAATACTGCAGGGCGCCGACCCGGTCCTGT
GCAGCCATCTTTCGGCTGGGACGCACGCGCGCTGCGTGTTCCCGCCCTGTGACTGCGCCGGCGA
TTACTGGGAGAGGATTGATGACGTCAACGTTCGGGTTCC (SEQ ID NO: 113)

sus_scrofa

GTAGGAAAACTGCTTCTGTGAGCACTTATAAAACTCCCATAAGTAGAGAGATTTCATAG TTATGGTGATTTCCCATAAGACATTGCGACATGCAAATATTGTGGCGCGTTCGTCCCCGTCCGG TGCAGGCAGCTTCGCTCCAGGACGCACGCGCAATACATGTTCCCGCCTTGAGACTGCGCCGGCA GATTCCTAGGAAGTGGTTGATGACGTCGATGTTAGGGATCC (SEQ ID NO: 114)

erinaceus_europaeus

GCCTAAACCGGCTCTTTCGACAGACTTATAAGGACCTCTTATCTTAGGACATTTTTTTG TTAGGGTAACTTCCCACGATGCATAGCGATATGTAAATATGGCGCCGCGAGTCTCTCCTAGGCG TCTCCCCAGGACGCAGGCGCACTGCTTGTTCCCGCGTTAACATTGCTGATTCTGGGAGACTGCT GATGACGTCAGCGTCCAGTCTAC (SEQ ID NO: 115)

choloepus_hoffmanni

AGAAAAAAATAGTTTATGCTGGATTTATAAGATTCCCAAATCTAAAGCCATTTCACAGT TACGGTGATTCCCCACTACACACGGCGATATGCAAATATAGCGGAAGTGTTCCTGAGGCGTGGT AAAGCGCGCGCGCGCTGAGAGTTCCCGCCCTGTGGTGCTGGGCTGGAGATGCCTGAGAACTGGC TGATGACGGCAACGTTCGGGCTCC (SEQ ID NO: 116)

dasypus_novemcinctus

AAAGCGATAGTTTTTTAAACTGGACTTATAAGGCACCCATATCTACGTATATTTCATGG TTAGGGTGATTTCCCACAACACATAGCGAAATGCAAATATTGGAGGGCGCTGAGGCGTGGTCGG GCGCAAGCGCGCTGCGACTTCCCGCCTTTCGGCCCTAGGCCCCAGATTCCTGGGAGCTGGATGA TGACGTTGACGTTCGGATACC (SEQ ID NO: 117)

loxodonta_africana

GGGAAGGAACAAATTCGTCAGGATTTATAAGACTCTCAGAGCTGTAGACATTTCACAGT TAGGGCGATGTCCCACAATACATAGCAACATGCAAATATTCTAGGAGGCCAGCCTCCCCGTCCG CGTGGTCATCTTCTCGCTAGGGCGCACGCCCGCTGCGTGTTCCCGCTCTGTGACCAGGCAGGCG ATTCCTGAGAACCGCTTGGTGACGTCAGTGTTCTGGCTCC (SEQ ID NO: 118)

procavia_capensis

AGGGTAAATCGGCGCTGCTCAGCATTTAAAAGAATCCCAAATGTGTCGCCATTTTACGC TTAGGGTGATATCCCACAAGACACAGCGACATGCAAATATCGTGAGTCTCTGTTTCCCTGTCCA CGAGGGCGTCCTCTCGCTGGGGCGCACGCGCGGTGTGTGTGCCCCCGTTGTGTGTTCCCGCGAT TCCAAAGAACTGGTTGATAACGTTAGACTTCCGGCTGC (SEQ ID NO: 119)

Таблица 6: Примеры терминаторов

SPA

AATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTG (SEQ ID NO: 120)

SPA и Pause

AATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGATAGT

ACTAACATACGCTCTCCATCAAAACAAAACGAAACAAAACAAACTAGCAAAATAGGCTGTCCCC AGTGCAAGTGCAGGTGCCAGAACATTTCTCT (SEQ ID NO: 121)

SV40 (240 п.н.)

ATCTAGATAACTGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAA AAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTT GTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCA TTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTA (SEQ ID NO: 122)

SV40-mini (120 п.н.)

TTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAA TAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAT (SEQ ID NO: 123)

Поли A bGH

CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTG ACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTC TGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGA AGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGG (SEQ ID NO: 124)

Поли A TK

GGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGGAAGGAACCCGCGCTATGACG GCAATAAAAAGACAGAATAAAACGCACGGGTGTTGGGTCGTTTGTTCATAAACGCGGGGTTCGG TCCCAGGGCTGGCACTCTGTCGATACCCCACCGAGACCCCATTGGGGCCAATACGCCCGCGTTT CTTCCTTTTCCCCACCCCACCCCCCAAGTTCGGGTGAAGGCCCAGGGCTCGCAGCCAACGTCGG GGCGGCAGGCCCTGCCATAG (SEQ ID NO: 125)

sNRP1

GGTATCAAATAAAATACGAAATGTGACAGATT (SEQ ID NO: 126)

sNRP1a

AAATAAAATACGAAATGTGACAGATT (SEQ ID NO: 127)

H4B гистонов

GGTTGCTGATTTCTCCACAGCTTGCATTTCTGAACCAAAGGCCCTTTTCAGGGCCGCCC AACTAAACAAAAGAAGAGCTGTATCCATTAAGTCAAGAAGC (SEQ ID NO: 128)

MALAT-1

GATTCGTCAGTAGGGTTGTAAAGGTTTTTCTTTTCCTGAGAAAACAACCTTTTGTTTTC TCAGGTTTTGCTTTTTGGCCTTTCCCTAGCTTTAAAAAAAAAAAAGCAAAAGACGCTGGTGGCT GGCACTCCTGGTTTCCAGGACGGGGTTCAAGTCCCTGCGGTGTCTTTGCTT (SEQ ID NO: 129)

MALAT-comp14

AAAGGTTTTTCTTTTCCTGAGAAATTTCTCAGGTTTTGCTTTTTAAAAAAAAAGCAAAA GACGCTGGTGGCTGGCACTCCTGGTTTCCAGGACGGGGTTCAAGTCCCTGCGGTGTCTTTGCTT (SEQ ID NO: 130)

Термин "гомологичный" относится к "% гомологии" и в настоящем описании используется взаимозаменяемо с термином "% идентичность", и относится к уровню идентичности последовательностей нуклеиновых кислот при выравнивании с использованием программы для выравнивания последовательностей.

Например, как используют в рамках изобретения, 80% гомология означает то же, что и 80% идентичность последовательностей, определяемая с помощью определенного алгоритма, и, таким образом. гомолог данной последовательности обладает более чем 80% идентичностью последовательности на протяжении длины данной последовательности. Предпочтительные уровни идентичности последовательностей включают, но не ограничиваются ими, приблизительно 60%, 65%, 70%,75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,1%, 99,2%, 99,3%, 99,4%, 99,5%, 99,6%, 99,7%, 99,8% или более идентичность последовательности с нуклеотидными последовательностями, указанными в SEQ ID NO: 1-82. Иллюстративные уровни идентичности последовательностей включают, но не ограничиваются ими, приблизительно 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99,1%, 99,2%, 99,3%, 99,4%, 99,5%, 99,6%, 99,7%, 99,8% или более идентичность последовательности с нуклеотидными последовательностями, указанными в SEQ ID NO: 1-82.

В некоторых вариантах осуществления описанное в настоящем описании изобретение относится к не встречающейся в природе системе CRISPR-Cas, содержащей один или несколько векторов, содержащих: a) промотор H1, функционально связанный по меньшей мере с одной нуклеотидной последовательностью, кодирующей гидовую РНК (гРНК) системы CRISPR-Cas, где гРНК гибридизуется с последовательностью ДНК, являющейся мишенью, в клетке, и где молекула ДНК кодирует один или несколько продуктов генов, экспрессируемых в клетке; и b) регуляторный элемент,

функциональный в клетке, функционально связанный с нуклеотидной последовательностью, кодирующей белок Cas9, где компоненты (a) и (b) расположены на одном и том же или на различных векторах системы, где гРНК осуществляет нацеливание на последовательность-мишень и гибридизуется с ней, и белок Cas9 расщепляет молекулу ДНК, изменяя экспрессию одного или нескольких продуктов генов.

В некоторых вариантах осуществления описанное в настоящем описании изобретение относится к не встречающейся в природе системе CRISPR-Cas, содержащей один или несколько векторов, содержащих: a) промотор H1, функционально связанный по меньшей мере с одной нуклеотидной последовательностью, кодирующей гидовую РНК (гРНК) системы CRISPR-Cas, где гРНК гибридизуется с последовательностью ДНК, являющейся мишенью, в эукариотической клетке, и где молекула ДНК кодирует один или несколько продуктов генов, экспрессируемых в эукариотической клетке; и b) регуляторный элемент, функциональный в эукариотической клетке, функционально связанный с нуклеотидной последовательностью, кодирующей белок Cas9 типа II, где компоненты (a) и (b) расположены на одном и том же или на различных векторах системы, причем гРНК осуществляет нацеливание на последовательность-мишень и гибридизуется с ней, и белок Cas9 расщепляет молекулу, и при этом происходит изменение экспрессии одного или нескольких продуктов генов. В одном аспекте последовательность-мишень может представлять собой последовательность-мишень, которая начинается с любого нуклеотида, например, $N_{20}NGG$. В некоторых вариантах осуществления последовательность-мишень содержит нуклеотидную последовательность $AN_{19}NGG$. В некоторых вариантах осуществления последовательность-мишень содержит нуклеотидную последовательность $GN_{19}NGG$. В некоторых вариантах осуществления последовательность-мишень содержит нуклеотидную последовательность $CN_{19}NGG$. В некоторых вариантах осуществления последовательность-мишень содержит нуклеотидную последовательность $TN_{19}NGG$. В некоторых вариантах осуществления последовательность-мишень содержит нуклеотидную последовательность $AN_{19}NGG$ или $GN_{19}NGG$. В другом аспекте белок

Cas9 кодон-оптимизирован для экспрессии в клетке. В другом аспекте белок Cas9 кодон-оптимизирован для экспрессии в эукариотической клетке. В следующем аспекте эукариотическая клетка представляет собой клетку млекопитающего или человека. В другом аспекте экспрессия одного или нескольких продуктов генов снижается.

Описанное в настоящем описании изобретение также относится к не встречающейся в природе системе CRISPR-Cas, содержащей вектор, содержащий двунаправленный промотор H1, где двунаправленный промотор H1 содержит: a) элементы контроля, которые обеспечивают транскрипцию в одном направлении по меньшей мере одной нуклеотидной последовательности, кодирующей гидовую РНК (гРНК) системы CRISPR-Cas, где гРНК гибридизуется с последовательностью ДНК, являющейся мишенью, в эукариотической клетке, и где молекула ДНК кодирует один или несколько продуктов, экспрессируемых в эукариотической клетке; и b) элементы контроля, которые обеспечивают транскрипцию в противоположном направлении нуклеотидной последовательности, кодирующей белок Cas9 типа II, причем гРНК нацелена на последовательность-мишень и гибридизуется с ней, и белок Cas9 расщепляет молекулу ДНК, и при этом происходит изменение экспрессии одного или нескольких продуктов генов. В одном аспекте последовательность-мишень может представлять собой последовательность-мишень, которая начинается с любого нуклеотида, например, N_{20}NGG. В некоторых вариантах осуществления последовательность-мишень содержит нуклеотидную последовательность AN_{19}NGG. В некоторых вариантах осуществления последовательность-мишень содержит нуклеотидную последовательность GN_{19}NGG. В некоторых вариантах осуществления последовательность-мишень содержит нуклеотидную последовательность CN_{19}NGG. В некоторых вариантах осуществления последовательность-мишень содержит нуклеотидную последовательность TN_{19}NGG. В некоторых вариантах осуществления последовательность-мишень содержит нуклеотидную последовательность AN_{19}NGG или GN_{19}NGG. В другом аспекте белок Cas9 является кодон-оптимизированным для экспрессии в клетке. В

другом аспекте белок Cas9 является кодон-оптимизированным для экспрессии в эукариотической клетке. В следующем аспекте эукариотическая клетка представляет собой клетку млекопитающего или человека. В другом аспекте, экспрессия одного или нескольких продуктов генов снижается.

В некоторых вариантах осуществления комплекс CRISPR содержит одну или несколько последовательностей ядерной локализации, имеющих достаточную длину для обеспечения накопления комплекса CRISPR в поддающемся обнаружению количестве в ядре клетки (например, эукариотической клетки). Без связи с теорией, полагают, что последовательность ядерной локализации не является необходимой для активности комплекса CRISPR у эукариот, но что включение таких последовательностей повышает активность системы, особенно в отношении нацеливания молекул нуклеиновой кислоты в ядро. В некоторых вариантах осуществления фермент CRISPR представляет собой фермент системы CRISPR типа II. В некоторых вариантах осуществления фермент CRISPR представляет собой фермент Cas9. В некоторых вариантах осуществления фермент Cas9 представляет собой Cas9 *C. jejuni*, *G. stearothermophilus*, *N. meningitidis*, *S. aureus*, *S. pneumoniae*, *S. pyogenes* или *S. thermophilus*, и может включать мутантный Cas9, происходящий из этих организмов. Фермент может представлять собой гомолог или ортолог Cas9, такой как Cpf1 или Cas13a/C2c2.

Обычно и на протяжении настоящего описания термин "вектор" относится к молекуле нуклеиновой кислоты, способной транспортировать другую нуклеиновую кислоту, с которой она связана. Векторы включают, но не ограничиваются ими, молекулы нуклеиновых кислот, которые являются одноцепочечными, двухцепочечными или частично двухцепочечными; молекулы нуклеиновых кислот, которые содержат один или несколько свободных концов, не содержат свободных концов (например, кольцевые); молекулы нуклеиновых кислот, которые содержат ДНК, РНК или обе из них; и другие типы полинуклеотидов, известные в данной области. Одним типом вектора является "плазмида", которая относится к кольцевой двухцепочечной петле ДНК, в которую могут быть встроены дополнительные сегменты ДНК, например,

стандартными способами молекулярного клонирования. Другим типом вектора является вирусный вектор, где последовательности ДНК или РНК вирусного происхождения присутствуют в векторе для упаковывания в вирус (например, ретровирусы, ретровирусы с дефектом репликации, аденовирусы, аденовирусы с дефектом репликации и аденоассоциированные вирусы). Вирусные векторы также включают полинуклеотиды, которые содержатся в вирусе, для трансфекции в клетку-хозяина.

Определенные векторы способны к автономной репликации в клетке-хозяине, в которую они введены (например, бактериальные векторы, имеющие бактериальный ориджин репликации и эписомальные векторы млекопитающих). Другие векторы (например, неэписомальные векторы млекопитающих) встраиваются в геном клетки-хозяина при введении в клетку-хозяина и тем самым реплицируются вдоль генома хозяина. Более того, определенные векторы способны направлять экспрессию генов, с которыми они функционально связаны. Такие векторы называют в настоящем описании "экспрессирующими векторами". Часто обычные экспрессирующие векторы, используемые в способах рекомбинантных ДНК, имеют форму плазмид.

Рекомбинантные экспрессирующие векторы могут содержать нуклеиновую кислоту по настоящему изобретению в форме, подходящей для экспрессии нуклеиновой кислоты в клетке-хозяине, что означает, что рекомбинантные экспрессирующие векторы включают один или несколько регуляторных элементов, которые могут быть выбраны, исходя из клеток-хозяев, используемых для экспрессии, которые функционально связаны с последовательностью нуклеиновой кислоты, подлежащей экспрессии.

В рекомбинантном экспрессирующем векторе "функционально связанный" означает, что представляющая интерес нуклеотидная последовательность связана с регуляторным элементом(ами) так, чтобы позволить экспрессию нуклеотидной последовательности (например, в системе транскрипции/трансляции *in vitro* или в клетке-хозяине, когда вектор вводят в клетку-хозяина).

Термин "регуляторный элемент" включает промоторы, энхансеры, внутренние участки связывания рибосомы (IRES) и другие элементы контроля экспрессии (например, сигналы

терминации транскрипции, такие как сигналы полиаденилирования и поли-U последовательности). Такие регуляторные элементы описаны, например, в Goeddel (1990) Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. Регуляторные элементы включают регуляторные элементы, которые обеспечивают конститутивную экспрессию нуклеотидной последовательности во многих типах клеток-хозяев, и регуляторные элементы, которые обеспечивает экспрессию нуклеотидной последовательности только в определенных клетках-хозяевах (например, тканеспецифические регуляторные последовательности). Тканеспецифический промотор может обеспечивать экспрессию в основном представляющей интерес ткани, такой как мышца, нейрон, кость, кожа, кровь, определенных органах (например, печень, поджелудочная железа) или конкретных типах клеток (например, лимфоциты). Регуляторные элементы также могут обеспечивать временную экспрессию, например, зависимым от клеточного цикла или от стадии развития образом, которая также может быть или может не быть тканеспецифической или специфической для типа клеток.

В некоторых вариантах осуществления вектор содержит один или несколько промоторов pol III, один или несколько промоторов pol II, один или несколько промоторов pol I, или их комбинации. Примеры промоторов pol III включают, но не ограничиваются ими, промоторы U6 и H1. Примеры промоторов pol II включают, но не ограничиваются ими, ретровирусный промотор LTR вируса саркомы Рауса (RSV) (необязательно с энхансером RSV), промотор цитомегаловируса (CMV) (необязательно с энхансером CMV) (например, Boshart et al. (1985) *Cell* 41:521-530), промотор SV40, промотор дигидрофолатредуктазы, промотор β-актина, промотор фосфоглицеринкиназы (PGK) и промотор EF1α.

Также термин "регуляторный элемент" охватывает энхансерные элементы, такие как WPRE; энхансеры CMV; сегмент R-U5' в LTR HTLV-I (Takebe et al. (1988) *Mol. Cell. Biol.*8:466-472); энхансер SV40; и последовательность интрона между экзонами 2 и 3 β-глобина кролика (O'Hare et al. (1981) *Proc. Natl. Acad. Sci.*

USA. 78(3):1527-31). Специалистам в данной области будет понятно, что конструкция экспрессирующего вектора может зависеть от таких факторов, как выбор клетки-хозяина, подлежащей трансформации, желаемый уровень экспрессии и т.д. Вектор можно вводить в клетки-хозяева, тем самым продуцируя транскрипты, белки или пептиды, включая слитые белки или пептиды, кодируемые нуклеиновыми кислотами, как описано в настоящем описании (например, транскрипты коротких палиндромных повторов, регулярно расположенных кластерами (CRISPR), белки, ферменты, их мутантные формы, их слитые белки, и т.д.). Преимущественные векторы включают лентивирусы и аденоассоциированные вирусы, и типы таких векторов также могут быть выбраны для нацеливания на конкретные типы клеток.

Термины "полинуклеотид", "нуклеотид", "нуклеотидная последовательность", "нуклеиновая кислота" и "олигонуклеотид" используются взаимозаменяемо. Относятся к полимерной форме нуклеотидов любой длины, либо к дезоксирибонуклеотидам, либо к рибонуклеотидам, или к их аналогам. Полинуклеотиды могут иметь любую трехмерную структуру и могут выполнять любую функцию, известную или неизвестную. Неограничивающие примеры полинуклеотидов являются следующими: кодирующие или некодирующие области гена или фрагмента гена, локусы (локус), определенные в анализе сцепления, экзоны, интроны, матричная РНК (мРНК), транспортная РНК, рибосомальная РНК, короткая интерферирующая РНК (киРНК), короткая шпилечная РНК (кшРНК), микро-РНК (мкРНК), рибозимы, кДНК, рекомбинантные полинуклеотиды, разветвленные полинуклеотиды, плазмиды, векторы, выделенная ДНК любой последовательности, выделенная РНК любой последовательности, зонды нуклеиновой кислоты и праймеры. Полинуклеотид может содержать один или несколько модифицированных нуклеотидов, таких как метилированные нуклеотиды и нуклеотидные аналоги. При наличии, модификации нуклеотидной структуры могут быть проведены до или после сборки полимера. Последовательность нуклеотидов может прерываться ненуклеотидными компонентами. Полинуклеотид может быть далее модифицирован после полимеризации, например, посредством конъюгации с компонентом для мечения.

В аспектах описанного в настоящем описании изобретения термины "химерная РНК", "химерная гидовая РНК", "гидовая РНК", "единичная гидовая РНК", "синтетическая гидовая РНК" и "cr-РНК" используются взаимозаменяемо и относятся к полинуклеотидной последовательности, содержащей гидовую последовательность. Термин "гидовая последовательность" относится к последовательности размером приблизительно 20 п.о. в гидовой РНК, которая определяет участок-мишень, и может использоваться взаимозаменяемо терминами "гид" или "спейсер".

Как используют в рамках изобретения термин "дикий тип" представляет собой термин, понятный специалистам в данной области, и он означает типичную форму организма, штамма, гена или характеристики, как она встречается в природе, в противоположность мутантной форме или варианту.

Как используют в рамках изобретения термин "вариант" означает проявление качеств, которые имеют характеристики, отличающиеся от характеристик, которые встречаются в природе.

Термины "не встречающийся в природе" или "сконструированный" используются взаимозаменяемо и указывают на вовлечение человека. Эти термины в отношении молекул нуклеиновой кислоты или полипептидов означают, что молекула нуклеиновой кислоты или полипептид по меньшей мере по существу свободны от по меньшей мере одного другого компонента, с которым они естественным образом ассоциированы и встречаются в природе.

"Комплементарность" относится к способности нуклеиновой кислоты образовывать водородную связь(и) с другой последовательностью нуклеиновой кислоты либо традиционного типа Уотсона-Крика, либо другого нетрадиционного типа. Процентная комплементарность указывает на процент остатков в молекуле нуклеиновой кислоты, которые могут образовывать водородные связи (например, образование пар по принципу Уотсона-Крика) со второй последовательностью нуклеиновой кислоты (например, 5, 6, 7, 8, 9, 10 из 10 являются комплементарными на 50%, 60%, 70%, 80%, 90% и 100%). "Полностью комплементарный" означает, что все последовательно расположенные остатки последовательности нуклеиновой кислоты будут образовывать водородную связь с тем же

количеством последовательно расположенных остатков во второй последовательности нуклеиновой кислоты. "По существу комплементарный", как используют в рамках изобретения, относится к степени комплементарности, которая составляет по меньшей мере 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%. 97%, 98%, 99% или 100% на протяжении области из 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50 или более нуклеотидов, или относится к двум нуклеиновым кислотам, которые гибридизуются в жестких условиях.

Как используют в рамках изобретения, "жесткие условия" гибридизации относятся к условиям, в которых нуклеиновая кислота, обладающая комплементарностью с последовательностью-мишенью, преимущественно гибридизуется с последовательностью-мишенью и по существу не гибридизуется с последовательностями, не являющимися мишенью. Жесткие условия обычно являются зависимыми от последовательности и варьируются в зависимости от ряда факторов. Как правило, чем более длинной является последовательность, тем более высокой является температура, при которой последовательность специфически гибридизуется с ее последовательностью-мишенью. Неограничивающие примеры жестких условиях подробно описаны в Tijssen (1993), Laboratory Techniques In Biochemistry And Molecular Biology-Hybridization With Nucleic Acid Probes Part 1, Second Chapter "Overview of principles of hybridization and the strategy of nucleic acid probe assay", Elsevier, N.Y.

"Гибридизация" относится к реакции, в которой один или несколько полинуклеотидов реагируют с образованием комплекса, который стабилизируется образованием водородных связей между основаниями нуклеотидных остатков. Образование водородных связей может происходить по принципу спаривания оснований Уотсона-Крика, по принципу связывания Хугстена, или любым другом специфическим для последовательности образом. Комплекс может содержать две цепи, образующих дуплексную структуру, три или более цепи, образующих многоцепочечный, одну самогибридизующуюся цепь или любую их комбинацию. Реакция гибридизации может представлять собой стадию в более широком процессе, таком как

инициация ПЦР или расщепление полинуклеотида ферментом. Последовательность, способная гибридизоваться с данной последовательностью, называется последовательностью, "комплементарной" данной последовательности.

Как используют в рамках изобретения, "экспрессия" относится к процессу, посредством которого полинуклеотид транскрибируется с ДНК-матрицы (например, в мРНК или другой РНК-транскрипт), и/или к процессу, посредством которого транскрибированная мРНК затем транслируется в пептиды, полипептиды или белки. Транскрипты и кодируемые полипептиды в совокупности могут называться "продуктом гена". Если полинуклеотид происходит из геномной ДНК, экспрессия может включать сплайсинг мРНК в эукариотической клетке.

Термины "полипептид", "пептид" и "белок" используют в настоящем описании взаимозаменяемо для обозначения полимеров из аминокислот любой длины. Полимер может быть линейным или разветвленным, он может содержать модифицированные аминокислоты, и он может прерываться не аминокислотами. Также термины охватывают полимер из аминокислот, который является модифицированным; например, посредством образования дисульфидных связей, гликозилирования, липидирования, ацетилирования, фосфорилирования или любой другой манипуляции, такой как конъюгация с компонентом для мечения.

Как используют в рамках изобретения, термин "аминокислота" включает природные, и/или неприродные или синтетические аминокислоты, включая глицин и как D, так L, оптические изомеры, и аналоги аминокислот и пептидомиметики.

Если нет иных указаний, для применения настоящего изобретения на практике используются общепринятый способы иммунологии, биохимии, химии, молекулярной биологии, клеточной биологии, геномики и рекомбинантных ДНК, которые входят в пределы квалификации в данной области (Sambrook, Fritsch and Maniatis (1989) Molecular Cloning: A Laboratory Manual, 2nd edition; Ausubel et al., eds. (1987) Current Protocols in Molecular Biology); MacPherson et al., eds. (1995) Methods in Enzymology (Academic Press, Inc.): PCR 2: A Practical Approach);

Harlow and Lane, eds. (1988) Antibodies, A Laboratory Manual; Freshney, ed. (1987) Animal Cell Culture).

Несколько аспектов настоящего изобретения относятся к векторным системам, содержащим один или несколько векторов, или к самим векторам. Векторы могут быть сконструированы для экспрессии транскриптов CRISPR (например, транскрипты нуклеиновых кислот, белки или ферменты) в прокариотических или эукариотических клетках. Например, транскрипты CRISPR могут экспрессироваться в бактериальных клетках, таких как *Escherichia coli*, клетках насекомых (с использованием бакуловирусных экспрессирующих векторов), клетках дрожжей или клетках млекопитающих. Подходящие клетки-хозяева дополнительно описаны в Goeddel (1990) Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. Альтернативно рекомбинантный экспрессирующий вектор может транскрибироваться и транслироваться *in vitro*, например с использованием регуляторных последовательностей промотора T7 и полимеразы T7.

Векторы могут быть введены и увеличены в количестве в прокариотическом организме. В некоторых вариантах осуществления прокариотический организм используют для увеличения количества копий вектора, подлежащего введению в эукариотическую клетку, или в качестве промежуточного вектора для получения вектора, подлежащего введению в эукариотическую клетку (например, амплификация плазмиды в качестве части системы упаковывания вирусного вектора). В некоторых вариантах осуществления прокариотический организм используют для увеличения количества копий вектора и экспрессии одной или нескольких нуклеиновых кислот, чтобы предоставить источник одного или нескольких белков для доставки в клетку-хозяина или организм хозяина. Экспрессию белков в прокариотах наиболее часто проводят в *Escherichia coli* с векторами, содержащими конститутивные или индуцибельные промоторы, обеспечивающие экспрессию либо слитых, либо неслитых белков.

Векторы для слияния добавляют некоторое количество аминокислот к белку, кодируемому ими, например, к N-концу рекомбинантного белка. Такие векторы для слияния могут служить

одной или нескольким целям, таким как: (i) повышение экспрессии рекомбинантного белка; (ii) повышение растворимости рекомбинантного белка; и (iii) облегчение очистки рекомбинантного белка путем действия в качестве лиганда при аффинной очистке. Часто, в экспрессирующих векторах для слияния участок протеолитического расщепления вносится в область контакта подлежащей слиянию части и рекомбинантного белка для обеспечения отделения рекомбинантного белка от подлежащей слиянию части после очистки слитого белка. Такие ферменты и распознаваемые ими последовательности включают фактор Xa, тромбин и энтерокиназу. Пример экспрессирующих векторов для слияния включают pGEX (Pharmacia Biotech Inc; Smith and Johnson (1988) *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) и pRIT5 (Pharmacia, Piscataway, N.J.), которые осуществляют слияние глутатион S-трансферазы (GST), связывающего мальтозу E белка или белка A, соответственно, с рекомбинантным белком-мишенью.

Примеры подходящих индуцибельных экспрессирующих векторов *E. coli*, не осуществляющих слияние, включают pTrc (Amrann et al. (1988) *Gene* 69:301-315) и pET 11d (Studier et al. (1990) Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif.).

В некоторых вариантах осуществления вектор представляет собой дрожжевой экспрессирующий вектор. Примеры векторов для экспрессии в дрожжах *Saccharomyces cerivisae* включают pYepSec1 (Baldari, et al.(1987) *EMBO J.* 6: 229-234), pMFa (Kuijan и Herskowitz (1982) *Cell* 30: 933-943), pJRY88 (Schultz et al. (1987) *Gene* 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.) и picZ (InVitrogen Corp, San Diego, Calif.).

В некоторых вариантах осуществления вектор способен обеспечивать экспрессию одной или нескольких последовательностей в клетках млекопитающих с использованием экспрессирующего вектора млекопитающих. Примеры экспрессирующих векторов млекопитающих включают pCDM8 (Seed (1987) *Nature* 329: 840) и pMT2PC (Kaufman et al. (1987) *EMBO J.* 6: 187-195). Когда его используют в клетках млекопитающих, контролирующие функции

экспрессирующего вектора, как правило, обеспечиваются одним или несколькими регуляторными элементами. Например, обычно используемые промоторы происходят из вируса полиомы, аденовируса 2, цитомегаловируса, вируса 40 обезьян и других, описанных в настоящем описании и известных в данной области. Для других подходящих экспрессирующих систем как для прокариотических, так и для эукариотических клеток, см., например, Chapters 16 and 17 Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

В некоторых вариантах осуществления рекомбинантный экспрессирующий вектор млекопитающих способен обеспечивать экспрессию нуклеиновой кислоты предпочтительно в конкретном типе клеток (например, для экспрессии нуклеиновой кислоты используются тканеспецифические регуляторные элементы). Тканеспецфические регуляторные элементы известны в данной области. Неограничивающие примеры подходящих тканеспецифических промоторов включают промотор альбумина (специфический для печени; Pinkert et al. (1987) *Genes Dev.* 1: 268-277), специфические для лимфоидной ткани промоторы (Calame and Eaton (1988) *Adv. Immunol.* 43: 235-275), в частности, промоторы Т-клеточных рецепторов (Winoto and Baltimore (1989) *EMBO J.*8: 729-733) и иммуноглобулины (Baneiji et al. (1983) *Cell* 33: 729-740; Queen and Baltimore (1983) *Cell* 33: 741-748), нейрон-специфические промоторы (например, промотор нейрофиламентов; Byrne and Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), специфические для поджелудочной железы промоторы (Edlund et al.(1985) *Science* 230: 912-916) и специфические для молочной железы промоторы (например, промотор молочной сыворотки; патент США № 4873316 и публикация заявки Европы № 264166). Также охватываются регулируемые стадией развития промоторы, например, промоторы hox мыши (Kessel and Gruss (1990) *Science* 249: 374-379) и промотор α-фетопротеина (Campes and Tilghman (1989) *Genes Dev.* 3: 537-546).

В некоторых вариантах осуществления регуляторный элемент

функционально связан с одним или несколькими элементами системы CRISPR, так что он контролирует экспрессию одного или нескольких элементов системы CRISPR. В общем, CRISPR (короткие палиндромные повторы, регулярно расположенные кластерами), также известны как SPIDR (прямые повторы с рассеянными спейсерами), представляют собой семейство локусов ДНК, которые обычно являются специфическими для конкретного вида бактерий. Локус CRISPR включает отдельный класс рассеянных повторов с короткой последовательностью (SSR), которые распознаются *E. coli* (Ishino et al. (1987) *J. Bacteriol.*, 169:5429-5433; и Nakata et al. (1989) J. *Bacteriol.*, 171:3553-3556), и ассоциированные с ними гены. Сходные рассеянные SSR были идентифицированы в *Haloferax mediterranei, Streptococcus pyogenes, Anabaena* и *Mycobacterium tuberculosis* (Groenen et al. (1993) *Mol. Microbiol.*, 10:1057-1065; Hoe et al. (1999) *Emerg. Infect. Dis.*, 5:254-263; Masepohl et al. (1996) *Biochim. Biophys. Acta* 1307:26-30; и Mojica et al. (1995) *Mol. Microbiol.*, 17:85-93). Локусы CRISPR, как правило, отличаются от других SSR структурой повторов, которые называются короткими регулярно расположенными повторами (SRSR) (Janssen et al. (2002) *OMICS J. Integ. Biol.*, 6:23-33; и Mojica et al. (2000) *Mol. Microbiol.*, 36:244-246). Как правило, повторы представляют собой короткие элементы, которые встречаются в кластерах, между которыми регулярно рассеяны уникальные встроенные последовательности по существу с постоянной длинной (Mojica et al. (2000) *Mol. Microbiol.*, 36:244-246). Хотя последовательности повторов являются высококонсервативными среди штаммов, количество рассеянных повторов и последовательностей спейсерных областей обычно различается от штамма к штамму (van Embden et al. (2000) *J. Bacteriol.*, 182:2393-2401). Локусы CRISPR были идентифицированы более чем в 40 прокариотических организмах (например, Jansen et al. (2002) *Mol. Microbiol.*, 43:1565-1575; и Mojica et al. (2005) *J. Mol. Evol.* 60:174-82), включая, но не ограничиваясь ими, *Aeropyrum, Pyrobaculum, Sulfolobus, Archaeoglobus, Halocarcula, Methanobacteriumn, Methanococcus, Methanosarcina, Methanopyrus, Pyrococcus, Picrophilus, Thernioplasnia, Corynebacterium, Mycobacterium,*

Streptomyces, Aquifrx, Porphvromonas, Chlorobium, Thermus, Bacillus, Listeria, Staphylococcus, Clostridium, Thermoanaerobacter, Mycoplasma, Fusobacterium, Azarcus, Chromobacterium, Neisseria, Nitrosomonas, Desulfovibrio, Geobacter, Myrococcus, Campylobacter, Wolinella, Acinetobacter, Erwinia, Escherichia, Legionella, Methylococcus, Pasteurella, Photobacterium, Salmonella, Xanthomonas, Yersinia, Treponema и Thermotoga.

Главным образом, "система CRISPR" относится в совокупности к транскриптам и другим элементам, вовлеченным в экспрессию или контролирующим активность CRISPR-ассоциированных ("Cas") генов, включая последовательности, кодирующие ген Cas, гидовую последовательность (также называемую "спейсером" в контексте эндогенной системы CRISPR), или другие последовательности и транскрипты из локуса CRISPR. В некоторых вариантах осуществления один или несколько элементов системы CRISPR происходят из системы CRISPR типа I, типа II или типа III. В некоторых вариантах осуществления один или несколько элементов системы CRISPR происходят из конкретного организма, содержащего эндогенную систему CRISPR, таких как *Streptococcus pyogenes*. Как правило, система CRISPR характеризуется элементами, которые способствуют образованию комплекса CRISPR в области последовательности-мишени (также называемой протоспейсером в контексте эндогенной системы CRISPR).

В контексте образования комплекса CRISPR, "последовательность-мишень" относится к последовательности, с которой гидовая последовательность имеет комплементарность, где гибридизация между последовательностью-мишенью и гидовой последовательностью способствует образованию комплекса CRISPR. Полная комплементарность не требуется при условии, что существует достаточная комплементарность для обеспечения гибридизации и способствования образованию комплекса CRISPR. Последовательность-мишень может содержать любой полинуклеотид, такой как ДНК- или РНК-полинуклеотиды. В некоторых вариантах осуществления последовательность-мишень расположена в ядре или цитоплазме клетки. В некоторых вариантах осуществления

последовательность-мишень может находиться в органелле эукариотической клетки, например, митохондрии или хлоропласте. Последовательность или матрицу, которые можно использовать для встраивания посредством рекомбинации в заданный локус, содержащий последовательности-мишени, называют "редактирующей матрицей", или "редактирующим полинуклеотидом", или "редактирующей последовательностью". В аспектах настоящего изобретения экзогенный матричный полинуклеотид может называться редактирующей матрицей. В одном аспекте настоящего изобретения рекомбинация представляет собой гомологичную рекомбинацию.

В некоторых вариантах осуществления вектор содержит один или несколько участков для встраивания, таких как распознаваемая эндонуклеазой рестрикции последовательность (также называемая "участком клонирования"). В некоторых вариантах осуществления один или несколько участков для встраивания (например, приблизительно или более чем приблизительно 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 или более участков для встраивания) расположены выше и/или ниже одного или нескольких элементов последовательности одного или нескольких векторов. Когда используют множество различных гидовых последовательностей, одну экспрессирующую конструкцию можно использовать для нацеливания активности CRISPR на множество различных соответствующих последовательностей-мишеней в клетке. Например, один вектор может содержать приблизительно или более чем приблизительно 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 или более гидовых последовательностей. В некоторых вариантах осуществления может быть предоставлено и необязательно доставлено в клетку приблизительно или более чем приблизительно 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 или более таких содержащих гидовую последовательность векторов.

В некоторых вариантах осуществления вектор содержит регуляторный элемент, функционально связанный с кодирующей фермент последовательностью, кодирующей фермент CRISPR, такой как белок Cas. Неограничивающие примеры белков Cas включают Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (также известна как Csn1 и Csx12), Cas10, Cas13a, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4,

Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, их гомологи или их модифицированные версии. Эти ферменты известны; например, аминокислотная последовательность белка Cas9 *S. pyogenes* может быть найдена в базе данных SwissProt под номером доступа Q99ZW2. В некоторых вариантах осуществления немодифицированный фермент CRISPR обладает активностью расщепления ДНК, такой как активность Cas9. В некоторых вариантах осуществления фермент CRISPR представляет собой Cas9 и может представлять собой Cas9 из *S. pyogenes* или *S. pneumoniae*.

В некоторых вариантах осуществления фермент CRISPR обеспечивает расщепление одной или обеих цепей в области последовательности-мишени, например, в последовательности-мишени и/или в последовательности, комплементарной последовательности-мишени. В некоторых вариантах осуществления фермент CRISPR обеспечивает расщепление одной или обеих цепей в пределах приблизительно 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500 или более пар оснований от первого или последнего нуклеотида последовательности-мишени. В некоторых вариантах осуществления вектор кодирует фермент CRISPR, который является мутантным относительно соответствующего фермента дикого типа, так что мутантный фермент CRISPR лишен способности расщеплять одну или обе цепи полинуклеотида-мишени, содержащего последовательность-мишень.

В некоторых вариантах осуществления фермент, кодирующий последовательность, кодирующую фермент CRISPR, является кодон-оптимизированным для экспрессии в конкретных клетках, таких как эукариотические клетки. Эукариотические клетки могут представлять собой клетки конкретного организма или могут происходить из конкретного организма, такого как млекопитающее, включая, но не ограничиваясь ими, человека, мышь, крысу, кролика, собаку или не являющегося человеком примата. Как правило, оптимизация кодонов относится к процессу модификации последовательности нуклеиновых кислот для усиления экспрессии в представляющих интерес клетках-хозяевах путем замены по меньшей

мере одного кодона (например, приблизительно или более чем приблизительно 1, 2, 3, 4, 5, 10, 15, 20, 25, 50 или более кодонов) нативной последовательности кодонами, которые более часто или наиболее часто используются в генах этой клетки-хозяина, при сохранении нативной аминокислотной последовательности. Различные виды демонстрируют конкретное предпочтение определенных кодонов для конкретной аминокислоты. Предпочтение кодонов (различие в использование кодонов между организмами) часто коррелирует с эффективностью трансляции матричной РНК (мРНК), которая, как полагают, в свою очередь зависит, среди прочего, от свойств транслируемых кодонов и доступности конкретных молекул транспортной РНК (тРНК) молекулы. Преобладание отдельных тРНК в клетке обычно отражает кодоны, используемые наиболее часто в синтезе пептидов. Таким образом, с помощью оптимизации кодонов гены могут быть приспособлены для оптимальной экспрессии в данном организме. Таблицы использования кодонов являются общедоступными, например, в "Codon Usage Database", и эти таблицы могут быть адаптированы рядом способов. См. Nakamura et al. (2000) *Nucl. Acids Res.* 28:292. Также доступны компьютерные алгоритмы для оптимизации кодонов конкретной последовательности для экспрессии в конкретной клетке-хозяине, такие как Gene Forge (Aptagen; Jacobus, Pa.). В некоторых вариантах осуществления один или несколько кодонов (например, 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, или более, или все кодоны) в последовательности, кодирующей фермент CRISPR, соответствуют наиболее часто используемому кодону для конкретной аминокислоты.

Как правило, гидовая последовательность представляет собой любую полинуклеотидную последовательность, обладающую достаточной комплементарностью с последовательностью полинуклеотида-мишени, чтобы гибридизоваться с последовательностью-мишенью и обеспечить специфическое для последовательности связывание комплекса CRISPR с последовательностью-мишенью. В некоторых вариантах осуществления степень комплементарности между гидовой последовательностью и соответствующей ей последовательностью-мишенью при оптимальном

выравнивании с использованием подходящего алгоритма выравнивания составляет или превышает приблизительно 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97,5%, 99% или более. Оптимальное выравнивание можно проводить с использованием любого подходящего алгоритма выравнивания последовательностей, неограничивающий пример которого включает алгоритм Smith-Waterman, алгоритм Needleman-Wunsch, алгоритмы на основе преобразования Burrows-Wheeler (например, Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies, ELAND (Illumina, San Diego, Calif.), SOAP (доступный на soap.genomics.org.cn) и Maq (доступный на maq.sourceforge.net). В некоторых вариантах осуществления гидовая последовательность имеет длину приблизительно или более чем приблизительно 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75 или более нуклеотидов. В некоторых вариантах осуществления гидовая последовательность имеет длину менее чем приблизительно 75, 50, 45, 40, 35, 30, 25, 20, 15, 12 или менее нуклеотидов.

Способность гидовой последовательности обеспечивать специфическое для последовательности связывание комплекса CRISPR с последовательностью-мишенью можно оценивать с помощью любого подходящего анализа. Например, компоненты системы CRISPR, достаточные для формирования комплекса CRISPR, включая подлежащую тестированию гидовую последовательность, могут быть предоставлены клетке-хозяину, имеющей соответствующую последовательность-мишень, например, путем трансфекции векторами, кодирующими компоненты последовательности CRISPR, с последующей оценкой предпочтительного расщепления в последовательности-мишени, например, с помощью анализа Surveyor, как описано в настоящем описании. Аналогично, расщепление последовательности полинуклеотида-мишени можно оценивать в пробирке путем предоставления последовательности-мишени, компонентов комплекса CRISPR, включая гидовую последовательность, подлежащую тестированию, и контрольную гидовую последовательность, отличную от исследуемой гидовой последовательности, и сравнения связывания и скорости

расщепления последовательности-мишени между исследуемой и контрольной гидовыми последовательностями. Возможны другие способы анализа, и они известны специалистам в данной области.

Гидовую последовательность можно выбирать для нацеливания на любую последовательность-мишень. В некоторых вариантах осуществления последовательность-мишень представляет собой последовательность в геноме клетки. Иллюстративные последовательности-мишени включают последовательности, которые являются уникальными в геноме-мишени.

В некоторых вариантах осуществления фермент CRISPR является частью слитого белка, содержащего один или несколько гетерологичных белковых доменов (например, приблизительно или более чем приблизительно 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 или более доменов в дополнение к ферменту CRISPR). Слитый с ферментом CRISPR белок может содержать любую дополнительную белковую последовательность, и необязательно линкерную последовательность между любыми двумя доменами. Примеры белковых доменов, которые могут быть слиты с ферментом CRISPR, включают, но не ограничиваются ими, эпитопные метки, последовательности репортерных генов и белковые домены, имеющие один или несколько из следующих видов активности: метилазная активность, деметилазная активность, активность активации транскрипции, активность репрессии транскрипции, активность высвобождения фактора транскрипции, активность модификации гистонов, активность расщепления РНК и активность связывания нуклеиновой кислоты. Неограничивающие примеры эпитопных меток включают гистидиновые (His) метки, V5-метки, FLAG-метки, метки на основе гемагглютинина (HA) вируса гриппа, Myc-метки, метки VSV-G и тиоредоксиновые (Trx) метки. Примеры репортерных генов включают, но не ограничиваются ими, гены глутатион-5-трансферазы (GST), пероксидазы хрена (HRP), хлорамфениколацетилтрансферазы (CAT) бета-галактозидазы, бета-глюкуронидазы, люциферазы, зеленого флуоресцентного белка (GFP), HcRed, DsRed, голубого флуоресцентного бела (CFP), желтого флуоресцентного белка (YFP) и аутофлуоресцентных белков, включая синий флуоресцентный белок (BFP). Фермент CRISPR может быть слит с последовательностью

гена, кодирующей белок или фрагмент белка, который связывает молекулы ДНК или связывает другие клеточные молекулы, включая, но не ограничиваясь ими, связывающий мальтозу белок (MBP), S-метку, слитые конструкции Lex A-ДНК-связывающий домен (DBD), слитые конструкции GAL4A-ДНК-связывающий домен и слитые конструкции вирус простого герпеса (HSV)-белок BP16. Дополнительные домены, которые могут образовывать часть слитого белка, содержащего фермент CRISPR, описаны в US20110059502, включенной в настоящее описание в качестве ссылки. В некоторых вариантах осуществления меченый фермент CRISPR используют для идентификации положения последовательности-мишени.

В одном аспекте настоящего изобретения в клетку может быть введен репортерный ген, который включает, но не ограничивается ими, ген глутатион-5-трансферазы (GST), пероксидазы хрена (HRP), хлорамфениколацетилтрансферазы (CAT), бета-галактозидазы, бета-глюкуронидазы, люциферазы, зеленого флуоресцентного белка (GFP), HcRed, DsRed, голубого флуоресцентного белка (CFP), желтого флуоресцентного белка (YFP) и аутофлуоресцентных белков, включая синий флуоресцентный белок (BFP), для кодирования продукта гена, который служит маркером, посредством которого измеряют изменение или модификацию экспрессии продукта гена. В следующем варианте осуществления настоящего изобретения молекула ДНК, кодирующая продукт гена, может быть введена в клетку с помощью вектора. В предпочтительном варианте осуществления настоящего изобретения продукт гена представляет собой люциферазу. В следующем варианте осуществления настоящего изобретения экспрессия продукта гена снижается.

Как правило, варианты осуществления промоторов настоящего изобретения включают: 1) полный промотор Pol III, который включает TATA-бокс, проксимальный элемент последовательности (PSE) и дистальный элемент последовательности (DSE); и 2) второй основной промотор Pol III, который включает PSE и TATA-бокс, слитые с 5′ концом DSE в обратной ориентации. TATA-бокс, который назван вследствие его нуклеотидной последовательности, является основной детерминантой специфичности Pol III. Он обычно расположен в положении между -23 и -30 нуклеотидами относительно

транскрибируемой последовательности, и он является основной детерминантой начала транскрибируемой последовательности. PSE обычно расположен между -45 и -66 нуклеотидами. DSE усиливает активность основного промотора Pol III. В промоторе H1 между PSE и DSE отсутствует пропуск.

Двунаправленные промоторы относятся к любому промотору (как правило, pol II), который ориентирован по принципу голова-хвост-хвост-голова - практически к любой контролирующей области, которая обеспечивает расходящуюся транскрипцию. В некоторых вариантах осуществления двунаправленный промотор H1 может представлять собой промотор pol II и pol II, или pol III и pol III, или их комбинации. Двунаправленный H1, также как 7Sk и другие, описанные в настоящем описании, являются уникальными тем, что они являются двунаправленными, что само по себе является необычным в эукариотических геномах, но наиболее важно в данном случае, контролируют кодирующий белок ген с одной стороны (pol II), и ген РНК с другой стороны (pol III).

В некоторых вариантах осуществления двунаправленные промоторы состоят из: 1) полного обычного однонаправленного промотора Pol III, который содержит 3 внешних элемента контроля: DSE, PSE и TATA-бокс; и 2) второго основного промотора Pol III, который включает PSE и TATA-бокс, слитые с 5'-концом DSE в обратной ориентации. TATA-бокс, который распознается TATA-связывающим белком, является необходимым для привлечения Pol III к промоторной области. Связывание TATA-связывающего белка с TATA-боксом стабилизируется взаимодействием SNAPc с PSE. Вместе эти элементы размещают Pol III правильным образом, так что она может транскрибировать экспрессируемую последовательность. DSE также необходим для полной активности промотора Pol III (Murphy et al. (1992) *Mol. Cell Biol.* 12:3247-3261; Mittal et al. (1996) *Mol. Cell Biol.* 16:1955-1965; Ford and Hernandez (1997) *J.Biol.Chem.*, 272:16048-16055; Ford et al. (1998) *Genes, Dev.*, 12:3528-3540; Hovde et al. (2002) *Genes Dev.* 16:2772-2777). Посредством взаимодействия факторов транскрипции Oct-1 и/или SBF/Staf с их мотивами в DSE транскрипция усиливается вплоть до 100 раз (Kunkel и Hixon (1998) *Nucl. Acid Res.*, 26:1536-1543).

Поскольку основные промоторы, имеющие прямую и обратную ориентацию, контролируют транскрипцию последовательностей на противоположных цепях двухцепочечных ДНК-матриц, положительная цепь основного промотора с обратной ориентации прилегает к 5′-концу отрицательной цепи DSE. Транскрипты, экспрессируемые под контролем промотора H1, терминируются непрерывной последовательностью из 4 или 5 T.

В промоторе H1 DSE прилегает к PSE и TATA-боксу (Myslinski et al. (2001) *Nucl. Acid Res.* 29:2502-2509). Для минимизации повторения последовательностей этот промотор сделали двунаправленным путем создания гибридного промотора, в котором транскрипция в обратном направлении контролируется прилегающими PSE и TATA-боксом, происходящими из промотора U6. Для облегчения конструирования двунаправленного промотора H1 между основным промотором в обратной ориентации и DSE может быть встроена небольшая спейсерная последовательность.

Примеры зависимых от клеточного цикла белков могут включать любые из следующих:

(a) Реакция репликации ДНК хромосом, включая инициацию репликации хромосомной ДНК и продвижение репликационной вилки

Mcm10, Orc1, Orc3, Cdc6, Cdt1, геминин, Mcm3, Mcm4, Mcm5, Mcm6, Mcm7, Mcm8, Cdc7, Cdc5, Psf1, Psf2, Psf3, Cdc45, Pola p180, Pola p70, Pola Spp1(Prim2a), RPA70, RPA34, PCNA, Elg1, лигаза 1, Pole Pol2, Pole Dpb3, топоизомераза I, TDP1, Orc2, Orc4, Orc5, Orc6, Mcm2, Dbf4, TopBP1, Sld5, Pola Spp2, RFC1, RFC2, RFC3, RFC4, RFC5, Pif1, Pold p50, Pole Dpb2, топоизомераза Iia, топоизомераза Iib, RPAI4, FEN1, DNA2, Pold p125, Pold p68, Pold p12, Pole Dpb4

(b) Точки контроля повреждения ДНК

ATR, Chk1, NBS1, Hus1, Rad1, Mad2, BubR1, ATM, Rad50, Mre11, Mdc1, 53BP1, Rad17, BubR1, ATRIP, Chk2, H2AX, RFC1, RFC2, RFC3, RFC4, RFC5, ATM, BRCA1, Chk1, Chk2, 14-3-3eta, 14-3-3sigma, cdc25A, cdc25c, wee1, ATR, ATRIP, Rad17, RFC2, RFC3, RFC4, RFC5, HUS1, Rad1, Rad9, P53, Rad50, Mre11, NBS1, TopBP1, 53BP1, H2AX

(c) Объединение и расхождение сестринских хроматид

Ctf18, Scc1, Scc3, Dcc1, Trf4-1, Trf4-2, Smc1, Smc3, Pds1(секурин), Mad-2, BubR1, Esp1

(d) Эксцизионная репарация оснований

UNG, MBD4, TDG, NTH1, NEIL2, NEIL3, APE2, PARP1, PNK, Polb, OGG1, APE1, XRCC1, Лигаза 3, SMUG1, TDG, MYH, MPG, NEIL1, ADPRT, ADPRTL2, MGMT, ABH1, ABH2, ABH3

(e) Эксцизионная репарация несоответствующих оснований

MSH2, PMS1, PMS2, MLH3, Экзонуклеаза 1, MSH3, MSH6, MSH5, MLH1, MSH4, PMS2L3, Trex1, Trex2, PMS2L4

(f) Эксцизионная репарация нуклеотидов

XPC, Rad23A, Rad23B, CSA, CSB, XPG, XPF, DDB1, DDB2, XAB2, XPB, ERCC1, XPD, XPA, DDB2, Mms19, CETN2, RPA70, RPA34, RPAI4, GTF2H1, GTF2H2, GTF2H3, GTF2H4, CDK7, CCNH, MNAT1, лигаза I, CSA, CSB

(g) Репарация посредством гомологичной рекомбинации

Rad51, Rad51L1, Rad51C, Rad51L3, DMC1, XRCC2, XRCC3, Rad52, Rad54L, Rad54B, BRCA1, BRCA2, Rad50, Mre11, NBS1, Топоизомераза IIIa, Топоизомераза IIIb, WHIP, WRN, BLM, RecQ1, RecQ5

(h) Репарация посредством негомологичного соединения концов (репарация посредством негомологичной рекомбинации)

Ku70, Ku80, DNA-pk, Лигаза 4, XRCC4, Artemis, WRN

i) Репарация двухцепочечных разрывов ДНК

Rad51, Rad51D, Xrcc2, Rad54, BRCA1, Ku80, XRCC4, Rad52, Rad51C, Dmc1, Rad54B, ДНК-pk, Ku70, Лигаза4, Rad51B, XRCC3, BRCA2, Artemis

(j) Пострепликационная репарация ДНК (толерантность к повреждению ДНК)

Rad6A, Rad6B, Rad18, Ubc13, FBH1

(k) Репарация повреждений посредством сшивок ДНК

FANCA, FANCC, FANCD2, FANCE, FANCF, FANCG

(l) Репарация повреждений посредством сшивок ДНК-белок

TDP1

(m) ДНК-полимераза

Poli, Polh, Polq, Polk, Polz (REV3), Poll, Polm, Rev1, Polb, Polg, Pold p50, Pole Pol2, REV7, Poln, Pola P180, Pola p70, Pola Spp1, Pola Spp2, Pold p68, Pold p12, Pole Dpb2, Pole

Dpb3, Pole Dpb4

(n) Нуклеотидная очистка

MTH1, DUT, p53R2

(o) Поддержание структуры хроматина

H2AX, Sir2, SIRT1 (сиртуин)

(p) Поддержание структуры теломер

Tin2, Sir2, hTert, TRF1, TRF2, танкираза, Pot1, Rap1, Pif1

Предпочтительные примеры генов, ассоциированных с каждой из вышеупомянутых функций (a)-(p), включают гены, примеры которых описаны ниже. Более конкретно, примерами таких генов являются следующие: APE2, ATR, BRCA1, Chk1, Cdc5, Cdc6, Cdc7, Cdc45, Cdt1, CSA, CSB, Ctf18, DDB1, DDB2, DNA2, DUT, Elg1, EndoV, Esp1, экзонуклеаза 1, FBH1, FEN1, геминин, Hus1, KNTC2 (NDC80), Ku80, Лигаза 1, Mad2, MBD4, Mcm3, Mcm4, Mcm5, Mcm6, Mcm7, Mcm8, Mcm10, MGMT, MLH3, Mms4, MPG, MSH2, Mus81, NBS1, NEIL2, NEIL3, NTH1, Orc1, Orc3, PARP1, PCNA, Pif1, PMS1, PMS2, PNK, Pola p180, Pola p70, Pola Spp1(Prim2a), Polb, Pold p125, Pole Dpb3, Pole Dpb4, Pole Pol2, Poli, Poll, Polm, Psf1, Psf2, Psf3, Rad1, Rad18, Rad23A, Rad23B, Rad51, Rad51D, Rad54, Rad6A, RPA34, RPA70, Scc1, Scc3, Sir2, SIRT1 (сиртуин), TDG, TDP1, TIMELESS, Tin2, топоизомераза I, топоизомераза IIIa, топоизомераза IIIb, Ubc13, UNG, XAB2, XPC, XPF, XPG, Xrcc2, и XRCC4.

Поскольку названия генов, описанных в настоящем описании, представляют собой названия, которые широко и общеизвестны, специалисты в данной области способны подходящим образом получить данные о нуклеотидных последовательностях указанных генов из общественной справочной базы данных или базы данных генов (например, GenBank), исходя из названия гена.

B. Способы

В некоторых вариантах осуществления описанное в настоящем описании изобретение также относится к способу изменения экспрессии одного или нескольких продуктов генов в эукариотической или неэукариотической клетке, где клетка содержит молекулу ДНК, кодирующую один или несколько продуктов генов, причем способ включает введение в клетку усовершенствованной и модифицированной не встречающейся в

природе системы CRISPR-Cas, описанной ранее в WO2015/195621 (включенной в настоящее описание в качестве ссылки в полном объеме). Такие усовершенствования включают не встречающуюся в природе нуклеазную систему (например, CRISPR-Cas9), содержащую один или несколько векторов, содержащих: a) промотор (например, ортологичный промотор H1 или 7sk), функционально связанный по меньшей мере с одной нуклеотидной последовательностью, кодирующей гидовую РНК (гРНК) нуклеазной системы, где гРНК гибридизуется с последовательностью молекулы ДНК, являющейся мишенью, в клетке, и где молекула ДНК кодирует один или несколько продуктов генов, экспрессируемых в клетке; и b) регуляторный элемент, функционирующий в клетке, функционально связанный с нуклеотидной последовательностью, кодирующей нацеленную на геном нуклеазу (например, модифицированный белок Cas9), где компоненты (a) и (b) расположены на одном и том же или на различных векторах системы, где гРНК осуществляет нацеливание на последовательность-мишень и гибридизуется с ней, и нуклеаза расщепляет молекулу ДНК, изменяя экспрессию одного или нескольких продуктов генов. В некоторых вариантах осуществления система упакована в одну частицу аденоассоциированного вируса (AAV). В некоторых вариантах осуществления система инактивирует один или несколько продуктов генов. В некоторых вариантах осуществления система вырезает по меньшей мере одну мутацию гена. В некоторых вариантах осуществления промотор H1 содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 12 или 32. В некоторых вариантах осуществления ортологичный промотор H1 происходит из мыши или крысы. В некоторых вариантах осуществления ортологичный промотор H1 содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 84-119. В некоторых вариантах осуществления ортологичный промотор H1 содержит нуклеотидные последовательности из группы, состоящей из SEQ ID

NO: 84-119. В некоторых вариантах осуществления промотор H1 содержит: a) элементы контроля, которые обеспечивают транскрипцию в одном направлении по меньшей мере одной нуклеотидной последовательности, кодирующей гРНК; и b) элементы контроля, которые обеспечивают транскрипцию в противоположном направлении нуклеотидной последовательности, кодирующей нацеленную на геном нуклеазу. В некоторых вариантах осуществления промотор представляет собой промотор 7sk человека, мыши или крысы. В некоторых вариантах осуществления 7sk выбран из группы, состоящей из 7sk1 (вариант 1 7sk, в котором используется эндогенная 5'UTR GSTA4), 7sk2 (вариант 2 7sk, в котором используется только последовательность Козак в 5'UTR) и 7sk3 (вариант 7sk, в котором используется 5'UTR бета-глобина). В некоторых вариантах осуществления 7sk1 человека содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 3. В некоторых вариантах осуществления 7sk1 мыши содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с SEQ ID NO: 6. В некоторых вариантах осуществления 7sk2 человека содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с SEQ ID NO: 4. В некоторых вариантах осуществления 7sk2 мыши содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с SEQ ID NO: 7. В некоторых вариантах осуществления 7sk3 человека содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с SEQ ID NO: 5. В некоторых вариантах осуществления 7sk3 мыши содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с SEQ ID NO: 8. В некоторых вариантах осуществления ортологичный промотор H1 или промотор 7sk имеет по меньшей мере одну модификацию. В некоторых вариантах осуществления по меньшей мере одна модификация

промотора содержит элемент, который позволяет зависимую от условий регуляцию. В некоторых вариантах осуществления элемент представляет собой отвечающий на tet промотор. В некоторых вариантах осуществления отвечающий на tet промотор содержит Tet-репрессор (TetR) и Tet-оператор (TetO) встроенный в промотор H1. В некоторых вариантах осуществления TetR содержит аминокислотную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с SEQ ID NO: 14. В некоторых вариантах осуществления TetO содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 11. В некоторых вариантах осуществления H1-TetO содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 13. В некоторых вариантах осуществления по меньшей мере одна модификация промотора включает участок, который позволяет ауторегуляцию. В некоторых вариантах осуществления участок ауторегуляции содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 33 или 34. В некоторых вариантах осуществления нуклеазная система дополнительно содержит по меньшей мере одну последовательность терминатора. В некоторых вариантах осуществления последовательность терминатора представляет собой любую последовательность нуклеиновой кислоты, выбранную из группы, состоящей из SEQ ID NO: 120-130. В некоторых вариантах осуществления по меньшей мере одна последовательность терминатора выбрана из группы, состоящей из последовательности SV40 или синтетических поли A (SPA) последовательностей. В некоторых вариантах осуществления последовательность терминатора представляет собой последовательность SV40 из 120 (SEQ ID NO: 123) или 240 (SEQ ID NO: 122) пар оснований. В некоторых вариантах осуществления последовательность терминатора представляет собой последовательность SPA из 49 пар оснований. В

некоторых вариантах осуществления нуклеазная система дополнительно содержит последовательности 5′-нетранслируемой области (5′UTR). В некоторых вариантах осуществления нуклеазная система дополнительно содержит последовательность Козак. В некоторых вариантах осуществления последовательность Козак содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с SEQ ID NO: 1 или 2. В некоторых вариантах осуществления нуклеазная система дополнительно содержит последовательность РНК, которая опосредует кэп-независимую инициацию трансляции. В некоторых вариантах осуществления последовательность РНК выбрана из группы, состоящей из 6.947 или 6.967 (Wellensiek I. (2013) *Nature Methods*, 10:747-750). В некоторых вариантах осуществления Cas9 содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 61. В некоторых вариантах осуществления Cas9 содержит по меньшей мере одну модификацию. В некоторых вариантах осуществления по меньшей мере одна модификация в Cas9 включает изменение расщепляющей последовательности. В некоторых вариантах осуществления нуклеаза выбрана из группы, состоящей из Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (также известной как Csn1 и Csx12), Cas10, Cas13a, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, Cpf1, C2c1, C2c2 и C2c3. В некоторых вариантах осуществления изменение в расщепляющей последовательности выбрано из группы, состоящей из T2A, P2A, E2A и F2A. В некоторых вариантах осуществления T2A содержит аминокислотную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 36. В некоторых вариантах осуществления P2A содержит аминокислотную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной

последовательностью, указанной в SEQ ID NO: 35. В некоторых вариантах осуществления P2A содержит аминокислотную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 35. В некоторых вариантах осуществления E2A содержит аминокислотную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 37. В некоторых вариантах осуществления F2A содержит аминокислотную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 38. В некоторых вариантах осуществления по меньшей мере одна модификация в Cas9 является кодон-оптимизированной для экспрессии в клетке. В некоторых вариантах осуществления Cas9 дополнительно содержит линкерную последовательность, функционально слитую в рамке считывания зависимым от клеточного цикла белком (слитая конструкция Cas9). В некоторых вариантах осуществления зависимый от клеточного цикла белок выбран из группы, состоящей из APE2, ATR, 53BP1, BRCA1, Chk1, Cdc5, Cdc6, Cdc7, Cdc45, Cdt1, CSA, CSB, Ctf18, CtIP, DDB1, DDB2, DNA2, DUT, Elg1, EndoV, Esp1, экзонуклеазы 1, FBH1, FEN1, геминина, Hus1, KNTC2 (NDC80), Ku80, лигазы 1, Mad2, MBD4, Mcm3, Mcm4, Mcm5, Mcm6, Mcm7, Mcm8, Mcm10, MGMT, MLH3, Mms4, MPG, MSH2, Mus81, NBS1, NEIL2, NEIL3, NTH1, Orc1, Orc3, PARP1, PCNA, Pif1, Pin1, PMS1, PMS2, PNK, Pola p180, Pola p70, Pola Spp1 (Prim2a), Polb, Pold p125, Pole Dpb3, Pole Dpb4, Pole Pol2, Poli, Poll, Polm, Psf1, Psf2, Psf3, Rad1, Rad18, Rad23A, Rad23B, Rad51, Rad51D, Rad54, Rad6A, Rag1, Rag2, RPA34, RPA70, Scc1, Scc3, Sir2, SIRT1 (сиртуин), TDG, TDP1, TIMELESS, Tin2, топоизомеразы I, топоизомеразы IIIa, топоизомеразы IIIb, Ubc13, UNG, XAB2, XPC, XPF, XPG, Xrcc2 и XRCC4. В некоторых вариантах осуществления зависимый от клеточного цикла белок представляет собой геминин. В некоторых вариантах осуществления зависимый от клеточного цикла белок представляет собой геминин человека. В некоторых вариантах осуществления геминин человека содержит аминокислоты из положений 1-110 (hGem(1-110)). В некоторых вариантах

осуществления hGem(1-110) содержит аминокислотную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с аминокислотной последовательностью, указанной в SEQ ID NO: 19. В некоторых вариантах осуществления зависимый от клеточного цикла белок представляет собой Cdt1. В некоторых вариантах осуществления зависимый от клеточного цикла белок представляет собой Cdt1 человека. В некоторых вариантах осуществления Cdt1 человека содержит аминокислоты из положений 30-120 (hCdt1(30-120)). В некоторых вариантах осуществления hCdt1(30-120) содержит аминокислотную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с аминокислотной последовательностью, указанной в SEQ ID NO: 18. В некоторых вариантах осуществления слитая конструкция Cas9 содержит аминокислотную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с аминокислотной последовательностью, выбранной из группы, состоящей из SEQ ID NO: 20-27. В некоторых вариантах осуществления Cas9 функционально слита в рамке считывания с убиквитином (Ub-Cas9). В некоторых вариантах осуществления Ub-Cas9 содержит по меньшей мере одну N-концевую модификацию. В некоторых вариантах осуществления Ub-Cas9 с N-концевой модификацией содержит аминокислотную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с аминокислотной последовательностью, выбранной из группы, состоящей из SEQ ID NO: 39-58. В некоторых вариантах осуществления белок убиквитин содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 60. В некоторых вариантах осуществления нуклеазная система дополнительно содержит никазу SaCas9. В некоторых вариантах осуществления нуклеазная система дополнительно содержит донорную матричную последовательность. В некоторых вариантах осуществления по меньшей мере один вектор содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%,

95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 62 или 67. В некоторых вариантах осуществления донорная матричная последовательность корректирует по меньшей мере одну мутацию гена. В некоторых вариантах осуществления по меньшей мере одна мутация гена представляет собой *rd10* или *rd12*. В некоторых вариантах осуществления по меньшей мере один вектор содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 63 или 64. В некоторых вариантах осуществления по меньшей мере один вектор содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 65 или 66. В некоторых вариантах осуществления промотор функционально связан по меньшей мере с одной, двумя, тремя, четырьмя, пятью, шестью, семью, восемью, девятью или десятью гРНК. В некоторых вариантах осуществления последовательность-мишень содержит нуклеотидную последовательность $AN_{19}NGG$, $GN_{19}NGG$, $CN_{19}NGG$ или $TN_{19}NGG$. В некоторых вариантах осуществления клетка представляет собой эукариотическую или неэукариотическую клетку. В некоторых вариантах осуществления эукариотическая клетка представляет собой клетку млекопитающего или человека. В некоторых вариантах осуществления эукариотическая клетка представляет собой фоторецепторную клетку сетчатки. В некоторых вариантах осуществления один или несколько продуктов генов представляют собой родопсин. В некоторых вариантах осуществления экспрессия одного или нескольких продуктов генов снижается. В некоторых вариантах осуществления клетка представляет собой ганглионарную клетку сетчатки. В некоторых вариантах осуществления эукариотическая клетка представляет собой злокачественную клетку. В некоторых вариантах осуществления пролиферация злокачественной клетки ингибируется или снижается. В некоторых вариантах осуществления в злокачественной клетке усиливается или повышается апоптоз.

В некоторых вариантах осуществления описанное в настоящем

описании изобретение также относится к способу изменения экспрессии одного или нескольких продуктов генов в клетке, где клетка содержит молекулу ДНК, кодирующую один или несколько продуктов генов, причем способ включает введение в клетку не встречающейся в природе системы CRISPR-Cas, содержащей один или несколько векторов, содержащих: a) промотор H1, функционально связанный по меньшей мере с одной нуклеотидной последовательностью, кодирующей гидовую РНК (гРНК) системы CRISPR-Cas, где гРНК гибридизуется с последовательностью-мишенью молекулы ДНК; и b) регуляторный элемент, функционирующий в клетке, функционально связанный с нуклеотидной последовательностью, кодирующей белок Cas9, где компоненты (a) и (b) расположены на одном и том же или на различных векторах системы, где гРНК осуществляет нацеливание на последовательность-мишень и гибридизуется с ней и белок Cas9 расщепляет молекулу ДНК, изменяя экспрессию одного или нескольких продуктов генов.

В некоторых вариантах осуществления настоящее изобретение также относится к способу изменения экспрессии одного или нескольких продуктов генов в эукариотической клетке, где клетка содержит молекулу ДНК, кодирующую один или несколько продуктов генов, причем способ включает введение в клетку не встречающейся в природе системы CRISPR-Cas, содержащей один или несколько векторов, содержащих: a) промотор H1, функционально связанный по меньшей мере с одной нуклеотидной последовательностью, кодирующей гидовую РНК (гРНК) системы CRISPR-Cas, где гРНК гибридизуется с последовательностью-мишенью молеклуы ДНК; и b) регуляторный элемент, функционирующий в эукариотической клетке, функционально связанной с нуклеотидной последовательностью, кодирующей белок Cas9 типа II, где компоненты (a) и (b) расположены на одном и том же или на различных векторах системы, причем гРНК осуществляет нацеливание на последовательность-мишень и гибридизуется с ней, и белок Cas9 расщепляет молекулу ДНК, и при этом экспрессия одного или нескольких продуктов генов изменяется. В одном аспекте последовательность-мишень может представлять собой последовательность-мишень, которая начинается

с любого нуклеотида, например, $N_{20}NGG$. В некоторых вариантах осуществления последовательность-мишень содержит нуклеотидную последовательность $AN_{19}NGG$. В некоторых вариантах осуществления последовательность-мишень содержит нуклеотидную последовательность $GN_{19}NGG$. В некоторых вариантах осуществления последовательность-мишень содержит нуклеотидную последовательность $CN_{19}NGG$. В некоторых вариантах осуществления последовательность-мишень содержит нуклеотидную последовательность $TN_{19}NGG$. В некоторых вариантах осуществления последовательность-мишень содержит нуклеотидную последовательность $AN_{19}NGG$ или $GN_{19}NGG$. В другом аспекте белок Cas9 является кодон-оптимизированным для экспрессии в клетке. В другом аспекте, белок Cas9 является кодон-оптимизированным для экспрессии в эукариотической клетке. В следующем аспекте, эукариотическая клетка представляет собой клетку млекопитающего или человека. В другом аспекте экспрессия одного или нескольких продуктов генов снижается.

Настоящее изобретение также относится к способу изменения экспрессии одного или нескольких продуктов генов в эукариотической клетке, где клетка содержит молекулу ДНК, кодирующую один или несколько продуктов генов, причем способ включает введение в клетку не встречающейся в природе системы CRISPR-Cas, содержащей вектор, содержащий двунаправленный промотор H1, где двунаправленный промотор H1 содержит: a) элементы контроля, которые обеспечивают транскрипцию в одном направлении по меньшей мере одной нуклеотидной последовательности, кодирующей гидовую РНК (гРНК) системы CRISPR-Cas, где гРНК гибридизуется с последовательность-мишенью молекулы ДНК; и b) элементы контроля, которые обеспечивают транскрипцию в противоположном направлении нуклеотидной последовательности, кодирующей белок Cas9 типа II, причем гРНК осуществляет нацеливание на последовательность-мишень и гибридизуется с ней, и белок Cas9 расщепляет молекулу ДНК, и при этом экспрессия одного или нескольких продуктов генов изменяется. В одном аспекте последовательность-мишень может представлять собой последовательность-мишень, которая начинается

с любого нуклеотида, например, $N_{20}NGG$. В некоторых вариантах осуществления последовательность-мишень содержит нуклеотидную последовательность $AN_{19}NGG$. В некоторых вариантах осуществления последовательность-мишень содержит нуклеотидную последовательность $GN_{19}NGG$. В некоторых вариантах осуществления последовательность-мишень содержит нуклеотидную последовательность $CN_{19}NGG$. В некоторых вариантах осуществления последовательность-мишень содержит нуклеотидную последовательность $TN_{19}NGG$. В другом аспекте последовательность-мишень содержит нуклеотидную последовательность $AN_{19}NGG$ или $GN_{19}NGG$. В другом аспекте белок Cas9 является кодон-оптимизированным для экспрессии в клетке. В другом аспекте белок Cas9 является кодон-оптимизированным для экспрессии в эукариотической клетке. В следующем аспекте эукариотическая клетка представляет собой клетку млекопитающего или человека. В другом аспекте экспрессия одного или нескольких продуктов генов снижается.

В некоторых аспектах настоящее изобретение относится к способам, включающим доставку одного или нескольких полинуклеотидов, таких как один или несколько векторов, как описано в настоящем описании, одного или нескольких его транскриптов и/или одного или нескольких белков, транслированных с них, в клетку-хозяина. В некоторых аспектах, настоящее изобретение, кроме того, относится к клеткам, продуцируемым такими способами, и к организмам (таким как животные. растения или грибы), содержащим или полученным из таких клеток. В некоторых вариантах осуществления фермент CRISPR в комбинации с (и необязательно в комплексе с) гидовой последовательностью доставляют в клетку. Для введения нуклеиновых кислот в клетки млекопитающих или ткани-мишени можно использовать общепринятые вирусные и невирусные способы переноса генов. Такие способы можно использовать для введения нуклеиновых кислот, кодирующих компоненты системы CRISPR в клетки в культуре или в организм-хозяин. Невирусные системы доставки векторов включают ДНК-плазмиды, РНК (например, транскрипт вектора, описанного в настоящем описании), голую нуклеиновую кислоту и нуклеиновую

кислоту в комплексе с носителем для доставки, таким как липосома. Вирусные системы доставки векторов включают ДНК- и РНК-вирусы, которые имеют либо эписомальные, либо встроенные геномы после доставки в клетку. Для обзора методик генной терапии см. Anderson (1992) *Science* 256:808-813; Nabel and Felgner (1993) TIBTECH 11:211-217; Mitani and Caskey (1993) TIBTECH 11:162-166; Dillon (1993) TIBTECH 11:167-175; Miller (1992) *Nature* 357:455-460; Van Brunt (1998) *Biotechnology* 6(10): 1149-1154; Vigne (1995) *Restorative Neurology and Neuroscience* 8:35-36; Kremer and Perricaudet (1995) *British Medical Bulletin* 51(1):31-44; Haddada et al. (1995) Current Topics in Microbiology and Immunology. Doerfler and Bohm (eds); and Yu et al. (1994) *Gene Therapy* 1:13-26.

Способы невирусной доставки нуклеиновых кислот включают липофекцию, нуклеофекцию, микроинъекцию, биолистические способы доставки, виросомы, липосомы, иммунолипосомы, конъюгаты поликатион или липид:нуклеиновая кислота, голую ДНК, искусственные вирионы и усиленный агентами захват ДНК. Липофекция описана, например, в патентах США № 5049386, 4946787 и 4897355) и реагенты для липофекции продаются коммерчески (например, Transfectam™ и Lipofectin™). Катионные и нейтральные липиды, которые являются пригодными для эффективной липофекции полинуклеотидов с помощью распознавания рецепторов включают липиды согласно Felgner, WO 91/17424; WO 91/16024. Доставка может осуществляться в клетки (например, введение *in vitro* или *ex vivo*) или ткани-мишени (например, введение *in vivo*).

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

Преимуществом применения систем на основе РНК- или ДНК-

вирусов для доставки нуклеиновых кислот являются высокоразвитые процессы нацеливания вируса на конкретные клетки в организме и транспорта вирусного груза в ядро. Вирусные векторы можно вводить непосредственно пациентам (*in vivo*), или их можно использовать для обработки клеток *in vitro*, и модифицированные клетки необязательно можно вводить пациентам (*ex vivo*). Общепринятые вирусные системы могут включать ретровирусные, лентивирусные, аденовирусные векторы для переноса генов, векторы на основе аденоассоциированного вируса и вируса простого герпеса. Встраивание в геном хозяина является возможным при использовании способов переноса генов на основе ретровируса, лентивируса и аденоассоциированного вируса, что часто приводит к длительной экспрессии встроенного трансгена. Кроме того, во многих различных типах клеток и тканей-мишеней наблюдают высокую эффективность трансдукции.

Тропизм ретровируса может быть изменен путем включения чужеродных белков оболочки, расширяющих потенциальную популяцию клеток-мишеней. Лентивирусные векторы представляют собой ретровирусные векторы, которые способны трансдуцировать или инфицировать неделящиеся клетки и, как правило, продуцируют высокие титры вируса. Таким образом, выбор ретровирусной системы переноса генов может зависеть от ткани-мишени. Ретровирусные векторы состоят из цис-действующих длинных концевых повторов со способностью к упаковке вплоть до 6-10 т.п.н. чужеродной последовательности. Минимальные цис-действующие LTR являются достаточными для репликации и упаковывания векторов, а затем они используются для встраивания терапевтического гена в клетку-мишень для обеспечения постоянной экспрессии трансгена. Широко используемые ретровирусные векторы включают векторы на основе вируса лейкоза мышей (MuLV), вируса лейкоза гиббонов (GaLV), вируса иммунодефицита обезьян (SIV), вируса иммунодефицита человека (ВИЧ) и их комбинаций (например, Buchscher et al. (1992) *J. Virol.* 66:2731-2739; Johann et al. (1992) *J. Virol.* 66:1635-1640; Sommnerfelt et al. (1990) *J. Virol.* 176:58-59; Wilson et al. (1989) *J. Virol.* 63:2374-2378; Miller et al. (1991) *J. Virol.* 65:2220-2224; PCT/US94/05700). В применениях, в

которых является предпочтительной временная экспрессия, можно использовать системы на основе аденовируса. Векторы на основе аденовируса способны к очень высокой эффективности трансдукции во многих типах клеток и не требуют деления клеток. С использованием таких факторов могут быть достигнуты высокие титры и уровни экспрессии. Этот вектор можно получать в больших количествах в относительно простой системе. Также для трансдукции клеток нуклеиновыми кислотами-мишенями, например, при продукции нуклеиновых кислот и пептидов *in vitro*, и для методик генной терапии *in vivo* и *ex vivo* можно использовать векторы на основе аденоассоциированного вируса ("AAV") (например, West et al. (1987) *Virology* 160:38-47; патент США № 4797368; WO 93/24641; Kotin (1994) *Human Gene Therapy* 5:793-801; Muzyczka (1994) *J. Clin. Invest.* 94:1351. Конструирование рекомбинантных векторов AAV описано в ряде публикаций, включая патент США № 5173414; Tratschin et al. (1985) *Mol. Cell. Biol.* 5:3251-3260; Tratschin et al. (1984) *Mol. Cell. Biol.* 4:2072-2081; Hermonat and Muzyczka (1984) *Proc. Natl. Acad. Sci. U.S.A.* 81:6466-6470; и Samulski et al. (1989) *J. Virol.* 63:03822-3828.

Для формирования вирусных частиц, которые способны инфицировать клетку-хозяина, обычно используют упаковывающие клетки. Такие клетки включают клетки 293, которые упаковывают аденовирус, и клетки ψ2 или клетки PA317, которые упаковывают ретровирус. Вирусные векторы, используемые в генной терапии, обычно получают путем получения клеточной линии, которая упаковывает вектор с нуклеиновой кислотой в вирусную частицу. Векторы, как правило, содержат минимальные вирусные последовательности, требуемые для упаковывания и последующего встраивания в хозяина, а другие вирусные последовательности заменены экспрессирующей кассетой для полинуклеотида(ов), подлежащего экспрессии. Недостающие вирусные функции, как правило, предоставляются в транс-форме упаковывающей клеточной линией. Например, векторы AAV, используемые для генной терапии, обычно имеют из генома AAV только последовательности ITR, которые требуются для упаковывания и встраивания в геном

хозяина. Вирусная ДНК упаковывается в клеточной линии, которая содержит плазмиду-помощника, кодирующую другие гены AAV, а именно, rep и cap, но лишенные последовательностей ITR. Клеточная линия также может быть инфицирована аденовирусом в качестве помощника. Вирус-помощник обеспечивает репликацию вектора AAV и экспрессию генов AAV с плазмиды-помощника. Плазмида-помощник не упаковывается в значительных количествах вследствие отсутствия последовательностей ITR. Контаминация аденовирусом может быть снижена, например, посредством тепловой обработки, к которой аденовирус является более чувствительным, чем AAV. Дополнительные способы доставки нуклеиновых кислот в клетки известны специалистам в данной области. См., например, US20030087817, включенную в настоящее описание в качестве ссылки.

В некоторых вариантах осуществления клетку-хозяина временно или не временно трансфицируют одним или несколькими векторами, описанными в настоящем описании. В некоторых вариантах осуществления клетка трансфицируется естественным образом у индивидуума. В некоторых вариантах осуществления клетка, которую трансфицируют, взята от индивидуума. В некоторых вариантах осуществления клетка происходит из клеток, взятых от индивидуума, как например, клеточная линия. В данной области известно широкое множество клеточных линий для культивирования тканей. Примеры клеточных линий включают, но не ограничиваются ими, C8161, CCRF-CEM, MOLT, mIMCD-3, NHDF, HeLa-S3, Huh1, Huh4, Huh7, HUVEC, HASMC, HEKn, HEKa, MiaPaCell, Panel, PC-3, TF1, CTLL-2, C1R, Rat6, CV1, RPTE, A10, T24, J82, A375, ARH-77, Calu1, SW480, SW620, SKOV3, SK-UT, CaCo2, P388D1, SEM-K2, WEHI-231, HB56, TIB55, Jurkat, J45.01, LRMB, Bcl-1, BC-3, IC21, DLD2, Raw264.7, NRK, NRK-52E, MRC5, MEF, Hep G2, HeLa B, HeLa T4, COS, COS-1, COS-6, COS-M6A, линию клеток эпителия почки обезьяны BS-C-1, фибробласты эмбриона мыши BALB/3T3, 3T3 Swiss, 3T3-L1, фетальные фибробласты человека 132-d5; фибробласты мыши 10.1, 293-T, 3T3, 721, 9L, A2780, A2780ADR, A2780cis, A172, A20, A253, A431, A-549, ALC, B16, B35, BCP-1 cells, BEAS-2B, bEnd.3, BHK-21, BR 293, BxPC3, C3H-10T1/2, C6/36, Cal-27, CHO, CHO-7, CHO-

IR, CHO-K1, CHO-K2, CHO-T, CHO Dhfr –/–, COR-L23, COR-L23/CPR, COR-L23/5010, COR-L23/R23, COS-7, COV-434, CML T1, CMT, CT26, D17, DH82, DU145, DuCaP, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10,0, FM3, H1299, H69, HB54, HB55, HCA2, HEK-293, HeLa, Hepa1c1c7, HL-60, HMEC, HT-29, Jurkat, клетки JY, клетки K562, Ku812, KCL22, KG1, KYO1, LNCap, Ma-MeI 1-48, MC-38, MCF-7, MCF-10A, MDA-MB-231, MDA-MB-468, MDA-MB-435, MDCK II, MDCK II, MOR/0,2R, MONO-MAC 6, MTD-1A, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NALM-1, NW-145, клеточные линии OPCN/OPCT, Peer, PNT-1A/PNT 2, RenCa, RIN-5F, RMA/RMAS, клетки Saos-2, Sf-9, SkBr3, T2, T-47D, T84, клеточную линию THP1, U373, U87, U937, VCaP, клетки Vero, WM39, WT-49, X63, YAC-1, YAR и их трансгенные варианты. Клеточные линии доступны из различных источников, известных специалисту в данной области (см., например, American Type Culture Collection (ATCC) (Manassus, Va.)). В некоторых вариантах осуществления клетку, трансфицированную одним или несколькими векторами, описанными в настоящем описании, используют для получения новой клеточной линии, содержащей одну или несколько происходящих из вектора последовательностей. В некоторых вариантах осуществления клетку, временно трансфицированную компонентами системы CRISPR, как описано в настоящем описании (например, посредством временной трансфекции одного или нескольких векторов или трансфекции РНК), и модифицированную посредством активности комплекса CRISPR, используют для получения новой клеточной линии, содержащей клетки, содержащие модификацию, но лишенные любой другой экзогенной последовательности. В некоторых вариантах осуществления клетки, временно или не временно трансфицированные одним или несколькими векторами, описанными в настоящем описании, или клеточные линии, происходящие из таких клеток, используют для оценки одного или нескольких исследуемых соединений.

В некоторых вариантах осуществления один или несколько векторов, описанных в настоящем описании, используют для получения не являющегося человеком трансгенного животного. В

некоторых вариантах осуществления трансгенное животное представляет собой млекопитающее, такое как мышь, крыса или кролик. В определенных вариантах осуществления организмом или индивидуумом является растение. Способы получения трансгенных животных известны в данной области и обычно начинаются со способа трансфекции клеток, как описано в настоящем описании.

В одном аспекте настоящее изобретение относится к способам модификации полинуклеотида-мишени в эукариотической клетке, которая может происходить *in vivo*, *ex vivo* или *in vitro*. В некоторых вариантах осуществления способ включает взятие образца клетки или популяции клеток от человека или не являющегося человеком животного и модификацию клетки или клеток. На любой стадии *ex vivo* может происходить культивирование. Клетку или клетки можно даже повторно вводить не являющемуся человеком животному.

В одном аспекте настоящее изобретение относится к способам модификации полинуклеотида-мишени в эукариотической клетке. В некоторых вариантах осуществления способ включает обеспечение возможности комплексу CRISPR связаться с полинуклеотидом-мишенью для обеспечения расщепления полинуклеотида-мишени, тем самым модифицируя полинуклеотид-мишень, где комплекс CRISPR содержит фермент CRISPR в комплексе с гидовой последовательностью, гибридизованной с последовательностью-мишенью в полинуклеотиде-мишени.

В одном аспекте настоящее изобретение относится к способу модификации экспрессии полинуклеотида в эукариотической клетке. В некоторых вариантах осуществления способ включает обеспечение возможности комплексу CRISPR связаться с полинуклеотидом, так что связывание приводит к увеличению или снижению экспрессии полинуклеотида; где комплекс CRISPR содержит фермент CRISPR в комплексе с гидовой последовательностью, гибридизованной с последовательностью-мишенью в полинуклеотиде.

В одном аспекте настоящее изобретение относится к способам применения одного или нескольких элементов системы CRISPR. Комплекс CRISPR по настоящему изобретению является эффективным средством для модификации полинуклеотида-мишени. Комплекс CRISPR

по настоящему изобретению имеет широкую применимость, включая модификацию (например, делецию, инсерцию, транслокацию, инактивацию, активацию) полинуклеотида-мишени в множестве типов клеток. Таким образом, комплекс CRISPR по настоящему изобретению имеет широкий спектр применений, например, в генной терапии, скрининге лекарственных средств, диагностике и прогнозировании заболеваний. Иллюстративный комплекс CRISPR содержит фермент CRISPR в комплексе с гидовой последовательностью, гибридизованной с полинуклеотидом-мишенью.

Полинуклеотид-мишень комплекса CRISPR может представлять собой любой полинуклеотид, являющийся эндогенным или экзогенным для эукариотической клетки. Например, полинуклеотид-мишень может представлять собой полинуклеотид, находящийся в ядре эукариотической клетки. Полинуклеотид-мишень может представлять собой последовательность, кодирующую продукт гена (например, белок) или некодирующую последовательность (например, регуляторный полинуклеотид или мусорную ДНК). Без связи с теорией полагают, что последовательность-мишень должна быть ассоциирована с PAM (примыкающий к протоспейсеру мотив); т.е. короткой последовательностью, распознаваемой комплексом CRISPR. Точные требования к последовательности и длине PAM различаются в зависимости от используемого фермента CRISPR, однако обычно PAM представляют собой последовательности из 2-5 пар оснований рядом с протоспейсером (т.е. последовательностью-мишенью). Примеры последовательностей PAM приведены в разделе примеров ниже, и специалист в данной области способен идентифицировать дополнительные последовательности PAM для применения с данным ферментом CRISPR.

Примеры полинуклеотидов-мишеней включают последовательность, ассоциированную с биохимическим каскадом передачи сигнала, например, ген или полинуклеотид, ассоциированный с биохимическим каскадом передачи сигнала. Примеры полинуклеотидов-мишеней включают ассоциированный с заболеванием ген или полинуклеотид. "Ассоциированный с заболеванием" ген или полинуклеотид относится к любому гену или полинуклеотиду, который обеспечивает продукты транскрипции или

трансляции на ненормальном уровне или в аномальной форме в клетках, происходящих из пораженных заболеванием тканей, по сравнению с контрольными тканями или клетками без заболевания. Он может представлять собой ген, который становится экспрессируемым на аномально высоком уровне; он может представлять собой ген, который становится экспрессируемым на аномально низком уровне, где измененная экспрессия коррелирует с возникновением и/или прогрессированием заболевания. Ассоциированный с заболеванием ген также относится к гену, обладающему мутацией(ями) или генетическим вариантом, который прямо ответственен за этиологию заболевания, или находится в неравновесии по сцеплению с геном(ами), который ответственен за этиологию заболевания. Транскрибированные или транслированные продукты могут быть известными или неизвестными, или могут быть на нормальном или аномальном уровне.

Варианты осуществления настоящего изобретения также относятся к способам и композициям, связанным с нокаутом генов, амплификацией генов и репарацией конкретных мутаций, ассоциированных с нестабильностью повторов ДНК и неврологическими нарушениями (Robert D. Wells, Tetsuo Ashizawa, Genetic Instabilities and Neurological Diseases, Second Edition, Academic Press, Oct. 13, 2011-Medical). Было обнаружено, что конкретные аспекты последовательностей с тандемными повторами ответственны за более чем двадцать заболеваний у человека (McIvor et al. (2010) *PHK Biol.* 7(5):551-8). Система CRISPR-Cas может быть приспособлена для коррекции этих дефектов, обусловленных геномной нестабильностью.

В другом аспекте настоящего изобретения систему CRISPR-Cas можно использовать для коррекции дефектов глаз, которые возникают в результате нескольких генетических мутаций, дополнительно описанных в Traboulsi, ed. (2012) Genetic Diseases of the Eye, Second Edition, Oxford University Press.

Несколько дополнительных аспектов настоящего изобретения касаются коррекции дефектов, ассоциированных с широким диапазоном генетических заболеваний. Например, генетические заболевания головного мозга могут включать, но не ограничиваются

ими, адренолейкодистрофию, аганез мозолистого тела, синдром Айкарди, болезнь Альперса, болезнь Альцгеймера, синдром Барта, болезнь Баттена, CADASIL, мозжечковую дегенерацию, болезнь Фабри, болезнь Герстманна-Штреусслера-Шейнкера, болезнь Гентингтона и другие обусловленные триплетными повторами нарушения, болезнь Ли, синдром Леш-Найхена, болезнь Менкеса, митохондриальные миопатии и кольпоцефалию NINDS.

В некоторых вариантах осуществления состояние может представлять собой новообразование. В некоторых вариантах осуществления состояние может представлять собой связанную со старением дегенерацию желтого пятна. В некоторых вариантах осуществления состояние может представлять собой шизофреническое расстройство. В некоторых вариантах осуществления состояние может представлять собой обусловленное тринуклеотидными повторами нарушение. В некоторых вариантах осуществления состояние может представлять собой синдром ломкой X-хромосомы. В некоторых вариантах осуществления состояние может представлять собой обусловленное секретазой нарушение. В некоторых вариантах осуществления состояние может представлять собой обусловленное прионами нарушение. В некоторых вариантах осуществления состояние может представлять собой ALS. В некоторых вариантах осуществления состояние может представлять собой наркотическую зависимость. В некоторых вариантах осуществления состояние может представлять собой аутизм. В некоторых вариантах осуществления состояние может представлять собой болезнь Альцгеймера. В некоторых вариантах осуществления состояние может представлять собой воспаление. В некоторых вариантах осуществления состояние может представлять собой болезнь Паркинсона.

Примеры белков, ассоциированных с болезнью Паркинсона, включают, но не ограничиваются ими, α-синуклеин, DJ-1, LRRK2, PINK1, паркин, UCHL1, синфилин-1 и NURR1.

Примеры связанных с пристрастием белков могут включать, например, ABAT.

Примеры связанных с воспалением белков могут включать, например, моноцитарный хемоаттрактантный белок 1 (MCP1),

кодируемый геном Ccr2, рецептор C-C хемокинов 5 типа (CCR5), кодируемый геном Ccr5, рецептор IgG IIB (FCGR2b, также называемый CD32), кодируемый геном Fcgr2b или белок Fc-эпсилон R1g (FCER1g), кодируемый геном Fcer1g.

Примеры белков, ассоциированных с сердечно-сосудистыми заболеваниями, могут включать, например, IL1B (интерлейкин 1 бета), XDH (ксантиндегидрогеназа), TP53 (белок опухоли p53), PTGIS (простагландин 12 (простациклин) синтаза), MB (миоглобин), IL4 (интерлейкин 4), ANGPT1 (ангиопоэтин 1), ABCG8 (ATP-связывающая кассета, подсемейство G (WHITE), представитель 8) или CTSK (катепсин K).

Примеры ассоциированых с болезнью Альцгеймера белков могут включать, например, белок липопротеинов очень низкой плотности (VLDLR), кодируемый геном VLDLR, фермент, активирующий убиквитин-подобный модификатор 1 (UBA1), кодируемый геном UBA1, или белок каталитической субъединицы NEDD8-активирующего фермента E1 (UBE1C), кодируемый геном UBA3.

Примеры белков, ассоциированных с расстройствами аутистического спектра, могут включать, например, белок 1, ассоциированный с бензодиазапиновым рецептором (периферическим) (BZRAP1), кодируемый геном BZRAP1, белок представителя 2 семейства AF4/FMR2 (AFF2), кодируемый геном AFF2 (также называемый MFR2), белок аутосомного гомолога 1 умственной отсталости, обусловленной ломкой X-хромосомой (FXR1), кодируемый геном FXR1, или белок аутосомного гомолога 2 умственной отсталости, обусловленной ломкой X-хромосомой (FXR2), кодируемый геном FXR2.

Примеры белков, ассоциированных с макулярной дегенерацией, могут включать, например, белок представителя 4 подсемейства A ATP-связывающей кассеты (ABC1) (ABCA4), кодируемый геном ABCR, белок аполипопротеина E (APOE), кодируемый геном APOE, или белок лиганда 2 хемокина (C-C мотив) (CCL2), кодируемый геном CCL2.

Примеры белков, ассоциированных с шизофренией, могут включать NRG1, ErbB4, CPLX1, TPH1, TPH2, NRXN1, GSK3A, BDNF, DISC1, GSK3B и их комбинации.

Примеры белков, вовлеченных в подавление опухоли, могут

включать, например, ATM (мутантный белок атаксии телеангиэктазии), ATR (белок атаксии телеангиэктазии и родственный Rad3), EGFR (эпидермальный фактор роста рецептор), ERBB2 (гомолог 2 вирусного онкогена эритробластного лейкоза v-erb-b2 2), ERBB3 (гомолог 3 вирусного онкогена эритробластного лейкоза v-erb-b2), ERBB4 (гомолог 4 вирусного онкогена эритробластного лейкоза v-erb-b2 4), Notch 1, Notch2, Notch 3 или Notch 4.

Примеры белков, ассоциированных с секретазным нарушением, могут включать, например, PSENEN (гомолог энхансера 2 пресенилина (*C. elegans*)), CTSB (катепсин B), PSEN1 (пресенилин 1), APP (белок-предшественник амилоида-бета (A4)), APH1B (гомолог B белка 1 дефекта переднего отдела глотки (*C. elegans*)), PSEN2 (пресенилин 2 (белок 4 болезни Альцгеймера)), или BACE1 (фермент 1 расщепления APP в бета-участке).

Примеры белков, ассоциированных с боковым амиотрофическим склерозом, могут включать SOD1 (супероксиддисмутаза 1), ALS2 (белок бокового амиотрофического склероза 2), FUS (белок, слитый при саркоме), TARDBP (белок, связывающий ДНК TAR), VAGFA (сосудисто-эндотелиальный фактор роста A), VAGFB (сосудисто-эндотелиальный фактор роста B) и VAGFC (сосудисто-эндотелиальный фактор роста C) и любую их комбинацию.

Примеры белков, ассоциированных с прионными заболеваниями, могут включать SOD1 (супероксиддисмутаза 1), ALS2 (белок бокового амиотрофического склероза 2), FUS (белок, слитый при саркоме), TARDBP (белок, связывающий ДНК TAR), VAGFA (сосудисто-эндотелиальный фактор роста A), VAGFB (сосудисто-эндотелиальный фактор роста B) и VAGFC (сосудисто-эндотелиальный фактор роста C) и любую их комбинацию.

Примеры белков, связанных с нейродегенеративными состояниями при прионных нарушениях, могут включать, например, A2M (альфа-2-макроглобулин), AATF (фактор транскрипции-антагонист апоптоза), ACPP (белок кислой фосфатазы предстательной железы), ACTA2 (белок актин альфа 2 гладкомышечных клеток аорты), ADAM22 (домен металлопептидазы ADAM), ADORA3 (рецептор аденозина A3) или ADRA1D (альфа-1D

адренергический рецептор или альфа-1D адренорецептор).

Примеры белков, ассоциированных с иммунодефицитом, могут включать, например, A2M [альфа-2-макроглобулин]; AANAT [арилалкиламин N-ацетилтрансфераза]; ABCA1 [представитель 1 подсемейства A АТР-связывающей кассеты (ABC1)]; ABCA2 [представитель 2 подсемейства A АТР-связывающей кассеты (ABC1)] или ABCA3 [представитель 3 подсемейства A АТР-связывающей кассеты (ABC1)].

Примеры белков, ассоциированных с обусловленными тринуклеотидными повторами нарушениями включают, например, AR (рецептор андрогенов), FMR1 (белок 1 умственной отсталости, связанной с ломкой X-хромосомой), HTT (хантингтин) или DMPK (протеинкиназа миотонической дистрофии), FXN (фратаксин), ATXN2 (атаксин 2).

Примеры белков, ассоциированных с нарушениями нейротрансмиссии, включают, например, SST (соматостатин), NOS1 (синтаза оксида азота 1 (нейрональная)), ADRA2A (адренергический рецептор альфа-2A), ADRA2C (адренергический рецептор альфа-2C), TACR1 (рецептор 1 тахикинина) или HTR2c (рецептор 2C 5-гидрокситриптамина (серотонина)).

Примеры ассоциированных с неврологическим развитием последовательностей включают, например, A2BP1 (атаксин 2-связывающий белок 1), AADAT (аминоадипатаминотрансфераза), AANAT (арилалкиламин N-ацетилтрансфераза), ABAT (4-аминобутиратаминотрансфераза), ABCA1 (представитель 1 подсемейства A АТР-связывающей кассеты (ABC1)) или ABCA13 (представитель 13 подсемейства A АТР-связывающей кассеты (ABC1)).

Следующие примеры предпочтительных состояний, поддающихся лечению с помощью системы по настоящему изобретению, могут быть выбраны из: синдрома Айкарди-Гутьерес; болезни Александера; Аллана-Херндона-Дадли; обусловленных POLG нарушений; альфа-маннозидоза (тип II и III); синдром Альстрема; синдрома Ангельмана; атаксии-телеангиэктазии; нейронного восковидного липофусциноза; бета-талассемии; билатеральной атрофии оптического нерва и (младенческой) атрофии зрительного нерва 1

типа; ретинобластомы (билатеральной); болезни Канавана; цереброокулофациоскелетного синдрома 1 (COFS1); церебросухожильного ксантоматоза; синдрома Корнелии де Ланге; обусловленных MAPT нарушений; генетических прионных заболеваний; синдрома Драве; семейной болезни Альцгеймера с ранним началом; атаксии Фридрейха [FRDA]; синдрома Фринса; фукозидоза; врожденной мышечной дистрофии Фукуяма; галактосиалидоза; болезни Гоше; органических ацидемий; гемофагоцитарного лимфогистиоцитоза; синдрома прогерии Хатчинсона-Гилфорда; муколипидоза II; младенческой болезни накопления свободной сиаловых кислот; PLA2G6-ассоциированной нейродегенерации; синдрома Джервелла и Ланге-Нильсена; пограничного буллезного эпидермолиза; болезни Хантингтона; болезни Краббе (младенческой); ассоциированного с митохондриальной ДНК синдрома Ли и NARP; синдрома Леш-Найхена; ассоциированной с LIS1 лиссэнцефалии; синдрома Лоу; болезни мочи с запахом кленового сиропа; синдрома дупликации MECP2; обусловленных ATP7A нарушений транспорта меди; обусловленной LAMA2 мышечной дистрофии; дефицита арилсульфатазы A; мукополисахаридоз типов I, II или III; нарушений биогенеза пероксисом, спектра синдромов Цельвегера; нейродегенерации с обусловленными накоплением железа в головном мозге нарушениями; дефицита кислой сфингомиелиназы; болезни Ниманна-Пика типа C; глициновой энцефалопатии; ARX-связанных нарушений; нарушений цикла мочевины; обусловленного COL1A 1/2 несовершенного остеогенеза; синдромов делеции митохондриальной ДНК; обусловленных PLP1 нарушений; синдрома Перри; синдрома Фелан-Макдермид; болезни накопления гликогена типа II (болезнь Помпе) (младенческая); обусловленных MAPT нарушения; обусловленных MECP2 нарушений; точечной эпифизарной хондродисплазии тазобедренного или плечевого сустава 1 типа; синдрома Робертса; болезни Сендхоффа; болезни Шиндлера 1 типа; дефицита аденозиндезаминазы; синдрома Смита-Лемли-Опица; спинальной мышечной атрофии, младенческой спинально-церебеллярной атаксии; дефицита гексозаминидазы A; летальной дисплазии 1 типа; обусловленных коллагеном типа VI нарушений; синдрома Ушера типа I; врожденной мышечной дистрофии; синдрома

Вольфа-Хиршхорна; дефицита лизосомальной кислой липазы и пигментной ксеродермы.

II. СПОСОБЫ ЛЕЧЕНИЯ НЕЙРОДЕГЕНЕРАТИВНЫХ ЗАБОЛЕВАНИЙ

Настоящее изобретение также относится к способам лечения заболевания, выбранного из группы, состоящей из дистрофии сетчатки, дистрофии роговицы, болезней экспансии микросателлитов, злокачественной опухоли и нейродегенеративных заболеваний у индивидуума, нуждающегося в этом, причем способ включает: (a) предоставление не встречающейся в природе нуклеазной системы (например, CRISPR-ассоциированная (Cas) 9 (системы CRISPR-Cas9, не Cas9 CRISPR, система CRISPR-Cpf-1 и т.п.)), содержащей один или несколько векторов, содержащих: a) промотор (например, ортологичный промотор H1 или 7sk), функционально связанн по меньшей мере с одной нуклеотидной последовательностью, кодирующей гидовую РНК (гРНК) нуклеазной системы, где гРНК гибридизуется с последовательностью ДНК, являющейся мишенью, в клетке, и где молекула ДНК кодирует один или несколько продуктов генов, экспрессируемых в клетке; и b) регуляторный элемент, функциональный в клетке, функционально связанный с нуклеотидной последовательностью, кодирующей нацеленную на геном нуклеазу (например, модифицированный белок Cas9), где компоненты (a) и (b) расположены на одном и том же или на различных векторах системы, где гРНК осуществляет нацеливание на последовательность-мишень и гибридизуется с ней, и нуклеаза расщепляет молекулу ДНК, изменяя экспрессию одного или нескольких продуктов; и (b) введение индивидууму эффективного количества системы. В некоторых вариантах осуществления система упакована в одну частицу аденоассоциированного вируса (AAV). В некоторых вариантах осуществления система инактивирует один или несколько продуктов генов. В некоторых вариантах осуществления система вырезает по меньшей мере одну мутацию гена. В некоторых вариантах осуществления промотор H1 содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 12 или 32. В

некоторых вариантах осуществления ортологичный промотор H1 происходит из мыши или крысы. В некоторых вариантах осуществления ортологичный промотор H1 содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 84-119. В некоторых вариантах осуществления ортологичный промотор H1 содержит нуклеотидные последовательности из группы, состоящей из SEQ ID NO: 84-119. В некоторых вариантах осуществления промотор H1 содержит: a) элементы контроля, которые обеспечивают транскрипцию в одном направлении по меньшей мере одной нуклеотидной последовательности, кодирующей гРНК; и b) элементы контроля, которые обеспечивают транскрипцию в противоположном направлении нуклеотидной последовательности, кодирующей нацеленную на геном нуклеазу. В некоторых вариантах осуществления промотор представляет собой промотор 7sk человека, мыши или крысы. В некоторых вариантах осуществления 7sk выбран из группы, состоящей из 7sk1 (вариант 1 7sk, в котором используется эндогенная 5'UTR GSTA4), 7sk2 (вариант 2 7sk, в котором используется только последовательность Козак в 5'UTR) и 7sk3 (вариант 7sk, в котором используется 5'UTR бета-глобина). В некоторых вариантах осуществления 7sk1 человека содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 3. В некоторых вариантах осуществления 7sk1 мыши содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с SEQ ID NO: 6. В некоторых вариантах осуществления 7sk2 человека содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с SEQ ID NO: 4. В некоторых вариантах осуществления 7sk2 мыши содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с SEQ ID NO: 7. В некоторых вариантах осуществления 7sk3 человека содержит нуклеотидную

последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с SEQ ID NO: 5. В некоторых вариантах осуществления 7sk3 мыши содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с SEQ ID NO: 8. В некоторых вариантах осуществления ортологичный промотор H1 или промотор 7sk имеет по меньшей мере одну модификацию. В некоторых вариантах осуществления по меньшей мере одна модификация промотора содержит элемент, который позволяет зависимую от условий регуляцию. В некоторых вариантах осуществления элемент представляет собой отвечающий на tet промотор. В некоторых вариантах осуществления отвечающий на tet промотор содержит Tet-репрессор (TetR) и Tet-оператор (TetO) встроенный в промотор H1. В некоторых вариантах осуществления TetR содержит аминокислотную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с SEQ ID NO: 14. В некоторых вариантах осуществления TetO содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 11. В некоторых вариантах осуществления H1-TetO содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 13. В некоторых вариантах осуществления по меньшей мере одна модификация промотора включает участок, который позволяет ауторегуляцию. В некоторых вариантах осуществления участок ауторегуляции содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 33 или 34. В некоторых вариантах осуществления нуклеазная система дополнительно содержит по меньшей мере одну последовательность терминатора. В некоторых вариантах осуществления последовательность терминатора представляет собой любую последовательность нуклеиновой кислоты, выбранную из группы, состоящей из SEQ ID NO: 120-130. В некоторых вариантах

осуществления по меньшей мере одна последовательность терминатора выбрана из группы, состоящей из последовательности SV40 или синтетических поли A (SPA) последовательностей. В некоторых вариантах осуществления последовательность терминатора представляет собой последовательность SV40 из 120 (SEQ ID NO: 123) или 240 (SEQ ID NO: 122) пар оснований. В некоторых вариантах осуществления последовательность терминатора представляет собой последовательность SPA из 49 пар оснований. В некоторых вариантах осуществления нуклеазная система дополнительно содержит последовательности 5′-нетранслируемой области (5′UTR). В некоторых вариантах осуществления нуклеазная система дополнительно содержит последовательность Козак. В некоторых вариантах осуществления последовательность Козак содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с SEQ ID NO: 1 или 2. В некоторых вариантах осуществления нуклеазная система дополнительно содержит последовательность РНК, которая опосредует кэп-независимую инициацию трансляции. В некоторых вариантах осуществления последовательность РНК выбрана из группы, состоящей из 6.947 или 6.967 (Wellensiek I. (2013) *Nature Methods*, 10:747-750). В некоторых вариантах осуществления Cas9 содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 61. В некоторых вариантах осуществления Cas9 содержит по меньшей мере одну модификацию. В некоторых вариантах осуществления по меньшей мере одна модификация в Cas9 включает изменение расщепляющей последовательности. В некоторых вариантах осуществления нуклеаза выбрана из группы, состоящей из Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (также известной как Csn1 и Csx12), Cas10, Cas13a, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, Cpf1, C2c1, C2c2 и C2c3. В некоторых вариантах осуществления изменение в расщепляющей

последовательности выбрано из группы, состоящей из T2A, P2A, E2A и F2A. В некоторых вариантах осуществления T2A содержит аминокислотную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 36. В некоторых вариантах осуществления P2A содержит аминокислотную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 35. В некоторых вариантах осуществления P2A содержит аминокислотную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 35. В некоторых вариантах осуществления E2A содержит аминокислотную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 37. В некоторых вариантах осуществления F2A содержит аминокислотную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 38. В некоторых вариантах осуществления по меньшей мере одна модификация в Cas9 является кодон-оптимизированной для экспрессии в клетке. В некоторых вариантах осуществления Cas9 дополнительно содержит линкерную последовательность, функционально слитую в рамке считывания зависимым от клеточного цикла белком (слитая конструкция Cas9). В некоторых вариантах осуществления зависимый от клеточного цикла белок выбран из группы, состоящей из APE2, ATR, 53BP1, BRCA1, Chk1, Cdc5, Cdc6, Cdc7, Cdc45, Cdt1, CSA, CSB, Ctf18, CtIP, DDB1, DDB2, DNA2, DUT, Elg1, EndoV, Esp1, экзонуклеазы 1, FBH1, FEN1, геминина, Hus1, KNTC2 (NDC80), Ku80, лигазы 1, Mad2, MBD4, Mcm3, Mcm4, Mcm5, Mcm6, Mcm7, Mcm8, Mcm10, MGMT, MLH3, Mms4, MPG, MSH2, Mus81, NBS1, NEIL2, NEIL3, NTH1, Orc1, Orc3, PARP1, PCNA, Pif1, Pin1, PMS1, PMS2, PNK, Pola p180, Pola p70, Pola Spp1 (Prim2a), Polb, Pold p125, Pole Dpb3, Pole Dpb4, Pole Pol2, Poli, Poll, Polm, Psf1, Psf2, Psf3, Rad1, Rad18, Rad23A, Rad23B, Rad51, Rad51D, Rad54, Rad6A, Rag1, Rag2, RPA34, RPA70, Scc1, Scc3, Sir2, SIRT1 (сиртуин), TDG, TDP1,

TIMELESS, Tin2, топоизомеразы I, топоизомеразы IIIa, топоизомеразы IIIb, Ubc13, UNG, XAB2, XPC, XPF, XPG, Xrcc2 и XRCC4. В некоторых вариантах осуществления зависимый от клеточного цикла белок представляет собой геминин. В некоторых вариантах осуществления зависимый от клеточного цикла белок представляет собой геминин человека. В некоторых вариантах осуществления геминин человека содержит аминокислоты из положений 1-110 (hGem(1-110)). В некоторых вариантах осуществления hGem(1-110) содержит аминокислотную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с аминокислотной последовательностью, указанной в SEQ ID NO: 19. В некоторых вариантах осуществления зависимый от клеточного цикла белок представляет собой Cdt1. В некоторых вариантах осуществления зависимый от клеточного цикла белок представляет собой Cdt1 человека. В некоторых вариантах осуществления Cdt1 человека содержит аминокислоты из положений 30-120 (hCdt1(30-120)). В некоторых вариантах осуществления hCdt1(30-120) содержит аминокислотную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с аминокислотной последовательностью, указанной в SEQ ID NO: 18. В некоторых вариантах осуществления слитая конструкция Cas9 содержит аминокислотную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с аминокислотной последовательностью, выбранной из группы, состоящей из SEQ ID NO: 20-27. В некоторых вариантах осуществления Cas9 функционально слита в рамке считывания с убиквитином (Ub-Cas9). В некоторых вариантах осуществления Ub-Cas9 содержит по меньшей мере одну N-концевую модификацию. В некоторых вариантах осуществления Ub-Cas9 с N-концевой модификацией содержит аминокислотную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с аминокислотной последовательностью, выбранной из группы, состоящей из SEQ ID NO: 39-58. В некоторых вариантах осуществления белок убиквитин содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%,

95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 60. В некоторых вариантах осуществления нуклеазная система дополнительно содержит никазу SaCas9. В некоторых вариантах осуществления нуклеазная система дополнительно содержит донорную матричную последовательность. В некоторых вариантах осуществления по меньшей мере один вектор содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 62 или 67. В некоторых вариантах осуществления донорная матричная последовательность корректирует по меньшей мере одну мутацию гена. В некоторых вариантах осуществления по меньшей мере одна мутация гена представляет собой *rd10* или *rd12*. В некоторых вариантах осуществления по меньшей мере один вектор содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 63 или 64. В некоторых вариантах осуществления по меньшей мере один вектор содержит нуклеотидную последовательность, обладающую по меньшей мере 80%, 85%, 90%, 95%, 98%, 99% или 100% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 65 или 66. В некоторых вариантах осуществления промотор функционально связан по меньшей мере с одной, двумя, тремя, четырьмя, пятью, шестью, семью, восемью, девятью или десятью гРНК. В некоторых вариантах осуществления последовательность-мишень содержит нуклеотидную последовательность $AN_{19}NGG$, $GN_{19}NGG$, $CN_{19}NGG$ или $TN_{19}NGG$. В некоторых вариантах осуществления клетка представляет собой эукариотическую или неэукариотическую клетку. В некоторых вариантах осуществления эукариотическая клетка представляет собой клетку млекопитающего или человека. В некоторых вариантах осуществления эукариотическая клетка представляет собой фоторецепторную клетку сетчатки. В некоторых вариантах осуществления один или несколько продуктов генов представляют собой родопсин. В некоторых вариантах осуществления экспрессия одного или нескольких продуктов генов снижается. В некоторых

вариантах осуществления клетка представляет собой ганглионарную клетку сетчатки. В некоторых вариантах осуществления эукариотическая клетка представляет собой злокачественную клетку. В некоторых вариантах осуществления пролиферация в злокачественной клетке ингибируется или снижается. В некоторых вариантах осуществления в злокачественной клетке усиливается или повышается апоптоз. В некоторых вариантах осуществления заболевание представляет собой дистрофию роговицы. В некоторых вариантах осуществления заболевание представляет собой дистрофию сетчатки. В некоторых вариантах осуществления дистрофия сетчатки выбрана из группы, состоящей из врожденного амавроза Лебера (LCA), пигментного ретинита (RP) и глаукомы. В некоторых вариантах осуществления дистрофия роговицы выбрана из группы, состоящей из дистрофии эпителия базальной мембраны роговицы, эпителиальной рецидивирующей эрозивной дистрофии, субэпителиальной муцинозной дистрофии роговицы, дистрофии роговицы Месманна, эпителиальной дистрофии роговицы Лиша, желатинозной каплевидной дистрофии роговицы, дистрофии роговицы Рейса-Бюклерса, дистрофии роговицы Тиля-Бенке, решетчатой дистрофии роговицы 1 типа (классическая), решетчатой дистрофии роговицы 2 типа, решетчатой дистрофии роговицы типа III, решетчатой дистрофии роговицы типа IIIA, решетчатой дистрофии роговицы типа I/IIIA, решетчатой дистрофии роговицы типа IV, полиморфного амилоидоза (роговицы), гранулярной дистрофии роговицы 1 типа, гранулярной дистрофии роговицы 2 типа, макулярной дистрофии роговицы, дистрофии роговицы Шнайдера, врожденной стромальной дистрофии роговицы, дистрофии роговицы Флека, задней аморфной дистрофии роговицы, центральной облаковидной дистрофии Франсуа, предесцеметовой дистрофии роговицы, эндотелиальной дистрофии роговицы Фукса, задней полиморфной дистрофии роговицы, врожденной наследственной эндотелиальной дистрофии и Х-сцепленной эндотелиальной дистрофии роговицы. В некоторых вариантах осуществления болезни экспансии микросателлитов выбраны из группы, состоящей из блефарофимоза, птоза и эпикантуса-инверсной синдактилии, ключично-черепной дисплазии, центрального врожденного гиповентиляционного

синдрома, синдрома DM (миотоническая дистрофия) Хаддада, FRAXA (синдром ломкой Х-хромосомы), FRAXE (задержка умственного развития, обусловленная ломкой XE), FRDA (атаксия Фридрейха), эндотелиальной дистрофии роговицы Фукса, FXTAS (ассоциированный с ломкой Х синдром тремора/атаксии), ладонно-подошвенно-генитального синдрома, HD (болезнь Гентингтона), голопрозэнцефалии, задержки умственного развития с дефицитом гормонов роста, задержки умственного развития, эпилепсии, синдрома веста, синдрома Партингтона, окулофарингеальной мышечной дистрофии, SBMA (спинальная и бульбарная мышечная атрофия), SCA1 (спинально-церебеллярная атаксия 1 типа), SCA12 (спинально-церебеллярная атаксия 12 типа), SCA17 (спинально-церебеллярная атаксия 17 типа), SCA2 (спинально-церебеллярная атаксия 2 типа), SCA3 (спинально-церебеллярная атаксия 3 типа или болезнь Мачадо-Джозефа), SCA6 (спинально-церебеллярная атаксия 6 типа), SCA7 (спинально-церебеллярная атаксия 7 типа), SCA8 (спинально-церебеллярная атаксия 8 типа) и синполидактилии. В некоторых вариантах осуществления введение индивидууму происходит посредством имплантации, инъекции или вирусным путем. В некоторых вариантах осуществления введение индивидууму осуществляется посредством субретинальной инъекции. В некоторых вариантах осуществления введение индивидууму осуществляется местным путем, внутрисосудистым путем, внутрикожным путем, трансдермально, парентерально, внутривенно, внутримышечно, интраназально, подкожно, регионарным путем, подкожно, интратрахеальным путем, внутрибрюшинно, интраартериально, внутрь мочевого пузыря, внутрь опухоли, перитуморальным путем, ингаляционным путем, системно, путем перфузии, посредством промывания, непосредственно посредством инъекции или пероральным введением состава. В некоторых вариантах осуществления введение индивидууму осуществляется местным путем на поверхность глаза. В некоторых вариантах осуществления введение индивидууму осуществляется на и снаружи роговицы, склеры, во внутриглазное, субконъюнктивальное, субтеноново или ретробульбарное пространство, или в или вокруг век. В некоторых вариантах осуществления индивидуума лечат по меньшей мере одним

дополнительным средством против злокачественной опухоли. В некоторых вариантах осуществления средство против злокачественной опухоли выбрано из группы, состоящей из паклитаксела, цисплатина, топотекана, гемцитабина, блеомицина, этопозида, карбоплатина, доцетаксела, доксорубицина, топотекана, циклофосфамида, трабектедина, олапариба, тамоксифена, лектрозола и бевацизумаба. В некоторых вариантах осуществления индивидуума лечат по меньшей мере одной дополнительной терапией против злокачественной опухоли. В некоторых вариантах осуществления терапия против злокачественной опухоли представляет собой лучевую терапию, химиотерапию или хирургическую операцию. В некоторых вариантах осуществления злокачественная опухоль представляет собой солидную опухоль. В некоторых вариантах осуществления злокачественная опухоль выбрана из группы, состоящей из злокачественной опухоли головного мозга, желудочно-кишечной злокачественной опухоли, рака полости рта, рака молочной железы, рака яичника, рака предстательной железы, рака поджелудочной железы, рака легкого, рака печени, рака горла, рака желудка и рака почки. В некоторых вариантах осуществления злокачественная опухоль представляет собой злокачественную опухоль головного мозга. В некоторых вариантах осуществления системное введение выбрано из группы, состоящей из перорального, внутривенного, внутрикожного, внутрибрюшинного, подкожного и внутримышечного введения. В некоторых вариантах осуществления индивидуумом является млекопитающее. В некоторых вариантах осуществления млекопитающим является человек.

В некоторых вариантах осуществления настоящее изобретение относится к способу лечения нейродегенеративного заболевания глаз у индивидуума, нуждающегося в этом, причем способ включает: (a) предоставление не встречающейся в природе системы CRISPR-Cas, содержащей один или несколько векторов, содержащих: i) промотор H1, функционально связанный по меньшей мере с одной нуклеотидной последовательностью, кодирующей гидовую РНК (гРНК) системы CRISPR-Cas, где гРНК гибридизуется с последовательностью ДНК, являющейся мишенью, в клетке субъекта, и где молекула ДНК кодирует один или несколько продуктов генов, экспрессируемых в

клетке; и ii) регуляторный элемент, функциональный в клетке, функционально связанный с нуклеотидной последовательностью, кодирующей белок Cas9, где компоненты (i) и (ii) расположены на одном и том же или на различных векторах системы, где гРНК осуществляет нацеливание на последовательность-мишень и гибридизуется с ней и белок Cas9 расщепляет молекулу ДНК, изменяя экспрессию одного или нескольких продуктов генов; и (b) введение индивидууму эффективного количества системы.

Под "нейродегенеративным заболеванием, нарушением или состоянием" подразумевают заболевание, нарушение или состояние (включая невропатию), ассоциированное с дегенерацией или дисфункцией нейронов или других нервных клеток, таких как фоторецепторные клетки сетчатки. Нейродегенеративное заболевание, нарушение или состояние может представлять собой любое заболевание, нарушение или состояние, при котором может происходить снижение функции или дисфункция нейронов, или потеря нейронов или других нервных клеток.

Такие заболевания, нарушения или состояния включают, но не ограничиваются ими, глаукому и нейродегенеративные заболевания, нарушения или состояния нервной системы, такие как боковой амиотрофический склероз (ALS), невралгию тройничного нерва, невралгию языкоглоточного нерва, паралич Белла, миастению, мышечную дистрофию, прогрессирующую мышечную атрофию, первичный латеральный склероз (PLS), псевдобульбарный паралич, прогрессирующий бульбарный паралич, спинальную мышечную атрофию, наследственную мышечную атрофию, синдромы мезпозвоночных дисков, спондилез шейного отдела, нарушения сплетений, компрессионный синдром верхней апертуры грудной клетки, периферические нейропатии, порфирию, болезнь Альцгеймера, болезнь Гентингтона, болезнь Паркинсона, заболевания типа паркинсонизм-плюс, полисистемную атрофию, прогрессирующий надъядерный паралич, кортикобазальную дегенерацию, деменция с тельцами Леви, лобно-височную деменцию, демиелинизирующие заболевания, синдром Гийена-Барре, рассеянный склероз, болезнь Шарко-Мари-Тута, прионные заболевания, болезнь Крейтцфельда-Якоба, синдром Герстманна-Штреусслера-Шейнкера (GSS), спорадическую фатальную

бессонницу (FFI), губчатую энцефалопатию крупного рогатого скота (BSE), болезнь Пика, эпилепсию и СПИД-дементный комплекс или заболевания, ассоциированные с ними.

Другие нейродегенеративные заболевания, нарушения или состояния нервной системы включают алкоголизм, болезнь Александера, болезнь Альперса, атаксию-телеангиэктазию, болезнь Баттена (также известную как болезнь Шпильмайера-Фогта-Шегрена-Баттена), болезнь Канавана, синдром Кокейна, диабетическую невропатию, лобно-височную дегенерацию, ВИЧ-ассоциированную деменцию, болезнь Кеннеди, болезнь Краббе, нефроборрелиоз, болезнь Мачадо-Джозефа (спинально-церебеллярная атаксия 3 типа), влажную или сухую макулярную дегенерацию, болезнь Ниманна-Пика, болезнь Пелицеуса-Мерцбахера, дегенеративные заболевания фоторецепторов, такие как пигментный ретинит и ассоциированные с ним заболевания, болезнь Рефсума, болезнь Сандгоффа, болезнь Шильдера, подострую комбинированную дегенерацию спинного мозга на фоне пернициозной анемии, болезнь Шпильмайера-Фогта-Шегрена-Баттена (также известную как болезнь Баттена), спинально-церебеллярную атаксию (множество типов с различными характеристиками), болезнь Стила-Ричардсона-Ольшевского заболевание и сухотку спинного мозга, или ассоциированы с ними.

Примеры окулярной нейродегенерации включают, но не ограничиваются ими, глаукому, решетчатую дистрофию, пигментный ретинит, связанную со старением дегенерацию желтого пятна (AMD), дегенерацию фоторецепторов, ассоциированную с влажной или сухой AMD, другую дегенерацию сетчатки, такую как пигментный ретинит (RP), друзы оптического нерва, оптическую невропатию и оптический неврит, такой как оптический неврит при рассеянном склерозе. В некоторых вариантах осуществления нейродегенеративное заболевание глаз выбрано из группы, состоящей из глаукомы, дегенерации сетчатки и связанной со старением дегенерации желтого пятна. В некоторых вариантах осуществления нейродегенеративное заболевание глаз представляет собой пигментный ретинит (RP).

Неограничивающие примеры различных типов глаукомы, которые можно предупреждать или лечить в соответствии с описанным в

настоящем описании изобретении включают первичную глаукому (также известную как открытоугольная глаукома, хроническая открытоугольная глаукома, хроническая простая глаукома и *glaucoma simplex*), глаукому низкого давления, первичную закрытоугольную глаукому (также известную как первичная закрытоугольная глаукома, узкоугольная глаукома, глаукома со зрачковым блоком и острая застойная глаукома), острую закрытоугольную глаукому, хроническую закрытоугольную глаукому, интермиттирующую закрытоугольную глаукому, хроническую открытоугольную глаукому, пигментарную глаукому, эксфолиативную глаукому (также известную как псевдоэксфолиативная глаукома или капсулярная глаукома), врожденную глаукому (например, первичная врожденная глаукома и инфантильная глаукома), вторичную глаукому (например, воспалительную глаукому (например, увеит и гетерохромный иридоциклит Фукса)), факогенную глаукому (например, закрытоугольная глаукома со зрелой катарактой, факоанафилактическая глаукома на фоне разрыва капсулы хрусталика, факолитическую глаукому вследствие факотоксической блокады сети, и подвывих хрусталика), глаукому на фоне внутриглазного кровоизлияния (например, гифема и гемолитическая глаукома, также известная как эритрокластическая глаукома), травматическую глаукому (например, глаукома со смещением угла, травматическое смещение угла передней камеры, послеаперационная глаукома, глаукома с афакическим зрачковым блоком и с блоком ресничного тела), неоваскулярную глаукому, индуцируемую лекарственными средствами глаукому (например, индуцируемую кортикостероидами глаукому и глаукома при использовании альфа-химотрипсина), токсическую глаукому и глаукому, ассоциированную с внутриглазными опухолями, отслоением сетчастки, тяжелыми химическими ожогами глаза и атрофией радужной оболочки. В определенных вариантах осуществления нейродегенеративное заболевание, нарушение или состояние представляет собой заболевание, нарушение или состояние, которые не ассоциированы с избыточным ангиогенезом, например, глаукому, которая не является неоваскулярной глаукомой.

Как используют в рамках изобретения, термин "нарушение",

главным образом, относится к любому состоянию, при котором является полезным лечение соединением против одной из идентифицированных мишеней или каскадов. включая любое заболевание, нарушение или состояние, которое можно лечить эффективным количеством соединения против одной из идентифицированных мишеней, или каскадов, или его фармацевтически приемлемой соли.

Как используют в рамках изобретения, термин "лечение" может включать обращение вспять, смягчение, ингибирование прогрессирования, предупреждение или уменьшение вероятности заболевания, нарушения или состояния, в отношении которых такой термин используется, или одного или нескольких симптомов или проявлений такого заболевания, нарушения или состояния (например, заболевания или нарушения, которое вызывает дисфункцию и/или гибель фоторецепторных клеток сетчатки). В некоторых вариантах осуществления лечение уменьшает дисфункцию и/или гибель фоторецепторных клеток сетчатки. Например, лечение может уменьшать дисфункцию и/или гибель фоторецепторных клеток сетчатки по меньшей мере на 5%, 10%, 15%, 20%, 25%, 30%, 33%, 35%, 40%, 45%, 50%, 55%, 60%, 66%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или более по сравнению с дисфункцией и/или гибелью фоторецепторных клеток сетчатки у индивидуума до лечения или у индивидуума, которому не проводят лечение. В некоторых вариантах осуществления лечение полностью ингибирует дисфункцию и/или гибель фоторецепторных клеток у индивидуума. Как используют в рамках изобретения, "фоторецепторная клетка сетчатки" представляет собой специализированный тип нейронов, встречающийся в сетчатке, который способен к фотопреобразованию. В некоторых вариантах осуществления по меньшей мере один продукт гена представляет собой родопсин.

В некоторых вариантах осуществления систему упаковывают в одну частицу аденоассоциированного вируса (AAV) перед введением индивидууму. В некоторых вариантах осуществления введение индивидууму происходит посредством субретинальной инъекции. Лечение, введение или терапия могут быть непрерывными или

прерывающимися. Непрерывное лечение, введение или терапия относятся к лечению по меньшей мере каждый день без прерывания лечения на одни или несколько суток. Прерывающееся лечение или введение, или лечение или введение прерывающимся образом, относится к лечению, которое не является непрерывным, а скорее является циклическим по своей природе. Лечение описанными в настоящем описании способами может приводить к полному смягчению или излечению от заболевания, нарушения или состоянию, или частичном облегчению одного или нескольких симптомов заболевания, нарушения или состояние, и оно может быть временным или постоянным. Термин "лечение" также охватывает профилактику, терапию и излечение.

Термин "эффективное количество" или "терапевтически эффективное количество" относится к количеству средства, которое является достаточным для обеспечения благоприятных или желательных результатов. Терапевтически эффективное количество может варьироваться в зависимости от одного или нескольких из индивидуума и заболевания, подвергаемого лечению, массы тела и возраста индивидуума, тяжести заболевания. пути введения и т.п., которые могут быть без труда определены специалистом в данной области. Термин также относится к дозе, которая обеспечит изображение для обнаружения любым из способов визуализации, описанных в настоящем описании. Конкретная доза может варьироваться в зависимости от одного или нескольких из конкретного выбранного средства, режима дозирования, которому следуют, введения в комбинации с другими соединениями, времени введения, подвергаемой визуализации ткани и физической системы доставки, в которой она содержится.

Термины "индивидуум" и "пациент" используются в настоящем описании взаимозаменяемо. Желательно, индивидуум, которого лечат описанными в настоящем описании способами в их множестве вариантов осуществления представляет собой человека, хотя следует понимать, что способы, описанные в настоящем описании, являются эффективными в отношении всех видов позвоночных, которые охватываются термином "индивидуум". Таким образом, "индивидуум" может включать человека для медицинских целей,

например, для лечения существующего состояния или заболевания или профилактического лечения для предупреждения возникновения состояния или заболевания, или животное для медицинских, ветеринарных целей или для экспериментальных целей. Подходящие животные включают млекопитающих, включая, но не ограничиваясь ими, приматов, например, человека, низших обезьян, человекообразных обезьян и т.п.; животных семейства бычьих, например, крупный рогатый скот, буйволы и т.п.; животные семейства овечьих, например, овцы и т.п.; животные семейства козьих, например, козы и т.п.; животные семейства свиных, например, свиньи, кабаны и т.п.; животные семейства лошадиных, например, лошади, ослы, зебры и т.п.; животные семейства кошачьих, включая диких и домашних кошек; животные семейства собачьих, включая собак; зайцеобразные, включая кроликов, зайцев и т.п.; и грызунов, включая мышей, крыс и т.п. Животное может представлять собой трансгенное животное. В некоторых вариантах осуществления индивидуум представляет собой человека, включая, но не ограничиваясь ими, плод, новорожденного, младенца, подростка и взрослого индивидуума. Кроме того, "индивидуум" может включать пациента, страдающего или предположительно страдающего состоянием или заболеванием.

IV. ОБЩИЕ ОПРЕДЕЛЕНИЯ

Хотя в настоящем описании используются конкретные термины, их используют только в обобщающем и описательном значении, а не для целей ограничения. Если не определено иначе, все технические и научные термины, используемые в настоящем описании, обладают тем же значением, которое обычно подразумевает специалист в области, к которой относится настоящее изобретение.

Согласно устоявшейся традиции патентного права форма единственного числа относятся к "одному или нескольким", когда они используются в настоящем описании и в том числе в формуле изобретения. Таким образом, например, указание на "индивидуума" включает множество индивидуумов, если контекст явно не указывает на обратное (например, множество индивидуумов), и т.д.

На протяжении настоящего описания и формулы изобретения термины "содержать", "содержит" и "содержащий" используются в

неисключающем значении, за исключением случаев, когда контекстом требуется иное. Аналогично, термин "включать" и его грамматические варианты является неограничивающим, так что перечисление положений в перечне не исключает других подобных положений, которые могут быть заменены или добавлены к перечисленным положениям.

Для целей настоящего описания и прилагаемой формулы изобретения, если нет иных указаний, все числа, выражающие величины, размеры, измерения, соотношения, формы, составы, параметры, проценты, параметры, количества, характеристики и другие числовые величины, используемые в описании и формуле изобретения, следует понимать как модифицированные во всех случаях термином "приблизительно", даже несмотря на то, что термин "приблизительно" может не присутствовать явно с величиной, количеством или диапазоном. Таким образом, если нет иных указаний, числовые параметры, указанные в приведенном ниже описании и прилагаемой формуле изобретения, не являются и не должны быть точными, а могут быть приблизительными и/или при желании большими или меньшими, отражая допуски, факторы преобразования, округление, погрешности измерения и т.п., и другие факторы, известные специалистам в данной области в зависимости от желаемых свойств, которых намереваются достигнуть для настоящего изобретения. Например, термин "приблизительно" при указании на величину может охватывать колебания в некоторых вариантах осуществления ± 100% в некоторых вариантах осуществления ± 50%, в некоторых вариантах осуществления ± 20%, в некоторых вариантах осуществления ± 10%, в некоторых вариантах осуществления ± 5%, в некоторых вариантах осуществления ±1%, в некоторых вариантах осуществления ± 0,5%, и в некоторых вариантах осуществления ± 0,1% от указанного количества, поскольку такие колебания являются пригодными для проведения описанных способов или использования описанных композиций.

Кроме того, следует понимать, что термин "приблизительно", когда его используют применительно к одному или нескольким числам и числовым диапазонам, относится ко всем таким числам в

диапазоне и модифицирует этот расширением границ выше и ниже указанных числовых величин. Перечисление числовых величин посредством конечных точек включает все числа, например, целые числа, включая их части, относящиеся к этому диапазону (например, перечисление от 1 до 5 включает 1, 2, 3, 4 и 5, а также их дробные части, например, 1,5, 2,25, 3,75, 4,1 и т.п.) и любой диапазон в этом диапазоне.

ПРИМЕРЫ

Приведенные ниже примеры включены для предоставления специалисту в данной области руководства по применению на практике репрезентативных вариантов осуществления настоящего изобретения. Ввиду настоящего изобретения и общего уровня техники, специалисту в данной области может быть понятно, что приведенные ниже примеры являются только иллюстративными и что многочисленные изменения, модификации и перестановки можно использовать без отклонения от объема настоящего изобретения. Описание синтеза и конкретные примеры, которые следуют далее, предназначены только для иллюстрации, а не для ограничения каким-либо образом получения соединений по настоящему изобретению другими способами.

ПРИМЕР 1

Уровень техники

Развитие технологии CRISPR-Cas9 произвело революцию в области редактирования генов и обеспечило в корне новый подход к лечению генетических заболеваний. Система CRISPR-Cas9 состоит из гидовой РНК (гРНК), которая осуществляет нацеливание нуклеазы Cas9 специфическим для последовательности образом. Расщепление посредством системы CRISPR требует комплементарного спаривания оснований гРНК с последовательностью ДНК и соответствующего примыкающего к протоспейсеру мотива (PAM) - короткого нуклеотидного мотива, находящегося с 3'-стороны от области-мишени (Dalkara, D. *et al. Science translational medicine* 5, 189ra176 (2013); Berns, KI *et al*. Fundamental Virology (ed B.N. Fields, и Knipe, D.M.) 545-562 (Raven Press, 1986).). В настоящее время имеющим наименьшие ограничения и наиболее часто используемым белком Cas9 является белок из *S. pyogenes*, который

распознает последовательность NGG, и, таким образом, последовательность, на которую нацеливается CRISPR, представляет собой N20NGG. Хотя в многочисленных исследованиях было показано, что может быть эффективно осуществлено нацеливание на обуславливающие заболевание мутации *in vitro*, разработке терапевтических средств на основе CRISPR-Cas9 для применения *in vivo* препятствовали проблемы безопасности и ограничения доставки.

В то время как было показано, что нацеливание CRISPR на обуславливающие заболевание мутации является эффективным в многочисленных условиях *in vitro*, а также *in vivo* посредством исследований на мышах и других животных, все современные подходы все еще далеки от клинического применения в основном в результате ограничений доставки. Векторы AAV являются наиболее часто и успешно используемыми вирусными векторами для инъекции в глаз при генной терапии (Swiech, L. *et al. Nature biotechnology* 33, 102-106 (2015); Jinek, M. *et al. Science* 337, 816-821 (2012); Cong, L. *et al. Science* 339, 819-823 (2013); Mali, P. *et al. Science* 339, 823-826 (2013)). Несколько признаков делают AAV наиболее привлекательным для выбора: вирус является непатогенным, он инфицирует как делящиеся, так и неделящиеся клетки, экспрессия может продолжаться в течение длительных периодов времени и, что особенно следует отметить, в результате его истории в отношении безопасности, эффективности и общего отсутствия токсичности в клинических испытаниях. Кроме того, конкретные серотипы AAV являются эффективными для нацеливания на фоторецепторные клетки после субретинальной инъекции. В то время как векторы AAV обеспечивают безопасное средство для доставки терапевтических компонентов CRISPR, существует только одно значительное техническое препятствие, которое ограничивает их применимость - их размер. Геномы AAV дикого типа имеют длину ~4,7 т.п.н., а рекомбинантные вирусы могут упаковывать вплоть до ~5,2 т.п.н. (Mancuso, K. *et al. Nature* 461, 784-787 (2009); Beltran, WA *et al. Proceedings of the National Academy of Sciences of the United States of America* 109, 2132-2137 (2012)). Эта способность к упаковыванию определяет верхний предел ДНК,

которая может использоваться для одного вирусного вектора.

ДНК, требуемая для экспрессии Cas9 и гРНК общепринятыми способами превышает 5,2 т.п.н.: промотор промотор Pol II (~0,5 т.п.н.), SpCas9 (~4,1 т.п.н.), терминатор Pol II (~0,2 т.п.н.), промотор U6 (~0,3 т.п.н.) и гРНК (~0,1 т.п.н.). Одним подходом к проблеме доставки AAV является двухвекторный подход: один вектор AAV для доставки Cas9, а другой вектор AAV для доставки гРНК (Petrs-Silva, H. *et al. Molecular therapy: the journal of the American Society of Gene Therapy* 19, 293-301 (2011)). Однако в двойном подходе AAV используется небольшой промотор Mecp2 мыши, гена, который экспрессируется в клетках сетчатки, ключевым исключением из которых являются палочки (Song, C. *et al. Epigenetics & chromatin* 7, 17 (2014); Jain, D. *et al. Pediatric neurology* 43, 35-40 (2010)), что указывает на то, что, помимо потенциальной токсичности вследствие увеличенной вирусной нагрузки, подход совместной доставки по-видимому не может осуществлять нацеливание на значительное большинство мутаций LCA априори. В то время как это является потенциально эффективным подходом для другого опосредуемого генной терапией геномного редактирования, в рамках настоящего изобретения предусматривается подход с одним вектором для редактирования генов сетчатки, который увеличивает эффективность, осуществляет специфическое нацеливание на фоторецепторы и снижает потенциальную токсичность вследствие вирусной нагрузки.

Недавно сообщалось, что применение промотора H1, вместо более традиционно используемого промотора U6 для контроля транскрипции гРНК позволяет приблизительно удвоить доступное пространство для генов нацеливания CRISPR (Ranganathan, V *et al. Nature communications* 5, 4516 (2014)). Примечательно, также была обнаружена более низкая склонность в отношении неспецифического расщепления, что указывает на то, что промотор H1 является более подходящим для терапевтических подходов. В ходе этих исследований было отмечено присутствие кодирующего белок гена (PARP-2) вблизи в геноме к эндогенному гену H1RNA (Baer, M. *et al. Nucleic acids research* 18, 97-103 (1990); Myslinski, E. *et al. Nucleic acids research research* 29, 2502-2509 (2001); Ame JC

et al. J Biol Chem. 276(14):11092-9 (2001)). Последовательность между началом РНК H1 (РНК-транскрипт pol III) и геном PARP-2 (транскрипт pol II) имеет размер 230 п.н. (фиг.2), что указывает на то, что эта относительно небольшая последовательность может выполнять функцию компактного двунаправленного промотора. Было предположено, что эти удачные свойства промотора H1 могут решить проблему размера при упаковывании обоих компонентов CRISPR в один AAV.

A. Усиление двунаправленной экспрессии pol II под контролем H1 с использованием модификаций 5'UTR.

Для разработки H1 в качестве двунаправленного промотора pol II/III и поскольку активность poll III уже хорошо охарактеризована, получали репортерную конструкцию с eGFP для лучшей оптимизации активности pol II (фиг.1A). Клетки человека (HEK293) и мыши (NIH3T3) продемонстрировали слабую, но отчетливо обнаруживаемую флуоресценцию GFP, что указывает на то, что промотор H1 мог обеспечивать экспрессию pol II. С использованием этой репортерной системы GFP проводили эксперименты по повышению экспрессии pol II при сохранении компактности путем оценки трех вариабельных компонентов в системе: (a) промоторная последовательность, (b) 5'UTR и (c) последовательность терминатора.

Для оценки различных последовательностей терминаторов тестировали семь различных последовательностей. Было обнаружено, что как терминаторы SV40 (240 п.н. и 120 п.н.), так и синтетическая последовательность поли(A) (SPA) размером 49 п.н. были функциональными в отношении экспрессии GFP, хотя последовательность SPA была менее эффективной.

Проводили исследование, улучшает ли оптимизация эффективности трансляции посредством модификации 5'UTR экспрессию репортера. Было обнаружено, что встраивание последовательности из 50 п.н., взятой из последовательности 5'UTR бета-глобина, было способно значительно улучшить экспрессию репортера и, в соответствии с этим, простое встраивание 9 оснований, кодирующих сильную последовательность Козак (5'-GCCGCCACC-3' SEQ ID NO: 1)

было достаточным для того, чтобы приблизиться к этим уровням.

Также было эффективным использование последовательностей, идентифицированных в качестве последовательностей РНК, которые опосредуют кэп-независимую инициацию трансляции (Wellensiek *et al. Nature Methods* 10, 747-750 (2013)) (6.947 и 6.967).

Обращая внимание на последовательность Козак, было обнаружено, что как сильная консенсусная последовательность из 9 оснований (5′-GCCGCCACC-3′ SEQ ID NO: 1), так и меньшая консенсусная последовательность из 6 оснований (5′-GCCACC-3′ SEQ ID NO: 2) могли усиливать экспрессию GFP (фиг.1). Последовательность из 9 оснований неизменно была несколько лучшей среди различных двунаправленных последовательностей H1.

B. Модулирование двунаправленной экспрессии с использованием различных ортологичных последовательностей.

Тестирование промоторных последовательностей H1 из различных организмов показало, что как последовательность мыши (176 п.н.), так и последовательность крысы (207 п.н.) была способна обеспечивать более сильную экспрессию GFP, чем промотор H1 человека (в ~7 и ~6 раз выше, соответственно). Этот результат противоположен тому, что было описано в других исследованиях, например, Hung *et al.* (2006) *Biochemical and Biophysical Research Communication* 339:1035-1042, в которых сообщалось, что последовательность человека экспрессировалась на более высоких уровнях.

Геномное выравнивание проводили для ортологичной области 36 плацентарных млекопитающих с использованием промоторной последовательности H1. Это обеспечивает дополнительные (SEQ ID NO: 84-119), которые можно использовать для тонкой регуляции двунаправленной экспрессии.

C. Новые компактные последовательности двунаправленного промотора с активностью в отношении как pol II, так и pol III

После установления того, что последовательность двунаправленного промотора H1 может эффективно контролировать экспрессию генов как pol II, так и pol III, проводили поиск в геноме потенциально компактных двунаправленных

последовательностей.

Был разработан специализированный perl-скрипт для сравнения 5′-участков инициации транскрипции гена pol III и гена pol II. Из результатов отбирали те, которые ориентированы в противоположных направлениях (расходящаяся транскрипция). Файлы входных данных могли представлять собой аннотированные файлы генома или данные транскрипции (участки ChIP pol II или pol III). Некоторые участки ChIP, которые были идентифицированы из генома человека, включают, но не ограничиваясь ими:

ALOXE3-tRNA(lys):

TCTTTCCGCTCCAGGACCGCCCTGGGCCTGCAGGATCCTGGGCGGGAGCCCAGGTGTCC
GGGATCTGGGCCACTAGGGACTGGGGAGGAACCTCTCAGAGAAGCCCATAGCCCGCAGCGGCCC
CGCGCGGCCGGTTCCGGCGCCGCACTGTTCCAGCCTCTACTATGGTACAGTCCCTGCGTCGCAG
CCTCGGCGGGGGCTCTAAGAACGGGAGGCAGAAAAAGCTCAATCAGCAGCAGGCGAGCTTCACC
CGCTGCTTCCAAATCTGTGCCAAAATATTCTATGCTGCACAGATAAAATCCTCTGTCGGTTCTA
CAAGCCTGGCTTTTCCTATAGAGAACCCTCTTATAAGCAAAAAGTAAAGCTCTCGTGAAGA;

C8orf41-SNORD13:

TCCTGACTGCAGCACCAGAAGGCTGGTCTCTCCCACAGAACGAGGATGGAGGCGGGGAG
GGATCCGTTGAAGAGGGAAGGAGCGATCACCCAAAGAGAACTAAAATCAAATAAAATAAAACAG
AGAGATGTCTTGGAGGAGGGGGCGAGTCTGACCGGGATAAGAATAAAGAGAAAGGGTGAACCCG
GGAGGCGGAGTTTGCAGTGAGCCGAGATCGCGCCACTGCACTCCAGCCTGGGCGACAGAGTGAG
ACTCCGTCTCAGTAAAAAAAAAAAAAAAAAAAAGAATAAAGAGGAAAGGACGCAAGAAAGGGAA
AGGGGACTCTCAGGGAGTAAAAGAGTCTTACACTTTTAACAGTGACGTTAAAAGACTACTGTTG
CCTTTCTGAAGACTAAAAAGAAAAAAAACTTAAAAATTTAAAGAAATAAACTTCTGAGCCATGT
CACCAACTTAACCACCCCCAGGTACCTGCAACGGCTCGCGCCCGCCGGTGTCTAACAGGATCCG
GACCTAGCTCATATTGCTGCCGCAAAACGCAAGGCTAGCTTCCGCCAGTACTGCCGCAACACCT
TCTTATTTCACGACGTATGGTCGTAAAGCAATAAAGATCCAGGCTCGGGAAAATGACGGAGAGG
TGGAACTATAGAGAATAAATTTGCATATATAATAATCCGCTCGCTAATTGTGTTTCTGTTTTCC
TTTGCTAAGGTAGAAACAAAAGAATAATCACAGAATCTCAGTGGGACTTTGAAAATATCCAGGA
TTTTATACGTGAAGAATGGATGTATCGCATTACGGTAGTCACCCTATGTGTAAATTAGTGGCAC
ATACTTGGCACTCCTTAATGTCAACTATAAGATG;

CGB1:

GTCTCTCTCTTAGCGGGATATCTTCCGCAAGCACTGGGAATGTGGACATGGAAAGTAAA
TTGAGTCTCCGTGGGGGAGTGAGACAGGGAGTGAGGGGTGTTGGACGCGGCACGGGAACCTGGC
CAGAGTCAGCGGACCCAATTGGCTGCTCTCTCTCAGATGCAGTTCCCCTTCCTCCCTCCAGGGG

GCGCCACGGAACGCAGGGCCCTCACTGGCCCTGGGGACTGGGTGACGTCAGGGATGAGCCTCTT
GTGATTGGCTCCATCACCCTGCGTAAGATCAAAGGGAAGAAAGGATGGGCCCGACAA;

CGB2:

GTCTCTCTCTTAGCGGGATATCTTCCGCAAGCACTGGGGATGTGGACATGGAAAGTAAA
TTGAGTCTCCGTGGGGGAGTGAGACAGGGAGTGAGGGGTGTTGGACGCGGCACGGGAACCCGGC
CGGAGTCAGCGGACCCAATTGGCTGCTCTCTCTCAGATACAGTTCCCCTTCCTCCCTCCAGGGG
GCGCCACGGAACGCAGGGCCCTCACTGGCCCTGGGGACTGGGTGACGTCAGGGGTGAGCCTCTC
CTGATTGGCTCCATCACCCTGCGTAAGGTCAAAAGGAAGAAAGGAGATCCCCGACAC;

DPP9-tRNA(gly):

TAACCGCTCAGCTGACCTCAGGAGGGCAGGGGTGCCTTCTAAAGGGTCCAGAGAGCCTC
CATTCCAGCTGCAGGCGTGGGACACAGACCGGGACGTGGGGCGGCGGCCGGACTGGGCAGGTCG
TCCCGGGTCCAGCGGCGCCTCACGGTCGCGGCTCCATGCCCGGGACTGCGACCCCGGAAGTGGC
GGGAGCGGGGGACGACAGCCGCGGCGGACACAGGGGACCCGCCGGCTCAGGCACCTTTGACCCG
GAAGTTGAGCGACCCAGGCGGCGGCCTGGGATTGGACACCACCAGGCACGTACCAAGGCGTCCG
CGGCGCTTGGGGGGGAGCCCGCGGCGCGGCGGCCTAAGGTGCGTAACGCCCCATGAACGACATC
TTCCGGTGGGTTAGGGAGAGACACCCCCCTGTGACTTGGTATCACTCAGTCAAACCCATGATCC
CCCACTATTAAGGATATCCGGAGAGGATGCTACCTATCAGG;

MED16-RNU6-9:

GAGGGCAGTCACCAGCTCCTGGCCCGTGCGCCAAGCTCAGCGGGCGTCCGCGGTGCGAT
CTTCCCTAGCGCCTCGGGTCTGGCGCCGCCATCTTCCTCGGTAACAACCAGTCGCCTGAGGCGT
GGGGCCGCCTCCCAAAGACTTCTGGGAGGGCGGTGCGGCTCAGGCTCTGCCCCGCCTCCGGGGC
TATTTGCATACGACCAT

Чтобы подтвердить это, проводили тестирование двунаправленного промотора 7sk или области между геном GSTA4 (ген pol II) и геном RN7SK (ген pol III) с использованием репортерного анализа GFP авторов изобретения. Активность этого промотора в отношении pol III документально подтверждена.

Одновременно тестировали несколько вариантов 5′-UTR; эндогенно аннотированная 5′UTR для гена GSTA4 человека разделена интроном, и ее объединяли при клонировании в репортерную систему, разработанную авторами изобретения (7sk1). Также тестировали укороченную форму 5′UTR или 5′UTR бета-глобина (7sk3). Все использованные конструкции включали сильную консенсусную последовательность Козак из 9 оснований (5′-GCCGCCACC-3′ SEQ ID NO: 1) (7sk2).

Репортерный анализ показал, что эта область была способна контролировать экспрессию pol II, подтвердив способность действовать в качестве двунаправленной промоторной области. Хотя эта область является несколько большей, чем область двунаправленного промотора H1, было обнаружено, что она является более активной.

Проводили тестирование ортологичных последовательностей в отношении обеспечения экспрессии. Последовательность мыши также была активной и несколько более сильной, чем последовательность человека.

Эти результаты подтвердили наличие двунаправленных промоторов с активностью в отношении как pol II, так и pol III, и возможность проведения поиска геномных последовательностей для этих областей.

D. Скрининг экспрессии для двунаправленных промоторов с активностью в отношении как РНК-pol II, так и РНК-pol III

Биоинформатика:

Последовательности двунаправленного промотора можно использовать в качестве начальных точек для идентификации различных вариантов активности pol II/pol III. Был разработан специализированный perl-скрипт для сравнения 5'-участков инициации транскрипции гена pol III и гена pol II. Из результатов отбирали те, которые ориентированы в противоположных направлениях (расходящаяся транскрипция). Файлы входных данных могли представлять собой аннотированные файлы генома или данные транскрипции (участки ChIP pol II или pol III). С использованием этой информации были идентифицированы двунаправленные промоторы с активностью как в отношении РНК-pol II, так и в отношении РНК-pol III. Эти последовательности можно было использовать для получения производных последовательностей.

Участки связывания фактора транскрипции, являющиеся мишенью:

Промоторные последовательности можно было использовать для идентификации транскрипционных участков связывания, или несколько промоторных последовательностей можно было выравнивать

для идентификации участков связывания факторов транскрипции, которые в свою очередь можно было использовать для конструирования двунаправленного промотора. Идентификацию участков связывания факторов транскрипции можно было проводить посредством консенсусных последовательностей или с использованием матрицы дифференциальных расстояний или многомерного масштабирования (De Bleser P *et al*. *Genome Biol*. 2007;8(5):R83.).

Библиотека случайных последовательностей:

Альтернативно синтетическую библиотеку случайных последовательностей можно было использовать в качестве начальной точки для скрининга последовательностей с двунаправленной активностью. Путем рандомизации, шаффлинга или внесения мутаций в двунаправленную последовательность можно проводить поиск последовательностей ДНК, которые имеют различную активность промотора pol II или pol III.

Скрининг можно было проводить с использованием кПЦР или флуоресценции для определения активности pol II и pol III. В простом флуоресцентном скрининге может быть использован репортер (GFP, mCherry и т.д.), как описано, и РНК-аптамер, такой как Broccoli, который становится флуоресцентным в присутствии низкомолекулярного соединения, такого как DFHB1-1T. В этом скрининге можно было использовать данные для белков, такие как резистентность к лекарственным средствам (например, антибиотики), антигенные пептиды или маркеры клеточной поверхности и т.д.

Альтернативными подходами могли быть функциональные комплексы, которые состоят как из белка, так и из РНК. Например, в РНК можно было включать структуру стебель-петля, связывающую белок оболочки MS2, и транскрипция pol II могла кодировать белок оболочки MS2, который в свою очередь связывался бы с последовательностью РНК. Если в РНК включить два участка связывания, тогда в качестве данных можно использовать комплементарность флуоресценции (комплементарность разделенного белка).

Скрининг также можно было проводить путем разделения и

оценки активности pol II и активности pol III. Наконец, затеи эти последовательности можно было комбинировать так, чтобы позволить расходящуюся транскрипцию как РНК-pol II, так и РНК-pol III.

Способы

Конструирование плазмиды: репортерные плазмиды с GFP конструировались Gibson Assembly (New England Biolabs, Ipswich, MA) (Gibson *et al.* (2009) *Nature* Способы 6:343-345) с небольшими модификациями.

Линию клеток почки эмбриона человека (HEK) 293T (Life Technologies, Grand Island, NY) поддерживали при 37°C с 5% CO_2/20% O_2 в модифицированной способом Дульбекко среде Игла (DMEM) (Invitrogen), дополненной 10% эмбриональной телячьей сывороткой (инактивированной нагреванием) (Gibco, Life Technologies, Grand Island, NY) и 2 мМ GlutaMAX (Invitrogen).

Линию клеток N2A (Life Technologies, Grand Island, NY) поддерживали при 37°C с 5% CO_2/20% O_2 в модифицированной способом Дульбекко среде Игла (DMEM) (Invitrogen), дополненной 10% эмбриональной телячьей сывороткой (Gibco, Life Technologies, Grand Island, NY) и 2 мМ GlutaMAX (Invitrogen).

Клетки трансфицировали (Lipofectamine 3000) каждой конструкцией. Экспрессию GFP анализировали с использованием проточной цитометрии.

7sk1 человека (GSTA4 5'UTR с последовательностью Козак):
GGTGGCGGCGATAGCTTTTCAGGCTTTCTGGAGTCCACTCGGAGGCCTGGAGCCGCACAAAGCG
CCAGGTCAGCGGTCCCGGCTGGGTGAGACCAGCAGGCGGCTCTAGCGCGCGGGAGCTGGGCGAG
GCTCCGGGACGACCTCACCAATGGAGACTGCAGTATTTAGCATGCCCCACCCATCTGCAAGGCA
TTCTGGATAGTGTCAAAACAGCCGGAAATCAAGTCCGTTTATCTCAAACTTTAGCATTTTGGGA
ATAAATGATATTTGCTATGCTGGTTAAATTAGATTTTAGTTAAATTTCCTGCTGAAGCTCTAGT
ACGATAAGCAACTTGACCTAAGTGTAAAGTTGAGACTTCCTTCAGGTTTATATAGCTTGTGCGC
CGCTT GGGTACCTC (SEQ ID NO: 3)

7sk2 человека (короткая 5'UTR):
GGTGGCGGCAGTATTTAGCATGCCCCACCCATCTGCAAGGCATTCTGGATAGTGTCAAAACAGC
CGGAAATCAAGTCCGTTTATCTCAAACTTTAGCATTTTGGGAATAAATGATATTTGCTATGCTG
GTTAAATTAGATTTTAGTTAAATTTCCTGCTGAAGCTCTAGTACGATAAGCAACTTGACCTAAG

TGTAAAGTTGAGACTTCCTTCAGGTTTATATAGCT TGTGCGCCGCTTGGGTACCTC (SEQ ID NO: 4)

7sk3 человека (5′UTR бета-глобина):

GGTGTCTGTTTGAGGTTGCTAGTGAACACAGTTGTGTCAGAAGCAAATGTAGTATTTAGCATGC
CCCACCCATCTGCAAGGCATTCTGGATAGTGTCAAAACAGCCGGAAATCAAGTCCGTTTATCTC
AAACTTTAGCATTTTGGGAATAAATGATATTTGCTATGCTGGTTAAATTAGATTTTAGTTAAAT
TTCCTGCTGAAGCTCTAGTACGATAAGCAACTTGACCTAAGTGTAAAGTTGAGACTTCCTTCAG
GTTTATATAGCTTGTGCGCCGCTTGG GTACCTC (SEQ ID NO: 5)

7sk1 мыши (GSTA4 5′UTR с последовательностью Козак):

GGTGGCGGCCTCGTATGGGACCGCACCGGACACGGGCGCCTGGGCCAGGAGCAGAGCCGGGCCG
TAGAATAGACATGGCCGTCGGGGGCGGGGCTTCGGAAGGTTTAACCAATCCAAACTGTTGTATT
TTGCATAGCCCCAAAGCATTTTGGTTAACAGTAAAAACATCCTAAATTTAAGTATTTTAATTTA
AACTTAGAACGAAGCGAGTATAAAAAGGATTATTTAACCCTAAAACGGATTCAGGATTTGTTAT
AATATCAAGTACAGTCGGCTACATAAGGTCACCACATGTGTAAAGTTACAAAATTCTATGGCCT
TATATACCTACCAAGA GCCTGCTTACTCTC (SEQ ID NO: 6)

7sk2 мыши (короткая 5′UTR):

GGTGGCGGCAGACATGGCCGTCGGGGGCGGGGCTTCGGAAGGTTTAACCAATCCAAACTGTTGT
ATTTTGCATAGCCCCAAAGCATTTTGGTTAACAGTAAAAACATCCTAAATTTAAGTATTTTAAT
TTAAACTTAGAACGAAGCGAGTATAAAAAGGATTATTTAACCCTAAAACGGATTCAGGATTTGT
TATAATATCAAGTACAGTCGGCTACATAAGGTCACCACATGTGTAAAGTTACAAAATTCTATGG
CCTTATATACCTACCAAGAGCCTGCTT ACTCTC (SEQ ID NO: 7)

7sk3 мыши (5′-UTR бета-глобина):

GGTGTCTGTTTGAGGTTGCTAGTGAACACAGTTGTGTCAGAAGCAAATGTAGACATGGCCGTCG
GGGGCGGGGCTTCGGAAGGTTTAACCAATCCAAACTGTTGTATTTTGCATAGCCCCAAAGCATT
TTGGTTAACAGTAAAAACATCCTAAATTTAAGTATTTTAATTTAAACTTAGAACGAAGCGAGTA
TAAAAAGGATTATTTAACCCTAAAACGGATTCAGGATTTGTTATAATATCAAGTACAGTCGGCT
ACATAAGGTCACCACATGTGTAAAGTTACAAAATTCTATGGCCTTATATACCTACCAAGAGCCT
GCTTACTCTC (SEQ ID NO: 8)

1. Dryja, TP *et al. The New England Journal of Medicine* 323, 1302-1307 (1990).

2. Dryja, TP *et al. Nature* 343, 364-366 (1990).

3. Doudna, JA *et al. Science* 346, 1258096 (2014).

4. Hsu, PD *et al. Cell* 157, 1262-1278 (2014).

5. Liang, Y. *et al. The Journal of Biological Chemistry* 279, 48189-48196 (2004).

6. Dalkara, D. & Sahel, J. A. *Comptes Rendus Biologies* 337, 185-192 (2014).

7. Day, TP. *et al. Advances in Experimental Medicine and Biology* 801, 687-693 (2014).

8. Willett, K. & Bennett, Frontiers in immunology 4, 261, (2013).

9. Dinculescu, A. *et al. Human Gene Therapy* 16, 649-663 (2005).

10. Kotterman, MA *et al. Gene Therapy* 22, 116-126 (2015).

11. Mowat, FM *et al. Gene Therapy* 21, 96-105 (2014).

12. Dalkara, D. *et al. Science Translational Medicine* 5, 189ra176 (2013).

13. Berns, KI *et al.* Fundamental Virology (ed B.N. Fields, и Knipe, D.M.) 545-562 (Raven Press, 1986).

14. Swiech, L. *et al. Nature Biotechnology* 33, 102-106 (2015).

15. Jinek, M. *et al. Science* 337, 816-821 (2012).

16. Cong, L. *et al. Science* 339, 819-823 (2013).

17. Mali, P. *et al. Science* 339, 823-826 (2013).

18. Mancuso, K. *et al. Nature* 461, 784-787 (2009).

19. Beltran, WA *et al. Proceedings of the National Academy of Sciences of the United States of America* 109, 2132-2137 (2012).

20. Petrs-Silva, H. *et al. Molecular therapy: The Journal of the American Society of Gene Therapy* 19, 293-301 (2011).

21. Song, C. *et al. Epigenetics & Chromatin* 7, 17 (2014).

22. Jain, D. *et al. Pediatric Neurology* 43, 35-40 (2010).

23. Ranganathan, V *et al. Nature Communications* 5, 4516 (2014).

24. Baer, M. *et al. Nucleic Acids Research* 18, 97-103 (1990).

25. Myslinski, E. *et al. Nucleic Acids Research* 29, 2502-2509 (2001).

26. Ame JC *et al. J Biol Chem.* 276(14):11092-9 (2001).

ПРИМЕР 2

Зависимая от условий экспрессия с двунаправленного

промотора pol II/pol III для регуляции ферментативной активности рибонуклеопротеина или РНК-направленных нуклеаз

В предыдущей работе было продемонстрировано, что активность промотора H1 pol III можно встраивать способами инженерии в отвечающий на tet промотор и можно регулировать тетрациклином. В присутствии последовательностей Tet-репрессора (TetR), последовательности Tet-оператора (TetO), встроенные в промотор H1, связываются с TetR и подавляют экспрессию pol III.

Эту систему ранее не комбинировали с двунаправленным компонентом промотора H1, системой, которая может позволить регуляцию транскриптов как pol II, так и pol III. Рибонуклеопротеиновые ферменты или РНК-направленные нуклеазы можно регулировать посредством тетрациклина с использованием сконструированной двунаправленной промоторной системы.

В настоящем описании было показано, что активность H1 в отношении pol II не изменялась значительным образом в результате присутствия участков оператора TetO. Это обеспечивает механизм индуцибельной активности рибонуклеопротеиновых ферментов или РНК-направленных нуклеаз, таких как Cas9/гРНК. Кроме того, вставка участков TetO может быть осуществлена таким образом, чтобы специфически подавлять транскрипты либо pol II, либо pol III, или оби из них.

Такая система может иметь значительные клинические преимущества для регуляции активности системы CRISPR, доставляемой с помощью вирусов AAV, и ее можно использовать для других компактных двунаправленных промоторов.

1. Matthess *et al*. *Oncogene*. 2005 Apr 21;24(18):2973-80.

Способы

Конструирование плазмиды: Репортерные плазмиды с GFP конструировались Gibson Assembly (New England Biolabs, Ipswich, MA) (Gibson *et al*. (2009) *Nature* Способы 6:343-345) с модификациями.

Линию клеток почки эмбриона человека (HEK) 293T (Life Technologies, Grand Island, NY) поддерживали при 37°C с 5% CO_2/20% O_2 в модифицированной способом Дульбекко среде Игла

(DMEM) (Invitrogen), дополненной 10% эмбриональной телячьей сывороткой (инактивированной нагреванием) (Gibco, Life Technologies, Grand Island, NY) и 2 мМ GlutaMAX (Invitrogen).

Линию клеток N2A (Life Technologies, Grand Island, NY) поддерживали при 37°C с 5% CO_2/20% O_2 в модифицированной способом Дульбекко среде Игла (DMEM) (Invitrogen), дополненной 10% эмбриональной телячьей сывороткой (Gibco, Life Technologies, Grand Island, NY) и 2 мМ GlutaMAX (Invitrogen).

Клетки трансфицировали (Lipofectamine 3000) каждой конструкцией. Экспрессию GFP анализировали с использованием проточной цитометрии.

последовательность-мишень mm079: GAAGAAGGTTCGAGATCTCA (SEQ ID NO: 9)

геномная последовательность-мишень mm079: GAAGAAGGTTCGAGATCTCAAGG (SEQ ID NO: 10)

Участок TetO: TCCCTATCAGTGATAGAGA (SEQ ID NO: 11)

H1_wt:

GGAATTCGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCCCAGTGTCACTA GGCGGGAACACCCAGCGCGCGTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAGGGGAGTGGC GCCCTGCAATATTTGCATGTCGCTATGTGTTCTGGGAAATCACCATAAACGTGAAATGTCTTTG GATTTGGGAATCTTATAAGTTCTGTATGAGACCACTTTTTCCC (SEQ ID NO: 12)

H1_TetO:

GGAATTCGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCCCAGTGTCACTA GGCGGGAACACCCAGCGCGCGTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAGGGGAGTGGC GCCCTGCAATATTTGCATGTCGCTATGTGTTCTGGGAAATCACCATAAACGTGAAATCCCTATC AGTGATAGAGACTTATAAGTTCCCTATCAGTGATAGAGATCCC (SEQ ID NO: 13)

Последовательность TetR:

MSRLDKSKVINSALELLNEVGIEGLTTRKLAQKLGVEQPTLYWHVKNKRALLDALAIEM LDRHHTHFCPLEGESWQDFLRNNAKSFRCALLSHRDGAKVHLGTRPTEKQYETLENQLAFLCQQ GFSLENALYALSAVGHFTLGCVLEDQEHQVAKEERETPTTDSMPPLLRQAIELFDHQGAEPAFL FGLELIICGLEKQLKCESGSAYSGSREF (SEQ ID NO: 14)

ПРИМЕР 3

A. Регуляция модифицирующих геном нуклеаз посредством постранскрипционной обусловленной клеточным циклом регуляции

Появление новых РНК-направленных нуклеаз произвело

революцию в технологии редактирования генов и биологическом исследовании трансформации. Эти технологии, хотя и обеспечивают беспрецедентные средства точной модификации геномов, ограничиваются клеточными процессами, которые модулируют репарацию разрывов ДНК. Для применений для модификации генома, в частности, способность модулировать клеточный ответ на разрывы ДНК имеет огромные последствия (см. Gutschner *et. al*. (2016) Cell Reports).

Клеточный ответ на разрывы ДНК происходит в основном в основном посредством одного из двух конкурирующих путей: негомологичное соединение концов (NHEJ) или направляемая гомологией репарация (HDR). NHEJ обычно считается подверженным ошибкам путем, который приводит к изменению последовательности вокруг точки разрыва. Как правило, NHEJ является более эффективным или преобладающим путем репарации ДНК. С другой стороны, пути HDR значительно меньше подвержены ошибкам, но требуют участков гомологии с матрицей и репарации.

Оба пути имеют свойства, которые способствуют различным исходам. Например, если целью является "нокаут" гена, предпочтительным путем является NHEJ, поскольку разрывы ДНК в этом гене по большей части приведут к изменениям последовательности. В лаборатории многие генетические скрининги зависят от устранения функции гена и для терапевтических подходов может быть желательным полное разрушение гена с доминантной мутацией или мутацией с приобретением функции. Однако, если целью является внесение определенного изменения последовательности ("нокин"), можно попытаться осуществить репарацию посредством пути HDR, таким образом, обеспечивая точное разрезание и вставку. Для многих научным подходов и большинства клинических применений этот каскад является предпочтительным, поскольку он может позволить точную репарацию мутации.

Также известно, что клеточный ответ на разрывы ДНК модулируется клеточным циклом (фиг.9A); фаза клеточного цикла в значительной степени определяет выбор пути репарации ДНК (фиг.9B). NHEJ преобладает в G1 клеточного цикла, в то время как

HDR преобладает в ходе фаз S и G2. Таким образом, способность модулировать время разрывов ДНК в отношении клеточного цикла может быть мощным способом смещения клетки в направлении конкретного пути репарации.

Множество белков регулируется зависимым от клеточного цикла образом, в основном посредством механизмов регуляции транскрипции и посттранскрипционных механизмов, в частности, опосредуемого убиквитином протеолиза. Как правило, регуляция посредством транскрипционных механизмов является более медленной, в то время как протеолитические механизмы являются быстрыми. Для применений для редактирования генома протеолитические механизмы, вероятно, обеспечивают более благоприятный способ регуляции.

Двумя хорошо охарактеризованными белками с противоположной и, таким образом, переменной, экспрессией в клетке являются геминин (Gem) и Cdt1. Cdt1 накапливается в фазе G1 клеточного цикла, в то время как Gem накапливается в ходе фаз S/G2/M. Были картированы конкретные области этих белков, которые требуются для регуляции клеточного цикла: аминокислоты 30-120 для hCdt1 и аминокислоты 1-110 для hGem (Sakaue-Sawano A *et al*. *Cell* 132, 487-498 (2008)). Более того, можно получать белки, слитые с этими доменами таким образом, чтобы они демонстрировали зависимую от клеточного цикла регуляцию, даже если соответствующая мРНК постоянно транскрибируется на протяжении клеточного цикла. В сущности, клеточные белки распознают домен только в ходе конкретных фаз клеточного цикла, что приводит к убиквитинилированию, а затем быстрой деградации слитого белка.

В данном случае, было показано, что технологии редактирования генома, такие как РНК-направленные нуклеазы, можно подвергать слиянию с доменами либо hGem, либо hCdt1, что обеспечивает зависимую от клеточного цикла регуляцию этих белков (фиг.9C). Следовательно, эта регуляция приводит к значительно отличающимся исходам репарации ДНК. С использованием системы CRISPR-Cas9 было показано, что как путь NHEJ, так и путь HDR, усиливался или подавлялся с использованием Cas9 слияния либо hGem, либо hCdt1. Эта систем является настолько надежной, что

она может повысить уровень HDR до уровней выше NHEJ, что является в высокой степени значительным с учетом того, что HDR является значительно менее эффективным в клетке.

Более того, описанный способ является в высокой степени впечатляющим своей элегантностью и простотой. Для других попыток регулировать исход технологий редактирования генома использовались трудоемкие способы остановки клеток в определенных фазах клеточного цикла, после чего следовали технически трудные и неэффективные способы обеспечения доставки белка/РНК. Вместо этого, способ, описанный в настоящем описании, основан на простой доставке плазмиды, мРНК или белка, кодирующих слитую конструкцию, с соответствующими регуляторными последовательностями.

Способы

Конструирование плазмиды: Для получения регулируемых клеточным циклом нуклеазных белков кодон-оптимизированный ген Cas9 человека из вектора Cas9:T2A:GFP (Addgene 44719) модифицировали для замены T2A лучшей последовательностью, расщепляющей P2A. Далее флуоресцентный репортер с GFP заменяли последовательностью, кодирующей флуоресцентный белок mKate2. Далее вектор линеаризовывали и gBlock, кодирующий гибкий линкер из 15 аминокислот, слитый либо с доменом hGem(1-110), либо с доменом hCdt1(30-120), встраивали в рамке считывания с получением Cas9:линкер:hGem и Cas9:линкер:hCdt1 (обозначаемых как Cas9-Gem или Cas9-Cdt, соответственно). Все стадии клонирования проводили с использованием Gibson Assembly (New England Biolabs, Ipswich, MA) (Gibson *et al*. (2009) *Nature Methods* 6:343-345) с небольшими модификациями.

Линию клеток почки эмбриона человека (HEK) 293T (Life Technologies, Grand Island, NY) поддерживали при 37°C с 5% CO_2/20% O_2 в модифицированной способом Дульбекко среде Игла (DMEM) (Invitrogen), дополненной 10% эмбриональной телячьей сывороткой (Gibco, Life Technologies, Grand Island, NY) и 2 мМ GlutaMAX (Invitrogen).

гРНК, нацеливающую на родопсин (hs086172175:

AGTACTGTGGGTACTCGAAGGGG (SEQ ID NO: 68)) (см. Jaskula-Ranga, V., & Zack, D. J. (2016). grID: A CRISPR-Cas9 guide RNA Database and Resource for Genome-Editing. *bioRxiv*, 097352) получали с использованием перекрывающихся олигонуклеотидов, которые подвергали отжигу и амплификации способом ПЦР с использованием ДНК-полимеразы Phusion Flash для двухэтапной амплификации (Thermo Fisher Scientific, Rockford, IL), а затем очищали с использованием Zymo DNA clean и концентрирующих колонок. Затем очищенные продукты ПЦР ресуспендировали в H2O и количественно определяли с использованием NanoDrop 1000 (Thermo Fisher Scientific). Экспрессирующие гРНК конструкции получали с использованием Gibson Assembly (New England Biolabs, Ipswich, MA) (Gibson *et al*. *Nature Methods* 6:343-345 (2009)) с небольшими модификациями. Общий объем реакции уменьшался от 20 мкл до 2 мкл.

Клетки HEK293 совместно трансфицировали посредством Cas9 (немодифицированная или регулируемая клеточным циклом версия) и конструкций гРНК, нацеливающей на родопсин. Через 48 ч после трансфекции геномную ДНК собирали и последовательность, окружающую участки разрезания мишени, амплифицировали в соответствии с праймерами, приведенными ниже. Затем продукты ПЦР очищали и количественно определяли перед проведением анализа с T7 Endo I. В кратком изложении, 200 нг продукта ПЦР денатурировали, а затем медленно обратно гибридизовали для образования гетеродуплексов, к продуктам ПЦР добавляли эндонуклеазу I T7 и инкубировали при 37C в течение 25 минут для расщепления гетеродуплексов, реакцию гасили нагрузочным красителем, и, наконец, реакционную смесь разделяли на 6% геле TBE PAGE для разделения продуктов. Гел окрашивали SYBR-Gold, визуализировали и подвергали количественному определению с использованием ImageJ. Частоты NHEJ вычисляли с использованием двучленного уравнения:

$$1-\sqrt{1-\frac{(a+b)}{(a+b+c)}}\ \mathbf{X}\ 100$$

где величины "a" и "b" равны интегрированной площади для расщепленных фрагментов после вычитания фонового сигнала и "c"

равен интегрированной площади для нерасщепленного продукта ПЦР после вычитания фонового сигнала. Для вычисления частот HDR тот же продукт ПЦР, который описан выше, перед формированием гетеродуплекса смешивали с EcoRI в течение 1 ч, затем гасили нагрузочным красителем и, наконец, реакцию проводили на 6% геле TBE PAGE для разделения продуктов.

Частоты HDR вычисляли с использованием уравнения:

$$\%\,\mathrm{HDR} = \frac{(a+b)}{(a+b+c)} * 100$$

где величины "a" и "b" равны интегрированной площади для расщепленных фрагментов после вычитания фонового сигнала и "c" равен интегрированной площади для нерасщепленного продукта ПЦР после вычитания фонового сигнала.

Олигонуклеотид HDR (EcoRI):

CAGAAGGCCCTAACTTCTACGTGCCCTTCTCCAATGCGACGGGTGTGGTACGCAGCCCC TGAATTCACCCACAGTACTACCTGGCTGAGCCATGGCAGTTCTCCATGCTGGCCGCCTACATGT TTC (SEQ ID NO: 15)

Праймеры T7EI:

Rho_HDR_F; TGGAGCCCTGAGTGGCTGAG (SEQ ID NO: 16)

Rho_HDR_R; CCACCTAGGACCATGAAGAGGTCAG (SEQ ID NO: 17)

Белковые последовательности:

hCdt1(30-120)

PSPARPALRAPASATSGSRKRARPPAAPGRDQARPPARRRLRLSVDEVSSPSTPEAPDI PACPSPGQKIKKSTPAAGQPPHLTSAQDQDTI (SEQ ID NO: 18)

hGem (1-110)

MNPSMKQKQEEIKENIKNSSVPRRTLKMIQPSASGSLVGRENELSAGLSKRKHRNDHLT STTSSPGVIVPESSENKNLGGVTQESFDLMIKENPSSQYWKEVAEKRRKAL (SEQ ID NO: 19)

Cas9:линкер:hCdt1

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIV DEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQ LVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPN FKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKA PLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPI LEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKI

LTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE

VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF
LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKD
KDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLI
NGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSP
AIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTR
SDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVE
TRQITKHVAQILDSRMNTKYDENDKLIRE

VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYG
DYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD
KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTV
AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSL
FELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD
EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
DRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKVRGGGGSGGGGSGGGGS
PSPARPALRAPASATSGSRKRARPPAAPGRDQARPPARRRLRLSVDEVSSPSTPEAPDIPACPS
PGQKIKKSTPAAGQPPHLTSAQDQDTI (SEQ ID NO: 20)

Cas9:линкер:hGem

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIV
DEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQ
LVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPN
FKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKA
PLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPI
LEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKI
LTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD
KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKG
ILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNR
GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQIT
KHVAQILDSRMNTKYDENDKLIRE

VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYG

DYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD
KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTV
AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSL
FELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD
EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
DRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKVRGGGGSGGGGSGGGGS
MNPSMKQKQEEIKENIKNSSVPRRTLKMIQPSASGSLVGRENELSAGLSKRKHRNDHLTSTTSS
PGVIVPESSENKNLGGVTQESFDLMIKENPSSQYWKEVAEKRRKAL (SEQ ID NO: 21)

Cdt1:линкер:Cas9

MPSPARPALRAPASATSGSRKRARPPAAPGRDQARPPARRRLRLSVDEVSSPSTPEAPD
IPACPSPGQKIKKSTPAAGQPPHLTSAQDQDTIGGGGSGGGGSGGGGSDKKYSIGLDIGTNSVG
WAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICY
LQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDS
TDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAK
AILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDD
DLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKA
LVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK
QRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWM
TRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY
VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT
YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT
GWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLH
EHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIE
EGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKD
DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDK
AGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVRE
INNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSN
IMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGF
SKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIM
ERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYV
NFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRID
LSQLGGDSRADPKKKRKVR (SEQ ID NO: 22)

Gem:линкер:Cas9

MNPSMKQKQEEIKENIKNSSVPRRTLKMIQPSASGSLVGRENELSAGLSKRKHRNDHLT
STTSSPGVIVPESSENKNLGGVTQESFDLMIKENPSSQYWKEVAEKRRKALGGGGSGGGGSGGG

GSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATR
LKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA
YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQT
YNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSN
FDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSA
SMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKM
DGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFR
IPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKH
SLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTY
AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLT
FKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAREN
QTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL

DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR
EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKV
YDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRD
FATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSV
LVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELE
NGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIE
QISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
YTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKVR (SEQ ID NO: 23)

Cas9:Cdt1:

ATGGACAAGAAGTACTCCATTGGGCTCGATATCGGCACAAACAGCGTCGGCTGGGCCGT
CATTACGGACGAGTACAAGGTGCCGAGCAAAAAATTCAAAGTTCTGGGCAATACCGATCGCCAC
AGCATAAAGAAGAACCTCATTGGCGCCCTCCTGTTCGACTCCGGGGAGACGGCCGAAGCCACGC
GGCTCAAAAGAACAGCACGGCGCAGATATACCCGCAGAAAGAATCGGATCTGCTACCTGCAGGA
GATCTTTAGTAATGAGATGGCTAAGGTGGATGACTCTTTCTTCCATAGGCTGGAGGAGTCCTTT
TTGGTGGAGGAGGATAAAAAGCACGAGCGCCACCCAATCTTTGGCAATATCGTGGACGAGGTGG
CGTACCATGAAAAGTACCCAACCATATATCATCTGAGGAAGAAGCTTGTAGACAGTACTGATAA
GGCTGACTTGCGGTTGATCTATCTCGCGCTGGCGCATATGATCAAATTTCGGGGACACTTCCTC
ATCGAGGGGGACCTGAACCCAGACAACAGCGATGTCGACAAACTCTTTATCCAACTGGTTCAGA
CTTACAATCAGCTTTTCGAAGAGAACCCGATCAACGCATCCGGAGTTGACGCCAAAGCAATCCT
GAGCGCTAGGCTGTCCAAATCCCGGCGGCTCGAAAACCTCATCGCACAGCTCCCTGGGGAGAAG
AAGAACGGCCTGTTTGGTAATCTTATCGCCCTGTCACTCGGGCTGACCCCCAACTTTAAATCTA
ACTTCGACCTGGCCGAAGATGCCAAGCTTCAACTGAGCAAAGACACCTACGATGATGATCTCGA

CAATCTGCTGGCCCAGATCGGCGACCAGTACGCAGACCTTTTTTTGGCGGCAAAGAACCTGTCA
GACGCCATTCTGC

TGAGTGATATTCTGCGAGTGAACACGGAGATCACCAAAGCTCCGCTGAGCGCTAGTATG
ATCAAGCGCTATGATGAGCACCACCAAGACTTGACTTTGCTGAAGGCCCTTGTCAGACAGCAAC
TGCCTGAGAAGTACAAGGAAATTTTCTTCGATCAGTCTAAAAATGGCTACGCCGGATACATTGA
CGGCGGAGCAAGCCAGGAGGAATTTTACAAATTTATTAAGCCCATCTTGGAAAAAATGGACGGC
ACCGAGGAGCTGCTGGTAAAGCTTAACAGAGAAGATCTGTTGCGCAAACAGCGCACTTTCGACA
ATGGAAGCATCCCCCACCAGATTCACCTGGGCGAACTGCACGCTATCCTCAGGCGGCAAGAGGA
TTTCTACCCCTTTTTGAAAGATAACAGGGAAAAGATTGAGAAAATCCTCACATTTCGGATACCC
TACTATGTAGGCCCCCTCGCCCGGGGAAATTCCAGATTCGCGTGGATGACTCGCAAATCAGAAG
AGACCATCACTCCCTGGAACTTCGAGGAAGTCGTGGATAAGGGGGCCTCTGCCCAGTCCTTCAT
CGAAAGGATGACTAACTTTGATAAAAATCTGCCTAACGAAAAGGTGCTTCCTAAACACTCTCTG
CTGTACGAGTACTTCACAGTTTATAACGAGCTCACCAAGGTCAAATACGTCACAGAAGGGATGA
GAAAGCCAGCATTCCTGTCTGGAGAGCAGAAGAAAGCTATCGTGGACCTCCTCTTCAAGACGAA
CCGGAAAGTTACCGTGAAACAGCTCAAAGAAGACTATTTCAAAAAGATTGAATGTTTCGACTCT
GTTGAAATCAGCGGAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGTATCACGATCTCCTGA
AAATCATTAAA

GACAAGGACTTCCTGGACAATGAGGAGAACGAGGACATTCTTGAGGACATTGTCCTCAC
CCTTACGTTGTTTGAAGATAGGGAGATGATTGAAGAACGCTTGAAAACTTACGCTCATCTCTTC
GACGACAAAGTCATGAAACAGCTCAAGAGGCGCCGATATACAGGATGGGGGCGGCTGTCAAGAA
AACTGATCAATGGGATCCGAGACAAGCAGAGTGGAAAGACAATCCTGGATTTTCTTAAGTCCGA
TGGATTTGCCAACCGGAACTTCATGCAGTTGATCCATGATGACTCTCTCACCTTTAAGGAGGAC
ATCCAGAAAGCACAAGTTTCTGGCCAGGGGGACAGTCTTCACGAGCACATCGCTAATCTTGCAG
GTAGCCCAGCTATCAAAAAGGGAATACTGCAGACCGTTAAGGTCGTGGATGAACTCGTCAAAGT
AATGGGA

AGGCATAAGCCCGAGAATATCGTTATCGAGATGGCCCGAGAGAACCAAACTACCCAGAA
GGGACAGAAGAACAGTAGGGAAAGGATGAAGAGGATTGAAGAGGGTATAAAAGAACTGGGGTCC
CAAATCCTTAAGGAACACCCAGTTGAAAACACCCAGCTTCAGAATGAGAAGCTCTACCTGTACT
ACCTGCAGAACGGCAGGGACATGTACGTGGATCAGGAACTGGACATCAATCGGCTCTCCGACTA
CGACGTGGATCATATCGTGCCCCAGTCTTTTCTCAAAGATGATTCTATTGATAATAAAGTGTTG
ACAAGATCCGATAAAAATAGAGGGAAGAGTGATAACGTCCCCTCAGAAGAAGTTGTCAAGAAAA
TGAAAAATTATTGGCGGCAGCTGCTGAACGCCAAACTGATCACACAACGGAAGTTCGATAATCT
GACTAAGGCTGAACGAGGTGGCCTGTCTGAGTTGGATAAAGCCGGCTTCATCAAAAGGCAGCTT
GTTGAGACACGCCAGATCACCAAGCACGTGGCCCAAATTCTCGATTCACGCATGAACACCAAGT
ACGATGAAAATGACAAACTGATTCGAGAGGTGAAAGTTATTACTCTGAAGTCTAAGCTGGTCTC
AGATTTCAGAAAGGACTTTCAGTTTTATAAGGTGAGAGAGATCAACAATTACCACCATGCGCAT

GATGCCTACCTGAATGCAGTGGTAGGCACTGCACTTATCAAAAAATATCCCAAGCTTGAATCTG
AATTTGTTTACGGAGACTATAAAGTGTACGATGTTAGGAAAATGATCGCAAAGTCTGAGCAGGA
AATAGGCAAGGCCACCGCTAAGTACTTCTTTTACAGCAATATTATGAATTTTTTCAAGACCGAG
ATTACACTGGCCAATGGAGAGATTCGGAAGCGACCACTTATCGAAACAAACGGAGAAACAGGAG
AAATCGTGTGGGACAAGGGTAGGGATTTCGCGACAGTCCGGAAGGTCCTGTCCATGCCGCAGGT
GAACATCGTTAAAAAGACCGAAGTACAGACCGGAGGCTTCTCCAAGGAAAGTATCCTCCCGAAA
AGGAACAGCGACAAGCTGATCGCACGCAAAAAAGATTGGGACCCCAAGAAATACGGCGGATTCG
ATTCTCCTACAGTCGCTTACAGTGTACTGGTTGTGGCCAAAGTGGAGAAAGGGAAGTCTAAAAA
ACTCAAAAGCGTCAAGGAACTGCTGGGCATCACAATCATGGAGCGATCAAGCTTCGAAAAAAAC
CCCATCGACTTTCTCGAGG

CGAAAGGATATAAAGAGGTCAAAAAAGACCTCATCATTAAGCTTCCCAAGTACTCTCTC
TTTGAGCTTGAAAACGGCCGGAAACGAATGCTCGCTAGTGCGGGCGAGCTGCAGAAAGGTAACG
AGCTGGCACTGCCCTCTAAATACGTTAATTTCTTGTATCTGGCCAGCCACTATGAAAAGCTCAA
AGGGTCTCCCGAAGATAATGAGCAGAAGCAGCTGTTCGTGGAACAACACAAACACTACCTTGAT
GAGATCATCGAGCAAATAAGCGAATTCTCCAAAAGAGTGATCCTCGCCGACGCTAACCTCGATA
AGGTGCTTTCTGCTTACAATAAGCACAGGGATAAGCCCATCAGGGAGCAGGCAGAAAACATTAT
CCACTTGTTTACTCTGACCAACTTGGGCGCGCCTGCAGCCTTCAAGTACTTCGACACCACCATA
GACAGAA

AGCGGTACACCTCTACAAAGGAGGTCCTGGACGCCACACTGATTCATCAGTCAATTACG
GGGCTCTATGAAACAAGAATCGACCTCTCTCAGCTCGGTGGAGACAGCAGGGCTGACCCCAAGA
AGAAGAGGAAGGTGAGGGGAGGCGGAGGATCTGGCGGAGGCGGAAGTGGCGGAGGGGGCAGCCC
ATCTCCTGCCAGACCTGCTCTGAGAGCCCCTGCCTCTGCCACAAGCGGCAGCAGAAAGAGAGCC
AGACCTCCTGCCGCCCCTGGCAGAGATCAGGCTAGACCTCCAGCTCGGCGGAGACTGAGACTGA
GCGTGGACGAGGTGTCCAGCCCTAGCACACCTGAGGCCCCTGATATCCCCGCCTGTCCTAGCCC
TGGCCAGAAGATCAAGAAGTCCACCCCTGCCGCCGGACAGCCTCCTCATCTGACATCTGCCCAG
GACCAGGACACCATC (SEQ ID NO: 24)

Cas9:Gem:

ATGGACAAGAAGTACTCCATTGGGCTCGATATCGGCACAAACAGCGTCGGCTGGGCCGT
CATTACGGACGAGTACAAGGTGCCGAGCAAAAAATTCAAAGTTCTGGGCAATACCGATCGCCAC
AGCATAAAGAAGAACCTCATTGGCGCCCTCCTGTTCGACTCCGGGGAGACGGCCGAAGCCACGC
GGCTCAAAAGAACAGCACGGCGCAGATATACCCGCAGAAAGAATCGGATCTGCTACCTGCAGGA
GATCTTTAGTAATGAGATGGCTAAGGTGGATGACTCTTTCTTCCATAGGCTGGAGGAGTCCTTT
TTGGTGGAGGAGGATAAAAAGCACGAGCGCCACCCAATCTTTGGCAATATCGTGGACGAGGTGG
CGTACCATGAAAAGTACCCAACCATATATCATCTGAGGAAGAAGCTTGTAGACAGTACTGATAA
GGCTGACTTGCGGTTGATCTATCTCGCGCTGGCGCATATGATCAAATTTCGGGGACACTTCCTC
ATCGAGGGGGACCTGAACCCAGACAACAGCGATGTCGACAAACTCTTTATCCAACTGGTTCAGA

CTTACAATCAGCTTTTCGAAGAGAACCCGATCAACGCATCCGGAGTTGACGCCAAAGCAATCCT
GAGCGCTAGGCTGTCCAAATCCCGGCGGCTCGAAAACCTCATCGCACAGCTCCCTGGGGAGAAG
AAGAACGGCCTGTTTGGTAATCTTATCGCCCTGTCACTCGGGCTGACCCCCAACTTTAAATCTA
ACTTCGACCTGGCCGAAGATGCCAAGCTTCAACTGAGCAAAGACACCTACGATGATGATCTCGA
CAATCTGCTGGCCCAGATCGGCGACCAGTACGCAGACCTTTTTTTGGCGGCAAAGAACCTGTCA
GACGCCATTCTGC

TGAGTGATATTCTGCGAGTGAACACGGAGATCACCAAAGCTCCGCTGAGCGCTAGTATG
ATCAAGCGCTATGATGAGCACCACCAAGACTTGACTTTGCTGAAGGCCCTTGTCAGACAGCAAC
TGCCTGAGAAGTACAAGGAAATTTTCTTCGATCAGTCTAAAAATGGCTACGCCGGATACATTGA
CGGCGGAGCAAGCCAGGAGGAATTTTACAAATTTATTAAGCCCATCTTGGAAAAAATGGACGGC
ACCGAGGAGCTGCTGGTAAAGCTTAACAGAGAAGATCTGTTGCGCAAACAGCGCACTTTCGACA
ATGGAAGCATCCCCCACCAGATTCACCTGGGCGAACTGCACGCTATCCTCAGGCGGCAAGAGGA
TTTCTACCCCTTTTTGAAAGATAACAGGGAAAAGATTGAGAAAATCCTCACATTTCGGATACCC
TACTATG

TAGGCCCCCTCGCCCGGGGAAATTCCAGATTCGCGTGGATGACTCGCAAATCAGAAGAG
ACCATCACTCCCTGGAACTTCGAGGAAGTCGTGGATAAGGGGGCCTCTGCCCAGTCCTTCATCG
AAAGGATGACTAACTTTGATAAAAATCTGCCTAACGAAAAGGTGCTTCCTAAACACTCTCTGCT
GTACGAGTACTTCACAGTTTATAACGAGCTCACCAAGGTCAAATACGTCACAGAAGGGATGAGA
AAGCCAGCATTCCTGTCTGGAGAGCAGAAGAAAGCTATCGTGGACCTCCTCTTCAAGACGAACC
GGAAAGTTACCGTGAAACAGCTCAAAGAAGACTATTTCAAAAAGATTGAATGTTTCGACTCTGT
TGAAATCAGCGGAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGTATCACGATCTCCTGAAA
ATCATTAAA

GACAAGGACTTCCTGGACAATGAGGAGAACGAGGACATTCTTGAGGACATTGTCCTCAC
CCTTACGTTGTTTGAAGATAGGGAGATGATTGAAGAACGCTTGAAAACTTACGCTCATCTCTTC
GACGACAAAGTCATGAAACAGCTCAAGAGGCGCCGATATACAGGATGGGGGCGGCTGTCAAGAA
AACTGATCAATGGGATCCGAGACAAGCAGAGTGGAAAGACAATCCTGGATTTTCTTAAGTCCGA
TGGATTTGCCAACCGGAACTTCATGCAGTTGATCCATGATGACTCTCTCACCTTTAAGGAGGAC
ATCCAGAAAGCACAAGTTTCTGGCCAGGGGGACAGTCTTCACGAGCACATCGCTAATCTTGCAG
GTAGCCCAGCTATCAAAAAGGGAATACTGCAGACCGTTAAGGTCGTGGATGAACTCGTCAAAGT
AATGGGA

AGGCATAAGCCCGAGAATATCGTTATCGAGATGGCCCGAGAGAACCAAACTACCCAGAA
GGGACAGAAGAACAGTAGGGAAAGGATGAAGAGGATTGAAGAGGGTATAAAAGAACTGGGGTCC
CAAATCCTTAAGGAACACCCAGTTGAAAACACCCAGCTTCAGAATGAGAAGCTCTACCTGTACT
ACCTGCAGAACGGCAGGGACATGTACGTGGATCAGGAACTGGACATCAATCGGCTCTCCGACTA
CGACGTGGATCATATCGTGCCCCAGTCTTTTCTCAAAGATGATTCTATTGATAATAAAGTGTTG
ACAAGATCCGATAAAAATAGAGGGAAGAGTGATAACGTCCCCTCAGAAGAAGTTGTCAAGAAAA

TGAAAAATTATTGGCGGCAGCTGCTGAACGCCAAACTGATCACACAACGGAAGTTCGATAATCT
GACTAAGGCTGAACGAGGTGGCCTGTCTGAGTTGGATAAAGCCGGCTTCATCAAAAGGCAGCTT
GTTGAGACACGCCAGATCACCAAGCACGTGGCCCAAATTCTCGATTCACGCATGAACACCAAGT
ACGATGAAAATGACAAACTGATTCGAGAGGTGAAAGTTATTACTCTGAAGTCTAAGCTGGTCTC
AGATTTCAGAAAGGACTTTCAGTTTTATAAGGTGAGAGAGATCAACAATTACCACCATGCGCAT
GATGCCTACCTGAATGCAGTGGTAGGCACTGCACTTATCAAAAAATATCCCAAGCTTGAATCTG
AATTTGTTTACGGAGACTATAAAGTGTACGATGTTAGGAAAATGATCGCAAAGTCTGAGCAGGA
AATAGGCAAGGCCACCGCTAAGTACTTCTTTTACAGCAATATTATGAATTTTTTCAAGACCGAG
ATTACACTGGCCAATGGAGAGATTCGGAAGCGACCACTTATCGAAACAAACGGAGAAACAGGAG
AAATCGTGTGGGACAAGGGTAGGGATTTCGCGACAGTCCGGAAGGTCCTGTCCATGCCGCAGGT
GAACATCGTTAAAAAGACCGAAGTACAGACCGGAGGCTTCTCCAAGGAAAGTATCCTCCCGAAA
AGGAACAGCGACAAGCTGATCGCACGCAAAAAAGATTGGGACCCCAAGAAATACGGCGGATTCG
ATTCTCCTACAGTCGCTTACAGTGTACTGGTTGTGGCCAAAGTGGAGAAAGGGAAGTCTAAAAA
ACTCAAAAGCGTCAAGGAACTGCTGGGCATCACAATCATGGAGCGATCAAGCTTCGAAAAAAAC
CCCATCGACTTTCTCGAGG

CGAAAGGATATAAAGAGGTCAAAAAAGACCTCATCATTAAGCTTCCCAAGTACTCTCTC
TTTGAGCTTGAAAACGGCCGGAAACGAATGCTCGCTAGTGCGGGCGAGCTGCAGAAAGGTAACG
AGCTGGCACTGCCCTCTAAATACGTTAATTTCTTGTATCTGGCCAGCCACTATGAAAAGCTCAA
AGGGTCTCCCGAAGATAATGAGCAGAAGCAGCTGTTCGTGGAACAACACAAACACTACCTTGAT
GAGATCATCGAGCAAATAAGCGAATTCTCCAAAAGAGTGATCCTCGCCGACGCTAACCTCGATA
AGGTGCTTTCTGCTTACAATAAGCACAGGGATAAGCCCATCAGGGAGCAGGCAGAAAACATTAT
CCACTTGTTTACTCTGACCAACTTGGGCGCGCCTGCAGCCTTCAAGTACTTCGACACCACCATA
GACAGAA

AGCGGTACACCTCTACAAAGGAGGTCCTGGACGCCACACTGATTCATCAGTCAATTACG
GGGCTCTATGAAACAAGAATCGACCTCTCTCAGCTCGGTGGAGACAGCAGGGCTGACCCCAAGA
AGAAGAGGAAGGTGAGGGGAGGCGGAGGATCTGGCGGAGGCGGAAGTGGCGGAGGGGGCAGCAT
GAACCCTAGCATGAAGCAGAAGCAGGAAGAGATCAAAGAGAACATCAAGAACAGCAGCGTGCCC
AGACGGACCCTGAAGATGATCCAGCCTAGCGCCAGCGGCAGCCTCGTGGGCAGAGAGAATGAAC
TGTCTGCCGGCCTGAGCAAGCGGAAGCACAGAAACGACCACCTGACCAGCACCACCAGCAGCCC
TGGCGTGATCGTGCCTGAGAGCAGCGAGAACAAGAACCTGGGCGGCGTGACCCAGGAATCCTTC
GACCTGATGATCAAAGAAAACCCCAGCAGCCAGTATTGGAAAGAGGTGGCCGAGAAGCGGCGGA
AGGCCCTG (SEQ ID NO: 25)

Cdt1:линкер:Cas9

ATGCCATCTCCTGCCAGACCTGCTCTGAGAGCCCCTGCCTCTGCCACAAGCGGCAGCAG
AAAGAGAGCCAGACCTCCTGCCGCCCCTGGCAGAGATCAGGCTAGACCTCCAGCTCGGCGGAGA
CTGAGACTGAGCGTGGACGAGGTGTCCAGCCCTAGCACACCTGAGGCCCCTGATATCCCCGCCT

GTCCTAGCCCTGGCCAGAAGATCAAGAAGTCCACCCCTGCCGCCGGACAGCCTCCTCATCTGAC
ATCTGCCCAGGACCAGGACACCATCGGAGGCGGAGGATCTGGCGGAGGCGGAAGTGGCGGAGGG
GGCAGCGACAAGAAGTACTCCATTGGGCTCGATATCGGCACAAACAGCGTCGGCTGGGCCGTCA
TTACGGACGAGTACAAGGTGCCGAGCAAAAAATTCAAAGTTCTGGGCAATACCGATCGCCACAG
CATAAA

GAAGAACCTCATTGGCGCCCTCCTGTTCGACTCCGGGGAGACGGCCGAAGCCACGCGGCTCAAA
AGAACAGCACGGCGCAGATATACCCGCAGAAAGAATCGGATCTGCTACCTGCAGGAGATCTTTA
GTAATGAGATGGCTAAGGTGGATGACTCTTTCTTCCATAGGCTGGAGGAGTCCTTTTTGGTGGA
GGAGGATAAAAAGCACGAGCGCCACCCAATCTTTGGCAATATCGTGGACGAGGTGGCGTACCAT
GAAAAGTACCCAACCATATATCATCTGAGGAAGAAGCTTGTAGACAGTACTGATAAGGCTGACT
TGCGGTTGATCTATCTCGCGCTGGCGCATATGATCAAATTTCGGGGACACTTCCTCATCGAGGG
GGACCTGAACCCAGACAACAGCGATGTCGACAAACTCTTTATCCAACTGGTTCAGACTTACAAT
CAGCTTTTCGAAGAGAACCCGATCAACGCATCCGGAGTTGACGCCAAAGCAATCCTGAGCGCTA
GGCTGTCCAAATCCCGGCGGCTCGAAAACCTCATCGCACAGCTCCCTGGGGAGAAGAAGAACGG
CCTGTTTGGTAATCTTATCGCCCTGTCACTCGGGCTGACCCCCAACTTTAAATCTAACTTCGAC
CTGGCCGAAGATGCCAAGCTTCAACTGAGCAAAGACACCTACGATGATGATCTCGACAATCTGC
TGGCCCAGATCGGCGACCAGTACGCAGACCTTTTTTTGGCGGCAAAGAACCTGTCAGACGCCAT
TCTGCTGAGTGATATTCTGCGAGTGAACACGGAGATCACCAAAGCTCCGCTGAGCGCTAGTATG
ATCAAGCGCTATGATGAGCACCACCAAGACTTGACTTTGCTGAAGGCCCTTGTCAGACAGCAAC
TGCCTGAGAAGTACAAGGAAATTTTCTTCGATCAGTCTAAAAATGGCTACGCCGGATACATTGA
CGGCGGAGCAAGCCAGGAGGAATTTTACAAATTTATTAAGCCCATCTTGGAAAAAATGGACGGC
ACCGAGGAGCTGCTGGTAAAGCTTAACAGAGAAGATCTGTTGCGCAAACAGCGCACTTTCGACA
ATGGAAGCATCCCCCACCAGATTCACCTGGGCGAACTGCACGCTATCCTCAGGCGGCAAGAGGA
TTTCTACCCCTTTTTGAAAGATAACAGGGAAAAGATTGAGAAAATCCTCACATTTCGGATACCC
TACTATGTAGGCCCCCTCGCCCGGGGAAATTCCAGATTCGCGTGGATGACTCGCAAATCAGAAG
AGACCATCACTCCCTGGAACTTCGAGGAAGTCGTGGATAAGGGGGCCTCTGCCCAGTCCTTCAT
CGAAAGGATGACTAACTTTGATAAAAATCTGCCTAACGAAAAGGTGCTTCCTAAACACTCTCTG
CTGTACGAGTACTTCACAGTTTATAACGAGCTCACCAAGGTCAAATACGTCACAGAAGGGATGA
GAAAGCCAGCATTCCTGTCTGGAGAGCAGAAGAAAGCTATCGTGGACCTCCTCTTCAAGACGAA
CCGGAAAGTTACCGTGAAACAGCTCAAAGAAGACTATTTCAAAAAGATTGAATGTTTCGACTCT
GTTGAAATCAGCGGAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGTATCACGATCTCCTGA
AAATCATTAAAGACAAGGACTTCCTGGACAATGAGGAGAACGAGGACATTCTTGAGGACATTGT
CCTCACCCTTACGTTGTTTGAAGATAGGGAGATGATTGAAGAACGCTTGAAAACTTACGCTCAT
CTCTTCGACGACAAAGTCATGAAACAGCTCAAGAGGCGCCGATATACAGGATGGGGGCGGCTGT
CAAGAAAACTGATCAATGGGATCCGAGACAAGCAGAGTGGAAAGACAATCCTGGATTTTCTTAA
GTCCGATGGATTTGCCAACCGGAACTTCATGCAGTTGATCCATGATGACTCTCTCACCTTTAAG

GAGGACATCCAGAAAGCACAAGTTTCTGGCCAGGGGGACAGTCTTCACGAGCACATCGCTAATC
TTGCAGGTAGCCCAGCTATCAAAAAGGGAATACTGCAGACCGTTAAGGTCGTGGATGAACTCGT
CAAAGTAATGGGAAGGCATAAGCCCGAGAATATCGTTATCGAGATGGCCCGAGAGAACCAAACT
ACCCAGAAGGGACAGAAGAACAGTAGGGAAAGGATGAAGAGGATTGAAGAGGGTATAAAAGAAC
TGGGGTCCCAAATCCTTAAGGAACACCCAGTTGAAAACACCCAGCTTCAGAATGAGAAGCTCTA
CCTGTACTACCTGCAGAACGGCAGGGACATGTACGTGGATCAGGAACTGGACATCAATCGGCTC
TCCGACTACGACGTGGATCATATCGTGCCCCAGTCTTTTCTCAAAGATGATTCTATTGATAATA
AAGTGTTGACAAGATCCGATAAAAATAGAGGGAAGAGTGATAACGTCCCCTCAGAAGAAGTTGT
CAAGAAAATGAAAAATTATTGGCGGCAGCTGCTGAACGCCAAACTGATCACACAACGGAAGTTC
GATAATCTGACTAAGGCTGAACGAGGTGGCCTGTCTGAGTTGGATAAAGCCGGCTTCATCAAAA
GGCAGCTTGTTGAGACACGCCAGATCACCAAGCACGTGGCCCAAATTCTCGATTCACGCATGAA
CACCAAGTACGATGAAAATGACAAACTGATTCGAGAGGTGAAAGTTATTACTCTGAAGTCTAAG
CTGGTCTCAGATTTCAGAAAGGACTTTCAGTTTTATAAGGTGAGAGAGATCAACAATTACCACC
ATGCGCATGATGCCTACCTGAATGCAGTGGTAGGCACTGCACTTATCAAAAAATATCCCAAGCT
TGAATCTGAATTTGTTTACGGAGACTATAAAGTGTACGATGTTAGGAAAATGATCGCAAAGTCT
GAGCAGGAAATAGGCAAGGCCACCGCTAAGTACTTCTTTTACAGCAATATTATGAATTTTTTCA
AGACCGAGATTACACTGGCCAATGGAGAGATTCGGAAGCGACCACTTATCGAAACAAACGGAGA
AACAGGAGAAATCGTGTGGGACAAGGGTAG

GGATTTCGCGACAGTCCGGAAGGTCCTGTCCATGCCGCAGGTGAACATCGTTAAAAAGA
CCGAAGTACAGACCGGAGGCTTCTCCAAGGAAAGTATCCTCCCGAAAAGGAACAGCGACAAGCT
GATCGCACGCAAAAAAGATTGGGACCCCAAGAAATACGGCGGATTCGATTCTCCTACAGTCGCT
TACAGTGTACTGGTTGTGGCCAAAGTGGAGAAAGGGAAGTCTAAAAAACTCAAAAGCGTCAAGG
AACTGCTGGGCATCACAATCATGGAGCGATCAAGCTTCGAAAAAAACCCCATCGACTTTCTCGA
GGCGAAAGGATATAAAGAGGTCAAAAAAGACCTCATCATTAAGCTTCCCAAGTACTCTCTCTTT
GAGCTTGAAAACGGCCGGAAACGAATGCTCGCTAGTGCGGGCGAGCTGCAGAAAGGTAACGAGC
TGGCACTGCCCTCTAAATACGTTAATTTCTTGTATCTGGCCAGCCACTATGAAAAGCTCAAAGG
GTCTCCCGAAGATAATGAGCAGAAGCAGCTGTTCGTGGAACAACACAAACACTACCTTGATGAG
ATCATCGAGCAAATAAGCGAATTCTCCAAAAGAGTGATCCTCGCCGACGCTAACCTCGATAAGG
TGCTTTCTGCTTACAATAAGCACAGGGATAAGCCCATCAGGGAGCAGGCAGAAAACATTATCCA
CTTGTTTACTCTGACCAACTTGGGCGCGCCTGCAGCCTTCAAGTACTTCGACACCACCATAGAC
AGAAAGCGGTACACCTCTACAAAGGAGGTCCTGGACGCCACACTGATTCATCAGTCAATTACGG
GGCTCTATGAAACAAGAATCGACCTCTCTCAGCTCGGTGGAGACAGCAGGGCTGACCCCAAGAA
GAAGAGGAAGGTGAGG (SEQ ID NO: 26)

Gem:линкер:Cas9

ATGAACCCTAGCATGAAGCAGAAGCAGGAAGAGATCAAAGAGAACATCAAGAACAGCAG
CGTGCCCAGACGGACCCTGAAGATGATCCAGCCTAGCGCCAGCGGCAGCCTCGTGGGCAGAGAG

AATGAACTGTCTGCCGGCCTGAGCAAGCGGAAGCACAGAAACGACCACCTGACCAGCACCACCA
GCAGCCCTGGCGTGATCGTGCCTGAGAGCAGCGAGAACAAGAACCTGGGCGGCGTGACCCAGGA
ATCCTTCGACCTGATGATCAAAGAAAACCCCAGCAGCCAGTATTGGAAAGAGGTGGCCGAGAAG
CGGCGGAAGGCCCTGGGAGGCGGAGGATCTGGCGGAGGCGGAAGTGGCGGAGGGGGCAGCGACA
AGAAGTACTCCATTGGGCTCGATATCGGCACAAACAGCGTCGGCTGGGCCGTCATTACGGACGA
G

TACAAGGTGCCGAGCAAAAAATTCAAAGTTCTGGGCAATACCGATCGCCACAGCATAAA
GAAGAACCTCATTGGCGCCCTCCTGTTCGACTCCGGGGAGACGGCCGAAGCCACGCGGCTCAAA
AGAACAGCACGGCGCAGATATACCCGCAGAAAGAATCGGATCTGCTACCTGCAGGAGATCTTTA
GTAATGAGATGGCTAAGGTGGATGACTCTTTCTTCCATAGGCTGGAGGAGTCCTTTTTGGTGGA
GGAGGATAAAAAGCACGAGCGCCACCCAATCTTTGGCAATATCGTGGACGAGGTGGCGTACCAT
GAAAAGTACCCAACCATATATCATCTGAGGAAGAAGCTTGTAGACAGTACTGATAAGGCTGACT
TGCGGTTGATCTATCTCGCGCTGGCGCATATGATCAAATTTCGGGGACACTTCCTCATCGAGGG
GGACCTGAACCCAGACAACAGCGATGTCGACAAACTCTTTATCCAACTGGTTCAGACTTACAAT
CAGCTTTTCGAAGAGAACCCGATCAACGCATCCGGAGTTGACGCCAAAGCAATCCTGAGCGCTA
GGCTGTCCAAATCCCGGCGGCTCGAAAACCTCATCGCACAGCTCCCTGGGGAGAAGAAGAACGG
CCTGTTTGGTAATCTTATCGCCCTGTCACTCGGGCTGACCCCCAACTTTAAATCTAACTTCGAC
CTGGCCGAAGATGCCAAGCTTCAACTGAGCAAAGACACCTACGATGATGATCTCGACAATCTGC
TGGCCCAGATCGGCGACCAGTACGCAGACCTTTTTTTGGCGGCAAAGAACCTGTCAGACGCCAT
TCTGCTGAGTGATATTCTGCGAGTGAACACGGAGATCACCAAAGCTCCGCTGAGCGCTAGTATG
ATCAAGCGCTATGATGAGCACCACCAAGACTTGACTTTGCTGAAGGCCCTTGTCAGACAGCAAC
TGCCTGAGAAGTACAAGGAAATTTTCTTCGATCAGTCTAAAAATGGCTACGCCGGATACATTGA
CGGCGGAGCAAGCCAGGAGGAATTTTACAAATTTATTAAGCCCATCTTGGAAAAAATGGACGGC
ACCGAGGAGCTGCTGGTAAAGCTTAACAGAGAAGATCTGTTGCGCAAACAGCGCACTTTCGACA
ATGGAAGCATCCCCCACCAGATTCACCTGGGCGAACTGCACGCTATCCTCAGGCGGCAAGAGGA
TTTCTACCCCTTTTTGAAAGATAACAGGGAAAAGATTGAGAAAATCCTCACATTTCGGATACCC
TACTATGTAGGCCCCCTCGCCCGGGGAAATTCCAGATTCGCGTGGATGACTCGCAAATCAGAAG
AGACCATCACTCCCTGGAACTTCGAGGAAGTCGTGGATAAGGGGGCCTCTGCCCAGTCCTTCAT
CGAAAGGATGACTAACTTTGATAAAAATCTGCCTAACGAAAAGGTGCTTCCTAAACACTCTCTG
CTGTACGAGTACTTCACAGTTTATAACGAGCTCACCAAGGTCAAATACGTCACAGAAGGGATGA
GAAAGCCAGCATTCCTGTCTGGAGAGCAGAAGAAAGCTATCGTGGACCTCCTCTTCAAGACGAA
CCGGAAAGTTACCGTGAAACAGCTCAAAGAAGACTATTTCAAAAAGATTGAATGTTTCGACTCT
GTTGAAATCAGCGGAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGTATCACGATCTCCTGA
AAATCATTAAAGACAAGGACTTCCTGGACAATGAGGAGAACGAGGACATTCTTGAGGACATTGT
CCTCACCCTTACGTTGTTTGAAGATAGGGAGATGATTGAAGAACGCTTGAAAACTTACGCTCAT
CTCTTCGACGACAAAGTCATGAAACAGCTCAAGAGGCGCCGATATACAGGATGGGGGCGGCTGT

CAAGAAAACTGATCAATGGGATCCGAGACAAGCAGAGTGGAAAGACAATCCTGGATTTTCTTAA
GTCCGATGGATTTGCCAACCGGAACTTCATGCAGTTGATCCATGATGACTCTCTCACCTTTAAG
GAGGACATCCAGAAAGCACAAGTTTCTGGCCAGGGGGACAGTCTTCACGAGCACATCGCTAATC
TTGCAGGTAGCCCAGCTATCAAAAAGGGAATACTGCAGACCGTTAAGGTCGTGGATGAACTCGT
CAAAGTAATGGGAAGGCATAAGCCCGAGAATATCGTTATCGAGATGGCCCGAGAGAACCAAACT
ACCCAGAAGGGACAGAAGAACAGTAGGGAAAGGATGAAGAGGATTGAAGAGGGTATAAAAGAAC
TGGGGTCCCAAATCCTTAAGGAACACCCAGTTGAAAACACCCAGCTTCAGAATGAGAAGCTCTA
CCTGTACTACCTGCAGAACGGCAGGGACATGTACGTGGATCAGGAACTGGACATCAATCGGCTC
TCCGACTACGACGTGGATCATATCGTGCCCCAGTCTTTTCTCAAAGATGATTCTATTGATAATA
AAGTGTTGACAAGATCCGATAAAAATAGAGGGAAGAGTGATAACGTCCCCTCAGAAGAAGTTGT
CAAGAAAATGAAAAATTATTGGCGGCAGCTGCTGAACGCCAAACTGATCACACAACGGAAGTTC
GATAATCTGACTAAGGCTGAACGAGGTGGCCTGTCTGAGTTGGATAAAGCCGGCTTCATCAAAA
GGCAGCTTGTTGAGACACGCCAGATCACCAAGCACGTGGCCCAAATTCTCGATTCACGCATGAA
CACCAAGTACGATGAAAATGACAAACTGATTCGAGAGGTGAAAGTTATTACTCTGAAGTCTAAG
CTGGTCTCAGATTTCAGAAAGGACTTTCAGTTTTATAAGGTGAGAGAGATCAACAATTACCACC
ATGCGCATGATGCCTACCTGAATGCAGTGGTAGGCACTGCACTTATCAAAAAATATCCCAAGCT
TGAATCTGAATTTGTTTACGGAGACTATAAAGTGTACGATGTTAGGAAAATGATCGCAAAGTCT
GAGCAGGAAATAGGCAAGGCCACCGCTAAGTACTTCTTTTACAGCAATATTATGAATTTTTTCA
AGACCGAGATTACACTGGCCAATGGAGAGATTCGGA

AGCGACCACTTATCGAAACAAACGGAGAAACAGGAGAAATCGTGTGGGACAAGGGTAGG
GATTTCGCGACAGTCCGGAAGGTCCTGTCCATGCCGCAGGTGAACATCGTTAAAAAGACCGAAG
TACAGACCGGAGGCTTCTCCAAGGAAAGTATCCTCCCGAAAAGGAACAGCGACAAGCTGATCGC
ACGCAAAAAAGATTGGGACCCCAAGAAATACGGCGGATTCGATTCTCCTACAGTCGCTTACAGT
GTACTGGTTGTGGCCAAAGTGGAGAAAGGGAAGTCTAAAAAACTCAAAAGCGTCAAGGAACTGC
TGGGCATCACAATCATGGAGCGATCAAGCTTCGAAAAAAACCCCATCGACTTTCTCGAGGCGAA
AGGATATAAAGAGGTCAAAAAAGACCTCATCATTAAGCTTCCCAAGTACTCTCTCTTTGAGCTT
GAAAACGGCCGGAAACGAATGCTCGCTAGTGCGGGCGAGCTGCAGAAAGGTAACGAGCTGGCAC
TGCCCTCTAAATACGTTAATTTCTTGTATCTGGCCAGCCACTATGAAAAGCTCAAAGGGTCTCC
CGAAGATAATGAGCAGAAGCAGCTGTTCGTGGAACAACACAAACACTACCTTGATGAGATCATC
GAGCAAATAAGCGAATTCTCCAAAAGAGTGATCCTCGCCGACGCTAACCTCGATAAGGTGCTTT
CTGCTTACAATAAGCACAGGGATAAGCCCATCAGGGAGCAGGCAGAAAACATTATCCACTTGTT
TACTCTGACCAACTTGGGCGCGCCTGCAGCCTTCAAGTACTTCGACACCACCATAGACAGAAAG
CGGTACACCTCTACAAAGGAGGTCCTGGACGCCACACTGATTCATCAGTCAATTACGGGGCTCT
ATGAAACAAGAATCGACCTCTCTCAGCTCGGTGGAGACAGCAGGGCTGACCCCAAGAAGAAGAG
GAAGGTGAGG (SEQ ID NO: 27)

ПРИМЕР 4

Ауторегуляция РНК-направляемых нуклеаз путем встраивания частичных участков-мишеней

Снижение активности РНК-направляемых нуклеаз может быть полезным для снижения потенциальных неспецифических эффектов. В клинических условиях это может иметь высокое значение в случае вирусных способов доставки, например, посредством AAV, которые характеризуются длительной или пожизненной экспрессией.

В данном примере описана саморегулирующаяся система для экспрессии Cas9, РНК-направленной нуклеазы, которая основана на применении неполных участков-мишеней, закодированных в самой системе экспрессии Cas9/гРНК. В многочисленных исследованиях было показано, что связывание Cas9 (без расщепления) является эффективным в отношении регуляции экспрессии генов. В большинстве исследований использовалась версия Cas9 без нуклеазной активности для предотвращения активности расщепления, однако является в высокой степени желательной система, которая позволяет как расщепление, так и ауторегуляцию.

С использованием сконструированных последовательностей, которые соответствуют частичным участкам мишени, активность Cas9 в отношении расщепления и связывания может быть разделена; частичная комплементарность последовательностей позволяет связывание Cas9 без расщепления ДНК. С использованием этой склонности к связыванию Cas9 может быть направлена обратно на регуляцию ее собственной экспрессии. Для SpCas9, участки, комплементарность которых обычно составляет менее 17 нуклеотидов, не приводят к расщеплению, и даже одно несоответствие может приводить к отсутствию расщепления с высокоточных или высокоспецифичных мутантов Cas9.

Важно, что вся эта регуляция происходит с минимальными изменениями последовательности и без дополнительных связывающих факторов, слияний или других белков, что позволяет оставить размер экспрессирующей кассеты достаточно малым, чтобы доставлять ее посредством AAV.

Способы

Конструирование плазмиды: Репортерные плазмиды с GFP конструировались Gibson Assembly (New England Biolabs, Ipswich,

MA) (Gibson *et al.* (2009) *Nature Methods* 6:343-345) с небольшими модификациями.

Линию клеток почки эмбриона человека (HEK) 293T (Life Technologies, Grand Island, NY) поддерживали при 37°C с 5% CO_2/20% O_2 в модифицированной способом Дульбекко среде Игла (DMEM) (Invitrogen), дополненной 10% эмбриональной телячьей сывороткой (инактивированной нагреванием) (Gibco, Life Technologies, Grand Island, NY) и 2 мМ GlutaMAX (Invitrogen).

Линию клеток N2A (Life Technologies, Grand Island, NY) поддерживали при 37°C с 5% CO_2/20% O_2 в модифицированной способом Дульбекко среде Игла (DMEM) (Invitrogen), дополненной 10% эмбриональной телячьей сывороткой (Gibco, Life Technologies, Grand Island, NY) и 2 мМ GlutaMAX (Invitrogen).

Клетки трансфицировали (Lipofectamine 3000) каждой конструкцией. Экспрессию GFP анализировали с использованием проточной цитометрии.

последовательность-мишень mm079: GAAGAAGGTTCGAGATCTCA (SEQ ID NO: 28)

геномная последовательность-мишень mm079: GAAGAAGGTTCGAGATCTCAAGG (SEQ ID NO: 29)

Участок Auto1: GTTCGAGATCTCAGGGAAT (SEQ ID NO: 30)

Участок Auto2: GTTCGAGATCTCAGGGTTT (SEQ ID NO: 31)

H1_wt:

GGAATTCGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCCCAGTGTCACTAGGCGG
GAACACCCAGCGCGCGTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAGGGGAGTGGCGCCCT
GCAATATTTGCATGTCGCTATGTGTTCTGGGAAATCACCATAAACGTGAAATGTCTTTGGATTT
GGGAATCTTATAAGTTCTGTATGAGACCA CTTTTTCCC (SEQ ID NO: 32)

H1_Auto1:

GGAATTCGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCCCAGTGTCACTAGGCGG
GAACACCCAGCGCGCGTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAGGGGAGTGGCGCCCT
GCAATATTTGCATGTCGCTATGTGTTCTGGGAAATCACCATAAACGTGAAAGTTCGAGATCTCA
GGGAATCTTATAAGTTCTGTATGAGACC ACTTTTTCCC (SEQ ID NO: 33)

H1_Auto2:

GGAATTCGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCCCAGTGTCACTAGGCGG
GAACACCCAGCGCGCGTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAGGGGAGTGGCGCCCT

GCAATATTTGCATGTCGCTATGTGTTCTGGGAAATCACCATAAACGTGAAATGTCTTTGGATTT GGGAATCTTATAAGTGTTCGAGATCTCAG GGTTTTCCC (SEQ ID NO: 34)

ПРИМЕР 5

Метионин (Met), за которым следует глицин (Gly), пролин (Pro), серин (Ser), треонин (Thr), аланин (Ala), валин (Val) или цистеин (Cys), процессируются метионинаминопептидазами, которые отщепляют N-концевой met (фиг.1A). Далее, тип аминокислоты в положении 2 либо стабилизирует, либо дестабилизирует весь белок. Таким образом, изменяя тип второй аминокислоты можно изменять время полужизни некоторых белков с помощью Met-аминопептидаз.

Несмотря на отличия от организма к организму и вариабельность между белками, правило N-конца позволяет оценить время полужизни белков в клетке (фиг.17B). Уровни белка модулируются протеолитической системой убиквитина (Ub). Убиквитин является в высокой степени консервативным белком из 76 аминокислот, который обычно ассоциирован с деградацией белков; молекулы Ub конъюгируются с белками-мишенями, маркируя эти белки для разрушения протеасомой.

Линейный полиубиквитин распознается клеткой ферментами деубиквитинилирования, которые отщепляют индивидуальные пептиды Ub. Этот процесс может быть использован для получения определенных N-концевых остатков путем слияния Ub с N-концом любого белка. (фиг.18). После попадания в клетку деубиквитинилирующие ферменты распознают и отщепляют часть Ub, высвобождая слитый белок. Этот процесс может быть использован для получения точных N-концевых аминокислотных остатков на данном белке, тип которых определяет время полужизни белков.

Как правило, контроль экспрессии либо РНК-направляемой нуклеазы, либо гидовой РНК может модулировать весь голоферментный комплекс; для системы CRISPR-Cas9 это может быть осуществлено посредством регуляции уровней либо Cas9, либо гРНК. Более того, простые изменения N-концевых аминокислот, посредством либо Met-аминопептидаз, либо деубиквитинилирования, или альтернативными способами, могут осуществлять эту регуляцию (фиг.19A).

Снижение времени полужизни Cas9 может быть полезным

посредством снижения потенциальных неспецифических эффектов. В клинических условиях это имеет высокое значение в случае вирусных способов доставки, например, посредством AAV, которые характеризуются длительной или пожизненной экспрессией. Напротив, в случае неэффективных способов доставки может быть желательным или необходимым увеличение времени полужизни белка для обеспечения эффекта. В различных условиях способность регулировать уровни нуклеазы является в высокой степени желательной. Кроме того, этот подход к регуляции является простым и может быть использован совместно с другими формами регуляции, такими как дегроны, индуцибельные дегроны или разделенные варианты РНК-направляемых нуклеаз.

В данном примере было показано, что уровни Cas9, РНК-направляемой нуклеазы, могут модулироваться типом N-концевой аминокислоты вплоть до 8-10 раз, что является значительным диапазоном для экспрессии нуклеазы (фиг.19A). Также показано, как широко используемые пептиды 2A также могут быть использованы для модулирования уровней белка в клетке (фиг.20).

Способы

Конструирование плазмиды: для получения слитых с убиквитином белков Cas9 проводили слияние убиквитина с N-концом Cas9 с использованием Gibson Assembly (New England Biolabs, Ipswich, MA) (Gibson *et al*. (2009) *Nature Methods* 6:343-345) с небольшими модификациями. В кратком изложении, праймеры, кодирующие все 20 аминокислот, использовали для амплификации gBlock, кодирующего убиквитин, который затем клонировали в рамке считывания с Cas9. Все конструкции подтверждали секвенированием.

Линию клеток почки эмбриона человека (HEK) 293T (Life Technologies, Grand Island, NY) поддерживали при 37°C с 5% CO_2/20% O_2 в модифицированной способом Дульбекко среде Игла (DMEM) (Invitrogen), дополненной 10% эмбриональной телячьей сывороткой (инактивированной нагреванием) (Gibco, Life Technologies, Grand Island, NY) и 2 мМ GlutaMAX (Invitrogen).

Клетки трансфицировали каждой конструкцией в количестве 2000 нг, 500 нг или 125 нг. Через 48 ч клетки собирали для

получения белка, а затем анализировали на устройстве Simple Simon для экспрессии Cas9 или GAPDH в качестве контроля. Для нормализации величин и для определения стабильности белка использовали соотношения.

1. Tasaki T *et al.* (2012) *Annu Rev Biochem.* 81:261-89.

2. Varshavsky A *et al.* (2011) *Protein Sci.* 20(8):1298-345.

3. Bachmair A *et al.* (1986) *Science.* 234(4773):179-86.

Таблица 1: Приблизительное время полужизни белков в клетках млекопитающих, исходя из N-концевого остатка

Val	100 ч
Met	30 ч
Gly	30 ч
Pro	20 ч
Ile	20 ч
Thr	7,2 ч
Leu	5,5 ч
Ala	4,4 ч
His	3,5 ч
Trp	2,8 ч
Tyr	2,8 ч
Ser	1,9 ч
Asn	1,4 ч
Lys	1,3 ч
Cys	1,2 ч
Asp	1,1 ч
Phe	1,1 ч
Glu	1,0 ч
Arg	1,0 ч
Gln	0,8 ч

Таблица 2: Перечень обычно используемых последовательностей 2A

P2A: GSGATNFSLLKQAGDVEENPGP (SEQ ID NO: 35)

T2A: GSGEGRGSLLTCGDVEENPGP (SEQ ID NO: 36)

E2A: GSGQCTNYALLKLAGDVESNPGP (SEQ ID NO: 37)

F2A: GSGVKQTLNFDLLKLAGDVESNPGP (SEQ ID NO: 38)

Белковые последовательности

Ub:Val::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDY
NIQKESTLHLVLRLRGGVDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL
IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK
KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGA
SQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM
TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV
TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG
FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM
GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY
LQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS
RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKY
PKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKD
LIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAP
AAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 39)

Ub:Met::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDY
NIQKESTLHLVLRLRGGMDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL
IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK
KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGA
SQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM

TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV
TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG
FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM
GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY
LQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKM
KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKY
DENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESE
FVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE
IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD
SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLP
KYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHK
HYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ ID NO:
40)

Ub:Gly::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDY
NIQKESTLHLVLRLRGGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL
IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK
KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGA
SQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM
TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV
TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG
FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM
GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY
LQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS
RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKY
PKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKD

LIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAP
AAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV(SEQ
ID NO: 41)

Ub:Pro::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDY
NIQKESTLHLVLRLRGGPDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL
IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK
KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGA
SQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM
TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV
TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG
FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM
GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY
LQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS
RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKY
PKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKD
LIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAP
AAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 42)

Ub:Ile::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDY
NIQKESTLHLVLRLRGGIDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL
IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK
KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS

DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGA
SQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM
TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV
TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG
FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM
GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY
LQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS
RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKY
PKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKD
LIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAP
AAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 43)

Ub:Thr::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDY
NIQKESTLHLVLRLRGGTDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL
IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK
KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGA
SQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM
TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV
TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG
FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM
GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY
LQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS
RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKY

PKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKD
LIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAP
AAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 44)

Ub:Leu::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDY
NIQKESTLHLVLRLRGGLDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL
IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK
KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGA
SQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM
TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV
TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG
FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM
GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY
LQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS
RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKY
PKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKD
LIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAP
AAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 45)

Ub:Ala::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDY
NIQKESTLHLVLRLRGGADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL
IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK

KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGA
SQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM
TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV
TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG
FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM
GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY
LQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS
RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKY
PKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKD
LIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAP
AAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 46)

Ub:His::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDY
NIQKESTLHLVLRLRGGHDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL
IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK
KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGA
SQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM
TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV
TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG
FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM
GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY

LQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK

MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS

RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKY

PKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET

NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK

KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKD

LIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL

FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAP

AAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ

ID NO: 47)

Ub:Trp::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDY

NIQKESTLHLVLRLRGGWDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL

IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK

KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN

PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG

NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS

DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGA

SQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP

FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM

TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV

TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL

TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG

FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM

GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY

LQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK

MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS

RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKY

PKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET

NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK

KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKD

LIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL

FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAP

AAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ

ID NO: 48)

Ub:Try::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDY
NIQKESTLHLVLRLRGGYDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL
IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK
KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGA
SQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM
TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV
TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG
FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM
GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY
LQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS
RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKY
PKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKD
LIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAP
AAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 49)

Ub:Ser::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDY
NIQKESTLHLVLRLRGGSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL
IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK
KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGA
SQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM
TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV
TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL

TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG
FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM
GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY
LQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS
RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKY
PKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKD
LIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAP
AAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 50)

Ub:Asn::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDY
NIQKESTLHLVLRLRGGNDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL
IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK
KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGA
SQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM
TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV
TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG
FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM
GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY
LQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS
RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKY
PKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKD
LIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAP

AAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ ID NO: 51)

Ub:Lys::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDY
NIQKESTLHLVLRLRGGKDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL
IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK
KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGA
SQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM
TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV
TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG
FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM
GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY
LQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS
RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKY
PKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKD
LIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAP
AAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ ID NO: 52)

Ub:Cys::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDY
NIQKESTLHLVLRLRGGCDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL
IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK
KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGA
SQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP

FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM
TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV
TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG
FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM
GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY
LQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS
RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKY
PKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKD
LIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAP
AAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 53)

Ub:Asp::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDY
NIQKESTLHLVLRLRGGDDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL
IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK
KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGA
SQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM
TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV
TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG
FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM
GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY
LQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS
RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKY
PKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK

KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKD
LIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAP
AAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 54)

Ub:Phe::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDY
NIQKESTLHLVLRLRGGFDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL
IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK
KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGA
SQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM
TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV
TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG
FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM
GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY
LQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS
RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKY
PKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKD
LIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAP
AAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 55)

Ub:Glu::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDY
NIQKESTLHLVLRLRGGEDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL
IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK
KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG

NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGA
SQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM
TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV
TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG
FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM
GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY
LQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS
RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKY
PKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKD
LIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAP
AAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 56)

Ub:Arg::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDY
NIQKESTLHLVLRLRGGRDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL
IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK
KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGA
SQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM
TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV
TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG
FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM
GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY
LQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS

RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKY
PKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKD
LIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAP
AAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 57)

Ub:Gln::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDY
NIQKESTLHLVLRLRGGQDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL
IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK
KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGA
SQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM
TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV
TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG
FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM
GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY
LQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS
RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKY
PKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKD
LIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAP
AAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 58)

Ub(G76V)::Cas9

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDY
NIQKESTLHLVLRLRGVDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLI

GALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP
DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGN
LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSD
ILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGAS
QEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPF
LKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT
VKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLT
LFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGF
ANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
QNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS
RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKY
PKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKD
LIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAP
AAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKV (SEQ
ID NO: 59)

Последовательности ДНК

Ub:

ACCACCTCTCAGACGCAGGACCAGGTGCAGGGTCGACTCCTTCTGGATGTTGTAGTCAG
AAAGAGTGCGGCCATCTTCCAGCTGCTTGCCTGCAAAGATGAGCCTCTGCTGGTCGGGAGGGAT
GCCTTCTTTATCCTGGATCTTGGCCTTCACATTTTCGATGGTGTCACTGGGCTCCACTTCCAGG
GTGATGGTCTTGCCGGTCAGGGTCTTCACGAAGATCTGCAT (SEQ ID NO: 60)

Cas9:

GACAAGAAGTACTCCATTGGGCTCGATATCGGCACAAACAGCGTCGGCTGGGCCGTCAT
TACGGACGAGTACAAGGTGCCGAGCAAAAAATTCAAAGTTCTGGGCAATACCGATCGCCACAGC
ATAAAGAAGAACCTCATTGGCGCCCTCCTGTTCGACTCCGGGGAGACGGCCGAAGCCACGCGGC
TCAAAAGAACAGCACGGCGCAGATATACCCGCAGAAAGAATCGGATCTGCTACCTGCAGGAGAT
CTTTAGTAATGAGATGGCTAAGGTGGATGACTCTTTCTTCCATAGGCTGGAGGAGTCCTTTTTG
GTGGAGGAGGATAAAAAGCACGAGCGCCACCCAATCTTTGGCAATATCGTGGACGAGGTGGCGT
ACCATGAAAAGTACCCAACCATATATCATCTGAGGAAGAAGCTTGTAGACAGTACTGATAAGGC

TGACTTGCGGTTGATCTATCTCGCGCTGGCGCATATGATCAAATTTCGGGGACACTTCCTCATC
GAGGGGGACCTGAACCCAGACAACAGCGATGTCGACAAACTCTTTATCCAACTGGTTCAGACTT
ACAATCAGCTTTTCGAAGAGAACCCGATCAACGCATCCGGAGTTGACGCCAAAGCAATCCTGAG
CGCTAGGCTGTCCAAATCCCGGCGGCTCGAAAACCTCATCGCACAGCTCCCTGGGGAGAAGAAG
AACGGCCTGTTTGGTAATCTTATCGCCCTGTCACTCGGGCTGACCCCCAACTTTAAATCTAACT
TCGACCTGGCCGAAGATGCCAAGCTTCAACTGAGCAAAGACACCTACGATGATGATCTCGACAA
TCTGCTGGCCCAGATCGGCGACCAGTACGCAGACCTTTTTTTGGCGGCAAAGAACCTGTCAGAC
GCCATTCTGCTGAGTGATATTCTGCGAGTGAACACGGAGATCACCAAAGCTCCGCTGAGCGCTA
GTATGATCAAGCGCTATGATGAGCACCACCAAGACTTGACTTTGCTGAAGGCCCTTGTCAGACA
GCAACTGCCTGAGAAGTACAAGGAAATTTTCTTCGATCAGTCTAAAAATGGCTACGCCGGATAC
ATTGACGGCGGAGCAAGCCAGGAGGAATTTTACAAATTTATTAAGCCCATCTTGGAAAAAATGG
ACGGCACCGAGGAGCTGCTGGTAAAGCTTAACAGAGAAGATCTGTTGCGCAAACAGCGCACTTT
CGACAATGGAAGCATCCCCCACCAGATTCACCTGGGCGAACTGCACGCTATCCTCAGGCGGCAA
GAGGATTTCTACCCCTTTTTGAAAGATAACAGGGAAAAGATTGAGAAAATCCTCACATTTCGGA
TACCCTACTATGTAGGCCCCCTCGCCCGGGGAAATTCCAGATTCGCGTGGATGACTCGCAAATC
AGAAGAGACCATCACTCCCTGGAACTTCGAGGAAGTCGTGGATAAGGGGGCCTCTGCCCAGTCC
TTCATCGAAAGGATGACTAACTTTGATAAAAATCTGCCTAACGAAAAGGTGCTTCCTAAACACT
CTCTGCTGTACGAGTACTTCACAGTTTATAACGAGCTCACCAAGGTCAAATACGTCACAGAAGG
GATGAGAAAGCCAGCATTCCTGTCTGGAGAGCAGAAGAAAGCTATCGTGGACCTCCTCTTCAAG
ACGAACCGGAAAGTTACCGTGAAACAGCTCAAAGAAGACTATTTCAAAAAGATTGAATGTTTCG
ACTCTGTTGAAATCAGCGGAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGTATCACGATCT
CCTGAAAATCATTAAAGACAAGGACTTCCTGGACAATGAGGAGAACGAGGACATTCTTGAGGAC
ATTGTCCTCACCCTTACGTTGTTTGAAGATAGGGAGATGATTGAAGAACGCTTGAAAACTTACG
CTCATCTCTTCGACGACAAAGTCATGAAACAGCTCAAGAGGCGCCGATATACAGGATGGGGGCG
GCTGTCAAGAAAACTGATCAATGGGATCCGAGACAAGCAGAGTGGAAAGACAATCCTGGATTTT
CTTAAGTCCGATGGATTTGCCAACCGGAACTTCATGCAGTTGATCCATGATGACTCTCTCACCT
TTAAGGAGGACATCCAGAAAGCACAAGTTTCTGGCCAGGGGGACAGTCTTCACGAGCACATCGC
TAATCTTGCAGGTAGCCCAGCTATCAAAAAGGGAATACTGCAGACCGTTAAGGTCGTGGATGAA
CTCGTCAAAGTAATGGGAAGGCATAAGCCCGAGAATATCGTTATCGAGATGGCCCGAGAGAACC
AAACTACCCAGAAGGGACAGAAGAACAGTAGGGAAAGGATGAAGAGGATTGAAGAGGGTATAAA
AGAACTGGGGTCCCAAATCCTTAAGGAACACCCAGTTGAAAACACCCAGCTTCAGAATGAGAAG
CTCTACCTGTACTACCTGCAGAACGGCAGGGACATGTACGTGGATCAGGAACTGGACATCAATC
GGCTCTCCGACTACGACGTGGATCATATCGTGCCCCAGTCTTTTCTCAAAGATGATTCTATTGA
TAATAAAGTGTTGACAAGATCCGATAAAAATAGAGGGAAGAGTGATAACGTCCCCTCAGAAGAA
GTTGTCAAGAAAATGAAAAATTATTGGCGGCAGCTGCTGAACGCCAAACTGATCACACAACGGA
AGTTCGATAATCTGACTAAGGCTGAACGAGGTGGCCTGTCTGAGTTGGATAAAGCCGGCTTCAT

CAAAAGGCAGCTTGTTGAGACACGCCAGATCACCAAGCACGTGGCCCAAATTCTCGATTCACGC
ATGAACACCAAGTACGATGAAAATGACAAACTGATTCGAGAGGTGAAAGTTATTACTCTGAAGT
CTAAGCTGGTCTCAGATTTCAGAAAGGACTTTCAGTTTTATAAGGTGAGAGAGATCAACAATTA
CCACCATGCGCATGATGCCTACCTGAATGCAGTGGTAGGCACTGCACTTATCAAAAAATATCCC
AAGCTTGAATCTGAATTTGTTTACGGAGACTATAAAGTGTACGATGTTAGGAAAATGATCGCAA
AGTCTGAGCAGGAAATAGGCAAGGCCACCGCTAAGTACTTCTTTTACAGCAATATTATGAATTT
TTTCAAGACCGAGATTACACTGGCCAATGGAGAGATTCGGAAGCGACCACTTATCGAAACAAAC
GGAGAAACAGGAGAAATCGTGTGGGACAAGGGTAGGGATTTCGCGACAGTCCGGAAGGTCCTGT
CCATGCCGCAGGTGAACATCGTTAAAAAGACCGAAGTACAGACCGGAGGCTTCTCCAAGGAAAG
TATCCTCCCGAAAAGGAACAGCGACAAGCTGATCGCACGCAAAAAAGATTGGGACCCCAAGAAA
TACGGCGGATTCGATTCTCCTACAGTCGCTTACAGTGTACTGGTTGTGGCCAAAGTGGAGAAAG
GGAAGTCTAAAAAACTCAAAAGCGTCAAGGAACTGCTGGGCATCACAATCATGGAGCGATCAAG
CTTCGAAAAAAACCCCATCGACTTTCTCGAGGCGAAAGGATATAAAGAGGTCAAAAAAGACCTC
ATCATTAAGCTTCCCAAGTACTCTCTCTTTGAGCTTGAAAACGGCCGGAAACGAATGCTCGCTA
GTGCGGGCGAGCTGCAGAAAGGTAACGAGCTGGCACTGCCCTCTAAATACGTTAATTTCTTGTA
TCTGGCCAGCCACTATGAAAAGCTCAAAGGGTCTCCCGAAGATAATGAGCAGAAGCAGCTGTTC
GTGGAACAACACAAACACTACCTTGATGAGATCATCGAGCAAATAAGCGAATTCTCCAAAAGAG
TGATCCTCGCCGACGCTAACCTCGATAAGGTGCTTTCTGCTTACAATAAGCACAGGGATAAGCC
CATCAGGGAGCAGGCAGAAAACATTATCCACTTGTTTACTCTGACCAACTTGGGCGCGCCTGCA
GCCTTCAAGTACTTCGACACCACCATAGACAGAAAGCGGTACACCTCTACAAAGGAGGTCCTGG
ACGCCACACTGATTCATCAGTCAATTACGGGGCTCTATGAAACAAGAATCGACCTCTCTCAGCT
CGGTGGAGACAGCAGGGCTGACCCCAAGAAGAAGAGGAAGGTG (SEQ ID NO: 61).

ПРИМЕР 6

Применение двунаправленного промотора для доставки РНК-направляемой нуклеазы, гидовой РНК и матриц HDR посредством одного вируса AAV

Доставка РНК-направляемой нуклеазы, гидовой РНК и матриц HDR посредством одного вируса AAV. Революционная технология редактирования генома, известная как CRISPR, изменила биологические исследования и привела к новой эре в генетической медицине. Каждая клетка в организме человека содержит 3 миллиарда пар оснований ДНК и каждое единственное изменение или мутация может вызвать широкое множество наследственных или приобретенных заболеваний. С использованием процесса, напоминающего хирургическую операцию на геноме, технология CRISPR позволяет исследователям осуществлять нацеливание на

ассоциированные с заболеванием мутации с беспрецедентной точностью, осуществляя лечение их коренной причины.

Исследовательские лаборатории по всему миру показали, что CRISPR может быть эффективно нацелена на обуславливающие заболевания мутации; однако до настоящего времени разработке терапевтических средств на основе CRISPR для применения у человека препятствовали ограничения, связанные с доставкой. Предпочтительной для выбора доставкой для тканеспецифической генной терапии *in vivo* являются рекомбинантные аденоассоциированные вирусы (AAV). AAV имеет примечательную историю безопасности, эффективности и отсутствия токсичности, что далее иллюстрируется возрастающим количеством одобренных FDA испытаний с использованием этого подхода - по всему миру проводится 150 испытаний с использованием AAV. Эти компактные вирусы могут доставлять материалы с высокой эффективностью и специфичностью к широкому множеств типов клеток и, хотя они и являются безопасными, их небольшой размер является значительным препятствием для упаковывания компонентов CRISPR.

Доминантные заболевания возникают, когда человек наследует один неблагоприятный ген, который затем "отравляет" нормальную копию. Хотя они являются неизлечимыми с использованием технологии генной терапии, эти мутации могут быть подавлены с использованием CRISPR, которая разрушает мутацию и позволяет нормальному гену функционировать. Ранее авторы настоящего изобретения описали способ упаковывания обоих компонентов CRISPR в один вирус AAV с использованием компактного двунаправленного промотора H1.

Рецессивные мутации, с другой стороны, являются результатом наследования двух неблагоприятных копий гена. Эти заболевания требуют доставки CRISPR, а также длинного участка ДНК, окружающей мутации (матрицы) чтобы клетка могла отредактировать мутации путем копирования правильной последовательности. Хотя значительное большинство заболеваний относятся к этой категории, не существуют способов доставки компонентов CRISPR и матричной ДНК посредством одного вируса AAV вследствие ограничений AAV по размеру.

Исторически, AAV служил в качестве способа доставки матриц для сайт-специфического изменения ДНК. По сравнению с другими известными способами, матрицы AAV являются наиболее рекомбиногенными, хотя эти уровни все еще были низкими и составляли порядка 1 из 104 до появления способов редактирования генов. Известно, что разрывы ДНК являются в высокой степени рекомбиногенными и что совместная доставка CRISPR-Cas9 с матрицей AAV способна к рекомбинации с высокой частотой. Также известно, что оцДНК-матрицы размером всего 40-60 нуклеотидов являются очень эффективными матрицами HDR, как и дцДНК-матрицы размером 200 нуклеотидов и больше. К сожалению, отсутствуют способы доставки компонентов CRISPR и матрицы HDR в одном вирусе AAV.

Однако поскольку двунаправленная промоторная система H1 является настолько компактной, можно доставлять как расщепляющие (Cas9 и гРНК), так и осуществляющие вставку (матрица HDR) элементы с использованием одного вируса AAV. Терапевтический потенциал для этого класса заболеваний является значительно большим с точки зрения как количества заболеваний, так и количества человек, которых можно лечить.

В данном примере показана конструкция вектора AAV для коррекции мутаций *in vivo* с использованием промотора H1 для экспрессии РНК-направляемой нуклеазы (Cas9), гидовой РНК и донорной матричной последовательности. Эти два примера коррекции мутаций иллюстрируются, например, 1) с использованием никазы SaCas9 с двумя гРНК и матрицей, корректирующей мутацию rd12 (фиг.25B) и 2) коррекцией посредством SpCas9 мутации rd10 (фиг.27).

Способы

Линию клеток эмбриона почки человека (HEK) 293T (Life Technologies, Grand Island, NY) поддерживали при 37°C с 5% CO_2/20% O_2 в модифицированной способом Дульбекко среде Игла (DMEM) (Invitrogen), дополненной 10% эмбриональной телячьей сывороткой (Gibco, Life Technologies, Grand Island, NY) и 2 мМ GlutaMAX (Invitrogen).

Схема конструкции. См.фиг.28. Конструкцию получали с использованием промотора H1 мыши для экспрессии как SpCas9, так и гРНК. Два участка рестрикции включали перед фланкирующими участками NotI, чтобы обеспечить клонирование Gibson HDR-матрицы: KpnI для левой HDR, и HpaI для правой HDR. Для встраивания после расщепления KpnI использовали следующую последовательность 5′-GAGAGTGCACCATAGCGGCCGCGNNNGTACCCACACAAAAAACCAACACACAG-3′ (SEQ ID NO: 69), где N обозначает матричную область, или после расщепления HpaI использовали следующую последовательность 5′-GGCACCGAGTCGGTGCTTTTTTGTTNNNNNNAACGCGGCCGCCTAGAGTCGAC-3′ (SEQ ID NO: 70), где N обозначает матричную область.

гРНК (см. Jaskula-Ranga, V., & Zack, D. J. (2016). grID: A CRISPR-Cas9 guide RNA Database and Resource for Genome-Editing. *bioRxiv*, 097352) получали с помощью перекрывающихся олигонуклеотидов, которые подвергали отжигу и амплификации способом ПЦР с использованием ДНК-полимеразы Phusion Flash для двухэтапной амплификации (Thermo Fisher Scientific, Rockford, IL), а затем очищали с использованием Zymo DNA clean и концентрирующих колонок. Затем очищенные продукты ПЦР ресуспендировали в H2O и количественно определяли с использованием NanoDrop 1000 (Thermo Fisher Scientific). Экспрессирующие гРНК конструкции получали с использованием Gibson Assembly (New England Biolabs, Ipswich, MA) (Gibson *et al*. *Nature Methods* 6:343-345 (2009)) с небольшими модификациями. Общий объем реакции уменьшался от 20 мкл до 2 мкл. Клоны подтверждали секвенированием по методу Сэнгера.

Клетки HEK293 совместно трансфицировали посредством Cas9 (немодифицированная или регулируемая клеточным циклом версия) и конструкций гРНК, нацеливающей на родопсин. Через 48 ч после трансфекции геномную ДНК собирали и последовательность, окружающую участки разрезания мишени, амплифицировали в соответствии с праймерами, приведенными ниже. Затем продукты ПЦР очищали и количественно определяли перед проведением анализа с T7 Endo I. В кратком изложении, 200 нг продукта ПЦР

денатурировали, а затем медленно обратно гибридизовали для образования гетеродуплексов, к продуктам ПЦР добавляли эндонуклеазу I T7 и инкубировали при 37C в течение 25 минут для расщепления гетеродуплексов, реакцию гасили нагрузочным красителем, и, наконец, реакционную смесь разделяли на 6% геле TBE PAGE для разделения продуктов. Гел окрашивали SYBR-Gold, визуализировали и подвергали количественному определению с использованием ImageJ. Частоты NHEJ вычисляли с использованием двучленного уравнения:

$$1-\sqrt{1-\frac{(a+b)}{(a+b+c)}}\ \mathbf{X}\ 100$$

где величины "a" и "b" равны интегрированной площади для расщепленных фрагментов после вычитания фонового сигнала и "c" равен интегрированной площади для нерасщепленного продукта ПЦР после вычитания фонового сигнала.

Последовательности

Исходная плазмида mSPA_HDR SpCas9

CCTGCTGTCTCCACCGAGCTGAGAGAGGTCGATTCTTGTTTCATAGAGCCCCGTAATTG
ACTGATGAATCAGTGTGGCGTCCAGGACCTCCTTTGTAGAGGTGTACCGCTTTCTGTCTATGGT
GGTGTCGAAGTACTTGAAGGCTGCAGGCGCGCCCAAGTTGGTCAGAGTAAACAAGTGGATAATG
TTTTCTGCCTGCTCCCTGATGGGCTTATCCCTGTGCTTATTGTAAGCAGAAAGCACCTTATCGA
GGTTAGCGTCGGCGAGGATCACTCTTTTGGAGAATTCGCTTATTTGCTCGATGATCTCATCAAG
GTAGTGTTTGTGTTGTTCCACGAACAGCTGCTTCTGCTCATTATCTTCGGGAGACCCTTTGAGC
TTTTCATAGTGGCTGGCCAGATACAAGAAATTAACGTATTTAGAGGGCAGTGCCAGCTCGTTAC
CTTTCTGCAGCTCGCCCGCACTAGCGAGCATTCGTTTCCGGCCGTTTTCAAGCTCAAAGAGAGA
GTACTTGGGAAGCTTAATGATGAGGTCTTTTTTGACCTCTTTATATCCTTTCGCCTCGAGAAAG
TCGATGGGGTTTTTTTCGAAGCTTGATCGCTCCATGATTGTGATGCCCAGCAGTTCCTTGACGC
TTTTGAGTTTTTTAGACTTCCCTTTCTCCACTTTGGCCACAACCAGTACACTGTAAGCGACTGT
AGGAGAATCGAATCCGCCGTATTTCTTGGGGTCCCAATCTTTTTTGCGTGCGATCAGCTTGTCG
CTGTTCCTTTTCGGGAGGATACTTTCCTTGGAGAAGCCTCCGGTCTGTACTTCGGTCTTTTTAA
CGATGTTCACCTGCGGCATGGACAGGACCTTCCGGACTGTCGCGAAATCCCTACCCTTGTCCCA
CACGATTTCTCCTGTTTCTCCGTTTGTTTCGATAAGTGGTCGCTTCCGAATCTCTCCATTGGCC
AGTGTAATCTCGGTCTTGAAAAAATTCATAATATTGCTGTAAAAGAAGTACTTAGCGGTGGCCT
TGCCTATTTCCTGCTCAGACTTTGCGATCATTTTCCTAACATCGTACACTTTATAGTCTCCGTA
AACAAATTCAGATTCAAGCTTGGGATATTTTTTGATAAGTGCAGTGCCTACCACTGCATTCAGG
TAGGCATCATGCGCATGGTGGTAATTGTTGATCTCTCTCACCTTATAAAACTGAAAGTCCTTTC

TGAAATCTGAGACCAGCTTAGACTTCAGAGTAATAACTTTCACCTCTCGAATCAGTTTGTCATT
TTCATCGTACTTGGTGTTCATGCGTGAATCGAGAATTTGGGCCACGTGCTTGGTGATCTGGCGT
GTCTCAACAAGCTGCCTTTTGATGAAGCCGGCTTTATCCAACTCAGACAGGCCACCTCGTTCAG
CCTTAGTCAGATTATCGAACTTCCGTTGTGTGATCAGTTTGGCGTTCAGCAGCTGCCGCCAATA
ATTTTTCATTTTCTTGACAACTTCTTCTGAGGGGACGTTATCACTCTTCCCTCTATTTTTATCG
GATCTTGTCAACACTTTATTATCAATAGAATCATCTTTGAGAAAAGACTGGGGCACGATATGAT
CCACGTCGTAGTCGGAGAGCCGATTGATGTCCAGTTCCTGATCCACGTACATGTCCCTGCCGTT
CTGCAGGTAGTACAGGTAGAGCTTCTCATTCTGAAGCTGGGTGTTTTCAACTGGGTGTTCCTTA
AGGATTTGGGACCCCAGTTCTTTTATACCCTCTTCAATCCTCTTCATCCTTTCCCTACTGTTCT
TCTGTCCCTTCTGGGTAGTTTGGTTCTCTCGGGCCATCTCGATAACGATATTCTCGGGCTTATG
CCTTCCCATTACTTTGACGAGTTCATCCACGACCTTAACGGTCTGCAGTATTCCCTTTTTGATA
GCTGGGCTACCTGCAAGATTAGCGATGTGCTCGTGAAGACTGTCCCCCTGGCCAGAAACTTGTG
CTTTCTGGATGTCCTCCTTAAAGGTGAGAGAGTCATCATGGATCAACTGCATGAAGTTCCGGTT
GGCAAATCCATCGGACTTAAGAAAATCCAGGATTGTCTTTCCACTCTGCTTGTCTCGGATCCCA
TTGATCAGTTTTCTTGACAGCCGCCCCCATCCTGTATATCGGCGCCTCTTGAGCTGTTTCATGA
CTTTGTCGTCGAAGAGATGAGCGTAAGTTTTCAAGCGTTCTTCAATCATCTCCCTATCTTCAAA
CAACGTAAGGGTGAGGACAATGTCCTCAAGAATGTCCTCGTTCTCCTCATTGTCCAGGAAGTCC
TTGTCTTTAATGATTTTCAGGAGATCGTGATACGTTCCCAGGGATGCGTTGAAGCGATCCTCCA
CTCCGCTGATTTCAACAGAGTCGAAACATTCAATCTTTTTGAAATAGTCTTCTTTGAGCTGTTT
CACGGTAACTTTCCGGTTCGTCTTGAAGAGGAGGTCCACGATAGCTTTCTTCTGCTCTCCAGAC
AGGAATGCTGGCTTTCTCATCCCTTCTGTGACGTATTTGACCTTGGTGAGCTCGTTATAAACTG
TGAAGTACTCGTACAGCAGAGAGTGTTTAGGAAGCACCTTTTCGTTAGGCAGATTTTTATCAAA
GTTAGTCATCCTTTCGATGAAGGACTGGGCAGAGGCCCCCTTATCCACGACTTCCTCGAAGTTC
CAGGGAGTGATGGTCTCTTCTGATTTGCGAGTCATCCACGCGAATCTGGAATTTCCCCGGGCGA
GGGGGCCTACATAGTAGGGTATCCGAAATGTGAGGATTTTCTCAATCTTTTCCCTGTTATCTTT
CAAAAAGGGGTAGAAATCCTCTTGCCGCCTGAGGATAGCGTGCAGTTCGCCCAGGTGAATCTGG
TGGGGGATGCTTCCATTGTCGAAAGTGCGCTGTTTGCGCAACAGATCTTCTCTGTTAAGCTTTA
CCAGCAGCTCCTCGGTGCCGTCCATTTTTTCCAAGATGGGCTTAATAAATTTGTAAAATTCCTC
CTGGCTTGCTCCGCCGTCAATGTATCCGGCGTAGCCATTTTTAGACTGATCGAAGAAAATTTCC
TTGTACTTCTCAGGCAGTTGCTGTCTGACAAGGGCCTTCAGCAAAGTCAAGTCTTGGTGGTGCT
CATCATAGCGCTTGATCATACTAGCGCTCAGCGGAGCTTTGGTGATCTCCGTGTTCACTCGCAG
AATATCACTCAGCAGAATGGCGTCTGACAGGTTCTTTGCCGCCAAAAAAAGGTCTGCGTACTGG
TCGCCGATCTGGGCCAGCAGATTGTCGAGATCATCATCGTAGGTGTCTTTGCTCAGTTGAAGCT
TGGCATCTTCGGCCAGGTCGAAGTTAGATTTAAAGTTGGGGGTCAGCCCGAGTGACAGGGCGAT
AAGATTACCAAACAGGCCGTTCTTCTTCTCCCCAGGGAGCTGTGCGATGAGGTTTTCGAGCCGC
CGGGATTTGGACAGCCTAGCGCTCAGGATTGCTTTGGCGTCAACTCCGGATGCGTTGATCGGGT

TCTCTTCGAAAAGCTGATTGTAAGTCTGAACCAGTTGGATAAAGAGTTTGTCGACATCGCTGTT
GTCTGGGTTCAGGTCCCCCTCGATGAGGAAGTGTCCCCGAAATTTGATCATATGCGCCAGCGCG
AGATAGATCAACCGCAAGTCAGCCTTATCAGTACTGTCTACAAGCTTCTTCCTCAGATGATATA
TGGTTGGGT

ACTTTTCATGGTACGCCACCTCGTCCACGATATTGCCAAAGATTGGGTGGCGCTCGTGC
TTTTTATCCTCCTCCACCAAAAAGGACTCCTCCAGCCTATGGAAGAAAGAGTCATCCACCTTAG
CCATCTCATTACTAAAGATCTCCTGCAGGTAGCAGATCCGATTCTTTCTGCGGGTATATCTGCG
CCGTGCTGTTCTTTTGAGCCGCGTGGCTTCGGCCGTCTCCCCGGAGTCGAACAGGAGGGCGCCA
ATGAGGTTCTTCTTTATGCTGTGGCGATCGGTATTGCCCAGAACTTTGAATTTTTTGCTCGGCA
CCTTGTACTCGTCCGTAATGACGGCCCAGCCGACGCTGTTTGTGCCGATATCGAGCCCAATGGA
GTACTTCTTGTCCATGGTGGCGGCTCTTGAAGGACGACGTCATCATCCCTTGCCCGGATGCGCG
GGCTTCTTGTCTAGCACAGGAGCCTGGGGTAGAGCGCATGCAAATTACGCGCTGTGCTTTGTGG
GAAATCACCCTAAACGAAAAATTTATTCCTCTTTCGAGCCTTATAGTGGCGGCCGGTCTACATC
CTAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGT
GGCACCGAGTCGGTGCTTTTTTGTTAACGCGGCCGCCTAGAGTCGACCTGCAGGCATGCAAGCT
TGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAA
CATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTA
ATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAA
TCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGA
CTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGG
TTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCA
GGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCA
CAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTT
CCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCG
CCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGT
GTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCC
TTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAG
CCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTG
GCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACC
TTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTT
TTGTTTGCAA

GCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCT
GACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCT
TCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAAC
TTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGT
TCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTG

GCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAA
CCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCT
ATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTG
CCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTC
CCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGT
CCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGC
ATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAA
GTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAAT
ACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAAC
TCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATC
TTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAG

GCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCT
TCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGA
ATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGAC
GTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTC
GTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCAC
AGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGC
GGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATAGCG
GCCGCGGTACCCACACAAAAAACCAACACACAGATGTAATGAAAATAAAGATATTTTATT

TCACACCTTCCTCTTCTTCTTGGGGTCAGC (SEQ ID NO: 62)

mSPA_HDR rd10(250) правая

CTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTT
ATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTA
ATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTG
TCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGC

TCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGT
ATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAAC
ATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCC
ATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAAC

CCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCC
TGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTT
TCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTG
TGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAA
CCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGG
TATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAG
TATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATC

CGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGA
AAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAA
ACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAA
TTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAA
TGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGAC
TCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGAT
ACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCC
GAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAG
CTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGC

TACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCC
AACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCC
TCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCAT
AATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGT
CA

TTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAA
TACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAA
CTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGAT
CTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGC

AAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAAT
ATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAA
AAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACC
ATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTT
TCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTA
AGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGC
TGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATAGCGGCCGCGGTACCC
ACACAAAAAACCAACACACAGATGTAATGAAAATAAAGATATTTTATTTCACACCTTCCTCTTC
TTCTTGGGGTCAGCCCTGCTGTCTCCACCGAGCTGAGAGAGGTCGATTCTTGTTTCATAGAGCC
CCGTAATTGACTGATGAATCAGTGTGGCGTCCAGGACCTCCTTTGTAGAGGTGTACCGCTTTCT
GTCTATGGTGGTGTCGAAGTACTTGAAGGCTGCAGGCGCGCCCAAGTTGGTCAGAGTAAACAAG
TGGATAATGTTTTCTGCCTGCTCCCTGATGGGCTTATCCCTGTGCTTATTGTAAGCAGAAAGCA
CCTTATCGAGGTTAGCGTCGGCGAGGATCACTCTTTTGGAGAATTCGCTTATTTGCTCGATGAT
CTCATCAAGGTAGTGTTTGTGTTGTTCCACGAACAGCTGCTTCTGCTCATTATCTTCGGGAGAC
CCTTTGAGCTTTTCATAGTGGCTGGCCAGATACAAGAAATTAACGTATTTAGAGGGCAGTGCCA
GCTCGTTACCTTTCTGCAGCTCGCCCGCACTAGCGAGCATTCGTTTCCGGCCGTTTTCAAGCTC
AAAGAGAGAGTACTTGGGAAGCTTAATGATGAGGTCTTTTTTGACCTCTTTATATCCTTTCGCC
TCGAGAAAGTCGATGGGGTTTTTTTCGAAGCTTGATCGCTCCATGATTGTGATGCCCAGCAGTT

CCTTGACGCTTTTGAGTTTTTTAGACTTCCCTTTCTCCACTTTGGCCACAACCAGTACACTGTA
AGCGACTGTAGGAGAATCGAATCCGCCGTATTTCTTGGGGTCCCAATCTTTTT

TGCGTGCGATCAGCTTGTCGCTGTTCCTTTTCGGGAGGATACTTTCCTTGGAGAAGCCT
CCGGTCTGTACTTCGGTCTTTTTAACGATGTTCACCTGCGGCATGGACAGGACCTTCCGGACTG
TCGCGAAATCCCTACCCTTGTCCCACACGATTTCTCCTGTTTCTCCGTTTGTTTCGATAAGTGG
TCGCTTCCGAATCTCTCCATTGGCCAGTGTAATCTCGGTCTTGAAAAAATTCATAATATTGCTG
TAA

AAGAAGTACTTAGCGGTGGCCTTGCCTATTTCCTGCTCAGACTTTGCGATCATTTTCCT
AACATCGTACACTTTATAGTCTCCGTAAACAAATTCAGATTCAAGCTTGGGATATTTTTTGATA
AGTGCAGTGCCTACCACTGCATTCAGGTAGGCATCATGCGCATGGTGGTAATTGTTGATCTCTC
TCACCTTATAAAACTGAAAGTCCTTTCTGAAATCTGAGACCAGCTTAGACTTCAGAGTAATAAC
TTTCACCTCTCGAATCAGTTTGTCATTTTCATCGTACTTGGTGTTCATGCGTGAATCGAGAATT
TGGGCCACGTGCTTGGTGATCTGGCGTGTCTCAACAAGCTGCCTTTTGATGAAGCCGGCTTTAT
CCAACTCAGACAGGCCACCTCGTTCAGCCTTAGTCAGATTATCGAACTTCCGTTGTGTGATCAG
TTTGGCGTTCAGCAGCTGCCGCCAATAATTTTTCATTTTCTTGACAACTTCTTCTGAGGGGACG
TTATCACTCTTCCCTCTATTTTTATCGGATCTTGTCAACACTTTATTATCAATAGAATCATCTT
TGAGAAAAGACTGGGGCACGATATGATCCACGTCGTAGTCGGAGAGCCGATTGATGTCCAGTTC
CTGATCCACGTACATGTCCCTGCCGTTCTGCAGGTAGTACAGGTAGAGCTTCTCATTCTGAAGC
TGGGTGTTTTCAACTGGGTGTTCCTTAAGGATTTGGGACCCCAGTTCTTTTATACCCTCTTCAA
TCCTCTTCATCCTTTCCCTACTGTTCTTCTGTCCCTTCTGGGTAGTTTGGTTCTCTCGGGCCAT
CTCGATAACGATATTCTCGGGCTTATGCCTTCCCATTACTTTGACGAGTTCATCCACGACCTTA
ACGGTCTGCAGTATTCCCTTTTTGATAGCTGGGCTACCTGCAAGATTAGCGATGTGCTCGTGAA
GACTGTCCCCCTGGCCAGAAACTTGTGCTTTCTGGATGTCCTCCTTAAAGGTGAGAGAGTCATC
ATGGATCAACTGCATGAAGTTCCGGTTGGCAAATCCATCGGACTTAAGAAAATCCAGGATTGTC
TTTCCACTCTGCTTGTCTCGGATCCCATTGATCAGTTTTCTTGACAGCCGCCCCCATCCTGTAT
ATCGGCGCCTCTTGAGCTGTTTCATGACTTTGTCGTCGAAGAGATGAGCGTAAGTTTTCAAGCG
TTCTTCAATCATCTCCCTATCTTCAAACAACGTAAGGGTGAGGACAATGTCCTCAAGAATGTCC
TCGTTCTCCTCATTGTCCAGGAAGTCCTTGTCTTTAATGATTTTCAGGAGATCGTGATACGTTC
CCAGGGATGCGTTGAAGCGATCCTCCACTCCGCTGATTTCAACAGAGTCGAAACATTCAATCTT
TTTGAAATAGTCTTCTTTGAGCTGTTTCACGGTAACTTTCCGGTTCGTCTTGAAGAGGAGGTCC
ACGATAGCTTTCTTCTGCTCTCCAGACAGGAATGCTGGCTTTCTCATCCCTTCTGTGACGTATT
TGACCTTGGTGAGCTCGTTATAAACTGTGAAGTACTCGTACAGCAGAGAGTGTTTAGGAAGCAC
CTTTTCGTTAGGCAGATTTTTATCAAAGTTAGTCATCCTTTCGATGAAGGACTGGGCAGAGGCC
CCCTTATCCACGACTTCCTCGAAGTTCCAGGGAGTGATGGTCTCTTCTGATTTGCGAGTCATCC
ACGCGAATCTGGAATTTCCCCGGGCGAGGGGGCCTACATAGTAGGGTATCCGAAATGTGAGGAT
TTTCTCAATCTTTTCCCTGTTATCTTTCAAAAAGGGGTAGAAATCCTCTTGCCGCCTGAGGATA

GCGTGCAGTTCGCCCAGGTGAATCTGGTGGGGGATGCTTCCATTGTCGAAAGTGCGCTGTTTGC
GCAACAGATCTTCTCTGTTAAGCTTTACCAGCAGCTCCTCGGTGCCGTCCATTTTTTCCAAGAT
GGGCTTAATAAATTTGTAAAATTCCTCCTGGCTTGCTCCGCCGTCAATGTATCCGGCGTAGCCA
TTTTTAGACTGATCGAAGAAAATTTCCTTGTACTTCTCAGGCAGTTGCTGTCTGACAAGGGCCT
TCAGCAAAGTCAAGTCTTGGTGGTGCTCATCATAGCGCTTGATCATACTAGCGCTCAGCGGAGC
TTTGGTGATCTCCGTGTTCACTCGCAGAATATCACTCAGCAGAATGGCGTCTGACAGGTTCTTT
GCCGCCAAAAAAAGGTCTGCGTACTGGTCGCCGATCTGGGCCAGCAGATTGTCGAGATCATCAT
CGTAGGTGTCTTTGCTCAGTTGAAGCTTGGCATCTTCGGCCAGGTCGAAGTTAGATTTAAAGTT
GGGGGTCAGCCCGAGTGACAGGGCGATAAGATTACCAAACAGGCCGTTCTTCTTCTCCCCAGGG
AGCTGTGCGATGAGGTTTTCGAGCCGCCGGGATTTGGACAGCCTAGCGCTCAGGATTGCTTTGG
CGTCAACTCCGGATGCGTTGATCGGGTTCTCTTCGAAAAGCTGATTGTAAGTCTGAACCAGTTG
GATAAAGAGTTTGTCGACATCGCTGTTGTCTGGGTTCAGGTCCCCCTCGATGAGGAAGTGTCCC
CGAAATTTGATCATATGCGCCAGCGCGAGATAGATCAACCGCAAGTCAGCCTTATCAGTACTGT
CTACAAGCTTCTTCCTCAGATGATATATGGTTGGGTACTTTTCATGGTACGCCACCTCGTCCAC
GATATTGCCAAAGATTGGGTGGCGCTCGTGCTTTTTATCCTCCTCCACCAAAAAGGACTCCTCC
AGCCTATGGAAGAAAGAGTCATCCACCTTAGCCATCTCATTACTAAAGATCTCCTGCAGGTAGC
AGATCCGATTCTTTCTGCGGGTATATCTGCGCCGTGCTGTTCTTTTGAGCCGCGTGGCTTCGGC
CGTCTCCCCGGAGTCGAACAGGAGGGCGCCAATGAGGTTCTTCTTTATGCTGTGGCGATCGGTA
TTGCCCAGAACTTTGAATTTTTTGCTCGGCACCTTGTACTCGTCCGTAATGACGGCCCAGCCGA
CGCTGTTTGTGCCGATATCGAGCCCAATGGAGTACTTCTTGTCCATGGTGGCGGCTCTTGAAGG
ACGACGTCATCATCCCTTGCCCGGATGCGCGGGCTTCTTGTCTAGCACAGGAGCCTGGGGTAGA
GCGCATGC

AAATTACGCGCTGTGCTTTGTGGGAAATCACCCTAAACGAAAAATTTATTCCTCTTTCG
AGCCTTATAGTGGCGGCCGGTCTACATCCTAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAA
GGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTGTTGATATGGTGCT
GTGTAGGCTCATATGTGGATCTCAGAACCCACATGTACTCTGCTCCCCAGGTCTTGGTGCGCTT
TCTATTCTCTGTCAGCAAAGCCTATCGAAGAATCACCTACCACAACTGGCGCCACGGCTTCAAT
GTAGCCCAGACCATGTTTACCCTACTCATGGTACGTATGTAAATTGGATGGGCTAGATGAATCA
GAGGGCTGGGGCAAGGACCACAGCTAACTATCTTCTGGCCCAAGGAACGCGGCCGCCTAGAGTC
GAC (SEQ ID NO: 63)

mSPA_HDR rd10(250) левая

ATGTAATGAAAATAAAGATATTTTATTTCACACCTTCCTCTTCTTCTTGGGGTCAGCCC
TGCTGTCTCCACCGAGCTGAGAGAGGTCGATTCTTGTTTCATAGAGCCCCGTAATTGACTGATG
AATCAGTGTGGCGTCCAGGACCTCCTTTGTAGAGGTGTACCGCTTTCTGTCTATGGTGGTGTCG
AAGTACTTGAAGGCTGCAGGCGCGCCCAAGTTGGTCAGAGTAAACAAGTGGATAATGTTTTCTG
CCTGCTCCCTGATGGGCTTATCCCTGTGCTTATTGTAAGCAGAAAGCACCTTATCGAGGTTAGC

GTCGGCGAGGATCACTCTTTTGGAGAATTCGCTTATTTGCTCGATGATCTCATCAAGGTAGTGT
TTGTGTTGTTCCACGAACAGCTGCTTCTGCTCATTATCTTCGGGAGACCCTTTGAGCTTTTCAT
AGTGGCTGGCCAGATACAAGAAATTAACGTATTTAGAGGGCAGTGCCAGCTCGTTACCTTTCTG
CAGCTCGCCCGCACTAGCGAGCATTCGTTTCCGGCCGTTTTCAAGCTCAAAGAGAGAGTACTTG
GGAAGCTTAATGATGAGGTCTTTTTTGACCTCTTTATATCCTTTCGCCTCGAGAAAGTCGATGG
GGTTTTTTTCGAAGCTTGATCGCTCCATGATTGTGATGCCCAGCAGTTCCTTGACGCTTTTGAG
TTTTTTAGACTTCCCTTTCTCCACTTTGGCCACAACCAGTACACTGTAAGCGACTGTAGGAGAA
TCGAATCCGCCGTATTTCTTGGGGTCCCAATCTTTTTTGCGTGCGATCAGCTTGTCGCTGTTCC
TTTTCGGGAGGATACTTTCCTTGGAGAAGCCTCCGGTCTGTACTTCGGTCTTTTTAACGATGTT
CACCTGCGGCATGGACAGGACCTTCCGGACTGTCGCGAAATCCCTACCCTTGTCCCACACGATT
TCTCCTGTTTCTCCGTTTGTTTCGATAAGTGGTCGCTTCCGAATCTCTCCATTGGCCAGTGTAA
TCTCGGTCTTGAAAAAATTCATAATATTGCTGTAAAAGAAGTACTTAGCGGTGGCCTTGCCTAT
TTCCTGCTCAGACTTTGCGATCATTTTCCTAACATCGTACACTTTATAGTCTCCGTAAACAAAT
TCAGATTCAAGCTTGGGATATTTTTTGATAAGTGCAGTGCCTACCACTGCATTCAGGTAGGCAT
CATGCGCATGGTGGTAATTGTTGATCTCTCTCACCTTATAAAACTGAAAGTCCTTTCTGAAATC
TGAGACCAGCTTAGACTTCAGAGTAATAACTTTCACCTCTCGAATCAGTTTGTCATTTTCATCG
TACTTGGTGTTCATGCGTGAATCGAGAATTTGGGCCACGTGCTTGGTGATCTGGCGTGTCTCAA
CAAGCTGCCTTTTGATGAAGCCGGCTTTATCCAACTCAGACAGGCCACCTCGTTCAGCCTTAGT
CAGATTATCGAACTTCCGTTGTGTGATCAGTTTGGCGTTCAGCAGCTGCCGCCAATAATTTTTC
ATTTTCTTGACAACTTCTTCTGAGGGGACGTTATCACTCTTCCCTCTATTTTTATCGGATCTTG
TCAACACTTTATTATCAATAGAATCATCTTTGAGAAAAGACTGGGGCACGATATGATCCACGTC
GTAGTCGGAGAGCCGATTGATGTCCAGTTCCTGATCCACGTACATGTCCCTGCCGTTCTGCAGG
TAGTACAGGTAGAGCTTCTCATTCTGAAGCTGGGTGTTTTCAACTGGGTGTTCCTTAAGGATTT
GGGACCCCAGTTCTTTTATACCCTCTTCAATCCTCTTCATCCTTTCCCTACTGTTCTTCTGTCC
CTTCTGGGTAGTTTGGTTCTCTCGGGCCATCTCGATAACGATATTCTCGGGCTTATGCCTTCCC
ATTACTTTGACGAGTTCATCCACGACCTTAACGGTCTGCAGTATTCCCTTTTTGATAGCTGGGC
TACCTGCAAGATTAGCGATGTGCTCGTGAAGACTGTCCCCCTGGCCAGAAACTTGTGCTTTCTG
GATGTCCTCCTTAAAGGTGAGAGAGTCATCATGGATCAACTGCATGAAGTTCCGGTTGGCAAAT
CCATCGGACTTAAGAAAATCCAGGATTGTCTTTCCACTCTGCTTGTCTCGGATCCCATTGATCA
GTTTTCTTGACAGCCGCCCCCATCCTGTATATCGGCGCCTCTTGAGCTGTTTCATGACTTTGTC
GTCGAAGAGATGAGCGTAAGTTTTCAAGCGTTCTTCAATCATCTCCCTATCTTCAAACAACGTA
AGGGTGAGGACAATGTCCTCAAGAATGTCCTCGTTCTCCTCATTGTCCAGGAAGTCCTTGTCTT
TAATGATTTTCAGGAGATCGTGATACGTTCCCAGGGATGCGTTGAAGCGATCCTCCACTCCGCT
GATTTCAACAGAGTCGAAACATTCAATCTTTTTGAAATAGTCTTCTTTGAGCTGTTTCACGGTA
ACTTTCCGGTTCGTCTTGAAGAGGAGGTCCACGATAGCTTTCTTCTGCTCTCCAGACAGGAATG
CTGGCTTTCTCATCCCTTCTGTGACGTATTTGACCTTGGTGAGCTCGTTATAAACTGTGAAGTA

CTCGTACAGCAGAGAGTGTTTAGGAAGCACCTTTTCGTTAGGCAGATTTTTATCAAAGTTAGTC
ATCCTTTCGATGAAGGACTGGGCAGAGGCCCCCTTATCCACGACTTCCTCGAAGTTCCAGGGAG
TGATGGTCTCTTCTGATTTGCGAGTCATCCACGCGAATCTGGAATTTCCCCGGGCGAGGGGGCC
TACATAGTAGGGTATCCGAAATGTGAGGATTTTCTCAATCTTTTCCCTGTTATCTTTCAAAAAG
GGGTAGAAATCCTCTTGCCGCCTGAGGATAGCGTGCAGTTCGCCCAGGTGAATCTGGTGGGGGA
TGCTTCCATTGTCGAAAGTGCGCTGTTTGCGCAACAGATCTTCTCTGTTAAGCTTTACCAGCAG
CTCCTCGGTGCCGTCCATTTTTTCCAAGATGGGCTTAATAAATTTGTAAAATTCCTCCTGGCTT
GCTCCGCCGTCAATGTATCCGGCGTAGCCATTTTTAGACTGATCGAAGAAAATTTCCTTGTACT
TCTCAGGCAGTTGCTGTCTGACAAGGGCCTTCAGCAAAGTCAAGTCTTGGTGGTGCTCATCATA
GCGCTTGATCATACTAGCGCTCAGCGGAGCTTTGGTGATCTCCGTGTTCACTCGCAGAATATCA
CTCAGCAGAATGGCGTCTGACAGGTTCTTTGCCGCCAAAAAAAGGTCTGCGTACTGGTCGCCGA
TCTGGGCCAGCAGATTGTCGAGATCATCATCGTAGGTGTCTTTGCTCAGTTGAAGCTTGGCATC
TTCGGCCAGGTCGAAGTTAGATTTAAAGTTGGGGGTCAGCCCGAGTGACAGGGCGATAAGATTA
CCAAACAGGCCGTTCTTCTTCTCCCCAGGGAGCTGTGCGATGAGGTTTTCGAGCCGCCGGGATT
TGGACAGCCTAGCGCTCAGGATTGCTTTGGCGTCAACTCCGGATGCGTTGATCGGGTTCTCTTC
GAAAAGCTGATTGTAAGTCTGAACCAGTTGGATAAAGAGTTTGTCGACATCGCTGTTGTCTGGG
TTCAGGTCCCCCTCGATGAGGAAGTGTCCCCGAAATTTGATCATATGCGCCAGCGCGAGATAGA
TCAACCGCAAGTCAGCCTTATCAGTACTGTCTACAAGCTTCTTCCTCAGATGATATATGGTTGG
GTACTTTTCATGGTACGCCACCTCGTCCACGATATTGCCAAAGATTGGGTGGCGCTCGTGCTTT
TTATCCTCCTCCACCAAAAAGGACTCCTCCAGCCTATGGAAGAAAGAGTCATCCACCTTAGCCA
TCTCATTACTAAAGATCTCCTGCAGGTAGCAGATCCGATTCTTTCTGCGGGTATATCTGCGCCG
TGCTGTTCTTTTGAGCCGCGTGGCTTCGGCCGTCTCCCCGGAGTCGAACAGGAGGGCGCCAATG
AGGTTCTTCTTTATGCTGTGGCGATCGGTATTGCCCAGAACTTTGAATTTTTTGCTCGGCACCT
TGTACTCGTCCGTAATGACGGCCCAGCCGACGCTGTTTGTGCCGATATCGAGCCCAATGGAGTA
CTTCTTGTCCATGGTGGCGGCTCTTGAAGGACGACGTCATCATCCCTTGCCCGGATGCGCGGGC
TTCTTGTCTAGCACAGGAGCCTGGGGTAGAGCGCATGCAAATTACGCGCTGTGCTTTGT
GGGAAATCACCCTAAACGAAAAATTTATTCCTCTTTCGAGCCTTATAGTGGCGGCCGGTCTACA
TCCTAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAA
GTGGCACCGAGTCGGTGCTTTTTTGTTAACGCGGCCGCCTAGAGTCGACCTGCAGGCATGCAAG
C
TTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCC
ACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTC
ACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATT
AATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCT
CACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGG
CGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCA

GCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCT
GACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAA
GATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCG
CTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCT
GTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGT
TCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGAC
TTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTA
CAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGC
TCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACC
GCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAG
AAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGAT
TTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTT
AAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGG
CACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGT
GTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGC
GAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCG
CAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGA
GTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGT
CACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATG
ATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAG
TTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCAT
CCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCG
GCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTA
AAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGA
GATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAG
CGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGG
AAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTC
TCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATT
TCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAAT
AGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACA
TGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCA
GGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATT
GTACTGAGAGTGCACCATAGCGGCCGCGGTACGATATGGTGCTGTGTAGGCTCATATGTGGATC
TCAGAACCCACATGTACTCTGCTCCCCAGGTCTTGGTGCGCTTTCTATTCTCTGTCAGCAAAGC
CTATCGAAGAATCACCTACCACAACTGGCGCCACGGCTTCAATGTAGCCCAGACCATGTTTACC
CTACTCATGGTACGTATGTAAATTGGATGGGCTAGATGAATCAGAGGGCTGGGGCAAGGACCAC

AGCTAACTATCTTCTGGCCCAAGGGTACCCACACAAAAAACCAACACACAG (SEQ ID NO: 64)

mSPA_HDR rd12(250) правая

CTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTT
ATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTA
ATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTG
TCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCT
CTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGC
TCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGA
GCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGC
TCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGG
ACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTG
CCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCAC
GCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCC
CGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACAC
GACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTG
CTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTG
CGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACC
ACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTC
AAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGG
GATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGT
TTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTG
AGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTA
GATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCA
CGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTG
GTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAG
TTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCG
TCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCA
TGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGC
AGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGA
TGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGA
GTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCT
CATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGT
TCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTG
GGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTG
AATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGC

GGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAA
AAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTAT
CACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTC
CCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGT
CAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAG
AGTGCACCATAGCGGCCGCGGTACCCACACAAAAAACCAACACACAGATGTAATGAAAATAAAG
ATATTTTATTTCACACCTTCCTCTTCTTCTTGGGGTCAGCCCTGCTGTCTCCACCGAGCTGAGA
GAGGTCGATTCTTGTTTCATAGAGCCCCGTAATTGACTGATGAATCAGTGTGGCGTCCAGGACC
TCCTTTGTAGAGGTGTACCGCTTTCTGTCTATGGTGGTGTCGAAGTACTTGAAGGCTGCAGGCG
CGCCCAAGTTGGTCAGAGTAAACAAGTGGATAATGTTTTCTGCCTGCTCCCTGATGGGCTTATC
CCTGTGCTTATTGTAAGCAGAAAGCACCTTATCGAGGTTAGCGTCGGCGAGGATCACTCTTTTG
GAGAATTCGCTTATTTGCTCGATGATCTCATCAAGGTAGTGTTTGTGTTGTTCCACGAACAGCT
GCTTCTGCTCATTATCTTCGGGAGACCCTTTGAGCTTTTCATAGTGGCTGGCCAGATACAAGAA
ATTAACGTATTTAGAGGGCAGTGCCAGCTCGTTACCTTTCTGCAGCTCGCCCGCACTAGCGAGC
ATTCGTTTCCGGCCGTTTTCAAGCTCAAAGAGAGAGTACTTGGGAAGCTTAATGATGAGGTCTT
TTTTGACCTCTTTATATCCTTTCGCCTCGAGAAAGTCGATGGGGTTTTTTTCGAAGCTTGATCG
CTCCATGATTGTGATGCCCAGCAGTTCCTTGACGCTTTTGAGTTTTTTAGACTTCCCTTTCTCC
ACTTTGGCCACAACCAGTACACTGTAAGCGACTGTAGGAGAATCGAATCCGCCGTATTTCTTGG
GGTCCCAATCTTTTTTGCGTGCGATCAGCTTGTCGCTGTTCCTTTTCGGGAGGATACTTTCCTT
GGAGAAGCCTCCGGTCTGTACTTCGGTCTTTTTAACGATGTTCACCTGCGGCATGGACAGGACC
TTCCGGACTGTCGCGAAATCCCTACCCTTGTCCCACACGATTTCTCCTGTTTCTCCGTTTGTTT
CGATAAGTGGTCGCTTCCGAATCTCTCCATTGGCCAGTGTAATCTCGGTCTTGAAAAAATTCAT
AATATTGCTGTAAAAGAAGTACTTAGCGGTGGCCTTGCCTATTTCCTGCTCAGACTTTGCGATC
ATTTTCCTAACATCGTACACTTTATAGTCTCCGTAAACAAATTCAGATTCAAGCTTGGGATATT
TTTTGATAAGTGCAGTGCCTACCACTGCATTCAGGTAGGCATCATGCGCATGGTGGTAATTGTT
GATCTCTCTCACCTTATAAAACTGAAAGTCCTTTCTGAAATCTGAGACCAGCTTAGACTTCAGA
GTAATAACTTTCACCTCTCGAATCAGTTTGTCATTTTCATCGTACTTGGTGTTCATGCGTGAAT
CGAGAATTTGGGCCACGTGCTTGGTGATCTGGCGTGTCTCAACAAGCTGCCTTTTGATGAAGCC
GGCTTTATCCAACTCAGACAGGCCACCTCGTTCAGCCTTAGTCAGATTATCGAACTTCCGTTGT
GTGATCAGTTTGGCGTTCAGCAGCTGCCGCCAATAATTTTTCATTTTCTTGACAACTTCTTCTG
AGGGGACGTTATCACTCTTCCCTCTATTTTTATCGGATCTTGTCAACACTTTATTATCAATAGA
ATCATCTTTGAGAAAAGACTGGGGCACGATATGATCCACGTCGTAGTCGGAGAGCCGATTGATG
TCCAGTTCCTGATCCACGTACATGTCCCTGCCGTTCTGCAGGTAGTACAGGTAGAGCTTCTCAT
TCTGAAGCTGGGTGTTTTCAACTGGGTGTTCCTTAAGGATTTGGGACCCCAGTTCTTTTATACC
CTCTTCAATCCTCTTCATCCTTTCCCTACTGTTCTTCTGTCCCTTCTGGGTAGTTTGGTTCTCT
CGGGCCATCTCGATAACGATATTCTCGGGCTTATGCCTTCCCATTACTTTGACGAGTTCATCCA

CGACCTTAACGGTCTGCAGTATTCCCTTTTTGATAGCTGGGCTACCTGCAAGATTAGCGATGTG
CTCGTGAAGACTGTCCCCCTGGCCAGAAACTTGTGCTTTCTGGATGTCCTCCTTAAAGGTGAGA
GAGTCATCATGGATCAACTGCATGAAGTTCCGGTTGGCAAATCCATCGGACTTAAGAAAATCCA
GGATTGTCTTTCCACTCTGCTTGTCTCGGATCCCATTGATCAGTTTTCTTGACAGCCGCCCCCA
TCCTGTATATCGGCGCCTCTTGAGCTGTTTCATGACTTTGTCGTCGAAGAGATGAGCGTAAGTT
TTCAAGCGTTCTTCAATCATCTCCCTATCTTCAAACAACGTAAGGGTGAGGACAATGTCCTCAA
GAATGTCCTCGTTCTCCTCATTGTCCAGGAAGTCCTTGTCTTTAATGATTTTCAGGAGATCGTG
ATACGTTCCCAGGGATGCGTTGAAGCGATCCTCCACTCCGCTGATTTCAACAGAGTCGAAACAT
TCAATCTTTTTGAAATAGTCTTCTTTGAGCTGTTTCACGGTAACTTTCCGGTTCGTCTTGAAGA
GGAGGTCCACGATAGCTTTCTTCTGCTCTCCAGACAGGAATGCTGGCTTTCTCATCCCTTCTGT
GACGTATTTGACCTTGGTGAGCTCGTTATAAACTGTGAAGTACTCGTACAGCAGAGAGTGTTTA
GGAAGCACCTTTTCGTTAGGCAGATTTTTATCAAAGTTAGTCATCCTTTCGATGAAGGACTGGG
CAGAGGCCCCCTTATCCACGACTTCCTCGAAGTTCCAGGGAGTGATGGTCTCTTCTGATTTGCG
AGTCATCCACGCGAATCTGGAATTTCCCCGGGCGAGGGGGCCTACATAGTAGGGTATCCGAAAT
GTGAGGATTTTCTCAATCTTTTCCCTGTTATCTTTCAAAAAGGGGTAGAAATCCTCTTGCCGCC
TGAGGATAGCGTGCAGTTCGCCCAGGTGAATCTGGTGGGGGATGCTTCCATTGTCGAAAGTGCG
CTGTTTGCGCAACAGATCTTCTCTGTTAAGCTTTACCAGCAGCTCCTCGGTGCCGTCCATTTTT
TCCAAGATGGGCTTAATAAATTTGTAAAATTCCTCCTGGCTTGCTCCGCCGTCAATGTATCCGG
CGTAGCCATTTTTAGACTGATCGAAGAAAATTTCCTTGTACTTCTCAGGCAGTTGCTGTCTGAC
AAGGGCCTTCAGCAAAGTCAAGTCTTGGTGGTGCTCATCATAGCGCTTGATCATACTAGCGCTC
AGCGGAGCTTTGGTGATCTCCGTGTTCACTCGCAGAATATCACTCAGCAGAATGGCGTCTGACA
GGTTCTTTGCCGCCAAAAAAAGGTCTGCGTACTGGTCGCCGATCTGGGCCAGCAGATTGTCGAG
ATCATCATCGTAGGTGTCTTTGCTCAGTTGAAGCTTGGCATCTTCGGCCAGGTCGAAGTTAGAT
TTAAAGTTGGGGGTCAGCCCGAGTGACAGGGCGATAAGATTACCAAACAGGCCGTTCTTCTTCT
CCCCAGGGAGCTGTGCGATGAGGTTTTCGAGC

CGCCGGGATTTGGACAGCCTAGCGCTCAGGATTGCTTTGGCGTCAACTCCGGATGCGTT
GATCGGGTTCTCTTCGAAAAGCTGATTGTAAGTCTGAACCAGTTGGATAAAGAGTTTGTCGACA
TCGCTGTTGTCTGGGTTCAGGTCCCCCTCGATGAGGAAGTGTCCCCGAAATTTGATCATATGCG
CCAGCGCGAGATAGATCAACCGCAAGTCAGCCTTATCAGTACTGTCTACAAGCTTCTTCCTCAG

ATGATATATGGTTGGGTACTTTTCATGGTACGCCACCTCGTCCACGATATTGCCAAAGA
TTGGGTGGCGCTCGTGCTTTTTATCCTCCTCCACCAAAAAGGACTCCTCCAGCCTATGGAAGAA
AGAGTCATCCACCTTAGCCATCTCATTACTAAAGATCTCCTGCAGGTAGCAGATCCGATTCTTT
CTGCGGGTATATCTGCGCCGTGCTGTTCTTTTGAGCCGCGTGGCTTCGGCCGTCTCCCCGGAGT
CGAACAGGAGGGCGCCAATGAGGTTCTTCTTTATGCTGTGGCGATCGGTATTGCCCAGAACTTT
GAATTTTTTGCTCGGCACCTTGTACTCGTCCGTAATGACGGCCCAGCCGACGCTGTTTGTGCCG
ATATCGAGCCCAATGGAGTACTTCTTGTCCATGGTGGCGGCTCTTGAAGGACGACGTCATCATC

CCTTGCCCGGATGCGCGGGCTTCTTGTCTAGCACAGGAGCCTGGGGTAGAGCGCATGCAAATTA
CGCGCTGTGCTTTGTGGGAAATCACCCTAAACGAAAAATTTATTCCTCTTTCGAGCCTTATAGT
GGCGGCCGGTCTACATCCTAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGT
TATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTGTTGTTCTGTCTATAGGTAAGCTGA
CAAATAACAAATAGGCACATAGAAAATCTAGTAAGTAGTACCACCTGATATCTCACTTTGCTGC
AGGCAGGATTCCCCTCTGGCTCACTGGCAGTCTCCTCCGGTGTGGGCCAGGGCTCTTTGAAGTT
GGATCTGAGCCTTTCTATCACCTGTTTGATGGACAAGCCCTTTTGCACAAGTTTGACTTCAAGG
AGGGCCATGTCACATACCACAGAAGGTAAGTCCATGAACGCGGCCGCCTAGAGTCGAC (SEQ
ID NO: 65)

mSPA_HDR rd12(250) левая

ATGTAATGAAAATAAAGATATTTTATTTCACACCTTCCTCTTCTTCTTGGGGTCAGCCC
TGCTGTCTCCACCGAGCTGAGAGAGGTCGATTCTTGTTTCATAGAGCCCCGTAATTGACTGATG
AATCAGTGTGGCGTCCAGGACCTCCTTTGTAGAGGTGTACCGCTTTCTGTCTATGGTGGTGTCG
AAGTACTTGAAGGCTGCAGGCGCGCCCAAGTTGGTCAGAGTAAACAAGTGGATAATGTTTTCTG
CCTGCTCCCTGATGGGCTTATCCCTGTGCTTATTGTAAGCAGAAAGCACCTTATCGAGGTTAGC
GTCGGCGAGGATCACTCTTTTGGAGAATTCGCTTATTTGCTCGATGATCTCATCAAGGTAGTGT
TTGTGTTGTTCCACGAACAGCTGCTTCTGCTCATTATCTTCGGGAGACCCTTTGAGCTTTTCAT
AGTGGCTGGCCAGATACAAGAAATTAACGTATTTAGAGGGCAGTGCCAGCTCGTTACCTTTCTG
CAGCTCGCCCGCACTAGCGAGCATTCGTTTCCGGCCGTTTTCAAGCTCAAAGAGAGAGTACTTG
GGAAGCTTAATGATGAGGTCTTTTTTGACCTCTTTATATCCTTTCGCCTCGAGAAAGTCGATGG
GGTTTTTTTCGAAGCTTGATCGCTCCATGATTGTGATGCCCAGCAGTTCCTTGACGCTTTTGAG
TTTTTTAGACTTCCCTTTCTCCACTTTGGCCACAACCAGTACACTGTAAGCGACTGTAGGAGAA
TCGAATCCGCCGTATTTCTTGGGGTCCCAATCTTTTTTGCGTGCGATCAGCTTGTCGCTGTTCC
TTTTCGGGAGGATACTTTCCTTGGAGAAGCCTCCGGTCTGTACTTCGGTCTTTTTAACGATGTT
CACCTGCGGCATGGACAGGACCTTCCGGACTGTCGCGAAATCCCTACCCTTGTCCCACACGATT
TCTCCTGTTTCTCCGTTTGTTTCGATAAGTGGTCGCTTCCGAATCTCTCCATTGGCCAGTGTAA
TCTCGGTCTTGAAAAAATTCATAATATTGCTGTAAAAGAAGTACTTAGCGGTGGCCTTGCCTAT
TTCCTGCTCAGACTTTGCGATCATTTTCCTAACATCGTACACTTTATAGTCTCCGTAAACAAAT
TCAGATTCAAGCTTGGGATATTTTTTGATAAGTGCAGTGCCTACCACTGCATTCAGGTAGGCAT
CATGCGCATGGTGGTAATTGTTGATCTCTCTCACCTTATAAAACTGAAAGTCCTTTCTGAAATC
TGAGACCAGCTTAGACTTCAGAGTAATAACTTTCACCTCTCGAATCAGTTTGTCATTTTCATCG
TACTTGGTGTTCATGCGTGAATCGAGAATTTGGGCCACGTGCTTGGTGATCTGGCGTGTCTCAA
CAAGCTGCCTTTTGATGAAGCCGGCTTTATCCAACTCAGACAGGCCACCTCGTTCAGCCTTAGT
CAGATTATCGAACTTCCGTTGTGTGATCAGTTTGGCGTTCAGCAGCTGCCGCCAATAATTTTTC
ATTTTCTTGACAACTTCTTCTGAGGGGACGTTATCACTCTTCCCTCTATTTTTATCGGATCTTG
TCAACACTTTATTATCAATAGAATCATCTTTGAGAAAAGACTGGGGCACGATATGATCCACGTC

GTAGTCGGAGAGCCGATTGATGTCCAGTTCCTGATCCACGTACATGTCCCTGCCGTTCTGCAGG
TAGTACAGGTAGAGCTTCTCATTCTGAAGCTGGGTGTTTTCAACTGGGTGTTCCTTAAGGATTT
GGGACCCCAGTTCTTTTATACCCTCTTCAATCCTCTTCATCCTTTCCCTACTGTTCTTCTGTCC
CTTCTGGGTAGTTTGGTTCTCTCGGGCCATCTCGATAACGATATTCTCGGGCTTATGCCTTCCC
ATTACTTTGACGAGTTCATCCACGACCTTAACGGTCTGCAGTATTCCCTTTTTGATAGCTGGGC
TACCTGCAAGATTAGCGATGTGCTCGTGAAGACTGTCCCCCTGGCCAGAAACTTGTGCTTTCTG
GATGTCCTCCTTAAAGGTGAGAGAGTCATCATGGATCAACTGCATGAAGTTCCGGTTGGCAAAT
CCATCGGACTTAAGAAAATCCAGGATTGTCTTTCCACTCTGCTTGTCTCGGATCCCATTGATCA
GTTTTCTTGACAGCCGCCCCCATCCTGTATATCGGCGCCTCTTGAGCTGTTTCATGACTTTGTC
GTCGAAGAGATGAGCGTAAGTTTTCAAGCGTTCTTCAATCATCTCCCTATCTTCAAACAACGTA
AGGGTGAGGACAATGTCCTCAAGAATGTCCTCGTTCTCCTCATTGTCCAGGAAGTCCTTGTCTT
TAATGATTTTCAGGAGATCGTGATACGTTCCCAGGGATGCGTTGAAGCGATCCTCCACTCCGCT
GATTTCAACAGAGTCGAAACATTCAATCTTTTTGAAATAGTCTTCTTTGAGCTGTTTCACGGTA
ACTTTCCGGTTCGTCTTGAAGAGGAGGTCCACGATAGCTTTCTTCTGCTCTCCAGACAGGAATG
CTGGCTTTCTCATCCCTTCTGTGACGTATTTGACCTTGGTGAGCTCGTTATAAACTGTGAAGTA
CTCGTACAGCAGAGAGTGTTTAGGAAGCACCTTTTCGTTAGGCAGATTTTTATCAAAGTTAGTC
ATCCTTTCGATGAAGGACTGGGCAGAGGCCCCCTTATCCACGACTTCCTCGAAGTTCCAGGGAG
TGATGGTCTCTTCTGATTTGCGAGTCATCCACGCGAATCTGGAATTTCCCCGGGCGAGGGGGCC
TACATAGTAGGGTATCCGAAATGTGAGGATTTTCTCAATCTTTTCCCTGTTATCTTTCAAAAAG
GGGTAGAAATCCTCTTGCCGCCTGAGGATAGCGTGCAGTTCGCCCAGGTGAATCTGGTGGGGGA
TGCTTCCATTGTCGAAAGTGCGCTGTTTGCGCAACAGATCTTCTCTGTTAAGCTTTACCAGCAG
CTCCTCGGTGCCGTCCATTTTTTCCAAGATGGGCTTAATAAATTTGTAAAATTCCTCCTGGCTT
GCTCCGCCGTCAATGTATCCGGCGTAGCCATTTTTAGACTGATCGAAGAAAATTTCCTTGTACT
TCTCAGGCAGTTGCTGTCTGACAAGGGCCTTCAGCAAAGTCAAGTCTTGGTGGTGCTCATCATA
GCGCTTGATCATACTAGCGCTCAGCGGAGCTTTGGTGATCTCCGTGTTCACTCGCAGAATATCA
CTCAGCAGAATGGCGTCTGACAGGTTCTTTGCCGCCAAAAAAAGGTCTGCGTACTGGTCGCCGA
TCTGGGCCAGCAGATTGTCGAGATCATCATCGTAGGTGTCTTTGCTCAGTTGAAGCTTGGCATC
TTCGGCCAGGTCGAAGTTAGATTTAAAGTTGGGGGTCAGCCCGAGTGACAGGGCGATAAGATTA
CCAAACAGGCCGTTCTTCTTCTCCCCAGGGAGCTGTGCGATGAGGTTTTCGAGCCGCCGGGATT
TGGACAGCCTAGCGCTCAGGATTGCTTTGGCGTCAACTCCGGATGCGTTGATCGGGTTCTCTTC
GAAAAGCTGATTGTAAGTCTGAACCAGTTGGATAAAGAGTTTGTCGACATCGCTGTTGTCTGGG
TTCAGGTCCCCCTCGATGAGGAAGTGTCCCCGAAATTTGATCATATGCGCCAGCGCGAGATAGA
TCAACCGCAAGTCAGCCTTATCAGTACTGTCTACAAGCTTCTTCCTCAGATGATATATGGTTGG
GTACTTTTCATGGTACGCCACCTCGTCCACGATATTGCCAAAGATTGGGTGGCGCTCGTGCTTT
TTATCCTCCTCCACCAAAAAGGACTCCTCCAGCCTATGGAAGAAAGAGTCATCCACCTTAGCCA
TCTCATTACTAAAGATCTCCTGCAGGTAGCAGATCCGATTCTTTCTGCGGGTATATCTGCGCCG

TGCTGTTCTTTTGAGCCGCGTGGCTTCGGCCGTCTCCCCGGAGTCGAACAGGAGGGCGCCAATG
AGGTTCTTCTTTATGCTGTGGCGATCGGTATTGCCCAGAACTTTGAATTTTTTGCTCGGCACCT
TGTACTCGTCCGTAATGACGGCCCAGCCGACGCTGTTTGTGCCGATATCGAGCCCAATGGAGTA
CTTCTTGTCCATGGTGGCGGCTCTTGAAGGACGACGTCATCATCCCTTGCCCGGATGCGCGGGC
TTCTTGTCTAGCACAGGAGCCTGGGGTAGAGCGCATGCAAATTACGCGCTGTGCTTTGT
GGGAAATCACCCTAAACGAAAAATTTATTCCTCTTTCGAGCCTTATAGTGGCGGCCGGTCTACA
TCCTAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAA
GTGGCACCGAGTCGGTGCTTTTTTGTTAACGCGGCCGCCTAGAGTCGACCTGCAGGCATGCAAG
C

TTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCC
ACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTC
ACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATT
AATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCT
CACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGG
CGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCA
GCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCT
GACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAA

GATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCG
CTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCT
GTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGT
TCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGAC
TTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTA
CAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGC
TCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACC
GCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAG
AAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGAT
TTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTT
AAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGG
CACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGT

GTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGC
GAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCG
CAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGA
GTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGT
CACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATG
ATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAG
TTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCAT

CCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCG
GCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTA
AAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGA
GATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAG
CGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGG
AAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTC
TCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATT
TCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAAT
AGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACA
TGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCA
GGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATT
GTACTGAGAGTGCACCATAGCGGCCGCGGTTCTGTCTATAGGTAAGCTGACAAATAACAAATAG
GCACATAGAAAATCTAGTAAGTAGTACCACCTGATATCTCACTTTGCT

GCAGGCAGGATTCCCCTCTGGCTCACTGGCAGTCTCCTCCGGTGTGGGCCAGGGCTCTT
TGAAGTTGGATCTGAGCCTTTCTATCACCTGTTTGATGGACAAGCCCTTTTGCACAAGTTTGAC
TTCAAGGAGGGCCATGTCACATACCACAGAAGGTAAGTCCATGGTACCCACACAAAAAACCAAC
ACACAG (SEQ ID NO: 66)

Исходная плазмида mSPA_HDR SaCas9

GCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATT
CCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAAC
TCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCA
TTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCG
CTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGG
TAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCA
AAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGAC
GAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACC
AGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATA
CCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTC
AGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACC
GCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACT
GGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTG
AAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGC
CAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGG
TGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTG
ATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGA
GATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTA

AAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCA
GCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATAC
GGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCC
AGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTA
TCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATA
GTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGC
TTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAA
GCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCA
TGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGAC
TGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCG
GCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAAC
GTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCAC
TCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACA
GGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCT
TCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGA
ATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGAC
GTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTC
GTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCAC
AGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGC
GGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATAGCG
GCCGCGGTACCCACACAAAAAACCAACACACAGATGTAATGAAAATAAAGATATTTTATTAGGC
TGATCAGCGAGCTCTAGGAATTCTTAGGATCCCTTTTTCTTTTTTGCCTGGCCGGCCTTTTTCG
TGGCCGCCGGCCTTTTGCCCTTTTTGATGATCTGAGGGTGCTTCTTAGATTTCACTTCATACAG
GTTGCCCAGAATGTCTGTGCTGTACTTCTTAATGCTCTGGGTCTTGGAGGCGATTGTCTTAATG
ATCCTGGGGGGCCTCTTGTCGTTCATGTTTTCCAGGTACTCGCGGTAGGTGATGTCGATCATGT
TCACTTCGATCCGGTTCAGCAGGTCGTTGTTCACGCCGATCACTCTATACAGCTCGCCGTTGAT
CTTGATCAGATCGTTGTTGTAGAAGGAGGCGATAAACTCGGCCTGGTTGCTGATCTTCTTCAGC
TTCTTAGCTTCCTCATAGCACTTGCTATTCACTTCGTAGTAGTTTTCTTTTTTGATCACATCCA
GATTCTTCACGGTCACGAACTTGTACACGCCATTGTCCAGGTACACGTCGAATCTGTAGGGCTT
CAGGGACAGCTTCACGACCTTGTTTCTGCTGTTGGGGTAGTCGTCGGTGATGTCCAGATGGGCG
TTCAGTTTGTTGCCGTAATACTTAATCTTCTTGATCACGGGGCCGTTGTCCTTTTTGGAGTACT
TGGTCAGGTAGTTCCCGGTTTCCTCGTAGTACTTGTACAGGGGATTCTTCTCGTCGCCGTACTG
TTCCATAATCAGCTTCAGTTTCTGGTAGGTCTGGGGGTCGTGGTGGTACATCAGCAGCTTTTCG
GGGCTCTTGTTGATCAGCTTTTTCAGCTTGTCATTGTCCTTGTCGTACAGGCCGTTCAGATTGT
TCACGATCAGGGTGTTGCCCTTGTCGTCCTTCCGGGTGGAGTACAGGGTGTCGTTAATCAGCTC
TCTATTAGGCTTCTTGTCCACCCGGTGGCTGTACTTGTAGTCCTTGAAGTCCTTAATGTGCTTG

ATCTGGTGGGGGGTGATGAAGATCTCTTTGTACTCCTGCTCGGTTTCGATCTCGGGCATGCTCT
CGGCCTGCTTTTCCTCGAACATCTGGTTTTCCATCACTTTTTTGGCCTTGTCCAGTTTCTTCCA
CTCTTTGAAGATGAAATCGGCGTTGGCAATGATCAGGGCGTCCTCGGCGTGGTGCTTGTACCCC
TTGTTCCGCTCTTTCTTAAACTTCCACTTCCGCCGCAGAAAGCTGGTGAAGCCGCCATTGATGG
ACTTCACTTTCACGTCCAGGTTGTTCACTCTGAAGTAGCTCCGCAGCAGGTTCATCAGGCCTCT
GGTGGCGTATCTGGTATCCACCAGGTTCCGGTTGATGAAGTCTTTCTGCACGGAGAACCTGTTG
ATGTCCCGTTCTTCCAGCAGATACTCTTTCTTGGTCTTGCTGATTCTGCCCTTGCCCTTGGCCA
GATTCAGGATGTGCTTCTTGAAGGTTTCGTAGCTGATCTTGCTGTCGCTGCTGCTCAGGTACTG
GAATGGGGTCCGGTTGCCCTTCTTGCTGTTTTCTTCCTGCTTCACGAGCACCTTGTTGTTGAAG
CTGTTGTCGAAGGACACGCTTCTGGGGATGATGTGGTCCACCTCATAGTTGAAGGGGTTGTTCA
GCAGATCTTCCAGAGGGATGGCTTCCAGGCTGTACAGGCACTTGCCTTCCTGCATGTCGTGCAG
CTTGATCTTCTCGATCAGGTACTTGGCGTTCTCTTTGCCGGTGGTCCGGATGATTTCCTCGATC
CGCTCGTTGGTCTGCCGGTTCCGCTTCTGCATCTCGTTGATCATTTTCTGGGCGTCCTTGGAGT
TCTTCTCGCGGGCCAGCTCGATAATGATGTCGTTGGGCAGGCCGTACTTCTTGATGATGGCGTT
GATCACTTTGATGCTCTGGATGAAGCTTCTCTTCACGACGGGGCTCAGGATGAAGTCGTCCACC
AGGGTGGTGGGGATCTCTTTCTGCTGGGACAGGTCCACCTTCTTGGGCACCAGCTTCAGCCGGT
TGAAGATAGCGATCTGGTTGTCGTTGGTGTGCCACAGCTCGTCCAGGATCAGGTTGATGGCCTT
CAGGCTCAGGTTGTGGGTGCCGGTATAGCCCTTCAGATTAGAGATCTGCTCGATCTCTTCCTGG
GTCAGCTCGGAGTTCAGATTGGTCAGTTCTTCCTGGATGTCCTCGCTGCTCTGGTAGATGGTCA
GGATCTTGGCAATCTGATCCAGCAGCTCGGCGTTCTCAATAATCTCTTTCCGGGCGGTAATGTC
CTTGATGTCGTGGTACACCTTCAGGTTGGTGAACTCGGGCTTGCCGGTGCTGGTCACTCTGTAG
CCCTTAATATCCTCTTCGTTCACGAGGATTTCTTTGGCGATCTGCTTCAGGGTGGGCTTCTTCT
TCTGCTTGAACACGTTCTCGATGATCTGGAACTTCTCGTAATATTCCAGCTTCTCGTTCTCGTC
CCTGGTGATCACGAGATTGTTCAGGTCGTTCAGGGCGTTGTACAGGTCGGCGTTGTAGGCGTAC
TTCACGCTCCGCAGTTCCTCGGGGAAGTAGGTGCAGTGGCCCATCAGCATCTCGTACCATTCTT
TGATGTCCTTCCAGCCGAAGGGGCTGCCCTCGCCAGGTCCCTCATAGTAGGTCCGCCGGGTTTC
CAGCAGGTCGATGTAGGTGTCGATGAAGCTCTGGTCCAGCTGGTGGTAGGCCTTCTGCACCTTC
AGCAGCTGTTTGGCTTCTTTCACGTAGTCGCTGGTCTTGAATCTGTTGATGCTGCCCCGCACTT
CGCCGTCTTTCTTCAGCCGTTCCAGCTGCAGTTCGGCCACGTATTTCTCTTCCAGGGCCTTGCT
GTTCCGGCTGATCTGCTCTTTGGTGGACAGCTCGTTGCCGGTGTCCTCTTCCACCTCGTTCACG
TTGT

GCACGCCTCTTCTCTTGGCCAGGTGCAGCAGGGCGGCAGAGAACTCTTCCTCGCTCAGC
TTCTGGCTCAGGCCCTTCACTCTGGCCTCGTAGGGGTTGATGCCGCTCAGCTCGCTGTGGTCGG
TCAGCAGGTTGTAGTCGAACAGCAGCTTCTTCACTCTCTGGATTCTATGCCGCCTCCGCCGCTT
CAGCCTTCTGGCGCCTCTCTTGCTCCGCCTGCCCTCGTTGTTTTCCACGTTGGCCTCTTTGAAC
AGCCGCACGCCGGCATCGATCACGTCCCGTGTCTCGTAGTCGATGATGCCGTAGCCCACGCTGG

TGATGCCGATGTCCAGGCCCAGGATGTAGTTCCGCTTGGCTGCTGGGACTCCGTGGATACCGAC
CTTCCGCTTCTTCTTTGGGGCCATGTGGCGGCTCTTGAAGGACGACGTCATCATCCCTTGCCCG
GATGCGCGGGCTTCTTGTCTAGCACAGGAGCCTGGGGTAGAGCGCATGCAAATTACGCGCTGTG
CTTTGTGGGAAATCACCCTAAACGAAAAATTTATTCCTCTTTCGAGCCTTATAGTGGCGGCCGG
TCTACATCCTAGGTTTTAGTACTCTGGAAACAGAATCTACTAAAACAAGGCAAAATGCCGTGTT
TATCTCGTCAACTTGTTGGCGAGATTTTTTGTTAACGCGGCCGCCTAGAGTCGACCTGCAGGCA
TGCAA (SEQ ID NO: 67)

ССЫЛКИ

Все публикации, патентные заявки, патенты и другие ссылки, упомянутые в описании, отражают уровень специалистов в области, к которой относится настоящее изобретение. Все публикации, патентные заявки, патенты и другие ссылки включены в настоящее описание в качестве ссылок, как если каждая индивидуальная публикация, патентная заявка, патент и другая ссылка были конкретно и индивидуально указаны как включенные в качестве ссылок. Будет понятно, что, хотя ряд патентных заявок, патентов и других ссылок упоминаются в настоящем описании, такие ссылки не подразумевают допущения, что какой-либо из этих документов составляет часть общих знаний в данной области.

Хотя описанное выше изобретение описано подробно посредством иллюстрации и примеров для ясности понимания, специалистам в данной области будет понятно, что определенные изменения и модификации можно применять на практике в пределах объема прилагаемой формулы изобретения.

СПИСОК ПОСЛЕДОВАТЕЛЬНОСТЕЙ

<110> ДЗЕ ДЖОНС ХОПКИНС ЮНИВЕРСИТИ

<120> КОМПОЗИЦИИ И СПОСОБЫ, ВКЛЮЧАЮЩИЕ УСОВЕРШЕНСТВОАНИЯ ГИДОВЫХ РНК CRISPR С ИСПОЛЬЗОВАНИЕМ ПРОМОТОРА H1

<130> JHV-15525

<140> PCT/US2017/040707
<141> 2017-07-05

<150> 62/358,335
<151> 2016-07-05

<160> 147

<170> PatentIn version 3.5

<210> 1
<211> 9
<212> РНК
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический олигонуклеотид

<400> 1
gccgccacc 9

<210> 2
<211> 6
<212> РНК
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический олигонуклеотид

<400> 2
gccacc 6

<210> 3
<211> 398
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полинуклеотид

<400> 3
ggtggcggcg atagcttttc aggctttctg gagtccactc ggaggcctgg agccgcacaa 60

agcgccaggt cagcggtccc ggctgggtga gaccagcagg cggctctagc gcgcgggagc 120

tgggcgaggc tccgggacga cctcaccaat ggagactgca gtatttagca tgccccaccc 180

atctgcaagg cattctggat agtgtcaaaa cagccggaaa tcaagtccgt ttatctcaaa 240

ctttagcatt ttgggaataa atgatatttg ctatgctggt taaattagat tttagttaaa 300

tttcctgctg aagctctagt acgataagca acttgaccta agtgtaaagt tgagacttcc 360

ttcaggttta tatagcttgt gcgccgcttg ggtacctc 398

<210> 4
<211> 248
<212> ДНК
<213> Homo sapiens

<400> 4
ggtggcggca gtatttagca tgccccaccc atctgcaagg cattctggat agtgtcaaaa 60

cagccggaaa tcaagtccgt ttatctcaaa ctttagcatt ttgggaataa atgatatttg 120

ctatgctggt taaattagat tttagttaaa tttcctgctg aagctctagt acgataagca 180

acttgaccta agtgtaaagt tgagacttcc ttcaggttta tatagcttgt gcgccgcttg 240

ggtacctc 248

<210> 5
<211> 289
<212> ДНК
<213> Homo sapiens

<400> 5
ggtgtctgtt tgaggttgct agtgaacaca gttgtgtcag aagcaaatgt agtatttagc 60

atgccccacc catctgcaag gcattctgga tagtgtcaaa acagccggaa atcaagtccg 120

tttatctcaa actttagcat tttgggaata aatgatattt gctatgctgg ttaaattaga 180

ttttagttaa atttcctgct gaagctctag tacgataagc aacttgacct aagtgtaaag 240

ttgagacttc cttcaggttt atatagcttg tgcgccgctt gggtacctc 289

<210> 6
<211> 350
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический
полинуклеотид

<400> 6
ggtggcggcc tcgtatggga ccgcaccgga cacgggcgcc tgggccagga gcagagccgg 60

gccgtagaat agacatggcc gtcgggggcg gggcttcgga aggtttaacc aatccaaact 120

gttgtatttt gcatagcccc aaagcatttt ggttaacagt aaaaacatcc taaatttaag 180

tattttaatt taaacttaga acgaagcgag tataaaaagg attatttaac cctaaaacgg 240

attcaggatt tgttataata tcaagtacag tcggctacat aaggtcacca catgtgtaaa 300

gttacaaaat tctatggcct tatataccta ccaagagcct gcttactctc 350

<210> 7

<211> 289
<212> ДНК
<213> Mus musculus

<400> 7
ggtggcggca gacatggccg tcgggggcgg ggcttcggaa ggtttaacca atccaaactg 60

ttgtattttg catagcccca aagcattttg gttaacagta aaaacatcct aaatttaagt 120

attttaattt aaacttagaa cgaagcgagt ataaaaagga ttatttaacc ctaaaacgga 180

ttcaggattt gttataatat caagtacagt cggctacata aggtcaccac atgtgtaaag 240

ttacaaaatt ctatggcctt atatacctac caagagcctg cttactctc 289

<210> 8
<211> 330
<212> ДНК
<213> Mus musculus

<400> 8
ggtgtctgtt tgaggttgct agtgaacaca gttgtgtcag aagcaaatgt agacatggcc 60

gtcgggggcg gggcttcgga aggtttaacc aatccaaact gttgtatttt gcatagcccc 120

aaagcatttt ggttaacagt aaaaacatcc taaatttaag tattttaatt taaacttaga 180

acgaagcgag tataaaaagg attatttaac cctaaaacgg attcaggatt tgttataata 240

tcaagtacag tcggctacat aaggtcacca catgtgtaaa gttacaaaat tctatggcct 300

tatataccta ccaagagcct gcttactctc 330

<210> 9
<211> 20
<212> ДНК
<213> Mus musculus

<400> 9
gaagaaggtt cgagatctca 20

<210> 10
<211> 23
<212> ДНК
<213> Mus musculus

<400> 10
gaagaaggtt cgagatctca agg 23

<210> 11
<211> 19
<212> ДНК
<213> Неизвестная

<220>
<223> Описание неизвестной:
последовательность сайта TetO

<400> 11
tccctatcag tgatagaga 19

<210> 12
<211> 230
<212> ДНК
<213> неизвестная

<220>
<223> Описание неизвестной:
H1 последовательность дикого типа

<400> 12
ggaattcgaa cgctgacgtc atcaacccgc tccaaggaat cgcgggccca gtgtcactag 60

gcgggaacac ccagcgcgcg tgcgccctgg caggaagatg gctgtgaggg acaggggagt 120

ggcgccctgc aatatttgca tgtcgctatg tgttctggga aatcaccata aacgtgaaat 180

gtctttggat ttgggaatct tataagttct gtatgagacc actttttccc 230

<210> 13
<211> 230
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический
полинуклеотид

<400> 13
ggaattcgaa cgctgacgtc atcaacccgc tccaaggaat cgcgggccca gtgtcactag 60

gcgggaacac ccagcgcgcg tgcgccctgg caggaagatg gctgtgaggg acaggggagt 120

ggcgccctgc aatatttgca tgtcgctatg tgttctggga aatcaccata aacgtgaaat 180

ccctatcagt gatagagact tataagttcc ctatcagtga tagagatccc 230

<210> 14
<211> 215
<212> Белок
<213> Неизвестная

<220>
<223> Описание неизвестной:
TetR последовательность

<400> 14
Met Ser Arg Leu Asp Lys Ser Lys Val Ile Asn Ser Ala Leu Glu Leu
1 5 10 15

Leu Asn Glu Val Gly Ile Glu Gly Leu Thr Thr Arg Lys Leu Ala Gln
20 25 30

Lys Leu Gly Val Glu Gln Pro Thr Leu Tyr Trp His Val Lys Asn Lys
35 40 45

Arg Ala Leu Leu Asp Ala Leu Ala Ile Glu Met Leu Asp Arg His His
50 55 60

Thr His Phe Cys Pro Leu Glu Gly Glu Ser Trp Gln Asp Phe Leu Arg
65 70 75 80

Asn Asn Ala Lys Ser Phe Arg Cys Ala Leu Leu Ser His Arg Asp Gly
85 90 95

Ala Lys Val His Leu Gly Thr Arg Pro Thr Glu Lys Gln Tyr Glu Thr
100 105 110

Leu Glu Asn Gln Leu Ala Phe Leu Cys Gln Gln Gly Phe Ser Leu Glu
115 120 125

Asn Ala Leu Tyr Ala Leu Ser Ala Val Gly His Phe Thr Leu Gly Cys
130 135 140

Val Leu Glu Asp Gln Glu His Gln Val Ala Lys Glu Glu Arg Glu Thr
145 150 155 160

Pro Thr Thr Asp Ser Met Pro Pro Leu Leu Arg Gln Ala Ile Glu Leu
165 170 175

Phe Asp His Gln Gly Ala Glu Pro Ala Phe Leu Phe Gly Leu Glu Leu
180 185 190

Ile Ile Cys Gly Leu Glu Lys Gln Leu Lys Cys Glu Ser Gly Ser Ala
195 200 205

Tyr Ser Gly Ser Arg Glu Phe
210 215

<210> 15
<211> 126
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полинуклеотид

<400> 15
cagaaggccc taacttctac gtgcccttct ccaatgcgac gggtgtggta cgcagcccct 60

gaattcaccc acagtactac ctggctgagc catggcagtt ctccatgctg gccgcctaca 120

tgtttc 126

<210> 16
<211> 20
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический праймер

<400> 16
tggagccctg agtggctgag 20

<210> 17
<211> 25
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический праймер

<400> 17
ccacctagga ccatgaagag gtcag 25

<210> 18
<211> 91
<212> Белок
<213> Homo sapiens

<400> 18
Pro Ser Pro Ala Arg Pro Ala Leu Arg Ala Pro Ala Ser Ala Thr Ser
1 5 10 15

Gly Ser Arg Lys Arg Ala Arg Pro Pro Ala Ala Pro Gly Arg Asp Gln
20 25 30

Ala Arg Pro Pro Ala Arg Arg Arg Leu Arg Leu Ser Val Asp Glu Val
35 40 45

Ser Ser Pro Ser Thr Pro Glu Ala Pro Asp Ile Pro Ala Cys Pro Ser
50 55 60

Pro Gly Gln Lys Ile Lys Lys Ser Thr Pro Ala Ala Gly Gln Pro Pro
65 70 75 80

His Leu Thr Ser Ala Gln Asp Gln Asp Thr Ile
85 90

<210> 19
<211> 110
<212> Белок
<213> Homo sapiens

<400> 19
Met Asn Pro Ser Met Lys Gln Lys Gln Glu Glu Ile Lys Glu Asn Ile
1 5 10 15

Lys Asn Ser Ser Val Pro Arg Arg Thr Leu Lys Met Ile Gln Pro Ser
20 25 30

```
Ala Ser Gly Ser Leu Val Gly Arg Glu Asn Glu Leu Ser Ala Gly Leu
        35                  40                  45

Ser Lys Arg Lys His Arg Asn Asp His Leu Thr Ser Thr Thr Ser Ser
    50                  55                  60

Pro Gly Val Ile Val Pro Glu Ser Ser Glu Asn Lys Asn Leu Gly Gly
65                  70                  75                  80

Val Thr Gln Glu Ser Phe Asp Leu Met Ile Lys Glu Asn Pro Ser Ser
                85                  90                  95

Gln Tyr Trp Lys Glu Val Ala Glu Lys Arg Arg Lys Ala Leu
            100                 105                 110
```

<210> 20
<211> 1486
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полипептид

<400> 20

```
Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1               5                   10                  15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
```

```
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
```

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

```
           900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050

Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065

Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080

Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095

Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110

Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                1120                1125

Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                1135                1140

Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
```

```
    1145                1150                1155

Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                1165                1170

Phe Glu  Lys Asn Pro Ile Asp  Phe Leu Glu Ala Lys  Gly Tyr Lys
    1175                1180                1185

Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                1195                1200

Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1205                1210                1215

Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1220                1225                1230

Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1235                1240                1245

Pro Glu  Asp Asn Glu Gln Lys  Gln Leu Phe Val Glu  Gln His Lys
    1250                1255                1260

His Tyr  Leu Asp Glu Ile Ile  Glu Gln Ile Ser Glu  Phe Ser Lys
    1265                1270                1275

Arg Val  Ile Leu Ala Asp Ala  Asn Leu Asp Lys Val  Leu Ser Ala
    1280                1285                1290

Tyr Asn  Lys His Arg Asp Lys  Pro Ile Arg Glu Gln  Ala Glu Asn
    1295                1300                1305

Ile Ile  His Leu Phe Thr Leu  Thr Asn Leu Gly Ala  Pro Ala Ala
    1310                1315                1320

Phe Lys  Tyr Phe Asp Thr Thr  Ile Asp Arg Lys Arg  Tyr Thr Ser
    1325                1330                1335

Thr Lys  Glu Val Leu Asp Ala  Thr Leu Ile His Gln  Ser Ile Thr
    1340                1345                1350

Gly Leu  Tyr Glu Thr Arg Ile  Asp Leu Ser Gln Leu  Gly Gly Asp
    1355                1360                1365

Ser Arg  Ala Asp Pro Lys Lys  Lys Arg Lys Val Arg  Gly Gly Gly
    1370                1375                1380

Gly Ser  Gly Gly Gly Gly Ser  Gly Gly Gly Gly Ser  Pro Ser Pro
```

```
    1385                1390                1395

Ala Arg  Pro Ala Leu Arg Ala  Pro Ala Ser Ala Thr  Ser Gly Ser
    1400                1405                1410

Arg Lys  Arg Ala Arg Pro Pro  Ala Ala Pro Gly Arg  Asp Gln Ala
    1415                1420                1425

Arg Pro  Pro Ala Arg Arg Arg  Leu Arg Leu Ser Val  Asp Glu Val
    1430                1435                1440

Ser Ser  Pro Ser Thr Pro Glu  Ala Pro Asp Ile Pro  Ala Cys Pro
    1445                1450                1455

Ser Pro  Gly Gln Lys Ile Lys  Lys Ser Thr Pro Ala  Ala Gly Gln
    1460                1465                1470

Pro Pro  His Leu Thr Ser Ala  Gln Asp Gln Asp Thr  Ile
    1475                1480                1485
```

<210> 21
<211> 1505
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полипептид

<400> 21

```
Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1               5                   10                  15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110
```

```
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365
```

```
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620
```

```
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880
```

```
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050

Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065

Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080

Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095

Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110

Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                1120                1125
```

```
Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                 1135                 1140

Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                 1150                 1155

Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                 1165                 1170

Phe Glu  Lys Asn Pro Ile Asp  Phe Leu Glu Ala Lys  Gly Tyr Lys
    1175                 1180                 1185

Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                 1195                 1200

Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1205                 1210                 1215

Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1220                 1225                 1230

Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1235                 1240                 1245

Pro Glu  Asp Asn Glu Gln Lys  Gln Leu Phe Val Glu  Gln His Lys
    1250                 1255                 1260

His Tyr  Leu Asp Glu Ile Ile  Glu Gln Ile Ser Glu  Phe Ser Lys
    1265                 1270                 1275

Arg Val  Ile Leu Ala Asp Ala  Asn Leu Asp Lys Val  Leu Ser Ala
    1280                 1285                 1290

Tyr Asn  Lys His Arg Asp Lys  Pro Ile Arg Glu Gln  Ala Glu Asn
    1295                 1300                 1305

Ile Ile  His Leu Phe Thr Leu  Thr Asn Leu Gly Ala  Pro Ala Ala
    1310                 1315                 1320

Phe Lys  Tyr Phe Asp Thr Thr  Ile Asp Arg Lys Arg  Tyr Thr Ser
    1325                 1330                 1335

Thr Lys  Glu Val Leu Asp Ala  Thr Leu Ile His Gln  Ser Ile Thr
    1340                 1345                 1350

Gly Leu  Tyr Glu Thr Arg Ile  Asp Leu Ser Gln Leu  Gly Gly Asp
    1355                 1360                 1365
```

```
Ser Arg  Ala Asp Pro Lys Lys  Lys Arg Lys Val Arg  Gly Gly Gly
    1370                1375                1380


Gly Ser  Gly Gly Gly Gly Ser  Gly Gly Gly Gly Ser  Met Asn Pro
    1385                1390                1395


Ser Met  Lys Gln Lys Gln Glu  Glu Ile Lys Glu Asn  Ile Lys Asn
    1400                1405                1410


Ser Ser  Val Pro Arg Arg Thr  Leu Lys Met Ile Gln  Pro Ser Ala
    1415                1420                1425


Ser Gly  Ser Leu Val Gly Arg  Glu Asn Glu Leu Ser  Ala Gly Leu
    1430                1435                1440


Ser Lys  Arg Lys His Arg Asn  Asp His Leu Thr Ser  Thr Thr Ser
    1445                1450                1455


Ser Pro  Gly Val Ile Val Pro  Glu Ser Ser Glu Asn  Lys Asn Leu
    1460                1465                1470


Gly Gly  Val Thr Gln Glu Ser  Phe Asp Leu Met Ile  Lys Glu Asn
    1475                1480                1485


Pro Ser  Ser Gln Tyr Trp Lys  Glu Val Ala Glu Lys  Arg Arg Lys
    1490                1495                1500


Ala Leu
    1505
```

<210> 22
<211> 1486
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический polypeptide

<400> 22

```
Met Pro Ser Pro Ala Arg Pro Ala Leu Arg Ala Pro Ala Ser Ala Thr
1               5                  10                  15


Ser Gly Ser Arg Lys Arg Ala Arg Pro Pro Ala Ala Pro Gly Arg Asp
            20                  25                  30


Gln Ala Arg Pro Pro Ala Arg Arg Arg Leu Arg Leu Ser Val Asp Glu
        35                  40                  45


Val Ser Ser Pro Ser Thr Pro Glu Ala Pro Asp Ile Pro Ala Cys Pro
    50                  55                  60
```

```
Ser Pro Gly Gln Lys Ile Lys Lys Ser Thr Pro Ala Ala Gly Gln Pro
65                  70                  75                  80

Pro His Leu Thr Ser Ala Gln Asp Gln Asp Thr Ile Gly Gly Gly Gly
                85                  90                  95

Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Asp Lys Lys Tyr Ser
            100                 105                 110

Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr
        115                 120                 125

Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr
    130                 135                 140

Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp
145                 150                 155                 160

Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg
                165                 170                 175

Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe
            180                 185                 190

Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu
        195                 200                 205

Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile
    210                 215                 220

Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr
225                 230                 235                 240

Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp
                245                 250                 255

Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly
            260                 265                 270

His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp
        275                 280                 285

Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu
    290                 295                 300

Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala
305                 310                 315                 320
```

```
Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro
                325                 330                 335

Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu
            340                 345                 350

Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala
        355                 360                 365

Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu
    370                 375                 380

Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys
385                 390                 395                 400

Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr
                405                 410                 415

Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp
            420                 425                 430

Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln
        435                 440                 445

Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly
    450                 455                 460

Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys
465                 470                 475                 480

Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu
                485                 490                 495

Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp
            500                 505                 510

Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile
        515                 520                 525

Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu
    530                 535                 540

Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro
545                 550                 555                 560

Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu
                565                 570                 575
```

```
Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala
            580                 585                 590

Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu
        595                 600                 605

Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe
    610                 615                 620

Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met
625                 630                 635                 640

Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp
                645                 650                 655

Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu
            660                 665                 670

Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly
        675                 680                 685

Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu
    690                 695                 700

Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp
705                 710                 715                 720

Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu
                725                 730                 735

Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys
            740                 745                 750

Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu
        755                 760                 765

Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr
    770                 775                 780

Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met
785                 790                 795                 800

Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys
                805                 810                 815

Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn
            820                 825                 830
```

```
Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys
        835                 840                 845

Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn
    850                 855                 860

Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln
865                 870                 875                 880

Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu
                885                 890                 895

Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu
            900                 905                 910

Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met
        915                 920                 925

Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val
    930                 935                 940

Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn
945                 950                 955                 960

Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val
                965                 970                 975

Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu
            980                 985                 990

Leu Asn Ala Lys Leu Ile Thr Gln  Arg Lys Phe Asp Asn  Leu Thr Lys
        995                 1000                1005

Ala Glu  Arg Gly Gly Leu Ser  Glu Leu Asp Lys Ala  Gly Phe Ile
    1010                1015                1020

Lys Arg  Gln Leu Val Glu Thr  Arg Gln Ile Thr Lys  His Val Ala
    1025                1030                1035

Gln Ile  Leu Asp Ser Arg Met  Asn Thr Lys Tyr Asp  Glu Asn Asp
    1040                1045                1050

Lys Leu  Ile Arg Glu Val Lys  Val Ile Thr Leu Lys  Ser Lys Leu
    1055                1060                1065

Val Ser  Asp Phe Arg Lys Asp  Phe Gln Phe Tyr Lys  Val Arg Glu
    1070                1075                1080
```

```
Ile Asn  Asn Tyr His His Ala  His Asp Ala Tyr Leu  Asn Ala Val
    1085                1090                1095

Val Gly  Thr Ala Leu Ile Lys  Lys Tyr Pro Lys Leu  Glu Ser Glu
    1100                1105                1110

Phe Val  Tyr Gly Asp Tyr Lys  Val Tyr Asp Val Arg  Lys Met Ile
    1115                1120                1125

Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala Thr Ala  Lys Tyr Phe
    1130                1135                1140

Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys Thr Glu  Ile Thr Leu
    1145                1150                1155

Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu Ile Glu  Thr Asn Gly
    1160                1165                1170

Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly Arg Asp  Phe Ala Thr
    1175                1180                1185

Val Arg  Lys Val Leu Ser Met  Pro Gln Val Asn Ile  Val Lys Lys
    1190                1195                1200

Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys Glu Ser  Ile Leu Pro
    1205                1210                1215

Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg Lys Lys  Asp Trp Asp
    1220                1225                1230

Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro Thr Val  Ala Tyr Ser
    1235                1240                1245

Val Leu  Val Val Ala Lys Val  Glu Lys Gly Lys Ser  Lys Lys Leu
    1250                1255                1260

Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr Ile Met  Glu Arg Ser
    1265                1270                1275

Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu Glu Ala  Lys Gly Tyr
    1280                1285                1290

Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys Leu Pro  Lys Tyr Ser
    1295                1300                1305

Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg Met Leu  Ala Ser Ala
    1310                1315                1320
```

```
Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala Leu Pro  Ser Lys Tyr
    1325                1330                1335

Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr Glu Lys  Leu Lys Gly
    1340                1345                1350

Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu Phe Val  Glu Gln His
    1355                1360                1365

Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln Ile Ser  Glu Phe Ser
    1370                1375                1380

Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu Asp Lys  Val Leu Ser
    1385                1390                1395

Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile Arg Glu  Gln Ala Glu
    1400                1405                1410

Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn Leu Gly  Ala Pro Ala
    1415                1420                1425

Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp Arg Lys  Arg Tyr Thr
    1430                1435                1440

Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu Ile His  Gln Ser Ile
    1445                1450                1455

Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu Ser Gln  Leu Gly Gly
    1460                1465                1470

Asp Ser  Arg Ala Asp Pro Lys  Lys Lys Arg Lys Val  Arg
    1475                1480                1485
```

<210> 23
<211> 1504
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полипептид
<400> 23

```
Met Asn Pro Ser Met Lys Gln Lys Gln Glu Glu Ile Lys Glu Asn Ile
1               5                  10                  15

Lys Asn Ser Ser Val Pro Arg Arg Thr Leu Lys Met Ile Gln Pro Ser
            20                  25                  30

Ala Ser Gly Ser Leu Val Gly Arg Glu Asn Glu Leu Ser Ala Gly Leu
```

```
        35                  40                  45

Ser Lys Arg Lys His Arg Asn Asp His Leu Thr Ser Thr Thr Ser Ser
    50                  55                  60

Pro Gly Val Ile Val Pro Glu Ser Ser Glu Asn Lys Asn Leu Gly Gly
65                  70                  75                  80

Val Thr Gln Glu Ser Phe Asp Leu Met Ile Lys Glu Asn Pro Ser Ser
                85                  90                  95

Gln Tyr Trp Lys Glu Val Ala Glu Lys Arg Arg Lys Ala Leu Gly Gly
            100                 105                 110

Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Asp Lys Lys
        115                 120                 125

Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val
    130                 135                 140

Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly
145                 150                 155                 160

Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu
                165                 170                 175

Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala
            180                 185                 190

Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu
        195                 200                 205

Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg
    210                 215                 220

Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His
225                 230                 235                 240

Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr
                245                 250                 255

Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys
            260                 265                 270

Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe
        275                 280                 285

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp
```

```
    290                 295                 300

Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe
305                 310                 315                 320

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu
                325                 330                 335

Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln
            340                 345                 350

Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu
        355                 360                 365

Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu
    370                 375                 380

Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp
385                 390                 395                 400

Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala
                405                 410                 415

Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val
            420                 425                 430

Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg
        435                 440                 445

Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg
    450                 455                 460

Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys
465                 470                 475                 480

Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe
                485                 490                 495

Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu
            500                 505                 510

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr
        515                 520                 525

Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His
    530                 535                 540

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn
```

```
545                 550                 555                 560

Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val
                565                 570                 575

Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys
            580                 585                 590

Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys
        595                 600                 605

Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys
    610                 615                 620

Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu
625                 630                 635                 640

Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu
                645                 650                 655

Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile
            660                 665                 670

Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu
        675                 680                 685

Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile
    690                 695                 700

Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp
705                 710                 715                 720

Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn
                725                 730                 735

Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp
            740                 745                 750

Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp
        755                 760                 765

Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly
    770                 775                 780

Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly
785                 790                 795                 800

Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn
```

```
                805                 810                 815

Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile
            820                 825                 830

Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile
        835                 840                 845

Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr
    850                 855                 860

Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro
865                 870                 875                 880

Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys
                885                 890                 895

Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile
            900                 905                 910

Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr
        915                 920                 925

Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg
    930                 935                 940

Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr
945                 950                 955                 960

Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile
                965                 970                 975

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp
            980                 985                 990

Asn Val Pro Ser Glu Glu Val Val  Lys Lys Met Lys Asn  Tyr Trp Arg
        995                 1000                1005

Gln Leu  Leu Asn Ala Lys Leu  Ile Thr Gln Arg Lys  Phe Asp Asn
    1010                1015                1020

Leu Thr  Lys Ala Glu Arg Gly  Gly Leu Ser Glu Leu  Asp Lys Ala
    1025                1030                1035

Gly Phe  Ile Lys Arg Gln Leu  Val Glu Thr Arg Gln  Ile Thr Lys
    1040                1045                1050

His Val  Ala Gln Ile Leu Asp  Ser Arg Met Asn Thr  Lys Tyr Asp
```

```
    1055                1060                1065

Glu Asn  Asp Lys Leu Ile Arg  Glu Val Lys Val Ile  Thr Leu Lys
    1070                1075                1080

Ser Lys  Leu Val Ser Asp Phe  Arg Lys Asp Phe Gln  Phe Tyr Lys
    1085                1090                1095

Val Arg  Glu Ile Asn Asn Tyr  His His Ala His Asp  Ala Tyr Leu
    1100                1105                1110

Asn Ala  Val Val Gly Thr Ala  Leu Ile Lys Lys Tyr  Pro Lys Leu
    1115                1120                1125

Glu Ser  Glu Phe Val Tyr Gly  Asp Tyr Lys Val Tyr  Asp Val Arg
    1130                1135                1140

Lys Met  Ile Ala Lys Ser Glu  Gln Glu Ile Gly Lys  Ala Thr Ala
    1145                1150                1155

Lys Tyr  Phe Phe Tyr Ser Asn  Ile Met Asn Phe Phe  Lys Thr Glu
    1160                1165                1170

Ile Thr  Leu Ala Asn Gly Glu  Ile Arg Lys Arg Pro  Leu Ile Glu
    1175                1180                1185

Thr Asn  Gly Glu Thr Gly Glu  Ile Val Trp Asp Lys  Gly Arg Asp
    1190                1195                1200

Phe Ala  Thr Val Arg Lys Val  Leu Ser Met Pro Gln  Val Asn Ile
    1205                1210                1215

Val Lys  Lys Thr Glu Val Gln  Thr Gly Gly Phe Ser  Lys Glu Ser
    1220                1225                1230

Ile Leu  Pro Lys Arg Asn Ser  Asp Lys Leu Ile Ala  Arg Lys Lys
    1235                1240                1245

Asp Trp  Asp Pro Lys Lys Tyr  Gly Gly Phe Asp Ser  Pro Thr Val
    1250                1255                1260

Ala Tyr  Ser Val Leu Val Val  Ala Lys Val Glu Lys  Gly Lys Ser
    1265                1270                1275

Lys Lys  Leu Lys Ser Val Lys  Glu Leu Leu Gly Ile  Thr Ile Met
    1280                1285                1290

Glu Arg  Ser Ser Phe Glu Lys  Asn Pro Ile Asp Phe  Leu Glu Ala
```

```
    1295                1300                1305

Lys Gly  Tyr Lys Glu Val Lys  Lys Asp Leu Ile Ile  Lys Leu Pro
    1310                1315                1320

Lys Tyr  Ser Leu Phe Glu Leu  Glu Asn Gly Arg Lys  Arg Met Leu
    1325                1330                1335

Ala Ser  Ala Gly Glu Leu Gln  Lys Gly Asn Glu Leu  Ala Leu Pro
    1340                1345                1350

Ser Lys  Tyr Val Asn Phe Leu  Tyr Leu Ala Ser His  Tyr Glu Lys
    1355                1360                1365

Leu Lys  Gly Ser Pro Glu Asp  Asn Glu Gln Lys Gln  Leu Phe Val
    1370                1375                1380

Glu Gln  His Lys His Tyr Leu  Asp Glu Ile Ile Glu  Gln Ile Ser
    1385                1390                1395

Glu Phe  Ser Lys Arg Val Ile  Leu Ala Asp Ala Asn  Leu Asp Lys
    1400                1405                1410

Val Leu  Ser Ala Tyr Asn Lys  His Arg Asp Lys Pro  Ile Arg Glu
    1415                1420                1425

Gln Ala  Glu Asn Ile Ile His  Leu Phe Thr Leu Thr  Asn Leu Gly
    1430                1435                1440

Ala Pro  Ala Ala Phe Lys Tyr  Phe Asp Thr Thr Ile  Asp Arg Lys
    1445                1450                1455

Arg Tyr  Thr Ser Thr Lys Glu  Val Leu Asp Ala Thr  Leu Ile His
    1460                1465                1470

Gln Ser  Ile Thr Gly Leu Tyr  Glu Thr Arg Ile Asp  Leu Ser Gln
    1475                1480                1485

Leu Gly  Gly Asp Ser Arg Ala  Asp Pro Lys Lys Lys  Arg Lys Val
    1490                1495                1500

Arg
```

<210> 24
<211> 4458
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полинуклеотид

<400> 24
atggacaaga agtactccat tgggctcgat atcggcacaa acagcgtcgg ctgggccgtc 60

attacggacg agtacaaggt gccgagcaaa aaattcaaag ttctgggcaa taccgatcgc 120

cacagcataa agaagaacct cattggcgcc ctcctgttcg actccgggga gacggccgaa 180

gccacgcggc tcaaaagaac agcacggcgc agatataccc gcagaaagaa tcggatctgc 240

tacctgcagg agatctttag taatgagatg gctaaggtgg atgactcttt cttccatagg 300

ctggaggagt ccttttggt ggaggaggat aaaaagcacg agcgccaccc aatctttggc 360

aatatcgtgg acgaggtggc gtaccatgaa aagtacccaa ccatatatca tctgaggaag 420

aagcttgtag acagtactga taaggctgac ttgcggttga tctatctcgc gctggcgcat 480

atgatcaaat ttcggggaca cttcctcatc gaggggggacc tgaacccaga caacagcgat 540

gtcgacaaac tctttatcca actggttcag acttacaatc agcttttcga agagaacccg 600

atcaacgcat ccggagttga cgccaaagca atcctgagcg ctaggctgtc caaatcccgg 660

cggctcgaaa acctcatcgc acagctccct ggggagaaga agaacggcct gtttggtaat 720

cttatcgccc tgtcactcgg gctgaccccc aactttaaat ctaacttcga cctggccgaa 780

gatgccaagc ttcaactgag caaagacacc tacgatgatg atctcgacaa tctgctggcc 840

cagatcggcg accagtacgc agacctttt ttggcggcaa agaacctgtc agacgccatt 900

ctgctgagtg atattctgcg agtgaacacg gagatcacca aagctccgct gagcgctagt 960

atgatcaagc gctatgatga gcaccaccaa gacttgactt tgctgaaggc ccttgtcaga 1020

cagcaactgc ctgagaagta caaggaaatt ttcttcgatc agtctaaaaa tggctacgcc 1080

ggatacattg acggcggagc aagccaggag gaattttaca aatttattaa gcccatcttg 1140

gaaaaaatgg acggcaccga ggagctgctg gtaaagctta acagagaaga tctgttgcgc 1200

aaacagcgca ctttcgacaa tggaagcatc ccccaccaga ttcacctggg cgaactgcac 1260

gctatcctca ggcggcaaga ggatttctac ccctttttga aagataacag ggaaaagatt 1320

gagaaaatcc tcacatttcg gataccctac tatgtaggcc ccctcgcccg gggaaattcc 1380

agattcgcgt ggatgactcg caaatcagaa gagaccatca ctccctggaa cttcgaggaa 1440

gtcgtggata agggggcctc tgcccagtcc ttcatcgaaa ggatgactaa ctttgataaa 1500

aatctgccta acgaaaaggt gcttcctaaa cactctctgc tgtacgagta cttcacagtt 1560

tataacgagc tcaccaaggt caaatacgtc acagaaggga tgagaaagcc agcattcctg 1620

tctggagagc agaagaaagc tatcgtggac ctcctcttca agacgaaccg gaaagttacc 1680

gtgaaacagc tcaaagaaga ctatttcaaa aagattgaat gtttcgactc tgttgaaatc 1740

agcggagtgg aggatcgctt caacgcatcc ctgggaacgt atcacgatct cctgaaaatc 1800

attaaagaca aggacttcct ggacaatgag gagaacgagg acattcttga ggacattgtc 1860

ctcaccctta cgttgtttga agatagggag atgattgaag aacgcttgaa aacttacgct 1920

catctcttcg acgacaaagt catgaaacag ctcaagaggc gccgatatac aggatggggg 1980

cggctgtcaa gaaaactgat caatgggatc cgagacaagc agagtggaaa gacaatcctg 2040

gattttctta agtccgatgg atttgccaac cggaacttca tgcagttgat ccatgatgac 2100

tctctcacct ttaaggagga catccagaaa gcacaagttt ctggccaggg ggacagtctt 2160

cacgagcaca tcgctaatct tgcaggtagc ccagctatca aaaagggaat actgcagacc 2220

gttaaggtcg tggatgaact cgtcaaagta atgggaaggc ataagcccga gaatatcgtt 2280

atcgagatgg cccgagagaa ccaaactacc cagaagggac agaagaacag tagggaaagg 2340

atgaagagga ttgaagaggg tataaaagaa ctggggtccc aaatccttaa ggaacaccca 2400

gttgaaaaca cccagcttca gaatgagaag ctctacctgt actacctgca gaacggcagg 2460

gacatgtacg tggatcagga actggacatc aatcggctct ccgactacga cgtggatcat 2520

atcgtgcccc agtcttttct caaagatgat tctattgata ataaagtgtt gacaagatcc 2580

gataaaaata gagggaagag tgataacgtc ccctcagaag aagttgtcaa gaaaatgaaa 2640

aattattggc ggcagctgct gaacgccaaa ctgatcacac aacggaagtt cgataatctg 2700

actaaggctg aacgaggtgg cctgtctgag ttggataaag ccggcttcat caaaaggcag 2760

cttgttgaga cacgccagat caccaagcac gtggcccaaa ttctcgattc acgcatgaac 2820

accaagtacg atgaaaatga caaactgatt cgagaggtga aagttattac tctgaagtct 2880

aagctggtct cagatttcag aaaggacttt cagttttata aggtgagaga gatcaacaat 2940

taccaccatg cgcatgatgc ctacctgaat gcagtggtag gcactgcact tatcaaaaaa 3000

tatcccaagc ttgaatctga atttgtttac ggagactata aagtgtacga tgttaggaaa 3060

atgatcgcaa agtctgagca ggaaataggc aaggccaccg ctaagtactt cttttacagc 3120

aatattatga atttttttcaa gaccgagatt acactggcca atggagagat tcggaagcga 3180

ccacttatcg aaacaaacgg agaaacagga gaaatcgtgt gggacaaggg tagggatttc 3240

gcgacagtcc ggaaggtcct gtccatgccg caggtgaaca tcgttaaaaa gaccgaagta 3300

cagaccggag gcttctccaa ggaaagtatc ctcccgaaaa ggaacagcga caagctgatc 3360

gcacgcaaaa aagattggga ccccaagaaa tacggcggat tcgattctcc tacagtcgct 3420

tacagtgtac tggttgtggc caaagtggag aaagggaagt ctaaaaaact caaaagcgtc 3480

aaggaactgc tgggcatcac aatcatggag cgatcaagct tcgaaaaaaa ccccatcgac 3540

tttctcgagg cgaaaggata taaagaggtc aaaaaagacc tcatcattaa gcttcccaag 3600

tactctctct ttgagcttga aaacggccgg aaacgaatgc tcgctagtgc gggcgagctg 3660

cagaaaggta acgagctggc actgccctct aaatacgtta atttcttgta tctggccagc 3720

cactatgaaa agctcaaagg gtctcccgaa gataatgagc agaagcagct gttcgtggaa 3780

caacacaaac actaccttga tgagatcatc gagcaaataa gcgaattctc caaaagagtg 3840

atcctcgccg acgctaacct cgataaggtg ctttctgctt acaataagca cagggataag 3900

cccatcaggg agcaggcaga aaacattatc cacttgttta ctctgaccaa cttgggcgcg 3960

cctgcagcct tcaagtactt cgacaccacc atagacagaa agcggtacac ctctacaaag 4020

gaggtcctgg acgccacact gattcatcag tcaattacgg ggctctatga aacaagaatc 4080

gacctctctc agctcggtgg agacagcagg gctgacccca agaagaagag gaaggtgagg 4140

ggaggcggag gatctggcgg aggcggaagt ggcggagggg gcagcccatc tcctgccaga 4200

cctgctctga gagcccctgc ctctgccaca agcggcagca gaaagagagc cagacctcct 4260

gccgcccctg gcagagatca ggctagacct ccagctcggc ggagactgag actgagcgtg 4320

gacgaggtgt ccagccctag cacacctgag gcccctgata tccccgcctg tcctagccct 4380

ggccagaaga tcaagaagtc cacccctgcc gccggacagc ctcctcatct gacatctgcc 4440

caggaccagg acaccatc 4458

<210> 25
<211> 4515
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полинуклеотид

<400> 25
atggacaaga agtactccat tgggctcgat atcggcacaa acagcgtcgg ctgggccgtc 60

attacggacg agtacaaggt gccgagcaaa aaattcaaag ttctgggcaa taccgatcgc 120

cacagcataa agaagaacct cattggcgcc ctcctgttcg actccgggga gacggccgaa 180

gccacgcggc tcaaaagaac agcacggcgc agatataccc gcagaaagaa tcggatctgc 240

tacctgcagg agatctttag taatgagatg gctaaggtgg atgactcttt cttccatagg 300

ctggaggagt ccttttggt ggaggaggat aaaaagcacg agcgccaccc aatctttggc 360

aatatcgtgg acgaggtggc gtaccatgaa aagtacccaa ccatatatca tctgaggaag 420

aagcttgtag acagtactga taaggctgac ttgcggttga tctatctcgc gctggcgcat 480

atgatcaaat ttcggggaca cttcctcatc gaggggggacc tgaacccaga caacagcgat 540

gtcgacaaac tctttatcca actggttcag acttacaatc agcttttcga agagaacccg 600

atcaacgcat ccggagttga cgccaaagca atcctgagcg ctaggctgtc caaatcccgg 660

cggctcgaaa acctcatcgc acagctccct ggggagaaga agaacggcct gtttggtaat 720

cttatcgccc tgtcactcgg gctgaccccc aactttaaat ctaacttcga cctggccgaa 780

gatgccaagc ttcaactgag caaagacacc tacgatgatg atctcgacaa tctgctggcc 840

cagatcggcg accagtacgc agacctttt ttggcggcaa agaacctgtc agacgccatt 900

ctgctgagtg atattctgcg agtgaacacg gagatcacca aagctccgct gagcgctagt 960

atgatcaagc gctatgatga gcaccaccaa gacttgactt tgctgaaggc ccttgtcaga 1020

cagcaactgc ctgagaagta caaggaaatt ttcttcgatc agtctaaaaa tggctacgcc 1080

ggatacattg acggcggagc aagccaggag gaattttaca aatttattaa gcccatcttg 1140

gaaaaaatgg acggcaccga ggagctgctg gtaaagctta acagagaaga tctgttgcgc 1200

aaacagcgca ctttcgacaa tggaagcatc ccccaccaga ttcacctggg cgaactgcac 1260

gctatcctca ggcggcaaga ggatttctac ccctttttga aagataacag ggaaaagatt 1320

gagaaaatcc tcacatttcg gataccctac tatgtaggcc ccctcgcccg gggaaattcc 1380

agattcgcgt ggatgactcg caaatcagaa gagaccatca ctccctggaa cttcgaggaa 1440

gtcgtggata agggggcctc tgcccagtcc ttcatcgaaa ggatgactaa ctttgataaa 1500

aatctgccta acgaaaaggt gcttcctaaa cactctctgc tgtacgagta cttcacagtt 1560

tataacgagc tcaccaaggt caaatacgtc acagaaggga tgagaaagcc agcattcctg 1620

tctggagagc agaagaaagc tatcgtggac ctcctcttca agacgaaccg gaaagttacc 1680

gtgaaacagc tcaaagaaga ctatttcaaa aagattgaat gtttcgactc tgttgaaatc 1740

agcggagtgg aggatcgctt caacgcatcc ctgggaacgt atcacgatct cctgaaaatc 1800

attaaagaca aggacttcct ggacaatgag gagaacgagg acattcttga ggacattgtc 1860

ctcaccctta cgttgtttga agatagggag atgattgaag aacgcttgaa aacttacgct 1920

catctcttcg acgacaaagt catgaaacag ctcaagaggc gccgatatac aggatggggg 1980

cggctgtcaa gaaaactgat caatgggatc cgagacaagc agagtggaaa gacaatcctg 2040

gattttctta agtccgatgg atttgccaac cggaacttca tgcagttgat ccatgatgac 2100

tctctcacct ttaaggagga catccagaaa gcacaagttt ctggccaggg ggacagtctt 2160

cacgagcaca tcgctaatct tgcaggtagc ccagctatca aaaagggaat actgcagacc 2220

gttaaggtcg tggatgaact cgtcaaagta atgggaaggc ataagcccga gaatatcgtt 2280

atcgagatgg cccgagagaa ccaaactacc cagaagggac agaagaacag tagggaaagg 2340

atgaagagga ttgaagaggg tataaaagaa ctggggtccc aaatccttaa ggaacaccca 2400

gttgaaaaca cccagcttca gaatgagaag ctctacctgt actacctgca gaacggcagg 2460

gacatgtacg tggatcagga actggacatc aatcggctct ccgactacga cgtggatcat 2520

atcgtgcccc agtcttttct caaagatgat tctattgata ataaagtgtt gacaagatcc 2580

gataaaaata gagggaagag tgataacgtc ccctcagaag aagttgtcaa gaaaatgaaa 2640

aattattggc ggcagctgct gaacgccaaa ctgatcacac aacggaagtt cgataatctg 2700

actaaggctg aacgaggtgg cctgtctgag ttggataaag ccggcttcat caaaaggcag 2760

cttgttgaga cacgccagat caccaagcac gtggcccaaa ttctcgattc acgcatgaac 2820

accaagtacg atgaaaatga caaactgatt cgagaggtga aagttattac tctgaagtct 2880

aagctggtct cagatttcag aaaggacttt cagttttata aggtgagaga gatcaacaat 2940

taccaccatg cgcatgatgc ctacctgaat gcagtggtag gcactgcact tatcaaaaaa 3000

tatcccaagc ttgaatctga atttgtttac ggagactata aagtgtacga tgttaggaaa 3060

atgatcgcaa agtctgagca ggaaataggc aaggccaccg ctaagtactt cttttacagc 3120

aatattatga atttttttcaa gaccgagatt acactggcca atggagagat tcggaagcga 3180

ccacttatcg aaacaaacgg agaaacagga gaaatcgtgt gggacaaggg tagggatttc 3240

gcgacagtcc ggaaggtcct gtccatgccg caggtgaaca tcgttaaaaa gaccgaagta 3300

cagaccggag gcttctccaa ggaaagtatc ctcccgaaaa ggaacagcga caagctgatc 3360

gcacgcaaaa aagattggga ccccaagaaa tacggcggat tcgattctcc tacagtcgct 3420

tacagtgtac tggttgtggc caaagtggag aaagggaagt ctaaaaaact caaaagcgtc 3480

aaggaactgc tgggcatcac aatcatggag cgatcaagct tcgaaaaaaa ccccatcgac 3540

tttctcgagg cgaaaggata taaagaggtc aaaaaagacc tcatcattaa gcttcccaag 3600

tactctctct ttgagcttga aaacggccgg aaacgaatgc tcgctagtgc gggcgagctg 3660

cagaaaggta acgagctggc actgccctct aaatacgtta atttcttgta tctggccagc 3720

cactatgaaa agctcaaagg gtctcccgaa gataatgagc agaagcagct gttcgtggaa 3780

caacacaaac actaccttga tgagatcatc gagcaaataa gcgaattctc caaaagagtg 3840

atcctcgccg acgctaacct cgataaggtg ctttctgctt acaataagca cagggataag 3900

cccatcaggg agcaggcaga aaacattatc cacttgttta ctctgaccaa cttgggcgcg 3960

cctgcagcct tcaagtactt cgacaccacc atagacagaa agcggtacac ctctacaaag 4020

gaggtcctgg acgccacact gattcatcag tcaattacgg ggctctatga aacaagaatc 4080

gacctctctc agctcggtgg agacagcagg gctgacccca agaagaagag gaaggtgagg 4140

ggaggcggag gatctggcgg aggcggaagt ggcggagggg gcagcatgaa ccctagcatg 4200

aagcagaagc aggaagagat caaagagaac atcaagaaca gcagcgtgcc cagacggacc 4260

ctgaagatga tccagcctag cgccagcggc agcctcgtgg gcagagagaa tgaactgtct 4320

gccggcctga gcaagcggaa gcacagaaac gaccacctga ccagcaccac cagcagccct 4380

ggcgtgatcg tgcctgagag cagcgagaac aagaacctgg gcggcgtgac ccaggaatcc 4440

ttcgacctga tgatcaaaga aaaccccagc agccagtatt ggaaagaggt ggccgagaag 4500

cggcggaagg ccctg 4515

<210> 26

<211> 4458
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полинуклеотид

<400> 26
atgccatctc ctgccagacc tgctctgaga gcccctgcct ctgccacaag cggcagcaga 60

aagagagcca gacctcctgc cgcccctggc agagatcagg ctagacctcc agctcggcgg 120

agactgagac tgagcgtgga cgaggtgtcc agccctagca cacctgaggc ccctgatatc 180

cccgcctgtc ctagccctgg ccagaagatc aagaagtcca cccctgccgc cggacagcct 240

cctcatctga catctgccca ggaccaggac accatcggag gcggaggatc tggcggaggc 300

ggaagtggcg gagggggcag cgacaagaag tactccattg ggctcgatat cggcacaaac 360

agcgtcggct gggccgtcat tacggacgag tacaaggtgc cgagcaaaaa attcaaagtt 420

ctgggcaata ccgatcgcca cagcataaag aagaacctca ttggcgccct cctgttcgac 480

tccggggaga cggccgaagc cacgcggctc aaaagaacag cacggcgcag atatacccgc 540

agaaagaatc ggatctgcta cctgcaggag atctttagta atgagatggc taaggtggat 600

gactctttct tccataggct ggaggagtcc tttttggtgg aggaggataa aaagcacgag 660

cgccacccaa tctttggcaa tatcgtggac gaggtggcgt accatgaaaa gtacccaacc 720

atatatcatc tgaggaagaa gcttgtagac agtactgata aggctgactt gcggttgatc 780

tatctcgcgc tggcgcatat gatcaaattt cggggacact tcctcatcga gggggacctg 840

aacccagaca acagcgatgt cgacaaactc tttatccaac tggttcagac ttacaatcag 900

cttttcgaag agaacccgat caacgcatcc ggagttgacg ccaaagcaat cctgagcgct 960

aggctgtcca aatcccggcg gctcgaaaac ctcatcgcac agctccctgg ggagaagaag 1020

aacggcctgt ttggtaatct tatcgccctg tcactcgggc tgacccccaa ctttaaatct 1080

aacttcgacc tggccgaaga tgccaagctt caactgagca aagacaccta cgatgatgat 1140

ctcgacaatc tgctggccca gatcggcgac cagtacgcag acctttttt ggcggcaaag 1200

aacctgtcag acgccattct gctgagtgat attctgcgag tgaacacgga gatcaccaaa 1260

gctccgctga gcgctagtat gatcaagcgc tatgatgagc accaccaaga cttgactttg 1320

ctgaaggccc ttgtcagaca gcaactgcct gagaagtaca aggaaatttt cttcgatcag 1380

tctaaaaatg gctacgccgg atacattgac ggcggagcaa gccaggagga attttacaaa 1440

tttattaagc ccatcttgga aaaaatggac ggcaccgagg agctgctggt aaagcttaac 1500

agagaagatc tgttgcgcaa acagcgcact ttcgacaatg gaagcatccc ccaccagatt 1560

cacctgggcg aactgcacgc tatcctcagg cggcaagagg atttctaccc ctttttgaaa 1620

gataacaggg aaaagattga gaaaatcctc acatttcgga taccctacta tgtaggcccc 1680

ctcgcccggg gaaattccag attcgcgtgg atgactcgca aatcagaaga gaccatcact 1740

ccctggaact tcgaggaagt cgtggataag ggggcctctg cccagtcctt catcgaaagg 1800

atgactaact ttgataaaaa tctgcctaac gaaaaggtgc ttcctaaaca ctctctgctg 1860

tacgagtact tcacagttta taacgagctc accaaggtca aatacgtcac agaagggatg 1920

agaaagccag cattcctgtc tggagagcag aagaaagcta tcgtggacct cctcttcaag 1980

acgaaccgga aagttaccgt gaaacagctc aaagaagact atttcaaaaa gattgaatgt 2040

ttcgactctg ttgaaatcag cggagtggag gatcgcttca acgcatccct gggaacgtat 2100

cacgatctcc tgaaaatcat taaagacaag gacttcctgg acaatgagga gaacgaggac 2160

attcttgagg acattgtcct cacccttacg ttgtttgaag atagggagat gattgaagaa 2220

cgcttgaaaa cttacgctca tctcttcgac gacaaagtca tgaaacagct caagaggcgc 2280

cgatatacag gatgggggcg gctgtcaaga aaactgatca atgggatccg agacaagcag 2340

agtggaaaga caatcctgga ttttcttaag tccgatggat ttgccaaccg gaacttcatg 2400

cagttgatcc atgatgactc tctcaccttt aaggaggaca tccagaaagc acaagtttct 2460

ggccaggggg acagtcttca cgagcacatc gctaatcttg caggtagccc agctatcaaa 2520

aagggaatac tgcagaccgt taaggtcgtg gatgaactcg tcaaagtaat gggaaggcat 2580

aagcccgaga atatcgttat cgagatggcc cgagagaacc aaactaccca gaagggacag 2640

aagaacagta gggaaaggat gaagaggatt gaagagggta taaaagaact ggggtcccaa 2700

atccttaagg aacacccagt tgaaaacacc cagcttcaga atgagaagct ctacctgtac 2760

tacctgcaga acggcaggga catgtacgtg gatcaggaac tggacatcaa tcggctctcc 2820

gactacgacg tggatcatat cgtgccccag tcttttctca aagatgattc tattgataat 2880

aaagtgttga caagatccga taaaaataga gggaagagtg ataacgtccc ctcagaagaa 2940

gttgtcaaga aaatgaaaaa ttattggcgg cagctgctga acgccaaact gatcacacaa 3000

cggaagttcg ataatctgac taaggctgaa cgaggtggcc tgtctgagtt ggataaagcc 3060

ggcttcatca aaaggcagct tgttgagaca cgccagatca ccaagcacgt ggcccaaatt 3120

ctcgattcac gcatgaacac caagtacgat gaaaatgaca aactgattcg agaggtgaaa 3180

gttattactc tgaagtctaa gctggtctca gatttcagaa aggactttca gttttataag 3240

gtgagagaga tcaacaatta ccaccatgcg catgatgcct acctgaatgc agtggtaggc 3300

actgcactta tcaaaaaata tcccaagctt gaatctgaat ttgtttacgg agactataaa 3360

gtgtacgatg ttaggaaaat gatcgcaaag tctgagcagg aaataggcaa ggccaccgct 3420

aagtacttct tttacagcaa tattatgaat tttttcaaga ccgagattac actggccaat 3480

ggagagattc ggaagcgacc acttatcgaa acaaacggag aaacaggaga aatcgtgtgg 3540

gacaagggta gggatttcgc gacagtccgg aaggtcctgt ccatgccgca ggtgaacatc 3600

gttaaaaaga ccgaagtaca gaccggaggc ttctccaagg aaagtatcct cccgaaaagg 3660

aacagcgaca agctgatcgc acgcaaaaaa gattgggacc ccaagaaata cggcggattc 3720

gattctccta cagtcgctta cagtgtactg gttgtggcca aagtggagaa agggaagtct 3780

aaaaaactca aaagcgtcaa ggaactgctg ggcatcacaa tcatggagcg atcaagcttc 3840

gaaaaaaacc ccatcgactt tctcgaggcg aaaggatata aagaggtcaa aaaagacctc 3900

atcattaagc ttcccaagta ctctctcttt gagcttgaaa acggccggaa acgaatgctc 3960

gctagtgcgg gcgagctgca gaaaggtaac gagctggcac tgccctctaa atacgttaat 4020

ttcttgtatc tggccagcca ctatgaaaag ctcaaagggt ctcccgaaga taatgagcag 4080

aagcagctgt tcgtggaaca acacaaacac taccttgatg agatcatcga gcaaataagc 4140

gaattctcca aaagagtgat cctcgccgac gctaacctcg ataaggtgct ttctgcttac 4200

aataagcaca gggataagcc catcagggag caggcagaaa acattatcca cttgtttact 4260

ctgaccaact tgggcgcgcc tgcagccttc aagtacttcg acaccaccat agacagaaag 4320

cggtacacct ctacaaagga ggtcctggac gccacactga ttcatcagtc aattacgggg 4380

ctctatgaaa caagaatcga cctctctcag ctcggtggag acagcagggc tgaccccaag 4440

aagaagagga aggtgagg 4458

<210> 27
<211> 4512
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полинуклеотид

<400> 27
atgaacccta gcatgaagca gaagcaggaa gagatcaaag agaacatcaa gaacagcagc 60

gtgcccagac ggaccctgaa gatgatccag cctagcgcca gcggcagcct cgtgggcaga 120

gagaatgaac tgtctgccgg cctgagcaag cggaagcaca gaaacgacca cctgaccagc 180

accaccagca gccctggcgt gatcgtgcct gagagcagcg agaacaagaa cctgggcggc 240

gtgacccagg aatccttcga cctgatgatc aaagaaaacc ccagcagcca gtattggaaa 300

gaggtggccg agaagcggcg gaaggccctg ggaggcggag gatctggcgg aggcggaagt 360

ggcggagggg gcagcgacaa gaagtactcc attgggctcg atatcggcac aaacagcgtc 420

ggctgggccg tcattacgga cgagtacaag gtgccgagca aaaaattcaa agttctgggc 480

aataccgatc gccacagcat aaagaagaac ctcattggcg ccctcctgtt cgactccggg 540

gagacggccg aagccacgcg gctcaaaaga acagcacggc gcagatatac ccgcagaaag 600

aatcggatct gctacctgca ggagatcttt agtaatgaga tggctaaggt ggatgactct 660

ttcttccata ggctggagga gtccttttg gtggaggagg ataaaaagca cgagcgccac 720

ccaatctttg gcaatatcgt ggacgaggtg gcgtaccatg aaaagtaccc aaccatatat 780

catctgagga agaagcttgt agacagtact gataaggctg acttgcggtt gatctatctc 840

gcgctggcgc atatgatcaa atttcgggga cacttcctca tcgaggggga cctgaaccca 900

gacaacagcg atgtcgacaa actctttatc caactggttc agacttacaa tcagcttttc 960

gaagagaacc cgatcaacgc atccggagtt gacgccaaag caatcctgag cgctaggctg 1020

tccaaatccc ggcggctcga aaacctcatc gcacagctcc ctggggagaa gaagaacggc 1080

ctgtttggta atcttatcgc cctgtcactc gggctgaccc ccaactttaa atctaacttc 1140

gacctggccg aagatgccaa gcttcaactg agcaaagaca cctacgatga tgatctcgac 1200

aatctgctgg cccagatcgg cgaccagtac gcagaccttt ttttggcggc aaagaacctg 1260

tcagacgcca ttctgctgag tgatattctg cgagtgaaca cggagatcac caaagctccg 1320

ctgagcgcta gtatgatcaa gcgctatgat gagcaccacc aagacttgac tttgctgaag 1380

gcccttgtca gacagcaact gcctgagaag tacaaggaaa ttttcttcga tcagtctaaa 1440

aatggctacg ccggatacat tgacggcgga gcaagccagg aggaatttta caaatttatt 1500

aagcccatct tggaaaaaat ggacggcacc gaggagctgc tggtaaagct taacagagaa 1560

gatctgttgc gcaaacagcg cactttcgac aatggaagca tcccccacca gattcacctg 1620

ggcgaactgc acgctatcct caggcggcaa gaggatttct acccctttt gaaagataac 1680

agggaaaaga ttgagaaaat cctcacattt cggataccct actatgtagg ccccctcgcc 1740

cggggaaatt ccagattcgc gtggatgact cgcaaatcag aagagaccat cactccctgg 1800

aacttcgagg aagtcgtgga taaggggcc tctgcccagt ccttcatcga aaggatgact 1860

aactttgata aaaatctgcc taacgaaaag gtgcttccta aacactctct gctgtacgag 1920

tacttcacag tttataacga gctcaccaag gtcaaatacg tcacagaagg gatgagaaag 1980

ccagcattcc tgtctggaga gcagaagaaa gctatcgtgg acctcctctt caagacgaac 2040

cggaaagtta ccgtgaaaca gctcaaagaa gactatttca aaaagattga atgtttcgac 2100

tctgttgaaa tcagcggagt ggaggatcgc ttcaacgcat ccctgggaac gtatcacgat 2160

ctcctgaaaa tcattaaaga caaggacttc ctggacaatg aggagaacga ggacattctt 2220

gaggacattg tcctcaccct tacgttgttt gaagataggg agatgattga agaacgcttg 2280

aaaacttacg ctcatctctt cgacgacaaa gtcatgaaac agctcaagag gcgccgatat 2340

acaggatggg ggcggctgtc aagaaaactg atcaatggga tccgagacaa gcagagtgga 2400

aagacaatcc tggattttct taagtccgat ggatttgcca accggaactt catgcagttg 2460

atccatgatg actctctcac ctttaaggag gacatccaga aagcacaagt ttctggccag 2520

ggggacagtc ttcacgagca catcgctaat cttgcaggta gcccagctat caaaaaggga 2580

atactgcaga ccgttaaggt cgtggatgaa ctcgtcaaag taatgggaag gcataagccc 2640

gagaatatcg ttatcgagat ggcccgagag aaccaaacta cccagaaggg acagaagaac 2700

agtagggaaa ggatgaagag gattgaagag ggtataaaag aactggggtc ccaaatcctt 2760

aaggaacacc cagttgaaaa cacccagctt cagaatgaga agctctacct gtactacctg 2820

cagaacggca gggacatgta cgtggatcag gaactggaca tcaatcggct ctccgactac 2880

gacgtggatc atatcgtgcc ccagtctttt ctcaaagatg attctattga taataaagtg 2940

ttgacaagat ccgataaaaa tagagggaag agtgataacg tcccctcaga agaagttgtc 3000

aagaaaatga aaaattattg gcggcagctg ctgaacgcca aactgatcac acaacggaag 3060

ttcgataatc tgactaaggc tgaacgaggt ggcctgtctg agttggataa agccggcttc 3120

atcaaaaggc agcttgttga gacacgccag atcaccaagc acgtggccca aattctcgat 3180

tcacgcatga acaccaagta cgatgaaaat gacaaactga ttcgagaggt gaaagttatt 3240

actctgaagt ctaagctggt ctcagatttc agaaaggact ttcagtttta taaggtgaga 3300

gagatcaaca attaccacca tgcgcatgat gcctacctga atgcagtggt aggcactgca 3360

cttatcaaaa aatatcccaa gcttgaatct gaatttgttt acggagacta taaagtgtac 3420

gatgttagga aaatgatcgc aaagtctgag caggaaatag gcaaggccac cgctaagtac 3480

ttcttttaca gcaatattat gaatttttc aagaccgaga ttacactggc caatggagag 3540

attcggaagc gaccacttat cgaaacaaac ggagaaacag gagaaatcgt gtgggacaag 3600

ggtagggatt tcgcgacagt ccggaaggtc ctgtccatgc cgcaggtgaa catcgttaaa 3660

aagaccgaag tacagaccgg aggcttctcc aaggaaagta tcctcccgaa aaggaacagc 3720

gacaagctga tcgcacgcaa aaaagattgg gaccccaaga aatacggcgg attcgattct 3780

cctacagtcg cttacagtgt actggttgtg gccaaagtgg agaaagggaa gtctaaaaaa 3840

ctcaaaagcg tcaaggaact gctgggcatc acaatcatgg agcgatcaag cttcgaaaaa 3900

aaccccatcg actttctcga ggcgaaagga tataaagagg tcaaaaaaga cctcatcatt 3960

aagcttccca agtactctct ctttgagctt gaaaacggcc ggaaacgaat gctcgctagt 4020

gcgggcgagc tgcagaaagg taacgagctg gcactgccct ctaaatacgt taatttcttg 4080

tatctggcca gccactatga aaagctcaaa gggtctcccg aagataatga gcagaagcag 4140

ctgttcgtgg aacaacacaa acactacctt gatgagatca tcgagcaaat aagcgaattc 4200

tccaaaagag tgatcctcgc cgacgctaac ctcgataagg tgctttctgc ttacaataag 4260

cacagggata agcccatcag ggagcaggca gaaaacatta tccacttgtt tactctgacc 4320

aacttgggcg cgcctgcagc cttcaagtac ttcgacacca ccatagacag aaagcggtac 4380

acctctacaa aggaggtcct ggacgccaca ctgattcatc agtcaattac ggggctctat 4440

gaaacaagaa tcgacctctc tcagctcggt ggagacagca gggctgaccc caagaagaag 4500

aggaaggtga gg 4512

<210> 28
<211> 20
<212> ДНК
<213> Mus musculus

<400> 28
gaagaaggtt cgagatctca 20

<210> 29
<211> 23
<212> ДНК
<213> Mus musculus

<400> 29
gaagaaggtt cgagatctca agg 23

<210> 30
<211> 19
<212> ДНК
<213> Неизвестная

<220>
<223> Описание неизвестной:
Последовательность сайта Auto1

<400> 30
gttcgagatc tcagggaat 19

<210> 31
<211> 19
<212> ДНК
<213> Неизвестная

<220>
<223> Описание неизвестной:
Последовательность сайта Auto2

<400> 31
gttcgagatc tcagggttt 19

<210> 32
<211> 230
<212> ДНК
<213> Неизвестная

<220>
<223> Описание неизвестной:
H1 последовательность дикого типа

<400> 32
ggaattcgaa cgctgacgtc atcaacccgc tccaaggaat cgcgggccca gtgtcactag 60

gcgggaacac ccagcgcgcg tgcgccctgg caggaagatg gctgtgaggg acaggggagt 120

ggcgccctgc aatatttgca tgtcgctatg tgttctggga aatcaccata aacgtgaaat 180

gtctttggat ttgggaatct tataagttct gtatgagacc actttttccc 230

<210> 33
<211> 230
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический
полинуклеотид

<400> 33
ggaattcgaa cgctgacgtc atcaacccgc tccaaggaat cgcgggccca gtgtcactag 60

gcgggaacac ccagcgcgcg tgcgccctgg caggaagatg gctgtgaggg acaggggagt 120

ggcgccctgc aatatttgca tgtcgctatg tgttctggga aatcaccata aacgtgaaag 180

ttcgagatct cagggaatct tataagttct gtatgagacc acttttttccc 230

<210> 34
<211> 230
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический
полинуклеотид

<400> 34
ggaattcgaa cgctgacgtc atcaacccgc tccaaggaat cgcgggccca gtgtcactag 60

gcgggaacac ccagcgcgcg tgcgccctgg caggaagatg gctgtgaggg acaggggagt 120

ggcgccctgc aatatttgca tgtcgctatg tgttctggga aatcaccata aacgtgaaat 180

gtctttggat ttgggaatct tataagtgtt cgagatctca gggttttccc 230

<210> 35
<211> 22
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический
пептид

<400> 35
Gly Ser Gly Ala Thr Asn Phe Ser Leu Leu Lys Gln Ala Gly Asp Val
1 5 10 15

Glu Glu Asn Pro Gly Pro
20

<210> 36
<211> 21
<212> Белок
<213> Искусственная последовательность

<220>

<223> Описание искусственной последовательности: Синтетический
пептид
<400> 36
Gly Ser Gly Glu Gly Arg Gly Ser Leu Leu Thr Cys Gly Asp Val Glu
1 5 10 15

Glu Asn Pro Gly Pro
20

<210> 37
<211> 23
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический
пептид

<400> 37
Gly Ser Gly Gln Cys Thr Asn Tyr Ala Leu Leu Lys Leu Ala Gly Asp
1 5 10 15

Val Glu Ser Asn Pro Gly Pro
20

<210> 38
<211> 25
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический
пептид

<400> 38
Gly Ser Gly Val Lys Gln Thr Leu Asn Phe Asp Leu Leu Lys Leu Ala
1 5 10 15

Gly Asp Val Glu Ser Asn Pro Gly Pro
20 25

<210> 39
<211> 1455
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический
полипептид

<400> 39
Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu
1 5 10 15

Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp
20 25 30

```
Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
        35                  40                  45

Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu
    50                  55                  60

Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly Val Asp Lys Lys
65                  70                  75                  80

Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val
                85                  90                  95

Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly
            100                 105                 110

Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu
        115                 120                 125

Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala
    130                 135                 140

Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu
145                 150                 155                 160

Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg
                165                 170                 175

Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His
            180                 185                 190

Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr
        195                 200                 205

Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys
    210                 215                 220

Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe
225                 230                 235                 240

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp
                245                 250                 255

Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe
            260                 265                 270

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu
        275                 280                 285
```

```
Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln
    290                 295                 300

Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu
305                 310                 315                 320

Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu
                325                 330                 335

Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp
            340                 345                 350

Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala
        355                 360                 365

Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val
    370                 375                 380

Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg
385                 390                 395                 400

Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg
                405                 410                 415

Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys
            420                 425                 430

Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe
        435                 440                 445

Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu
    450                 455                 460

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr
465                 470                 475                 480

Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His
                485                 490                 495

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn
            500                 505                 510

Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val
        515                 520                 525

Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys
    530                 535                 540
```

```
Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys
545                 550                 555                 560

Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys
                565                 570                 575

Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu
            580                 585                 590

Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu
        595                 600                 605

Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile
    610                 615                 620

Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu
625                 630                 635                 640

Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile
                645                 650                 655

Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp
            660                 665                 670

Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn
        675                 680                 685

Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp
    690                 695                 700

Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp
705                 710                 715                 720

Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly
                725                 730                 735

Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly
            740                 745                 750

Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn
        755                 760                 765

Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile
    770                 775                 780

Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile
785                 790                 795                 800
```

```
Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr
                805                 810                 815

Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro
            820                 825                 830

Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys
        835                 840                 845

Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile
    850                 855                 860

Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr
865                 870                 875                 880

Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg
                885                 890                 895

Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr
            900                 905                 910

Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile
        915                 920                 925

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp
    930                 935                 940

Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
945                 950                 955                 960

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu
                965                 970                 975

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe
            980                 985                 990

Ile Lys Arg Gln Leu Val Glu Thr  Arg Gln Ile Thr Lys  His Val Ala
        995                 1000                1005

Gln Ile  Leu Asp Ser Arg Met  Asn Thr Lys Tyr Asp  Glu Asn Asp
    1010                1015                1020

Lys Leu  Ile Arg Glu Val Lys  Val Ile Thr Leu Lys  Ser Lys Leu
    1025                1030                1035

Val Ser  Asp Phe Arg Lys Asp  Phe Gln Phe Tyr Lys  Val Arg Glu
    1040                1045                1050
```

```
Ile Asn  Asn Tyr His His Ala  His Asp Ala Tyr Leu  Asn Ala Val
    1055                1060                1065

Val Gly  Thr Ala Leu Ile Lys  Lys Tyr Pro Lys Leu  Glu Ser Glu
    1070                1075                1080

Phe Val  Tyr Gly Asp Tyr Lys  Val Tyr Asp Val Arg  Lys Met Ile
    1085                1090                1095

Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala Thr Ala  Lys Tyr Phe
    1100                1105                1110

Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys Thr Glu  Ile Thr Leu
    1115                1120                1125

Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu Ile Glu  Thr Asn Gly
    1130                1135                1140

Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly Arg Asp  Phe Ala Thr
    1145                1150                1155

Val Arg  Lys Val Leu Ser Met  Pro Gln Val Asn Ile  Val Lys Lys
    1160                1165                1170

Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys Glu Ser  Ile Leu Pro
    1175                1180                1185

Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg Lys Lys  Asp Trp Asp
    1190                1195                1200

Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro Thr Val  Ala Tyr Ser
    1205                1210                1215

Val Leu  Val Val Ala Lys Val  Glu Lys Gly Lys Ser  Lys Lys Leu
    1220                1225                1230

Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr Ile Met  Glu Arg Ser
    1235                1240                1245

Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu Glu Ala  Lys Gly Tyr
    1250                1255                1260

Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys Leu Pro  Lys Tyr Ser
    1265                1270                1275

Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg Met Leu  Ala Ser Ala
    1280                1285                1290
```

```
Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala Leu Pro  Ser Lys Tyr
    1295                1300                1305

Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr Glu Lys  Leu Lys Gly
    1310                1315                1320

Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu Phe Val  Glu Gln His
    1325                1330                1335

Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln Ile Ser  Glu Phe Ser
    1340                1345                1350

Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu Asp Lys  Val Leu Ser
    1355                1360                1365

Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile Arg Glu  Gln Ala Glu
    1370                1375                1380

Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn Leu Gly  Ala Pro Ala
    1385                1390                1395

Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp Arg Lys  Arg Tyr Thr
    1400                1405                1410

Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu Ile His  Gln Ser Ile
    1415                1420                1425

Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu Ser Gln  Leu Gly Gly
    1430                1435                1440

Asp Ser  Arg Ala Asp Pro Lys  Lys Lys Arg Lys Val
    1445                1450                1455
```

<210> 40
<211> 1455
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полипептид

<400> 40

```
Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu
1               5                  10                  15

Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp
            20                  25                  30

Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
```

```
        35                  40                  45

Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu
    50                  55                  60

Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly Met Asp Lys Lys
65                  70                  75                  80

Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val
                85                  90                  95

Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly
            100                 105                 110

Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu
        115                 120                 125

Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala
    130                 135                 140

Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu
145                 150                 155                 160

Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg
                165                 170                 175

Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His
            180                 185                 190

Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr
        195                 200                 205

Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys
    210                 215                 220

Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe
225                 230                 235                 240

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp
                245                 250                 255

Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe
            260                 265                 270

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu
        275                 280                 285

Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln
```

```
    290                 295                 300

Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu
305                 310                 315                 320

Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu
                325                 330                 335

Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp
            340                 345                 350

Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala
        355                 360                 365

Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val
    370                 375                 380

Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg
385                 390                 395                 400

Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg
                405                 410                 415

Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys
            420                 425                 430

Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe
        435                 440                 445

Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu
    450                 455                 460

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr
465                 470                 475                 480

Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His
                485                 490                 495

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn
            500                 505                 510

Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val
        515                 520                 525

Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys
    530                 535                 540

Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys
```

```
545                 550                 555                 560

Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys
                565                 570                 575

Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu
            580                 585                 590

Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu
        595                 600                 605

Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile
    610                 615                 620

Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu
625                 630                 635                 640

Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile
                645                 650                 655

Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp
            660                 665                 670

Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn
        675                 680                 685

Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp
    690                 695                 700

Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp
705                 710                 715                 720

Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly
                725                 730                 735

Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly
            740                 745                 750

Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn
        755                 760                 765

Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile
    770                 775                 780

Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile
785                 790                 795                 800

Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr
```

```
                805                 810                 815

Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro
            820                 825                 830

Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys
        835                 840                 845

Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile
    850                 855                 860

Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr
865                 870                 875                 880

Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg
                885                 890                 895

Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr
            900                 905                 910

Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile
        915                 920                 925

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp
    930                 935                 940

Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
945                 950                 955                 960

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu
                965                 970                 975

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe
            980                 985                 990

Ile Lys Arg Gln Leu Val Glu Thr  Arg Gln Ile Thr Lys  His Val Ala
        995                 1000                1005

Gln Ile  Leu Asp Ser Arg Met  Asn Thr Lys Tyr Asp  Glu Asn Asp
    1010                1015                1020

Lys Leu  Ile Arg Glu Val Lys  Val Ile Thr Leu Lys  Ser Lys Leu
    1025                1030                1035

Val Ser  Asp Phe Arg Lys Asp  Phe Gln Phe Tyr Lys  Val Arg Glu
    1040                1045                1050

Ile Asn  Asn Tyr His His Ala  His Asp Ala Tyr Leu  Asn Ala Val
```

```
    1055                1060                1065

Val Gly  Thr Ala Leu Ile Lys  Lys Tyr Pro Lys Leu  Glu Ser Glu
    1070                1075                1080

Phe Val  Tyr Gly Asp Tyr Lys  Val Tyr Asp Val Arg  Lys Met Ile
    1085                1090                1095

Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala Thr Ala  Lys Tyr Phe
    1100                1105                1110

Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys Thr Glu  Ile Thr Leu
    1115                1120                1125

Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu Ile Glu  Thr Asn Gly
    1130                1135                1140

Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly Arg Asp  Phe Ala Thr
    1145                1150                1155

Val Arg  Lys Val Leu Ser Met  Pro Gln Val Asn Ile  Val Lys Lys
    1160                1165                1170

Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys Glu Ser  Ile Leu Pro
    1175                1180                1185

Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg Lys Lys  Asp Trp Asp
    1190                1195                1200

Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro Thr Val  Ala Tyr Ser
    1205                1210                1215

Val Leu  Val Val Ala Lys Val  Glu Lys Gly Lys Ser  Lys Lys Leu
    1220                1225                1230

Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr Ile Met  Glu Arg Ser
    1235                1240                1245

Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu Glu Ala  Lys Gly Tyr
    1250                1255                1260

Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys Leu Pro  Lys Tyr Ser
    1265                1270                1275

Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg Met Leu  Ala Ser Ala
    1280                1285                1290

Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala Leu Pro  Ser Lys Tyr
```

```
    1295                1300                1305

Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr Glu Lys  Leu Lys Gly
    1310                1315                1320

Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu Phe Val  Glu Gln His
    1325                1330                1335

Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln Ile Ser  Glu Phe Ser
    1340                1345                1350

Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu Asp Lys  Val Leu Ser
    1355                1360                1365

Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile Arg Glu  Gln Ala Glu
    1370                1375                1380

Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn Leu Gly  Ala Pro Ala
    1385                1390                1395

Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp Arg Lys  Arg Tyr Thr
    1400                1405                1410

Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu Ile His  Gln Ser Ile
    1415                1420                1425

Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu Ser Gln  Leu Gly Gly
    1430                1435                1440

Asp Ser  Arg Ala Asp Pro Lys  Lys Lys Arg Lys Val
    1445                1450                1455
```

<210> 41
<211> 1455
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полипептид

<400> 41

```
Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu
1               5                  10                  15

Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp
            20                  25                  30

Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
        35                  40                  45
```

```
Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu
    50                  55                  60

Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly Gly Asp Lys Lys
65                  70                  75                  80

Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val
                85                  90                  95

Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly
            100                 105                 110

Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu
        115                 120                 125

Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala
    130                 135                 140

Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu
145                 150                 155                 160

Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg
                165                 170                 175

Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His
            180                 185                 190

Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr
        195                 200                 205

Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys
    210                 215                 220

Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe
225                 230                 235                 240

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp
                245                 250                 255

Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe
            260                 265                 270

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu
        275                 280                 285

Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln
    290                 295                 300
```

```
Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu
305                 310                 315                 320

Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu
                325                 330                 335

Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp
            340                 345                 350

Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala
        355                 360                 365

Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val
    370                 375                 380

Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg
385                 390                 395                 400

Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg
                405                 410                 415

Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys
            420                 425                 430

Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe
        435                 440                 445

Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu
    450                 455                 460

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr
465                 470                 475                 480

Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His
                485                 490                 495

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn
            500                 505                 510

Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val
        515                 520                 525

Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys
    530                 535                 540

Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys
545                 550                 555                 560
```

```
Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys
                565                 570                 575

Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu
            580                 585                 590

Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu
        595                 600                 605

Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile
    610                 615                 620

Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu
625                 630                 635                 640

Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile
                645                 650                 655

Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp
            660                 665                 670

Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn
        675                 680                 685

Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp
    690                 695                 700

Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp
705                 710                 715                 720

Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly
                725                 730                 735

Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly
            740                 745                 750

Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn
        755                 760                 765

Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile
    770                 775                 780

Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile
785                 790                 795                 800

Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr
                805                 810                 815
```

```
Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro
            820                 825                 830

Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys
        835                 840                 845

Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile
    850                 855                 860

Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr
865                 870                 875                 880

Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg
                885                 890                 895

Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr
            900                 905                 910

Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile
        915                 920                 925

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp
    930                 935                 940

Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
945                 950                 955                 960

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu
                965                 970                 975

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe
            980                 985                 990

Ile Lys Arg Gln Leu Val Glu Thr  Arg Gln Ile Thr Lys  His Val Ala
        995                 1000                1005

Gln Ile  Leu Asp Ser Arg Met  Asn Thr Lys Tyr Asp  Glu Asn Asp
    1010                1015                1020

Lys Leu  Ile Arg Glu Val Lys  Val Ile Thr Leu Lys  Ser Lys Leu
    1025                1030                1035

Val Ser  Asp Phe Arg Lys Asp  Phe Gln Phe Tyr Lys  Val Arg Glu
    1040                1045                1050

Ile Asn  Asn Tyr His His Ala  His Asp Ala Tyr Leu  Asn Ala Val
    1055                1060                1065
```

```
Val Gly  Thr Ala Leu Ile Lys  Lys Tyr Pro Lys Leu  Glu Ser Glu
    1070                1075                1080

Phe Val  Tyr Gly Asp Tyr Lys  Val Tyr Asp Val Arg  Lys Met Ile
    1085                1090                1095

Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala Thr Ala  Lys Tyr Phe
    1100                1105                1110

Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys Thr Glu  Ile Thr Leu
    1115                1120                1125

Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu Ile Glu  Thr Asn Gly
    1130                1135                1140

Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly Arg Asp  Phe Ala Thr
    1145                1150                1155

Val Arg  Lys Val Leu Ser Met  Pro Gln Val Asn Ile  Val Lys Lys
    1160                1165                1170

Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys Glu Ser  Ile Leu Pro
    1175                1180                1185

Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg Lys Lys  Asp Trp Asp
    1190                1195                1200

Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro Thr Val  Ala Tyr Ser
    1205                1210                1215

Val Leu  Val Val Ala Lys Val  Glu Lys Gly Lys Ser  Lys Lys Leu
    1220                1225                1230

Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr Ile Met  Glu Arg Ser
    1235                1240                1245

Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu Glu Ala  Lys Gly Tyr
    1250                1255                1260

Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys Leu Pro  Lys Tyr Ser
    1265                1270                1275

Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg Met Leu  Ala Ser Ala
    1280                1285                1290

Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala Leu Pro  Ser Lys Tyr
    1295                1300                1305
```

```
Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr Glu Lys  Leu Lys Gly
    1310                1315                1320

Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu Phe Val  Glu Gln His
    1325                1330                1335

Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln Ile Ser  Glu Phe Ser
    1340                1345                1350

Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu Asp Lys  Val Leu Ser
    1355                1360                1365

Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile Arg Glu  Gln Ala Glu
    1370                1375                1380

Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn Leu Gly  Ala Pro Ala
    1385                1390                1395

Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp Arg Lys  Arg Tyr Thr
    1400                1405                1410

Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu Ile His  Gln Ser Ile
    1415                1420                1425

Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu Ser Gln  Leu Gly Gly
    1430                1435                1440

Asp Ser  Arg Ala Asp Pro Lys  Lys Lys Arg Lys Val
    1445                1450                1455
```

<210> 42
<211> 1455
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полипептид

<400> 42

```
Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu
1               5                  10                  15

Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp
            20                  25                  30

Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
        35                  40                  45

Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu
```

```
    50                  55                  60

Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly Pro Asp Lys Lys
65                  70                  75                  80

Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val
                85                  90                  95

Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly
            100                 105                 110

Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu
        115                 120                 125

Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala
    130                 135                 140

Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu
145                 150                 155                 160

Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg
                165                 170                 175

Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His
            180                 185                 190

Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr
        195                 200                 205

Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys
    210                 215                 220

Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe
225                 230                 235                 240

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp
                245                 250                 255

Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe
            260                 265                 270

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu
        275                 280                 285

Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln
    290                 295                 300

Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu
```

305 310 315 320

Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu
325 330 335

Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp
340 345 350

Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala
355 360 365

Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val
370 375 380

Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg
385 390 395 400

Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg
405 410 415

Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys
420 425 430

Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe
435 440 445

Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu
450 455 460

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr
465 470 475 480

Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His
485 490 495

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn
500 505 510

Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val
515 520 525

Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys
530 535 540

Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys
545 550 555 560

Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys

565 570 575

Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu
580 585 590

Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu
595 600 605

Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile
610 615 620

Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu
625 630 635 640

Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile
645 650 655

Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp
660 665 670

Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn
675 680 685

Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp
690 695 700

Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp
705 710 715 720

Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly
725 730 735

Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly
740 745 750

Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn
755 760 765

Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile
770 775 780

Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile
785 790 795 800

Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr
805 810 815

Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro

```
            820                 825                 830

Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys
        835                 840                 845

Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile
    850                 855                 860

Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr
865                 870                 875                 880

Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg
                885                 890                 895

Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr
            900                 905                 910

Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile
        915                 920                 925

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp
    930                 935                 940

Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
945                 950                 955                 960

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu
                965                 970                 975

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe
            980                 985                 990

Ile Lys Arg Gln Leu Val Glu Thr  Arg Gln Ile Thr Lys  His Val Ala
        995                 1000                1005

Gln Ile  Leu Asp Ser Arg Met  Asn Thr Lys Tyr Asp  Glu Asn Asp
    1010                1015                1020

Lys Leu  Ile Arg Glu Val Lys  Val Ile Thr Leu Lys  Ser Lys Leu
    1025                1030                1035

Val Ser  Asp Phe Arg Lys Asp  Phe Gln Phe Tyr Lys  Val Arg Glu
    1040                1045                1050

Ile Asn  Asn Tyr His His Ala  His Asp Ala Tyr Leu  Asn Ala Val
    1055                1060                1065

Val Gly  Thr Ala Leu Ile Lys  Lys Tyr Pro Lys Leu  Glu Ser Glu
```

```
    1070                1075                1080

Phe Val  Tyr Gly Asp Tyr Lys  Val Tyr Asp Val Arg  Lys Met Ile
    1085                1090                1095

Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala Thr Ala  Lys Tyr Phe
    1100                1105                1110

Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys Thr Glu  Ile Thr Leu
    1115                1120                1125

Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu Ile Glu  Thr Asn Gly
    1130                1135                1140

Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly Arg Asp  Phe Ala Thr
    1145                1150                1155

Val Arg  Lys Val Leu Ser Met  Pro Gln Val Asn Ile  Val Lys Lys
    1160                1165                1170

Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys Glu Ser  Ile Leu Pro
    1175                1180                1185

Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg Lys Lys  Asp Trp Asp
    1190                1195                1200

Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro Thr Val  Ala Tyr Ser
    1205                1210                1215

Val Leu  Val Val Ala Lys Val  Glu Lys Gly Lys Ser  Lys Lys Leu
    1220                1225                1230

Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr Ile Met  Glu Arg Ser
    1235                1240                1245

Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu Glu Ala  Lys Gly Tyr
    1250                1255                1260

Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys Leu Pro  Lys Tyr Ser
    1265                1270                1275

Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg Met Leu  Ala Ser Ala
    1280                1285                1290

Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala Leu Pro  Ser Lys Tyr
    1295                1300                1305

Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr Glu Lys  Leu Lys Gly
```

```
    1310                1315                1320

Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu Phe Val  Glu Gln His
    1325                1330                1335

Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln Ile Ser  Glu Phe Ser
    1340                1345                1350

Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu Asp Lys  Val Leu Ser
    1355                1360                1365

Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile Arg Glu  Gln Ala Glu
    1370                1375                1380

Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn Leu Gly  Ala Pro Ala
    1385                1390                1395

Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp Arg Lys  Arg Tyr Thr
    1400                1405                1410

Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu Ile His  Gln Ser Ile
    1415                1420                1425

Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu Ser Gln  Leu Gly Gly
    1430                1435                1440

Asp Ser  Arg Ala Asp Pro Lys  Lys Lys Arg Lys Val
    1445                1450                1455
```

<210> 43
<211> 1455
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полипептид

<400> 43

```
Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu
1               5                  10                  15

Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp
            20                  25                  30

Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
        35                  40                  45

Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu
    50                  55                  60
```

```
Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly Ile Asp Lys Lys
65                  70                  75                  80

Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val
                85                  90                  95

Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly
            100                 105                 110

Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu
        115                 120                 125

Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala
    130                 135                 140

Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu
145                 150                 155                 160

Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg
                165                 170                 175

Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His
            180                 185                 190

Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr
        195                 200                 205

Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys
    210                 215                 220

Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe
225                 230                 235                 240

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp
                245                 250                 255

Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe
            260                 265                 270

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu
        275                 280                 285

Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln
    290                 295                 300

Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu
305                 310                 315                 320
```

```
Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu
                325                 330                 335

Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp
            340                 345                 350

Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala
        355                 360                 365

Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val
    370                 375                 380

Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg
385                 390                 395                 400

Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg
                405                 410                 415

Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys
            420                 425                 430

Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe
        435                 440                 445

Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu
    450                 455                 460

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr
465                 470                 475                 480

Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His
                485                 490                 495

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn
            500                 505                 510

Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val
        515                 520                 525

Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys
    530                 535                 540

Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys
545                 550                 555                 560

Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys
                565                 570                 575
```

```
Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu
            580                 585                 590

Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu
        595                 600                 605

Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile
    610                 615                 620

Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu
625                 630                 635                 640

Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile
                645                 650                 655

Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp
            660                 665                 670

Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn
        675                 680                 685

Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp
    690                 695                 700

Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp
705                 710                 715                 720

Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly
                725                 730                 735

Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly
            740                 745                 750

Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn
        755                 760                 765

Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile
    770                 775                 780

Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile
785                 790                 795                 800

Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr
                805                 810                 815

Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro
            820                 825                 830
```

```
Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys
        835                 840                 845

Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile
    850                 855                 860

Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr
865                 870                 875                 880

Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg
                885                 890                 895

Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr
            900                 905                 910

Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile
        915                 920                 925

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp
    930                 935                 940

Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
945                 950                 955                 960

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu
                965                 970                 975

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe
            980                 985                 990

Ile Lys Arg Gln Leu Val Glu Thr  Arg Gln Ile Thr Lys  His Val Ala
        995                 1000                1005

Gln Ile  Leu Asp Ser Arg Met  Asn Thr Lys Tyr Asp  Glu Asn Asp
    1010                1015                1020

Lys Leu  Ile Arg Glu Val Lys  Val Ile Thr Leu Lys  Ser Lys Leu
    1025                1030                1035

Val Ser  Asp Phe Arg Lys Asp  Phe Gln Phe Tyr Lys  Val Arg Glu
    1040                1045                1050

Ile Asn  Asn Tyr His His Ala  His Asp Ala Tyr Leu  Asn Ala Val
    1055                1060                1065

Val Gly  Thr Ala Leu Ile Lys  Lys Tyr Pro Lys Leu  Glu Ser Glu
    1070                1075                1080
```

```
Phe Val  Tyr Gly Asp Tyr Lys  Val Tyr Asp Val Arg  Lys Met Ile
    1085                1090                1095

Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala Thr Ala  Lys Tyr Phe
    1100                1105                1110

Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys Thr Glu  Ile Thr Leu
    1115                1120                1125

Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu Ile Glu  Thr Asn Gly
    1130                1135                1140

Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly Arg Asp  Phe Ala Thr
    1145                1150                1155

Val Arg  Lys Val Leu Ser Met  Pro Gln Val Asn Ile  Val Lys Lys
    1160                1165                1170

Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys Glu Ser  Ile Leu Pro
    1175                1180                1185

Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg Lys Lys  Asp Trp Asp
    1190                1195                1200

Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro Thr Val  Ala Tyr Ser
    1205                1210                1215

Val Leu  Val Val Ala Lys Val  Glu Lys Gly Lys Ser  Lys Lys Leu
    1220                1225                1230

Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr Ile Met  Glu Arg Ser
    1235                1240                1245

Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu Glu Ala  Lys Gly Tyr
    1250                1255                1260

Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys Leu Pro  Lys Tyr Ser
    1265                1270                1275

Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg Met Leu  Ala Ser Ala
    1280                1285                1290

Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala Leu Pro  Ser Lys Tyr
    1295                1300                1305

Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr Glu Lys  Leu Lys Gly
    1310                1315                1320
```

```
Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu Phe Val  Glu Gln His
    1325                1330                1335

Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln Ile Ser  Glu Phe Ser
    1340                1345                1350

Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu Asp Lys  Val Leu Ser
    1355                1360                1365

Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile Arg Glu  Gln Ala Glu
    1370                1375                1380

Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn Leu Gly  Ala Pro Ala
    1385                1390                1395

Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp Arg Lys  Arg Tyr Thr
    1400                1405                1410

Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu Ile His  Gln Ser Ile
    1415                1420                1425

Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu Ser Gln  Leu Gly Gly
    1430                1435                1440

Asp Ser  Arg Ala Asp Pro Lys  Lys Lys Arg Lys Val
    1445                1450                1455
```

<210> 44
<211> 1455
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полипептид

<400> 44

```
Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu
1               5                  10                  15

Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp
            20                  25                  30

Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
        35                  40                  45

Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu
    50                  55                  60

Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly Thr Asp Lys Lys
65                  70                  75                  80
```

```
Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val
                85                  90                  95

Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly
            100                 105                 110

Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu
        115                 120                 125

Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala
    130                 135                 140

Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu
145                 150                 155                 160

Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg
                165                 170                 175

Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His
            180                 185                 190

Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr
        195                 200                 205

Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys
    210                 215                 220

Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe
225                 230                 235                 240

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp
                245                 250                 255

Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe
            260                 265                 270

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu
        275                 280                 285

Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln
    290                 295                 300

Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu
305                 310                 315                 320

Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu
                325                 330                 335
```

```
Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp
            340                 345                 350

Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala
        355                 360                 365

Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val
    370                 375                 380

Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg
385                 390                 395                 400

Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg
                405                 410                 415

Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys
            420                 425                 430

Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe
        435                 440                 445

Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu
    450                 455                 460

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr
465                 470                 475                 480

Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His
                485                 490                 495

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn
            500                 505                 510

Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val
        515                 520                 525

Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys
    530                 535                 540

Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys
545                 550                 555                 560

Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys
                565                 570                 575

Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu
            580                 585                 590
```

```
Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu
        595                 600                 605

Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile
    610                 615                 620

Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu
625                 630                 635                 640

Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile
                645                 650                 655

Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp
            660                 665                 670

Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn
        675                 680                 685

Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp
    690                 695                 700

Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp
705                 710                 715                 720

Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly
                725                 730                 735

Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly
            740                 745                 750

Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn
        755                 760                 765

Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile
    770                 775                 780

Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile
785                 790                 795                 800

Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr
                805                 810                 815

Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro
            820                 825                 830

Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys
        835                 840                 845
```

```
Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile
    850                 855                 860

Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr
865                 870                 875                 880

Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg
                885                 890                 895

Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr
            900                 905                 910

Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile
        915                 920                 925

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp
    930                 935                 940

Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
945                 950                 955                 960

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu
                965                 970                 975

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe
            980                 985                 990

Ile Lys Arg Gln Leu Val Glu Thr  Arg Gln Ile Thr Lys  His Val Ala
        995                 1000                1005

Gln Ile  Leu Asp Ser Arg Met  Asn Thr Lys Tyr Asp  Glu Asn Asp
    1010                1015                1020

Lys Leu  Ile Arg Glu Val Lys  Val Ile Thr Leu Lys  Ser Lys Leu
    1025                1030                1035

Val Ser  Asp Phe Arg Lys Asp  Phe Gln Phe Tyr Lys  Val Arg Glu
    1040                1045                1050

Ile Asn  Asn Tyr His His Ala  His Asp Ala Tyr Leu  Asn Ala Val
    1055                1060                1065

Val Gly  Thr Ala Leu Ile Lys  Lys Tyr Pro Lys Leu  Glu Ser Glu
    1070                1075                1080

Phe Val  Tyr Gly Asp Tyr Lys  Val Tyr Asp Val Arg  Lys Met Ile
    1085                1090                1095
```

```
Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala Thr Ala  Lys Tyr Phe
    1100                 1105                 1110

Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys Thr Glu  Ile Thr Leu
    1115                 1120                 1125

Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu Ile Glu  Thr Asn Gly
    1130                 1135                 1140

Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly Arg Asp  Phe Ala Thr
    1145                 1150                 1155

Val Arg  Lys Val Leu Ser Met  Pro Gln Val Asn Ile  Val Lys Lys
    1160                 1165                 1170

Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys Glu Ser  Ile Leu Pro
    1175                 1180                 1185

Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg Lys Lys  Asp Trp Asp
    1190                 1195                 1200

Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro Thr Val  Ala Tyr Ser
    1205                 1210                 1215

Val Leu  Val Val Ala Lys Val  Glu Lys Gly Lys Ser  Lys Lys Leu
    1220                 1225                 1230

Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr Ile Met  Glu Arg Ser
    1235                 1240                 1245

Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu Glu Ala  Lys Gly Tyr
    1250                 1255                 1260

Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys Leu Pro  Lys Tyr Ser
    1265                 1270                 1275

Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg Met Leu  Ala Ser Ala
    1280                 1285                 1290

Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala Leu Pro  Ser Lys Tyr
    1295                 1300                 1305

Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr Glu Lys  Leu Lys Gly
    1310                 1315                 1320

Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu Phe Val  Glu Gln His
    1325                 1330                 1335
```

```
Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln Ile Ser  Glu Phe Ser
    1340                1345                1350

Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu Asp Lys  Val Leu Ser
    1355                1360                1365

Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile Arg Glu  Gln Ala Glu
    1370                1375                1380

Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn Leu Gly  Ala Pro Ala
    1385                1390                1395

Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp Arg Lys  Arg Tyr Thr
    1400                1405                1410

Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu Ile His  Gln Ser Ile
    1415                1420                1425

Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu Ser Gln  Leu Gly Gly
    1430                1435                1440

Asp Ser  Arg Ala Asp Pro Lys  Lys Lys Arg Lys Val
    1445                1450                1455
```

<210> 45
<211> 1455
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полипептид

<400> 45

```
Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu
1               5                  10                  15

Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp
            20                  25                  30

Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
        35                  40                  45

Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu
    50                  55                  60

Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly Leu Asp Lys Lys
65                  70                  75                  80
```

```
Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val
                85                  90                  95

Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly
            100                 105                 110

Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu
        115                 120                 125

Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala
    130                 135                 140

Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu
145                 150                 155                 160

Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg
                165                 170                 175

Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His
            180                 185                 190

Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr
        195                 200                 205

Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys
    210                 215                 220

Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe
225                 230                 235                 240

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp
                245                 250                 255

Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe
            260                 265                 270

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu
        275                 280                 285

Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln
    290                 295                 300

Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu
305                 310                 315                 320

Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu
                325                 330                 335
```

```
Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp
            340                 345                 350

Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala
        355                 360                 365

Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val
    370                 375                 380

Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg
385                 390                 395                 400

Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg
                405                 410                 415

Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys
            420                 425                 430

Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe
        435                 440                 445

Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu
    450                 455                 460

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr
465                 470                 475                 480

Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His
                485                 490                 495

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn
            500                 505                 510

Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val
        515                 520                 525

Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys
    530                 535                 540

Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys
545                 550                 555                 560

Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys
                565                 570                 575

Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu
            580                 585                 590
```

```
Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu
        595                 600                 605

Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile
    610                 615                 620

Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu
625                 630                 635                 640

Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile
                645                 650                 655

Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp
            660                 665                 670

Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn
        675                 680                 685

Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp
    690                 695                 700

Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp
705                 710                 715                 720

Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly
                725                 730                 735

Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly
            740                 745                 750

Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn
        755                 760                 765

Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile
    770                 775                 780

Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile
785                 790                 795                 800

Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr
                805                 810                 815

Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro
            820                 825                 830

Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys
        835                 840                 845
```

```
Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile
    850                 855                 860

Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr
865                 870                 875                 880

Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg
                885                 890                 895

Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr
            900                 905                 910

Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile
        915                 920                 925

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp
    930                 935                 940

Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
945                 950                 955                 960

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu
                965                 970                 975

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe
            980                 985                 990

Ile Lys Arg Gln Leu Val Glu Thr  Arg Gln Ile Thr Lys  His Val Ala
        995                 1000                1005

Gln Ile  Leu Asp Ser Arg Met  Asn Thr Lys Tyr Asp  Glu Asn Asp
    1010                1015                1020

Lys Leu  Ile Arg Glu Val Lys  Val Ile Thr Leu Lys  Ser Lys Leu
    1025                1030                1035

Val Ser  Asp Phe Arg Lys Asp  Phe Gln Phe Tyr Lys  Val Arg Glu
    1040                1045                1050

Ile Asn  Asn Tyr His His Ala  His Asp Ala Tyr Leu  Asn Ala Val
    1055                1060                1065

Val Gly  Thr Ala Leu Ile Lys  Lys Tyr Pro Lys Leu  Glu Ser Glu
    1070                1075                1080

Phe Val  Tyr Gly Asp Tyr Lys  Val Tyr Asp Val Arg  Lys Met Ile
    1085                1090                1095
```

```
Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala Thr Ala  Lys Tyr Phe
    1100                 1105                 1110

Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys Thr Glu  Ile Thr Leu
    1115                 1120                 1125

Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu Ile Glu  Thr Asn Gly
    1130                 1135                 1140

Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly Arg Asp  Phe Ala Thr
    1145                 1150                 1155

Val Arg  Lys Val Leu Ser Met  Pro Gln Val Asn Ile  Val Lys Lys
    1160                 1165                 1170

Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys Glu Ser  Ile Leu Pro
    1175                 1180                 1185

Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg Lys Lys  Asp Trp Asp
    1190                 1195                 1200

Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro Thr Val  Ala Tyr Ser
    1205                 1210                 1215

Val Leu  Val Val Ala Lys Val  Glu Lys Gly Lys Ser  Lys Lys Leu
    1220                 1225                 1230

Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr Ile Met  Glu Arg Ser
    1235                 1240                 1245

Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu Glu Ala  Lys Gly Tyr
    1250                 1255                 1260

Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys Leu Pro  Lys Tyr Ser
    1265                 1270                 1275

Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg Met Leu  Ala Ser Ala
    1280                 1285                 1290

Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala Leu Pro  Ser Lys Tyr
    1295                 1300                 1305

Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr Glu Lys  Leu Lys Gly
    1310                 1315                 1320

Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu Phe Val  Glu Gln His
    1325                 1330                 1335
```

```
Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln Ile Ser  Glu Phe Ser
    1340                1345                1350

Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu Asp Lys  Val Leu Ser
    1355                1360                1365

Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile Arg Glu  Gln Ala Glu
    1370                1375                1380

Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn Leu Gly  Ala Pro Ala
    1385                1390                1395

Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp Arg Lys  Arg Tyr Thr
    1400                1405                1410

Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu Ile His  Gln Ser Ile
    1415                1420                1425

Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu Ser Gln  Leu Gly Gly
    1430                1435                1440

Asp Ser  Arg Ala Asp Pro Lys  Lys Lys Arg Lys Val
    1445                1450                1455
```

<210> 46
<211> 1455
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полипептид

<400> 46

```
Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu
1               5                  10                  15

Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp
            20                  25                  30

Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
        35                  40                  45

Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu
    50                  55                  60

Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly Ala Asp Lys Lys
65                  70                  75                  80

Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val
                85                  90                  95
```

```
Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly
            100                 105                 110

Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu
        115                 120                 125

Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala
    130                 135                 140

Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu
145                 150                 155                 160

Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg
                165                 170                 175

Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His
            180                 185                 190

Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr
        195                 200                 205

Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys
    210                 215                 220

Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe
225                 230                 235                 240

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp
                245                 250                 255

Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe
            260                 265                 270

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu
        275                 280                 285

Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln
    290                 295                 300

Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu
305                 310                 315                 320

Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu
                325                 330                 335

Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp
            340                 345                 350
```

```
Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala
        355                 360                 365

Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val
    370                 375                 380

Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg
385                 390                 395                 400

Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg
                405                 410                 415

Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys
            420                 425                 430

Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe
        435                 440                 445

Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu
    450                 455                 460

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr
465                 470                 475                 480

Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His
                485                 490                 495

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn
            500                 505                 510

Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val
        515                 520                 525

Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys
    530                 535                 540

Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys
545                 550                 555                 560

Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys
                565                 570                 575

Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu
            580                 585                 590

Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu
        595                 600                 605
```

```
Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile
    610                 615                 620

Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu
625                 630                 635                 640

Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile
                645                 650                 655

Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp
            660                 665                 670

Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn
        675                 680                 685

Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp
    690                 695                 700

Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp
705                 710                 715                 720

Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly
                725                 730                 735

Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly
            740                 745                 750

Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn
        755                 760                 765

Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile
    770                 775                 780

Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile
785                 790                 795                 800

Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr
                805                 810                 815

Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro
            820                 825                 830

Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys
        835                 840                 845

Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile
    850                 855                 860
```

```
Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr
865                 870                 875                 880

Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg
                885                 890                 895

Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr
            900                 905                 910

Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile
        915                 920                 925

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp
    930                 935                 940

Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
945                 950                 955                 960

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu
                965                 970                 975

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe
            980                 985                 990

Ile Lys Arg Gln Leu Val Glu Thr  Arg Gln Ile Thr Lys  His Val Ala
        995                 1000                1005

Gln Ile  Leu Asp Ser Arg Met  Asn Thr Lys Tyr Asp  Glu Asn Asp
    1010                1015                1020

Lys Leu  Ile Arg Glu Val Lys  Val Ile Thr Leu Lys  Ser Lys Leu
    1025                1030                1035

Val Ser  Asp Phe Arg Lys Asp  Phe Gln Phe Tyr Lys  Val Arg Glu
    1040                1045                1050

Ile Asn  Asn Tyr His His Ala  His Asp Ala Tyr Leu  Asn Ala Val
    1055                1060                1065

Val Gly  Thr Ala Leu Ile Lys  Lys Tyr Pro Lys Leu  Glu Ser Glu
    1070                1075                1080

Phe Val  Tyr Gly Asp Tyr Lys  Val Tyr Asp Val Arg  Lys Met Ile
    1085                1090                1095

Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala Thr Ala  Lys Tyr Phe
    1100                1105                1110
```

```
Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys Thr Glu  Ile Thr Leu
    1115                1120                1125

Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu Ile Glu  Thr Asn Gly
    1130                1135                1140

Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly Arg Asp  Phe Ala Thr
    1145                1150                1155

Val Arg  Lys Val Leu Ser Met  Pro Gln Val Asn Ile  Val Lys Lys
    1160                1165                1170

Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys Glu Ser  Ile Leu Pro
    1175                1180                1185

Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg Lys Lys  Asp Trp Asp
    1190                1195                1200

Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro Thr Val  Ala Tyr Ser
    1205                1210                1215

Val Leu  Val Val Ala Lys Val  Glu Lys Gly Lys Ser  Lys Lys Leu
    1220                1225                1230

Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr Ile Met  Glu Arg Ser
    1235                1240                1245

Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu Glu Ala  Lys Gly Tyr
    1250                1255                1260

Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys Leu Pro  Lys Tyr Ser
    1265                1270                1275

Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg Met Leu  Ala Ser Ala
    1280                1285                1290

Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala Leu Pro  Ser Lys Tyr
    1295                1300                1305

Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr Glu Lys  Leu Lys Gly
    1310                1315                1320

Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu Phe Val  Glu Gln His
    1325                1330                1335

Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln Ile Ser  Glu Phe Ser
    1340                1345                1350
```

```
Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu Asp Lys  Val Leu Ser
    1355                1360                1365

Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile Arg Glu  Gln Ala Glu
    1370                1375                1380

Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn Leu Gly  Ala Pro Ala
    1385                1390                1395

Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp Arg Lys  Arg Tyr Thr
    1400                1405                1410

Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu Ile His  Gln Ser Ile
    1415                1420                1425

Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu Ser Gln  Leu Gly Gly
    1430                1435                1440

Asp Ser  Arg Ala Asp Pro Lys  Lys Lys Arg Lys Val
    1445                1450                1455
```

<210> 47
<211> 1455
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полипептид

<400> 47

```
Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu
1               5                  10                  15

Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp
            20                  25                  30

Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
        35                  40                  45

Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu
    50                  55                  60

Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly His Asp Lys Lys
65                  70                  75                  80

Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val
                85                  90                  95
```

```
Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly
            100                 105                 110

Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu
        115                 120                 125

Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala
    130                 135                 140

Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu
145                 150                 155                 160

Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg
                165                 170                 175

Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His
            180                 185                 190

Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr
        195                 200                 205

Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys
    210                 215                 220

Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe
225                 230                 235                 240

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp
                245                 250                 255

Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe
            260                 265                 270

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu
        275                 280                 285

Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln
    290                 295                 300

Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu
305                 310                 315                 320

Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu
                325                 330                 335

Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp
            340                 345                 350
```

```
Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala
        355                 360                 365

Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val
    370                 375                 380

Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg
385                 390                 395                 400

Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg
                405                 410                 415

Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys
            420                 425                 430

Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe
        435                 440                 445

Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu
    450                 455                 460

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr
465                 470                 475                 480

Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His
                485                 490                 495

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn
            500                 505                 510

Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val
        515                 520                 525

Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys
    530                 535                 540

Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys
545                 550                 555                 560

Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys
                565                 570                 575

Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu
            580                 585                 590

Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu
        595                 600                 605
```

```
Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile
    610                 615                 620

Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu
625                 630                 635                 640

Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile
                645                 650                 655

Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp
            660                 665                 670

Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn
        675                 680                 685

Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp
    690                 695                 700

Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp
705                 710                 715                 720

Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly
                725                 730                 735

Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly
            740                 745                 750

Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn
        755                 760                 765

Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile
    770                 775                 780

Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile
785                 790                 795                 800

Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr
                805                 810                 815

Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro
            820                 825                 830

Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys
        835                 840                 845

Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile
    850                 855                 860
```

```
Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr
865                 870                 875                 880

Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg
                885                 890                 895

Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr
            900                 905                 910

Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile
        915                 920                 925

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp
    930                 935                 940

Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
945                 950                 955                 960

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu
                965                 970                 975

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe
            980                 985                 990

Ile Lys Arg Gln Leu Val Glu Thr  Arg Gln Ile Thr Lys  His Val Ala
        995                 1000                1005

Gln Ile  Leu Asp Ser Arg Met  Asn Thr Lys Tyr Asp  Glu Asn Asp
    1010                1015                1020

Lys Leu  Ile Arg Glu Val Lys  Val Ile Thr Leu Lys  Ser Lys Leu
    1025                1030                1035

Val Ser  Asp Phe Arg Lys Asp  Phe Gln Phe Tyr Lys  Val Arg Glu
    1040                1045                1050

Ile Asn  Asn Tyr His His Ala  His Asp Ala Tyr Leu  Asn Ala Val
    1055                1060                1065

Val Gly  Thr Ala Leu Ile Lys  Lys Tyr Pro Lys Leu  Glu Ser Glu
    1070                1075                1080

Phe Val  Tyr Gly Asp Tyr Lys  Val Tyr Asp Val Arg  Lys Met Ile
    1085                1090                1095

Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala Thr Ala  Lys Tyr Phe
    1100                1105                1110
```

```
Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys Thr Glu  Ile Thr Leu
    1115                1120                1125

Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu Ile Glu  Thr Asn Gly
    1130                1135                1140

Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly Arg Asp  Phe Ala Thr
    1145                1150                1155

Val Arg  Lys Val Leu Ser Met  Pro Gln Val Asn Ile  Val Lys Lys
    1160                1165                1170

Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys Glu Ser  Ile Leu Pro
    1175                1180                1185

Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg Lys Lys  Asp Trp Asp
    1190                1195                1200

Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro Thr Val  Ala Tyr Ser
    1205                1210                1215

Val Leu  Val Val Ala Lys Val  Glu Lys Gly Lys Ser  Lys Lys Leu
    1220                1225                1230

Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr Ile Met  Glu Arg Ser
    1235                1240                1245

Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu Glu Ala  Lys Gly Tyr
    1250                1255                1260

Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys Leu Pro  Lys Tyr Ser
    1265                1270                1275

Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg Met Leu  Ala Ser Ala
    1280                1285                1290

Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala Leu Pro  Ser Lys Tyr
    1295                1300                1305

Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr Glu Lys  Leu Lys Gly
    1310                1315                1320

Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu Phe Val  Glu Gln His
    1325                1330                1335

Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln Ile Ser  Glu Phe Ser
    1340                1345                1350
```

```
Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu Asp Lys  Val Leu Ser
    1355                1360                1365

Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile Arg Glu  Gln Ala Glu
    1370                1375                1380

Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn Leu Gly  Ala Pro Ala
    1385                1390                1395

Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp Arg Lys  Arg Tyr Thr
    1400                1405                1410

Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu Ile His  Gln Ser Ile
    1415                1420                1425

Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu Ser Gln  Leu Gly Gly
    1430                1435                1440

Asp Ser  Arg Ala Asp Pro Lys  Lys Lys Arg Lys Val
    1445                1450                1455
```

<210> 48
<211> 1455
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полипептид

<400> 48

```
Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu
1               5                  10                  15

Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp
            20                  25                  30

Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
        35                  40                  45

Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu
    50                  55                  60

Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly Trp Asp Lys Lys
65                  70                  75                  80

Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val
                85                  90                  95

Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly
            100                 105                 110
```

```
Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu
        115                 120                 125

Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala
    130                 135                 140

Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu
145                 150                 155                 160

Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg
                165                 170                 175

Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His
            180                 185                 190

Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr
        195                 200                 205

Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys
    210                 215                 220

Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe
225                 230                 235                 240

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp
                245                 250                 255

Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe
            260                 265                 270

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu
        275                 280                 285

Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln
    290                 295                 300

Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu
305                 310                 315                 320

Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu
                325                 330                 335

Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp
            340                 345                 350

Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala
        355                 360                 365
```

```
Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val
    370                 375                 380

Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg
385                 390                 395                 400

Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg
                405                 410                 415

Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys
            420                 425                 430

Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe
        435                 440                 445

Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu
    450                 455                 460

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr
465                 470                 475                 480

Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His
                485                 490                 495

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn
            500                 505                 510

Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val
        515                 520                 525

Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys
    530                 535                 540

Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys
545                 550                 555                 560

Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys
                565                 570                 575

Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu
            580                 585                 590

Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu
        595                 600                 605

Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile
    610                 615                 620
```

```
Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu
625                 630                 635                 640

Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile
                645                 650                 655

Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp
            660                 665                 670

Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn
        675                 680                 685

Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp
    690                 695                 700

Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp
705                 710                 715                 720

Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly
                725                 730                 735

Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly
            740                 745                 750

Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn
        755                 760                 765

Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile
    770                 775                 780

Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile
785                 790                 795                 800

Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr
                805                 810                 815

Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro
            820                 825                 830

Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys
        835                 840                 845

Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile
    850                 855                 860

Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr
865                 870                 875                 880
```

```
Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg
                885                 890                 895

Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr
            900                 905                 910

Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile
        915                 920                 925

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp
    930                 935                 940

Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
945                 950                 955                 960

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu
                965                 970                 975

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe
            980                 985                 990

Ile Lys Arg Gln Leu Val Glu Thr  Arg Gln Ile Thr Lys  His Val Ala
        995                 1000                1005

Gln Ile  Leu Asp Ser Arg Met  Asn Thr Lys Tyr Asp  Glu Asn Asp
    1010                1015                1020

Lys Leu  Ile Arg Glu Val Lys  Val Ile Thr Leu Lys  Ser Lys Leu
    1025                1030                1035

Val Ser  Asp Phe Arg Lys Asp  Phe Gln Phe Tyr Lys  Val Arg Glu
    1040                1045                1050

Ile Asn  Asn Tyr His His Ala  His Asp Ala Tyr Leu  Asn Ala Val
    1055                1060                1065

Val Gly  Thr Ala Leu Ile Lys  Lys Tyr Pro Lys Leu  Glu Ser Glu
    1070                1075                1080

Phe Val  Tyr Gly Asp Tyr Lys  Val Tyr Asp Val Arg  Lys Met Ile
    1085                1090                1095

Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala Thr Ala  Lys Tyr Phe
    1100                1105                1110

Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys Thr Glu  Ile Thr Leu
    1115                1120                1125
```

```
Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu Ile Glu  Thr Asn Gly
    1130                1135                1140

Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly Arg Asp  Phe Ala Thr
    1145                1150                1155

Val Arg  Lys Val Leu Ser Met  Pro Gln Val Asn Ile  Val Lys Lys
    1160                1165                1170

Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys Glu Ser  Ile Leu Pro
    1175                1180                1185

Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg Lys Lys  Asp Trp Asp
    1190                1195                1200

Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro Thr Val  Ala Tyr Ser
    1205                1210                1215

Val Leu  Val Val Ala Lys Val  Glu Lys Gly Lys Ser  Lys Lys Leu
    1220                1225                1230

Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr Ile Met  Glu Arg Ser
    1235                1240                1245

Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu Glu Ala  Lys Gly Tyr
    1250                1255                1260

Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys Leu Pro  Lys Tyr Ser
    1265                1270                1275

Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg Met Leu  Ala Ser Ala
    1280                1285                1290

Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala Leu Pro  Ser Lys Tyr
    1295                1300                1305

Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr Glu Lys  Leu Lys Gly
    1310                1315                1320

Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu Phe Val  Glu Gln His
    1325                1330                1335

Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln Ile Ser  Glu Phe Ser
    1340                1345                1350

Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu Asp Lys  Val Leu Ser
    1355                1360                1365
```

```
Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile Arg Glu  Gln Ala Glu
    1370                1375                1380

Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn Leu Gly  Ala Pro Ala
    1385                1390                1395

Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp Arg Lys  Arg Tyr Thr
    1400                1405                1410

Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu Ile His  Gln Ser Ile
    1415                1420                1425

Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu Ser Gln  Leu Gly Gly
    1430                1435                1440

Asp Ser  Arg Ala Asp Pro Lys  Lys Lys Arg Lys Val
    1445                1450                1455
```

<210> 49
<211> 1455
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полипептид

<400> 49

```
Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu
1               5                  10                  15

Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp
            20                  25                  30

Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
        35                  40                  45

Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu
    50                  55                  60

Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly Tyr Asp Lys Lys
65                  70                  75                  80

Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val
                85                  90                  95

Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly
            100                 105                 110
```

```
Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu
        115                 120                 125

Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala
    130                 135                 140

Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu
145                 150                 155                 160

Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg
                165                 170                 175

Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His
            180                 185                 190

Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr
        195                 200                 205

Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys
    210                 215                 220

Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe
225                 230                 235                 240

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp
                245                 250                 255

Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe
            260                 265                 270

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu
        275                 280                 285

Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln
    290                 295                 300

Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu
305                 310                 315                 320

Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu
                325                 330                 335

Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp
            340                 345                 350

Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala
        355                 360                 365
```

```
Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val
    370                 375                 380

Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg
385                 390                 395                 400

Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg
                405                 410                 415

Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys
            420                 425                 430

Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe
        435                 440                 445

Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu
    450                 455                 460

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr
465                 470                 475                 480

Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His
                485                 490                 495

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn
            500                 505                 510

Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val
        515                 520                 525

Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys
    530                 535                 540

Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys
545                 550                 555                 560

Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys
                565                 570                 575

Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu
            580                 585                 590

Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu
        595                 600                 605

Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile
    610                 615                 620
```

```
Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu
625                 630                 635                 640

Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile
                645                 650                 655

Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp
            660                 665                 670

Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn
        675                 680                 685

Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp
    690                 695                 700

Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp
705                 710                 715                 720

Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly
                725                 730                 735

Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly
            740                 745                 750

Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn
        755                 760                 765

Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile
    770                 775                 780

Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile
785                 790                 795                 800

Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr
                805                 810                 815

Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro
            820                 825                 830

Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys
        835                 840                 845

Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile
    850                 855                 860

Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr
865                 870                 875                 880
```

```
Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg
                885                 890                 895

Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr
            900                 905                 910

Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile
        915                 920                 925

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp
    930                 935                 940

Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
945                 950                 955                 960

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu
                965                 970                 975

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe
            980                 985                 990

Ile Lys Arg Gln Leu Val Glu Thr  Arg Gln Ile Thr Lys  His Val Ala
        995                 1000                1005

Gln Ile  Leu Asp Ser Arg Met  Asn Thr Lys Tyr Asp  Glu Asn Asp
    1010                1015                1020

Lys Leu  Ile Arg Glu Val Lys  Val Ile Thr Leu Lys  Ser Lys Leu
    1025                1030                1035

Val Ser  Asp Phe Arg Lys Asp  Phe Gln Phe Tyr Lys  Val Arg Glu
    1040                1045                1050

Ile Asn  Asn Tyr His His Ala  His Asp Ala Tyr Leu  Asn Ala Val
    1055                1060                1065

Val Gly  Thr Ala Leu Ile Lys  Lys Tyr Pro Lys Leu  Glu Ser Glu
    1070                1075                1080

Phe Val  Tyr Gly Asp Tyr Lys  Val Tyr Asp Val Arg  Lys Met Ile
    1085                1090                1095

Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala Thr Ala  Lys Tyr Phe
    1100                1105                1110

Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys Thr Glu  Ile Thr Leu
    1115                1120                1125
```

```
Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu Ile Glu  Thr Asn Gly
    1130                1135                1140

Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly Arg Asp  Phe Ala Thr
    1145                1150                1155

Val Arg  Lys Val Leu Ser Met  Pro Gln Val Asn Ile  Val Lys Lys
    1160                1165                1170

Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys Glu Ser  Ile Leu Pro
    1175                1180                1185

Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg Lys Lys  Asp Trp Asp
    1190                1195                1200

Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro Thr Val  Ala Tyr Ser
    1205                1210                1215

Val Leu  Val Val Ala Lys Val  Glu Lys Gly Lys Ser  Lys Lys Leu
    1220                1225                1230

Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr Ile Met  Glu Arg Ser
    1235                1240                1245

Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu Glu Ala  Lys Gly Tyr
    1250                1255                1260

Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys Leu Pro  Lys Tyr Ser
    1265                1270                1275

Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg Met Leu  Ala Ser Ala
    1280                1285                1290

Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala Leu Pro  Ser Lys Tyr
    1295                1300                1305

Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr Glu Lys  Leu Lys Gly
    1310                1315                1320

Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu Phe Val  Glu Gln His
    1325                1330                1335

Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln Ile Ser  Glu Phe Ser
    1340                1345                1350

Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu Asp Lys  Val Leu Ser
    1355                1360                1365
```

```
Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile Arg Glu  Gln Ala Glu
    1370                1375                1380


Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn Leu Gly  Ala Pro Ala
    1385                1390                1395


Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp Arg Lys  Arg Tyr Thr
    1400                1405                1410


Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu Ile His  Gln Ser Ile
    1415                1420                1425


Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu Ser Gln  Leu Gly Gly
    1430                1435                1440


Asp Ser  Arg Ala Asp Pro Lys  Lys Lys Arg Lys Val
    1445                1450                1455
```

<210> 50
<211> 1455
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полипептид

<400> 50

```
Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu
1               5                  10                  15


Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp
            20                  25                  30


Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
        35                  40                  45


Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu
    50                  55                  60


Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly Ser Asp Lys Lys
65                  70                  75                  80


Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val
                85                  90                  95


Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly
            100                 105                 110


Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu
        115                 120                 125
```

```
Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala
    130                 135                 140

Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu
145                 150                 155                 160

Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg
                165                 170                 175

Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His
            180                 185                 190

Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr
        195                 200                 205

Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys
    210                 215                 220

Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe
225                 230                 235                 240

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp
                245                 250                 255

Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe
            260                 265                 270

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu
        275                 280                 285

Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln
    290                 295                 300

Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu
305                 310                 315                 320

Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu
                325                 330                 335

Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp
            340                 345                 350

Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala
        355                 360                 365

Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val
    370                 375                 380
```

```
Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg
385                 390                 395                 400

Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg
                405                 410                 415

Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys
            420                 425                 430

Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe
        435                 440                 445

Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu
    450                 455                 460

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr
465                 470                 475                 480

Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His
                485                 490                 495

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn
            500                 505                 510

Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val
        515                 520                 525

Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys
    530                 535                 540

Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys
545                 550                 555                 560

Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys
                565                 570                 575

Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu
            580                 585                 590

Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu
        595                 600                 605

Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile
    610                 615                 620

Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu
625                 630                 635                 640
```

```
Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile
                645                 650                 655

Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp
            660                 665                 670

Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn
        675                 680                 685

Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp
    690                 695                 700

Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp
705                 710                 715                 720

Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly
                725                 730                 735

Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly
            740                 745                 750

Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn
        755                 760                 765

Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile
    770                 775                 780

Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile
785                 790                 795                 800

Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr
                805                 810                 815

Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro
            820                 825                 830

Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys
        835                 840                 845

Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile
    850                 855                 860

Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr
865                 870                 875                 880

Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg
                885                 890                 895
```

```
Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr
            900                 905                 910

Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile
        915                 920                 925

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp
    930                 935                 940

Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
945                 950                 955                 960

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu
                965                 970                 975

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe
            980                 985                 990

Ile Lys Arg Gln Leu Val Glu Thr  Arg Gln Ile Thr Lys  His Val Ala
        995                 1000                1005

Gln Ile  Leu Asp Ser Arg Met  Asn Thr Lys Tyr Asp  Glu Asn Asp
    1010                1015                1020

Lys Leu  Ile Arg Glu Val Lys  Val Ile Thr Leu Lys  Ser Lys Leu
    1025                1030                1035

Val Ser  Asp Phe Arg Lys Asp  Phe Gln Phe Tyr Lys  Val Arg Glu
    1040                1045                1050

Ile Asn  Asn Tyr His His Ala  His Asp Ala Tyr Leu  Asn Ala Val
    1055                1060                1065

Val Gly  Thr Ala Leu Ile Lys  Lys Tyr Pro Lys Leu  Glu Ser Glu
    1070                1075                1080

Phe Val  Tyr Gly Asp Tyr Lys  Val Tyr Asp Val Arg  Lys Met Ile
    1085                1090                1095

Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala Thr Ala  Lys Tyr Phe
    1100                1105                1110

Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys Thr Glu  Ile Thr Leu
    1115                1120                1125

Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu Ile Glu  Thr Asn Gly
    1130                1135                1140
```

```
Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly Arg Asp  Phe Ala Thr
    1145                1150                1155

Val Arg  Lys Val Leu Ser Met  Pro Gln Val Asn Ile  Val Lys Lys
    1160                1165                1170

Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys Glu Ser  Ile Leu Pro
    1175                1180                1185

Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg Lys Lys  Asp Trp Asp
    1190                1195                1200

Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro Thr Val  Ala Tyr Ser
    1205                1210                1215

Val Leu  Val Val Ala Lys Val  Glu Lys Gly Lys Ser  Lys Lys Leu
    1220                1225                1230

Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr Ile Met  Glu Arg Ser
    1235                1240                1245

Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu Glu Ala  Lys Gly Tyr
    1250                1255                1260

Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys Leu Pro  Lys Tyr Ser
    1265                1270                1275

Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg Met Leu  Ala Ser Ala
    1280                1285                1290

Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala Leu Pro  Ser Lys Tyr
    1295                1300                1305

Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr Glu Lys  Leu Lys Gly
    1310                1315                1320

Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu Phe Val  Glu Gln His
    1325                1330                1335

Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln Ile Ser  Glu Phe Ser
    1340                1345                1350

Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu Asp Lys  Val Leu Ser
    1355                1360                1365

Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile Arg Glu  Gln Ala Glu
    1370                1375                1380
```

```
Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn Leu Gly  Ala Pro Ala
    1385                1390                1395

Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp Arg Lys  Arg Tyr Thr
    1400                1405                1410

Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu Ile His  Gln Ser Ile
    1415                1420                1425

Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu Ser Gln  Leu Gly Gly
    1430                1435                1440

Asp Ser  Arg Ala Asp Pro Lys  Lys Lys Arg Lys Val
    1445                1450                1455
```

<210> 51
<211> 1455
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полипептид

<400> 51

```
Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu
1               5                   10                  15

Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp
            20                  25                  30

Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
        35                  40                  45

Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu
    50                  55                  60

Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly Asn Asp Lys Lys
65                  70                  75                  80

Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val
                85                  90                  95

Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly
            100                 105                 110

Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu
        115                 120                 125
```

```
Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala
    130                 135                 140

Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu
145                 150                 155                 160

Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg
                165                 170                 175

Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His
            180                 185                 190

Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr
        195                 200                 205

Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys
    210                 215                 220

Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe
225                 230                 235                 240

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp
                245                 250                 255

Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe
            260                 265                 270

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu
        275                 280                 285

Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln
    290                 295                 300

Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu
305                 310                 315                 320

Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu
                325                 330                 335

Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp
            340                 345                 350

Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala
        355                 360                 365

Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val
    370                 375                 380
```

```
Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg
385                 390                 395                 400

Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg
                405                 410                 415

Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys
            420                 425                 430

Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe
        435                 440                 445

Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu
    450                 455                 460

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr
465                 470                 475                 480

Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His
                485                 490                 495

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn
            500                 505                 510

Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val
        515                 520                 525

Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys
    530                 535                 540

Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys
545                 550                 555                 560

Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys
                565                 570                 575

Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu
            580                 585                 590

Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu
        595                 600                 605

Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile
    610                 615                 620

Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu
625                 630                 635                 640
```

```
Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile
                645                 650                 655

Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp
            660                 665                 670

Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn
        675                 680                 685

Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp
    690                 695                 700

Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp
705                 710                 715                 720

Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly
                725                 730                 735

Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly
            740                 745                 750

Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn
        755                 760                 765

Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile
    770                 775                 780

Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile
785                 790                 795                 800

Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr
                805                 810                 815

Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro
            820                 825                 830

Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys
        835                 840                 845

Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile
    850                 855                 860

Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr
865                 870                 875                 880

Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg
                885                 890                 895
```

```
Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr
            900                 905                 910

Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile
        915                 920                 925

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp
    930                 935                 940

Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
945                 950                 955                 960

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu
                965                 970                 975

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe
            980                 985                 990

Ile Lys Arg Gln Leu Val Glu Thr  Arg Gln Ile Thr Lys  His Val Ala
        995                 1000                1005

Gln Ile  Leu Asp Ser Arg Met  Asn Thr Lys Tyr Asp  Glu Asn Asp
    1010                1015                1020

Lys Leu  Ile Arg Glu Val Lys  Val Ile Thr Leu Lys  Ser Lys Leu
    1025                1030                1035

Val Ser  Asp Phe Arg Lys Asp  Phe Gln Phe Tyr Lys  Val Arg Glu
    1040                1045                1050

Ile Asn  Asn Tyr His His Ala  His Asp Ala Tyr Leu  Asn Ala Val
    1055                1060                1065

Val Gly  Thr Ala Leu Ile Lys  Lys Tyr Pro Lys Leu  Glu Ser Glu
    1070                1075                1080

Phe Val  Tyr Gly Asp Tyr Lys  Val Tyr Asp Val Arg  Lys Met Ile
    1085                1090                1095

Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala Thr Ala  Lys Tyr Phe
    1100                1105                1110

Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys Thr Glu  Ile Thr Leu
    1115                1120                1125

Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu Ile Glu  Thr Asn Gly
    1130                1135                1140
```

```
Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly Arg Asp  Phe Ala Thr
    1145                1150                1155

Val Arg  Lys Val Leu Ser Met  Pro Gln Val Asn Ile  Val Lys Lys
    1160                1165                1170

Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys Glu Ser  Ile Leu Pro
    1175                1180                1185

Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg Lys Lys  Asp Trp Asp
    1190                1195                1200

Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro Thr Val  Ala Tyr Ser
    1205                1210                1215

Val Leu  Val Val Ala Lys Val  Glu Lys Gly Lys Ser  Lys Lys Leu
    1220                1225                1230

Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr Ile Met  Glu Arg Ser
    1235                1240                1245

Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu Glu Ala  Lys Gly Tyr
    1250                1255                1260

Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys Leu Pro  Lys Tyr Ser
    1265                1270                1275

Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg Met Leu  Ala Ser Ala
    1280                1285                1290

Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala Leu Pro  Ser Lys Tyr
    1295                1300                1305

Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr Glu Lys  Leu Lys Gly
    1310                1315                1320

Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu Phe Val  Glu Gln His
    1325                1330                1335

Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln Ile Ser  Glu Phe Ser
    1340                1345                1350

Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu Asp Lys  Val Leu Ser
    1355                1360                1365

Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile Arg Glu  Gln Ala Glu
    1370                1375                1380
```

```
Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn Leu Gly  Ala Pro Ala
    1385                1390                1395


Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp Arg Lys  Arg Tyr Thr
    1400                1405                1410


Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu Ile His  Gln Ser Ile
    1415                1420                1425


Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu Ser Gln  Leu Gly Gly
    1430                1435                1440


Asp Ser  Arg Ala Asp Pro Lys  Lys Lys Arg Lys Val
    1445                1450                1455
```

<210> 52
<211> 1455
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полипептид
<400> 52

```
Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu
1               5                   10                  15


Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp
            20                  25                  30


Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
        35                  40                  45


Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu
    50                  55                  60


Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly Lys Asp Lys Lys
65                  70                  75                  80


Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val
                85                  90                  95


Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly
            100                 105                 110


Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu
        115                 120                 125


Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala
    130                 135                 140
```

```
Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu
145                 150                 155                 160

Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg
                165                 170                 175

Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His
            180                 185                 190

Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr
        195                 200                 205

Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys
    210                 215                 220

Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe
225                 230                 235                 240

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp
                245                 250                 255

Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe
            260                 265                 270

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu
        275                 280                 285

Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln
    290                 295                 300

Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu
305                 310                 315                 320

Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu
                325                 330                 335

Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp
            340                 345                 350

Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala
        355                 360                 365

Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val
    370                 375                 380

Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg
385                 390                 395                 400
```

```
Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg
                405                 410                 415

Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys
            420                 425                 430

Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe
        435                 440                 445

Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu
    450                 455                 460

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr
465                 470                 475                 480

Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His
                485                 490                 495

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn
            500                 505                 510

Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val
        515                 520                 525

Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys
    530                 535                 540

Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys
545                 550                 555                 560

Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys
                565                 570                 575

Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu
            580                 585                 590

Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu
        595                 600                 605

Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile
    610                 615                 620

Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu
625                 630                 635                 640

Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile
                645                 650                 655
```

```
Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp
            660                 665                 670

Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn
        675                 680                 685

Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp
    690                 695                 700

Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp
705                 710                 715                 720

Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly
                725                 730                 735

Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly
            740                 745                 750

Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn
        755                 760                 765

Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile
    770                 775                 780

Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile
785                 790                 795                 800

Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr
                805                 810                 815

Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro
            820                 825                 830

Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys
        835                 840                 845

Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile
    850                 855                 860

Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr
865                 870                 875                 880

Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg
                885                 890                 895

Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr
            900                 905                 910
```

```
Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile
        915                 920                 925

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp
    930                 935                 940

Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
945                 950                 955                 960

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu
                965                 970                 975

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe
            980                 985                 990

Ile Lys Arg Gln Leu Val Glu Thr  Arg Gln Ile Thr Lys  His Val Ala
        995                 1000                1005

Gln Ile  Leu Asp Ser Arg Met  Asn Thr Lys Tyr Asp  Glu Asn Asp
    1010                1015                1020

Lys Leu  Ile Arg Glu Val Lys  Val Ile Thr Leu Lys  Ser Lys Leu
    1025                1030                1035

Val Ser  Asp Phe Arg Lys Asp  Phe Gln Phe Tyr Lys  Val Arg Glu
    1040                1045                1050

Ile Asn  Asn Tyr His His Ala  His Asp Ala Tyr Leu  Asn Ala Val
    1055                1060                1065

Val Gly  Thr Ala Leu Ile Lys  Lys Tyr Pro Lys Leu  Glu Ser Glu
    1070                1075                1080

Phe Val  Tyr Gly Asp Tyr Lys  Val Tyr Asp Val Arg  Lys Met Ile
    1085                1090                1095

Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala Thr Ala  Lys Tyr Phe
    1100                1105                1110

Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys Thr Glu  Ile Thr Leu
    1115                1120                1125

Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu Ile Glu  Thr Asn Gly
    1130                1135                1140

Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly Arg Asp  Phe Ala Thr
    1145                1150                1155
```

```
Val Arg  Lys Val Leu Ser Met  Pro Gln Val Asn Ile  Val Lys Lys
    1160                1165                1170

Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys Glu Ser  Ile Leu Pro
    1175                1180                1185

Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg Lys Lys  Asp Trp Asp
    1190                1195                1200

Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro Thr Val  Ala Tyr Ser
    1205                1210                1215

Val Leu  Val Val Ala Lys Val  Glu Lys Gly Lys Ser  Lys Lys Leu
    1220                1225                1230

Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr Ile Met  Glu Arg Ser
    1235                1240                1245

Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu Glu Ala  Lys Gly Tyr
    1250                1255                1260

Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys Leu Pro  Lys Tyr Ser
    1265                1270                1275

Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg Met Leu  Ala Ser Ala
    1280                1285                1290

Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala Leu Pro  Ser Lys Tyr
    1295                1300                1305

Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr Glu Lys  Leu Lys Gly
    1310                1315                1320

Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu Phe Val  Glu Gln His
    1325                1330                1335

Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln Ile Ser  Glu Phe Ser
    1340                1345                1350

Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu Asp Lys  Val Leu Ser
    1355                1360                1365

Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile Arg Glu  Gln Ala Glu
    1370                1375                1380

Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn Leu Gly  Ala Pro Ala
    1385                1390                1395
```

Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr
1400 1405 1410

Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile
1415 1420 1425

Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly
1430 1435 1440

Asp Ser Arg Ala Asp Pro Lys Lys Lys Arg Lys Val
1445 1450 1455

<210> 53
<211> 1455
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полипептид

<400> 53
Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu
1 5 10 15

Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp
20 25 30

Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
35 40 45

Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu
50 55 60

Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly Cys Asp Lys Lys
65 70 75 80

Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val
85 90 95

Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly
100 105 110

Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu
115 120 125

Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala
130 135 140

Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu

```
145                 150                 155                 160

Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg
                165                 170                 175

Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His
            180                 185                 190

Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr
        195                 200                 205

Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys
    210                 215                 220

Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe
225                 230                 235                 240

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp
                245                 250                 255

Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe
            260                 265                 270

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu
        275                 280                 285

Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln
    290                 295                 300

Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu
305                 310                 315                 320

Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu
                325                 330                 335

Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp
            340                 345                 350

Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala
        355                 360                 365

Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val
    370                 375                 380

Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg
385                 390                 395                 400

Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg
```

405 410 415

Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys
420 425 430

Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe
435 440 445

Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu
450 455 460

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr
465 470 475 480

Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His
485 490 495

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn
500 505 510

Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val
515 520 525

Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys
530 535 540

Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys
545 550 555 560

Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys
565 570 575

Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu
580 585 590

Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu
595 600 605

Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile
610 615 620

Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu
625 630 635 640

Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile
645 650 655

Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp

```
            660                 665                 670

Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn
        675                 680                 685

Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp
    690                 695                 700

Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp
705                 710                 715                 720

Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly
                725                 730                 735

Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly
            740                 745                 750

Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn
        755                 760                 765

Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile
    770                 775                 780

Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile
785                 790                 795                 800

Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr
                805                 810                 815

Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro
            820                 825                 830

Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys
        835                 840                 845

Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile
    850                 855                 860

Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr
865                 870                 875                 880

Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg
                885                 890                 895

Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr
            900                 905                 910

Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile
```

```
        915                 920                 925

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp
    930                 935                 940

Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
945                 950                 955                 960

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu
                965                 970                 975

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe
            980                 985                 990

Ile Lys Arg Gln Leu Val Glu Thr  Arg Gln Ile Thr Lys  His Val Ala
        995                 1000                1005

Gln Ile  Leu Asp Ser Arg Met  Asn Thr Lys Tyr Asp  Glu Asn Asp
    1010                1015                1020

Lys Leu  Ile Arg Glu Val Lys  Val Ile Thr Leu Lys  Ser Lys Leu
    1025                1030                1035

Val Ser  Asp Phe Arg Lys Asp  Phe Gln Phe Tyr Lys  Val Arg Glu
    1040                1045                1050

Ile Asn  Asn Tyr His His Ala  His Asp Ala Tyr Leu  Asn Ala Val
    1055                1060                1065

Val Gly  Thr Ala Leu Ile Lys  Lys Tyr Pro Lys Leu  Glu Ser Glu
    1070                1075                1080

Phe Val  Tyr Gly Asp Tyr Lys  Val Tyr Asp Val Arg  Lys Met Ile
    1085                1090                1095

Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala Thr Ala  Lys Tyr Phe
    1100                1105                1110

Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys Thr Glu  Ile Thr Leu
    1115                1120                1125

Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu Ile Glu  Thr Asn Gly
    1130                1135                1140

Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly Arg Asp  Phe Ala Thr
    1145                1150                1155

Val Arg  Lys Val Leu Ser Met  Pro Gln Val Asn Ile  Val Lys Lys
```

```
    1160                1165                1170

Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys Glu Ser  Ile Leu Pro
    1175                1180                1185

Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg Lys Lys  Asp Trp Asp
    1190                1195                1200

Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro Thr Val  Ala Tyr Ser
    1205                1210                1215

Val Leu  Val Val Ala Lys Val  Glu Lys Gly Lys Ser  Lys Lys Leu
    1220                1225                1230

Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr Ile Met  Glu Arg Ser
    1235                1240                1245

Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu Glu Ala  Lys Gly Tyr
    1250                1255                1260

Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys Leu Pro  Lys Tyr Ser
    1265                1270                1275

Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg Met Leu  Ala Ser Ala
    1280                1285                1290

Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala Leu Pro  Ser Lys Tyr
    1295                1300                1305

Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr Glu Lys  Leu Lys Gly
    1310                1315                1320

Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu Phe Val  Glu Gln His
    1325                1330                1335

Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln Ile Ser  Glu Phe Ser
    1340                1345                1350

Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu Asp Lys  Val Leu Ser
    1355                1360                1365

Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile Arg Glu  Gln Ala Glu
    1370                1375                1380

Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn Leu Gly  Ala Pro Ala
    1385                1390                1395

Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp Arg Lys  Arg Tyr Thr
```

1400 1405 1410

Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile
1415 1420 1425

Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly
1430 1435 1440

Asp Ser Arg Ala Asp Pro Lys Lys Lys Arg Lys Val
1445 1450 1455

<210> 54
<211> 1455
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полипептид

<400> 54
Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu
1 5 10 15

Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp
20 25 30

Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
35 40 45

Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu
50 55 60

Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly Asp Asp Lys Lys
65 70 75 80

Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val
85 90 95

Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly
100 105 110

Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu
115 120 125

Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala
130 135 140

Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu
145 150 155 160

```
Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg
                165                 170                 175

Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His
            180                 185                 190

Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr
        195                 200                 205

Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys
    210                 215                 220

Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe
225                 230                 235                 240

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp
                245                 250                 255

Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe
            260                 265                 270

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu
        275                 280                 285

Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln
    290                 295                 300

Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu
305                 310                 315                 320

Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu
                325                 330                 335

Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp
            340                 345                 350

Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala
        355                 360                 365

Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val
    370                 375                 380

Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg
385                 390                 395                 400

Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg
                405                 410                 415
```

```
Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys
            420                 425                 430

Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe
        435                 440                 445

Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu
    450                 455                 460

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr
465                 470                 475                 480

Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His
                485                 490                 495

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn
            500                 505                 510

Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val
        515                 520                 525

Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys
    530                 535                 540

Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys
545                 550                 555                 560

Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys
                565                 570                 575

Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu
            580                 585                 590

Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu
        595                 600                 605

Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile
    610                 615                 620

Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu
625                 630                 635                 640

Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile
                645                 650                 655

Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp
            660                 665                 670
```

```
Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn
        675                 680                 685

Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp
    690                 695                 700

Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp
705                 710                 715                 720

Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly
                725                 730                 735

Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly
            740                 745                 750

Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn
        755                 760                 765

Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile
    770                 775                 780

Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile
785                 790                 795                 800

Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr
                805                 810                 815

Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro
            820                 825                 830

Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys
        835                 840                 845

Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile
    850                 855                 860

Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr
865                 870                 875                 880

Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg
                885                 890                 895

Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr
            900                 905                 910

Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile
        915                 920                 925
```

```
Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp
    930                 935                 940

Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
945                 950                 955                 960

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu
                965                 970                 975

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe
            980                 985                 990

Ile Lys Arg Gln Leu Val Glu Thr  Arg Gln Ile Thr Lys  His Val Ala
        995                 1000                1005

Gln Ile  Leu Asp Ser Arg Met  Asn Thr Lys Tyr Asp  Glu Asn Asp
    1010                1015                1020

Lys Leu  Ile Arg Glu Val Lys  Val Ile Thr Leu Lys  Ser Lys Leu
    1025                1030                1035

Val Ser  Asp Phe Arg Lys Asp  Phe Gln Phe Tyr Lys  Val Arg Glu
    1040                1045                1050

Ile Asn  Asn Tyr His His Ala  His Asp Ala Tyr Leu  Asn Ala Val
    1055                1060                1065

Val Gly  Thr Ala Leu Ile Lys  Lys Tyr Pro Lys Leu  Glu Ser Glu
    1070                1075                1080

Phe Val  Tyr Gly Asp Tyr Lys  Val Tyr Asp Val Arg  Lys Met Ile
    1085                1090                1095

Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala Thr Ala  Lys Tyr Phe
    1100                1105                1110

Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys Thr Glu  Ile Thr Leu
    1115                1120                1125

Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu Ile Glu  Thr Asn Gly
    1130                1135                1140

Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly Arg Asp  Phe Ala Thr
    1145                1150                1155

Val Arg  Lys Val Leu Ser Met  Pro Gln Val Asn Ile  Val Lys Lys
    1160                1165                1170
```

```
Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys Glu Ser  Ile Leu Pro
    1175                1180                1185

Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg Lys Lys  Asp Trp Asp
    1190                1195                1200

Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro Thr Val  Ala Tyr Ser
    1205                1210                1215

Val Leu  Val Val Ala Lys Val  Glu Lys Gly Lys Ser  Lys Lys Leu
    1220                1225                1230

Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr Ile Met  Glu Arg Ser
    1235                1240                1245

Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu Glu Ala  Lys Gly Tyr
    1250                1255                1260

Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys Leu Pro  Lys Tyr Ser
    1265                1270                1275

Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg Met Leu  Ala Ser Ala
    1280                1285                1290

Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala Leu Pro  Ser Lys Tyr
    1295                1300                1305

Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr Glu Lys  Leu Lys Gly
    1310                1315                1320

Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu Phe Val  Glu Gln His
    1325                1330                1335

Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln Ile Ser  Glu Phe Ser
    1340                1345                1350

Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu Asp Lys  Val Leu Ser
    1355                1360                1365

Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile Arg Glu  Gln Ala Glu
    1370                1375                1380

Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn Leu Gly  Ala Pro Ala
    1385                1390                1395

Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp Arg Lys  Arg Tyr Thr
    1400                1405                1410
```

Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile
1415 1420 1425

Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly
1430 1435 1440

Asp Ser Arg Ala Asp Pro Lys Lys Lys Arg Lys Val
1445 1450 1455

<210> 55
<211> 1455
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полипептид

<400> 55
Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu
1 5 10 15

Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp
20 25 30

Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
35 40 45

Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu
50 55 60

Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly Phe Asp Lys Lys
65 70 75 80

Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val
85 90 95

Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly
100 105 110

Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu
115 120 125

Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala
130 135 140

Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu
145 150 155 160

Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg

```
                165                 170                 175

Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His
            180                 185                 190

Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr
        195                 200                 205

Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys
    210                 215                 220

Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe
225                 230                 235                 240

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp
                245                 250                 255

Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe
            260                 265                 270

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu
        275                 280                 285

Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln
    290                 295                 300

Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu
305                 310                 315                 320

Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu
                325                 330                 335

Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp
            340                 345                 350

Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala
        355                 360                 365

Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val
    370                 375                 380

Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg
385                 390                 395                 400

Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg
                405                 410                 415

Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys
```

```
            420                 425                 430

Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe
        435                 440                 445

Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu
    450                 455                 460

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr
465                 470                 475                 480

Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His
                485                 490                 495

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn
            500                 505                 510

Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val
        515                 520                 525

Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys
    530                 535                 540

Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys
545                 550                 555                 560

Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys
                565                 570                 575

Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu
            580                 585                 590

Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu
        595                 600                 605

Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile
    610                 615                 620

Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu
625                 630                 635                 640

Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile
                645                 650                 655

Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp
            660                 665                 670

Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn
```

```
        675                 680                 685

Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp
    690                 695                 700

Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp
705                 710                 715                 720

Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly
                725                 730                 735

Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly
            740                 745                 750

Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn
        755                 760                 765

Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile
    770                 775                 780

Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile
785                 790                 795                 800

Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr
                805                 810                 815

Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro
            820                 825                 830

Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys
        835                 840                 845

Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile
    850                 855                 860

Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr
865                 870                 875                 880

Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg
                885                 890                 895

Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr
            900                 905                 910

Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile
        915                 920                 925

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp
```

```
    930                 935                 940

Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
945                 950                 955                 960

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu
                965                 970                 975

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe
            980                 985                 990

Ile Lys Arg Gln Leu Val Glu Thr  Arg Gln Ile Thr Lys  His Val Ala
        995                 1000                1005

Gln Ile  Leu Asp Ser Arg Met  Asn Thr Lys Tyr Asp  Glu Asn Asp
    1010                1015                1020

Lys Leu  Ile Arg Glu Val Lys  Val Ile Thr Leu Lys  Ser Lys Leu
    1025                1030                1035

Val Ser  Asp Phe Arg Lys Asp  Phe Gln Phe Tyr Lys  Val Arg Glu
    1040                1045                1050

Ile Asn  Asn Tyr His His Ala  His Asp Ala Tyr Leu  Asn Ala Val
    1055                1060                1065

Val Gly  Thr Ala Leu Ile Lys  Lys Tyr Pro Lys Leu  Glu Ser Glu
    1070                1075                1080

Phe Val  Tyr Gly Asp Tyr Lys  Val Tyr Asp Val Arg  Lys Met Ile
    1085                1090                1095

Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala Thr Ala  Lys Tyr Phe
    1100                1105                1110

Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys Thr Glu  Ile Thr Leu
    1115                1120                1125

Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu Ile Glu  Thr Asn Gly
    1130                1135                1140

Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly Arg Asp  Phe Ala Thr
    1145                1150                1155

Val Arg  Lys Val Leu Ser Met  Pro Gln Val Asn Ile  Val Lys Lys
    1160                1165                1170

Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys Glu Ser  Ile Leu Pro
```

```
    1175                1180                1185

Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg Lys Lys  Asp Trp Asp
    1190                1195                1200

Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro Thr Val  Ala Tyr Ser
    1205                1210                1215

Val Leu  Val Val Ala Lys Val  Glu Lys Gly Lys Ser  Lys Lys Leu
    1220                1225                1230

Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr Ile Met  Glu Arg Ser
    1235                1240                1245

Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu Glu Ala  Lys Gly Tyr
    1250                1255                1260

Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys Leu Pro  Lys Tyr Ser
    1265                1270                1275

Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg Met Leu  Ala Ser Ala
    1280                1285                1290

Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala Leu Pro  Ser Lys Tyr
    1295                1300                1305

Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr Glu Lys  Leu Lys Gly
    1310                1315                1320

Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu Phe Val  Glu Gln His
    1325                1330                1335

Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln Ile Ser  Glu Phe Ser
    1340                1345                1350

Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu Asp Lys  Val Leu Ser
    1355                1360                1365

Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile Arg Glu  Gln Ala Glu
    1370                1375                1380

Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn Leu Gly  Ala Pro Ala
    1385                1390                1395

Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp Arg Lys  Arg Tyr Thr
    1400                1405                1410

Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu Ile His  Gln Ser Ile
```

1415 1420 1425

Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly
1430 1435 1440

Asp Ser Arg Ala Asp Pro Lys Lys Lys Arg Lys Val
1445 1450 1455

<210> 56
<211> 1455
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полипептид

<400> 56
Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu
1 5 10 15

Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp
20 25 30

Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
35 40 45

Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu
50 55 60

Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly Glu Asp Lys Lys
65 70 75 80

Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val
85 90 95

Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly
100 105 110

Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu
115 120 125

Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala
130 135 140

Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu
145 150 155 160

Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg
165 170 175

```
Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His
            180                 185                 190

Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr
        195                 200                 205

Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys
    210                 215                 220

Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe
225                 230                 235                 240

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp
                245                 250                 255

Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe
            260                 265                 270

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu
        275                 280                 285

Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln
    290                 295                 300

Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu
305                 310                 315                 320

Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu
                325                 330                 335

Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp
            340                 345                 350

Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala
        355                 360                 365

Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val
    370                 375                 380

Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg
385                 390                 395                 400

Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg
                405                 410                 415

Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys
            420                 425                 430
```

```
Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe
        435                 440                 445

Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu
    450                 455                 460

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr
465                 470                 475                 480

Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His
                485                 490                 495

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn
            500                 505                 510

Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val
        515                 520                 525

Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys
    530                 535                 540

Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys
545                 550                 555                 560

Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys
                565                 570                 575

Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu
            580                 585                 590

Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu
        595                 600                 605

Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile
    610                 615                 620

Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu
625                 630                 635                 640

Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile
                645                 650                 655

Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp
            660                 665                 670

Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn
        675                 680                 685
```

```
Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp
    690                 695                 700

Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp
705                 710                 715                 720

Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly
                725                 730                 735

Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly
            740                 745                 750

Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn
        755                 760                 765

Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile
    770                 775                 780

Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile
785                 790                 795                 800

Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr
                805                 810                 815

Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro
            820                 825                 830

Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys
        835                 840                 845

Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile
    850                 855                 860

Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr
865                 870                 875                 880

Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg
                885                 890                 895

Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr
            900                 905                 910

Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile
        915                 920                 925

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp
    930                 935                 940
```

```
Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
945                 950                 955                 960

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu
                965                 970                 975

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe
            980                 985                 990

Ile Lys Arg Gln Leu Val Glu Thr  Arg Gln Ile Thr Lys  His Val Ala
        995                 1000                1005

Gln Ile  Leu Asp Ser Arg Met  Asn Thr Lys Tyr Asp  Glu Asn Asp
    1010                1015                1020

Lys Leu  Ile Arg Glu Val Lys  Val Ile Thr Leu Lys  Ser Lys Leu
    1025                1030                1035

Val Ser  Asp Phe Arg Lys Asp  Phe Gln Phe Tyr Lys  Val Arg Glu
    1040                1045                1050

Ile Asn  Asn Tyr His His Ala  His Asp Ala Tyr Leu  Asn Ala Val
    1055                1060                1065

Val Gly  Thr Ala Leu Ile Lys  Lys Tyr Pro Lys Leu  Glu Ser Glu
    1070                1075                1080

Phe Val  Tyr Gly Asp Tyr Lys  Val Tyr Asp Val Arg  Lys Met Ile
    1085                1090                1095

Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala Thr Ala  Lys Tyr Phe
    1100                1105                1110

Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys Thr Glu  Ile Thr Leu
    1115                1120                1125

Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu Ile Glu  Thr Asn Gly
    1130                1135                1140

Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly Arg Asp  Phe Ala Thr
    1145                1150                1155

Val Arg  Lys Val Leu Ser Met  Pro Gln Val Asn Ile  Val Lys Lys
    1160                1165                1170

Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys Glu Ser  Ile Leu Pro
    1175                1180                1185
```

```
Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg Lys Lys  Asp Trp Asp
    1190                1195                1200

Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro Thr Val  Ala Tyr Ser
    1205                1210                1215

Val Leu  Val Val Ala Lys Val  Glu Lys Gly Lys Ser  Lys Lys Leu
    1220                1225                1230

Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr Ile Met  Glu Arg Ser
    1235                1240                1245

Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu Glu Ala  Lys Gly Tyr
    1250                1255                1260

Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys Leu Pro  Lys Tyr Ser
    1265                1270                1275

Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg Met Leu  Ala Ser Ala
    1280                1285                1290

Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala Leu Pro  Ser Lys Tyr
    1295                1300                1305

Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr Glu Lys  Leu Lys Gly
    1310                1315                1320

Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu Phe Val  Glu Gln His
    1325                1330                1335

Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln Ile Ser  Glu Phe Ser
    1340                1345                1350

Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu Asp Lys  Val Leu Ser
    1355                1360                1365

Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile Arg Glu  Gln Ala Glu
    1370                1375                1380

Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn Leu Gly  Ala Pro Ala
    1385                1390                1395

Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp Arg Lys  Arg Tyr Thr
    1400                1405                1410

Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu Ile His  Gln Ser Ile
    1415                1420                1425
```

Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly
1430 1435 1440

Asp Ser Arg Ala Asp Pro Lys Lys Lys Arg Lys Val
1445 1450 1455

<210> 57
<211> 1455
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полипептид

<400> 57
Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu
1 5 10 15

Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp
20 25 30

Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
35 40 45

Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu
50 55 60

Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly Arg Asp Lys Lys
65 70 75 80

Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val
85 90 95

Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly
100 105 110

Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu
115 120 125

Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala
130 135 140

Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu
145 150 155 160

Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg
165 170 175

Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His

```
            180                 185                 190

Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr
        195                 200                 205

Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys
    210                 215                 220

Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe
225                 230                 235                 240

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp
                245                 250                 255

Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe
            260                 265                 270

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu
        275                 280                 285

Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln
    290                 295                 300

Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu
305                 310                 315                 320

Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu
                325                 330                 335

Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp
            340                 345                 350

Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala
        355                 360                 365

Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val
    370                 375                 380

Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg
385                 390                 395                 400

Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg
                405                 410                 415

Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys
            420                 425                 430

Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe
```

```
        435                 440                 445

Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu
    450                 455                 460

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr
465                 470                 475                 480

Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His
                485                 490                 495

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn
            500                 505                 510

Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val
        515                 520                 525

Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys
    530                 535                 540

Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys
545                 550                 555                 560

Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys
                565                 570                 575

Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu
            580                 585                 590

Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu
        595                 600                 605

Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile
    610                 615                 620

Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu
625                 630                 635                 640

Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile
                645                 650                 655

Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp
            660                 665                 670

Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn
        675                 680                 685

Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp
```

```
    690                 695                 700

Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp
705                 710                 715                 720

Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly
                725                 730                 735

Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly
            740                 745                 750

Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn
        755                 760                 765

Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile
    770                 775                 780

Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile
785                 790                 795                 800

Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr
                805                 810                 815

Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro
            820                 825                 830

Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys
        835                 840                 845

Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile
    850                 855                 860

Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr
865                 870                 875                 880

Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg
                885                 890                 895

Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr
            900                 905                 910

Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile
        915                 920                 925

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp
    930                 935                 940

Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
```

```
945                 950                 955                 960

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu
                965                 970                 975

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe
            980                 985                 990

Ile Lys Arg Gln Leu Val Glu Thr  Arg Gln Ile Thr Lys  His Val Ala
        995                 1000                1005

Gln Ile  Leu Asp Ser Arg Met  Asn Thr Lys Tyr Asp  Glu Asn Asp
    1010                1015                1020

Lys Leu  Ile Arg Glu Val Lys  Val Ile Thr Leu Lys  Ser Lys Leu
    1025                1030                1035

Val Ser  Asp Phe Arg Lys Asp  Phe Gln Phe Tyr Lys  Val Arg Glu
    1040                1045                1050

Ile Asn  Asn Tyr His His Ala  His Asp Ala Tyr Leu  Asn Ala Val
    1055                1060                1065

Val Gly  Thr Ala Leu Ile Lys  Lys Tyr Pro Lys Leu  Glu Ser Glu
    1070                1075                1080

Phe Val  Tyr Gly Asp Tyr Lys  Val Tyr Asp Val Arg  Lys Met Ile
    1085                1090                1095

Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala Thr Ala  Lys Tyr Phe
    1100                1105                1110

Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys Thr Glu  Ile Thr Leu
    1115                1120                1125

Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu Ile Glu  Thr Asn Gly
    1130                1135                1140

Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly Arg Asp  Phe Ala Thr
    1145                1150                1155

Val Arg  Lys Val Leu Ser Met  Pro Gln Val Asn Ile  Val Lys Lys
    1160                1165                1170

Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys Glu Ser  Ile Leu Pro
    1175                1180                1185

Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg Lys Lys  Asp Trp Asp
```

```
    1190                1195                1200

Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro Thr Val  Ala Tyr Ser
    1205                1210                1215

Val Leu  Val Val Ala Lys Val  Glu Lys Gly Lys Ser  Lys Lys Leu
    1220                1225                1230

Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr Ile Met  Glu Arg Ser
    1235                1240                1245

Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu Glu Ala  Lys Gly Tyr
    1250                1255                1260

Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys Leu Pro  Lys Tyr Ser
    1265                1270                1275

Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg Met Leu  Ala Ser Ala
    1280                1285                1290

Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala Leu Pro  Ser Lys Tyr
    1295                1300                1305

Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr Glu Lys  Leu Lys Gly
    1310                1315                1320

Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu Phe Val  Glu Gln His
    1325                1330                1335

Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln Ile Ser  Glu Phe Ser
    1340                1345                1350

Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu Asp Lys  Val Leu Ser
    1355                1360                1365

Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile Arg Glu  Gln Ala Glu
    1370                1375                1380

Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn Leu Gly  Ala Pro Ala
    1385                1390                1395

Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp Arg Lys  Arg Tyr Thr
    1400                1405                1410

Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu Ile His  Gln Ser Ile
    1415                1420                1425

Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu Ser Gln  Leu Gly Gly
```

1430 1435 1440

Asp Ser Arg Ala Asp Pro Lys Lys Lys Arg Lys Val
1445 1450 1455

<210> 58
<211> 1455
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полипептид

<400> 58
Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu
1 5 10 15

Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp
20 25 30

Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
35 40 45

Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu
50 55 60

Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly Gln Asp Lys Lys
65 70 75 80

Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val
85 90 95

Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly
100 105 110

Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu
115 120 125

Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala
130 135 140

Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu
145 150 155 160

Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg
165 170 175

Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His
180 185 190

```
Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr
        195                 200                 205

Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys
    210                 215                 220

Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe
225                 230                 235                 240

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp
                245                 250                 255

Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe
            260                 265                 270

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu
        275                 280                 285

Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln
    290                 295                 300

Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu
305                 310                 315                 320

Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu
                325                 330                 335

Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp
            340                 345                 350

Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala
        355                 360                 365

Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val
    370                 375                 380

Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg
385                 390                 395                 400

Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg
                405                 410                 415

Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys
            420                 425                 430

Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe
        435                 440                 445
```

```
Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu
    450                 455                 460

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr
465                 470                 475                 480

Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His
                485                 490                 495

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn
            500                 505                 510

Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val
        515                 520                 525

Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys
    530                 535                 540

Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys
545                 550                 555                 560

Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys
                565                 570                 575

Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu
            580                 585                 590

Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu
        595                 600                 605

Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile
    610                 615                 620

Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu
625                 630                 635                 640

Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile
                645                 650                 655

Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp
            660                 665                 670

Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn
        675                 680                 685

Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp
    690                 695                 700
```

```
Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp
705                 710                 715                 720

Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly
                725                 730                 735

Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly
            740                 745                 750

Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn
        755                 760                 765

Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile
    770                 775                 780

Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile
785                 790                 795                 800

Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr
                805                 810                 815

Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro
            820                 825                 830

Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys
        835                 840                 845

Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile
    850                 855                 860

Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr
865                 870                 875                 880

Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg
                885                 890                 895

Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr
            900                 905                 910

Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile
        915                 920                 925

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp
    930                 935                 940

Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
945                 950                 955                 960
```

```
Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu
               965                 970                 975

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe
           980                 985                 990

Ile Lys Arg Gln Leu Val Glu Thr  Arg Gln Ile Thr Lys  His Val Ala
        995                1000                1005

Gln Ile  Leu Asp Ser Arg Met  Asn Thr Lys Tyr Asp  Glu Asn Asp
    1010                1015                1020

Lys Leu  Ile Arg Glu Val Lys  Val Ile Thr Leu Lys  Ser Lys Leu
    1025                1030                1035

Val Ser  Asp Phe Arg Lys Asp  Phe Gln Phe Tyr Lys  Val Arg Glu
    1040                1045                1050

Ile Asn  Asn Tyr His His Ala  His Asp Ala Tyr Leu  Asn Ala Val
    1055                1060                1065

Val Gly  Thr Ala Leu Ile Lys  Lys Tyr Pro Lys Leu  Glu Ser Glu
    1070                1075                1080

Phe Val  Tyr Gly Asp Tyr Lys  Val Tyr Asp Val Arg  Lys Met Ile
    1085                1090                1095

Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala Thr Ala  Lys Tyr Phe
    1100                1105                1110

Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys Thr Glu  Ile Thr Leu
    1115                1120                1125

Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu Ile Glu  Thr Asn Gly
    1130                1135                1140

Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly Arg Asp  Phe Ala Thr
    1145                1150                1155

Val Arg  Lys Val Leu Ser Met  Pro Gln Val Asn Ile  Val Lys Lys
    1160                1165                1170

Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys Glu Ser  Ile Leu Pro
    1175                1180                1185

Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg Lys Lys  Asp Trp Asp
    1190                1195                1200
```

```
Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro Thr Val  Ala Tyr Ser
    1205                 1210                 1215

Val Leu  Val Val Ala Lys Val  Glu Lys Gly Lys Ser  Lys Lys Leu
    1220                 1225                 1230

Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr Ile Met  Glu Arg Ser
    1235                 1240                 1245

Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu Glu Ala  Lys Gly Tyr
    1250                 1255                 1260

Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys Leu Pro  Lys Tyr Ser
    1265                 1270                 1275

Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg Met Leu  Ala Ser Ala
    1280                 1285                 1290

Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala Leu Pro  Ser Lys Tyr
    1295                 1300                 1305

Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr Glu Lys  Leu Lys Gly
    1310                 1315                 1320

Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu Phe Val  Glu Gln His
    1325                 1330                 1335

Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln Ile Ser  Glu Phe Ser
    1340                 1345                 1350

Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu Asp Lys  Val Leu Ser
    1355                 1360                 1365

Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile Arg Glu  Gln Ala Glu
    1370                 1375                 1380

Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn Leu Gly  Ala Pro Ala
    1385                 1390                 1395

Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp Arg Lys  Arg Tyr Thr
    1400                 1405                 1410

Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu Ile His  Gln Ser Ile
    1415                 1420                 1425

Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu Ser Gln  Leu Gly Gly
    1430                 1435                 1440
```

Asp Ser Arg Ala Asp Pro Lys Lys Lys Arg Lys Val
1445 1450 1455

<210> 59
<211> 1454
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полипептид

<400> 59
Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu
1 5 10 15

Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp
20 25 30

Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
35 40 45

Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu
50 55 60

Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Val Asp Lys Lys Tyr
65 70 75 80

Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile
85 90 95

Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn
100 105 110

Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe
115 120 125

Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg
130 135 140

Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile
145 150 155 160

Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu
165 170 175

Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro
180 185 190

Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro

```
        195                 200                 205

Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala
    210                 215                 220

Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg
225                 230                 235                 240

Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val
                245                 250                 255

Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu
            260                 265                 270

Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser
        275                 280                 285

Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu
    290                 295                 300

Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser
305                 310                 315                 320

Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp
                325                 330                 335

Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn
            340                 345                 350

Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala
        355                 360                 365

Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn
    370                 375                 380

Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr
385                 390                 395                 400

Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln
                405                 410                 415

Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn
            420                 425                 430

Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr
        435                 440                 445

Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu
```

```
    450                 455                 460

Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe
465                 470                 475                 480

Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His Ala
                485                 490                 495

Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg
            500                 505                 510

Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly
        515                 520                 525

Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser
    530                 535                 540

Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly
545                 550                 555                 560

Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn
                565                 570                 575

Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr
            580                 585                 590

Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly
        595                 600                 605

Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val
    610                 615                 620

Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys
625                 630                 635                 640

Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser
                645                 650                 655

Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu
            660                 665                 670

Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu
        675                 680                 685

Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg
    690                 695                 700

Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp
```

705 710 715 720

Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg
725 730 735

Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys
740 745 750

Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe
755 760 765

Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln
770 775 780

Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile Ala
785 790 795 800

Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val
805 810 815

Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu
820 825 830

Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly
835 840 845

Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys
850 855 860

Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln
865 870 875 880

Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp
885 890 895

Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp
900 905 910

Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp
915 920 925

Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn
930 935 940

Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln
945 950 955 960

Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr

```
                965                 970                 975

Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile
            980                 985                 990

Lys Arg Gln Leu Val Glu Thr Arg  Gln Ile Thr Lys His  Val Ala Gln
        995                 1000                1005

Ile Leu  Asp Ser Arg Met Asn  Thr Lys Tyr Asp Glu  Asn Asp Lys
    1010                1015                1020

Leu Ile  Arg Glu Val Lys Val  Ile Thr Leu Lys Ser  Lys Leu Val
    1025                1030                1035

Ser Asp  Phe Arg Lys Asp Phe  Gln Phe Tyr Lys Val  Arg Glu Ile
    1040                1045                1050

Asn Asn  Tyr His His Ala His  Asp Ala Tyr Leu Asn  Ala Val Val
    1055                1060                1065

Gly Thr  Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
    1070                1075                1080

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1085                1090                1095

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1100                1105                1110

Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1115                1120                1125

Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1130                1135                1140

Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1145                1150                1155

Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1160                1165                1170

Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1175                1180                1185

Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1190                1195                1200

Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr Val Ala  Tyr Ser Val
```

```
    1205                1210                1215

Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1220                1225                1230

Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1235                1240                1245

Phe Glu  Lys Asn Pro Ile Asp  Phe Leu Glu Ala Lys  Gly Tyr Lys
    1250                1255                1260

Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1265                1270                1275

Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1280                1285                1290

Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1295                1300                1305

Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1310                1315                1320

Pro Glu  Asp Asn Glu Gln Lys  Gln Leu Phe Val Glu  Gln His Lys
    1325                1330                1335

His Tyr  Leu Asp Glu Ile Ile  Glu Gln Ile Ser Glu  Phe Ser Lys
    1340                1345                1350

Arg Val  Ile Leu Ala Asp Ala  Asn Leu Asp Lys Val  Leu Ser Ala
    1355                1360                1365

Tyr Asn  Lys His Arg Asp Lys  Pro Ile Arg Glu Gln  Ala Glu Asn
    1370                1375                1380

Ile Ile  His Leu Phe Thr Leu  Thr Asn Leu Gly Ala  Pro Ala Ala
    1385                1390                1395

Phe Lys  Tyr Phe Asp Thr Thr  Ile Asp Arg Lys Arg  Tyr Thr Ser
    1400                1405                1410

Thr Lys  Glu Val Leu Asp Ala  Thr Leu Ile His Gln  Ser Ile Thr
    1415                1420                1425

Gly Leu  Tyr Glu Thr Arg Ile  Asp Leu Ser Gln Leu  Gly Gly Asp
    1430                1435                1440

Ser Arg  Ala Asp Pro Lys Lys  Lys Arg Lys Val
```

1445 1450

<210> 60
<211> 228
<212> ДНК
<213> Неизвестная

<220>
<223> Описание неизвестной:
последовательность убиквитина

<400> 60
accacctctc agacgcagga ccaggtgcag ggtcgactcc ttctggatgt tgtagtcaga 60

aagagtgcgg ccatcttcca gctgcttgcc tgcaaagatg agcctctgct ggtcgggagg 120

gatgccttct ttatcctgga tcttggcctt cacattttcg atggtgtcac tgggctccac 180

ttccagggtg atggtcttgc cggtcagggt cttcacgaag atctgcat 228

<210> 61
<211> 4134
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический
полинуклеотид

<400> 61
gacaagaagt actccattgg gctcgatatc ggcacaaaca gcgtcggctg ggccgtcatt 60

acggacgagt acaaggtgcc gagcaaaaaa ttcaaagttc tgggcaatac cgatcgccac 120

agcataaaga agaacctcat tggcgccctc ctgttcgact ccggggagac ggccgaagcc 180

acgcggctca aaagaacagc acggcgcaga tatacccgca gaaagaatcg gatctgctac 240

ctgcaggaga tctttagtaa tgagatggct aaggtggatg actctttctt ccataggctg 300

gaggagtcct ttttggtgga ggaggataaa aagcacgagc gccacccaat ctttggcaat 360

atcgtggacg aggtggcgta ccatgaaaag tacccaacca tatatcatct gaggaagaag 420

cttgtagaca gtactgataa ggctgacttg cggttgatct atctcgcgct ggcgcatatg 480

atcaaatttc ggggacactt cctcatcgag ggggacctga acccagacaa cagcgatgtc 540

gacaaactct ttatccaact ggttcagact tacaatcagc ttttcgaaga gaacccgatc 600

aacgcatccg gagttgacgc caaagcaatc ctgagcgcta ggctgtccaa atcccggcgg 660

ctcgaaaacc tcatcgcaca gctccctggg gagaagaaga acggcctgtt tggtaatctt 720

atcgccctgt cactcgggct gacccccaac tttaaatcta acttcgacct ggccgaagat 780

gccaagcttc aactgagcaa agacacctac gatgatgatc tcgacaatct gctggcccag 840

atcggcgacc agtacgcaga ccttttttttg gcggcaaaga acctgtcaga cgccattctg 900

ctgagtgata ttctgcgagt gaacacggag atcaccaaag ctccgctgag cgctagtatg 960

atcaagcgct atgatgagca ccaccaagac ttgactttgc tgaaggccct tgtcagacag 1020

caactgcctg agaagtacaa ggaaattttc ttcgatcagt ctaaaaatgg ctacgccgga 1080

tacattgacg gcggagcaag ccaggaggaa ttttacaaat ttattaagcc catcttggaa 1140

aaaatggacg gcaccgagga gctgctggta aagcttaaca gagaagatct gttgcgcaaa 1200

cagcgcactt tcgacaatgg aagcatcccc caccagattc acctgggcga actgcacgct 1260

atcctcaggc ggcaagagga tttctacccc tttttgaaag ataacaggga aaagattgag 1320

aaaatcctca catttcggat accctactat gtaggccccc tcgcccgggg aaattccaga 1380

ttcgcgtgga tgactcgcaa atcagaagag accatcactc cctggaactt cgaggaagtc 1440

gtggataagg gggcctctgc ccagtccttc atcgaaagga tgactaactt tgataaaaat 1500

ctgcctaacg aaaaggtgct tcctaaacac tctctgctgt acgagtactt cacagtttat 1560

aacgagctca ccaaggtcaa atacgtcaca gaagggatga gaaagccagc attcctgtct 1620

ggagagcaga agaaagctat cgtggacctc ctcttcaaga cgaaccggaa agttaccgtg 1680

aaacagctca aagaagacta tttcaaaaag attgaatgtt tcgactctgt tgaaatcagc 1740

ggagtggagg atcgcttcaa cgcatccctg ggaacgtatc acgatctcct gaaaatcatt 1800

aaagacaagg acttcctgga caatgaggag aacgaggaca ttcttgagga cattgtcctc 1860

acccttacgt tgtttgaaga tagggagatg attgaagaac gcttgaaaac ttacgctcat 1920

ctcttcgacg acaaagtcat gaaacagctc aagaggcgcc gatatacagg atgggggcgg 1980

ctgtcaagaa aactgatcaa tgggatccga gacaagcaga gtggaaagac aatcctggat 2040

tttcttaagt ccgatggatt tgccaaccgg aacttcatgc agttgatcca tgatgactct 2100

ctcaccttta aggaggacat ccagaaagca caagtttctg gccaggggga cagtcttcac 2160

gagcacatcg ctaatcttgc aggtagccca gctatcaaaa agggaatact gcagaccgtt 2220

aaggtcgtgg atgaactcgt caaagtaatg ggaaggcata agcccgagaa tatcgttatc 2280

gagatggccc gagagaacca aactacccag aagggacaga agaacagtag ggaaaggatg 2340

aagaggattg aagagggtat aaaagaactg gggtcccaaa tccttaagga acacccagtt 2400

gaaaacaccc agcttcagaa tgagaagctc tacctgtact acctgcagaa cggcagggac 2460

atgtacgtgg atcaggaact ggacatcaat cggctctccg actacgacgt ggatcatatc 2520

gtgccccagt cttttctcaa agatgattct attgataata aagtgttgac aagatccgat 2580

aaaaatagag ggaagagtga taacgtcccc tcagaagaag ttgtcaagaa aatgaaaaat 2640

tattggcggc agctgctgaa cgccaaactg atcacacaac ggaagttcga taatctgact 2700

aaggctgaac gaggtggcct gtctgagttg gataaagccg gcttcatcaa aaggcagctt 2760

gttgagacac gccagatcac caagcacgtg gcccaaattc tcgattcacg catgaacacc 2820

aagtacgatg aaaatgacaa actgattcga gaggtgaaag ttattactct gaagtctaag 2880

ctggtctcag atttcagaaa ggactttcag ttttataagg tgagagagat caacaattac 2940

caccatgcgc atgatgccta cctgaatgca gtggtaggca ctgcacttat caaaaaatat 3000

cccaagcttg aatctgaatt tgtttacgga gactataaag tgtacgatgt taggaaaatg 3060

atcgcaaagt ctgagcagga aataggcaag gccaccgcta agtacttctt ttacagcaat 3120

attatgaatt ttttcaagac cgagattaca ctggccaatg gagagattcg gaagcgacca 3180

cttatcgaaa caaacggaga aacaggagaa atcgtgtggg acaagggtag ggatttcgcg 3240

acagtccgga aggtcctgtc catgccgcag gtgaacatcg ttaaaaagac cgaagtacag 3300

accggaggct tctccaagga aagtatcctc ccgaaaagga acagcgacaa gctgatcgca 3360

cgcaaaaaag attgggaccc caagaaatac ggcggattcg attctcctac agtcgcttac 3420

agtgtactgg ttgtggccaa agtggagaaa gggaagtcta aaaaactcaa aagcgtcaag 3480

gaactgctgg gcatcacaat catggagcga tcaagcttcg aaaaaaaccc catcgacttt 3540

ctcgaggcga aaggatataa agaggtcaaa aaagacctca tcattaagct tcccaagtac 3600

tctctctttg agcttgaaaa cggccggaaa cgaatgctcg ctagtgcggg cgagctgcag 3660

aaaggtaacg agctggcact gccctctaaa tacgttaatt tcttgtatct ggccagccac 3720

tatgaaaagc tcaaagggtc tcccgaagat aatgagcaga agcagctgtt cgtggaacaa 3780

cacaaacact accttgatga gatcatcgag caaataagcg aattctccaa aagagtgatc 3840

ctcgccgacg ctaacctcga taaggtgctt tctgcttaca ataagcacag ggataagccc 3900

atcagggagc aggcagaaaa cattatccac ttgtttactc tgaccaactt gggcgcgcct 3960

gcagccttca agtacttcga caccaccata gacagaaagc ggtacacctc tacaaaggag 4020

gtcctggacg ccacactgat tcatcagtca attacggggc tctatgaaac aagaatcgac 4080

ctctctcagc tcggtggaga cagcagggct gaccccaaga agaagaggaa ggtg 4134

<210> 62
<211> 6929
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полинуклеотид

<400> 62
cctgctgtct ccaccgagct gagagaggtc gattcttgtt tcatagagcc ccgtaattga 60

ctgatgaatc agtgtggcgt ccaggacctc ctttgtagag gtgtaccgct ttctgtctat 120

ggtggtgtcg aagtacttga aggctgcagg cgcgcccaag ttggtcagag taaacaagtg 180

gataatgttt tctgcctgct ccctgatggg cttatccctg tgcttattgt aagcagaaag 240

caccttatcg aggttagcgt cggcgaggat cactcttttg gagaattcgc ttatttgctc 300

gatgatctca tcaaggtagt gtttgtgttg ttccacgaac agctgcttct gctcattatc 360

ttcgggagac cctttgagct tttcatagtg gctggccaga tacaagaaat taacgtattt 420

agagggcagt gccagctcgt tacctttctg cagctcgccc gcactagcga gcattcgttt 480

ccggccgttt tcaagctcaa agagagagta cttgggaagc ttaatgatga ggtctttttt 540

gacctcttta tatcctttcg cctcgagaaa gtcgatgggg ttttttttcga agcttgatcg 600

ctccatgatt gtgatgccca gcagttcctt gacgcttttg agtttttttag acttcccttt 660

ctccactttg gccacaacca gtacactgta agcgactgta ggagaatcga atccgccgta 720

tttcttgggg tcccaatctt ttttgcgtgc gatcagcttg tcgctgttcc ttttcgggag 780

gatactttcc ttggagaagc ctccggtctg tacttcggtc tttttaacga tgttcacctg 840

cggcatggac aggaccttcc ggactgtcgc gaaatcccta cccttgtccc acacgatttc 900

tcctgtttct ccgtttgttt cgataagtgg tcgcttccga atctctccat tggccagtgt 960

aatctcggtc ttgaaaaaat tcataatatt gctgtaaaag aagtacttag cggtggcctt 1020

gcctatttcc tgctcagact ttgcgatcat tttcctaaca tcgtacactt tatagtctcc 1080

gtaaacaaat tcagattcaa gcttgggata ttttttgata agtgcagtgc ctaccactgc 1140

attcaggtag gcatcatgcg catggtggta attgttgatc tctctcacct tataaaactg 1200

aaagtccttt ctgaaatctg agaccagctt agacttcaga gtaataactt tcacctctcg 1260

aatcagtttg tcattttcat cgtacttggt gttcatgcgt gaatcgagaa tttgggccac 1320

gtgcttggtg atctggcgtg tctcaacaag ctgccttttg atgaagccgg ctttatccaa 1380

ctcagacagg ccacctcgtt cagccttagt cagattatcg aacttccgtt gtgtgatcag 1440

tttggcgttc agcagctgcc gccaataatt tttcattttc ttgacaactt cttctgaggg 1500

gacgttatca ctcttccctc tatttttatc ggatcttgtc aacactttat tatcaataga 1560

atcatctttg agaaaagact ggggcacgat atgatccacg tcgtagtcgg agagccgatt 1620

gatgtccagt tcctgatcca cgtacatgtc cctgccgttc tgcaggtagt acaggtagag 1680

cttctcattc tgaagctggg tgttttcaac tgggtgttcc ttaaggattt gggaccccag 1740

ttcttttata ccctcttcaa tcctcttcat cctttcccta ctgttcttct gtcccttctg 1800

ggtagtttgg ttctctcggg ccatctcgat aacgatattc tcgggcttat gccttcccat 1860

tactttgacg agttcatcca cgaccttaac ggtctgcagt attccctttt tgatagctgg 1920

gctacctgca agattagcga tgtgctcgtg aagactgtcc ccctggccag aaacttgtgc 1980

tttctggatg tcctccttaa aggtgagaga gtcatcatgg atcaactgca tgaagttccg 2040

gttggcaaat ccatcggact taagaaaatc caggattgtc tttccactct gcttgtctcg 2100

gatcccattg atcagttttc ttgacagccg cccccatcct gtatatcggc gcctcttgag 2160

ctgtttcatg actttgtcgt cgaagagatg agcgtaagtt ttcaagcgtt cttcaatcat 2220

ctccctatct tcaaacaacg taagggtgag gacaatgtcc tcaagaatgt cctcgttctc 2280

ctcattgtcc aggaagtcct tgtctttaat gattttcagg agatcgtgat acgttcccag 2340

ggatgcgttg aagcgatcct ccactccgct gatttcaaca gagtcgaaac attcaatctt 2400

tttgaaatag tcttctttga gctgtttcac ggtaactttc cggttcgtct tgaagaggag 2460

gtccacgata gctttcttct gctctccaga caggaatgct ggctttctca tcccttctgt 2520

gacgtatttg accttggtga gctcgttata aactgtgaag tactcgtaca gcagagagtg 2580

tttaggaagc acctttтcgt taggcagatt tttatcaaag ttagtcatcc tttcgatgaa 2640

ggactgggca gaggcccccт tatccacgac ttcctcgaag ttccagggag tgatggtctc 2700

ttctgatttg cgagtcatcc acgcgaatct ggaatttccc cgggcgaggg ggcctacata 2760

gtagggtatc cgaaatgtga ggattttctc aatcttttcc ctgttatctt tcaaaaaggg 2820

gtagaaatcc tcttgccgcc tgaggatagc gtgcagttcg cccaggtgaa tctggtgggg 2880

gatgcttcca ttgtcgaaag tgcgctgttt gcgcaacaga tcttctctgt taagctttac 2940

cagcagctcc tcggtgccgt ccatttttтc caagatgggc ttaataaatt tgtaaaattc 3000

ctcctggctt gctccgccgt caatgtatcc ggcgtagcca tttttagact gatcgaagaa 3060

aatttccttg tacttctcag gcagttgctg tctgacaagg gccttcagca aagtcaagtc 3120

ttggtggtgc tcatcatagc gcttgatcat actagcgctc agcggagctt tggtgatctc 3180

cgtgttcact cgcagaatat cactcagcag aatggcgtct gacaggttct ttgccgccaa 3240

aaaaaggtct gcgtactggt cgccgatctg ggccagcaga ttgtcgagat catcatcgta 3300

ggtgtctttg ctcagttgaa gcttggcatc ttcggccagg tcgaagttag atttaaagtt 3360

gggggtcagc ccgagtgaca gggcgataag attaccaaac aggccgttct tcttctcccc 3420

agggagctgt gcgatgaggt tttcgagccg ccgggatttg gacagcctag cgctcaggat 3480

tgctttggcg tcaactccgg atgcgttgat cgggttctct tcgaaaagct gattgtaagt 3540

ctgaaccagt tggataaaga gtttgtcgac atcgctgttg tctgggttca ggtccccctc 3600

gatgaggaag tgtccccgaa atttgatcat atgcgccagc gcgagataga tcaaccgcaa 3660

gtcagcctta tcagtactgt ctacaagctt cttcctcaga tgatatatgg ttgggtactt 3720

ttcatggtac gccacctcgt ccacgatatt gccaaagatt gggtggcgct cgtgcttttt 3780

atcctcctcc accaaaaagg actcctccag cctatggaag aaagagtcat ccaccttagc 3840

catctcatta ctaaagatct cctgcaggta gcagatccga ttctttctgc gggtatatct 3900

gcgccgtgct gttcttttga gccgcgtggc ttcggccgtc tccccggagt cgaacaggag 3960

ggcgccaatg aggttcttct ttatgctgtg gcgatcggta ttgcccagaa ctttgaattt 4020

tttgctcggc accttgtact cgtccgtaat gacggcccag ccgacgctgt ttgtgccgat 4080

atcgagccca atggagtact tcttgtccat ggtggcggct cttgaaggac gacgtcatca 4140

tcccttgccc ggatgcgcgg gcttcttgtc tagcacagga gcctggggta gagcgcatgc 4200

aaattacgcg ctgtgctttg tgggaaatca ccctaaacga aaaatttatt cctctttcga 4260

gccttatagt ggcggccggt ctacatccta ggttttagag ctagaaatag caagttaaaa 4320

taaggctagt ccgttatcaa cttgaaaaag tggcaccgag tcggtgcttt tttgttaacg 4380

cggccgccta gagtcgacct gcaggcatgc aagcttggcg taatcatggt catagctgtt 4440

tcctgtgtga aattgttatc cgctcacaat tccacacaac atacgagccg gaagcataaa 4500

gtgtaaagcc tggggtgcct aatgagtgag ctaactcaca ttaattgcgt tgcgctcact 4560

gcccgctttc cagtcgggaa acctgtcgtg ccagctgcat taatgaatcg gccaacgcgc 4620

ggggagaggc ggtttgcgta ttgggcgctc ttccgcttcc tcgctcactg actcgctgcg 4680

ctcggtcgtt cggctgcggc gagcggtatc agctcactca aaggcggtaa tacggttatc 4740

cacagaatca ggggataacg caggaaagaa catgtgagca aaaggccagc aaaaggccag 4800

gaaccgtaaa aaggccgcgt tgctggcgtt tttccatagg ctccgccccc ctgacgagca 4860

tcacaaaaat cgacgctcaa gtcagaggtg gcgaaacccg acaggactat aaagatacca 4920

ggcgtttccc cctggaagct ccctcgtgcg ctctcctgtt ccgaccctgc cgcttaccgg 4980

atacctgtcc gcctttctcc cttcgggaag cgtggcgctt tctcatagct cacgctgtag 5040

gtatctcagt tcggtgtagg tcgttcgctc caagctgggc tgtgtgcacg aacccccccgt 5100

tcagcccgac cgctgcgcct tatccggtaa ctatcgtctt gagtccaacc cggtaagaca 5160

cgacttatcg ccactggcag cagccactgg taacaggatt agcagagcga ggtatgtagg 5220

cggtgctaca gagttcttga agtggtggcc taactacggc tacactagaa gaacagtatt 5280

tggtatctgc gctctgctga agccagttac cttcggaaaa agagttggta gctcttgatc 5340

cggcaaacaa accaccgctg gtagcggtgg ttttttttgtt tgcaagcagc agattacgcg 5400

cagaaaaaaa ggatctcaag aagatccttt gatcttttct acggggtctg acgctcagtg 5460

gaacgaaaac tcacgttaag ggattttggt catgagatta tcaaaaagga tcttcaccta 5520

gatcctttta aattaaaaat gaagttttaa atcaatctaa agtatatatg agtaaacttg 5580

gtctgacagt taccaatgct taatcagtga ggcacctatc tcagcgatct gtctatttcg 5640

ttcatccata gttgcctgac tccccgtcgt gtagataact acgatacggg agggcttacc 5700

atctggcccc agtgctgcaa tgataccgcg agacccacgc tcaccggctc cagatttatc 5760

agcaataaac cagccagccg gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc 5820

ctccatccag tctattaatt gttgccggga agctagagta agtagttcgc cagttaatag 5880

tttgcgcaac gttgttgcca ttgctacagg catcgtggtg tcacgctcgt cgtttggtat 5940

ggcttcattc agctccggtt cccaacgatc aaggcgagtt acatgatccc ccatgttgtg 6000

caaaaaagcg gttagctcct tcggtcctcc gatcgttgtc agaagtaagt tggccgcagt 6060

gttatcactc atggttatgg cagcactgca taattctctt actgtcatgc catccgtaag 6120

atgcttttct gtgactggtg agtactcaac caagtcattc tgagaatagt gtatgcggcg 6180

accgagttgc tcttgcccgg cgtcaatacg ggataatacc gcgccacata gcagaacttt 6240

aaaagtgctc atcattggaa aacgttcttc ggggcgaaaa ctctcaagga tcttaccgct 6300

gttgagatcc agttcgatgt aacccactcg tgcacccaac tgatcttcag catcttttac 6360

tttcaccagc gtttctgggt gagcaaaaac aggaaggcaa aatgccgcaa aaaagggaat 6420

aagggcgaca cggaaatgtt gaatactcat actcttcctt tttcaatatt attgaagcat 6480

ttatcagggt tattgtctca tgagcggata catatttgaa tgtatttaga aaaataaaca 6540

aatagggttt ccgcgcacat ttccccgaaa agtgccacct gacgtctaag aaaccattat 6600

tatcatgaca ttaacctata aaaataggcg tatcacgagg ccctttcgtc tcgcgcgttt 6660

cggtgatgac ggtgaaaacc tctgacacat gcagctcccg gagacggtca cagcttgtct 6720

gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg ttggcgggtg 6780

tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc accatagcgg 6840

ccgcggtacc cacacaaaaa accaacacac agatgtaatg aaaataaaga tattttattt 6900

cacaccttcc tcttcttctt ggggtcagc 6929

<210> 63
<211> 7177
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полинуклеотид

<400> 63
ctgcaggcat gcaagcttgg cgtaatcatg gtcatagctg tttcctgtgt gaaattgtta 60

tccgctcaca attccacaca acatacgagc cggaagcata aagtgtaaag cctggggtgc 120

ctaatgagtg agctaactca cattaattgc gttgcgctca ctgcccgctt tccagtcggg 180

aaacctgtcg tgccagctgc attaatgaat cggccaacgc gcggggagag gcggtttgcg 240

tattgggcgc tcttccgctt cctcgctcac tgactcgctg cgctcggtcg ttcggctgcg 300

gcgagcggta tcagctcact caaaggcggt aatacggtta tccacagaat caggggataa 360

cgcaggaaag aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta aaaaggccgc 420

gttgctggcg tttttccata ggctccgccc ccctgacgag catcacaaaa atcgacgctc 480

aagtcagagg tggcgaaacc cgacaggact ataaagatac caggcgtttc cccctggaag 540

ctccctcgtg cgctctcctg ttccgaccct gccgcttacc ggatacctgt ccgcctttct 600

cccttcggga agcgtggcgc tttctcatag ctcacgctgt aggtatctca gttcggtgta 660

ggtcgttcgc tccaagctgg gctgtgtgca cgaaccccc gttcagcccg accgctgcgc 720

cttatccggt aactatcgtc ttgagtccaa cccggtaaga cacgacttat cgccactggc 780

agcagccact ggtaacagga ttagcagagc gaggtatgta ggcggtgcta cagagttctt 840

gaagtggtgg cctaactacg gctacactag aagaacagta tttggtatct gcgctctgct 900

gaagccagtt accttcggaa aaagagttgg tagctcttga tccggcaaac aaaccaccgc 960

tggtagcggt ggtttttttg tttgcaagca gcagattacg cgcagaaaaa aaggatctca 1020

agaagatcct ttgatctttt ctacggggtc tgacgctcag tggaacgaaa actcacgtta 1080

agggattttg gtcatgagat tatcaaaaag gatcttcacc tagatccttt taaattaaaa 1140

atgaagtttt aaatcaatct aaagtatata tgagtaaact tggtctgaca gttaccaatg 1200

cttaatcagt gaggcaccta tctcagcgat ctgtctattt cgttcatcca tagttgcctg 1260

actccccgtc gtgtagataa ctacgatacg ggagggctta ccatctggcc ccagtgctgc 1320

aatgataccg cgagacccac gctcaccggc tccagattta tcagcaataa accagccagc 1380

cggaagggcc gagcgcagaa gtggtcctgc aactttatcc gcctccatcc agtctattaa 1440

ttgttgccgg gaagctagag taagtagttc gccagttaat agtttgcgca acgttgttgc 1500

cattgctaca ggcatcgtgg tgtcacgctc gtcgtttggt atggcttcat tcagctccgg 1560

ttcccaacga tcaaggcgag ttacatgatc ccccatgttg tgcaaaaaag cggttagctc 1620

cttcggtcct ccgatcgttg tcagaagtaa gttggccgca gtgttatcac tcatggttat 1680

ggcagcactg cataattctc ttactgtcat gccatccgta agatgctttt ctgtgactgg 1740

tgagtactca accaagtcat tctgagaata gtgtatgcgg cgaccgagtt gctcttgccc 1800

ggcgtcaata cgggataata ccgcgccaca tagcagaact ttaaaagtgc tcatcattgg 1860

aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat ccagttcgat 1920

gtaacccact cgtgcaccca actgatcttc agcatctttt actttcacca gcgtttctgg 1980

gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga ataagggcga cacggaaatg 2040

ttgaatactc atactcttcc tttttcaata ttattgaagc atttatcagg gttattgtct 2100

catgagcgga tacatatttg aatgtattta gaaaaataaa caaatagggg ttccgcgcac 2160

atttccccga aaagtgccac ctgacgtcta agaaaccatt attatcatga cattaaccta 2220

taaaaatagg cgtatcacga ggccctttcg tctcgcgcgt ttcggtgatg acggtgaaaa 2280

cctctgacac atgcagctcc cggagacggt cacagcttgt ctgtaagcgg atgccgggag 2340

cagacaagcc cgtcagggcg cgtcagcggg tgttggcggg tgtcggggct ggcttaacta 2400

tgcggcatca gagcagattg tactgagagt gcaccatagc ggccgcggta cccacacaaa 2460

aaaccaacac acagatgtaa tgaaaataaa gatattttat ttcacacctt cctcttcttc 2520

ttggggtcag ccctgctgtc tccaccgagc tgagagaggt cgattcttgt ttcatagagc 2580

cccgtaattg actgatgaat cagtgtggcg tccaggacct cctttgtaga ggtgtaccgc 2640

tttctgtcta tggtggtgtc gaagtacttg aaggctgcag gcgcgcccaa gttggtcaga 2700

gtaaacaagt ggataatgtt ttctgcctgc tccctgatgg gcttatccct gtgcttattg 2760

taagcagaaa gcaccttatc gaggttagcg tcggcgagga tcactctttt ggagaattcg 2820

cttatttgct cgatgatctc atcaaggtag tgtttgtgtt gttccacgaa cagctgcttc 2880

tgctcattat cttcgggaga ccctttgagc ttttcatagt ggctggccag atacaagaaa 2940

ttaacgtatt tagagggcag tgccagctcg ttacctttct gcagctcgcc cgcactagcg 3000

agcattcgtt tccggccgtt ttcaagctca aagagagagt acttgggaag cttaatgatg 3060

aggtcttttt tgacctcttt atatcctttc gcctcgagaa agtcgatggg gtttttttcg 3120

aagcttgatc gctccatgat tgtgatgccc agcagttcct tgacgctttt gagtttttta 3180

gacttccctt tctccacttt ggccacaacc agtacactgt aagcgactgt aggagaatcg 3240

aatccgccgt atttcttggg gtcccaatct tttttgcgtg cgatcagctt gtcgctgttc 3300

cttttcggga ggatactttc cttggagaag cctccggtct gtacttcggt ctttttaacg 3360

atgttcacct gcggcatgga caggaccttc cggactgtcg cgaaatccct acccttgtcc 3420

cacacgattt ctcctgtttc tccgtttgtt tcgataagtg gtcgcttccg aatctctcca 3480

ttggccagtg taatctcggt cttgaaaaaa ttcataatat tgctgtaaaa gaagtactta 3540

gcggtggcct tgcctatttc ctgctcagac tttgcgatca ttttcctaac atcgtacact 3600

ttatagtctc cgtaaacaaa ttcagattca agcttgggat atttttgat aagtgcagtg 3660

cctaccactg cattcaggta ggcatcatgc gcatggtggt aattgttgat ctctctcacc 3720

ttataaaact gaaagtcctt tctgaaatct gagaccagct tagacttcag agtaataact 3780

ttcacctctc gaatcagttt gtcattttca tcgtacttgg tgttcatgcg tgaatcgaga 3840

atttgggcca cgtgcttggt gatctggcgt gtctcaacaa gctgcctttt gatgaagccg 3900

gctttatcca actcagacag gccacctcgt tcagccttag tcagattatc gaacttccgt 3960

tgtgtgatca gtttggcgtt cagcagctgc cgccaataat ttttcatttt cttgacaact 4020

tcttctgagg ggacgttatc actcttccct ctatttttat cggatcttgt caacacttta 4080

ttatcaatag aatcatcttt gagaaaagac tggggcacga tatgatccac gtcgtagtcg 4140

gagagccgat tgatgtccag ttcctgatcc acgtacatgt ccctgccgtt ctgcaggtag 4200

tacaggtaga gcttctcatt ctgaagctgg gtgttttcaa ctgggtgttc cttaaggatt 4260

tgggacccca gttcttttat accctcttca atcctcttca tcctttccct actgttcttc 4320

tgtcccttct gggtagtttg gttctctcgg gccatctcga taacgatatt ctcgggctta 4380

tgccttccca ttactttgac gagttcatcc acgaccttaa cggtctgcag tattcccttt 4440

ttgatagctg ggctacctgc aagattagcg atgtgctcgt gaagactgtc cccctggcca 4500

gaaacttgtg ctttctggat gtcctcctta aaggtgagag agtcatcatg gatcaactgc 4560

atgaagttcc ggttggcaaa tccatcggac ttaagaaaat ccaggattgt ctttccactc 4620

tgcttgtctc ggatcccatt gatcagtttt cttgacagcc gcccccatcc tgtatatcgg 4680

cgcctcttga gctgtttcat gactttgtcg tcgaagagat gagcgtaagt tttcaagcgt 4740

tcttcaatca tctccctatc ttcaaacaac gtaagggtga ggacaatgtc ctcaagaatg 4800

tcctcgttct cctcattgtc caggaagtcc ttgtctttaa tgattttcag gagatcgtga 4860

tacgttccca gggatgcgtt gaagcgatcc tccactccgc tgatttcaac agagtcgaaa 4920

cattcaatct ttttgaaata gtcttctttg agctgtttca cggtaacttt ccggttcgtc 4980

ttgaagagga ggtccacgat agctttcttc tgctctccag acaggaatgc tggctttctc 5040

atcccttctg tgacgtattt gaccttggtg agctcgttat aaactgtgaa gtactcgtac 5100

agcagagagt gtttaggaag caccttttcg ttaggcagat ttttatcaaa gttagtcatc 5160

ctttcgatga aggactgggc agaggccccc ttatccacga cttcctcgaa gttccaggga 5220

gtgatggtct cttctgattt gcgagtcatc cacgcgaatc tggaatttcc ccgggcgagg 5280

gggcctacat agtagggtat ccgaaatgtg aggattttct caatcttttc cctgttatct 5340

ttcaaaaagg ggtagaaatc ctcttgccgc ctgaggatag cgtgcagttc gcccaggtga 5400

atctggtggg ggatgcttcc attgtcgaaa gtgcgctgtt tgcgcaacag atcttctctg 5460

ttaagcttta ccagcagctc ctcggtgccg tccatttttt ccaagatggg cttaataaat 5520

ttgtaaaatt cctcctggct tgctccgccg tcaatgtatc cggcgtagcc atttttagac 5580

tgatcgaaga aaatttcctt gtacttctca ggcagttgct gtctgacaag ggccttcagc 5640

aaagtcaagt cttggtggtg ctcatcatag cgcttgatca tactagcgct cagcggagct 5700

ttggtgatct ccgtgttcac tcgcagaata tcactcagca gaatggcgtc tgacaggttc 5760

tttgccgcca aaaaaaggtc tgcgtactgg tcgccgatct gggccagcag attgtcgaga 5820

tcatcatcgt aggtgtcttt gctcagttga agcttggcat cttcggccag gtcgaagtta 5880

gatttaaagt tgggggtcag cccgagtgac agggcgataa gattaccaaa caggccgttc 5940

ttcttctccc cagggagctg tgcgatgagg ttttcgagcc gccgggattt ggacagccta 6000

gcgctcagga ttgctttggc gtcaactccg gatgcgttga tcgggttctc ttcgaaaagc 6060

tgattgtaag tctgaaccag ttggataaag agtttgtcga catcgctgtt gtctgggttc 6120

aggtcccccc cgatgaggaa gtgtccccga aatttgatca tatgcgccag cgcgagatag 6180

atcaaccgca agtcagcctt atcagtactg tctacaagct tcttcctcag atgatatatg 6240

gttgggtact tttcatggta cgccacctcg tccacgatat tgccaaagat tgggtggcgc 6300

tcgtgctttt tatcctcctc caccaaaaag gactcctcca gcctatggaa gaaagagtca 6360

tccaccttag ccatctcatt actaaagatc tcctgcaggt agcagatccg attctttctg 6420

cgggtatatc tgcgccgtgc tgttcttttg agccgcgtgg cttcggccgt ctccccggag 6480

tcgaacagga gggcgccaat gaggttcttc tttatgctgt ggcgatcggt attgcccaga 6540

actttgaatt ttttgctcgg caccttgtac tcgtccgtaa tgacggccca gccgacgctg 6600

tttgtgccga tatcgagccc aatggagtac ttcttgtcca tggtggcggc tcttgaagga 6660

cgacgtcatc atcccttgcc cggatgcgcg ggcttcttgt ctagcacagg agcctggggt 6720

agagcgcatg caaattacgc gctgtgcttt gtgggaaatc accctaaacg aaaaatttat 6780

tcctctttcg agccttatag tggcggccgg tctacatcct aggttttaga gctagaaata 6840

gcaagttaaa ataaggctag tccgttatca acttgaaaaa gtggcaccga gtcggtgctt 6900

ttttgttgat atggtgctgt gtaggctcat atgtggatct cagaacccac atgtactctg 6960

ctccccaggt cttggtgcgc tttctattct ctgtcagcaa agcctatcga agaatcacct 7020

accacaactg gcgccacggc ttcaatgtag cccagaccat gtttacccta ctcatggtac 7080

gtatgtaaat tggatgggct agatgaatca gagggctggg gcaaggacca cagctaacta 7140

tcttctggcc caaggaacgc ggccgcctag agtcgac 7177

<210> 64
<211> 7181
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полинуклеотид

<400> 64
atgtaatgaa aataaagata ttttatttca caccttcctc ttcttcttgg ggtcagccct 60

gctgtctcca ccgagctgag agaggtcgat tcttgtttca tagagccccg taattgactg 120

atgaatcagt gtggcgtcca ggacctcctt tgtagaggtg taccgctttc tgtctatggt 180

ggtgtcgaag tacttgaagg ctgcaggcgc gcccaagttg gtcagagtaa acaagtggat 240

aatgttttct gcctgctccc tgatgggctt atccctgtgc ttattgtaag cagaaagcac 300

cttatcgagg ttagcgtcgg cgaggatcac tcttttggag aattcgctta tttgctcgat 360

gatctcatca aggtagtgtt tgtgttgttc cacgaacagc tgcttctgct cattatcttc 420

gggagaccct ttgagctttt catagtggct ggccagatac aagaaattaa cgtatttaga 480

gggcagtgcc agctcgttac ctttctgcag ctcgcccgca ctagcgagca ttcgtttccg 540

gccgttttca agctcaaaga gagagtactt gggaagctta atgatgaggt ctttttgac 600

ctctttatat cctttcgcct cgagaaagtc gatggggttt ttttcgaagc ttgatcgctc 660

catgattgtg atgcccagca gttccttgac gcttttgagt tttttagact tcccttctc 720

cactttggcc acaaccagta cactgtaagc gactgtagga gaatcgaatc cgccgtattt 780

cttggggtcc caatcttttt tgcgtgcgat cagcttgtcg ctgttccttt tcgggaggat 840

actttccttg gagaagcctc cggtctgtac ttcggtcttt ttaacgatgt tcacctgcgg 900

catggacagg accttccgga ctgtcgcgaa atccctaccc ttgtcccaca cgatttctcc 960

tgtttctccg tttgtttcga taagtggtcg cttccgaatc tctccattgg ccagtgtaat 1020

ctcggtcttg aaaaaattca taatattgct gtaaaagaag tacttagcgg tggccttgcc 1080

tatttcctgc tcagactttg cgatcatttt cctaacatcg tacactttat agtctccgta 1140

aacaaattca gattcaagct tgggatattt tttgataagt gcagtgccta ccactgcatt 1200

caggtaggca tcatgcgcat ggtggtaatt gttgatctct ctcaccttat aaaactgaaa 1260

gtcctttctg aaatctgaga ccagcttaga cttcagagta ataactttca cctctcgaat 1320

cagtttgtca ttttcatcgt acttggtgtt catgcgtgaa tcgagaattt gggccacgtg 1380

cttggtgatc tggcgtgtct caacaagctg ccttttgatg aagccggctt tatccaactc 1440

agacaggcca cctcgttcag ccttagtcag attatcgaac ttccgttgtg tgatcagttt 1500

ggcgttcagc agctgccgcc aataattttt cattttcttg acaacttctt ctgaggggac 1560

gttatcactc ttccctctat ttttatcgga tcttgtcaac actttattat caatagaatc 1620

atctttgaga aaagactggg gcacgatatg atccacgtcg tagtcggaga gccgattgat 1680

gtccagttcc tgatccacgt acatgtccct gccgttctgc aggtagtaca ggtagagctt 1740

ctcattctga agctgggtgt tttcaactgg gtgttcctta aggatttggg accccagttc 1800

ttttataccc tcttcaatcc tcttcatcct ttccctactg ttcttctgtc ccttctgggt 1860

agtttggttc tctcgggcca tctcgataac gatattctcg ggcttatgcc ttcccattac 1920

tttgacgagt tcatccacga ccttaacggt ctgcagtatt ccctttttga tagctgggct 1980

acctgcaaga ttagcgatgt gctcgtgaag actgtccccc tggccagaaa cttgtgcttt 2040

ctggatgtcc tccttaaagg tgagagagtc atcatggatc aactgcatga agttccggtt 2100

ggcaaatcca tcggacttaa gaaaatccag gattgtcttt ccactctgct tgtctcggat 2160

cccattgatc agttttcttg acagccgccc ccatcctgta tatcggcgcc tcttgagctg 2220

tttcatgact ttgtcgtcga agagatgagc gtaagttttc aagcgttctt caatcatctc 2280

cctatcttca aacaacgtaa gggtgaggac aatgtcctca agaatgtcct cgttctcctc 2340

attgtccagg aagtccttgt ctttaatgat tttcaggaga tcgtgatacg ttcccaggga 2400

tgcgttgaag cgatcctcca ctccgctgat ttcaacagag tcgaaacatt caatcttttt 2460

gaaatagtct tctttgagct gtttcacggt aactttccgg ttcgtcttga agaggaggtc 2520

cacgatagct ttcttctgct ctccagacag gaatgctggc tttctcatcc cttctgtgac 2580

gtatttgacc ttggtgagct cgttataaac tgtgaagtac tcgtacagca gagagtgttt 2640

aggaagcacc ttttcgttag gcagattttt atcaaagtta gtcatccttt cgatgaagga 2700

ctgggcagag gccccttat ccacgacttc ctcgaagttc cagggagtga tggtctcttc 2760

tgatttgcga gtcatccacg cgaatctgga atttccccgg gcgaggggggc ctacatagta 2820

gggtatccga aatgtgagga ttttctcaat cttttccctg ttatctttca aaaaggggta 2880

gaaatcctct tgccgcctga ggatagcgtg cagttcgccc aggtgaatct ggtgggggat 2940

gcttccattg tcgaaagtgc gctgtttgcg caacagatct tctctgttaa gctttaccag 3000

cagctcctcg gtgccgtcca ttttttccaa gatgggctta ataaatttgt aaaattcctc 3060

ctggcttgct ccgccgtcaa tgtatccggc gtagccattt ttagactgat cgaagaaaat 3120

ttccttgtac ttctcaggca gttgctgtct gacaagggcc ttcagcaaag tcaagtcttg 3180

gtggtgctca tcatagcgct tgatcatact agcgctcagc ggagctttgg tgatctccgt 3240

gttcactcgc agaatatcac tcagcagaat ggcgtctgac aggttctttg ccgccaaaaa 3300

aaggtctgcg tactggtcgc cgatctgggc cagcagattg tcgagatcat catcgtaggt 3360

gtctttgctc agttgaagct tggcatcttc ggccaggtcg aagttagatt taaagttggg 3420

ggtcagcccg agtgacaggg cgataagatt accaaacagg ccgttcttct tctccccagg 3480

gagctgtgcg atgaggtttt cgagccgccg ggatttggac agcctagcgc tcaggattgc 3540

tttggcgtca actccggatg cgttgatcgg gttctcttcg aaaagctgat tgtaagtctg 3600

aaccagttgg ataaagagtt tgtcgacatc gctgttgtct gggttcaggt ccccctcgat 3660

gaggaagtgt ccccgaaatt tgatcatatg cgccagcgcg agatagatca accgcaagtc 3720

agccttatca gtactgtcta caagcttctt cctcagatga tatatggttg ggtacttttc 3780

atggtacgcc acctcgtcca cgatattgcc aaagattggg tggcgctcgt gctttttatc 3840

ctcctccacc aaaaaggact cctccagcct atggaagaaa gagtcatcca ccttagccat 3900

ctcattacta aagatctcct gcaggtagca gatccgattc tttctgcggg tatatctgcg 3960

ccgtgctgtt cttttgagcc gcgtggcttc ggccgtctcc ccggagtcga acaggagggc 4020

gccaatgagg ttcttcttta tgctgtggcg atcggtattg cccagaactt tgaatttttt 4080

gctcggcacc ttgtactcgt ccgtaatgac ggcccagccg acgctgtttg tgccgatatc 4140

gagcccaatg gagtacttct tgtccatggt ggcggctctt gaaggacgac gtcatcatcc 4200

cttgcccgga tgcgcgggct tcttgtctag cacaggagcc tggggtagag cgcatgcaaa 4260

ttacgcgctg tgctttgtgg gaaatcaccc taaacgaaaa atttattcct ctttcgagcc 4320

ttatagtggc ggccggtcta catcctaggt tttagagcta gaaatagcaa gttaaaataa 4380

ggctagtccg ttatcaactt gaaaaagtgg caccgagtcg gtgctttttt gttaacgcgg 4440

ccgcctagag tcgacctgca ggcatgcaag cttggcgtaa tcatggtcat agctgtttcc 4500

tgtgtgaaat tgttatccgc tcacaattcc acacaacata cgagccggaa gcataaagtg 4560

taaagcctgg ggtgcctaat gagtgagcta actcacatta attgcgttgc gctcactgcc 4620

cgctttccag tcgggaaacc tgtcgtgcca gctgcattaa tgaatcggcc aacgcgcggg 4680

gagaggcggt ttgcgtattg ggcgctcttc cgcttcctcg ctcactgact cgctgcgctc 4740

ggtcgttcgg ctgcggcgag cggtatcagc tcactcaaag gcggtaatac ggttatccac 4800

agaatcaggg gataacgcag gaaagaacat gtgagcaaaa ggccagcaaa aggccaggaa 4860

ccgtaaaaag gccgcgttgc tggcgttttt ccataggctc cgcccccctg acgagcatca 4920

caaaaatcga cgctcaagtc agaggtggcg aaacccgaca ggactataaa gataccaggc 4980

gtttccccct ggaagctccc tcgtgcgctc tcctgttccg accctgccgc ttaccggata 5040

cctgtccgcc tttctccctt cgggaagcgt ggcgctttct catagctcac gctgtaggta 5100

tctcagttcg gtgtaggtcg ttcgctccaa gctgggctgt gtgcacgaac cccccgttca 5160

gcccgaccgc tgcgccttat ccggtaacta tcgtcttgag tccaacccgg taagacacga 5220

cttatcgcca ctggcagcag ccactggtaa caggattagc agagcgaggt atgtaggcgg 5280

tgctacagag ttcttgaagt ggtggcctaa ctacggctac actagaagaa cagtatttgg 5340

tatctgcgct ctgctgaagc cagttacctt cggaaaaaga gttggtagct cttgatccgg 5400

caaacaaacc accgctggta gcggtggttt ttttgtttgc aagcagcaga ttacgcgcag 5460

aaaaaaagga tctcaagaag atcctttgat cttttctacg gggtctgacg ctcagtggaa 5520

cgaaaactca cgttaaggga ttttggtcat gagattatca aaaaggatct tcacctagat 5580

ccttttaaat taaaaatgaa gttttaaatc aatctaaagt atatatgagt aaacttggtc 5640

tgacagttac caatgcttaa tcagtgaggc acctatctca gcgatctgtc tatttcgttc 5700

atccatagtt gcctgactcc ccgtcgtgta gataactacg atacgggagg gcttaccatc 5760

tggccccagt gctgcaatga taccgcgaga cccacgctca ccggctccag atttatcagc 5820

aataaaccag ccagccggaa gggccgagcg cagaagtggt cctgcaactt tatccgcctc 5880

catccagtct attaattgtt gccgggaagc tagagtaagt agttcgccag ttaatagttt 5940

gcgcaacgtt gttgccattg ctacaggcat cgtggtgtca cgctcgtcgt ttggtatggc 6000

ttcattcagc tccggttccc aacgatcaag gcgagttaca tgatccccca tgttgtgcaa 6060

aaaagcggtt agctccttcg gtcctccgat cgttgtcaga agtaagttgg ccgcagtgtt 6120

atcactcatg gttatggcag cactgcataa ttctcttact gtcatgccat ccgtaagatg 6180

cttttctgtg actggtgagt actcaaccaa gtcattctga gaatagtgta tgcggcgacc 6240

gagttgctct tgcccggcgt caatacggga taataccgcg ccacatagca gaactttaaa 6300

agtgctcatc attggaaaac gttcttcggg gcgaaaactc tcaaggatct taccgctgtt 6360

gagatccagt tcgatgtaac ccactcgtgc acccaactga tcttcagcat cttttacttt 6420

caccagcgtt tctgggtgag caaaaacagg aaggcaaaat gccgcaaaaa agggaataag 6480

ggcgacacgg aaatgttgaa tactcatact cttccttttt caatattatt gaagcattta 6540

tcagggttat tgtctcatga gcggatacat atttgaatgt atttagaaaa ataaacaaat 6600

aggggttccg cgcacatttc cccgaaaagt gccacctgac gtctaagaaa ccattattat 6660

catgacatta acctataaaa ataggcgtat cacgaggccc tttcgtctcg cgcgtttcgg 6720

tgatgacggt gaaaacctct gacacatgca gctcccggag acggtcacag cttgtctgta 6780

agcggatgcc gggagcagac aagcccgtca gggcgcgtca gcgggtgttg gcgggtgtcg 6840

gggctggctt aactatgcgg catcagagca gattgtactg agagtgcacc atagcggccg 6900

cggtacgata tggtgctgtg taggctcata tgtggatctc agaacccaca tgtactctgc 6960

tccccaggtc ttggtgcgct ttctattctc tgtcagcaaa gcctatcgaa gaatcaccta 7020

ccacaactgg cgccacggct tcaatgtagc ccagaccatg tttaccctac tcatggtacg 7080

tatgtaaatt ggatgggcta gatgaatcag agggctgggg caaggaccac agctaactat 7140

cttctggccc aagggtaccc acacaaaaaa ccaacacaca g 7181

<210> 65
<211> 7179
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полинуклеотид

<400> 65
ctgcaggcat gcaagcttgg cgtaatcatg gtcatagctg tttcctgtgt gaaattgtta 60

tccgctcaca attccacaca acatacgagc cggaagcata aagtgtaaag cctggggtgc 120

ctaatgagtg agctaactca cattaattgc gttgcgctca ctgcccgctt tccagtcggg 180

aaacctgtcg tgccagctgc attaatgaat cggccaacgc gcggggagag gcggtttgcg 240

tattgggcgc tcttccgctt cctcgctcac tgactcgctg cgctcggtcg ttcggctgcg 300

gcgagcggta tcagctcact caaaggcggt aatacggtta tccacagaat caggggataa 360

cgcaggaaag aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta aaaaggccgc 420

gttgctggcg tttttccata ggctccgccc ccctgacgag catcacaaaa atcgacgctc 480

aagtcagagg tggcgaaacc cgacaggact ataaagatac caggcgtttc cccctggaag 540

ctccctcgtg cgctctcctg ttccgaccct gccgcttacc ggatacctgt ccgcctttct 600

cccttcggga agcgtggcgc tttctcatag ctcacgctgt aggtatctca gttcggtgta 660

ggtcgttcgc tccaagctgg gctgtgtgca cgaaccccc gttcagcccg accgctgcgc 720

cttatccggt aactatcgtc ttgagtccaa cccggtaaga cacgacttat cgccactggc 780

agcagccact ggtaacagga ttagcagagc gaggtatgta ggcggtgcta cagagttctt 840

gaagtggtgg cctaactacg gctacactag aagaacagta tttggtatct gcgctctgct 900

gaagccagtt accttcggaa aaagagttgg tagctcttga tccggcaaac aaaccaccgc 960

tggtagcggt ggtttttttg tttgcaagca gcagattacg cgcagaaaaa aaggatctca 1020

agaagatcct ttgatctttt ctacggggtc tgacgctcag tggaacgaaa actcacgtta 1080

agggattttg gtcatgagat tatcaaaaag gatcttcacc tagatccttt taaattaaaa 1140

atgaagtttt aaatcaatct aaagtatata tgagtaaact tggtctgaca gttaccaatg 1200

cttaatcagt gaggcaccta tctcagcgat ctgtctattt cgttcatcca tagttgcctg 1260

actccccgtc gtgtagataa ctacgatacg ggagggctta ccatctggcc ccagtgctgc 1320

aatgataccg cgagacccac gctcaccggc tccagattta tcagcaataa accagccagc 1380

cggaagggcc gagcgcagaa gtggtcctgc aactttatcc gcctccatcc agtctattaa 1440

ttgttgccgg gaagctagag taagtagttc gccagttaat agtttgcgca acgttgttgc 1500

cattgctaca ggcatcgtgg tgtcacgctc gtcgtttggt atggcttcat tcagctccgg 1560

ttcccaacga tcaaggcgag ttacatgatc ccccatgttg tgcaaaaaag cggttagctc 1620

cttcggtcct ccgatcgttg tcagaagtaa gttggccgca gtgttatcac tcatggttat 1680

ggcagcactg cataattctc ttactgtcat gccatccgta agatgctttt ctgtgactgg 1740

tgagtactca accaagtcat tctgagaata gtgtatgcgg cgaccgagtt gctcttgccc 1800

ggcgtcaata cgggataata ccgcgccaca tagcagaact ttaaaagtgc tcatcattgg 1860

aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat ccagttcgat 1920

gtaacccact cgtgcaccca actgatcttc agcatctttt actttcacca gcgtttctgg 1980

gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga ataagggcga cacggaaatg 2040

ttgaatactc atactcttcc tttttcaata ttattgaagc atttatcagg gttattgtct 2100

catgagcgga tacatatttg aatgtattta gaaaaataaa caaatagggg ttccgcgcac 2160

atttccccga aaagtgccac ctgacgtcta agaaaccatt attatcatga cattaaccta 2220

taaaaatagg cgtatcacga ggccctttcg tctcgcgcgt ttcggtgatg acggtgaaaa 2280

cctctgacac atgcagctcc cggagacggt cacagcttgt ctgtaagcgg atgccgggag 2340

cagacaagcc cgtcagggcg cgtcagcggg tgttggcggg tgtcggggct ggcttaacta 2400

tgcggcatca gagcagattg tactgagagt gcaccatagc ggccgcggta cccacacaaa 2460

aaaccaacac acagatgtaa tgaaaataaa gatattttat ttcacacctt cctcttcttc 2520

ttggggtcag ccctgctgtc tccaccgagc tgagagaggt cgattcttgt ttcatagagc 2580

cccgtaattg actgatgaat cagtgtggcg tccaggacct cctttgtaga ggtgtaccgc 2640

tttctgtcta tggtggtgtc gaagtacttg aaggctgcag gcgcgcccaa gttggtcaga 2700

gtaaacaagt ggataatgtt ttctgcctgc tccctgatgg gcttatccct gtgcttattg 2760

taagcagaaa gcaccttatc gaggttagcg tcggcgagga tcactctttt ggagaattcg 2820

cttatttgct cgatgatctc atcaaggtag tgtttgtgtt gttccacgaa cagctgcttc 2880

tgctcattat cttcgggaga ccctttgagc ttttcatagt ggctggccag atacaagaaa 2940

ttaacgtatt tagagggcag tgccagctcg ttacctttct gcagctcgcc cgcactagcg 3000

agcattcgtt tccggccgtt ttcaagctca aagagagagt acttgggaag cttaatgatg 3060

aggtcttttt tgacctcttt atatcctttc gcctcgagaa agtcgatggg gtttttttcg 3120

aagcttgatc gctccatgat tgtgatgccc agcagttcct tgacgctttt gagtttttta 3180

gacttccctt tctccacttt ggccacaacc agtacactgt aagcgactgt aggagaatcg 3240

aatccgccgt atttcttggg gtcccaatct tttttgcgtg cgatcagctt gtcgctgttc 3300

cttttcggga ggatactttc cttggagaag cctccggtct gtacttcggt ctttttaacg 3360

atgttcacct gcggcatgga caggaccttc cggactgtcg cgaaatccct acccttgtcc 3420

cacacgattt ctcctgtttc tccgtttgtt tcgataagtg gtcgcttccg aatctctcca 3480

ttggccagtg taatctcggt cttgaaaaaa ttcataatat tgctgtaaaa gaagtactta 3540

gcggtggcct tgcctatttc ctgctcagac tttgcgatca ttttcctaac atcgtacact 3600

ttatagtctc cgtaaacaaa ttcagattca agcttgggat atttttttgat aagtgcagtg 3660

cctaccactg cattcaggta ggcatcatgc gcatggtggt aattgttgat ctctctcacc 3720

ttataaaact gaaagtcctt tctgaaatct gagaccagct tagacttcag agtaataact 3780

ttcacctctc gaatcagttt gtcattttca tcgtacttgg tgttcatgcg tgaatcgaga 3840

atttgggcca cgtgcttggt gatctggcgt gtctcaacaa gctgcctttt gatgaagccg 3900

gctttatcca actcagacag gccacctcgt tcagccttag tcagattatc gaacttccgt 3960

tgtgtgatca gtttggcgtt cagcagctgc cgccaataat ttttcatttt cttgacaact 4020

tcttctgagg ggacgttatc actcttccct ctatttttat cggatcttgt caacacttta 4080

ttatcaatag aatcatcttt gagaaaagac tggggcacga tatgatccac gtcgtagtcg 4140

gagagccgat tgatgtccag ttcctgatcc acgtacatgt ccctgccgtt ctgcaggtag 4200

tacaggtaga gcttctcatt ctgaagctgg gtgttttcaa ctgggtgttc cttaaggatt 4260

tgggacccca gttcttttat accctcttca atcctcttca tcctttccct actgttcttc 4320

tgtcccttct gggtagtttg gttctctcgg gccatctcga taacgatatt ctcgggctta 4380

tgccttccca ttactttgac gagttcatcc acgaccttaa cggtctgcag tattcccttt 4440

ttgatagctg ggctacctgc aagattagcg atgtgctcgt gaagactgtc cccctggcca 4500

gaaacttgtg ctttctggat gtcctcctta aaggtgagag agtcatcatg gatcaactgc 4560

atgaagttcc ggttggcaaa tccatcggac ttaagaaaat ccaggattgt ctttccactc 4620

tgcttgtctc ggatcccatt gatcagtttt cttgacagcc gcccccatcc tgtatatcgg 4680

cgcctcttga gctgtttcat gactttgtcg tcgaagagat gagcgtaagt tttcaagcgt 4740

tcttcaatca tctccctatc ttcaaacaac gtaagggtga ggacaatgtc ctcaagaatg 4800

tcctcgttct cctcattgtc caggaagtcc ttgtctttaa tgattttcag gagatcgtga 4860

tacgttccca gggatgcgtt gaagcgatcc tccactccgc tgatttcaac agagtcgaaa 4920

cattcaatct ttttgaaata gtcttctttg agctgtttca cggtaacttt ccggttcgtc 4980

ttgaagagga ggtccacgat agctttcttc tgctctccag acaggaatgc tggctttctc 5040

atcccttctg tgacgtattt gaccttggtg agctcgttat aaactgtgaa gtactcgtac 5100

agcagagagt gtttaggaag caccttttcg ttaggcagat ttttatcaaa gttagtcatc 5160

ctttcgatga aggactgggc agaggccccc ttatccacga cttcctcgaa gttccaggga 5220

gtgatggtct cttctgattt gcgagtcatc cacgcgaatc tggaatttcc ccgggcgagg 5280

gggcctacat agtagggtat ccgaaatgtg aggattttct caatcttttc cctgttatct 5340

ttcaaaaagg ggtagaaatc ctcttgccgc ctgaggatag cgtgcagttc gcccaggtga 5400

atctggtggg ggatgcttcc attgtcgaaa gtgcgctgtt tgcgcaacag atcttctctg 5460

ttaagcttta ccagcagctc ctcggtgccg tccatttttt ccaagatggg cttaataaat 5520

ttgtaaaatt cctcctggct tgctccgccg tcaatgtatc cggcgtagcc atttttagac 5580

tgatcgaaga aaatttcctt gtacttctca ggcagttgct gtctgacaag ggccttcagc 5640

aaagtcaagt cttggtggtg ctcatcatag cgcttgatca tactagcgct cagcggagct 5700

ttggtgatct ccgtgttcac tcgcagaata tcactcagca gaatggcgtc tgacaggttc 5760

tttgccgcca aaaaaaggtc tgcgtactgg tcgccgatct gggccagcag attgtcgaga 5820

tcatcatcgt aggtgtcttt gctcagttga agcttggcat cttcggccag gtcgaagtta 5880

gatttaaagt tgggggtcag cccgagtgac agggcgataa gattaccaaa caggccgttc 5940

ttcttctccc cagggagctg tgcgatgagg ttttcgagcc gccgggattt ggacagccta 6000

gcgctcagga ttgctttggc gtcaactccg gatgcgttga tcgggttctc ttcgaaaagc 6060

tgattgtaag tctgaaccag ttggataaag agtttgtcga catcgctgtt gtctgggttc 6120

aggtcccccc cgatgaggaa gtgtccccga aatttgatca tatgcgccag cgcgagatag 6180

atcaaccgca agtcagcctt atcagtactg tctacaagct tcttcctcag atgatatatg 6240

gttgggtact tttcatggta cgccacctcg tccacgatat tgccaaagat tgggtggcgc 6300

tcgtgctttt tatcctcctc caccaaaaag gactcctcca gcctatggaa gaaagagtca 6360

tccaccttag ccatctcatt actaaagatc tcctgcaggt agcagatccg attctttctg 6420

cgggtatatc tgcgccgtgc tgttcttttg agccgcgtgg cttcggccgt ctccccggag 6480

tcgaacagga gggcgccaat gaggttcttc tttatgctgt ggcgatcggt attgcccaga 6540

actttgaatt ttttgctcgg caccttgtac tcgtccgtaa tgacggccca gccgacgctg 6600

tttgtgccga tatcgagccc aatggagtac ttcttgtcca tggtggcggc tcttgaagga 6660

cgacgtcatc atcccttgcc cggatgcgcg ggcttcttgt ctagcacagg agcctggggt 6720

agagcgcatg caaattacgc gctgtgcttt gtgggaaatc accctaaacg aaaaatttat 6780

tcctctttcg agccttatag tggcggccgg tctacatcct aggttttaga gctagaaata 6840

gcaagttaaa ataaggctag tccgttatca acttgaaaaa gtggcaccga gtcggtgctt 6900

ttttgttgtt ctgtctatag gtaagctgac aaataacaaa taggcacata gaaaatctag 6960

taagtagtac cacctgatat ctcactttgc tgcaggcagg attcccctct ggctcactgg 7020

cagtctcctc cggtgtgggc cagggctctt tgaagttgga tctgagcctt tctatcacct 7080

gtttgatgga caagcccttt tgcacaagtt tgacttcaag gagggccatg tcacatacca 7140

cagaaggtaa gtccatgaac gcggccgcct agagtcgac 7179

<210> 66
<211> 7179
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полинуклеотид

<400> 66
atgtaatgaa aataaagata ttttatttca caccttcctc ttcttcttgg ggtcagccct 60

gctgtctcca ccgagctgag agaggtcgat tcttgtttca tagagccccg taattgactg 120

atgaatcagt gtggcgtcca ggacctcctt tgtagaggtg taccgctttc tgtctatggt 180

ggtgtcgaag tacttgaagg ctgcaggcgc gcccaagttg gtcagagtaa acaagtggat 240

aatgttttct gcctgctccc tgatgggctt atccctgtgc ttattgtaag cagaaagcac 300

cttatcgagg ttagcgtcgg cgaggatcac tcttttggag aattcgctta tttgctcgat 360

gatctcatca aggtagtgtt tgtgttgttc cacgaacagc tgcttctgct cattatcttc 420

gggagaccct ttgagctttt catagtggct ggccagatac aagaaattaa cgtatttaga 480

gggcagtgcc agctcgttac ctttctgcag ctcgcccgca ctagcgagca ttcgtttccg 540

gccgttttca agctcaaaga gagagtactt gggaagctta atgatgaggt ctttttgac 600

ctctttatat cctttcgcct cgagaaagtc gatggggttt ttttcgaagc ttgatcgctc 660

catgattgtg atgcccagca gttccttgac gcttttgagt tttttagact tccctttctc 720

cactttggcc acaaccagta cactgtaagc gactgtagga gaatcgaatc cgccgtattt 780

cttggggtcc caatcttttt tgcgtgcgat cagcttgtcg ctgttccttt tcgggaggat 840

actttccttg gagaagcctc cggtctgtac ttcggtcttt ttaacgatgt tcacctgcgg 900

catggacagg accttccgga ctgtcgcgaa atccctaccc ttgtcccaca cgatttctcc 960

tgtttctccg tttgtttcga taagtggtcg cttccgaatc tctccattgg ccagtgtaat 1020

ctcggtcttg aaaaaattca taatattgct gtaaaagaag tacttagcgg tggccttgcc 1080

tatttcctgc tcagactttg cgatcatttt cctaacatcg tacactttat agtctccgta 1140

aacaaattca gattcaagct tgggatattt tttgataagt gcagtgccta ccactgcatt 1200

caggtaggca tcatgcgcat ggtggtaatt gttgatctct ctcaccttat aaaactgaaa 1260

gtcctttctg aaatctgaga ccagcttaga cttcagagta ataactttca cctctcgaat 1320

cagtttgtca ttttcatcgt acttggtgtt catgcgtgaa tcgagaattt gggccacgtg 1380

cttggtgatc tggcgtgtct caacaagctg ccttttgatg aagccggctt tatccaactc 1440

agacaggcca cctcgttcag ccttagtcag attatcgaac ttccgttgtg tgatcagttt 1500

ggcgttcagc agctgccgcc aataattttt cattttcttg acaacttctt ctgaggggac 1560

gttatcactc ttccctctat ttttatcgga tcttgtcaac actttattat caatagaatc 1620

atctttgaga aaagactggg gcacgatatg atccacgtcg tagtcggaga gccgattgat 1680

gtccagttcc tgatccacgt acatgtccct gccgttctgc aggtagtaca ggtagagctt 1740

ctcattctga agctgggtgt tttcaactgg gtgttcctta aggatttggg accccagttc 1800

ttttataccc tcttcaatcc tcttcatcct ttccctactg ttcttctgtc ccttctgggt 1860

agtttggttc tctcgggcca tctcgataac gatattctcg ggcttatgcc ttcccattac 1920

tttgacgagt tcatccacga ccttaacggt ctgcagtatt ccctttttga tagctgggct 1980

acctgcaaga ttagcgatgt gctcgtgaag actgtccccc tggccagaaa cttgtgcttt 2040

ctggatgtcc tccttaaagg tgagagagtc atcatggatc aactgcatga agttccggtt 2100

ggcaaatcca tcggacttaa gaaaatccag gattgtcttt ccactctgct tgtctcggat 2160

cccattgatc agttttcttg acagccgccc ccatcctgta tatcggcgcc tcttgagctg 2220

tttcatgact ttgtcgtcga agagatgagc gtaagttttc aagcgttctt caatcatctc 2280

cctatcttca aacaacgtaa gggtgaggac aatgtcctca agaatgtcct cgttctcctc 2340

attgtccagg aagtccttgt ctttaatgat tttcaggaga tcgtgatacg ttcccaggga 2400

tgcgttgaag cgatcctcca ctccgctgat ttcaacagag tcgaaacatt caatcttttt 2460

gaaatagtct tctttgagct gtttcacggt aactttccgg ttcgtcttga agaggaggtc 2520

cacgatagct ttcttctgct ctccagacag gaatgctggc tttctcatcc cttctgtgac 2580

gtatttgacc ttggtgagct cgttataaac tgtgaagtac tcgtacagca gagagtgttt 2640

aggaagcacc ttttcgttag gcagattttt atcaaagtta gtcatccttt cgatgaagga 2700

ctgggcagag gccccttat ccacgacttc ctcgaagttc cagggagtga tggtctcttc 2760

tgatttgcga gtcatccacg cgaatctgga atttccccgg gcgaggggggc ctacatagta 2820

gggtatccga aatgtgagga ttttctcaat cttttccctg ttatctttca aaaaggggta 2880

gaaatcctct tgccgcctga ggatagcgtg cagttcgccc aggtgaatct ggtgggggat 2940

gcttccattg tcgaaagtgc gctgtttgcg caacagatct tctctgttaa gctttaccag 3000

cagctcctcg gtgccgtcca ttttttccaa gatgggctta ataaatttgt aaaattcctc 3060

ctggcttgct ccgccgtcaa tgtatccggc gtagccattt ttagactgat cgaagaaaat 3120

ttccttgtac ttctcaggca gttgctgtct gacaagggcc ttcagcaaag tcaagtcttg 3180

gtggtgctca tcatagcgct tgatcatact agcgctcagc ggagctttgg tgatctccgt 3240

gttcactcgc agaatatcac tcagcagaat ggcgtctgac aggttctttg ccgccaaaaa 3300

aaggtctgcg tactggtcgc cgatctgggc cagcagattg tcgagatcat catcgtaggt 3360

gtctttgctc agttgaagct tggcatcttc ggccaggtcg aagttagatt taaagttggg 3420

ggtcagcccg agtgacaggg cgataagatt accaaacagg ccgttcttct tctccccagg 3480

gagctgtgcg atgaggtttt cgagccgccg ggatttggac agcctagcgc tcaggattgc 3540

tttggcgtca actccggatg cgttgatcgg gttctcttcg aaaagctgat tgtaagtctg 3600

aaccagttgg ataaagagtt tgtcgacatc gctgttgtct gggttcaggt ccccctcgat 3660

gaggaagtgt ccccgaaatt tgatcatatg cgccagcgcg agatagatca accgcaagtc 3720

agccttatca gtactgtcta caagcttctt cctcagatga tatatggttg ggtacttttc 3780

atggtacgcc acctcgtcca cgatattgcc aaagattggg tggcgctcgt gctttttatc 3840

ctcctccacc aaaaaggact cctccagcct atggaagaaa gagtcatcca ccttagccat 3900

ctcattacta aagatctcct gcaggtagca gatccgattc tttctgcggg tatatctgcg 3960

ccgtgctgtt cttttgagcc gcgtggcttc ggccgtctcc ccggagtcga acaggagggc 4020

gccaatgagg ttcttcttta tgctgtggcg atcggtattg cccagaactt tgaatttttt 4080

gctcggcacc ttgtactcgt ccgtaatgac ggcccagccg acgctgtttg tgccgatatc 4140

gagcccaatg gagtacttct tgtccatggt ggcggctctt gaaggacgac gtcatcatcc 4200

cttgcccgga tgcgcgggct tcttgtctag cacaggagcc tggggtagag cgcatgcaaa 4260

ttacgcgctg tgctttgtgg gaaatcaccc taaacgaaaa atttattcct ctttcgagcc 4320

ttatagtggc ggccggtcta catcctaggt tttagagcta gaaatagcaa gttaaaataa 4380

ggctagtccg ttatcaactt gaaaaagtgg caccgagtcg gtgctttttt gttaacgcgg 4440

ccgcctagag tcgacctgca ggcatgcaag cttggcgtaa tcatggtcat agctgtttcc 4500

tgtgtgaaat tgttatccgc tcacaattcc acacaacata cgagccggaa gcataaagtg 4560

taaagcctgg ggtgcctaat gagtgagcta actcacatta attgcgttgc gctcactgcc 4620

cgctttccag tcgggaaacc tgtcgtgcca gctgcattaa tgaatcggcc aacgcgcggg 4680

gagaggcggt ttgcgtattg ggcgctcttc cgcttcctcg ctcactgact cgctgcgctc 4740

ggtcgttcgg ctgcggcgag cggtatcagc tcactcaaag gcggtaatac ggttatccac 4800

agaatcaggg gataacgcag gaaagaacat gtgagcaaaa ggccagcaaa aggccaggaa 4860

ccgtaaaaag gccgcgttgc tggcgttttt ccataggctc cgccccccctg acgagcatca 4920

caaaaatcga cgctcaagtc agaggtggcg aaacccgaca ggactataaa gataccaggc 4980

gtttcccccct ggaagctccc tcgtgcgctc tcctgttccg accctgccgc ttaccggata 5040

cctgtccgcc tttctccctt cgggaagcgt ggcgctttct catagctcac gctgtaggta 5100

tctcagttcg gtgtaggtcg ttcgctccaa gctgggctgt gtgcacgaac cccccgttca 5160

gcccgaccgc tgcgccttat ccggtaacta tcgtcttgag tccaacccgg taagacacga 5220

cttatcgcca ctggcagcag ccactggtaa caggattagc agagcgaggt atgtaggcgg 5280

tgctacagag ttcttgaagt ggtggcctaa ctacggctac actagaagaa cagtatttgg 5340

tatctgcgct ctgctgaagc cagttacctt cggaaaaaga gttggtagct cttgatccgg 5400

caaacaaacc accgctggta gcggtggttt ttttgtttgc aagcagcaga ttacgcgcag 5460

aaaaaaagga tctcaagaag atcctttgat cttttctacg gggtctgacg ctcagtggaa 5520

cgaaaactca cgttaaggga ttttggtcat gagattatca aaaaggatct tcacctagat 5580

ccttttaaat taaaaatgaa gttttaaatc aatctaaagt atatatgagt aaacttggtc 5640

tgacagttac caatgcttaa tcagtgaggc acctatctca gcgatctgtc tatttcgttc 5700

atccatagtt gcctgactcc ccgtcgtgta gataactacg atacgggagg gcttaccatc 5760

tggccccagt gctgcaatga taccgcgaga cccacgctca ccggctccag atttatcagc 5820

aataaaccag ccagccggaa gggccgagcg cagaagtggt cctgcaactt tatccgcctc 5880

catccagtct attaattgtt gccgggaagc tagagtaagt agttcgccag ttaatagttt 5940

gcgcaacgtt gttgccattg ctacaggcat cgtggtgtca cgctcgtcgt ttggtatggc 6000

ttcattcagc tccggttccc aacgatcaag gcgagttaca tgatccccca tgttgtgcaa 6060

aaaagcggtt agctccttcg gtcctccgat cgttgtcaga agtaagttgg ccgcagtgtt 6120

atcactcatg gttatggcag cactgcataa ttctcttact gtcatgccat ccgtaagatg 6180

cttttctgtg actggtgagt actcaaccaa gtcattctga gaatagtgta tgcggcgacc 6240

gagttgctct tgcccggcgt caatacggga taataccgcg ccacatagca gaactttaaa 6300

agtgctcatc attggaaaac gttcttcggg gcgaaaactc tcaaggatct taccgctgtt 6360

gagatccagt tcgatgtaac ccactcgtgc acccaactga tcttcagcat cttttacttt 6420

caccagcgtt tctgggtgag caaaaacagg aaggcaaaat gccgcaaaaa agggaataag 6480

ggcgacacgg aaatgttgaa tactcatact cttccttttt caatattatt gaagcattta 6540

tcagggttat tgtctcatga gcggatacat atttgaatgt atttagaaaa ataaacaaat 6600

aggggttccg cgcacatttc cccgaaaagt gccacctgac gtctaagaaa ccattattat 6660

catgacatta acctataaaa ataggcgtat cacgaggccc tttcgtctcg cgcgtttcgg 6720

tgatgacggt gaaaacctct gacacatgca gctcccggag acggtcacag cttgtctgta 6780

agcggatgcc gggagcagac aagcccgtca gggcgcgtca gcgggtgttg gcgggtgtcg 6840

gggctggctt aactatgcgg catcagagca gattgtactg agagtgcacc atagcggccg 6900

cggttctgtc tataggtaag ctgacaaata acaaataggc acatagaaaa tctagtaagt 6960

agtaccacct gatatctcac tttgctgcag gcaggattcc cctctggctc actggcagtc 7020

tcctccggtg tgggccaggg ctctttgaag ttggatctga gcctttctat cacctgtttg 7080

atggacaagc ccttttgcac aagtttgact tcaaggaggg ccatgtcaca taccacagaa 7140

ggtaagtcca tggtacccac acaaaaaacc aacacacag 7179

<210> 67
<211> 6079
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полинуклеотид

<400> 67
gcttggcgta atcatggtca tagctgtttc ctgtgtgaaa ttgttatccg ctcacaattc 60

cacacaacat acgagccgga agcataaagt gtaaagcctg gggtgcctaa tgagtgagct 120

aactcacatt aattgcgttg cgctcactgc ccgctttcca gtcgggaaac ctgtcgtgcc 180

agctgcatta atgaatcggc caacgcgcgg ggagaggcgg tttgcgtatt gggcgctctt 240

ccgcttcctc gctcactgac tcgctgcgct cggtcgttcg gctgcggcga gcggtatcag 300

ctcactcaaa ggcggtaata cggttatcca cagaatcagg ggataacgca ggaaagaaca 360

tgtgagcaaa aggccagcaa aaggccagga accgtaaaaa ggccgcgttg ctggcgtttt 420

tccataggct ccgcccccct gacgagcatc acaaaaatcg acgctcaagt cagaggtggc 480

gaaacccgac aggactataa agataccagg cgtttccccc tggaagctcc ctcgtgcgct 540

ctcctgttcc gaccctgccg cttaccggat acctgtccgc ctttctccct tcgggaagcg 600

tggcgctttc tcatagctca cgctgtaggt atctcagttc ggtgtaggtc gttcgctcca 660

agctgggctg tgtgcacgaa ccccccgttc agcccgaccg ctgcgcctta tccggtaact 720

atcgtcttga gtccaacccg gtaagacacg acttatcgcc actggcagca gccactggta 780

acaggattag cagagcgagg tatgtaggcg gtgctacaga gttcttgaag tggtggccta 840

actacggcta cactagaaga acagtatttg gtatctgcgc tctgctgaag ccagttacct 900

tcggaaaaag agttggtagc tcttgatccg gcaaacaaac caccgctggt agcggtggtt 960

tttttgtttg caagcagcag attacgcgca gaaaaaaagg atctcaagaa gatcctttga 1020

tcttttctac ggggtctgac gctcagtgga acgaaaactc acgttaaggg attttggtca 1080

tgagattatc aaaaaggatc ttcacctaga tccttttaaa ttaaaaatga agttttaaat 1140

caatctaaag tatatatgag taaacttggt ctgacagtta ccaatgctta atcagtgagg 1200

cacctatctc agcgatctgt ctatttcgtt catccatagt tgcctgactc cccgtcgtgt 1260

agataactac gatacgggag ggcttaccat ctggccccag tgctgcaatg ataccgcgag 1320

acccacgctc accggctcca gatttatcag caataaacca gccagccgga agggccgagc 1380

gcagaagtgg tcctgcaact ttatccgcct ccatccagtc tattaattgt tgccgggaag 1440

ctagagtaag tagttcgcca gttaatagtt tgcgcaacgt tgttgccatt gctacaggca 1500

tcgtggtgtc acgctcgtcg tttggtatgg cttcattcag ctccggttcc caacgatcaa 1560

ggcgagttac atgatccccc atgttgtgca aaaaagcggt tagctccttc ggtcctccga 1620

tcgttgtcag aagtaagttg gccgcagtgt tatcactcat ggttatggca gcactgcata 1680

attctcttac tgtcatgcca tccgtaagat gcttttctgt gactggtgag tactcaacca 1740

agtcattctg agaatagtgt atgcggcgac cgagttgctc ttgcccggcg tcaatacggg 1800

ataataccgc gccacatagc agaactttaa aagtgctcat cattggaaaa cgttcttcgg 1860

ggcgaaaact ctcaaggatc ttaccgctgt tgagatccag ttcgatgtaa cccactcgtg 1920

cacccaactg atcttcagca tcttttactt tcaccagcgt ttctgggtga gcaaaaacag 1980

gaaggcaaaa tgccgcaaaa aagggaataa gggcgacacg gaaatgttga atactcatac 2040

tcttcctttt tcaatattat tgaagcattt atcagggtta ttgtctcatg agcggataca 2100

tatttgaatg tatttagaaa aataaacaaa taggggttcc gcgcacattt ccccgaaaag 2160

tgccacctga cgtctaagaa accattatta tcatgacatt aacctataaa aataggcgta 2220

tcacgaggcc ctttcgtctc gcgcgtttcg gtgatgacgg tgaaaacctc tgacacatgc 2280

agctcccgga gacggtcaca gcttgtctgt aagcggatgc cgggagcaga caagcccgtc 2340

agggcgcgtc agcgggtgtt ggcgggtgtc ggggctggct taactatgcg gcatcagagc 2400

agattgtact gagagtgcac catagcggcc gcggtaccca cacaaaaaac caacacacag 2460

atgtaatgaa aataaagata ttttattagg ctgatcagcg agctctagga attcttagga 2520

tccctttttc ttttttgcct ggccggcctt tttcgtggcc gccggccttt tgccctttt 2580

gatgatctga gggtgcttct tagatttcac ttcatacagg ttgcccagaa tgtctgtgct 2640

gtacttctta atgctctggg tcttggaggc gattgtctta atgatcctgg ggggcctctt 2700

gtcgttcatg ttttccaggt actcgcggta ggtgatgtcg atcatgttca cttcgatccg 2760

gttcagcagg tcgttgttca cgccgatcac tctatacagc tcgccgttga tcttgatcag 2820

atcgttgttg tagaaggagg cgataaactc ggcctggttg ctgatcttct tcagcttctt 2880

agcttcctca tagcacttgc tattcacttc gtagtagttt tctttttga tcacatccag 2940

attcttcacg gtcacgaact tgtacacgcc attgtccagg tacacgtcga atctgtaggg 3000

cttcagggac agcttcacga ccttgtttct gctgttgggg tagtcgtcgg tgatgtccag 3060

atgggcgttc agtttgttgc cgtaatactt aatcttcttg atcacggggc cgttgtcctt 3120

tttggagtac ttggtcaggt agttcccggt ttcctcgtag tacttgtaca ggggattctt 3180

ctcgtcgccg tactgttcca taatcagctt cagtttctgg taggtctggg ggtcgtggtg 3240

gtacatcagc agcttttcgg ggctcttgtt gatcagcttt ttcagcttgt cattgtcctt 3300

gtcgtacagg ccgttcagat tgttcacgat cagggtgttg cccttgtcgt ccttccgggt 3360

ggagtacagg gtgtcgttaa tcagctctct attaggcttc ttgtccaccc ggtggctgta 3420

cttgtagtcc ttgaagtcct taatgtgctt gatctggtgg ggggtgatga agatctcttt 3480

gtactcctgc tcggtttcga tctcgggcat gctctcggcc tgcttttcct cgaacatctg 3540

gttttccatc actttttgg ccttgtccag tttcttccac tctttgaaga tgaaatcggc 3600

gttggcaatg atcagggcgt cctcggcgtg gtgcttgtac cccttgttcc gctctttctt 3660

aaacttccac ttccgccgca gaaagctggt gaagccgcca ttgatggact tcactttcac 3720

gtccaggttg ttcactctga agtagctccg cagcaggttc atcaggcctc tggtggcgta 3780

tctggtatcc accaggttcc ggttgatgaa gtctttctgc acggagaacc tgttgatgtc 3840

ccgttcttcc agcagatact ctttcttggt cttgctgatt ctgcccttgc ccttggccag 3900

attcaggatg tgcttcttga aggtttcgta gctgatcttg ctgtcgctgc tgctcaggta 3960

ctggaatggg gtccggttgc ccttcttgct gttttcttcc tgcttcacga gcaccttgtt 4020

gttgaagctg ttgtcgaagg acacgcttct ggggatgatg tggtccacct catagttgaa 4080

ggggttgttc agcagatctt ccagagggat ggcttccagg ctgtacaggc acttgccttc 4140

ctgcatgtcg tgcagcttga tcttctcgat caggtacttg gcgttctctt tgccggtggt 4200

ccggatgatt tcctcgatcc gctcgttggt ctgccggttc cgcttctgca tctcgttgat 4260

cattttctgg gcgtccttgg agttcttctc gcgggccagc tcgataatga tgtcgttggg 4320

caggccgtac ttcttgatga tggcgttgat cactttgatg ctctggatga agcttctctt 4380

cacgacgggg ctcaggatga agtcgtccac cagggtggtg gggatctctt tctgctggga 4440

caggtccacc ttcttgggca ccagcttcag ccggttgaag atagcgatct ggttgtcgtt 4500

ggtgtgccac agctcgtcca ggatcaggtt gatggccttc aggctcaggt tgtgggtgcc 4560

ggtatagccc ttcagattag agatctgctc gatctcttcc tgggtcagct cggagttcag 4620

attggtcagt tcttcctgga tgtcctcgct gctctggtag atggtcagga tcttggcaat 4680

ctgatccagc agctcggcgt tctcaataat ctctttccgg gcggtaatgt ccttgatgtc 4740

gtggtacacc ttcaggttgg tgaactcggg cttgccggtg ctggtcactc tgtagccctt 4800

aatatcctct tcgttcacga ggatttcttt ggcgatctgc ttcagggtgg gcttcttctt 4860

ctgcttgaac acgttctcga tgatctggaa cttctcgtaa tattccagct tctcgttctc 4920

gtccctggtg atcacgagat tgttcaggtc gttcagggcg ttgtacaggt cggcgttgta 4980

ggcgtacttc acgctccgca gttcctcggg gaagtaggtg cagtggccca tcagcatctc 5040

gtaccattct ttgatgtcct tccagccgaa ggggctgccc tcgccaggtc cctcatagta 5100

ggtccgccgg gtttccagca ggtcgatgta ggtgtcgatg aagctctggt ccagctggtg 5160

gtaggccttc tgcaccttca gcagctgttt ggcttctttc acgtagtcgc tggtcttgaa 5220

tctgttgatg ctgccccgca cttcgccgtc tttcttcagc cgttccagct gcagttcggc 5280

cacgtatttc tcttccaggg ccttgctgtt ccggctgatc tgctctttgg tggacagctc 5340

gttgccggtg tcctcttcca cctcgttcac gttgtgcacg cctcttctct tggccaggtg 5400

cagcagggcg gcagagaact cttcctcgct cagcttctgg ctcaggccct tcactctggc 5460

ctcgtagggg ttgatgccgc tcagctcgct gtggtcggtc agcaggttgt agtcgaacag 5520

cagcttcttc actctctgga ttctatgccg cctccgccgc ttcagccttc tggcgcctct 5580

cttgctccgc ctgccctcgt tgttttccac gttggcctct ttgaacagcc gcacgccggc 5640

atcgatcacg tcccgtgtct cgtagtcgat gatgccgtag cccacgctgg tgatgccgat 5700

gtccaggccc aggatgtagt tccgcttggc tgctgggact ccgtggatac cgaccttccg 5760

cttcttcttt ggggccatgt ggcggctctt gaaggacgac gtcatcatcc cttgcccgga 5820

tgcgcgggct tcttgtctag cacaggagcc tggggtagag cgcatgcaaa ttacgcgctg 5880

tgctttgtgg gaaatcaccc taaacgaaaa atttattcct ctttcgagcc ttatagtggc 5940

ggccggtcta catcctaggt tttagtactc tggaaacaga atctactaaa acaaggcaaa 6000

atgccgtgtt tatctcgtca acttgttggc gagatttttt gttaacgcgg ccgcctagag 6060

tcgacctgca ggcatgcaa 6079

<210> 68
<211> 23
<212> ДНК
<213> Homo sapiens

<400> 68
agtactgtgg gtactcgaag ggg 23

<210> 69
<211> 53
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический олигонуклеотид

<220>
<221> модифицированное_основание
<222> (24)..(26)
<223> a, c, t, g, неизвестная или другая

<400> 69
gagagtgcac catagcggcc gcgnnngtac ccacacaaaa aaccaacaca cag 53

<210> 70
<211> 53
<212> ДНК

<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический
олигонуклеотид

<220>
<221> модифицированное_основание
<222> (26)..(31)
<223> a, c, t, g, неизвестная или другая

<400> 70
ggcaccgagt cggtgctttt ttgttnnnnn naacgcggcc gcctagagtc gac 53

<210> 71
<211> 230
<212> ДНК
<213> Homo sapiens

<400> 71
ggaattcgaa cgctgacgtc atcaacccgc tccaaggaat cgcgggccca gtgtcactag 60

gcgggaacac ccagcgcgcg tgcgccctgg caggaagatg gctgtgaggg acaggggagt 120

ggcgccctgc aatatttgca tgtcgctatg tgttctggga aatcaccata aacgtgaaat 180

gtctttggat ttgggaatct tataagttct gtatgagacc actttttccc 230

<210> 72
<211> 233
<212> ДНК
<213> Homo sapiens

<400> 72
cttgctctca gcagtgcaac gaggtaaaag gaagaagctg gcccacgcat gcgctcttca 60

aatttttgag acagtttacc cagaatgcag tgctcaaagg aaacgcgtgc gcagtgtggt 120

caggttgttt cgctgggtga gtaaaatgaa atcttagagg cgttgtgggc tggcccagtt 180

gatgacgtca ccataccaca gcttctagtg ctattctgcg ccggtatccg acc 233

<210> 73
<211> 239
<212> ДНК
<213> Homo sapiens

<400> 73
agtatttagc atgccccacc catctgcaag gcattctgga tagtgtcaaa acagccggaa 60

atcaagtccg tttatctcaa actttagcat tttgggaata aatgatattt gctatgctgg 120

ttaaattaga ttttagttaa atttcctgct gaagctctag tacgataagc aacttgacct 180

aagtgtaaag ttgagacttc cttcaggttt atatagcttg tgcgccgctt gggtacctc 239

<210> 74
<211> 308
<212> ДНК

<213> Homo sapiens

<400> 74
gtctctctct tagcgggata tcttccgcaa gcactgggaa tgtggacatg gaaagtaaat 60

tgagtctccg tgggggagtg agacagggag tgaggggtgt tggacgcggc acgggaacct 120

ggccagagtc agcggaccca attggctgct ctctctcaga tgcagttccc cttcctccct 180

ccaggggggcg ccacggaacg cagggccctc actggccctg gggactgggt gacgtcaggg 240

atgagcctct tgtgattggc tccatcaccc tgcgtaagat caaagggaag aaaggatggg 300

cccgacaa 308

<210> 75
<211> 308
<212> ДНК
<213> Homo sapiens

<400> 75
gtctctctct tagcgggata tcttccgcaa gcactgggga tgtggacatg gaaagtaaat 60

tgagtctccg tgggggagtg agacagggag tgaggggtgt tggacgcggc acgggaaccc 120

ggccggagtc agcggaccca attggctgct ctctctcaga tacagttccc cttcctccct 180

ccaggggggcg ccacggaacg cagggccctc actggccctg gggactgggt gacgtcaggg 240

gtgagcctct cctgattggc tccatcaccc tgcgtaaggt caaaaggaag aaaggagatc 300

cccgacac 308

<210> 76
<211> 361
<212> ДНК
<213> Homo sapiens

<400> 76
tgccggccca cgggtggagg gatcgggcgg gcggtgccga agcggtccgg cattggccgg 60

ccgccccaac gcgcacgcgc acgcgagcag gccggccggc tccggggagg ccacgcccac 120

tccccgtagg gcggggccag accatatttg cataagatag tgtcattcta gctttcctgt 180

atttgttcat ttcgtgtcta ttagctattc tgctagccac aatgcctctg aaagcctata 240

gtcttagaaa gttatgcccg aaaacggttt ttttaatctc acgccaccaa ctttctcacc 300

ctaatcataa aacacaattt ctttagggct ataaaatact actctgtgaa gctgaggacg 360

t 361

<210> 77
<211> 376
<212> ДНК
<213> Homo sapiens

<400> 77
tctttccgct ccaggaccgc cctgggcctg caggatcctg ggcgggagcc caggtgtccg 60

ggatctgggc cactagggac tggggaggaa cctctcagag aagcccatag cccgcagcgg 120

ccccgcgcgg ccggttccgg cgccgcactg ttccagcctc tactatggta cagtccctgc 180

gtcgcagcct cggcgggggc tctaagaacg ggaggcagaa aaagctcaat cagcagcagg 240

cgagcttcac ccgctgcttc caaatctgtg ccaaaatatt ctatgctgca cagataaaat 300

cctctgtcgg ttctacaagc ctggcttttc ctatagagaa ccctcttata agcaaaaagt 360

aaagctctcg tgaaga 376

<210> 78
<211> 412
<212> ДНК
<213> Homo sapiens

<400> 78
gagggcagtc accagctcct ggcccgtgcg ccaagctcag cgggcgtccg cggtgcgatc 60

ttccctagcg cctcgggtct ggcgccgcca tcttcctcgg taacaaccag tcgcctgagg 120

cgtggggccg cctcccaaag acttctggga gggcggtgcg gctcaggctc tgccccgcct 180

ccggggctat ttgcatacga ccatttccag taattcccag cagccaccgt agctatattt 240

ggtagaacaa cgagcacttt ctcaactcca gtcaataact acgttagttg cattacacat 300

tgggctaata taaatagagg ttaaatctct aggtcattta agagaagtcg gcctatgtgt 360

acagacattt gttccagggg ctttaaatag ctggtggtgg aactcaatat tc 412

<210> 79
<211> 484
<212> ДНК
<213> Homo sapiens

<400> 79
taaccgctca gctgacctca ggagggcagg ggtgccttct aaagggtcca gagagcctcc 60

attccagctg caggcgtggg acacagaccg ggacgtgggg cggcggccgg actgggcagg 120

tcgtcccggg tccagcggcg cctcacggtc gcggctccat gcccgggact gcgaccccgg 180

aagtggcggg agcgggggac gacagccgcg gcggacacag gggacccgcc ggctcaggca 240

cctttgaccc ggaagttgag cgacccaggc ggcggcctgg gattggacac caccaggcac 300

gtaccaaggc gtccgcggcg cttgggggggg agcccgcggc gcggcggcct aaggtgcgta 360

acgccccatg aacgacatct tccggtgggt tagggagaga cacccccctg tgacttggta 420

tcactcagtc aaacccatga tcccccacta ttaaggatat ccggagagga tgctacctat 480

cagg 484

<210> 80
<211> 861
<212> ДНК
<213> Homo sapiens

<400> 80
tcctgactgc agcaccagaa ggctggtctc tcccacagaa cgaggatgga ggcggggagg 60

gatccgttga agagggaagg agcgatcacc caaagagaac taaaatcaaa taaaataaaa 120

cagagagatg tcttggagga gggggcgagt ctgaccggga taagaataaa gagaaagggt 180

gaacccggga ggcggagttt gcagtgagcc gagatcgcgc cactgcactc cagcctgggc 240

gacagagtga gactccgtct cagtaaaaaa aaaaaaaaaa aaaagaataa agaggaaagg 300

acgcaagaaa gggaaagggg actctcaggg agtaaaagag tcttacactt ttaacagtga 360

cgttaaaaga ctactgttgc ctttctgaag actaaaaaga aaaaaaactt aaaaatttaa 420

agaaataaac ttctgagcca tgtcaccaac ttaaccaccc ccaggtacct gcaacggctc 480

gcgcccgccg gtgtctaaca ggatccggac ctagctcata ttgctgccgc aaaacgcaag 540

gctagcttcc gccagtactg ccgcaacacc ttcttatttc acgacgtatg gtcgtaaagc 600

aataaagatc caggctcggg aaaatgacgg agaggtggaa ctatagagaa taaatttgca 660

tatataataa tccgctcgct aattgtgttt ctgttttcct ttgctaaggt agaaacaaaa 720

gaataatcac agaatctcag tgggactttg aaaatatcca ggattttata cgtgaagaat 780

ggatgtatcg cattacggta gtcaccctat gtgtaaatta gtggcacata cttggcactc 840

cttaatgtca actataagat g 861

<210> 81
<211> 172
<212> ДНК
<213> Mus musculus

<400> 81
cgctcttgaa ggacgacgtc atcatccctt gcccggatgc gcgggcttct tgtctagcac 60

aggagcctgg ggtagagcgc atgcaaatta cgcgctgtgc tttgtgggaa atcaccctaa 120

acgaaaaatt tattcctctt tcgagcctta tagtggcggc cggtctacat cc 172

<210> 82
<211> 201
<212> ДНК
<213> Rattus norvegicus

<400> 82
ggctgatgag cttcccccgc ccactaggag tgtgaagacc tgccgccata ataagactcc 60

aaaagacagt gaatttaaca cttacggtga cttcccacaa agcacagcgt gtaatttgca 120

tgcgctctag cccaggctcc agctccggac cagaagcccg cgcatcccgg caaagggtga 180

tgacgtcgtc cttcaagcgc t 201

<210> 83
<211> 525
<212> ДНК
<213> Homo sapiens

<400> 83
gcctcccagc gtcgcgccct aacgacccgc aagtgtccga gggcgcctcc cggccgccat 60

cggccgccct cgcagccgcc gctctcctca cggcctcccg gccgccgccg ccatcttccg 120

ctttctcgtc cggctgcggc gctgctgacg ctagcgagtc gccacgccgg gcaagagcgg 180

ccccccctgcg cccgcagaga acgctgggat gccagcggcg cccgcggagg cctcaccccc 240

tacctcggcc gctccagggg gcgggcctgc atctgggcca cctcttttgc atattggcac 300

ccacaatcca ccgcggctat gaggccagta taaggcggta aaattacgat aagatatggg 360

attttacgtg atcgaagaca tcaaagtaag cgtaagcacg aaagttgttc tgcaacatac 420

cactgtagga aattatgcta aatatgaaac cgaccataag ttatcctaac caaaagatga 480

tttgattgaa gggcttaaaa taggtgtgac agtaaccctt gagtc 525

<210> 84
<211> 176
<212> ДНК
<213> Mus musculus

<400> 84
ttcaggatgt agaccggccg ccactataag gctcgaaaga ggaataaatt tttcgtttag 60

ggtgatttcc cacaaagcac agcgcgtaat ttgcatgcgc tctaccccag gctcctgtgc 120

tagacaagaa gcccgcgcat ccgggcaagg gatgatgacg tcgtccttca agagcg 176

<210> 85
<211> 176
<212> ДНК
<213> Rattus norvegicus

<400> 85
aggagtgtga agacctgccg ccataataag actccaaaag acagtgaatt taacacttac 60

ggtgacttcc cacaaagcac agcgtgtaat ttgcatgcgc tctagcccag gctccagctc 120

cggaccagaa gcccgcgcat cccggcaaag ggtgatgacg tcgtccttca agcgct 176

<210> 86
<211> 225
<212> ДНК
<213> Dipodomys ordii

<400> 86
aggaaagact tcgctgaggc agactttata aggctcccgc gcagaaagaa actttatagt 60

tatggtgatt tcccacaagc cactgcgtca tgcaaataaa gcagggtacg gcttccatgt 120

accttaaggt ttttttctag gccgcgtacg ctctgcgtat tcagccacgt gaccctgagc 180

cagtggttgt tgggagcacg ttgtggacct ctgcgtttgg attcc 225

<210> 87
<211> 222

<212> ДНК
<213> Ictidomys tridecemlineatus

<400> 87
gaaagggact ccgcacaagc agagtttata aggctcccat ctgtacagcc atttctcggt 60

catggtaact acccacaaca cacagcgata tgcaaatata gcagagcgtg tcttcccgcg 120

cgcgcctggt cgtctcggcg ccggcgcgct gcgtggggcg gaactgtgac agagaccctg 180

cgattcctgg gagctggctg atgacatcag tgtctaacct cc 222

<210> 88
<211> 225
<212> ДНК
<213> Cavia porcellus

<400> 88
gagaaagaaa ggctcaaacc tagccttata aggctcccaa atgtcggtat atttttggt 60

tatggtgact tcccacaatg catagcgata tgtagatatt gccaggagta cctcccactt 120

ctggtcctgt cagctctttt ctaggacgcg cgcgctgcag gtttccagcc tgtgattggg 180

ccagcaattc cgggaatgaa ttgatgacgt cagcgtttga attcc 225

<210> 89
<211> 211
<212> ДНК
<213> Ochotona princeps

<400> 89
gggggaagct gggctcgatc agcctttata aagctccaaa aactcaagac atttttctgt 60

tacggtggct tcccacagta cacagcgaca tgcaaatagc ttgccaatga attcgcggac 120

cgcttcccgc cccggcgcag gcgcgcggac gctgtctccc ctggacgcgc gctcgcggtt 180

cccgggagct ggctgatgac gttcggtctc c 211

<210> 90
<211> 211
<212> ДНК
<213> Oryctolagus cuniculus

<400> 90
ggggagaggt ggatccgaac agactttata aagctccgaa agcccaaggc atctttccct 60

tacggtagct tcccacaaga catagcgaca tgcaaatttc ttgaagtatg cttcagacgc 120

gcttctcgcc acagcgcaag cgcgctgtgt gctgacgcgg gaacgggcca gggcgcggtt 180

cccgggagcg ggttgatgac gttagatctc c 211

<210> 91
<211> 229
<212> ДНК
<213> Callithrix jacchus

<400> 91

gaggaaaagt agtcccacag acaacttata agattcccat accctaagac atttcacgat 60

tatggtgact tcccagaaga cacagcgaca tgcaaatatt gcaggtcgtg tttcgcctgt 120

ccctcacagt cgtcttcctg ccagggcgca cgcgcgctgg gtttcccgcc aactgacgct 180

gggctcgcga ttccttggag cgggttgatg acgtcagcgt ttgaattcc 229

<210> 92
<211> 230
<212> ДНК
<213> Chlorocebus sabaeus

<400> 92
ggggaagggt ggtcccttac agaacttata agattcccaa actcaaagac atttcacgtt 60

tatggtgact tcccagaaga catagcgaca tgcaaatatt gcagggcgtc acacccctct 120

ccctcacagt catcttcctg ccagggcgca cgcgcgctgg gtgttctcgc gtagtgacac 180

tgggcccgcg attccttgga gcgggttgat gacgtcagcg ttcgaattcc 230

<210> 93
<211> 230
<212> ДНК
<213> Macaca mulatta

<400> 93
ggggaagggt ggtcccacac agaacttata agattcccat actcaaagac atttctcgtt 60

tatggtgact tcccagaaga cacagcgaca tgcaaatatt gtagggcgtc acacccctgt 120

ccctcacagt catcttcctg ccagggcgca cgcgcgctgg gtgttcccgc gtagtgacac 180

tgggcccgcg attccttgga gcgggttgat gacgtcagcg ttcgaattcc 230

<210> 94
<211> 231
<212> ДНК
<213> Papio anubis

<400> 94
ggggaaaggt ggtaccatac agaacttata agattcccat actcaaagac atttcacgat 60

tatggtgact tcccagaaga cacagcgaca tgcaaatatt gtagggcgtc acacccccctg 120

tccctcacag tcatcttcct gccagggcgc acgcgcgctg ggtgttcccg cgtagtgaca 180

ctgggcccgc gattccttgg agcgggttga tgacgtcagc gttcgaattc c 231

<210> 95
<211> 230
<212> ДНК
<213> Gorilla gorilla

<400> 95
gggaaagggt ggtcccacac agaacttata agactcccat atccaaagac atttcacggt 60

tatggtgatt tcccagaaca catagcgaca tgtaaatatt gcagggcgcc actccccagt 120

ccctcacagc catcttcctg ccagggcgca cgcgcgctgg gtgttcccgc ctagtgacac 180

tgggcccgcg attccttgga gcgggttgat gacgtcagcg ttcgaattcc 230

<210> 96
<211> 230
<212> ДНК
<213> Homo sapiens

<400> 96
gggaaaaagt ggtctcatac agaacttata agattcccaa atccaaagac atttcacgtt 60

tatggtgatt tcccagaaca catagcgaca tgcaaatatt gcagggcgcc actcccctgt 120

ccctcacagc catcttcctg ccagggcgca cgcgcgctgg gtgttcccgc ctagtgacac 180

tgggcccgcg attccttgga gcgggttgat gacgtcagcg ttcgaattcc 230

<210> 97
<211> 230
<212> ДНК
<213> Pan troglodytes

<400> 97
gggaaagggt ggtgccacac agaacttata agattcccat atgcaaagac atttcacgtt 60

tatggtgatt tcccagaaca catagcgaca tgcaaatatt gcagggcgcc actcccctgt 120

ccctcactgc catcttcctg ccagggcgca cgcgcgctgg gtgttcccgc ctagtgacac 180

tgggcccgcg attccttgga gcgggttgat gacgtcagcg ttcgaattcc 230

<210> 98
<211> 229
<212> ДНК
<213> Pongo abelii

<400> 98
gagaaagggt ggtcccgtcc agaacttata agattcccat acccaaagac atttcacgtt 60

tatggtgact tcccagaatg catagcgaca tgcaaatatt gcagggcgtc actcccctgt 120

ccctcacagc catcttcctg ccagggcgcc cgcgcgctgg tgttcccgcc tagtgacact 180

gggcccacga ttccttggag cgggttgatg acgtcagcgc tcgtattcc 229

<210> 99
<211> 223
<212> ДНК
<213> Nomascus leucogenys

<400> 99
ggggaaaagt agtagacctt ataagattcc caaacccaaa gacatttctc gtttatggtg 60

acttcccaga agacatagcg acatgcaaat attgcagggc gccactcccc tgtccctcac 120

agccatcttc ctgccagggc gcacgcgcgc tgggtgttcc cgcctagtga cactcggccc 180

gcgattcctt ggagcgggtt gatgacgtca gcgttcgaat tcc 223

<210> 100
<211> 227
<212> ДНК
<213> Tarsius syrichta

<400> 100
gcgagagggt gggtccacac agagcttata aggttcacaa gtaaagatat ttcacggtga 60

cggtgacttc ccacaataca ctgcgacatg caaatatagc cgggcgtgcc tccccgatcc 120

cggaagagcg actcctagcc agtgcgcacg cgcgctgcgt gttcgcgtcc taggtcgctg 180

ggcccgcggt tcctgggagc gggtggtgac gtcagcggcc cagcttc 227

<210> 101
<211> 230
<212> ДНК
<213> Otolemur garnettii

<400> 101
gcctaaaagg gcgcttgcac agaatttata aggttcccaa acagagacac atttcattat 60

tatggtgact tcccacaatg cacagcgcca tgcaaatatg ctaggacgcc tccccccgct 120

accttaaggt cgtcaactaa ccagtgcgcg cgcgcactgc gcgtttcccg ccggtgactc 180

aatgcccgcg tttggtggga gctagttggt gacctcagtt ctggaggctc 230

<210> 102
<211> 201
<212> ДНК
<213> Tupaia belangeri

<400> 102
gggggaagct gggtccactg agttcttata aggtttccag tcctagagcg attttaccat 60

tgcggtgatt tcccagcatc cgtagctaca tgcaaatagc gcggggcgcg tctctcaggt 120

ccctccccgc cctctcactg tacgtacccg cgtcctaggg acgccgcgcc cggggttccc 180

ggacgtcagc gttccgacgc a 201

<210> 103
<211> 226
<212> ДНК
<213> Ailuropoda melanoleuca

<400> 103
agggaaagcc gcgcctgggg cggatttata aggcttccat atctaaaggc atttcacagt 60

catggtgact tcccacaata catagcaaca tgcaaatatc gcggggagaa cctcccctgt 120

cccttgtacg cggcttctaa agacgcacgc gcgctctgtg ttcccgccct gtgactctag 180

gcgggcaatt cctgggacag tgttctgacg ggaacgttca ggctcc 226

<210> 104
<211> 227
<212> ДНК

<213> Mustela putorius

<400> 104
gggaaagggt ggacccaccg agcatttata aggctcccgc atctaaagac attttacagt 60

tatggtgact tcccacaacg cgtagcaaca tgcaaatatc gtggagagta ccgcccctgt 120

cccatgcacg cgtcttctca gcagcacgca cgcgcgctgt gttcccgccc tgtgactcca 180

ggcgggtatt tccaggggcg ggttgctgac aggaacgttc aggcttc 227

<210> 105
<211> 229
<212> ДНК
<213> Canis familiaris

<400> 105
gcagcgcagc cctctcgccg cttataaagt gccgcccgca cggcccttct cgctcacggc 60

gacttcccat aacacacagc agcatgcaaa taccgcgggg agccccgccc cgccccggcc 120

cccgcaccgc ctcgggacgc atgcgccggc tctccgttcc cgccttgggc cggcggcggg 180

cgggcgggcg agcgggcggg agcggctccg gcggggacga gcgggcgcc 229

<210> 106
<211> 228
<212> ДНК
<213> Felis catus

<400> 106
gggaaagggt ggccccgccg agcatttata agactcccat acctaaagac atttctcagt 60

tatggtgatt tcccacaaca cacagcaaca tgcaaatatc gaggggtgta ccgcccctgt 120

cctttgtaga cgtcttctct ccaggacgca cgcgcgctgt attcccgcct tgtgactcta 180

ggcgggcgat tcctgggaga gggttgatga cgtccaagtt ctggcttc 228

<210> 107
<211> 227
<212> ДНК
<213> Equus caballus

<400> 107
gggggaaaac agcccatggc tgcatttata agactcacag atctaaagcc atttcacgaa 60

tagggtgact tcccacaata cacagcgaca tgcaaacata gcggggcgtg cctttcctgt 120

accctgtggg catctctcct ggacgcacgc gcgccgggtg ttcccgcgct gtgactctag 180

gcaagcgctt cctgggagag agttgatgac ggcagcattc gggctcc 227

<210> 108
<211> 225
<212> ДНК
<213> Myotis lucifugus

<400> 108
gggagaagga ggcgtagagg atatataagg ccccccttatg tgtagtcctt ttacggttag 60

ggtgacttcc cacaacgcat agcgacatgc aaatttgacg ggcgtgcctc ctctgtccct 120

gcgggcaact tctctcctgg acgcgcgcgc gctgcgtgtt cccgcctttt gactccagcc 180

gagcgaatcc tgggagaggg caggtgacgt caacagtcag gctcg 225

<210> 109
<211> 227
<212> ДНК
<213> Pteropus vampyrus

<400> 109
gcgagaaaaa ttcttcacgc agaatatata aggatcccat atctgaagac attttacgat 60

tacggcgatt tcccacaaca catagcgaca tgtaaatgta gtggggcatg cctcccctgt 120

cccttgtggg cagcttctcg ccagaacgca cgcgcggtgc gtgttcccgc cttgtgacta 180

agttggcgag tcagggagga gattgatgac gtcagctcac ccgctcc 227

<210> 110
<211> 218
<212> ДНК
<213> Bos taurus

<400> 110
ggcaaacacc gcacgcaaat agcacttata atgtgctcat acctagagcc acttttcggt 60

tacggtgact tctcaaaaag acagtggaac atgcaaatat tacagtgcgt cccgcccctg 120

gtaggtctac gctaggacgc acgcgcacta cggttcccgc ctatagactg cgctggcgat 180

tcctgggagc ggactgatga cgtcagcgtt cgggatcc 218

<210> 111
<211> 216
<212> ДНК
<213> Ovis aries

<400> 111
ggcgaacaat gcgcgcaaac agcatttata atgagctcat acctaaagcc actttacggt 60

tacggtgact tcccacaaga cattgcggca tgcaaatatt ttagtgcgtc ccgcccctgg 120

tagttccacg ctaggacgca cacgcactac ggttcccgcc tttagactgc gctggcgatt 180

ccaggagcgg actgatgacg tcagcgttgg ggctcc 216

<210> 112
<211> 218
<212> ДНК
<213> Tursiops truncatus

<400> 112
gccgaaaacc aggctcaagc cacatttata aggctcccaa atctaagtac atttgtcggt 60

tatggtgact tcccgcacca cattgcgaca tgcaaatact gcggagcgtc cctcccctgg 120

caactcctcg ctgggacgca cgcgcgctac gtgctcccgc cttttgactg cgccggcgat 180

acttgggaga gggttgatga cgtcagcgtt ctggctcc 218

<210> 113
<211> 226
<212> ДНК
<213> Vicugna pacos

<400> 113
gggaaagggt gggctcacgc agcctttata gactcccaa acttaaagac atttctcggt 60

tatggcgact tcccacaaga catagcgaca tgcaaatact gcagggcgcc gacccggtcc 120

tgtgcagcca tctttcggct gggacgcacg cgcgctgcgt gttcccgccc tgtgactgcg 180

ccggcgatta ctgggagagg attgatgacg tcaacgttcg ggttcc 226

<210> 114
<211> 228
<212> ДНК
<213> Sus scrofa

<400> 114
gtaggaaaac tgcttctgtg agcacttata aaactcccat aagtagagag atttcatagt 60

tatggtgatt tcccataaga cattgcgaca tgcaaatatt gtggcgcgtt cgtccccgtc 120

cggtgcaggc agcttcgctc caggacgcac gcgcaataca tgttcccgcc ttgagactgc 180

gccggcagat tcctaggaag tggttgatga cgtcgatgtt agggatcc 228

<210> 115
<211> 210
<212> ДНК
<213> Erinaceus europaeus

<400> 115
gcctaaaccg gctctttcga cagacttata aggacctctt atcttaggac atttttttgt 60

tagggtaact tcccacgatg catagcgata tgtaaatatg gcgccgcgag tctctcctag 120

gcgtctcccc aggacgcagg cgcactgctt gttcccgcgt taacattgct gattctggga 180

gactgctgat gacgtcagcg tccagtctac 210

<210> 116
<211> 211
<212> ДНК
<213> Choloepus hoffmanni

<400> 116
agaaaaaaat agtttatgct ggatttataa gattcccaaa tctaaagcca tttcacagtt 60

acggtgattc cccactacac acggcgatat gcaaatatag cggaagtgtt cctgaggcgt 120

ggtaaagcgc gcgcgcgctg agagttcccg ccctgtggtg ctgggctgga gatgcctgag 180

aactggctga tgacggcaac gttcgggctc c 211

<210> 117
<211> 208
<212> ДНК
<213> Dasypus novemcinctus

<400> 117
aaagcgatag ttttttaaac tggacttata aggcacccat atctacgtat atttcatggt 60

tagggtgatt tcccacaaca catagcgaaa tgcaaatatt ggagggcgct gaggcgtggt 120

cgggcgcaag cgcgctgcga cttcccgcct ttcggcccta ggccccagat tcctgggagc 180

tggatgatga cgttgacgtt cggatacc 208

<210> 118
<211> 227
<212> ДНК
<213> Loxodonta africana

<400> 118
gggaaggaac aaattcgtca ggatttataa gactctcaga gctgtagaca tttcacagtt 60

agggcgatgt cccacaatac atagcaacat gcaaatattc taggaggcca gcctccccgt 120

ccgcgtggtc atcttctcgc tagggcgcac gcccgctgcg tgttcccgct ctgtgaccag 180

gcaggcgatt cctgagaacc gcttggtgac gtcagtgttc tggctcc 227

<210> 119
<211> 225
<212> ДНК
<213> Procavia capensis

<400> 119
agggtaaatc ggcgctgctc agcatttaaa agaatcccaa atgtgtcgcc attttacgct 60

tagggtgata tcccacaaga cacagcgaca tgcaaatatc gtgagtctct gtttccctgt 120

ccacgagggc gtcctctcgc tggggcgcac gcgcggtgtg tgtgcccccg ttgtgtgttc 180

ccgcgattcc aaagaactgg ttgataacgt tagacttccg gctgc 225

<210> 120
<211> 49
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический олигонуклеотид

<400> 120
aataaaatat ctttattttc attacatctg tgtgttggtt ttttgtgtg 49

<210> 121
<211> 154
<212> ДНК
<213> Искусственная последовательность

<220>

<223> Описание искусственной последовательности: Синтетический
полинуклеотид

<400> 121
aataaaatat ctttattttc attacatctg tgtgttggtt ttttgtgtga atcgatagta 60

ctaacatacg ctctccatca aaacaaaacg aaacaaaaca aactagcaaa ataggctgtc 120

cccagtgcaa gtgcaggtgc cagaacattt ctct 154

<210> 122
<211> 240
<212> ДНК
<213> Simian virus 40

<400> 122
atctagataa ctgatcataa tcagccatac cacatttgta gaggttttac ttgctttaaa 60

aaacctccca cacctccccc tgaacctgaa acataaaatg aatgcaattg ttgttgttaa 120

cttgtttatt gcagcttata atggttacaa ataaagcaat agcatcacaa atttcacaaa 180

taaagcattt ttttcactgc attctagttg tggtttgtcc aaactcatca atgtatctta 240

<210> 123
<211> 120
<212> ДНК
<213> Simian virus 40

<400> 123
ttgtttattg cagcttataa tggttacaaa taaagcaata gcatcacaaa tttcacaaat 60

aaagcatttt tttcactgca ttctagttgt ggtttgtcca aactcatcaa tgtatcttat 120

<210> 124
<211> 228
<212> ДНК
<213> Неизвестная

<220>
<223> Описание неизвестной:
терминаяльная последовательность

<400> 124
cgactgtgcc ttctagttgc cagccatctg ttgtttgccc ctccccccgtg ccttccttga 60

ccctggaagg tgccactccc actgtccttt cctaataaaa tgaggaaatt gcatcgcatt 120

gtctgagtag gtgtcattct attctggggg gtggggtggg gcaggacagc aaggggggagg 180

attgggaaga caatagcagg catgctgggg atgcggtggg ctctatgg 228

<210> 125
<211> 271
<212> ДНК
<213> Неизвестная

<220>
<223> Описание неизвестной:
терминальная последовательность

<400> 125
gggggaggct aactgaaaca cggaaggaga caataccgga aggaacccgc gctatgacgg 60

caataaaaag acagaataaa acgcacgggt gttgggtcgt ttgttcataa acgcggggtt 120

cggtcccagg gctggcactc tgtcgatacc ccaccgagac cccattgggg ccaatacgcc 180

cgcgtttctt ccttttcccc accccacccc ccaagttcgg gtgaaggccc agggctcgca 240

gccaacgtcg gggcggcagg ccctgccata g 271

<210> 126
<211> 32
<212> ДНК
<213> неизвестная

<220>
<223> Описание неизвестной:
терминальная последовательность

<400> 126
ggtatcaaat aaaatacgaa atgtgacaga tt 32

<210> 127
<211> 26
<212> ДНК
<213> Неизвестная

<220>
<223> Описание неизвестной:
терминальная последовательность

<400> 127
aaataaaata cgaaatgtga cagatt 26

<210> 128
<211> 100
<212> ДНК
<213> Неизвестная

<220>
<223> Описание неизвестной:
терминальная последовательность

<400> 128
ggttgctgat ttctccacag cttgcatttc tgaaccaaag gccctttca gggccgccca 60

actaaacaaa agaagagctg tatccattaa gtcaagaagc 100

<210> 129
<211> 174
<212> ДНК
<213> неизвестная

<220>
<223> Описание неизвестной:
терминальная последовательность

<400> 129

gattcgtcag tagggttgta aaggtttttc ttttcctgag aaaacaacct tttgttttct 60

caggttttgc tttttggcct ttccctagct ttaaaaaaaa aaaagcaaaa gacgctggtg 120

gctggcactc ctggtttcca ggacggggtt caagtccctg cggtgtcttt gctt 174

<210> 130
<211> 123
<212> ДНК
<213> Неизвестная

<220>
<223> Описание неизвестной:
терминальная последовательность

<400> 130
aaaggttttt cttttcctga gaaatttctc aggttttgct ttttaaaaaa aaagcaaaag 60

acgctggtgg ctggcactcc tggtttccag gacggggttc aagtccctgc ggtgtctttg 120

ctt 123

<210> 131
<211> 204
<212> ДНК
<213> Homo sapiens

<400> 131
gagggcagtc accagctcct ggcccgtgcg ccaagctcag cgggcgtccg cggtgcgatc 60

ttccctagcg cctcgggtct ggcgccgcca tcttcctcgg taacaaccag tcgcctgagg 120

cgtggggccg cctcccaaag acttctggga gggcggtgcg gctcaggctc tgccccgcct 180

ccggggctat ttgcatacga ccat 204

<210> 132
<211> 50
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический
олигонуклеотид

<400> 132
acgggtgtgg tacgcagccc ctgaattcac ccacagtact acctggctga 50

<210> 133
<211> 50
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический
олигонуклеотид

<400> 133
acgggtgtgg tacgcagccc cttcgagtac ccacagtact acctggctga 50

<210> 134
<211> 351
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический
полинуклеотид

<220>
<221> CDS
<222> (1)..(102)

<220>
<221> CDS
<222> (106)..(177)

<220>
<221> CDS
<222> (181)..(327)

<220>
<221> CDS
<222> (331)..(351)

<400> 134

```
tgg agc cct gag tgg ctg agc tca ggc ctt cgc agc att ctt ggg tgg         48
Trp Ser Pro Glu Trp Leu Ser Ser Gly Leu Arg Ser Ile Leu Gly Trp
1               5                  10                  15

gag cag cca cgg gtc agc cac aag ggc cac agc cat gaa tgg cac aga         96
Glu Gln Pro Arg Val Ser His Lys Gly His Ser His Glu Trp His Arg
            20                  25                  30

agg ccc taa ctt cta cgt gcc ctt ctc caa tgc gac ggg tgt ggt acg        144
Arg Pro     Leu Leu Arg Ala Leu Leu Gln Cys Asp Gly Cys Gly Thr
            35                  40                  45

cag ccc ctg aat tca ccc aca gta cta cct ggc tga gcc atg gca gtt        192
Gln Pro Leu Asn Ser Pro Thr Val Leu Pro Gly     Ala Met Ala Val
        50                  55                      60

ctc cat gct ggc cgc cta cat gtt tct gct gat cgt gct ggg ctt ccc        240
Leu His Ala Gly Arg Leu His Val Ser Ala Asp Arg Ala Gly Leu Pro
        65                  70                  75

cat caa ctt cct cac gct cta cgt cac cgt cca gca caa gaa gct gcg        288
His Gln Leu Pro His Ala Leu Arg His Arg Pro Ala Gln Glu Ala Ala
    80                  85                  90

cac gcc tct caa cta cat cct gct caa cct agc cgt ggc tga cct ctt        336
His Ala Ser Gln Leu His Pro Ala Gln Pro Ser Arg Gly     Pro Leu
95                  100                 105

cat ggt cct agg tgg                                                    351
His Gly Pro Arg Trp
110
```

<210> 135
<211> 34
<212> Белок

<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полипептид

<400> 135
Trp Ser Pro Glu Trp Leu Ser Ser Gly Leu Arg Ser Ile Leu Gly Trp
1 5 10 15

Glu Gln Pro Arg Val Ser His Lys Gly His Ser His Glu Trp His Arg
20 25 30

Arg Pro

<210> 136
<211> 24
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический пептид

<400> 136
Leu Leu Arg Ala Leu Leu Gln Cys Asp Gly Cys Gly Thr Gln Pro Leu
1 5 10 15

Asn Ser Pro Thr Val Leu Pro Gly
20

<210> 137
<211> 49
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полипептид

<400> 137
Ala Met Ala Val Leu His Ala Gly Arg Leu His Val Ser Ala Asp Arg
1 5 10 15

Ala Gly Leu Pro His Gln Leu Pro His Ala Leu Arg His Arg Pro Ala
20 25 30

Gln Glu Ala Ala His Ala Ser Gln Leu His Pro Ala Gln Pro Ser Arg
35 40 45

Gly

<210> 138

<211> 7
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический
пептид

<400> 138
Pro Leu His Gly Pro Arg Trp
1 5

<210> 139
<211> 151
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический
полинуклеотид

<220>
<221> CDS
<222> (3)..(149)

<400> 139
gc agg att ccc ctc tgg ctc act ggc agt ctc ctc cga tgt ggg cca 47
Arg Ile Pro Leu Trp Leu Thr Gly Ser Leu Leu Arg Cys Gly Pro
1 5 10 15

ggg ctc ttt gaa gtt gga tct gag cct ttc tat cac ctg ttt gat gga 95
Gly Leu Phe Glu Val Gly Ser Glu Pro Phe Tyr His Leu Phe Asp Gly
20 25 30

caa gcc ctt ttg cac aag ttt gac ttc aag gag ggc cat gtc aca tac 143
Gln Ala Leu Leu His Lys Phe Asp Phe Lys Glu Gly His Val Thr Tyr
35 40 45

cac aga ag 151
His Arg

<210> 140
<211> 49
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический
полипептид

<400> 140
Arg Ile Pro Leu Trp Leu Thr Gly Ser Leu Leu Arg Cys Gly Pro Gly
1 5 10 15

Leu Phe Glu Val Gly Ser Glu Pro Phe Tyr His Leu Phe Asp Gly Gln
20 25 30

Ala Leu Leu His Lys Phe Asp Phe Lys Glu Gly His Val Thr Tyr His

```
        35                  40                  45

Arg


<210> 141
<211> 324
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический
      полинуклеотид


<220>
<221> CDS
<222> (1)..(12)

<220>
<221> CDS
<222> (16)..(39)

<220>
<221> CDS
<222> (43)..(225)

<220>
<221> CDS
<222> (229)..(243)

<220>
<221> CDS
<222> (247)..(324)

<400> 141
tat aac aga ctc taa tgc aag cag tat gag agg ctt gga tag gct ctg         48
Tyr Asn Arg Leu     Cys Lys Gln Tyr Glu Arg Leu Gly     Ala Leu
1                   5                   10

ata tgg tgc tgt gta ggc tca tat gtg gat ctc aga acc cac atg tac         96
Ile Trp Cys Cys Val Gly Ser Tyr Val Asp Leu Arg Thr His Met Tyr
15                  20                  25                  30

tct gct ccc cag gtc ttg gtg cgc ttt cta ttc tct gtc agc aaa gcc        144
Ser Ala Pro Gln Val Leu Val Arg Phe Leu Phe Ser Val Ser Lys Ala
                35                  40                  45

tat cga aga atc acc tac cac aac tgg tgc cac ggc ttc aat gta gcc        192
Tyr Arg Arg Ile Thr Tyr His Asn Trp Cys His Gly Phe Asn Val Ala
            50                  55                  60

cag acc atg ttt acc cta ctc atg gta cgt atg taa att gga tgg gct        240
Gln Thr Met Phe Thr Leu Leu Met Val Arg Met     Ile Gly Trp Ala
        65                  70                      75

aga tga atc aga ggg ctg ggg caa gga cca cag cta act atc ttc tgg        288
Arg     Ile Arg Gly Leu Gly Gln Gly Pro Gln Leu Thr Ile Phe Trp
            80                  85                  90

ccc aag gat gcc aat tgt gtg tat cca gtc cta gca                        324
Pro Lys Asp Ala Asn Cys Val Tyr Pro Val Leu Ala
```

95 100

<210> 142
<211> 4
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический пептид

<400> 142
Tyr Asn Arg Leu
1

<210> 143
<211> 8
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический пептид

<400> 143
Cys Lys Gln Tyr Glu Arg Leu Gly
1 5

<210> 144
<211> 61
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический полипептид

<400> 144
Ala Leu Ile Trp Cys Cys Val Gly Ser Tyr Val Asp Leu Arg Thr His
1 5 10 15

Met Tyr Ser Ala Pro Gln Val Leu Val Arg Phe Leu Phe Ser Val Ser
20 25 30

Lys Ala Tyr Arg Arg Ile Thr Tyr His Asn Trp Cys His Gly Phe Asn
35 40 45

Val Ala Gln Thr Met Phe Thr Leu Leu Met Val Arg Met
50 55 60

<210> 145
<211> 5
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический

пептид

<400> 145
Ile Gly Trp Ala Arg
1 5

<210> 146
<211> 26
<212> Белок
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический пептид

<400> 146
Ile Arg Gly Leu Gly Gln Gly Pro Gln Leu Thr Ile Phe Trp Pro Lys
1 5 10 15

Asp Ala Asn Cys Val Tyr Pro Val Leu Ala
20 25

<210> 147
<211> 85
<212> РНК
<213> Искусственная последовательность

<220>
<223> Описание искусственной последовательности: Синтетический олигонуклеотид

<400> 147
ggacgcgacc gaaaugguga aggacggggu ccagugcgaa acacgcacug uugaguagag 60

ugugagcucc guaacugguc gcguc 85

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Не встречающаяся в природе нуклеазная система, содержащая один или несколько векторов, содержащих:

a) промотор, функционально связанный по меньшей мере с одной нуклеотидной последовательностью, кодирующей гидовую РНК (гРНК) нуклеазной системы, где гРНК гибридизуется с последовательностью-мишенью из молекулы ДНК или молекулы РНК в клетке и где молекула ДНК или молекула РНК кодирует один или несколько продуктов генов, экспрессируемых в клетке; и b) регуляторный элемент, функционирующий в клетке, функционально связанный с нуклеотидной последовательностью, кодирующей нацеленную на геном нуклеазу,

где компоненты (a) и (b) расположены на одном и том же или на различных векторах системы, где гРНК осуществляет нацеливание на последовательность-мишень и гибридизуется с ней, и нуклеаза расщепляет молекулу ДНК или РНК для изменения экспрессии одного или нескольких продуктов генов.

2. Система по п.1, где система представляет собой CRISPR.

3. Система по п.1, где система упакована в одну частицу аденоассоциированного вируса (AAV).

4. Система по п.3, в которой упаковывающий аденоассоциированный вирус выбран из аденовируса серотипа 2, аденовируса серотипа 5 или аденовируса серотипа 35.

5. Система по п.7, в которой аденоассоциированный упаковывающий вирус представляет собой аденовирус серотипа 5.

6. Система по п.1, где система инактивирует один или несколько продуктов генов.

7. Система по п.1, где нуклеазная система вырезает по меньшей мере одну мутацию гена.

8. Система по п.1, в которой выбран из группы, состоящей из промотора H1, 7sk, RPPH1-PARP2 человека, SRP-RPS29, 7sk1-GSTA4, SNAR-G-1-CGB1, SNAR- CGB2, RMRP-CCDC107, tRNA(Lys)-ALOXE3, RNU6-9-MED16: tRNA (Gly)-DPP9, RNU6-2-THEM259, SNORD13-C8orf41, RPPH1-PARP2 мыши и RPPH1-PARP2 крысы.

9. Система по п.8, в которой промотор представляет собой промотор H1.

10. Система по п.9, в которой промотор H1 содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 12 или 32.

11. Система по п.10, в которой промотор H1 содержит нуклеотидную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 12.

12. Система по п.10, в которой промотор H1 содержит нуклеотидную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 32.

13. Система по любому из пп.9-12, в которой промотор является ортологом промотора H1.

14. Система по п.13, в которой ортологичный промотор H1 происходит из мыши или крысы.

15. Система по любому из пп.9-14, в которой промотор H1 является двунаправленным.

16. Система по любому из пп.9-15, в которой промотор H1 содержит:

a) элементы контроля, которые обеспечивают транскрипцию в одном направлении по меньшей мере одной нуклеотидной последовательности, кодирующей гРНК; и

b) элементы контроля, которые обеспечивают транскрипцию в противоположном направлении нуклеотидной последовательности, кодирующей нацеленную на геном нуклеазу.

17. Система по п.8, в которой промотор представляет собой промотор 7sk.

18. Система по п.17, в которой промотор 7sk происходит из человека, крысы или мыши.

19. Система по любому из пп.17-18, в которой промотор 7sk выбран из группы, состоящей из 7sk1, 7sk2 и 7sk3.

20. Система по п.19, в которой промотор 7sk представляет собой 7sk1.

21. Система по п.20, в которой 7sk1 происходит из человека.

22. Система по любому из пп.17-21, в которой 7sk1 человека содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с нуклеотидной последовательностью,

указанной в SEQ ID NO: 3.

23. Система по п.22, в которой 7sk1 человека содержит нуклеотидную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 3.

24. Система по п.19, в которой 7sk1 происходит из мыши.

25. Система по п.24, в которой 7sk1 мыши содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с SEQ ID NO: 6.

26. Система по любому из пп.24-25, в которой 7sk1 мыши содержит нуклеотидную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 6.

27. Система по п.19, в которой промотор 7sk представляет собой 7sk2.

28. Система по п.27, в которой 7sk2 происходит из человека.

29. Система по п.28, в которой 7sk2 человека содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с SEQ ID NO: 4.

30. Система по п.29, в которой 7sk2 человека содержит нуклеотидную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 4.

31. Система по п.27, в которой 7sk2 происходит из мыши.

32. Система по п.31, в которой 7sk2 мыши содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с SEQ ID NO: 7.

33. Система по п.32, в которой 7sk2 мыши содержит нуклеотидную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 7.

34. Система по п.19, в которой промотор 7sk представляет собой 7sk3.

35. Система по п.34, в которой 7sk3 происходит из человека.

36. Система по п.35, в которой 7sk3 человека содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с SEQ ID NO: 5.

37. Система по п.36, в которой 7sk3 человека содержит нуклеотидную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 5.

38. Система по п.34, в которой 7sk3 происходит из мыши.

39. Система по п.38, в которой 7sk3 мыши содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с SEQ ID NO: 8.

40. Система по п.39, в которой 7sk3 мыши содержит нуклеотидную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 8.

41. Система по любому из пп.8-40, в которой промотор имеет по меньшей мере одну модификацию.

42. Система по п.41, в которой по меньшей мере одна модификация промотора содержит элемент, который позволяет регуляцию в зависимости от условий.

43. Система по п.42, в которой элемент представляет собой отвечающий на tet промотор.

44. Система по п.43, в которой отвечающий на tet промотор содержит Tet-репрессор (TetR) и Tet-оператор (TetO), встроенные в промотор промотор H1.

45. Система по п.44, в которой TetR содержит аминокислотную последовательность, обладающую по меньшей мере 80% идентичностью с SEQ ID NO: 14.

46. Система по п.45, в которой TetR содержит аминокислотную последовательность, обладающую аминокислотной последовательностью, указанной в SEQ ID NO: 14.

47. Система по п.45, в которой TetO содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 11.

48. Система по п.47, в которой TetO содержит нуклеотидную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 11.

49. Система по п.44, в которой H1-TetO содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 13.

50. Система по п.49, в которой H1-TetO содержит нуклеотидную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 13.

51. Система по п.41, в которой по меньшей мере одна модификация промотора включает область, которая позволяет ауторегуляцию.

52. Система по п.51, в которой область ауторегуляции содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 33 или 34.

53. Система по п.52, в которой область ауторегуляции содержит нуклеотидную последовательность, указанную в SEQ ID NO: 33.

54. Система по п.52, в которой область ауторегуляции содержит нуклеотидную последовательность, указанную в SEQ ID NO: 34.

55. Система по п.1, дополнительно содержащая по меньшей мере одну последовательность терминатора.

56. Система по п.55, в которой по меньшей мере одна последовательность терминатора выбрана из группы, состоящей из последовательности SV40 или синтетической поли A (SPA) последовательности.

57. Система по п.56, в которой последовательность терминатора представляет собой последовательность SV40 из 120 (SEQ ID NO: 123) или 240 (SEQ ID NO: 122) пар оснований.

58. Система по п.56, в которой последовательность терминатора представляет собой последовательность из SPA 49 пар оснований.

59. Система по п.1, дополнительно содержащая последовательности 5'-нетранслируемой области (5'UTR).

60. Система по п.1, дополнительно содержащая последовательность Козак.

61. Система по п.60, в которой последовательность Козак содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с SEQ ID NO: 1 или 2.

62. Система по п.61, в которой последовательность Козак содержит нуклеотидную последовательность, указанную в SEQ ID NO: 1.

63. Система по п.61, в которой последовательность Козак содержит нуклеотидную последовательность, указанную в SEQ ID NO: 2.

64. Система по п.1, дополнительно содержащая последовательность РНК, которая опосредует кэп-независимую инициацию трансляции.

65. Система по п.64, в которой последовательность РНК выбрана из группы, состоящей из 6.947 или 6.967.

66. Система по п.1, в которой нацеленная на геном нуклеаза представляет собой белок Cas9.

67. Система по п.66, в которой Cas9 содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 61.

68. Система по п.67, в которой Cas9 содержит нуклеотидную последовательность, указанную в SEQ ID NO: 61.

69. Система по любому из пп.66-68, в которой Cas9 содержит по меньшей мере одну модификацию.

70. Система по п.69, в которой по меньшей мере одна модификация в Cas9 содержит изменение расщепляющей последовательности.

71. Система по п.70, в которой изменение расщепляющей последовательности выбрано из группы, состоящей из T2A, P2A, E2A и F2A.

72. Система по п.71, в которой T2A содержит аминокислотную последовательность, обладающую по меньшей мере 80% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 36.

73. Система по п.72, в которой T2A содержит аминокислотную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 36.

74. Система по п.71, в которой P2A содержит аминокислотную последовательность, обладающую по меньшей мере 80% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 35.

75. Система по п.74, в которой P2A содержит аминокислотную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 35.

76. Система по п.71, в которой E2A содержит аминокислотную

последовательность, обладающую по меньшей мере 80% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 37.

77. Система по п.76, в которой E2A содержит аминокислотную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 37.

78. Система по п.71, в которой F2A содержит аминокислотную последовательность, обладающую по меньшей мере 80% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 38.

79. Система по п.78, в которой F2A содержит аминокислотную последовательность, обладающую нуклеотидной последовательностью, указанной в SEQ ID NO: 38.

80. Система по п.69, в которой по меньшей мере одна модификация в Cas9 является кодон-оптимизированной для экспрессии в клетке.

81. Система по любому из пп.66-80, в которой Cas9 дополнительно содержит линкерную последовательность, функционально слитую в рамке считывания с зависимым от клеточного цикла белком (слитая конструкция Cas9).

82. Система по п.81, в которой зависимый от клеточного цикла белок выбран из группы, состоящей из APE2, ATR, BRCA1, Chk1, Cdc5, Cdc6, Cdc7, Cdc45, Cdt1, CSA, CSB, Ctf18, DDB1, DDB2, DNA2, DUT, Elg1, EndoV, Esp1, экзонуклеазы 1, FBH1, FEN1, геминина, Hus1, KNTC2 (NDC80), Ku80, лигазы 1, Mad2, MBD4, Mcm3, Mcm4, Mcm5, Mcm6, Mcm7, Mcm8, Mcm10, MGMT, MLH3, Mms4, MPG, MSH2, Mus81, NBS1, NEIL2, NEIL3, NTH1, Orc1, Orc3, PARP1, PCNA, Pif1, PMS1, PMS2, PNK, Pola p180, Pola p70, Pola Spp1 (Prim2a), Polb, Pold p125, Pole Dpb3, Pole Dpb4, Pole Pol2, Poli, Poll, Polm, Psf1, Psf2, Psf3, Rad1, Rad18, Rad23A, Rad23B, Rad51, Rad51D, Rad54, Rad6A, RPA34, RPA70, Scc1, Scc3, Sir2, SIRT1 (сиртуин), TDG, TDP1, TIMELESS, Tin2, топоизомеразы I, топоизомеразы IIIa, топоизомеразы IIIb, Ubc13, UNG, XAB2, XPC, XPF, XPG, Xrcc2 и XRCC4.

83. Система по п.82, в которой зависимый от клеточного цикла белок представляет собой геминин.

84. Система по п.83, в которой зависимый от клеточного цикла белок представляет собой геминин человека.

85. Система по п.84, в которой геминин человека содержит аминокислоты из положений 1-110 (hGem(1-110)).

86. Система по п.85, в которой hGem(1-110) содержит аминокислотную последовательность, обладающую по меньшей мере 80% идентичностью с аминокислотной последовательностью, указанной в SEQ ID NO: 19.

87. Система по п.86, в которой hGem(1-110) содержит аминокислотную последовательность, указанную в SEQ ID NO: 19.

88. Система по п.82, в которой зависимый от клеточного цикла белок представляет собой Cdt1.

90. Система по п.88, в которой зависимый от клеточного цикла белок представляет собой Cdt1 человека.

91. Система по п.90, в которой Cdt1 человека содержит аминокислоты из положений 30-120 (hCdt1(30-120)).

92. Система по п.91, в которой hCdt1(30-120) содержит аминокислотную последовательность, обладающую по меньшей мере 80% идентичностью с аминокислотной последовательностью, указанной в SEQ ID NO: 18.

93. Система по п.92, в которой hCdt1(30-120) содержит аминокислотную последовательность, указанную в SEQ ID NO: 18.

94. Система по любому из пп.81-93, в которой слитая конструкция Cas9 содержит аминокислотную последовательность, обладающую по меньшей мере 80% идентичностью с аминокислотной последовательностью, выбранной из группы, состоящей из SEQ ID NO: 20-27.

95. Система по п.94, в которой слитая конструкция Cas9 содержит аминокислотную последовательность, выбранную из группы, состоящей из SEQ ID NO: 20-27.

96. Система по любому из пп.66-95, в которой Cas9 функционально слит в рамке считывания с белком убиквитином (Ub-Cas9).

97. Система по п.96, в которой Ub-Cas9 имеет по меньшей мере одну N-концевую модификацию.

98. Система по п.97, в которой Ub-Cas9 с N-концевой модификацией содержит аминокислотную последовательность, обладающую по меньшей мере 80% идентичностью с аминокислотной

последовательностью, выбранной из группы, состоящей из SEQ ID NO: 39-58

99. Система по п.98, в которой Ub-Cas9 с N-концевой модификацией содержит аминокислотную последовательность, выбранную из группы, состоящей из SEQ ID NO: 39-58.

100. Система по п.96, в которой белок убиквитин содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 60.

101. Система по п.100, в которой белок убиквитин содержит нуклеотидную последовательность, указанную в SEQ ID NO: 60.

102. Система по п.1, дополнительно содержащая никазу SaCas9.

103. Система по п.102, дополнительно содержащая донорную матричную последовательность.

104. Система по любому из пп.102-103, в которой по меньшей мере один вектор содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 62 или 67.

105. Система по п.104, в которой по меньшей мере один вектор содержит нуклеотидную последовательность, указанную в SEQ ID NO: 62.

106. Система по п.104, в которой по меньшей мере один вектор содержит нуклеотидную последовательность, указанную в SEQ ID NO: 67.

107. Система по любому из пп.103-106, в которой донорная матричная последовательность корректирует по меньшей мере одну мутацию гена.

108. Система по п.107, в которой по меньшей мере одна мутация гена представляет собой *rd10* или *rd12*.

109. Система по п.108, в которой по меньшей мере один вектор содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 63 или 64.

110. Система по п.109, в которой по меньшей мере один вектор содержит нуклеотидную последовательность, указанную в SEQ

ID NO: 63.

111. Система по п.109, в которой по меньшей мере один вектор содержит нуклеотидную последовательность, указанную в SEQ ID NO: 64.

112. Система по п.108, в которой по меньшей мере один вектор содержит нуклеотидную последовательность, обладающую по меньшей мере 80% идентичностью с нуклеотидной последовательностью, указанной в SEQ ID NO: 65 или 66.

113. Система по п.112, в которой по меньшей мере один вектор содержит нуклеотидную последовательность, указанную в SEQ ID NO: 65.

114. Система по п.112, в которой по меньшей мере один вектор содержит нуклеотидную последовательность, указанную в SEQ ID NO: 66.

115. Система по любому из пп.8-54, в которой промотор функционально связан по меньшей мере с одной, двумя, тремя, четырьмя, пятью, шестью, семью, восемью, девятью или десятью гРНК.

116. Система по п.1, в которой последовательность-мишень содержит нуклеотидную последовательность $AN_{19}NGG$, $GN_{19}NGG$, $CN_{19}NGG$ или $TN_{19}NGG$.

117. Система по п.1, в которой клетка представляет собой эукариотическую или неэукариотическую клетку.

118. Система по п.117, в которой эукариотическая клетка представляет собой клетку млекопитающего или человека.

119. Система по п.5, в которой эукариотическая клетка представляет собой фоторецепторную клетку сетчатки.

120. Система по п.1, в которой один или нескольких продуктов генов представляют собой родопсин.

121. Система по п.1, в которой экспрессия одного или нескольких продуктов генов снижается.

122. Способ изменения экспрессии одного или нескольких продуктов генов в клетке, где клетка содержит молекулу ДНК, кодирующую один или несколько продуктов генов, причем способ включает введение в клетку не встречающейся в природе нуклеазной системы по п.1.

123. Способ по п.122, в котором система инактивирует один или несколько продуктов генов.

124. Способ по п.122, в котором нуклеазная система вырезает по меньшей мере одну мутацию гена.

125. Способ по п.122, в котором экспрессия одного или нескольких продуктов генов снижается.

126. Способ по п.122, в котором клетка представляет собой эукариотическую или неэукариотическую клетку.

127. Способ по п.126, в котором эукариотическая клетка представляет собой клетку млекопитающего или человека.

128. Способ по п.122, в котором клетка представляет собой фоторецепторную клетку сетчатки.

129. Способ по п.122, в котором клетка представляет собой ганглионарную клетку сетчатки.

130. Способ по п.127, в котором эукариотическая клетка представляет собой злокачественную клетку.

131. Способ по п.130, в котором пролиферация злокачественной клетки ингибируется или снижается.

132. Способ по п.130, в котором апоптоз в злокачественной клетки усиливается или повышается.

133. Способ лечения заболевания, выбранного из группы, состоящей из дистрофии сетчатки, дистрофии роговицы, болезни экспансии микросателлитов и злокачественной опухоли у индивидуума, нуждающегося в этом, причем способ включает:

(a) предоставление не встречающейся в природе нуклеазной системы по п.1; и

(b) введение индивидууму эффективного количества системы.

134. Способ по п.133, в котором заболевание представляет собой дистрофию роговицы.

135. Способ по п.133, в котором заболевание представляет собой дистрофию сетчатки.

136. Способ по п.133, в котором дистрофия сетчатки выбрана из группы, состоящей из врожденного амавроза Лебера (LCA), пигментного ретинита (RP) и глаукомы.

137. Способ по п.133, в котором дистрофия роговицы выбрана из группы, состоящей из дистрофии эпителия базальной мембраны

роговицы, эпителиальной рецидивирующей эрозивной дистрофии, субэпителиальной муцинозной дистрофии роговицы, дистрофии роговицы Месманна, эпителиальной дистрофии роговицы Лиша, желатинозной каплевидной дистрофии роговицы, дистрофии роговицы Рейса-Бюклерса, дистрофии роговицы Тиля-Бенке, решетчатой дистрофии роговицы 1 типа (классическая), решетчатой дистрофии роговицы 2 типа, решетчатой дистрофии роговицы типа III, решетчатой дистрофии роговицы типа IIIA, решетчатой дистрофии роговицы типа I/IIIA, решетчатой дистрофии роговицы типа IV, полиморфного амилоидоза (роговицы), гранулярной дистрофии роговицы 1 типа, гранулярной дистрофии роговицы 2 типа, макулярной дистрофии роговицы, дистрофии роговицы Шнайдера, врожденной стромальной дистрофии роговицы, дистрофии роговицы Флека, задней аморфной дистрофии роговицы, центральной облаковидной дистрофии Франсуа, предесцеметовой дистрофии роговицы, эндотелиальной дистрофии роговицы Фукса, задней полиморфной дистрофии роговицы, врожденной наследственной эндотелиальной дистрофии и Х-сцепленной эндотелиальной дистрофии роговицы.

138. Способ по п.133, в котором болезни экспансии микросателлитов выбраны из группы, состоящей из блефарофимоза, птоза и эпикантуса-инверсной синдактилии, ключично-черепной дисплазии, центрального врожденного гиповентиляционного синдрома, синдрома DM (миотоническая дистрофия) Хаддада, FRAXA (синдром ломкой Х-хромосомы), FRAXE (задержка умственного развития, обусловленная ломкой XE), FRDA (атаксия Фридрейха), эндотелиальной дистрофии роговицы Фукса, FXTAS (ассоциированный с ломкой Х синдром тремора/атаксии), ладонно-подошвенно-генитального синдрома, HD (болезнь Гентингтона), голопрозэнцефалии, задержки умственного развития с дефицитом гормонов роста, задержки умственного развития, эпилепсии, синдрома веста, синдрома Партингтона, окулофарингеальной мышечной дистрофии, SBMA (спинальная и бульбарная мышечная атрофия), SCA1 (спинально-церебеллярная атаксия 1 типа), SCA12 (спинально-церебеллярная атаксия 12 типа), SCA17 (спинально-церебеллярная атаксия 17 типа), SCA2 (спинально-церебеллярная

атаксия 2 типа), SCA3 (спинально-церебеллярная атаксия 3 типа или болезнь Мачадо-Джозефа), SCA6 (спинально-церебеллярная атаксия 6 типа), SCA7 (спинально-церебеллярная атаксия 7 типа), SCA8 (спинально-церебеллярная атаксия 8 типа) и синполидактилии.

139. Способ по п.133, в котором введение индивидууму происходит посредством имплантации, инъекции или вирусным путем.

140. Способ по п.139, в котором введение индивидууму происходит посредством субретинальной инъекции.

141. Способ по п.133, в котором введение индивидууму осуществляется местным путем, внутрисосудистым путем, внутрикожным путем, трансдермально, парентерально, внутривенно, внутримышечно, интраназально, подкожно, регионарным путем, подкожно, интратрахеальным путем, внутрибрюшинно, интраартериально, внутрь мочевого пузыря, внутрь опухоли, перитуморальным путем, ингаляционным путем, системно, путем перфузии, посредством промывания, непосредственно посредством инъекции или посредством перорального введения и состава.

142. Способ по п.141, в котором введение индивидууму осуществляется местным путем на поверхность глаза.

143. Способ по п.133, в котором введение индивидууму осуществляется на и снаружи роговицы, склеры, во внутриглазное, субконъюнктивальное, субтеноново или ретробульбарное пространство, или в или вокруг век.

144. Способ по п.133, в котором индивидуума лечат по меньшей мере одним дополнительным средством против злокачественной опухоли.

145. Способ по п.134, в котором средство против злокачественной опухоли выбрано из группы, состоящей из паклитаксела, цисплатина, топотекана, гемцитабина, блеомицина, этопозида, карбоплатина, доцетаксела, доксорубицина, топотекана, циклофосфамида, трабектедина, олапариба, тамоксифена, лектрозола и бевацизумаба.

146. Способ по любому из пп.144-145, в которой индивидуума лечат по меньшей мере одной дополнительной терапией против злокачественной опухоли.

147. Способ по п.146, в котором терапия против

злокачественной опухоли представляет собой лучевую терапию, химиотерапию или хирургическую операцию.

148. Способ по п.133, в котором злокачественная опухоль представляет собой солидную опухоль.

149. Способ по п.133, в котором злокачественная опухоль выбрана из группы, состоящей из злокачественной опухоли головного мозга, желудочно-кишечного рака, рака полости рта, рака молочной железы, рака яичника, рака предстательной железы, рака поджелудочной железы, рака легкого, рака печени, рака горла, рака желудка и рака почки.

150. Способ по п.149, в котором злокачественная опухоль представляет собой злокачественную опухоль головного мозга.

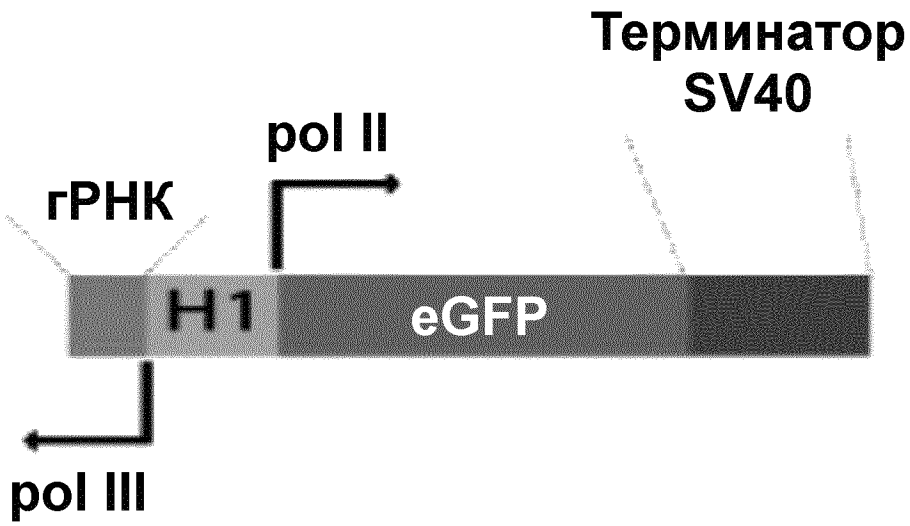
151. Способ по п.141, в котором системное введение выбрано из группы, состоящей из перорального, внутривенного, внутрикожного, внутрибрюшинного, подкожного и внутримышечного введения.

152. Способ по любому из пп.133-153, в котором индивидуумом является млекопитающее.

153. Способ по п.154, в котором млекопитающее представляет собой человека.

По доверенности

ФИГ.1А

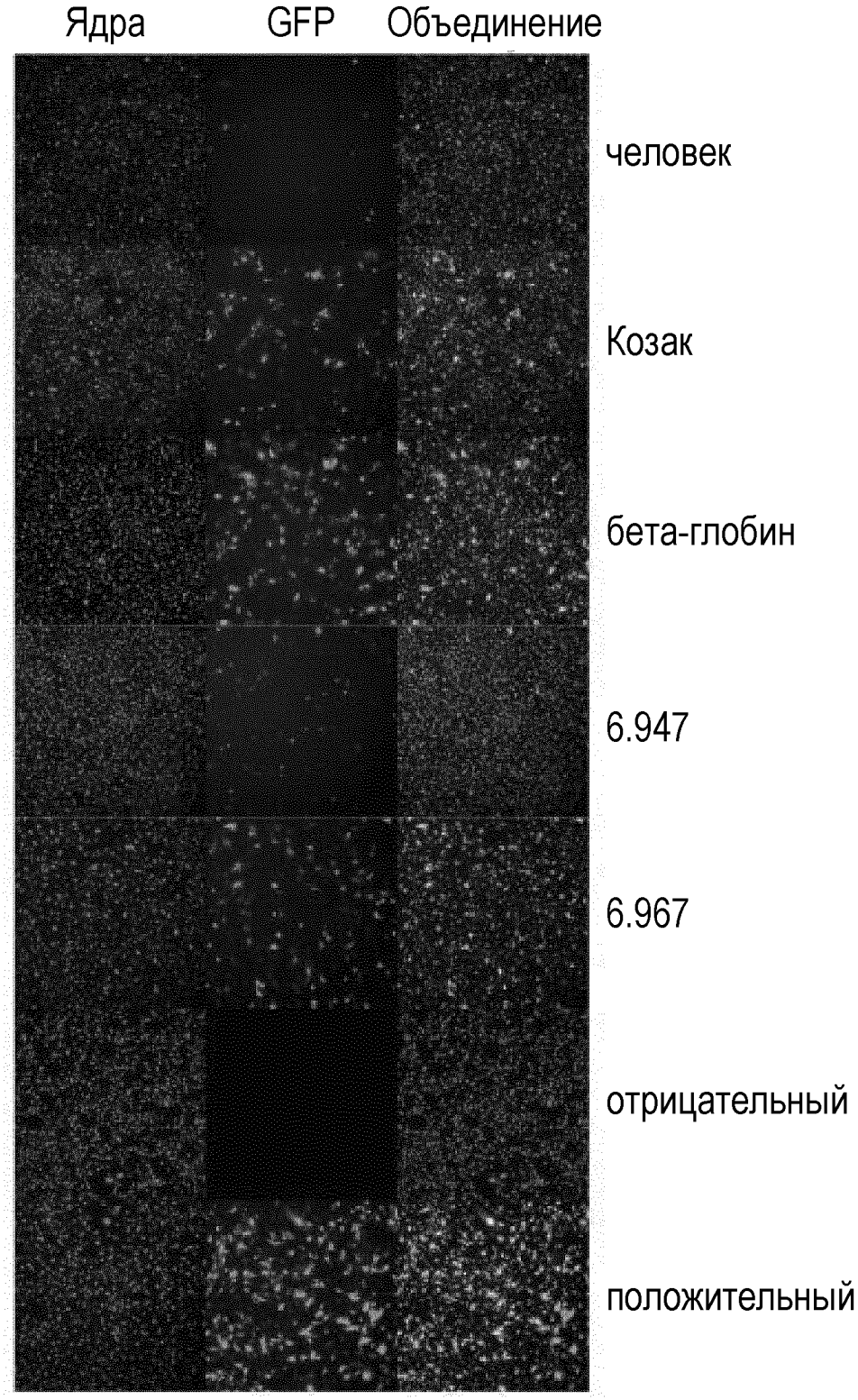


ФИГ.1В

5’UTR и энхансеры трансляции

Сильная последовательность Козак	9 п.н.
5’UTR бета-глобина	50 п.н.
6.947^4	91 п.н.
6.967^4	75 п.н.

ФИГ.1C

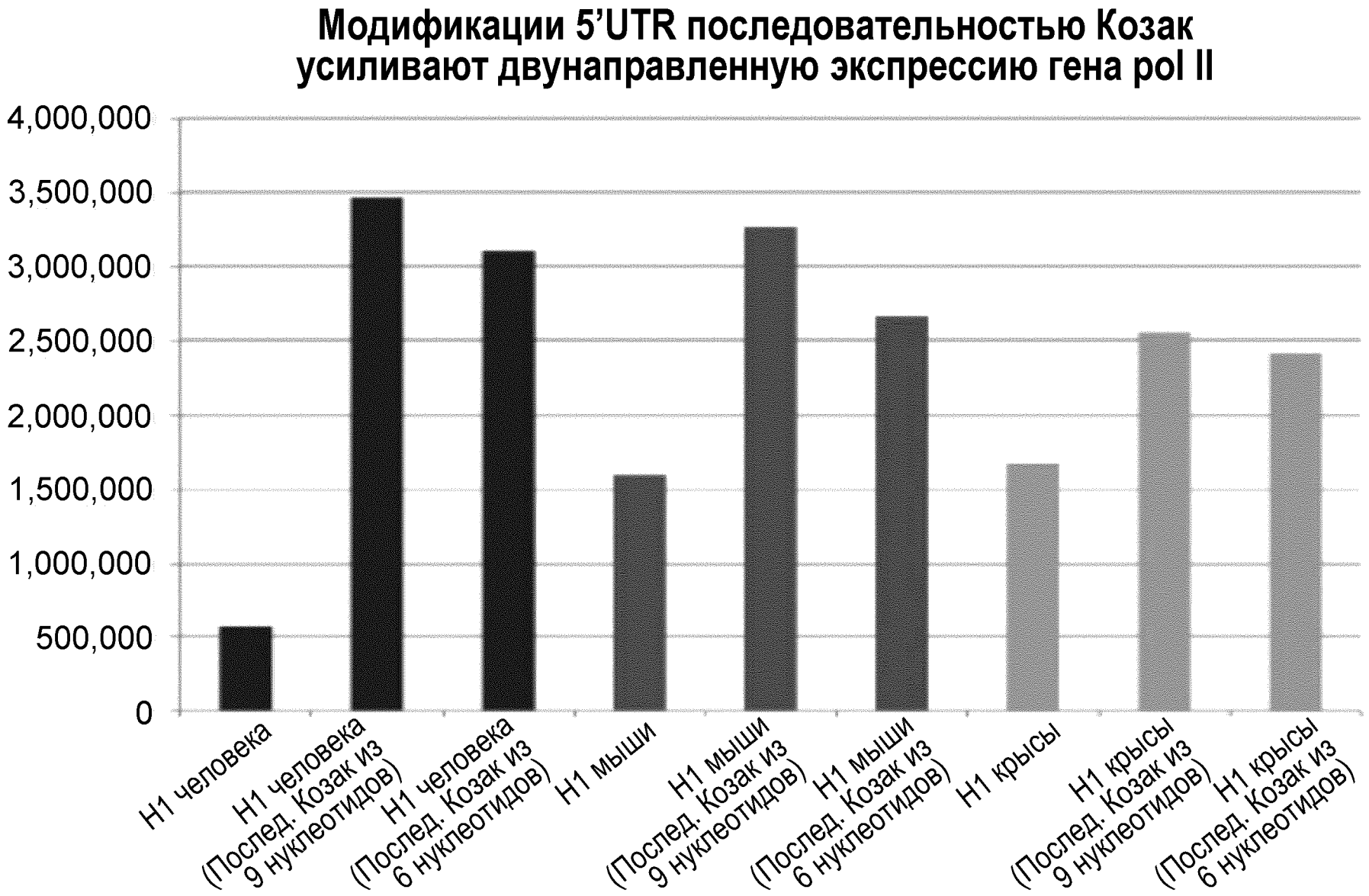


ФИГ.1D

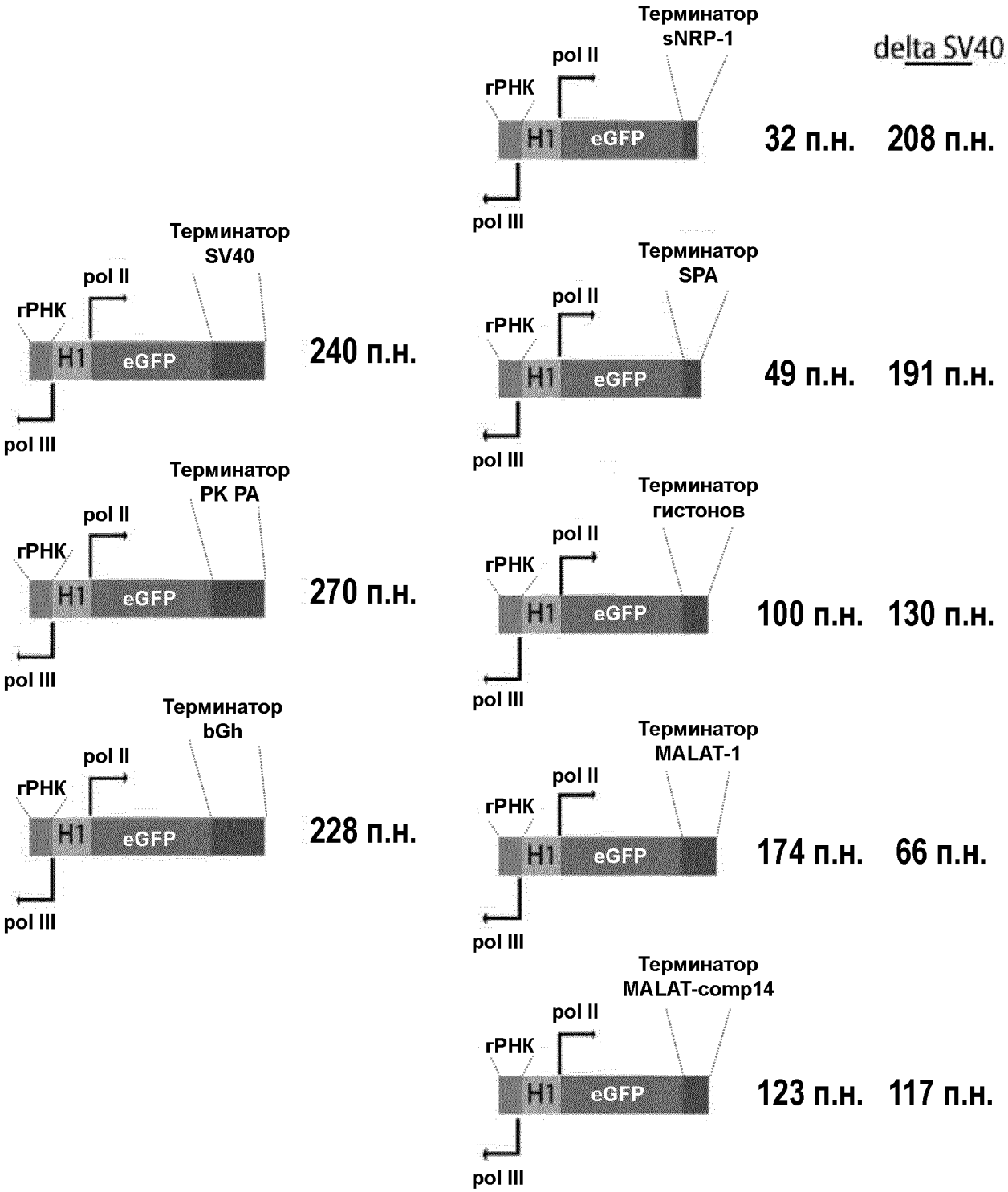
Последовательность Козак из 9 нуклеотидов: GCCGCCACC

Последовательность Козак из 6 нуклеотидов: GCCGACC

ФИГ.1E


Модификации 5'UTR последовательностью Козак усиливают двунаправленную экспрессию гена pol II
4,000,000
3,500,000
3,000,000
2,500,000
2,000,000
1,500,000
1,000,000
500,000
0
H1 человека
H1 человека (Послед. Козак из 9 нуклеотидов)
H1 человека (Послед. Козак из 6 нуклеотидов)
H1 мыши
H1 мыши (Послед. Козак из 9 нуклеотидов)
H1 мыши (Послед. Козак из 6 нуклеотидов)
H1 крысы
H1 крысы (Послед. Козак из 9 нуклеотидов)
H1 крысы (Послед. Козак из 6 нуклеотидов)

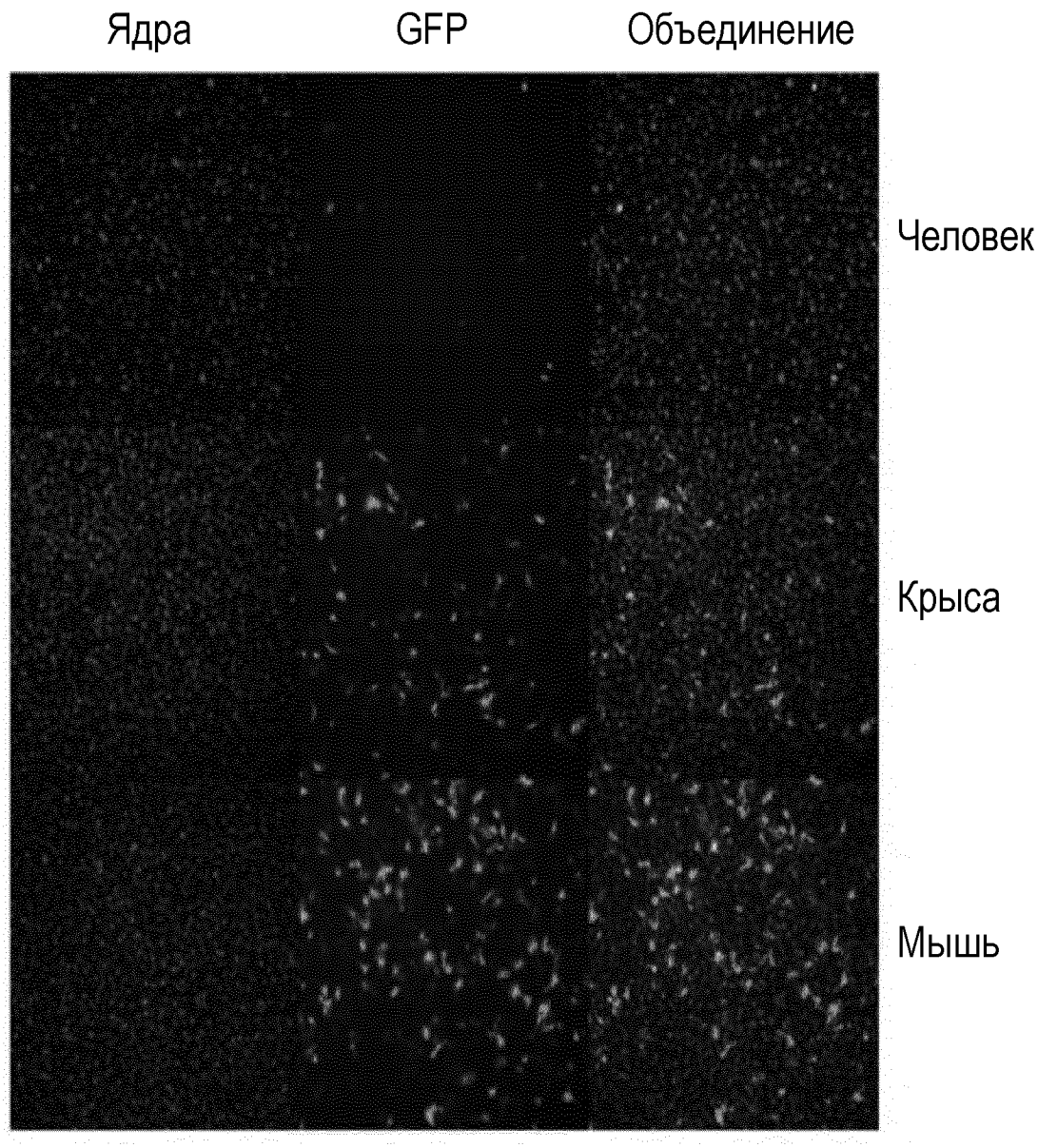
ФИГ.1F


Терминатор SV40
pol II
rРНК
H1
eGFP
pol III
240 п.н.
Терминатор PK PA
270 п.н.
Терминатор bGh
228 п.н.
Терминатор sNRP-1
delta SV40
32 п.н.
208 п.н.
Терминатор SPA
49 п.н.
191 п.н.
Терминатор гистонов
100 п.н.
130 п.н.
Терминатор MALAT-1
174 п.н.
66 п.н.
Терминатор MALAT-comp14
123 п.н.
117 п.н.

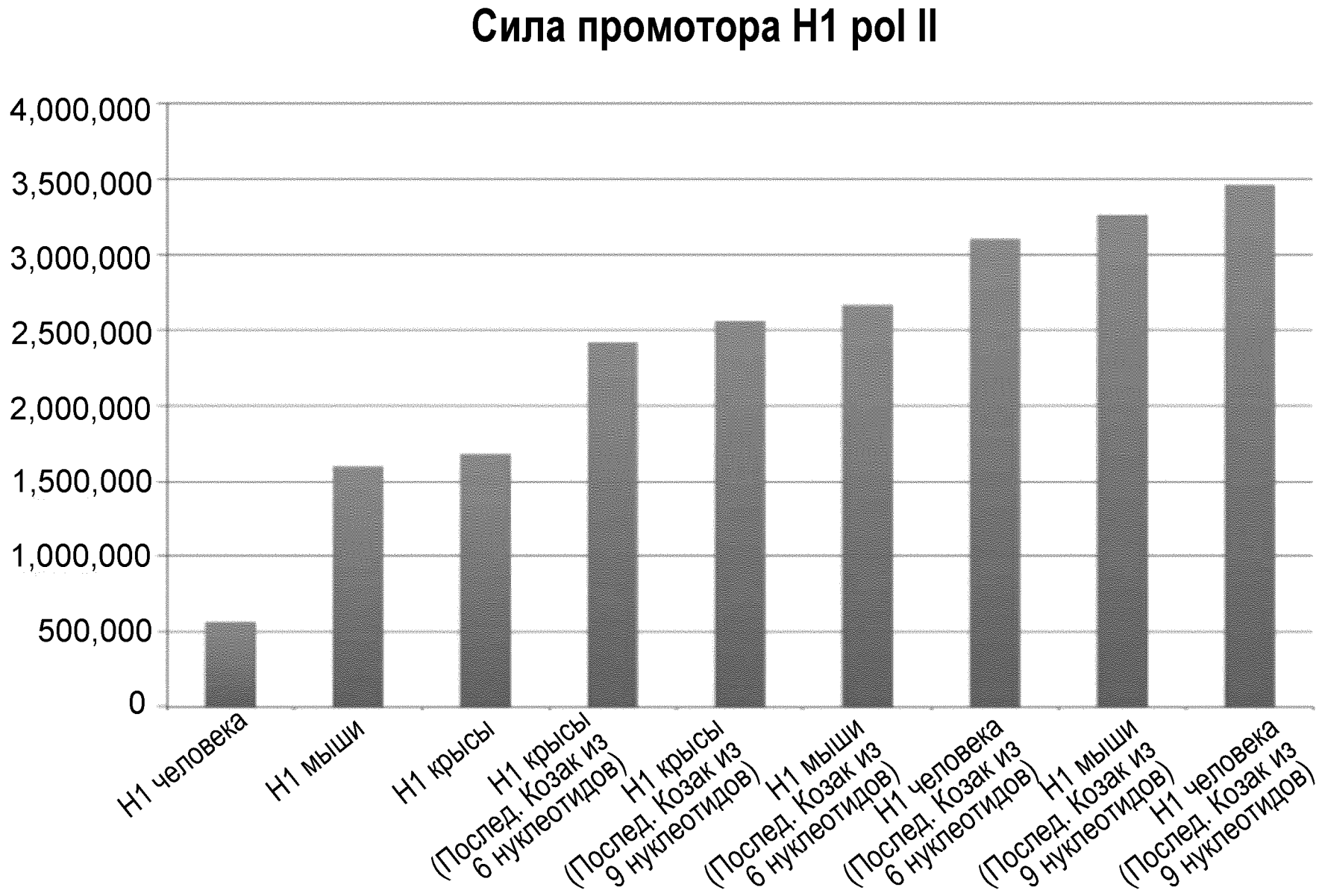
ФИГ.2А

Промотор H1	Человек	230 п.н.
Промотор H1	Мышь	176 п.н.
Промотор H1	Крыса	207 п.н.

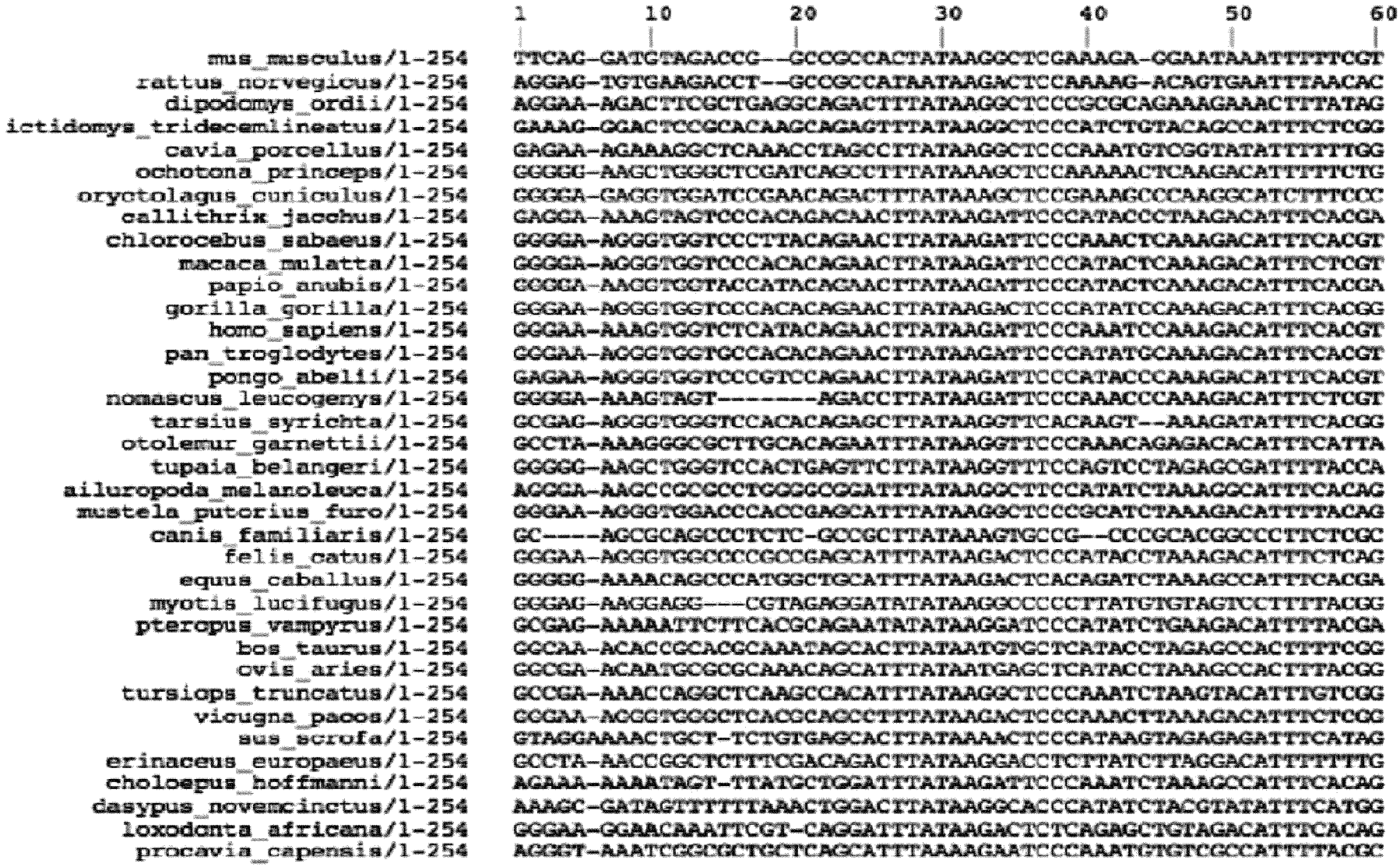
ФИГ.2В



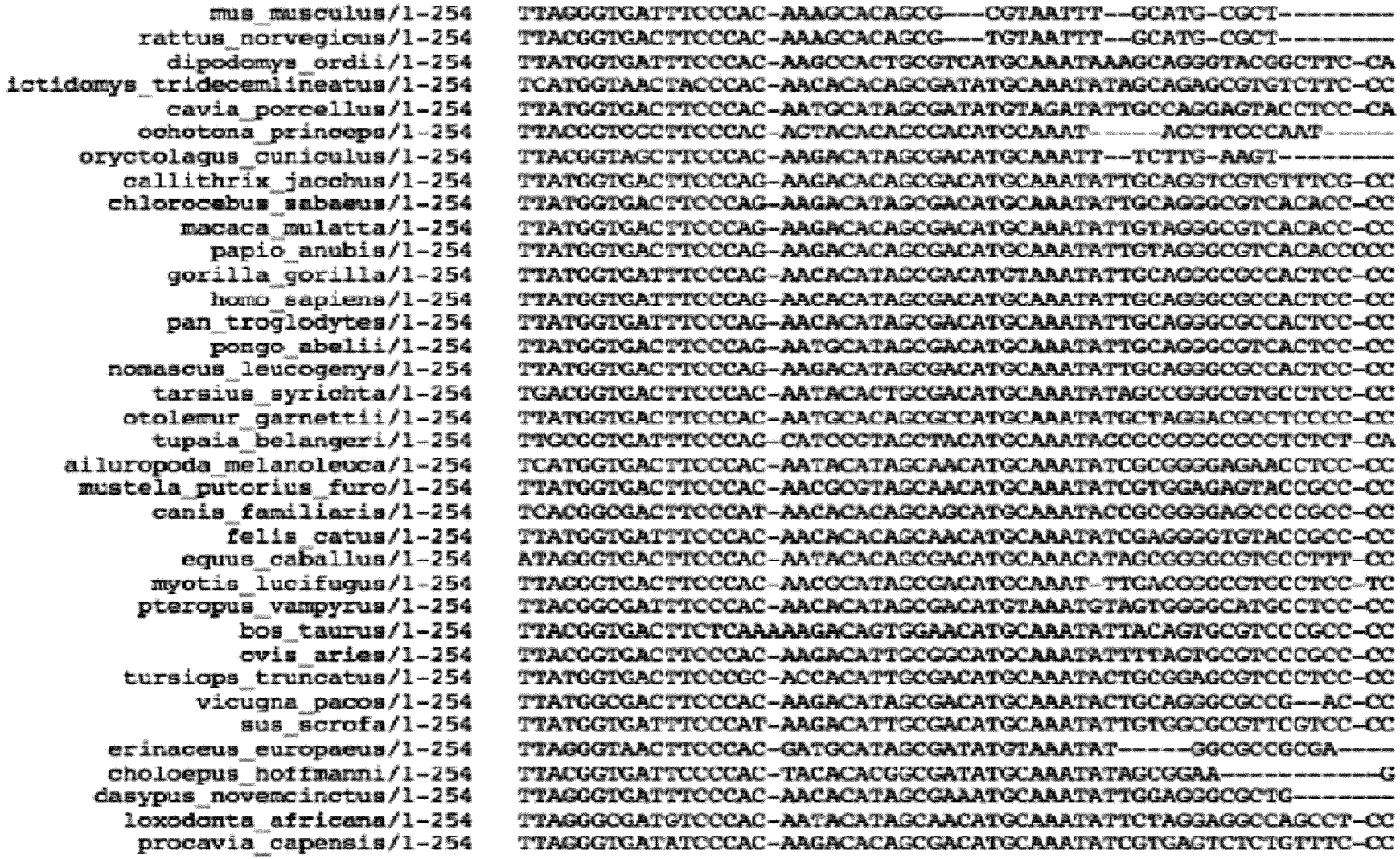
ФИГ.2С


Сила промотора H1 pol II
4,000,000
3,500,000
3,000,000
2,500,000
2,000,000
1,500,000
1,000,000
500,000
0
H1 человека
H1 мыши
H1 крысы
H1 крысы (Послед. Козак из 6 нуклеотидов)
H1 крысы (Послед. Козак из 9 нуклеотидов)
H1 мыши (Послед. Козак из 6 нуклеотидов)
H1 человека (Послед. Козак из 6 нуклеотидов)
H1 мыши (Послед. Козак из 9 нуклеотидов)
H1 человека (Послед. Козак из 9 нуклеотидов)

ФИГ.3А


1 10 20 30 40 50 60
mus_musculus/1-254 TTCAG-GATGTAGACCG--GCCGCCACTATAAGGCTCGAAAGA-GGAATAAATTTTTCGT
rattus_norvegicus/1-254 AGGAG-TGTGAAGACCT--GCCGCCATAATAAGACTCCAAAAG-ACAGTGAATTTAACAC
dipodomys_ordii/1-254 AGGAA-AGACTTCGCTGAGGCAGACTTTATAAGGCTCCCGCGCAGAAAGAAACTTTATAG
ictidomys_tridecemlineatus/1-254 GAAAG-GGACTCCGCACAAGCAGAGTTTATAAGGCTCCCATCTGTACAGCCATTTCTCGG
cavia_porcellus/1-254 GAGAA-AGAAAGGCTCAAACCTAGCCTTATAAGGCTCCCAAATGTCGGTATATTTTTTGG
ochotona_princeps/1-254 GGGGG-AAGCTGGGCTCGATCAGCCTTTATAAAGCTCCAAAAACTCAAGACATTTTTCTG
oryctolagus_cuniculus/1-254 GGGGA-GAGGTGGATCCGAACAGACTTTATAAAGCTCCGAAAGCCCAAGGCATCTTTCCC
callithrix_jacchus/1-254 GAGGA-AAAGTAGTCCCACAGACAACTTATAAGATTCCCATACCCTAAGACATTTCACGA
chlorocebus_sabaeus/1-254 GGGGA-AGGGTGGTCCCTTACAGAACTTATAAGATTCCCAAACTCAAAGACATTTCACGT
macaca_mulatta/1-254 GGGGA-AGGGTGGTCCCACACAGAACTTATAAGATTCCCATACTCAAAGACATTTCTCGT
papio_anubis/1-254 GGGGA-AAGGTGGTACCATACAGAACTTATAAGATTCCCATACTCAAAGACATTTCACGA
gorilla_gorilla/1-254 GGGAA-AGGGTGGTCCCACACAGAACTTATAAGACTCCCATATCCAAAGACATTTCACGG
homo_sapiens/1-254 GGGAA-AAAGTGGTCTCATACAGAACTTATAAGATTCCCAAATCCAAAGACATTTCACGT
pan_troglodytes/1-254 GGGAA-AGGGTGGTGCCACACAGAACTTATAAGATTCCCATATGCAAAGACATTTCACGT
pongo_abelii/1-254 GAGAA-AGGGTGGTCCCGTCCAGAACTTATAAGATTCCCATACCCAAAGACATTTCACGT
nomascus_leucogenys/1-254 GGGGA-AAAGTAGT--------AGACCTTATAAGATTCCCAAACCCAAAGACATTTCTCGT
tarsius_syrichta/1-254 GCGAG-AGGGTGGGTCCACACAGAGCTTATAAGGTTCACAAGT--AAAGATATTTCACGG
otolemur_garnettii/1-254 GCCTA-AAAGGGCGCTTGCACAGAATTTATAAGGTTCCCAAACAGAGACACATTTCATTA
tupaia_belangeri/1-254 GGGGG-AAGCTGGGTCCACTGAGTTCTTATAAGGTTTCCAGTCCTAGAGCGATTTTACCA
ailuropoda_melanoleuca/1-254 AGGGA-AAGCCGCGCCTGGGGCGGATTTATAAGGCTTCCATATCTAAAGGCATTTCACAG
mustela_putorius_furo/1-254 GGGAA-AGGGTGGACCCACCGAGCATTTATAAGGCTCCCGCATCTAAAGACATTTTACAG
canis_familiaris/1-254 GC-----AGCGCAGCCCTCTC-GCCGCTTATAAAGTGCCG--CCCGCACGGCCCTTCTCGC
felis_catus/1-254 GGGAA-AGGGTGGCCCCCGCCGAGCATTTATAAGACTCCCATACCTAAAGACATTTCTCAG
equus_caballus/1-254 GGGGG-AAAACAGCCCATGGCTGCATTTATAAGACTCACAGATCTAAAGCCATTTCACGA
myotis_lucifugus/1-254 GGGAG-AAGGAGG---CGTAGAGGATATATAAGGCCCCCTTATGTGTAGTCCTTTTACGG
pteropus_vampyrus/1-254 GCGAG-AAAAATTCTTCACGCAGAATATATAAGGATCCCATATCTGAAGACATTTTACGA
bos_taurus/1-254 GGCAA-ACACCGCACGCAAATAGCACTTATAATGTGCTCATACCTAGAGCCACTTTTCGG
ovis_aries/1-254 GGCGA-ACAATGCGCGCAAACAGCATTTATAATGAGCTCATACCTAAAGCCACTTTACGG
tursiops_truncatus/1-254 GCCGA-AAACCAGGCTCAAGCCACATTTATAAGGCTCCCAAATCTAAGTACATTTGTCGG
vicugna_pacos/1-254 GCGAA-ACGGTGGGCTCACGCAGCCTTTATAAGACTCCCAAACTTAAAGACATTTCTCGG
sus_scrofa/1-254 GTAGGAAAACTGCT-TCTGTGAGCACTTATAAAACTCCCATAAGTAGAGAGATTTCATAG
erinaceus_europaeus/1-254 GCCTA-AACCGGCTCTTTCGACAGACTTATAAGGACCTCTTATCTTAGGACATTTTTTTG
choloepus_hoffmanni/1-254 AGAAA-AAAATAGT-TTATGCTGGATTTATAAGATTCCCAAATCTAAAGCCATTTCACAG
dasypus_novemcinctus/1-254 AAAGC-GATAGTTTTTTAAACTGGACTTATAAGGCACCCATATCTACGTATATTTCATGG
loxodonta_africana/1-254 GGGAA-GGAACAAATTCGT-CAGGATTTATAAGACTCTCAGAGCTGTAGACATTTCACAG
procavia_capensis/1-254 AGGGT-AAATCGGCGCTGCTCAGCATTTAAAAGAATCCCAAATGTGTCGCCATTTTACGC

ФИГ.3А (продолжение)



ФИГ.3В

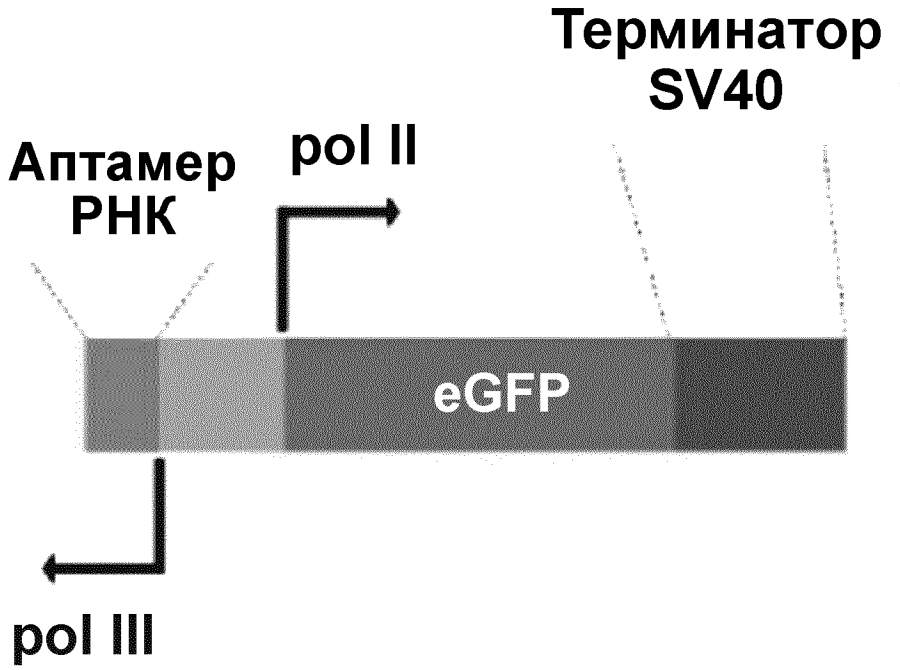
mus_musculus/1-254 -----CTACCCCAGGC----------TCCT----------GTGCTA---
rattus_norvegicus/1-254 -----CTAGCCCAGGC----------TCCA----------GCTCCG---
dipodomys_ordii/1-254 T----GTACCTTAAGG----TTTTTTTCTAGCCCGCGTACGC----TCTGCGTATT-CAG
ictidomys_tridecemlineatus/1-254 G----CGCGCGCCTGGTCGTCTC--------GGCGCCGGCGC----GCTGCGTGGGGCGG
cavia_porcellus/1-254 C----TTCTGGTCCTGTCAGCTCTTTTCTAGGACG--CGCGC----GCTGCAGGTTTCCA
ochotona_princeps/1-254 ------GAATTCGCGGACCGCTTCCCGCCCCGGCGCAGGCGC----GC-GGACGCTGTCT
oryctolagus_cuniculus/1-254 -----ATGCTTCAGACGC-GCTTCTCGCCACAGCGCAAGCGC----GCTGTGTGCTGACG
callithrix_jacchus/1-254 T----GTCCCTCACAGTCGTCTTCCTGCCAGCGCGCACGCGC----GCTGGGT-TTCCCG
chlorocebus_sabaeus/1-254 T----CTCCCTCACAGTCATCTTCCTGCCAGGGCGCACGCGC----GCTGGGTGTTCTCG
macaca_mulatta/1-254 T----GTCCCTCACAGTCATCTTCCTGCCAGGGCGCACGCGC----GCTGGGTGTTCCCG
papio_anubis/1-254 T----GTCCCTCACAGTCATCTTCCTGCCAGCGCGCACGCGC----GCTGGGTGTTCCCG
gorilla_gorilla/1-254 A----GTCCCTCACAGCCATCTTCCTGCCAGGGCGCACGCGC----GCTGGGTGTTCCCG
homo_sapiens/1-254 T----GTCCCTCACAGCCATCTTCCTGCCAGGGCGCACGCGC----GCTGGGTGTTCCCG
pan_troglodytes/1-254 T----GTCCCTCACTGCCATCTTCCTGCCAGGGCGCACGCGC----GCTGGGTGTTCCCG
pongo_abelii/1-254 T----GTCCCTCACAGCCATCTTCCTGCCAGGGCGCCCGCGC----GCTG-GTGTTCCCG
nomascus_leucogenys/1-254 T----GTCCCTCACAGCCATCTTCCTGCCAGGGCGCACGCGC----GCTGGGTGTTCCCG
tarsius_syrichta/1-254 G----ATCCCGGAAGAGCGACTCCTAGCCAGTGCGCACCGCC----GCTGCGTGTTCGCG
otolemur_garnettii/1-254 G----CTACCTTAAGGTCGTCAACTAACCAGTGCGCGCGCGC----ACTGCGCGTTTCCC
tupaia_belangeri/1-254 G----GTCCCTCCCCGCCCTCTCACTGT----ACGTACCCGC----GT----------
ailuropoda_melanoleuca/1-254 T----GTCCCTTGTACGCGGCTTC---TAAAGACGCACGCGC----GCTCTGTGTTCCCG
mustela_putorius_furo/1-254 T----GTCCCATGCACGCGTCTTCTCAGCAGCACGCACGCGC----GCT--GTGTTCCCG
canis_familiaris/1-254 GCCCCGGCCCCCGCA--------CCGCCTCGGGACGCATGCGCCGGCTCT--CCGTTCCCG
felis_catus/1-254 T----GTCCTTTGTAGACGTCTTCTCTCCAGGACGCACGCGC----GCT--GTATTCCCG
equus_caballus/1-254 T----GTACCCTGTGGGCA---TCTCTCCTGGACGCACGCGC----GCCGGGTGTTCCCG
myotis_lucifugus/1-254 T----GTCCCT-GCGGGCAACTTCTCTCCTGGACGCGCGCGC----GCTGCGTGTTCCCG
pteropus_vampyrus/1-254 T----GTCCCTTGTGGGCAGCTTCTCGCCAGAACGCACGCGC----GGTGCGTGTTCCCG
bos_taurus/1-254 ----------TGGTAGGTCTACGCTAGGACGCACGCGC----ACTACG-GTTCCCG
ovis_aries/1-254 ----------TGGTAGTTCCACGCTAGGACGCACACGC----ACTACG-GTTCCCG
tursiops_truncatus/1-254 T----GG----------CAACTCCTCGCTGGGACGCACGCGC----GCTACGTGCTCCCG
vicugna_pacos/1-254 G----GTCCTGTGCAGCCATCTTTCGGCTGGGACGCACGCGC----GCTGCGTGTTCCCG
sus_scrofa/1-254 -----GTCCGGTGCAGGCAGCTTCGCTCCAGGACGCACGCGC----AATACATGTTCCCG
erinaceus_europaeus/1-254 -----GTCTCTCCTAGGCGTCTCC---CCAGGACGCAGGCGC----ACTGCTTGTTCCCG
choloepus_hoffmanni/1-254 T----GTTCCTGACG---------CGTGCTAAAGCCGCCGCGC----GCTGACAGTTCCCG
dasypus_novemcinctus/1-254 ----------AGGCGTGGTC------GGGCGCAAGCGC----GCTGCGACTTCCCG
loxodonta_africana/1-254 C----CGTCCGCGTGGTCATCTTCTCGCTAGGGCGCACGCCC----GCTGCGTGTTCCCG
procavia_capensis/1-254 T----GTCCA-CGAGGGCGTCCTCTCGCTGGGGCGCACGCGC----GGTGTGTGTGCCCC

ФИГ.3В (продолжение)

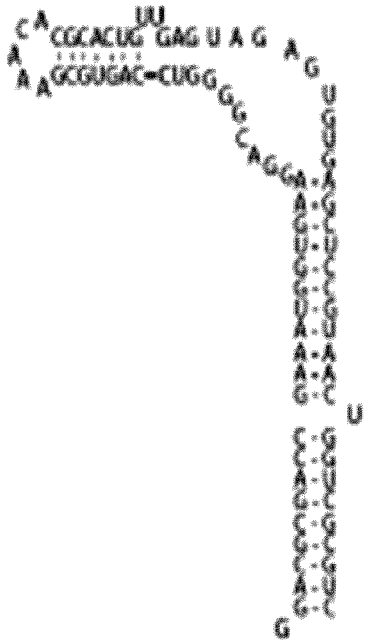
ФИГ.3В (продолжение)

mus_musculus/1-254	GTCCTTCAAGAGCG
rattus_norvegicus/1-254	GTCCTTCAAGCGCT
dipodomys_ordii/1-254	TGCGTTTGGATTCC
ictidomys_tridecemlineatus/1-254	AGTGTCTAACCTCC
cavia_porcellus/1-254	AGCGTTTGAATTCC
ochotona_princeps/1-254	--CGTTCGGTCTCC
oryctolagus_cuniculus/1-254	------AGATCTCC
callithrix_jacchus/1-254	AGCGTTTGAATTCC
chlorocebus_sabaeus/1-254	AGCGTTCGAATTCC
macaca_mulatta/1-254	AGCGTTCGAATTCC
papio_anubis/1-254	AGCGTTCGAATTCC
gorilla_gorilla/1-254	AGCGTTCGAATTCC
homo_sapiens/1-254	AGCGTTCGAATTCC
pan_troglodytes/1-254	AGCGTTCGAATTCC
pongo_abelii/1-254	AGCGCTCGTATTCC
nomascus_leucogenys/1-254	AGCGTTCGAATTCC
tarsius_syrichta/1-254	AGCGGCCCAGCTTC
otolemur_garnettii/1-254	AGTTCTGGAGGCTC
tupaia_belangeri/1-254	AGCGTTCCGACGCA
ailuropoda_melanoleuca/1-254	AACCTTCACCCTCC
mustela_putorius_furo/1-254	AACGTTCAGGCTTC
canis_familiaris/1-254	GACGAGCGGGCGCC
felis_catus/1-254	CAAGTTCTGGCTTC
equus_caballus/1-254	AGCATTCGGGCTCC
myotis_lucifugus/1-254	AACAGTCAGGCTCG
pteropus_vampyrus/1-254	AGCTCACCCGCTCC
bos_taurus/1-254	AGCGTTCGGGATCC
ovis_aries/1-254	AGCGTTGGGGCTCC
tursiops_truncatus/1-254	AGCGTTCTGGCTCC
vicugna_pacos/1-254	AACGTTCGGGTTCC
sus_scrofa/1-254	GATGTTAGGGATCC
erinaceus_europaeus/1-254	AGCGTCCAGTCTAC
choloepus_hoffmanni/1-254	AACCTTCCGCCTCC
dasypus_novemcinctus/1-254	GACGTTCGGATACC
loxodonta_africana/1-254	AGTGTTCTGGCTCC
procavia_capensis/1-254	AGACTTCCGGCTGC

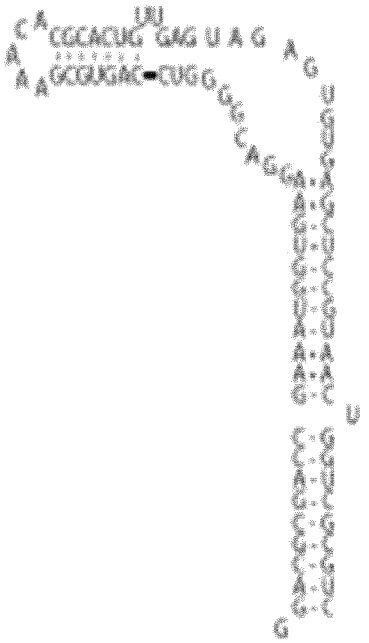
ФИГ.4А



ФИГ.4В



DFHB1-1T



Аптамер РНК Broccoli

Аптамер РНК Broccoli
+
DFHB1-1T

ФИГ.5

>mus_musculus/1-254
TTCAG-GATGTAGACCG--GCCGCCACTATAAGGCTCGAAAGA-GGAATAAATTTTTCGTTTAGGGTGATTTCCCAC-AA
AGCACAGCG---CGTAATTT--GCATG-CGCT----------CTACCCCAGGC----------TCCT--
----------GTGCTA---------GACAAG---A----AGC----CCGC-GCATC-CGGGCAA-GGGATGATGACGTC
GTCCTTCAAGAGCG
>rattus_norvegicus/1-254
AGGAG-TGTGAAGACCT--GCCGCCATAATAAGACTCCAAAAG-ACAGTGAATTTAACACTTACGGTGACTTCCCAC-AA
AGCACAGCG---TGTAATTT--GCATG-CGCT----------CTAGCCCAGGC----------TCCA--
----------GCTCCG---------GACCAG---A----AGC----CCGC-GCATC-CCGGCAA-AGGGTGATGACGTC
GTCCTTCAAGCGCT
>dipodomys_ordii/1-254
AGGAA-AGACTTCGCTGAGGCAGACTTTATAAGGCTCCCGCGCAGAAAGAAACTTTATAGTTATGGTGATTTCCCAC-AA
GCCACTGCGTCATGCAAATAAAGCAGGGTACGGCTTC-CAT----GTACCTTAAGG----TTTTTTTCTAGGCCGCGTAC
GC----TCTGCGTATT-CAGCCACGTGACCCT---G----AGC----CAGT-GGTTG-TTGGGAGCACGTTGTGGACCTC
TGCGTTTGGATTCC
>ictidomys_tridecemlineatus/1-254
GAAAG-GGACTCCGCACAAGCAGAGTTTATAAGGCTCCCATCTGTACAGCCATTTCTCGGTCATGGTAACTACCCAC-AA
CACACAGCGATATGCAAATATAGCAGAGCGTGTCTTC-CCG----CGCGCGCCTGGTCGTCTC--------GGCGCCGGC
GC----GCTGCGTGGGGCGGAACTGTGACAGA---G----ACC----CTGC-GATTC-CTGGGAGCTGGCTGATGACATC
AGTGTCTAACCTCC
>cavia_porcellus/1-254
GAGAA-AGAAAGGCTCAAACCTAGCCTTATAAGGCTCCCAAATGTCGGTATATTTTTTGGTTATGGTGACTTCCCAC-AA
TGCATAGCGATATGTAGATATTGCCAGGAGTACCTCC-CAC----TTCTGGTCCTGTCAGCTCTTTTCTAGGACG--CGC
GC----GCTGCAGGTTTCCAGCCTGTG--ATT---G----GGC----CAGC-AATTC-C-GGGAATGAATTGATGACGTC
AGCGTTTGAATTCC
>ochotona_princeps/1-254
GGGGG-AAGCTGGGCTCGATCAGCCTTTATAAAGCTCCAAAAACTCAAGACATTTTTCTGTTACGGTGGCTTCCCAC-AG
TACACAGCGACATGCAAAT-----AGCTTGCCAAT----------GAATTCGCGGACCGCTTCCCGCCCCGGCGCAGGC
GC----GC-GGACGCTGTCTCCCCTGGACGC----G----CGC----TCGC-GGTTC-CCGGGAGCTGGCTGATGA----
--CGTTCGGTCTCC
>oryctolagus_cuniculus/1-254
GGGGA-GAGGTGGATCCGAACAGACTTTATAAAGCTCCGAAAGCCCAAGGCATCTTTCCCTTACGGTAGCTTCCCAC-AA
GACATAGCGACATGCAAATT--TCTTG-AAGT----------ATGCTTCAGACGC-GCTTCTCGCCACAGCGCAAGC
GC----GCTGTGTGCTGACGC-----GGGAAC---G----GGCCAGGGCGC-GGTTC-CCGGGAGCGGGTTGATGACGTT
------AGATCTCC

ФИГ.5 (продолжение)

```
>callithrix_jacchus/1-254
GAGGA-AAAGTAGTCCCACAGACAACTTATAAGATTCCCATACCCTAAGACATTTCACGATTATGGTGACTTCCCAG-AA
GACACAGCGACATGCAAATATTGCAGGTCGTGTTTCG-CCT----GTCCCTCACAGTCGTCTTCCTGCCAGGGCGCACGC
GC----GCTGGGT-TTCCCGCCAACTGACGCT---G----GGC----TCGC-GATTC-CTTGGAGCGGGTTGATGACGTC
AGCGTTTGAATTCC
>chlorocebus_sabaeus/1-254
GGGGA-AGGGTGGTCCCTTACAGAACTTATAAGATTCCCAAACTCAAAGACATTTCACGTTTATGGTGACTTCCCAG-AA
GACATAGCGACATGCAAATATTGCAGGGCGTCACACC-CCT----CTCCCTCACAGTCATCTTCCTGCCAGGGCGCACGC
GC----GCTGGGTGTTCTCGCGTAGTGACACT---G----GGC----CCGC-GATTC-CTTGGAGCGGGTTGATGACGTC
AGCGTTCGAATTCC
>macaca_mulatta/1-254
GGGGA-AGGGTGGTCCCACACAGAACTTATAAGATTCCCATACTCAAAGACATTTCTCGTTTATGGTGACTTCCCAG-AA
GACACAGCGACATGCAAATATTGTAGGGCGTCACACC-CCT----GTCCCTCACAGTCATCTTCCTGCCAGGGCGCACGC
GC----GCTGGGTGTTCCCGCGTAGTGACACT---G----GGC----CCGC-GATTC-CTTGGAGCGGGTTGATGACGTC
AGCGTTCGAATTCC
>papio_anubis/1-254
GGGGA-AAGGTGGTACCATACAGAACTTATAAGATTCCCATACTCAAAGACATTTCACGATTATGGTGACTTCCCAG-AA
GACACAGCGACATGCAAATATTGTAGGGCGTCACACCCCCT----GTCCCTCACAGTCATCTTCCTGCCAGGGCGCACGC
GC----GCTGGGTGTTCCCGCGTAGTGACACT---G----GGC----CCGC-GATTC-CTTGGAGCGGGTTGATGACGTC
AGCGTTCGAATTCC
>gorilla_gorilla/1-254
GGGAA-AGGGTGGTCCCACACAGAACTTATAAGACTCCCATATCCAAAGACATTTCACGGTTATGGTGATTTCCCAG-AA
CACATAGCGACATGTAAATATTGCAGGGCGCCACTCC-CCA----GTCCCTCACAGCCATCTTCCTGCCAGGGCGCACGC
GC----GCTGGGTGTTCCCGCCTAGTGACACT---G----GGC----CCGC-GATTC-CTTGGAGCGGGTTGATGACGTC
AGCGTTCGAATTCC
>homo_sapiens/1-254
GGGAA-AAAGTGGTCTCATACAGAACTTATAAGATTCCCAAATCCAAAGACATTTCACGTTTATGGTGATTTCCCAG-AA
CACATAGCGACATGCAAATATTGCAGGGCGCCACTCC-CCT----GTCCCTCACAGCCATCTTCCTGCCAGGGCGCACGC
GC----GCTGGGTGTTCCCGCCTAGTGACACT---G----GGC----CCGC-GATTC-CTTGGAGCGGGTTGATGACGTC
AGCGTTCGAATTCC
>pan_troglodytes/1-254
GGGAA-AGGGTGGTGCCACACAGAACTTATAAGATTCCCATATGCAAAGACATTTCACGTTTATGGTGATTTCCCAG-AA
CACATAGCGACATGCAAATATTGCAGGGCGCCACTCC-CCT----GTCCCTCACTGCCATCTTCCTGCCAGGGCGCACGC
GC----GCTGGGTGTTCCCGCCTAGTGACACT---G----GGC----CCGC-GATTC-CTTGGAGCGGGTTGATGACGTC
AGCGTTCGAATTCC
```

ФИГ.5 (продолжение)

```
>pongo_abelii/1-254
GAGAA-AGGGTGGTCCCGTCCAGAACTTATAAGATTCCCATACCCAAAGACATTTCACGTTTATGGTGACTTCCCAG-AA
TGCATAGCGACATGCAAATATTGCAGGGCGTCACTCC-CCT----GTCCCTCACAGCCATCTTCCTGCCAGGGCGCCCGC
GC----GCTG-GTGTTCCCGCCTAGTGACACT---G----GGC----CCAC-GATTC-CTTGGAGCGGGTTGATGACGTC
AGCGCTCGTATTCC
>nomascus_leucogenys/1-254
GGGGA-AAAGTAGT-------AGACCTTATAAGATTCCCAAACCCAAAGACATTTCTCGTTTATGGTGACTTCCCAG-AA
GACATAGCGACATGCAAATATTGCAGGGCGCCACTCC-CCT----GTCCCTCACAGCCATCTTCCTGCCAGGGCGCACGC
GC----GCTGGGTGTTCCCGCCTAGTGACACT---C----GGC----CCGC-GATTC-CTTGGAGCGGGTTGATGACGTC
AGCGTTCGAATTCC
>tarsius_syrichta/1-254
GCGAG-AGGGTGGGTCCACACAGAGCTTATAAGGTTCACAAGT--AAAGATATTTCACGGTGACGGTGACTTCCCAC-AA
TACACTGCGACATGCAAATATAGCCGGGCGTGCCTCC-CCG----ATCCCGGAAGAGCGACTCCTAGCCAGTGCGCACGC
GC----GCTGCGTGTTCGCGTCCTAGGTCGCT---G----GGC----CCGC-GGTTC-CTGGGAGCGGG-TGGTGACGTC
AGCGGCCCAGCTTC
>otolemur_garnettii/1-254
GCCTA-AAAGGGCGCTTGCACAGAATTTATAAGGTTCCCAAACAGAGACACATTTCATTATTATGGTGACTTCCCAC-AA
TGCACAGCGCCATGCAAATATGCTAGGACGCCTCCCC-CCG----CTACCTTAAGGTCGTCAACTAACCAGTGCGCGCGC
GC----ACTGCGCGTTTCCCGCCGGTGACTCA---A----TGC----CCGC-GTTTG-GTGGGAGCTAGTTGGTGACCTC
AGTTCTGGAGGCTC
>tupaia_belangeri/1-254
GGGGG-AAGCTGGGTCCACTGAGTTCTTATAAGGTTTCCAGTCCTAGAGCGATTTTACCATTGCGGTGATTTCCCAG-CA
TCCGTAGCTACATGCAAATAGCGCGGGGCGCGTCTCT-CAG----GTCCCTCCCCGCCCTCTCACTGT----ACGTACCC
GC----GT------------CCTAGGGACGCC---G----CGC----CCGG-GGTTC-CC-------------GGACGTC
AGCGTTCCGACGCA
>ailuropoda_melanoleuca/1-254
AGGGA-AAGCCGCGCCTGGGGCGGATTTATAAGGCTTCCATATCTAAAGGCATTTCACAGTCATGGTGACTTCCCAC-AA
TACATAGCAACATGCAAATATCGCGGGGAGAACCTCC-CCT----GTCCCTTGTACGCGGCTTC---TAAAGACGCACGC
GC----GCTCTGTGTTCCCGCCCTGTGACTCT---A----GGC----GGGC-AATTC-CTGGGACAGTGTT-CTGACGGG
AACGTTCAGGCTCC
>mustela_putorius_furo/1-254
GGGAA-AGGGTGGACCCACCGAGCATTTATAAGGCTCCCGCATCTAAAGACATTTTACAGTTATGGTGACTTCCCAC-AA
CGCGTAGCAACATGCAAATATCGTGGAGAGTACCGCC-CCT----GTCCCATGCACGCGTCTTCTCAGCAGCACGCACGC
GC----GCT--GTGTTCCCGCCCTGTGACTCC---A----GGC----GGGT--ATTT-CCAGGGGCGGGTTGCTGACAGG
AACGTTCAGGCTTC
```

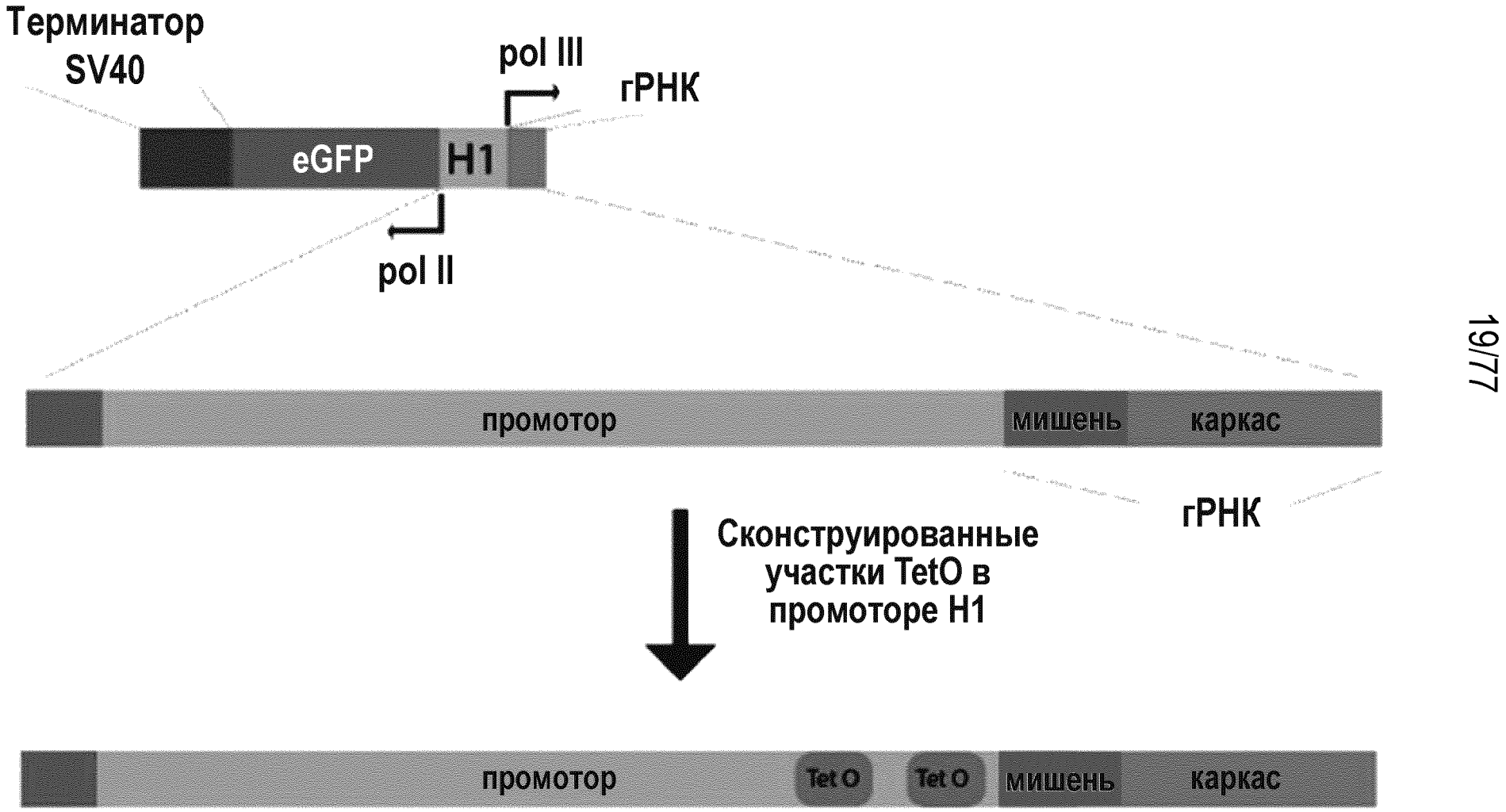
ФИГ.5 (продолжение)

```
>canis_familiaris/1-254
GC----AGCGCAGCCCTCTC-GCCGCTTATAAAGTGCCG--CCCGCACGGCCCTTCTCGCTCACGGCGACTTCCCAT-AA
CACACAGCAGCATGCAAATACCGCGGGGAGCCCCGCC-CCGCCCCGGCCCCCGCA-------CCGCCTCGGGACGCATGC
GCCGGCTCT--CCGTTCCCGCCTTGGGCCGGCGGCGGGCGGGC----GGGC-GAGCGGGCGGGAGCGGCTC--CGGCGGG
GACGAGCGGGCGCC
>felis_catus/1-254
GGGAA-AGGGTGGCCCCGCCGAGCATTTATAAGACTCCCATACCTAAAGACATTTCTCAGTTATGGTGATTTCCCAC-AA
CACACAGCAACATGCAAATATCGAGGGGTGTACCGCC-CCT----GTCCTTTGTAGACGTCTTCTCTCCAGGACGCACGC
GC----GCT--GTATTCCCGCCTTGTGACTCT---A----GGC----GGGC-GATTC-CTGGGAGAGGGTTGATGACGTC
CAAGTTCTGGCTTC
>equus_caballus/1-254
GGGGG-AAAACAGCCCATGGCTGCATTTATAAGACTCACAGATCTAAAGCCATTTCACGAATAGGGTGACTTCCCAC-AA
TACACAGCGACATGCAAACATAGCGGGGCGTGCCTTT-CCT----GTACCCTGTGGGCA---TCTCTCCTGGACGCACGC
GC----GCCGGGTGTTCCCGCGCTGTGACTCT---A----GGC----AAGC-GCTTC-CTGGGAGAGAGTTGATGACGGC
AGCATTCGGGCTCC
>myotis_lucifugus/1-254
GGGAG-AAGGAGG---CGTAGAGGATATATAAGGCCCCCTTATGTGTAGTCCTTTTACGGTTAGGGTGACTTCCCAC-AA
CGCATAGCGACATGCAAAT-TTGACGGGCGTGCCTCC-TCT----GTCCCT-GCGGGCAACTTCTCTCCTGGACGCGCGC
GC----GCTGCGTGTTCCCGCCTTTTGACTCC---A----GCC----GAGC-GAATC-CTGGGAGAGGGCAGGTGACGTC
AACAGTCAGGCTCG
>pteropus_vampyrus/1-254
GCGAG-AAAAATTCTTCACGCAGAATATATAAGGATCCCATATCTGAAGACATTTTACGATTACGGCGATTTCCCAC-AA
CACATAGCGACATGTAAATGTAGTGGGGCATGCCTCC-CCT----GTCCCTTGTGGGCAGCTTCTCGCCAGAACGCACGC
GC----GGTGCGTGTTCCCGCCTTGTG--ACT---A----AGT----TGGC-GAGTC-A-GGGAGGAGATTGATGACGTC
AGCTCACCCGCTCC
>bos_taurus/1-254
GGCAA-ACACCGCACGCAAATAGCACTTATAATGTGCTCATACCTAGAGCCACTTTTCGGTTACGGTGACTTCTCAAAAA
GACAGTGGAACATGCAAATATTACAGTGCGTCCCGCC-CC--------------TGGTAGGTCTACGCTAGGACGCACGC
GC----ACTACG-GTTCCCGCCTATAGACT-----G----CGC----TGGC-GATTC-CTGGGAGCGGACTGATGACGTC
AGCGTTCGGGATCC
>ovis_aries/1-254
GGCGA-ACAATGCGCGCAAACAGCATTTATAATGAGCTCATACCTAAAGCCACTTTACGGTTACGGTGACTTCCCAC-AA
GACATTGCGGCATGCAAATATTTTAGTGCGTCCCGCC-CC--------------TGGTAGTTCCACGCTAGGACGCACAC
GC----ACTACG-GTTCCCGCCTTTAGACT-----G----CGC----TGGC-GATTC-CA-GGAGCGGACTGATGACGTC
AGCGTTGGGGCTCC
>tursiops_truncatus/1-254
GCCGA-AAACCAGGCTCAAGCCACATTTATAAGGCTCCCAAATCTAAGTACATTTGTCGGTTATGGTGACTTCCCGC-AC
CACATTGCGACATGCAAATACTGCGGAGCGTCCCTCC-CCT----GG----------CAACTCCTCGCTGGGACGCACGC
```

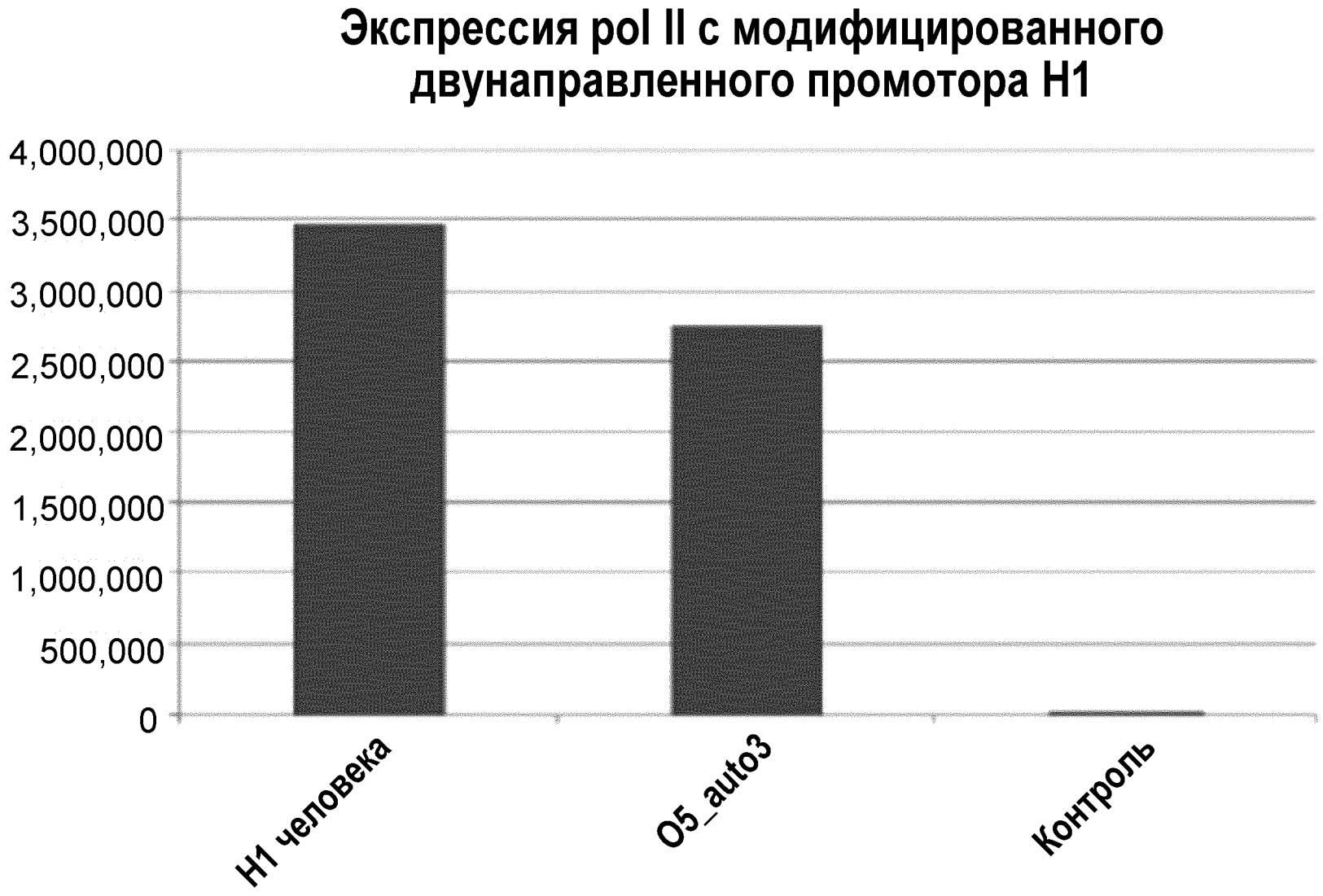
ФИГ.5 (продолжение)

```
GC----GCTACGTGCTCCCGCCTTTTG--ACT---G----CGC----CGGC-GATAC-TTGGGAGAGGGTTGATGACGTC
AGCGTTCTGGCTCC
>vicugna_pacos/1-254
GGGAA-AGGGTGGGCTCACGCAGCCTTTATAAGACTCCCAAACTTAAAGACATTTCTCGGTTATGGCGACTTCCCAC-AA
GACATAGCGACATGCAAATACTGCAGGGCGCCG--AC-CCG----GTCCTGTGCAGCCATCTTTCGGCTGGGACGCACGC
GC----GCTGCGTGTTCCCGCCCTGTG--ACT---G----CGC----CGGC-GATTA-CTGGGAGAGGATTGATGACGTC
AACGTTCGGGTTCC
>sus_scrofa/1-254
GTAGGAAAACTGCT-TCTGTGAGCACTTATAAAACTCCCATAAGTAGAGAGATTTCATAGTTATGGTGATTTCCCAT-AA
GACATTGCGACATGCAAATATTGTGGCGCGTTCGTCC-CC-----GTCCGGTGCAGGCAGCTTCGCTCCAGGACGCACGC
GC----AATACATGTTCCCGCCTTGAGACT-----G----CGC----CGGCAGATTC-CTAGGAAGTGGTTGATGACGTC
GATGTTAGGGATCC
>erinaceus_europaeus/1-254
GCCTA-AACCGGCTCTTTCGACAGACTTATAAGGACCTCTTATCTTAGGACATTTTTTTGTTAGGGTAACTTCCCAC-GA
TGCATAGCGATATGTAAATAT-----GGCGCCGCGA---------GTCTCTCCTAGGCGTCTCC---CCAGGACGCAGGC
GC----ACTGCTTGTTCCCGCGT--TAACATT----------------GC-TGATT-CTGGGAGACTGCTGATGACGTC
AGCGTCCAGTCTAC
>choloepus_hoffmanni/1-254
AGAAA-AAAATAGT-TTATGCTGGATTTATAAGATTCCCAAATCTAAAGCCATTTCACAGTTACGGTGATTCCCCAC-TA
CACACGGCGATATGCAAATATAGCGGAA----------GT----GTTCCTGAGG--------CGTGGTAAAGCGCGCGC
GC----GCTGAGAGTTCCCGCCCTGTGGTGCT---G----GGC----TGGA-GATGC-CTGAGAACTGGCTGATGACGGC
AACGTTCGGGCTCC
>dasypus_novemcinctus/1-254
AAAGC-GATAGTTTTTTAAACTGGACTTATAAGGCACCCATATCTACGTATATTTCATGGTTAGGGTGATTTCCCAC-AA
CACATAGCGAAATGCAAATATTGGAGGGCGCTG----------------------AGGCGTGGTC------GGGCGCAAGC
GC----GCTGCGACTTCCCGCCTTTCGGCCCT---A----GGC----CCCA-GATTC-CTGGGAGCTGGATGATGACGTT
GACGTTCGGATACC
>loxodonta_africana/1-254
GGGAA-GGAACAAATTCGT-CAGGATTTATAAGACTCTCAGAGCTGTAGACATTTCACAGTTAGGGCGATGTCCCAC-AA
TACATAGCAACATGCAAATATTCTAGGAGGCCAGCCT-CCC----CGTCCGCGTGGTCATCTTCTCGCTAGGGCGCACGC
CC----GCTGCGTGTTCCCGCTCTGTG--ACC---A----GGC----AGGC-GATTC-CTGAGAACCGCTTGGTGACGTC
AGTGTTCTGGCTCC
>procavia_capensis/1-254
AGGGT-AAATCGGCGCTGCTCAGCATTTAAAAGAATCCCAAATGTGTCGCCATTTTACGCTTAGGGTGATATCCCAC-AA
GACACAGCGACATGCAAATATCGTGAGTCTCTGTTTC-CCT----GTCCA-CGAGGGCGTCCTCTCGCTGGGGCGCACGC
GC----GGTGTGTGTGCCCCCGTTGTG----T---G----TTC----CCGC-GATTC-CAAAGAACTGGTTGATAACGTT
AGACTTCCGGCTGC
```

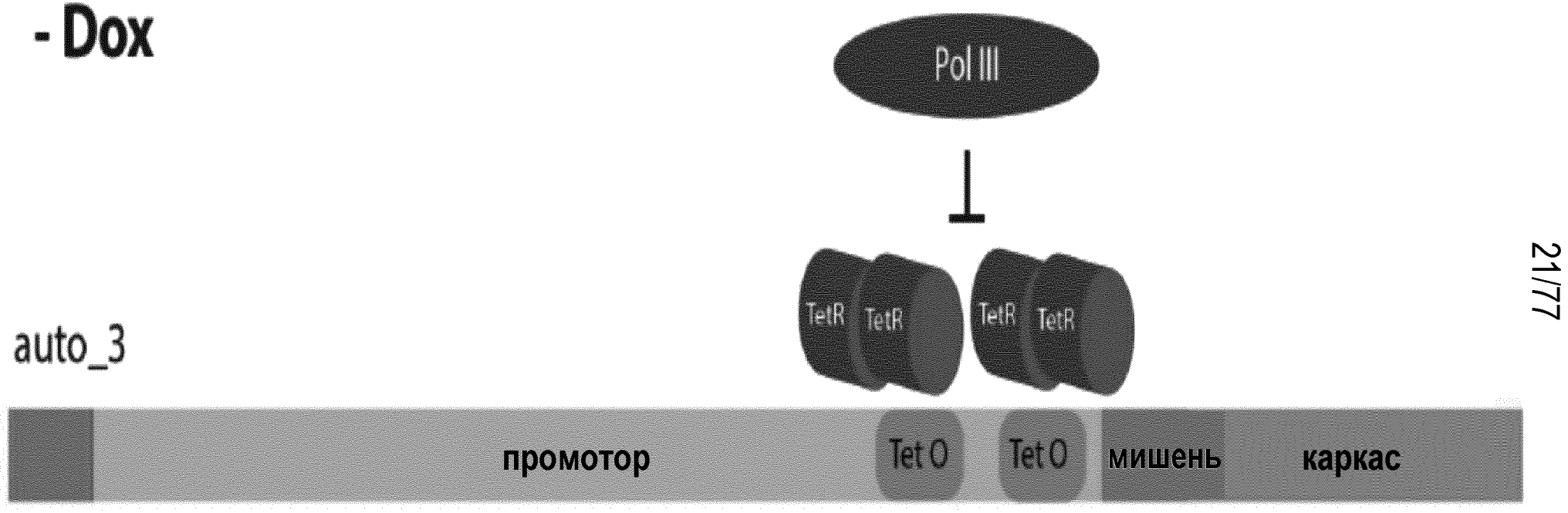
ФИГ.6А


Терминатор
SV40
pol III
гРНК
eGFP
H1
pol II
промотор
мишень
каркас
гРНК
Сконструированные
участки TetO в
промоторе H1
промотор
Tet O
Tet O
мишень
каркас

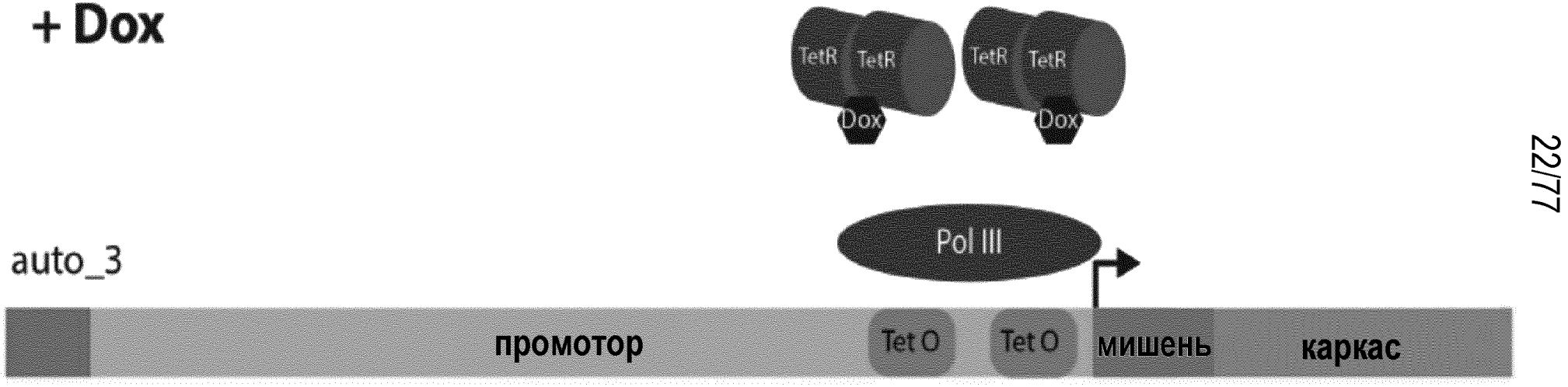
ФИГ.6В


Экспрессия pol II с модифицированного двунаправленного промотора H1
4,000,000
3,500,000
3,000,000
2,500,000
2,000,000
1,500,000
1,000,000
500,000
0
H1 человека
O5_auto3
Контроль

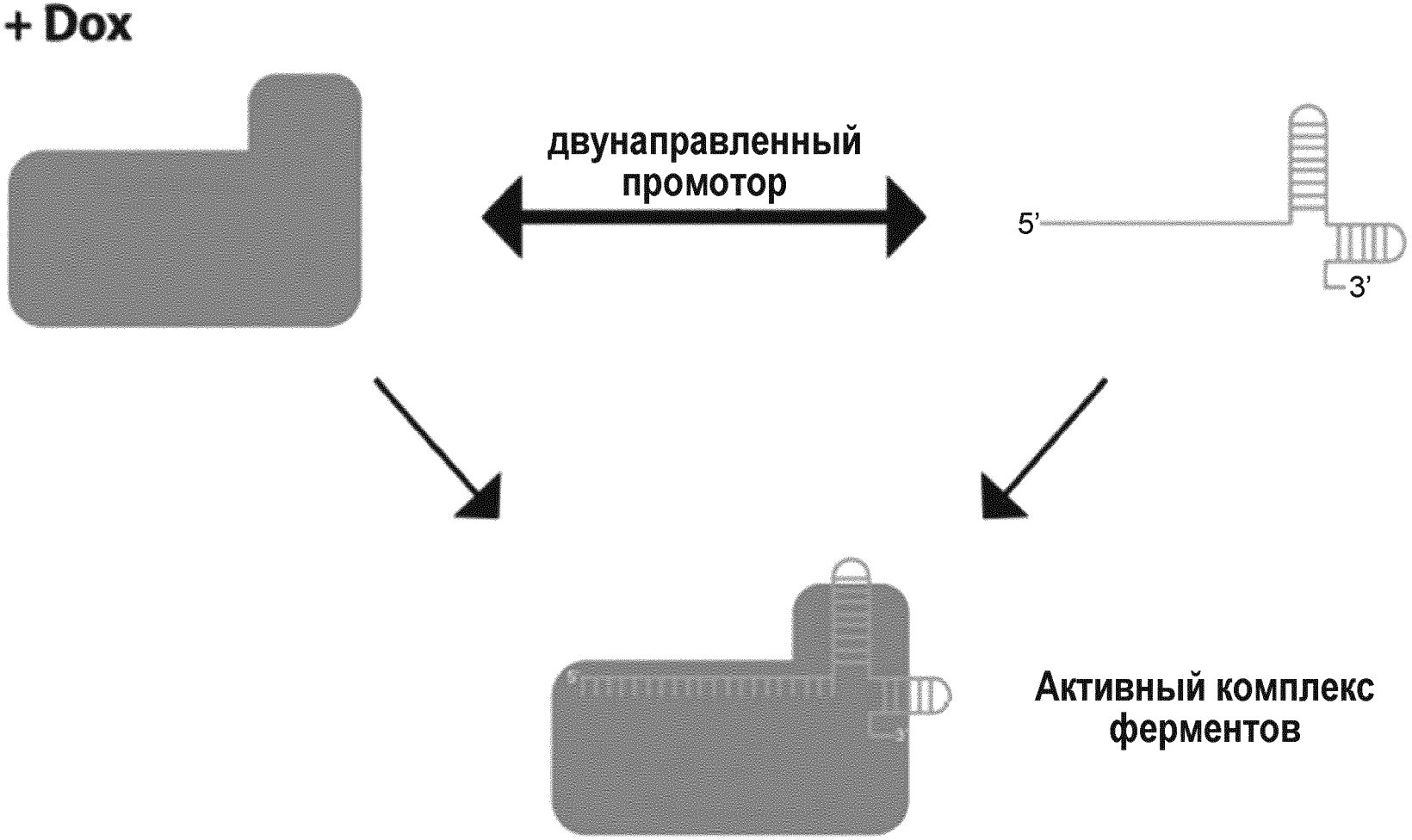
ФИГ.7А


- Dox
Pol III
TetR
TetR
TetR
TetR
auto_3
промотор
Tet O
Tet O
мишень
каркас

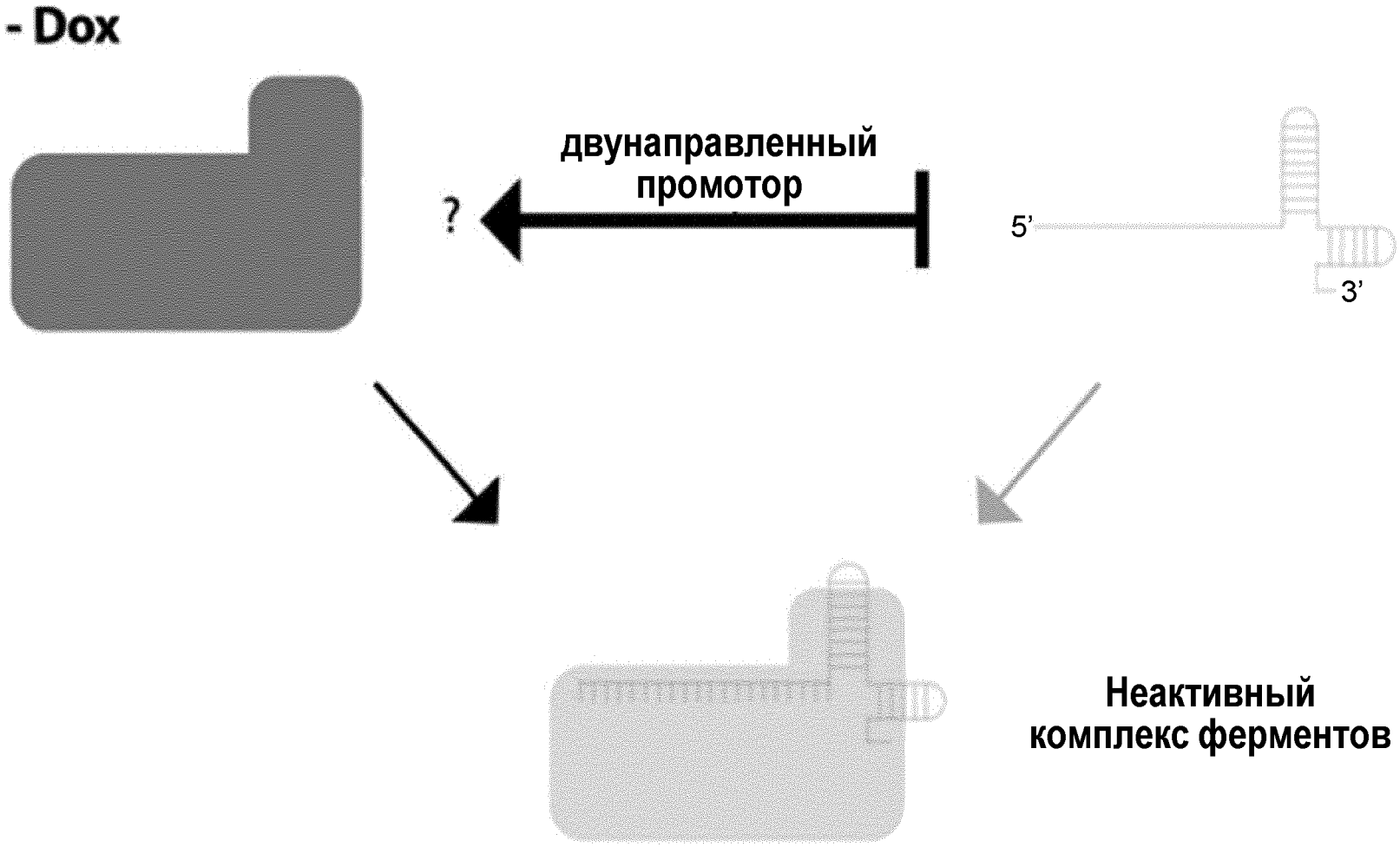
ФИГ.7В


+ Dox
TetR
TetR
Dox
TetR
TetR
Dox
Pol III
auto_3
промотор
Tet O
Tet O
мишень
каркас

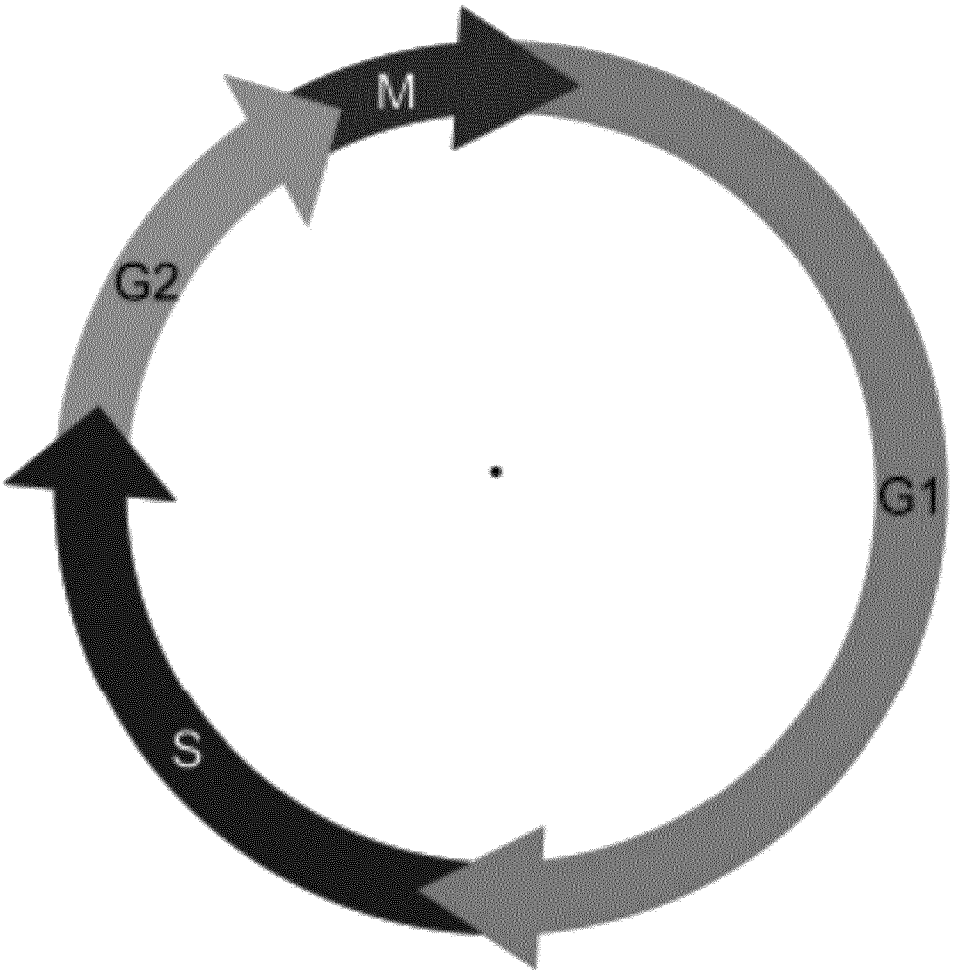
ФИГ.8А


+ Dox
двунаправленный промотор
5’
3’
Активный комплекс ферментов

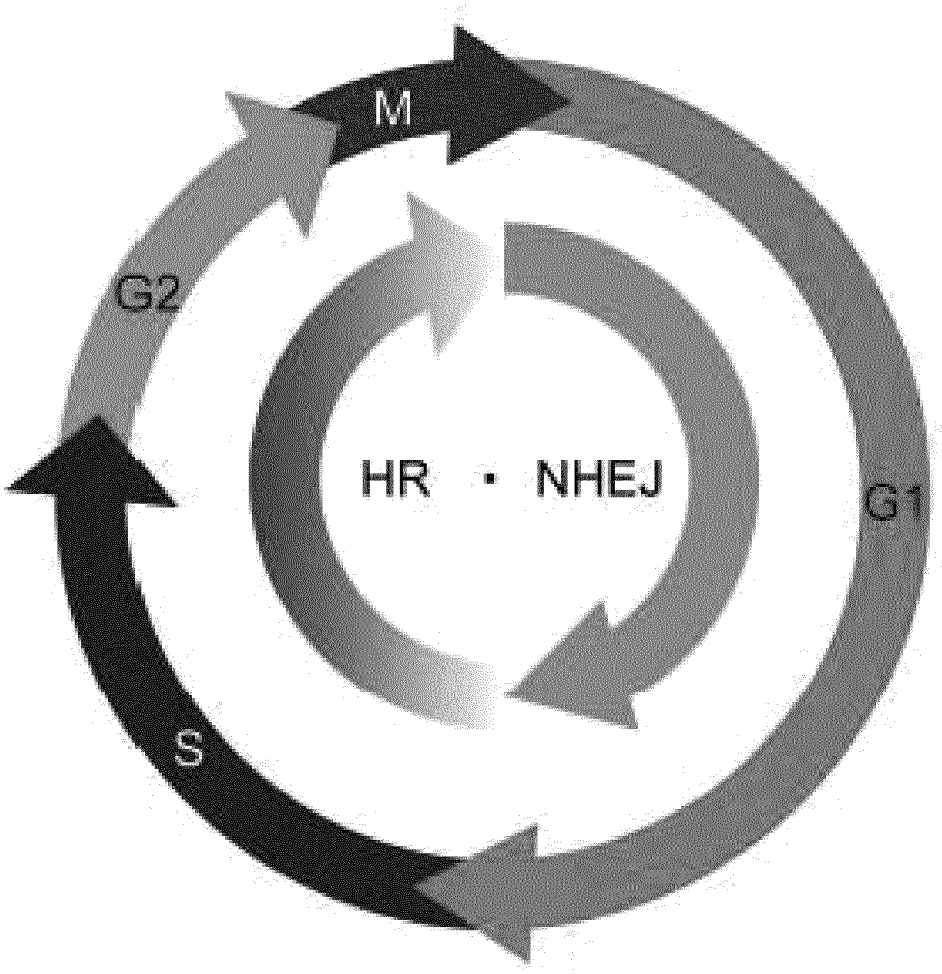
ФИГ.8B


- Dox
двунаправленный
промотор
?
5’
3’
Неактивный
комплекс ферментов

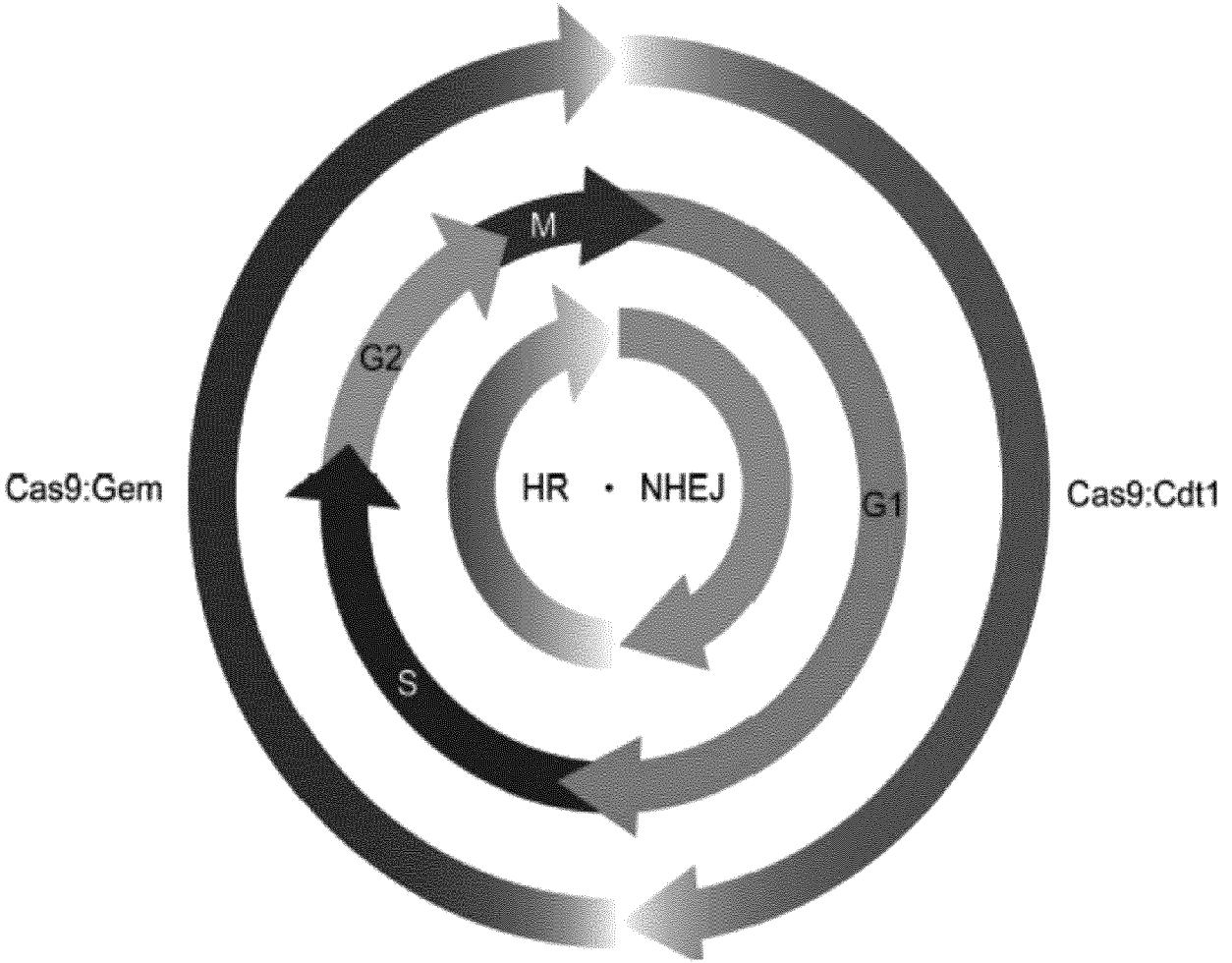
ФИГ.9А



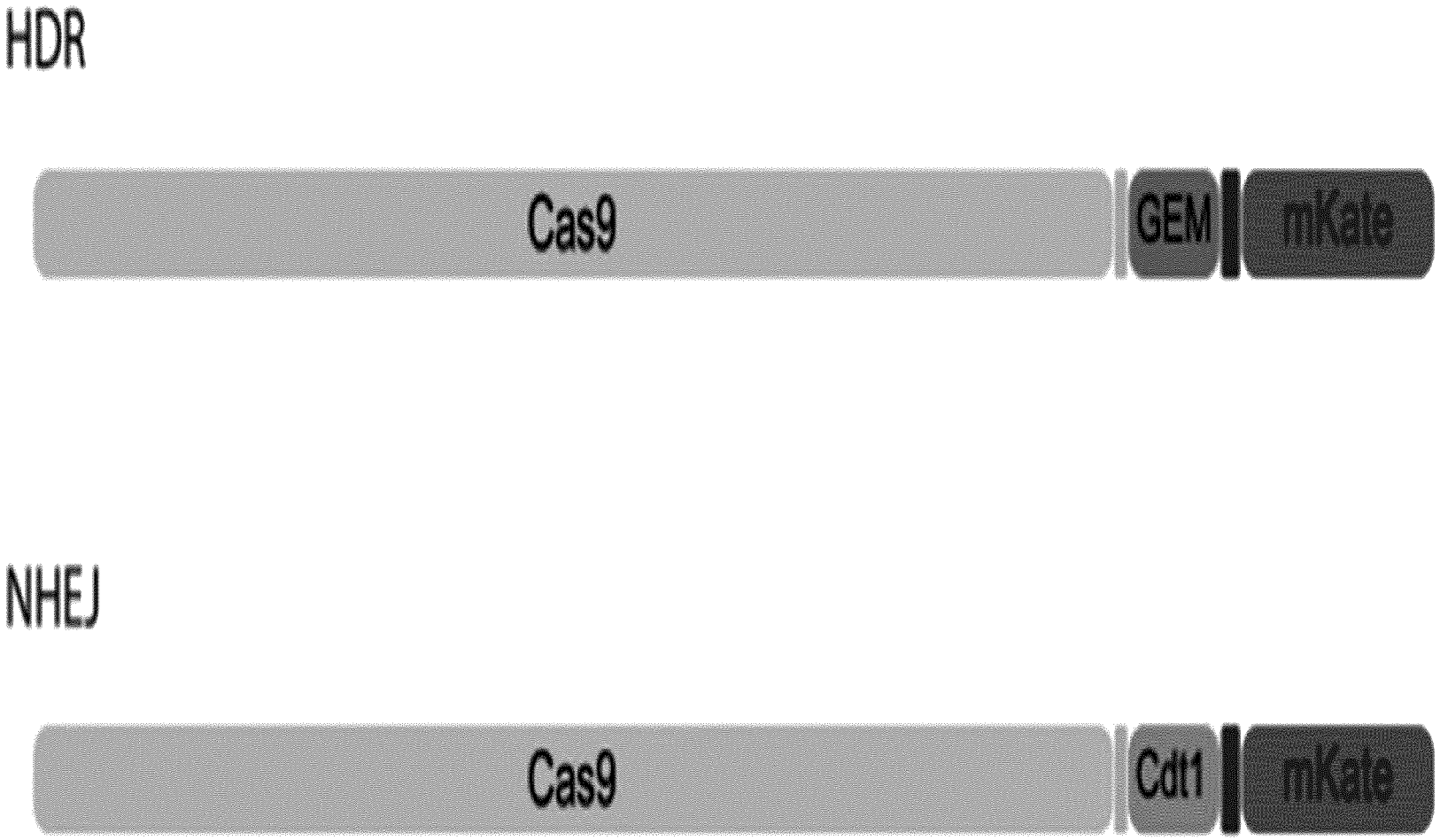
ФИГ.9В



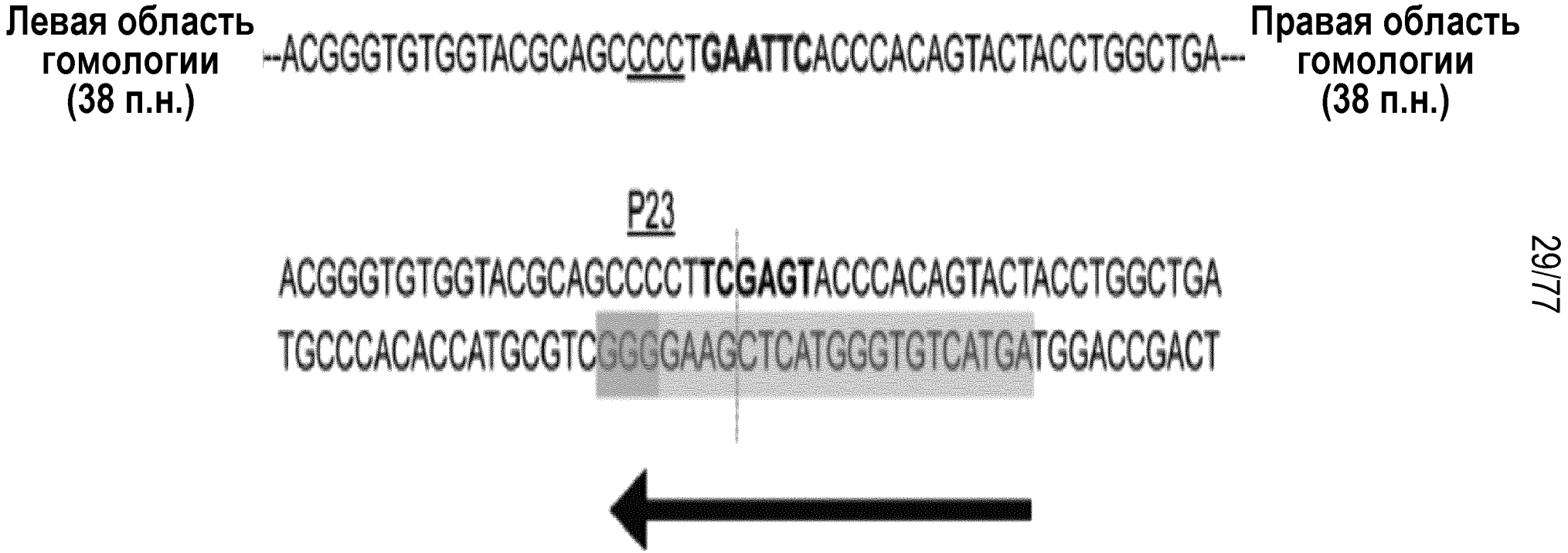
ФИГ.9C



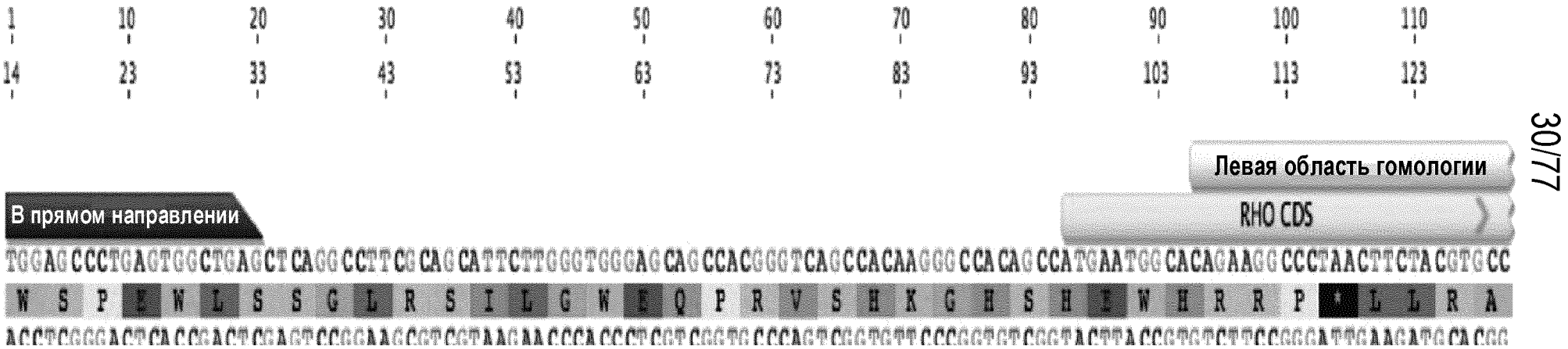
ФИГ.10А



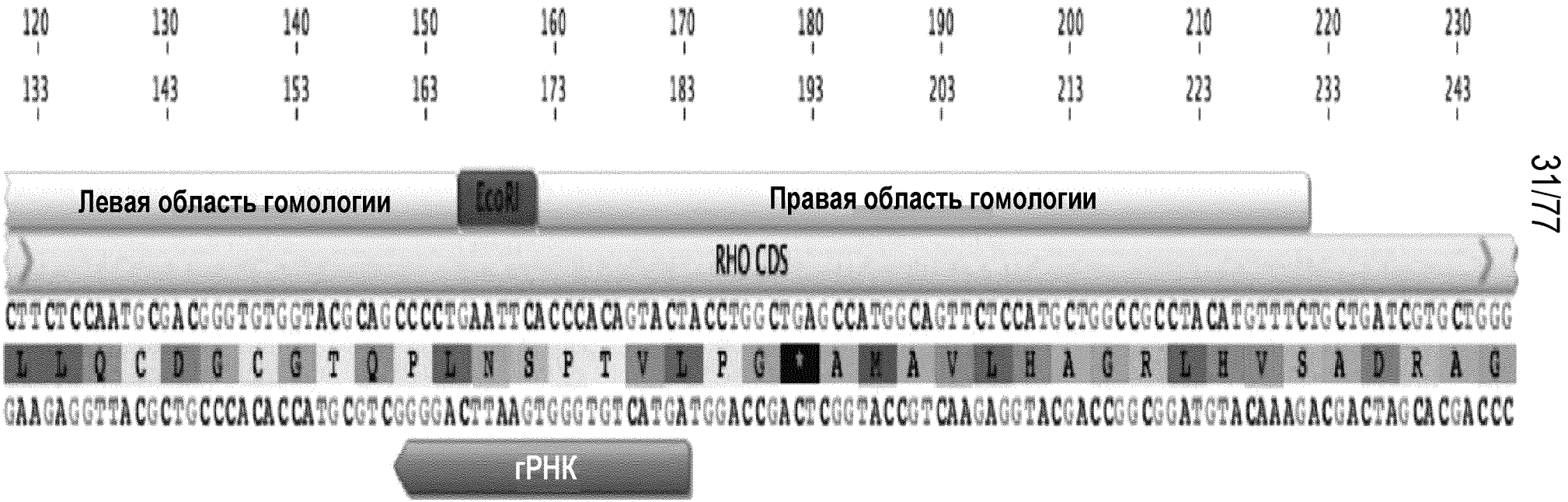
ФИГ.10B


Левая область гомологии (38 п.н.)
--ACGGGTGTGGTACGCAGCCCCTGAATTCACCCACAGTACTACCTGGCTGA---
Правая область гомологии (38 п.н.)
P23
ACGGGTGTGGTACGCAGCCCCTTCGAGTACCCACAGTACTACCTGGCTGA
TGCCCACACCATGCGTCGGGGAAGCTCATGGGTGTCATGATGGACCGACT

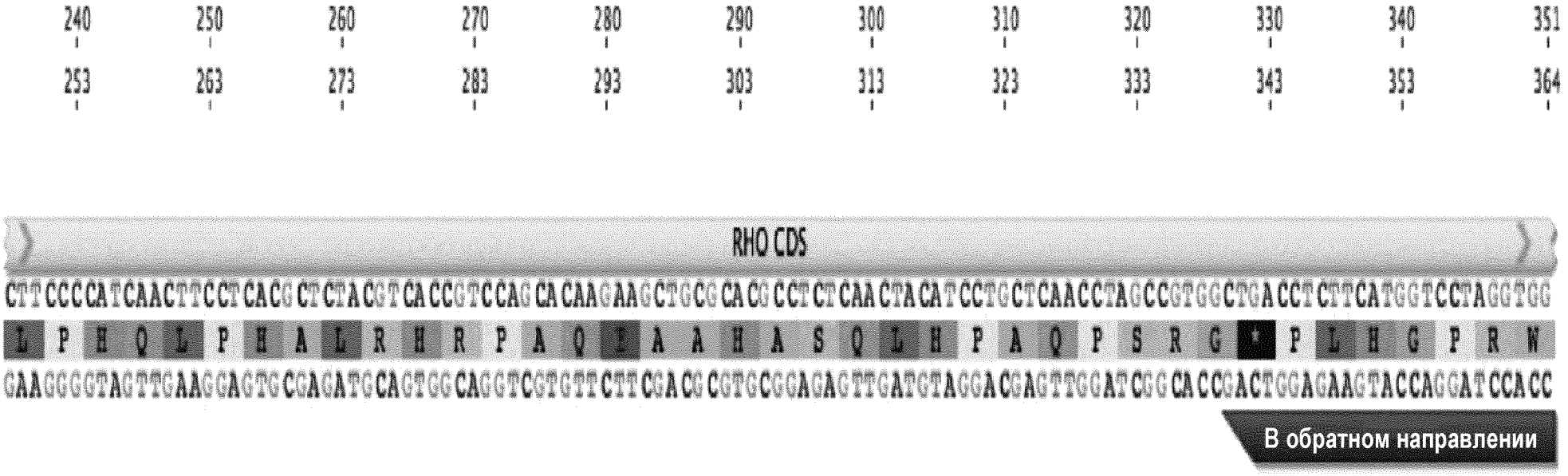
ФИГ.10C


1
10
20
30
40
50
60
70
80
90
100
110
14
23
33
43
53
63
73
83
93
103
113
123
Левая область гомологии
RHO CDS
В прямом направлении
TGGAGCCCTGAGTGGCTGAGCTCAGGCCTTCGCAGCATTCTTGGGTGGGAGCAGCCACGGGTCAGCCACAAGGGCCACAGCCATGAATGGCACAGAAGGCCCTAACTTCTACGTGCC
W S P E W L S S G L R S I L G W E Q P R V S H K G H S H E W H R R P * L L R A
ACCTCGGGACTCACGACTCGAGTCGGGAAGCGTCGTAAGAACGCACGCTCGTCGGTGCCCAGTCGGTGTTCGCGGTGTCGGTACTTACGGTGTCTTCGGGGATTGAAGATGCACGG

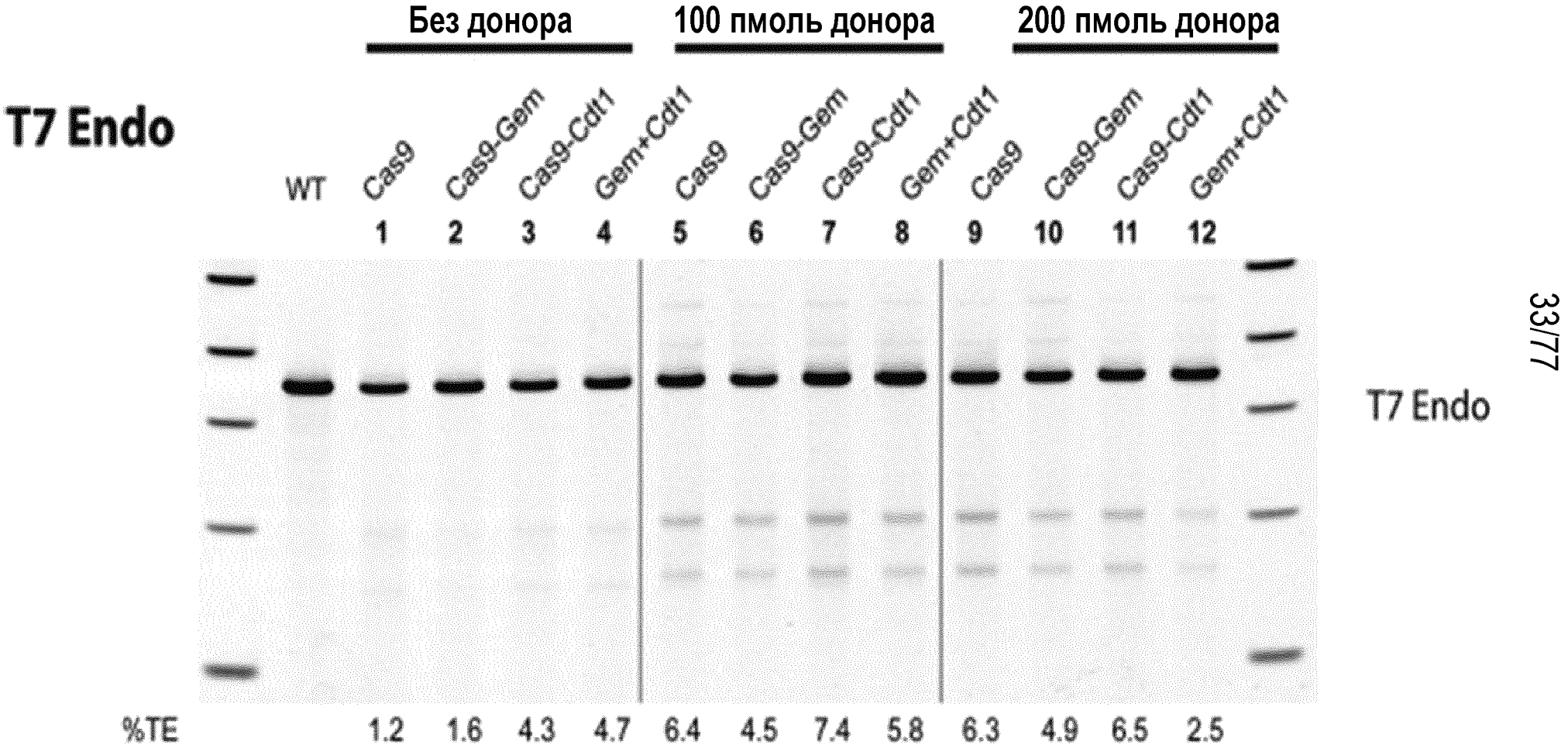
ФИГ.10C (продолжение)



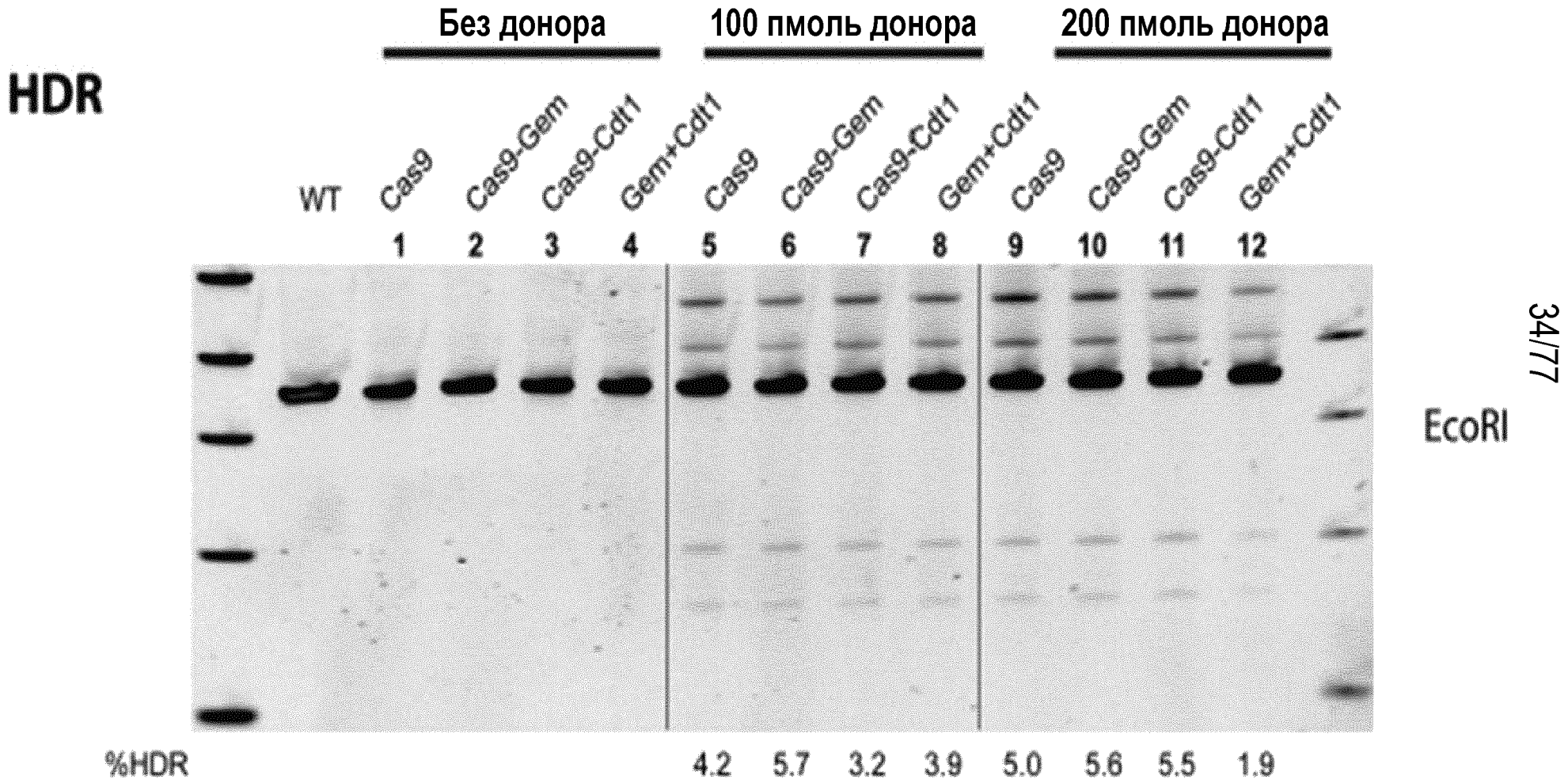
ФИГ.10C (продолжение)



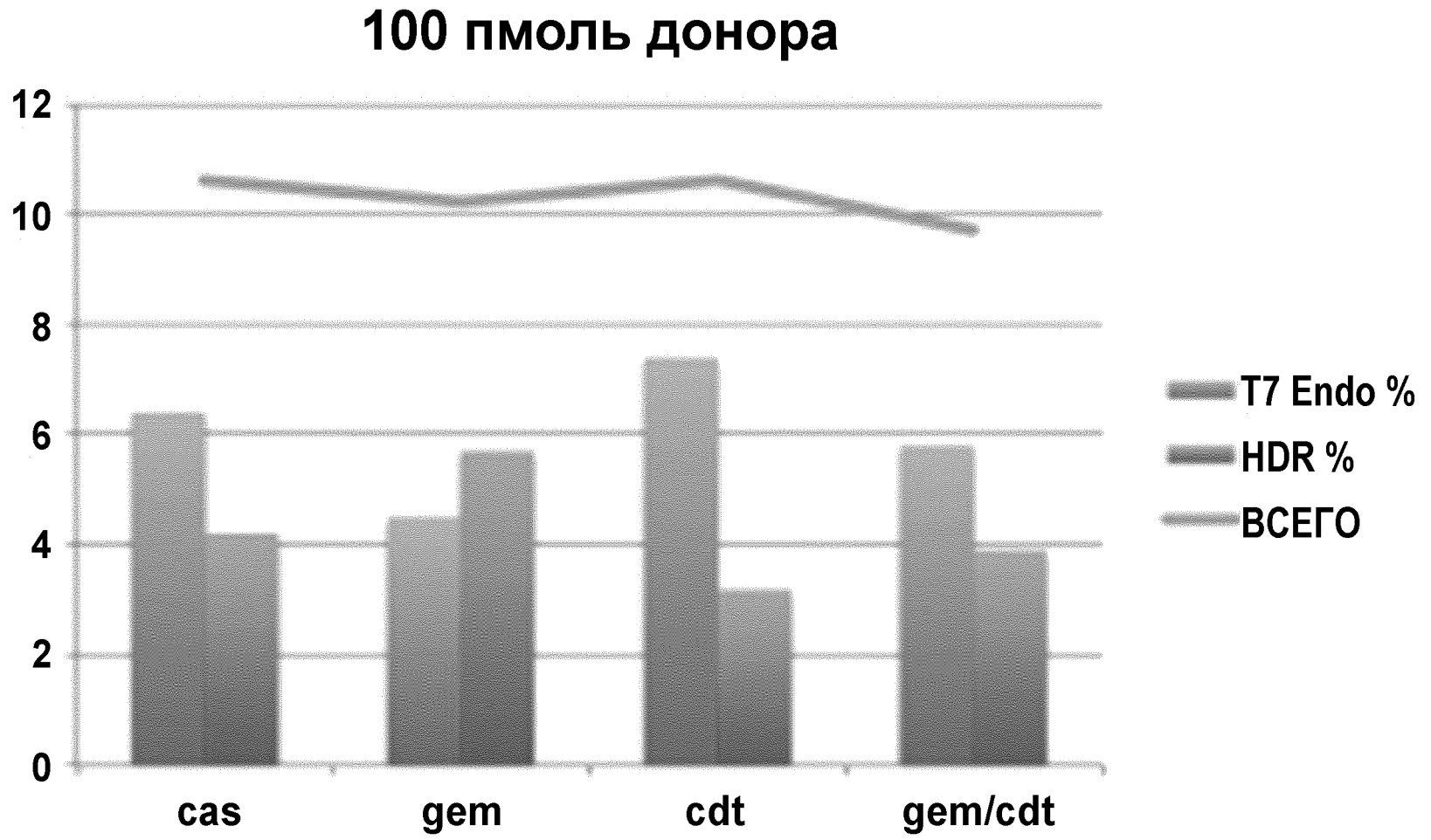
ФИГ.11А


Без донора
100 пмоль донора
200 пмоль донора
T7 Endo
WT
Cas9
Cas9-Gem
Cas9-Cdt1
Gem+Cdt1
1 2 3 4 5 6 7 8 9 10 11 12
%TE
1.2 1.6 4.3 4.7 6.4 4.5 7.4 5.8 6.3 4.9 6.5 2.5
T7 Endo

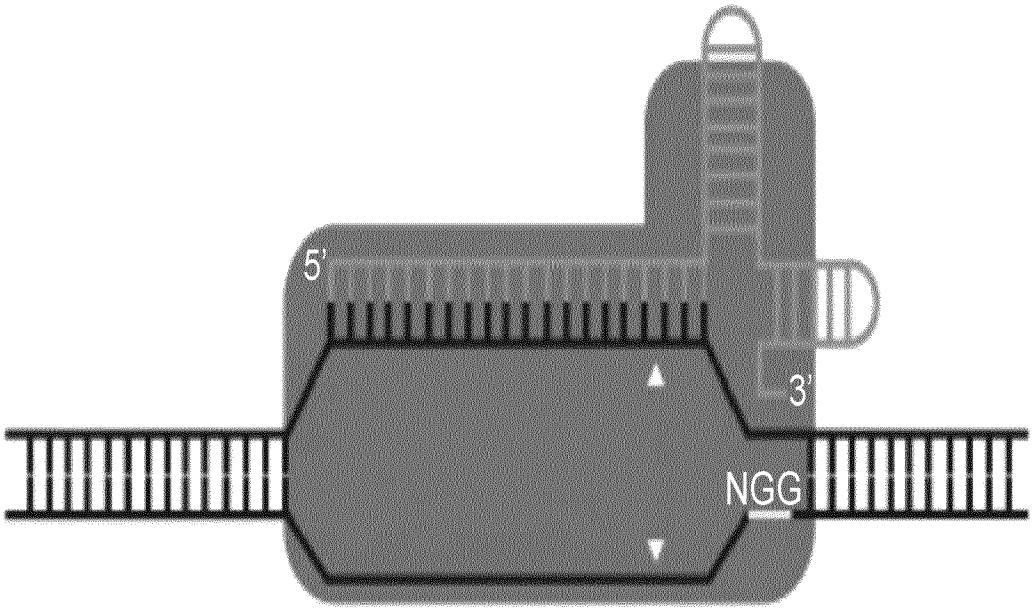
ФИГ.11B


HDR
Без донора
100 пмоль донора
200 пмоль донора
WT
Cas9
Cas9-Gem
Cas9-Cdt1
Gem+Cdt1
1
2
3
4
5
6
7
8
9
10
11
12
EcoRI
%HDR
4.2
5.7
3.2
3.9
5.0
5.6
5.5
1.9

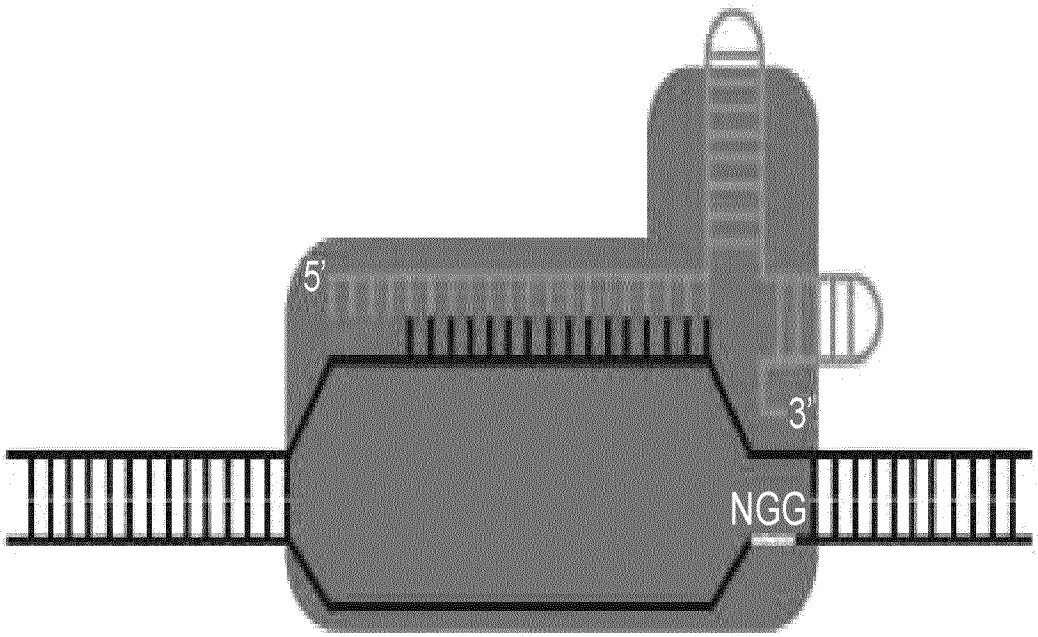
ФИГ.11C


100 пмоль донора
12
10
8
6
4
2
0
cas
gem
cdt
gem/cdt
T7 Endo %
HDR %
ВСЕГО

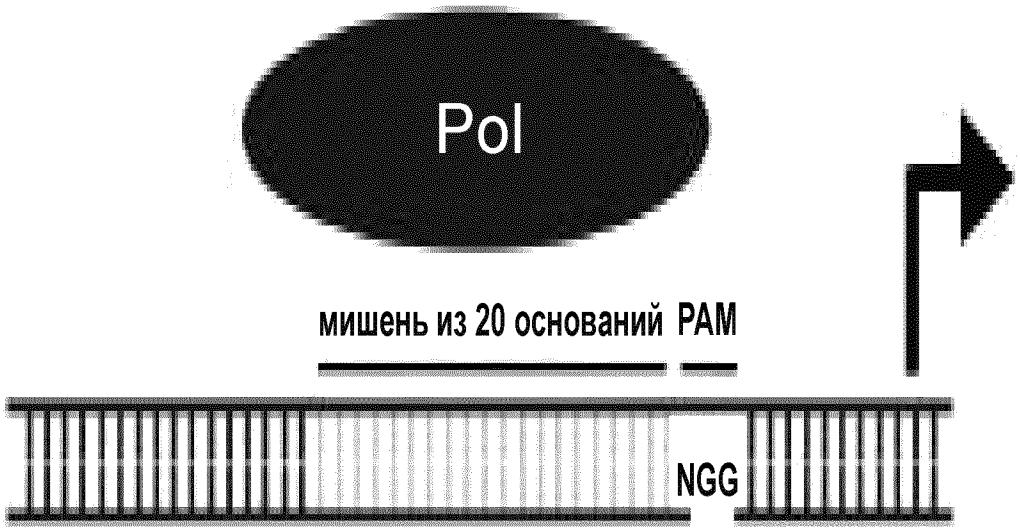
ФИГ.12A



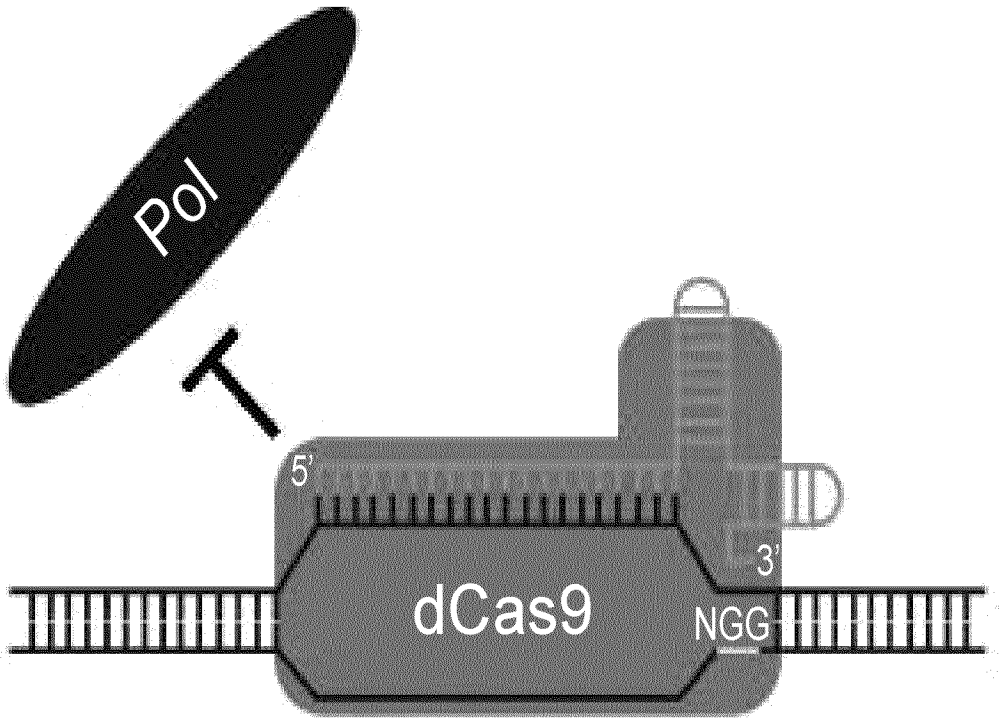
ФИГ.12B



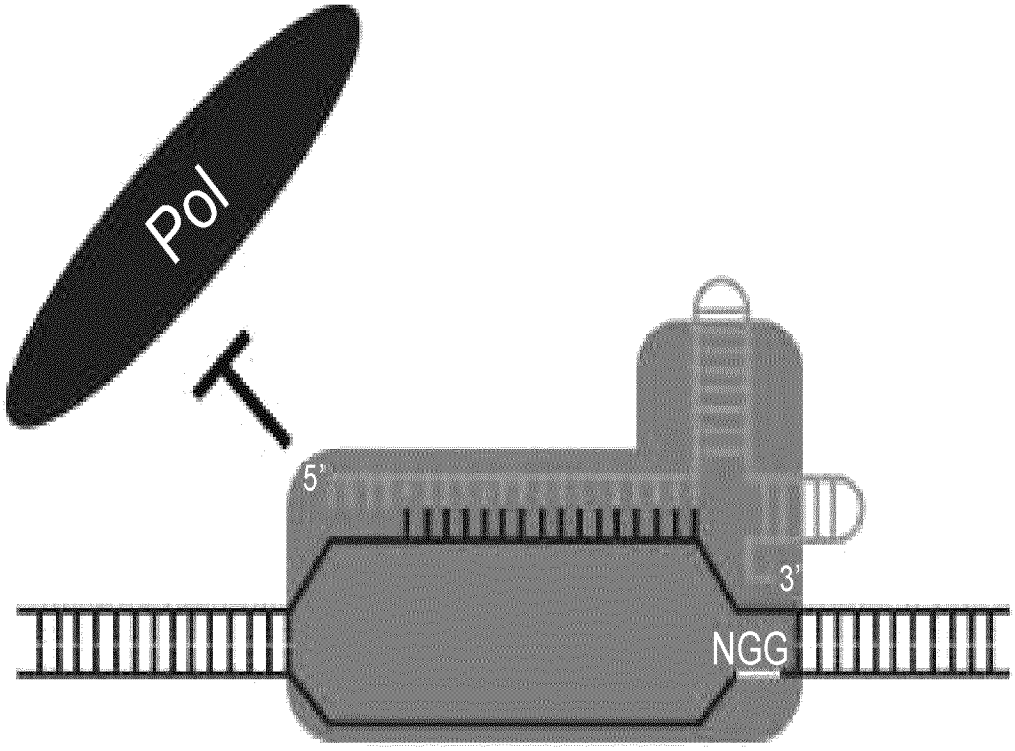
ФИГ.13A



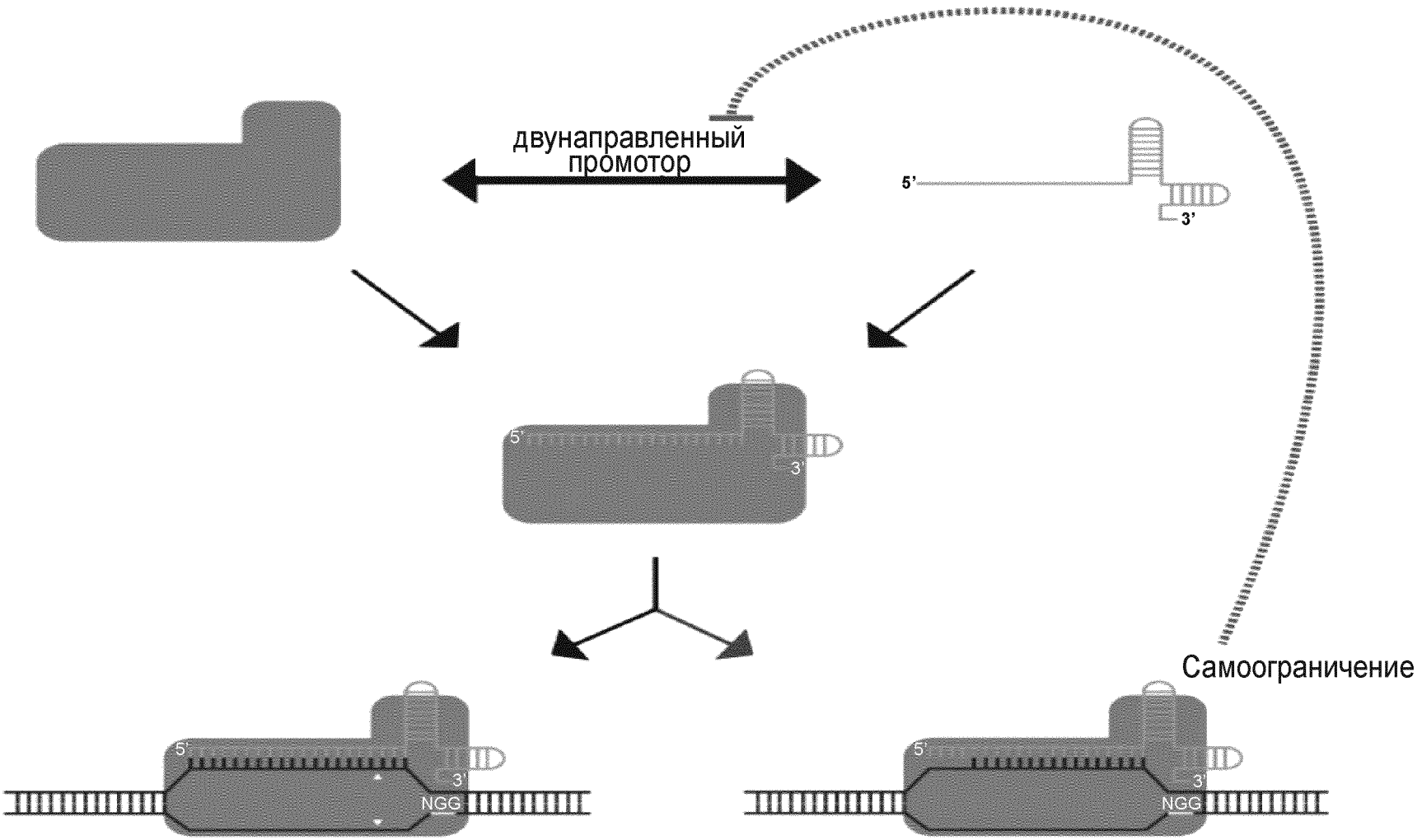
ФИГ.13B



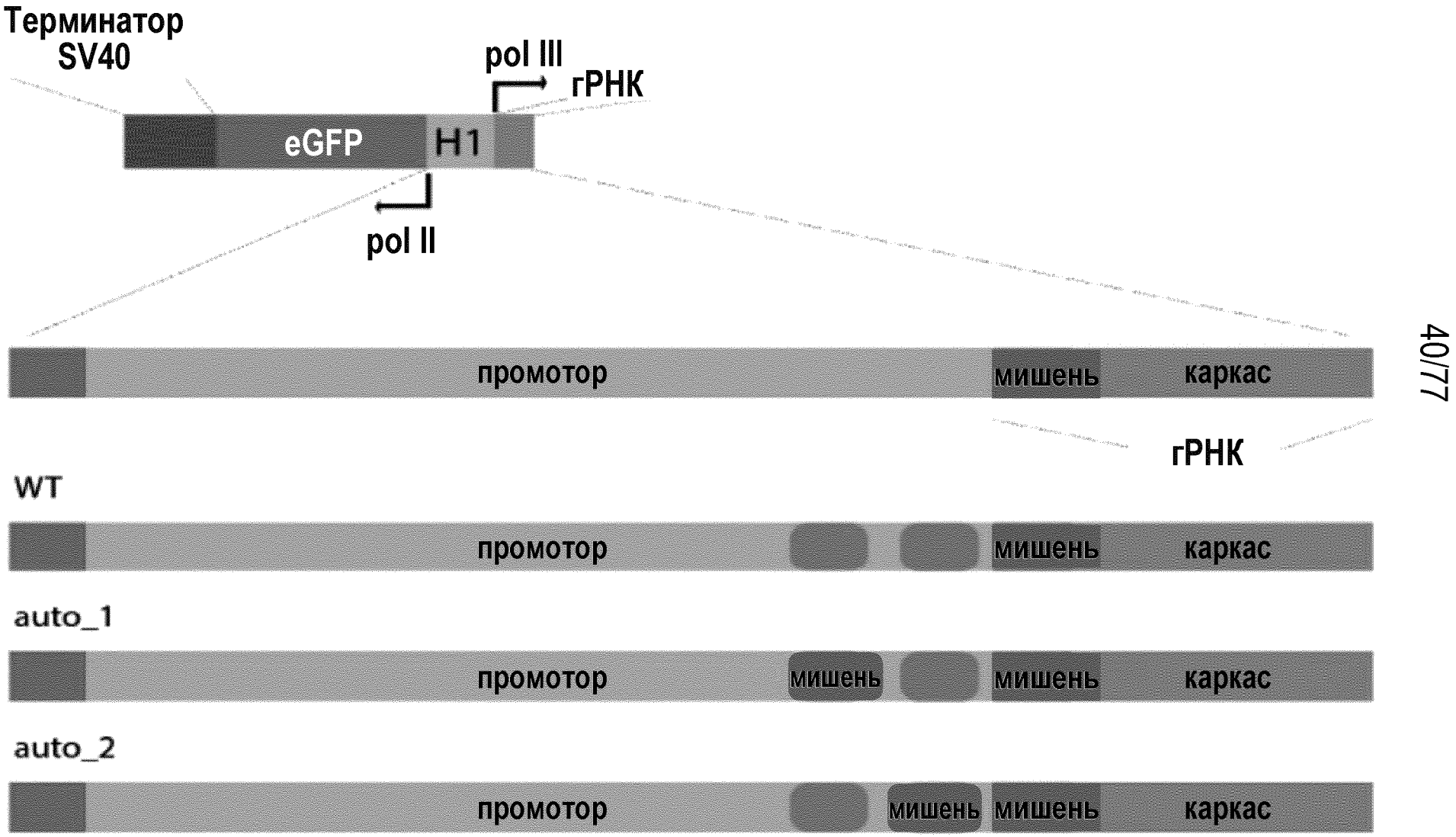
ФИГ.13C



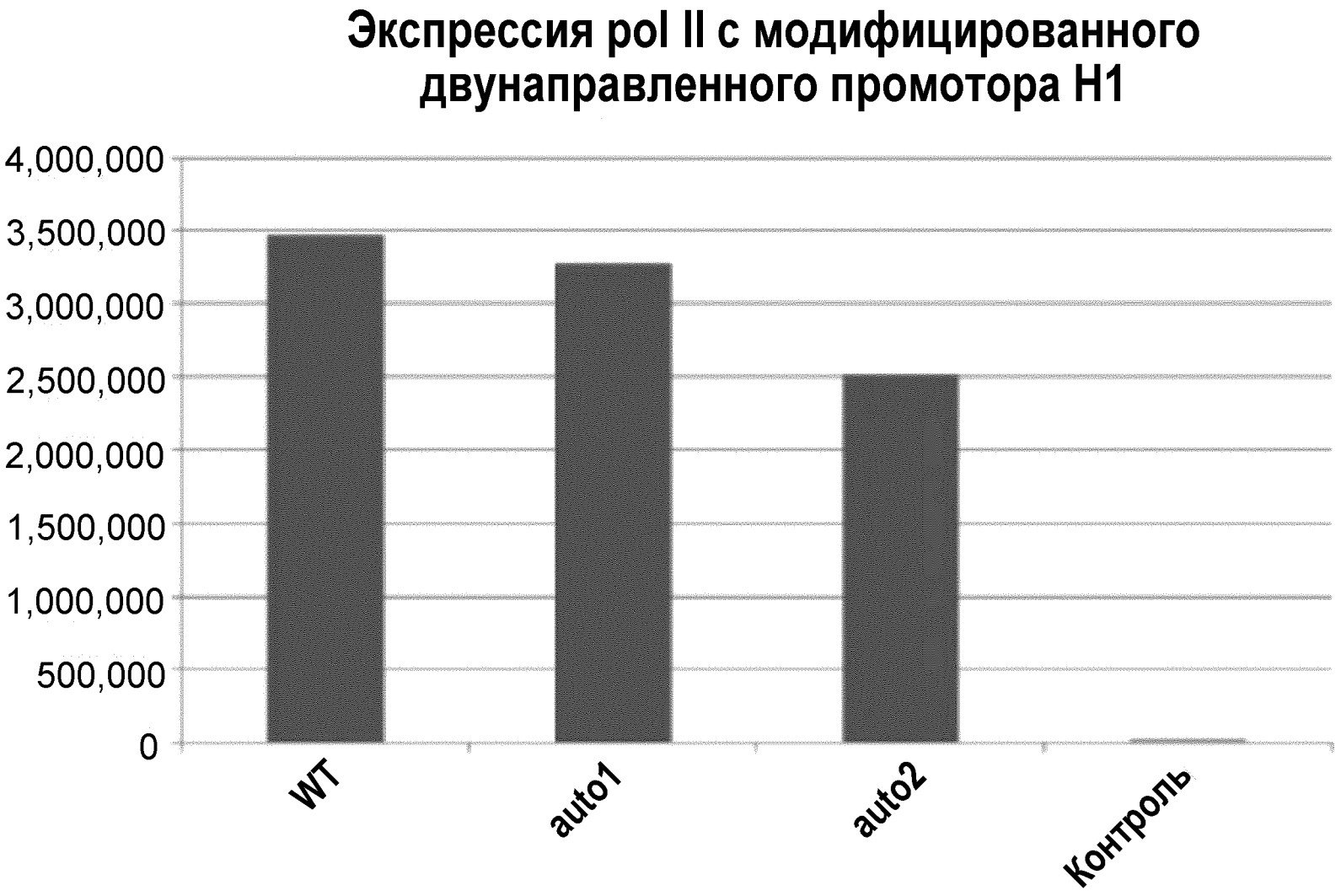
ФИГ.14


двунаправленный
промотор
5'
3'
5'
3'
Самоограничение
5'
3'
NGG
5'
3'
NGG

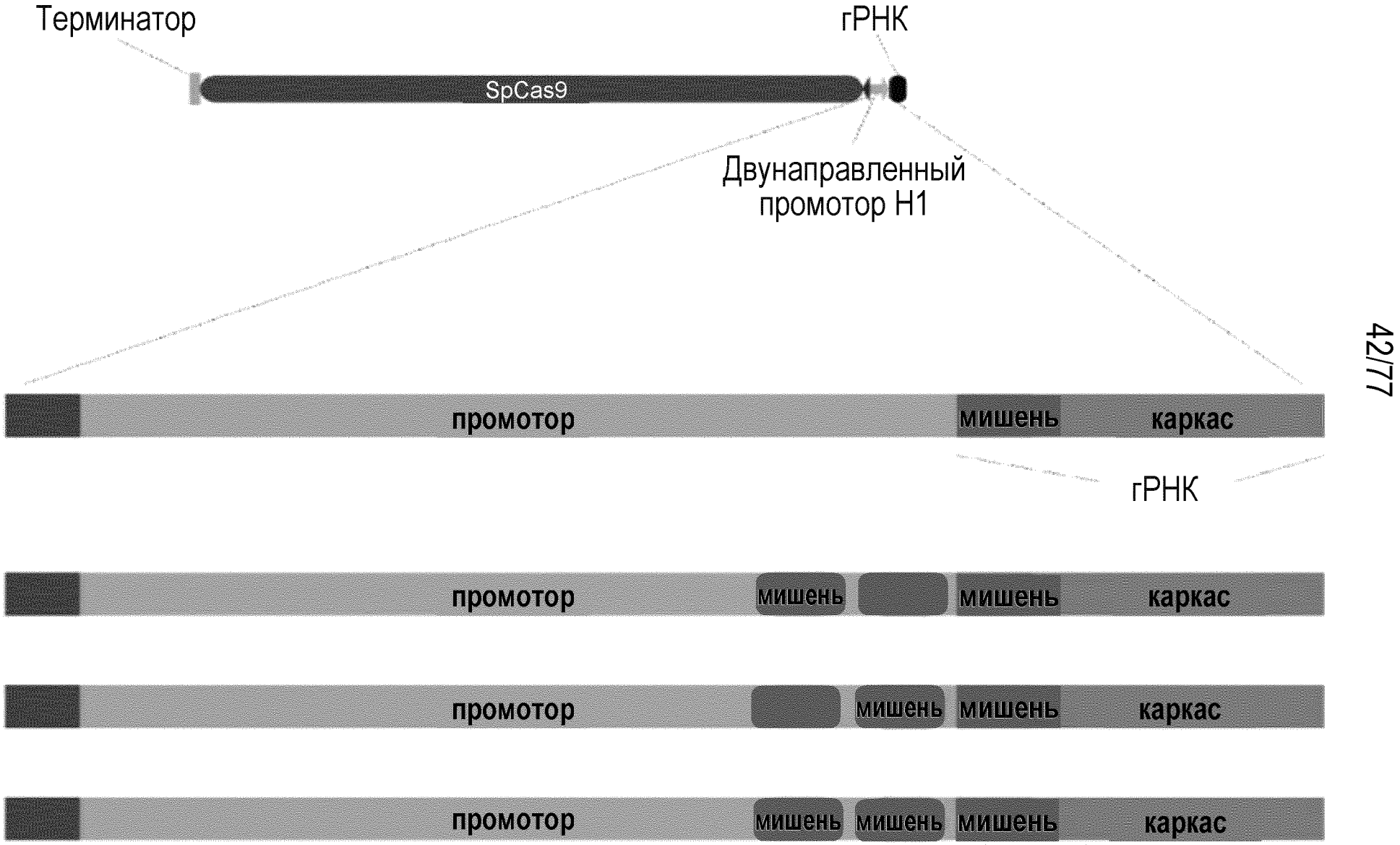
ФИГ.15


Терминатор
SV40
pol III
гРНК
eGFP
H1
pol II
промотор
мишень
каркас
гРНК
WT
промотор
мишень
каркас
auto_1
промотор
мишень
мишень
каркас
auto_2
промотор
мишень
мишень
каркас

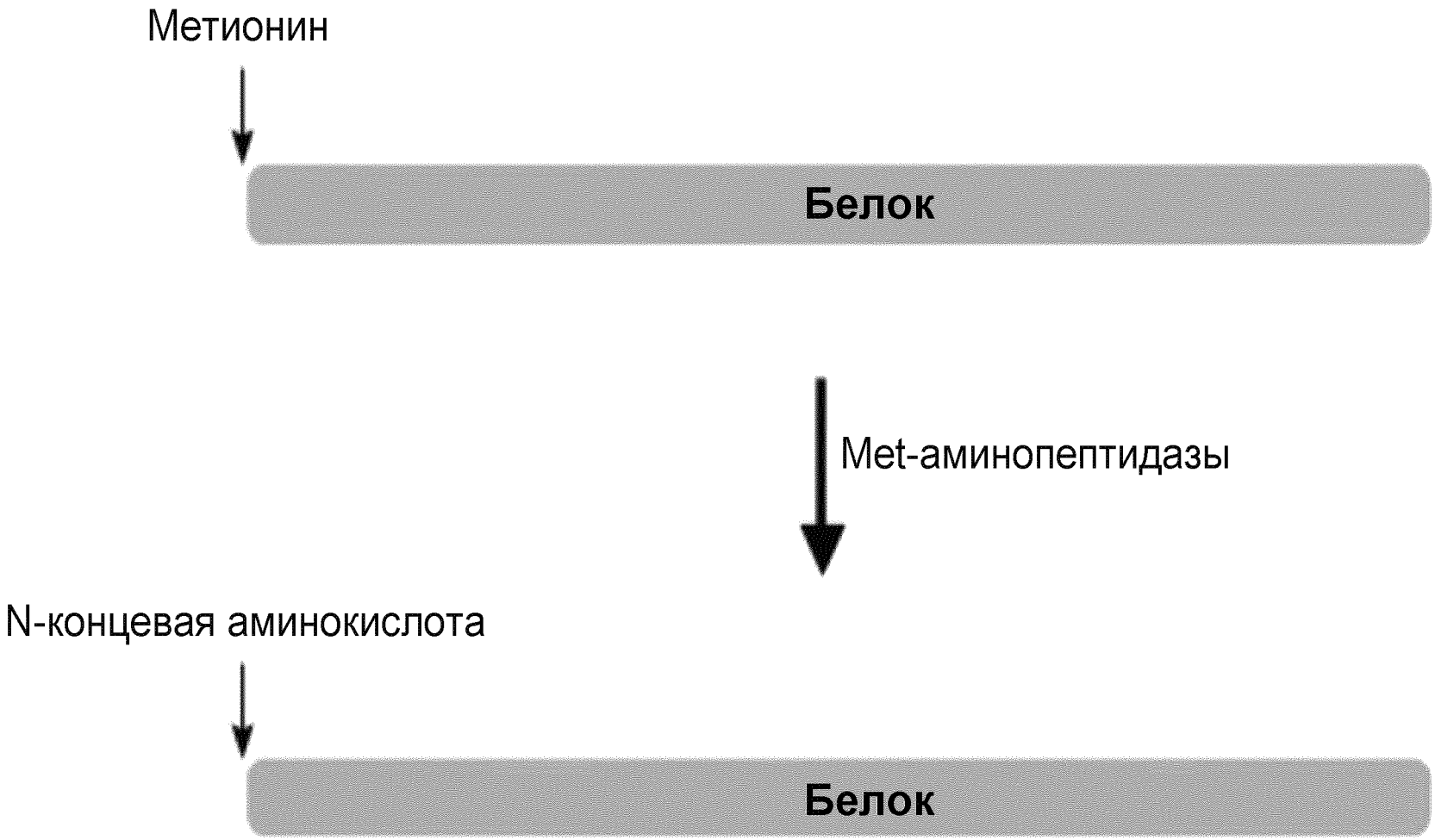
ФИГ.15 (продолжение)



ФИГ.16


Терминатор
гРНК
SpCas9
Двунаправленный промотор H1
промотор
мишень
каркас
гРНК
промотор
мишень
мишень
каркас
промотор
мишень
мишень
каркас
промотор
мишень
мишень
мишень
каркас

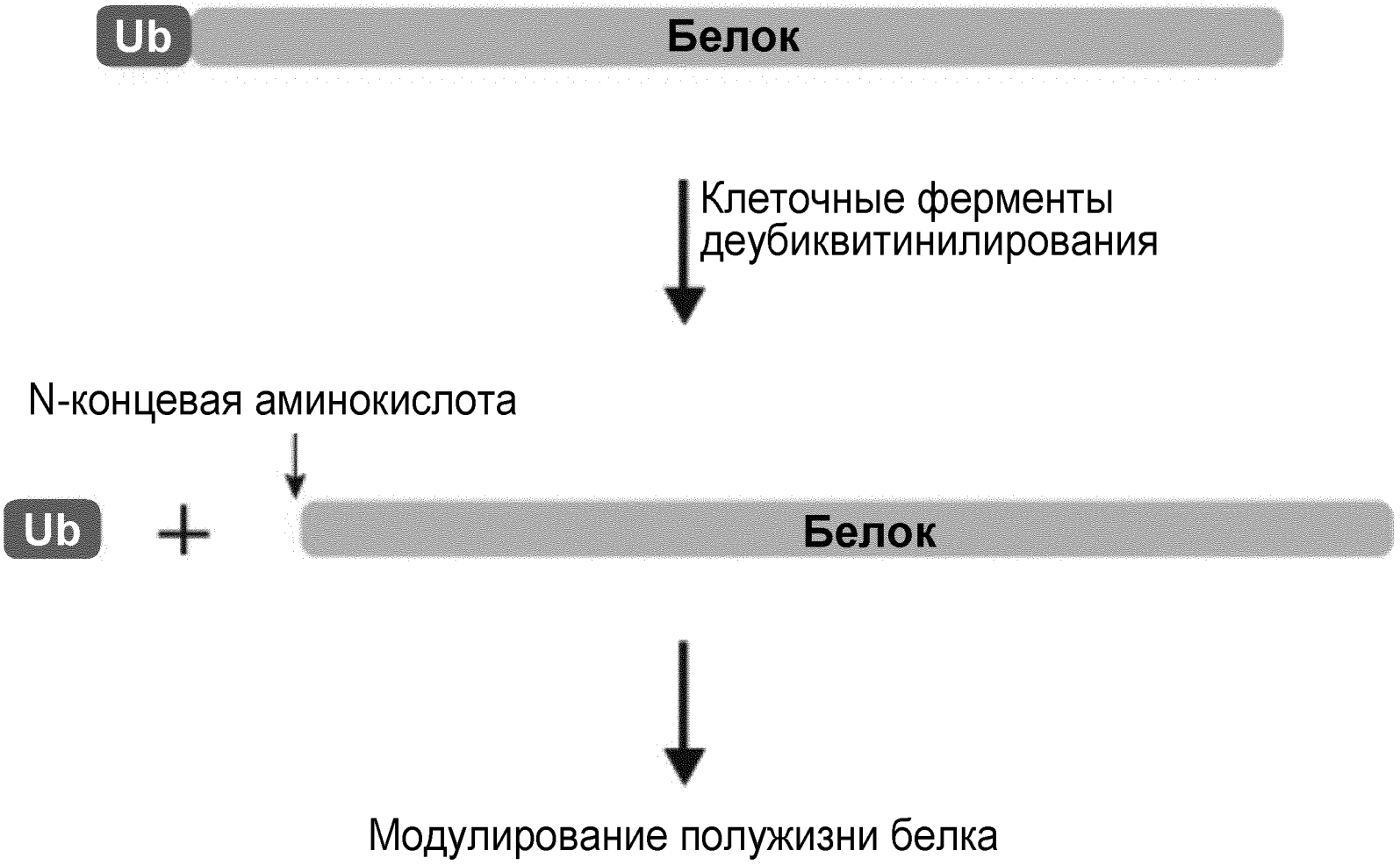
ФИГ.17А


Метионин
Белок
Met-аминопептидазы
N-концевая аминокислота
Белок

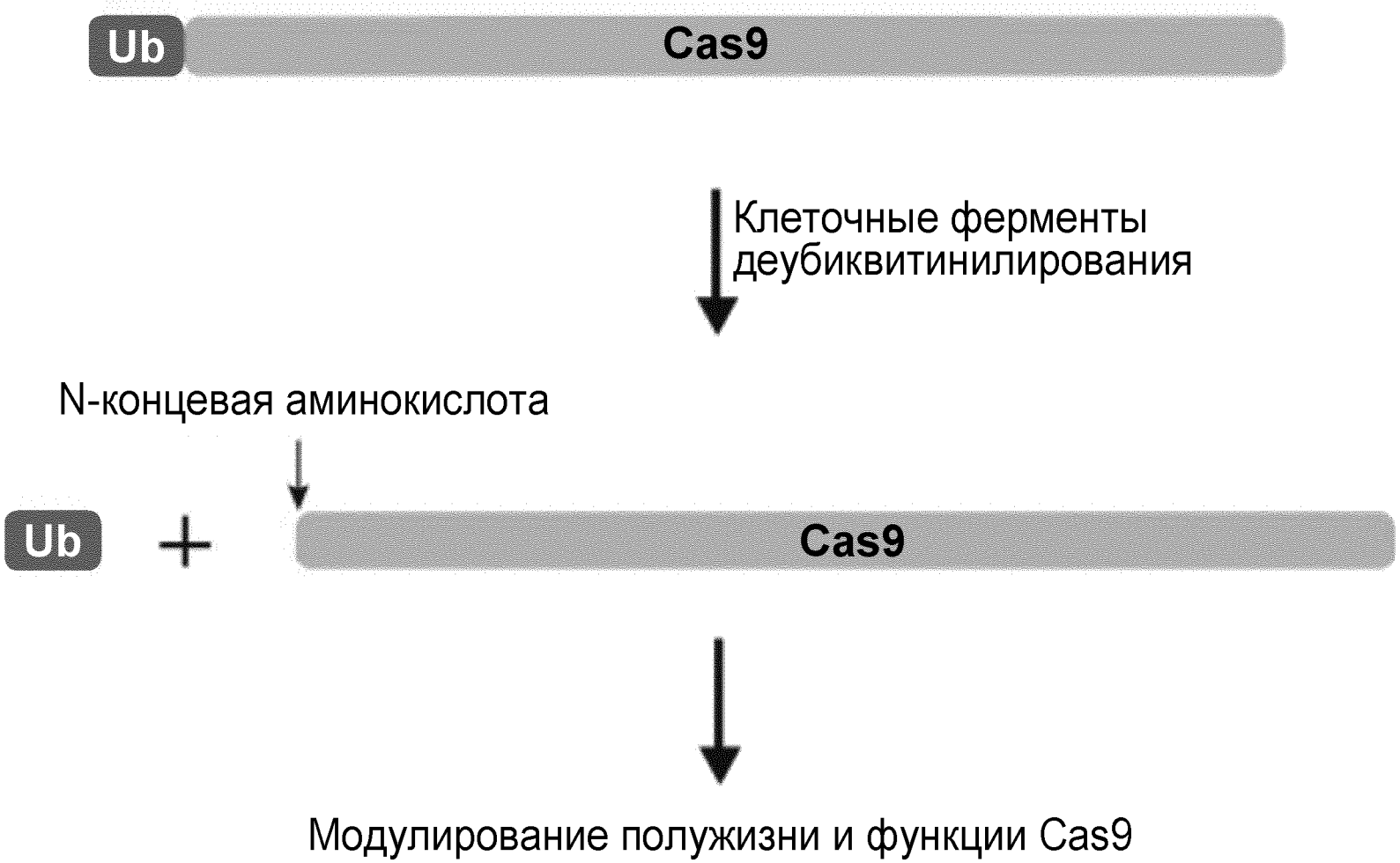
ФИГ.17B

Время полужизни белка в клетках млекопитающих
120
100
80
60
40
20
0
Время полужизни (мин)
Val
Met
Gly
Pro
Ile
Thr
Leu
Ala
His
Trp
Tyr
Ser
Asn
Lys
Cys
Asp
Phe
Glu
Arg
Gln
N-концевая аминокислота

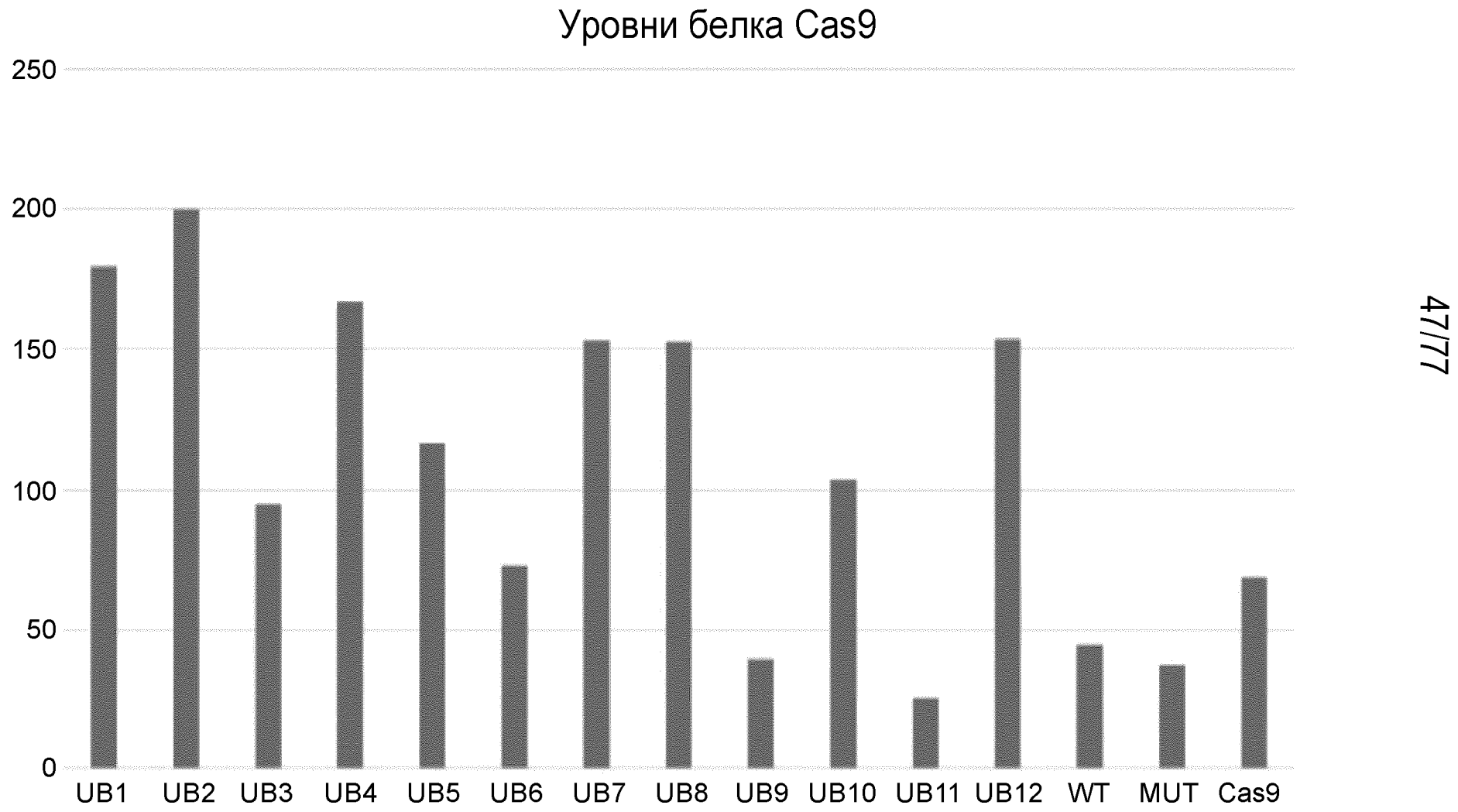
ФИГ.18


Ub
Белок
Клеточные ферменты
деубиквитинилирования
N-концевая аминокислота
Ub
+
Белок
Модулирование полужизни белка

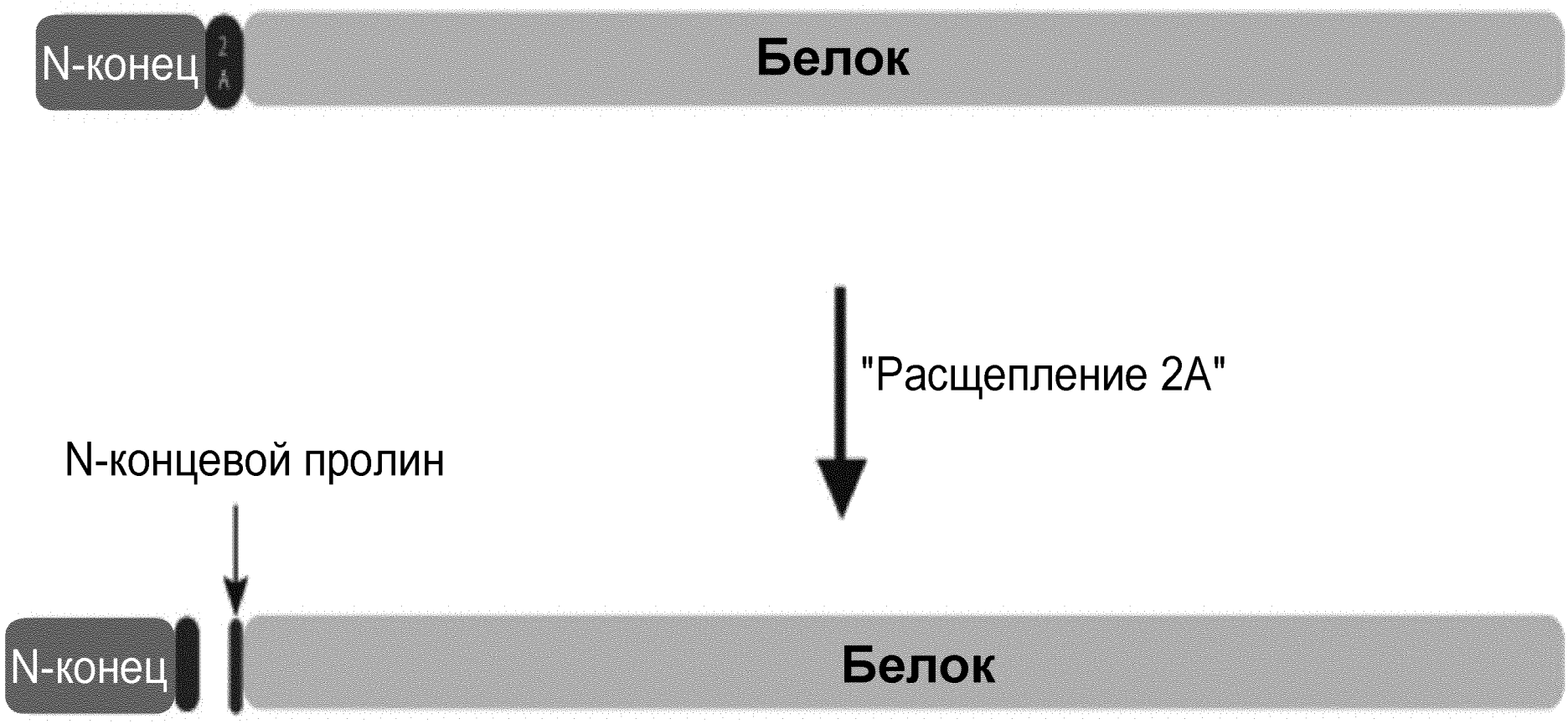
ФИГ.19А


Ub
Cas9
Клеточные ферменты
деубиквитинилирования
N-концевая аминокислота
Ub
+
Cas9
Модулирование полужизни и функции Cas9

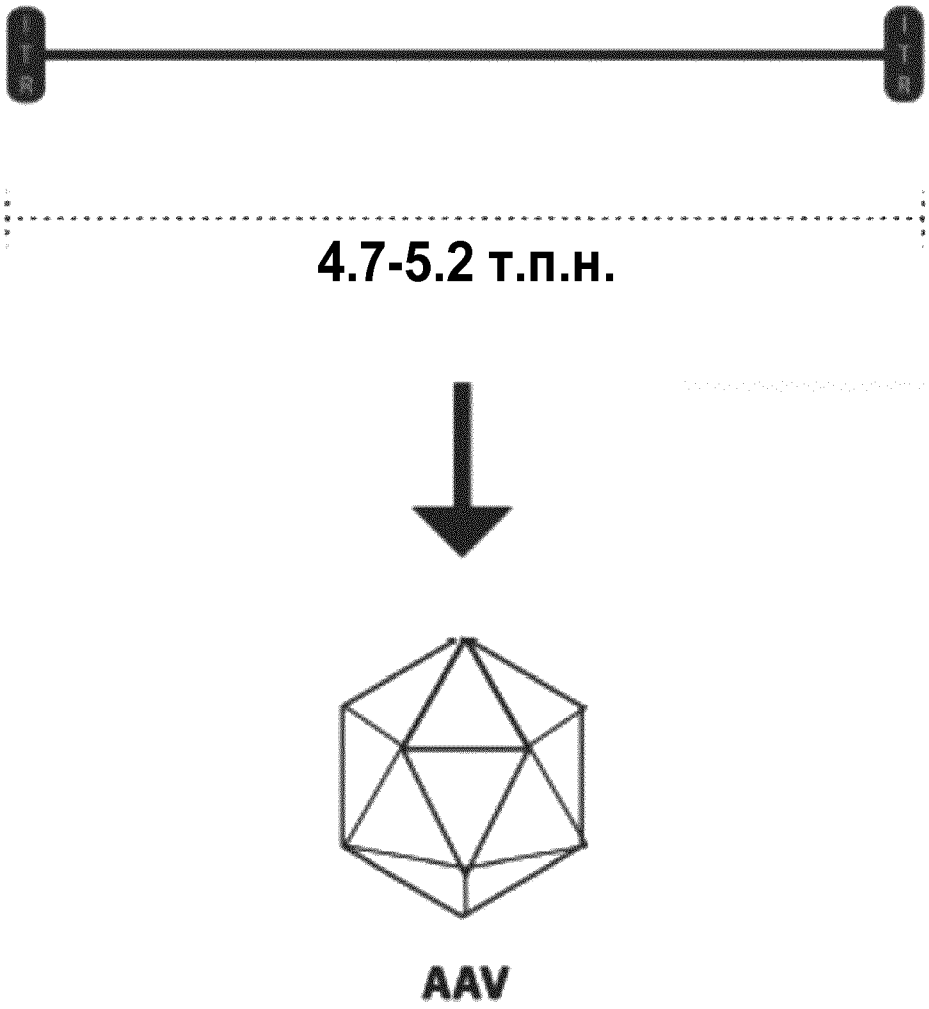
ФИГ.19B


Уровни белка Cas9
250
200
150
100
50
0
UB1
UB2
UB3
UB4
UB5
UB6
UB7
UB8
UB9
UB10
UB11
UB12
WT
MUT
Cas9

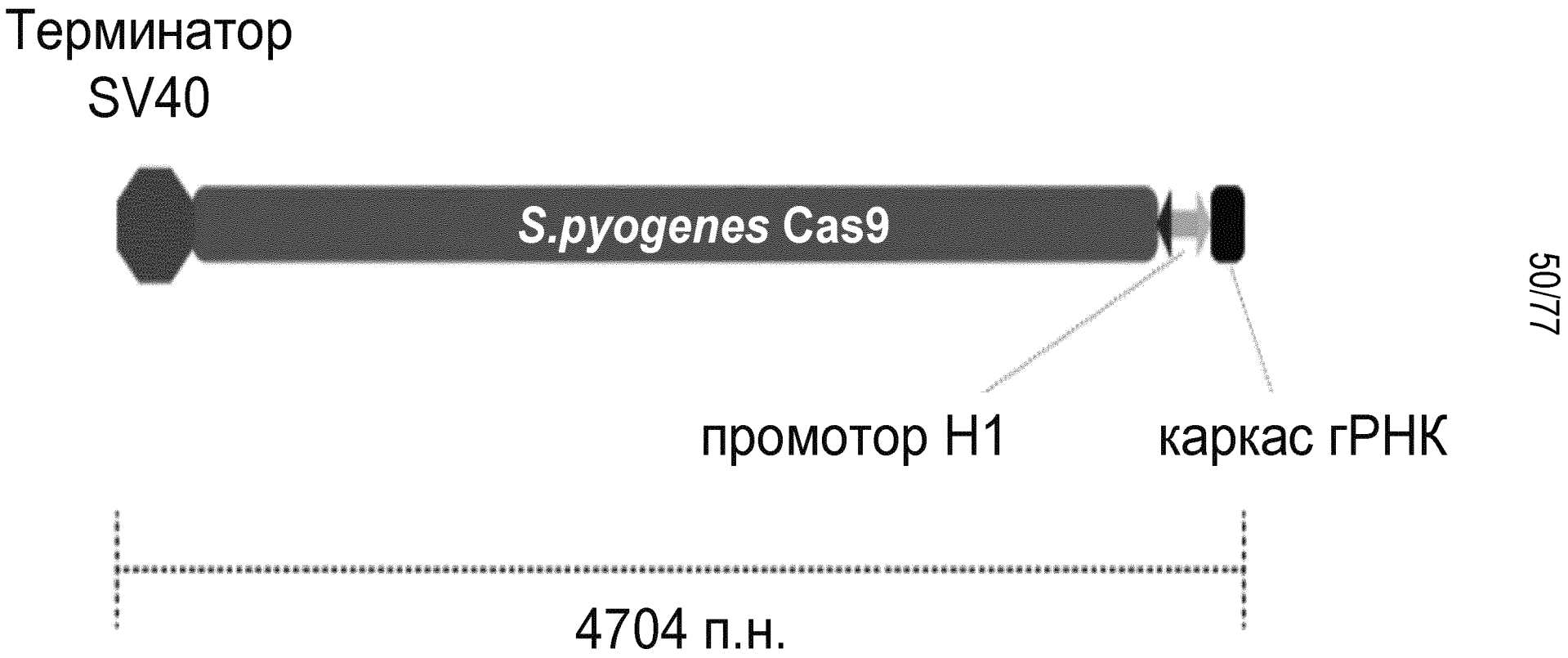
ФИГ.20


N-конец
2A
Белок
"Расщепление 2A"
N-концевой пролин
N-конец
Белок

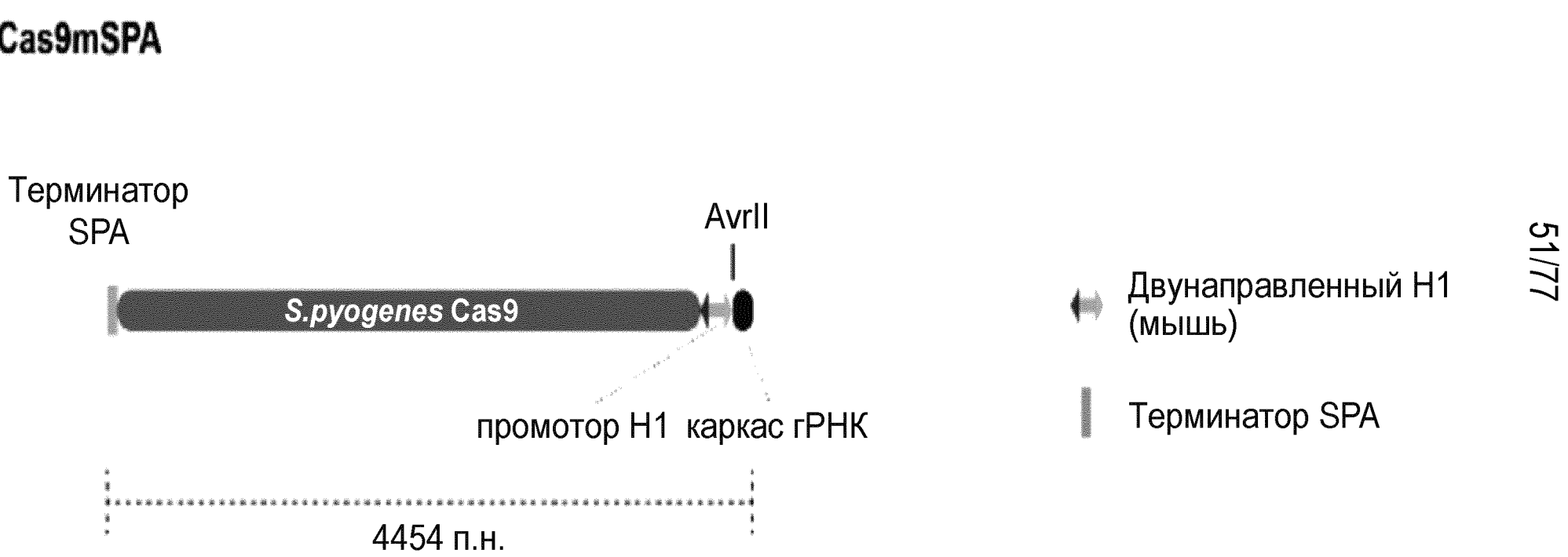
ФИГ.21А



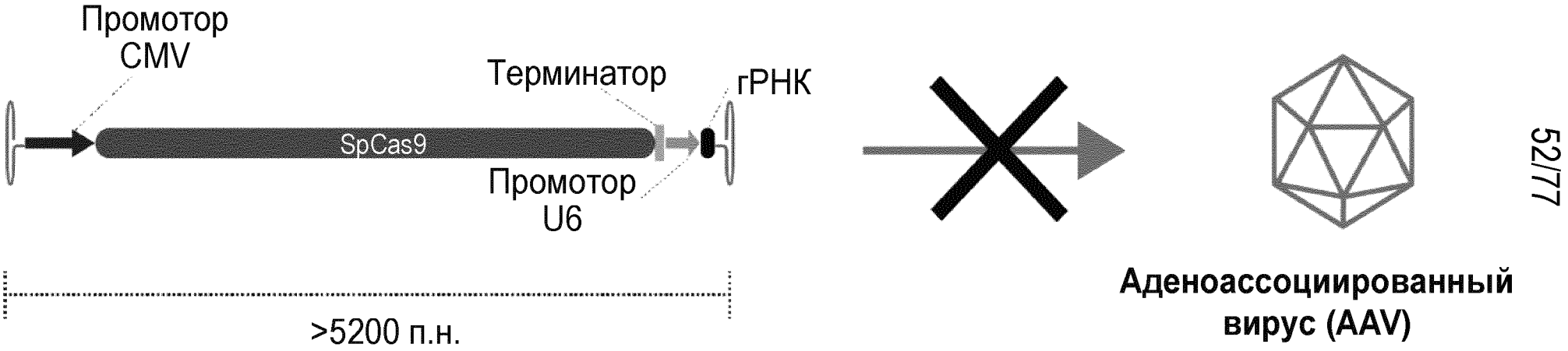
ФИГ.21B


Терминатор
SV40
S.pyogenes Cas9
промотор H1
каркас гРНК
4704 п.н.

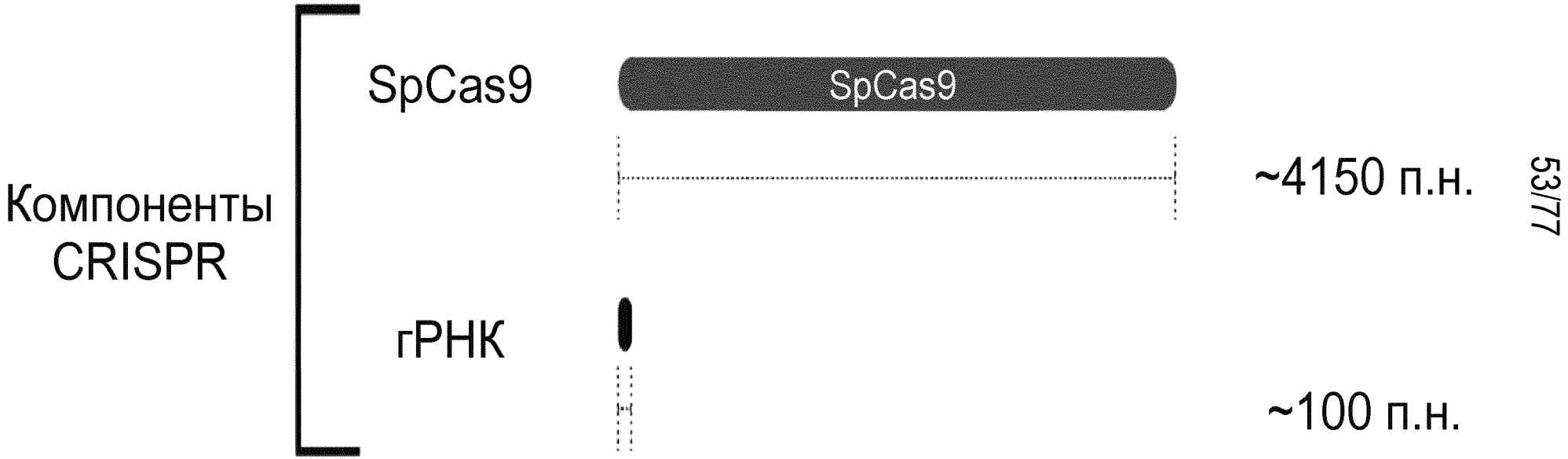
ФИГ.21C


pCas9mSPA
Терминатор SPA
AvrII
S.pyogenes Cas9
промотор H1
каркас гРНК
4454 п.н.
Двунаправленный H1 (мышь)
Терминатор SPA

ФИГ.22А

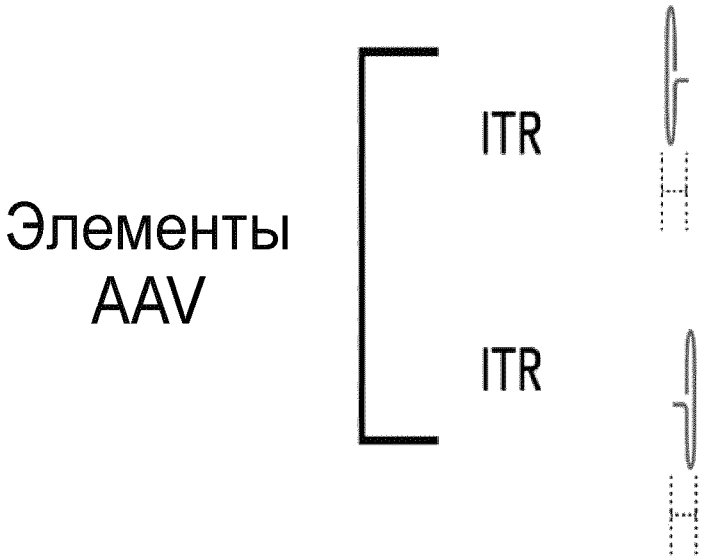

Промотор
CMV
Терминатор
гРНК
SpCas9
Промотор
U6
>5200 п.н.
Аденоассоциированный
вирус (AAV)

ФИГ.22A (продолжение)



ФИГ.22А (продолжение)

ФИГ.22A (продолжение)

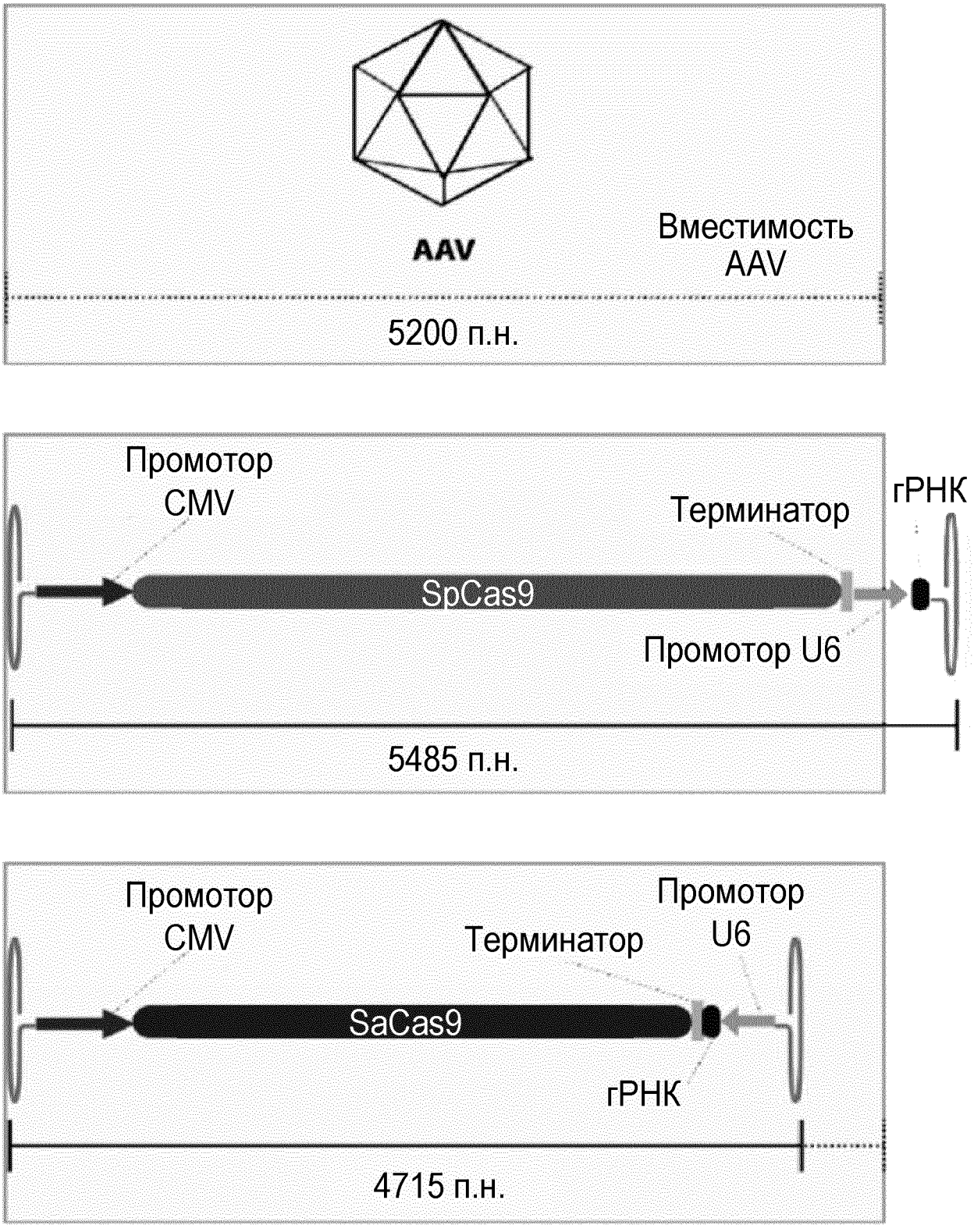


~150 п.н.

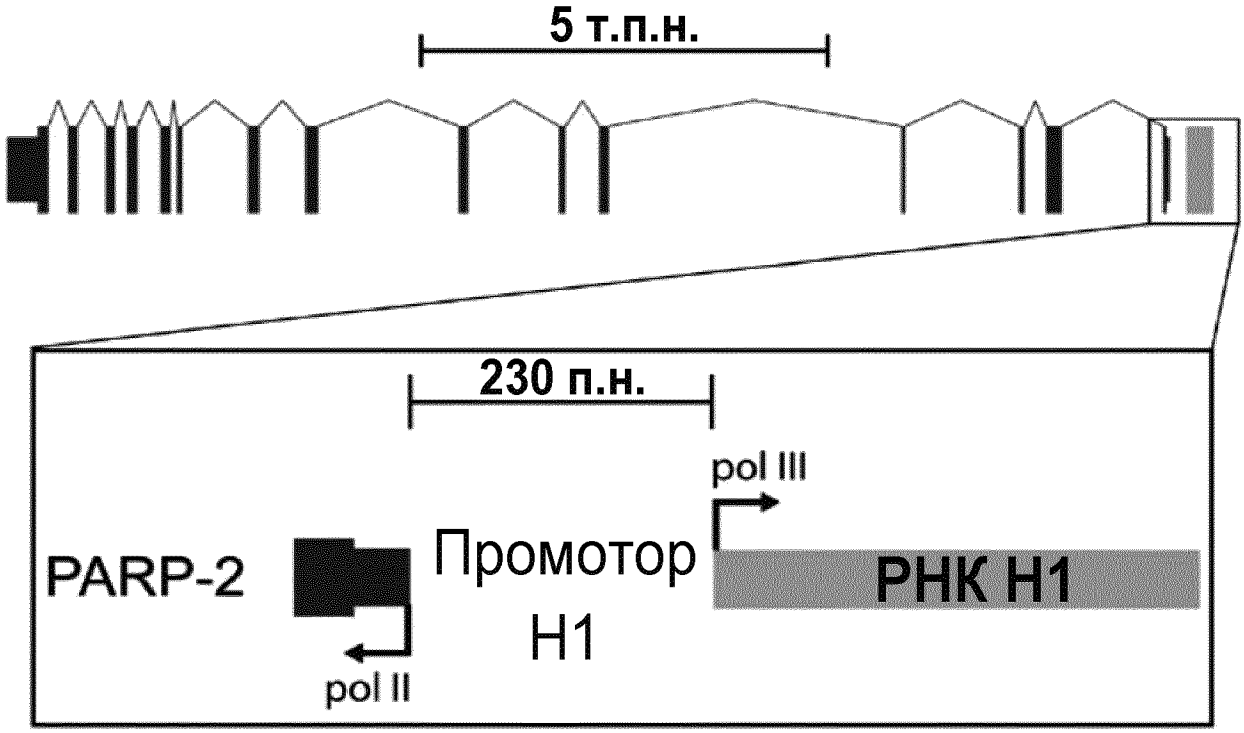
~150 п.н.

>5200 п.н.

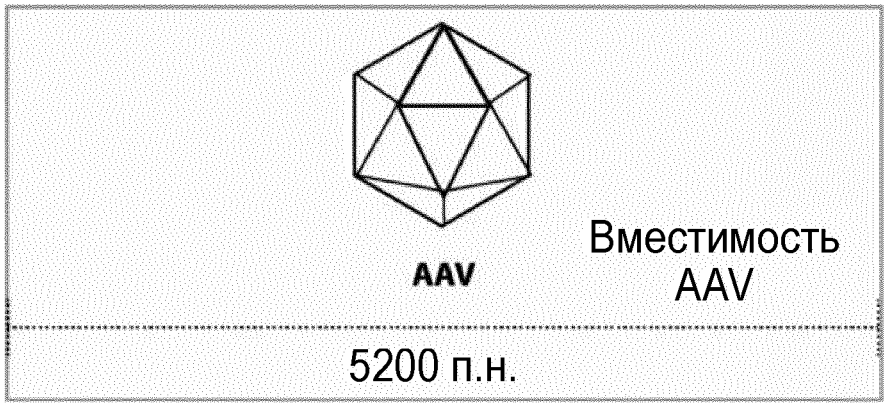
ФИГ.22В

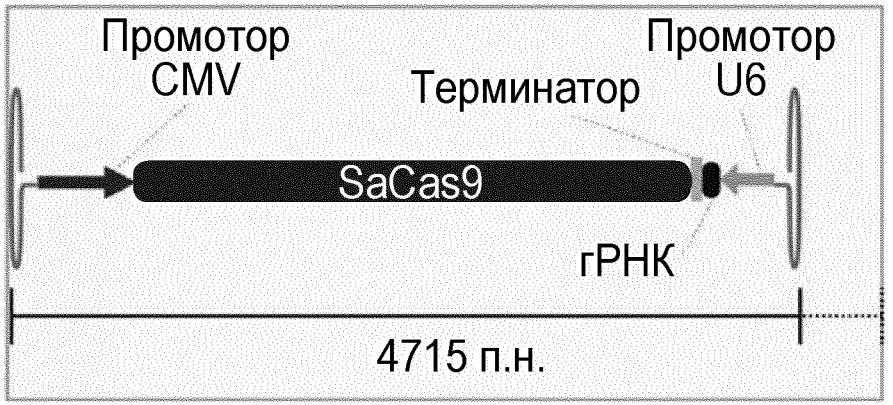

AAV
Вместимость AAV
5200 п.н.
Промотор CMV
Терминатор
гРНК
SpCas9
Промотор U6
5485 п.н.
Промотор CMV
Терминатор
Промотор U6
SaCas9
гРНК
4715 п.н.

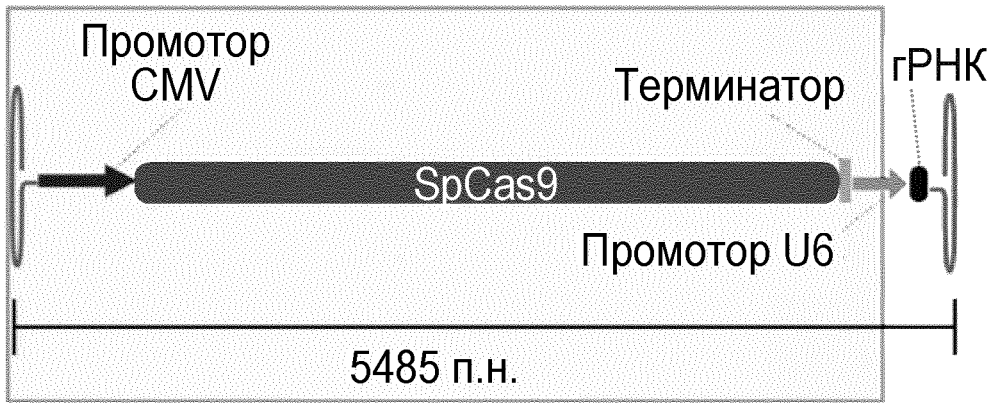
ФИГ.22C



ФИГ.22D

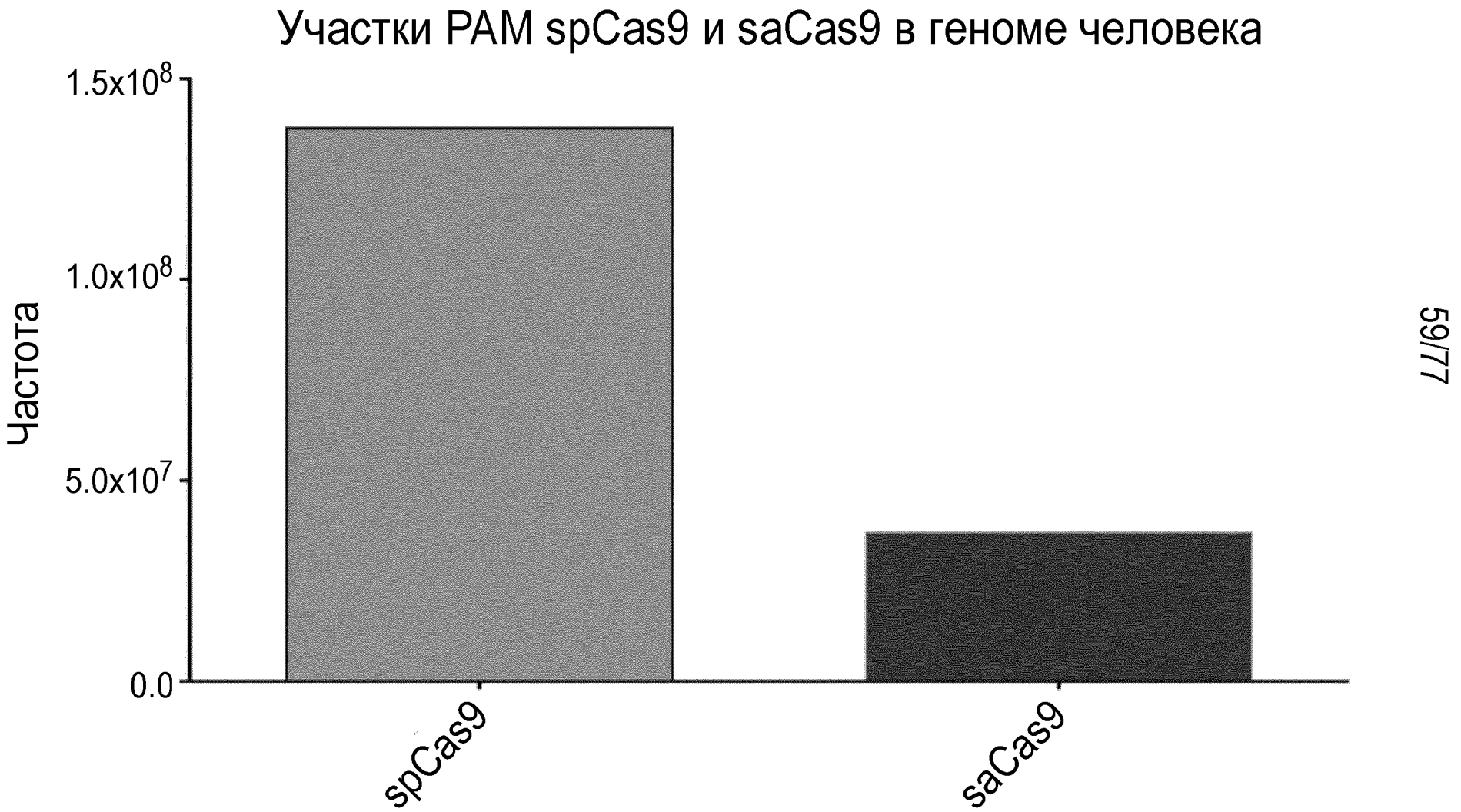

AAV
Вместимость AAV
5200 п.н.


Промотор CMV
Терминатор
Промотор U6
SaCas9
гРНК
4715 п.н.

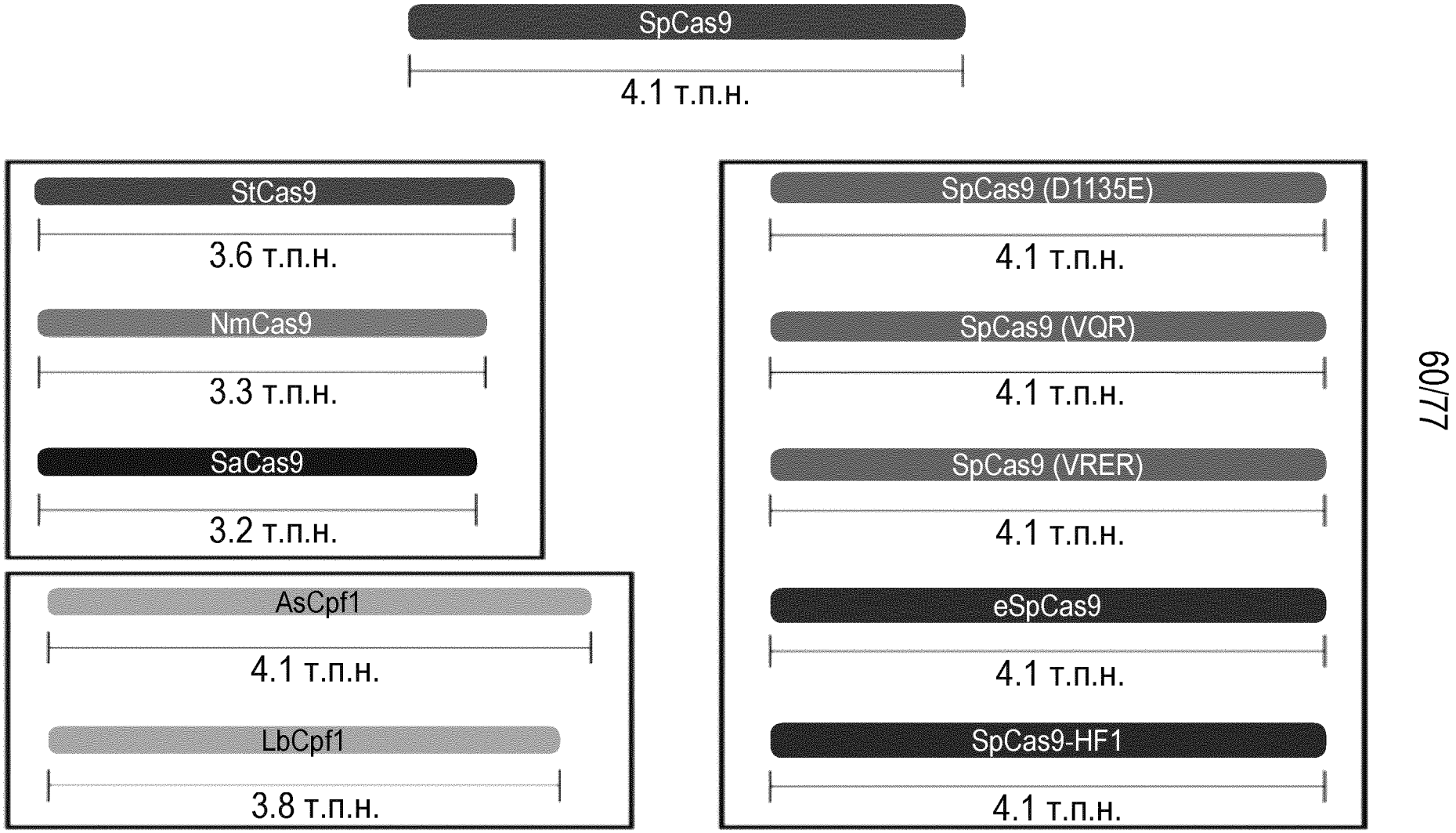

Промотор CMV
Терминатор
гРНК
SpCas9
Промотор U6
5485 п.н.

Двунаправленный промотор H1
Терминатор
SpCas9
гРНК
4855 п.н.

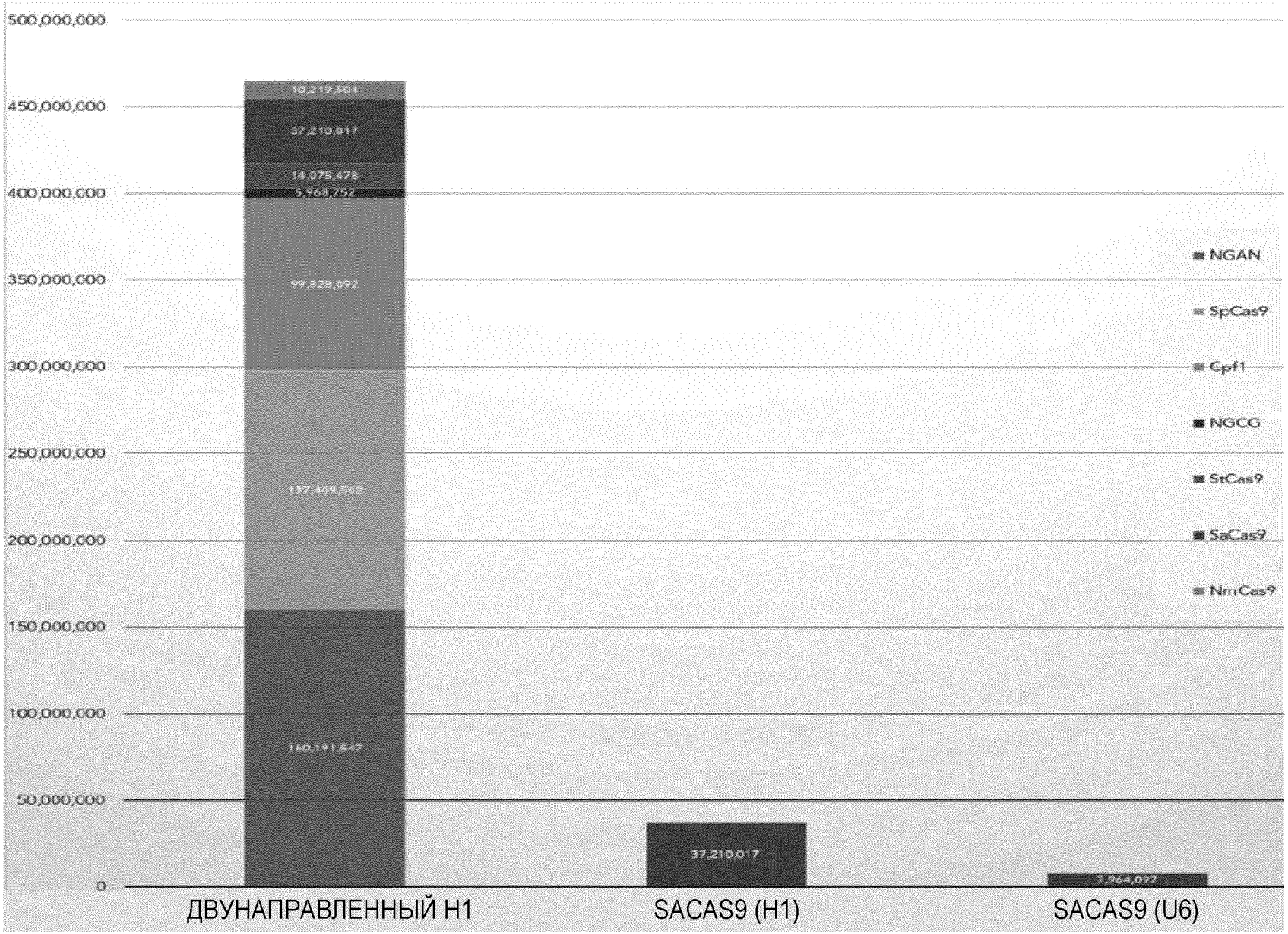
ФИГ.22D (продолжение)



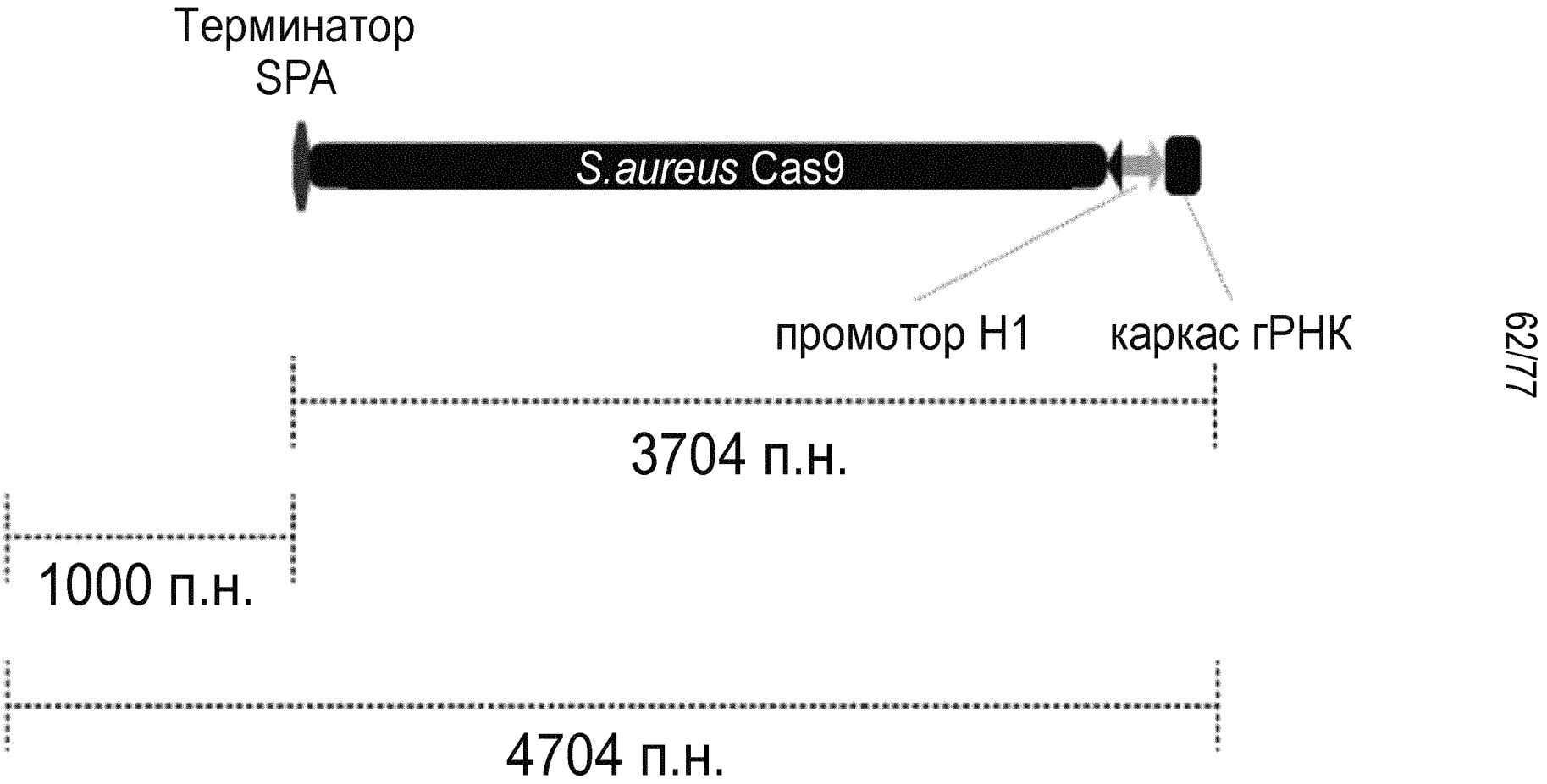
ФИГ.22E


SpCas9
4.1 т.п.н.
StCas9
3.6 т.п.н.
NmCas9
3.3 т.п.н.
SaCas9
3.2 т.п.н.
AsCpf1
4.1 т.п.н.
LbCpf1
3.8 т.п.н.
SpCas9 (D1135E)
4.1 т.п.н.
SpCas9 (VQR)
4.1 т.п.н.
SpCas9 (VRER)
4.1 т.п.н.
eSpCas9
4.1 т.п.н.
SpCas9-HF1
4.1 т.п.н.

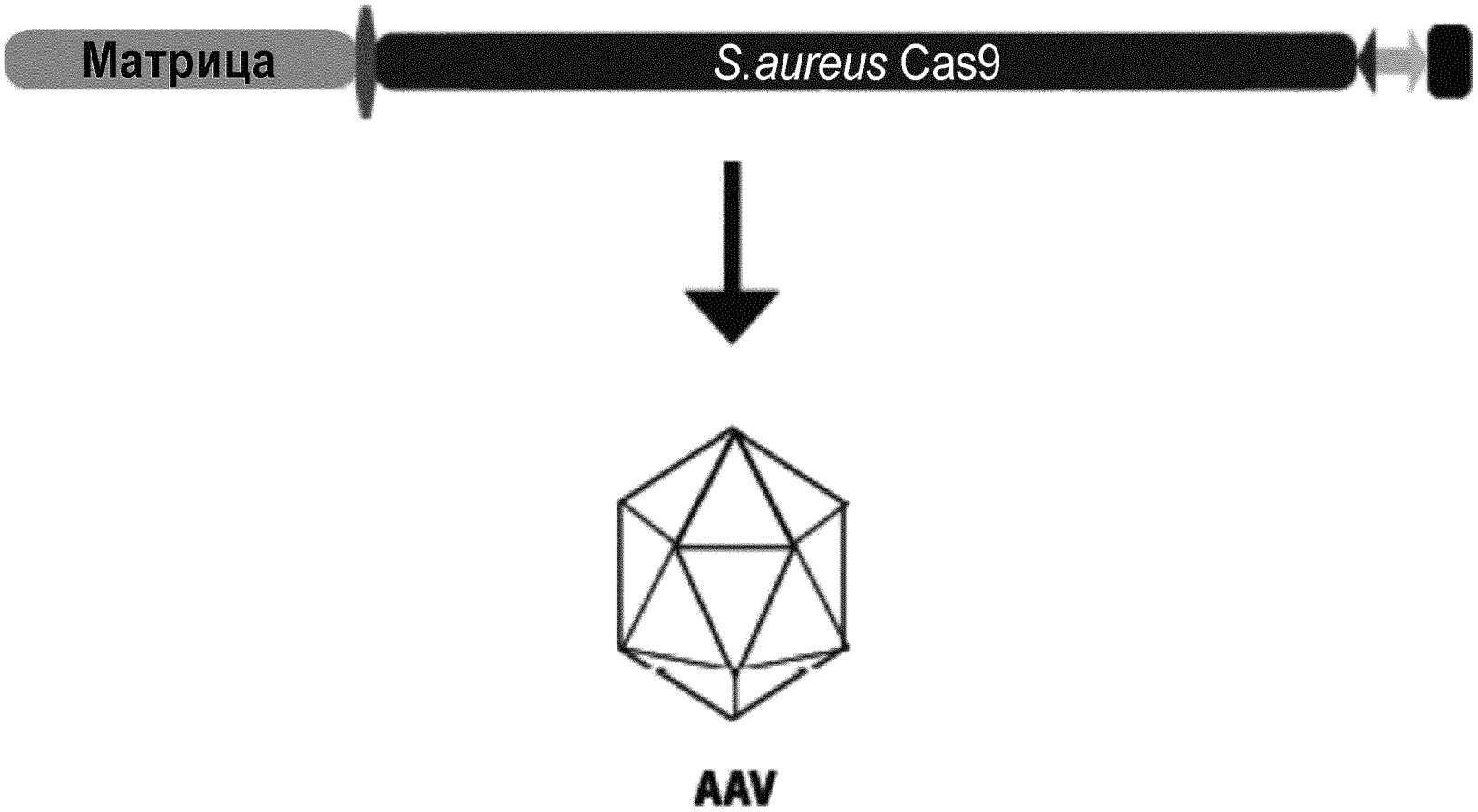
ФИГ.22F


500,000,000
450,000,000
400,000,000
350,000,000
300,000,000
250,000,000
200,000,000
150,000,000
100,000,000
50,000,000
0
10,219,504
37,210,017
14,075,478
5,168,752
99,828,092
137,469,562
160,191,547
37,210,017
7,964,092
NGAN
SpCas9
Cpf1
NGCG
StCas9
SaCas9
NmCas9
ДВУНАПРАВЛЕННЫЙ H1
SACAS9 (H1)
SACAS9 (U6)

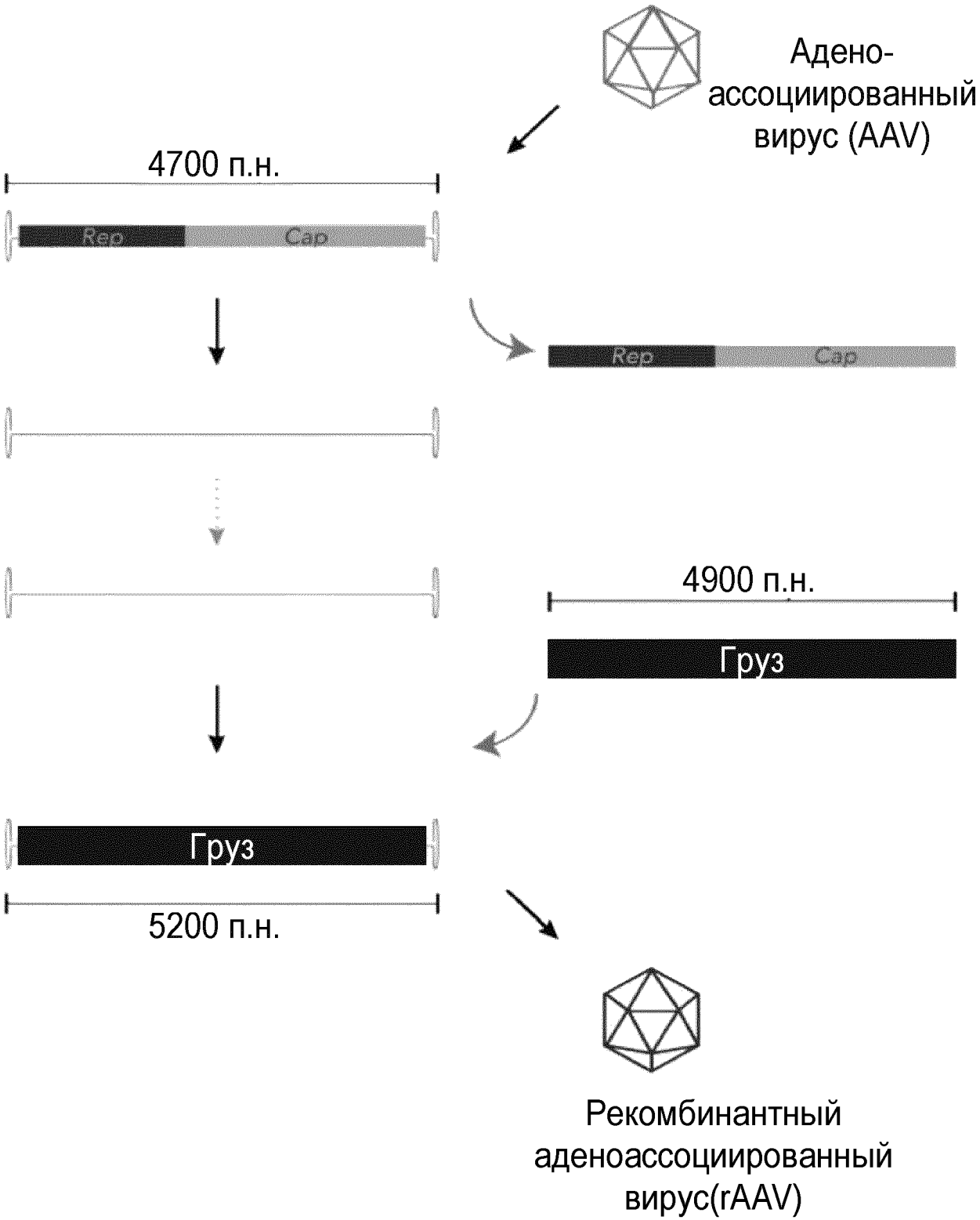
ФИГ.22G


Терминатор
SPA
S.aureus Cas9
промотор H1
каркас гРНК
3704 п.н.
1000 п.н.
4704 п.н.

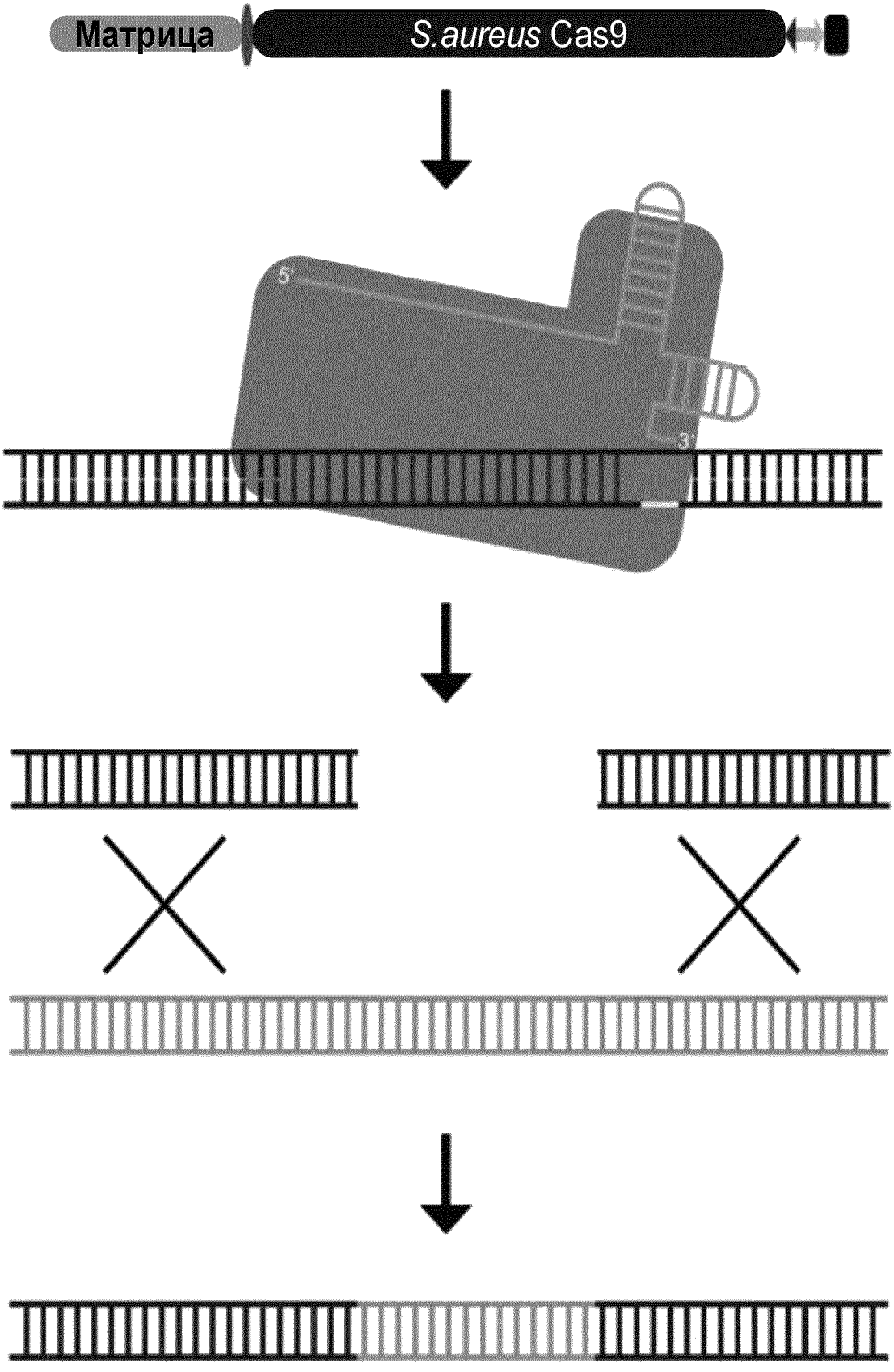
ФИГ.22H



ФИГ.22I

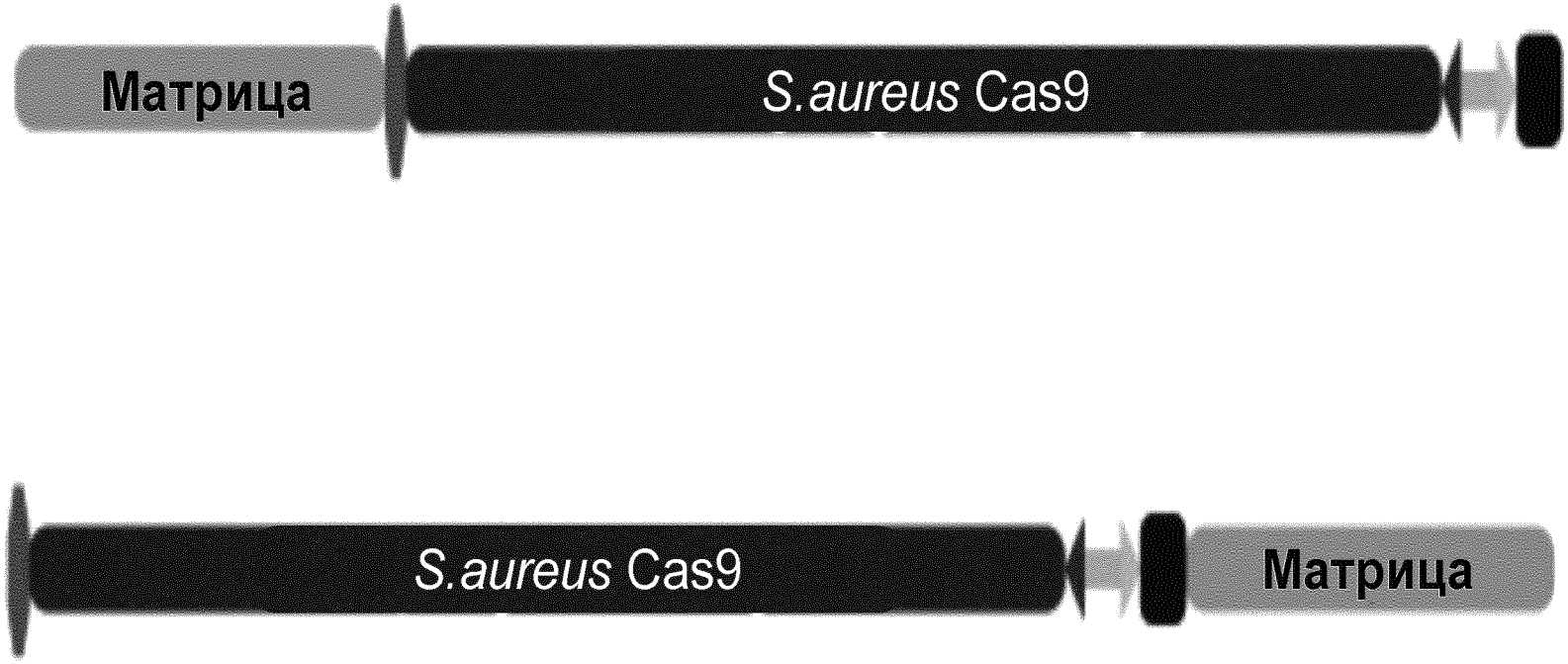

Адено-
ассоциированный
вирус (AAV)
4700 п.н.
Rep
Cap
Rep
Cap
4900 п.н.
Груз
Груз
5200 п.н.
Рекомбинантный
аденоассоциированный
вирус(rAAV)

ФИГ.23



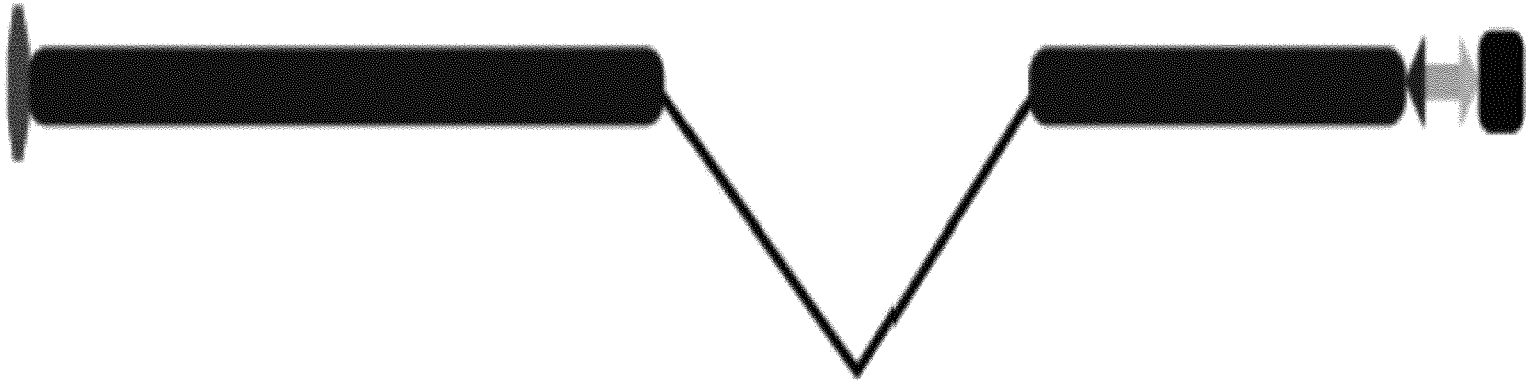
ФИГ.24А

Некоторые потенциальные конфигурации:



ФИГ.24B

Мультиэкзонный Cas9:



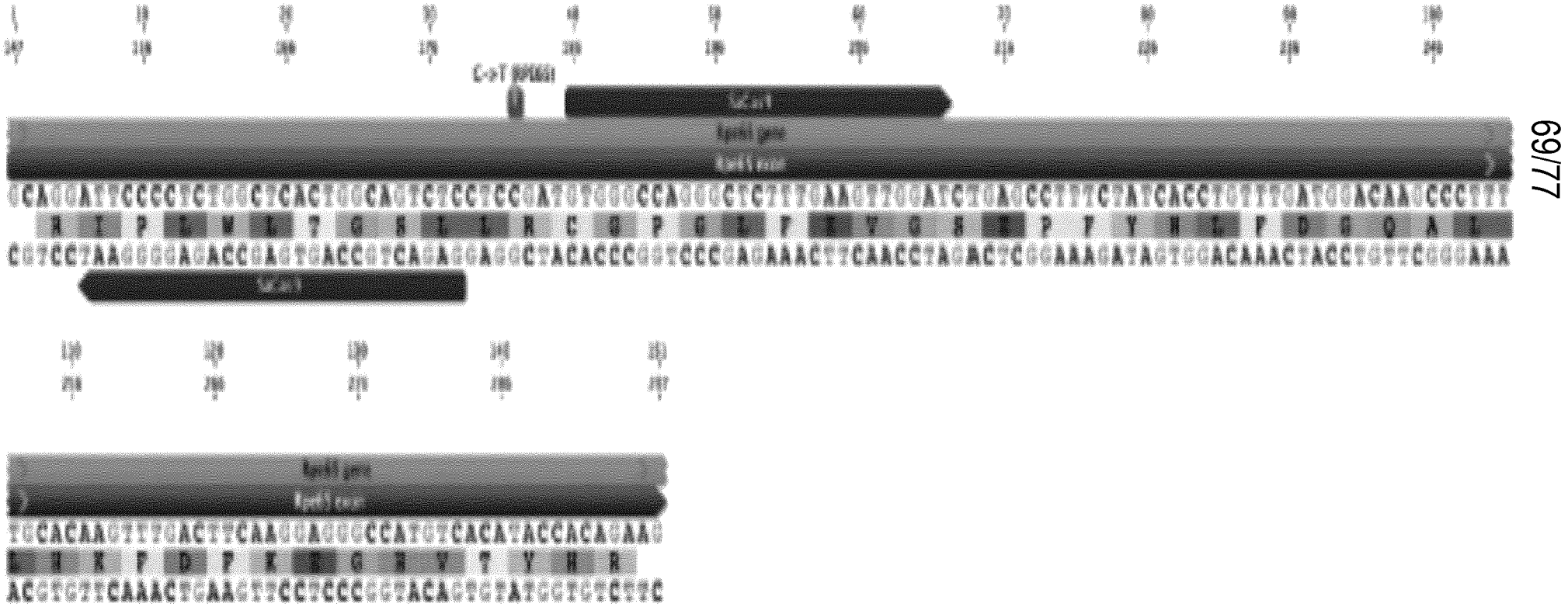
Донор интронов

ФИГ.25А

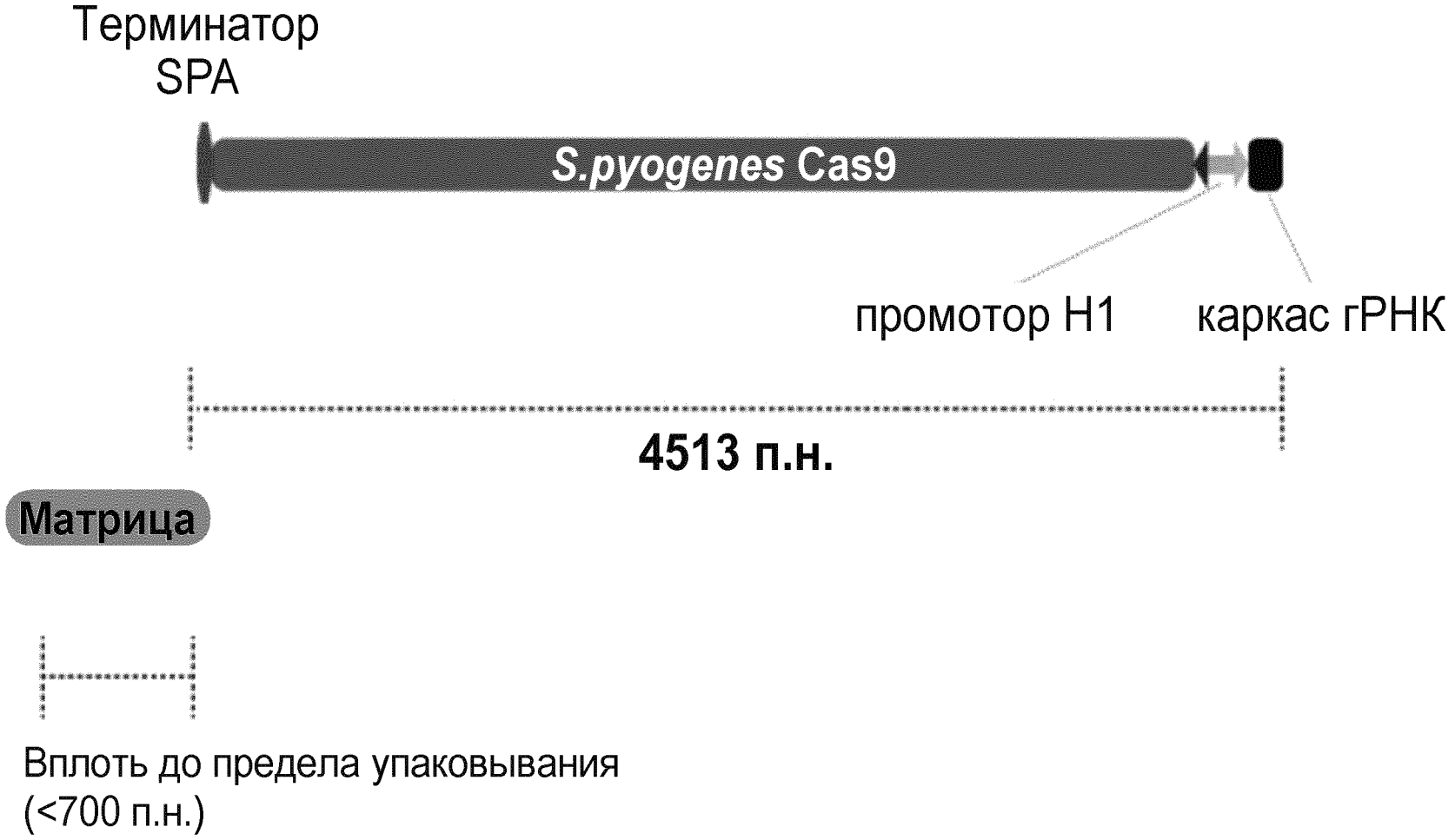
Несколько гРНК/никаза/HDR

ФИГ.25B

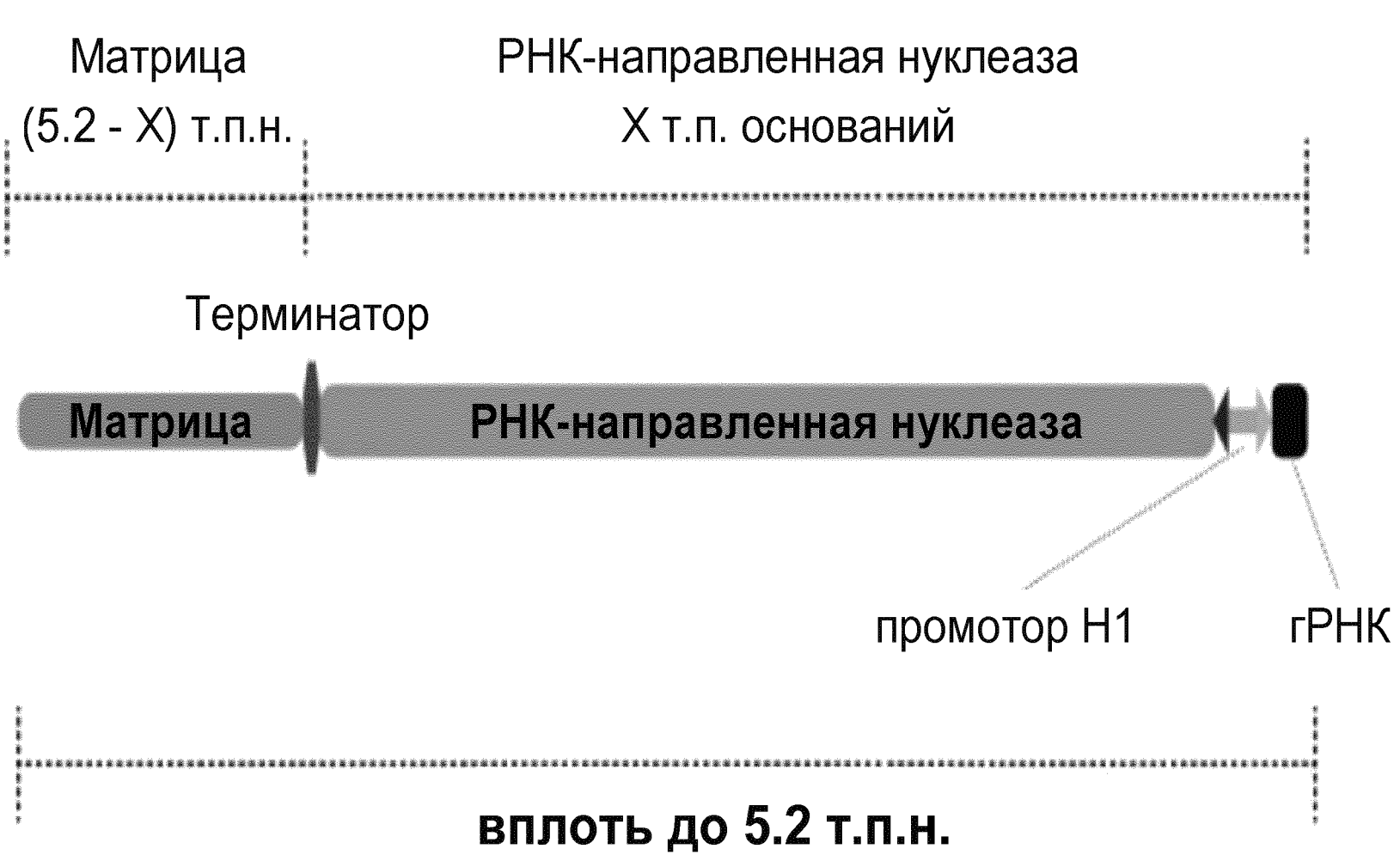
Пример RPE65: rd12 с никазой SaCas9



ФИГ.26A

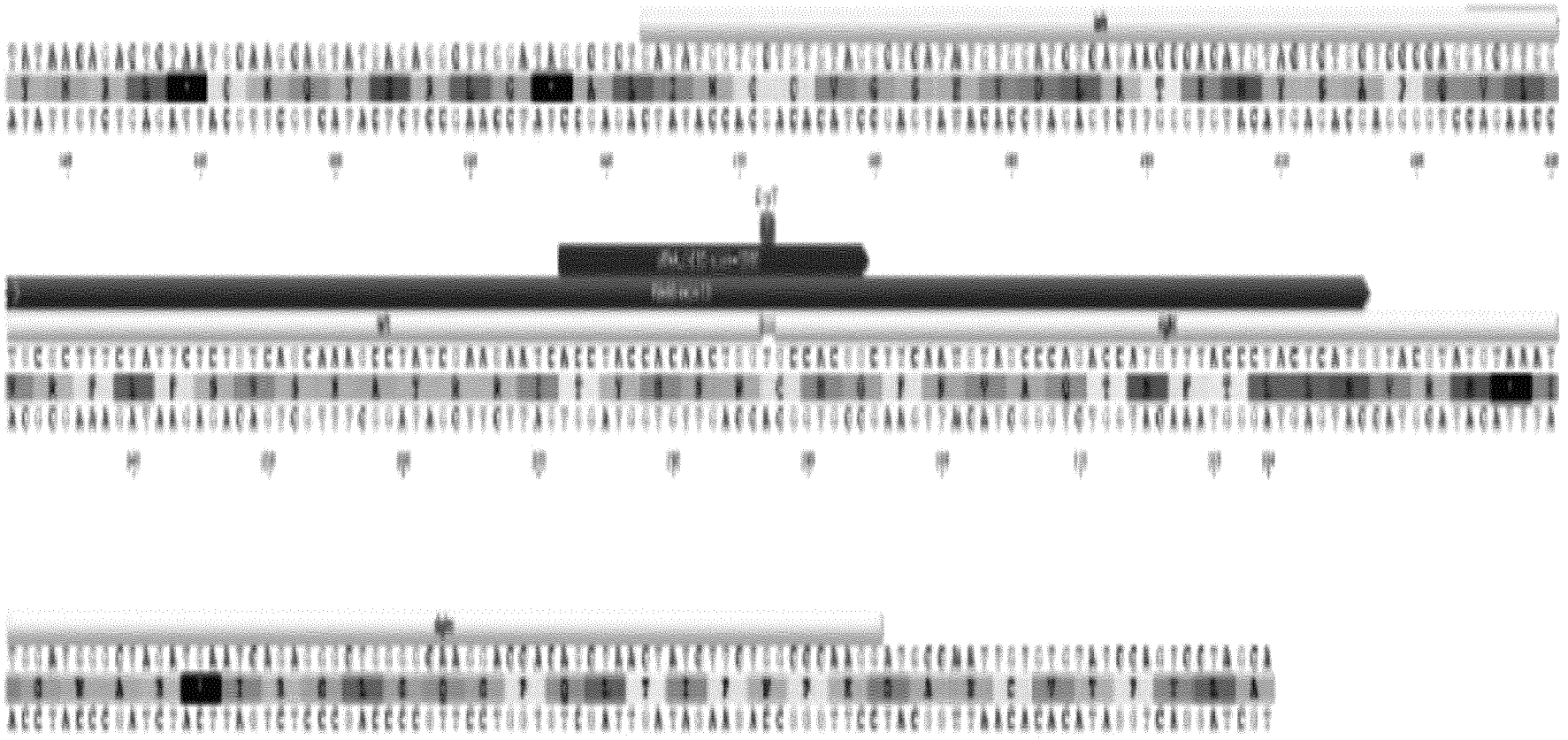

Терминатор
SPA
S.pyogenes Cas9
промотор H1
каркас гРНК
4513 п.н.
Матрица
Вплоть до предела упаковывания
(<700 п.н.)

ФИГ.26В


Матрица
(5.2 - X) т.п.н.
РНК-направленная нуклеаза
X т.п. оснований
Терминатор
Матрица
РНК-направленная нуклеаза
промотор H1
гРНК
вплоть до 5.2 т.п.н.

ФИГ.27

Пример rd10 с SpCas9



ФИГ.28

Клонирование HDR-матриц по методу Гибсона - продукт расщепления либо KpnI (слева), либо HpaI (справа)

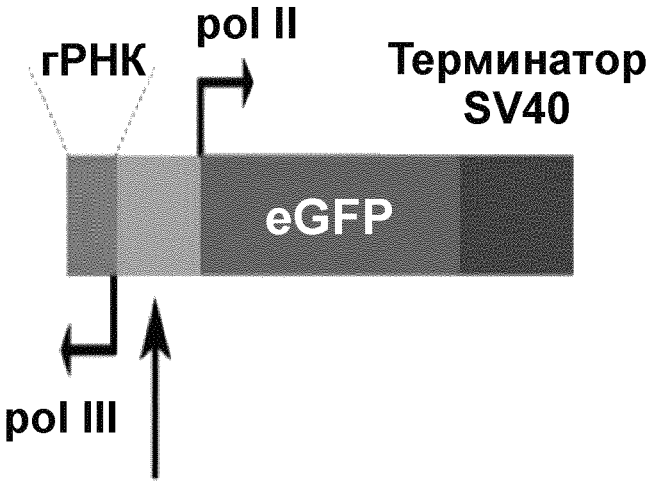
KpnI (N = матрица HDR)

GAGAGTGCACCATAGCGGCCGCGNNNGTACCCACACAAAAAACCAACACACAG

HpaI (N = матрица HDR)

GGCACCGAGTCGGTGCTTTTTTGTTNNNNNNAACGCGGCCGCCTAGAGTCGAC

ФИГ.29А



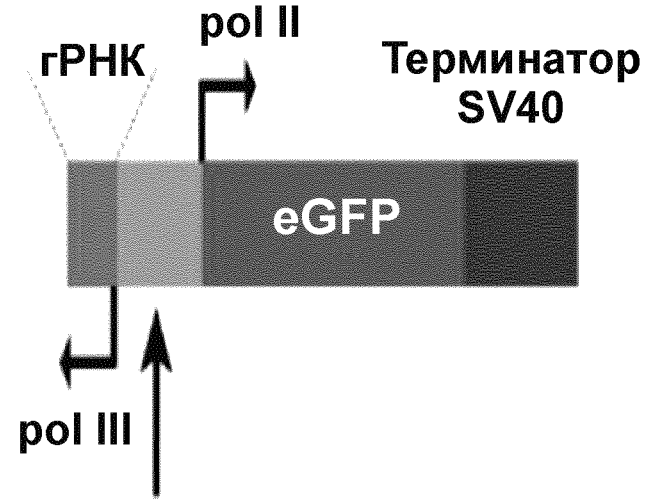
идентифицированные последовательности двунаправленных промоторов

Ген нкРНК	Ген белка	Расстояние
RPPH1	PARP2	230 п.н.
SRP	RPS29	233 п.н.
7sk1	GSTA4	239 п.н.
SNAR-G1	CGB1	308 п.н.
SNAR	CGB2	308 п.н.
RMRP	CCDC107	361 п.н.
tRNA(Lys)	ALOXE3	376 п.н.
RNU6-9	MED16	412 п.н.
tRNA (Gly)	DPP9	484 п.н.
SNORD13	C8orf41	847 п.н.

ФИГ.29B

Двунаправленная экспрессия GFP
9
8
7
6
5
4
3
2
1
0
отр. контроль
hH1Koz
h5k1
biSRP
biCGB1
biCGB2
biAloxE3
biDPP9
biTHEM
biORF

ФИГ.30А



идентифицированные последовательности двунаправленных промоторов

Ген нкРНК	Ген белка	Расстояние	Организм
RPPH1	PARP2	230 п.н.	Человек
RPPH1	PARP2	172 п.н.	Мышь
RPPH1	PARP2	201 п.н.	Крыса

ФИГ.30B