

(19)



**Евразийское
патентное
ведомство**

(11)

027693

(13)

B1

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ

(45) Дата публикации
и выдачи патента: **2017.08.31**

(21) Номер заявки: **201170492**

(22) Дата подачи: **2009.09.26**

(51) Int. Cl. *C12N 15/52* (2006.01)
C12N 5/10 (2006.01)
C12N 15/867 (2006.01)

(54) РЕКОМБИНАНТНЫЙ РЕПЛИКАЦИОННО-КОМПЕТЕНТНЫЙ РЕТРОВИРУС С ЦИТОЗИНДЕЗАМИНАЗНОЙ АКТИВНОСТЬЮ ДЛЯ ГЕННОЙ ТЕРАПИИ НАРУШЕНИЙ КЛЕТОЧНОЙ ПРОЛИФЕРАЦИИ

(31) **61/100,666; 61/120,618; 61/186,823**

(32) **2008.09.26; 2008.12.08; 2009.06.13**

(33) **US**

(43) **2011.10.31**

(86) **PCT/US2009/058510**

(87) **WO 2010/045002 2010.04.22**

(71)(73) Заявитель и патентовладелец:
ТОКАДЖЕН ИНК. (US)

(72) Изобретатель:
**Груббер Гарри Е., Джолли Дуглас, Перес
Омар, Логг Кристофер Р. (US)**

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

(56) PHILIPPE ERBS et al. In Vivo Cancer Gene Therapy by Adenovirus-mediated Transfer of a Bifunctional Yeast Cytosine Deaminase/Uracil Phosphoribosyltransferase Fusion Gene. *CANCER RESEARCH*. 2000, vol. 60, no. 14, p. 3813-3822. See the whole document, especially, abstract, p. 3814-3818

GenBank Accession Number 1YSB_A, "Chain A, Yeast Cytosine Deaminase Triple Mutant", 24-SEP-2008, http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=protein&dopt=GenPept&RID=YENPGVH016&log%24=protalign&blast_rank=1&list_uids=67464318

GenBank Accession Number AAG33626, "Cytosine deaminase-uracil phosphoribosyltransferase fusion protein [synthetic construct]", 21-NOV-2000, http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=protein&dopt=GenPept&RID=YENPGVMH016&log%24=protalign&blast_rank=3&list_uids=11245466

(57) Изобретение относится к рекомбинантному репликационно-компетентному ретровирусу и его применению. Рекомбинантный ретровирус по настоящему изобретению способен трансформировать клетку-мишень человека полинуклеотидом, кодирующим полипептид с цитозиндезаминазной активностью и повышенной стабильностью в клетках человека, который катализирует превращение экзогенного 5-фторцитозина в 5-фторурацил, обеспечивая цитотоксическое действие. Указанный рекомбинантный ретровирус применяется с 5-фторцитозином для получения лекарственного средства для лечения индивида с нарушением клеточной пролиферации.

B1**027693****027693****B1**

Перекрестная ссылка на родственные заявки

По заявке на данное изобретение испрашивается приоритет предварительной заявки США с серийным номером 61/100666, поданной 26 сентября 2008 г., предварительной заявки США с серийным номером 61/120618, поданной 8 декабря 2008 г., и предварительной заявки США с серийным номером 61/186823, поданной 13 июня 2009 г., описания которых включены в настоящий документ путем ссылки.

Предпосылки к созданию изобретения

Дрожжевая, или бактериальная, цитозиндезаминаза превращает нетоксичное антибиотическое пролекарство 5-FC в цитотоксическое химиотерапевтическое средство 5-фторурацил (5-FU). У людей (и млекопитающих вообще) не известно наличие природного гена, кодирующего фермент со значительной цитозиндезаминазной активностью. Цитозиндезаминаза дрожжей или бактерий получила признание в лечении злокачественных заболеваний, путем использования доставки гена и вирусных векторов для доставки этого фермента, с последующей обработкой 5-FC, который затем превращался под действием этого фермента в цитотоксическое лекарственное средство (Miller et al., Can Res 62:773-780 2002; Kievit et al., Can Res 59:1417-1421, 1999).

Краткое изложение сущности изобретения

Цитозиндезаминазу или слитую конструкцию по настоящему изобретению доставляют с использованием системы для доставки генов (GDS). В другом аспекте полинуклеотид, кодирующий цитозиндезаминазу, доставляют с помощью GDS, которая представляет собой вирусный вектор или вектор, производный вируса. Вирусный вектор может быть реплицирующимся или нереплицирующимся и может представлять собой аденовирусный вектор, коревой вектор, вектор вируса герпеса, ретровирусный вектор (в том числе лентивирусный вектор), рабдовирусный вектор, такой как вектор вируса визиккулярного стоматита, реовирусный вектор, вектор вируса Сенеки Валли, поксовирусный вектор (в том числе векторы, полученные из вируса оспы животных или коровьей оспы), парвовирусный вектор (в том числе AAV вектор), альфа-вирусный вектор или другой вирусный вектор, известный специалисту в данной области. В одном варианте осуществления вирусный вектор может представлять собой репликативно-компетентный ретровирусный вектор, способный инфицировать реплицирующиеся клетки млекопитающих.

Репликативно-компетентный ретровирусный вектор может содержать орторетровирус или, более характерно, гамма-ретровирусный вектор. В одном аспекте репликационно-компетентный ретровирусный вектор содержит внутренний сайт связывания рибосомы (IRES) 5' с полинуклеотидом, кодирующим цитозиндезаминазу по настоящему изобретению. В одном варианте осуществления полинуклеотид, кодирующий цитозиндезаминазу, находится в положении 3' по отношению к ENV полинуклеотиду ретровирусного вектора.

Настоящее изобретение относится к рекомбинантному репликационно-компетентному ретровирусу (RCR), содержащему: ретровирусный белок GAG; ретровирусный белок POL; оболочку ретровируса; ретровирусный полинуклеотид, содержащий последовательности длинных концевых повторов (LTR) на 3'-конце полинуклеотидной последовательности ретровируса, промоторную последовательность на 5'-конце ретровирусного полинуклеотида, указанный промотор подходит для экспрессии в клетках млекопитающих, домен нуклеиновой кислоты gag, домен нуклеиновой кислоты pol и домен нуклеиновой кислоты env; кассету, содержащую внутренний сайт связывания рибосомы (IRES) оперативно связанный с гетерологичным полинуклеотидом, где указанная кассета расположена от 5' к 3' LTR и 3' к домену нуклеиновой кислоты env, кодирующему оболочку ретровируса; и последовательности, действующие в cis-положении, необходимые для обратной транскрипции, упаковки и интеграции в клетку-мишень, где RCR сохраняет более высокую репликативную компетентность после 6 пассажей по сравнению с вектором pACE (SEQ ID NO: 21). В одном варианте осуществления полинуклеотидная последовательность ретровируса получена из вируса лейкоза мышей (MLV), вируса мышинного лейкоза Молони (MoMLV), вируса лейкоза кошачьих или вируса лейкоза гиббонов (GALV). В другом варианте осуществления MLV представляет собой амфотропный MLV. Еще в одном варианте осуществления ретровирус представляет собой онкоретровирус. Еще в одном варианте осуществления клетка-мишень представляет собой клетку, имеющую нарушение клеточной пролиферации. Нарушение клеточной пролиферации может быть выбрано из группы, состоящей, но не ограничивающейся этим, из рака легких, рака толстой и прямой кишки, рака груди, рака предстательной железы, рака мочевыводящих путей, рака матки, рака головного мозга, рака тканей головы и шеи, рака поджелудочной железы, меланомы, рака желудка и рака яичников, ревматоидного артрита и других аутоиммунных заболеваний. В одном варианте осуществления промотор содержит CMV промотор, имеющий последовательность SEQ ID NO: 19, 20 или 22 от нуклеотида 1 до примерно нуклеотида 582 и может включать модификацию одного или нескольких оснований нуклеиновых кислот, и который способен направлять и инициировать транскрипцию. Еще в одном варианте осуществления промотор содержит последовательность SEQ ID NO: 19, 20 или 22 от нуклеотида 1 до примерно нуклеотида 582. В дополнительном варианте осуществления промотор содержит полинуклеотид с доменом CMV-R-U5. В одном варианте осуществления домен CMV-R-U5 содержит ранний промотор из цитомегаловируса человека, связанный с областью MLV R-U5. Еще в одном варианте осуществления полинуклеотид с доменом CMV-R-U5 содержит последовательность SEQ ID NO: 19, 20 или 22

примерно от нуклеотида 1 до примерно нуклеотида 1202 или последовательности, которые по меньшей мере на 95% идентичны последовательности SEQ ID NO: 19, 20 или 22, где указанный полинуклеотид активизирует транскрипцию молекулы нуклеиновой кислоты, оперативно связанной с ним. В другом варианте осуществления gag и pol полинуклеотида получены из онкоретровируса. Домен gag нуклеиновой кислоты может содержать последовательность примерно от нуклеотида номер 1203 до примерно нуклеотида 2819 последовательности SEQ ID NO: 19 или последовательность, по меньшей мере на 95%, 98, 99 или 99,8% идентичную ей. Домен pol может содержать последовательность примерно от нуклеотида номер 2820 до примерно нуклеотида 6358 последовательности SEQ ID NO: 19 или последовательность, по меньшей мере на 95, 98, 99 или 99,9% идентичную ей. В одном варианте осуществления домен env кодирует амфотерный белок env. Домен env может содержать последовательность примерно от нуклеотида номер 6359 до примерно нуклеотида 8323 последовательности SEQ ID NO: 19 или последовательность, по меньшей мере на 95, 98, 99 или 99,8% идентичную ей. Домен IRES может представлять собой любой IRES, однако в одном варианте осуществления IRES получают из вируса энцефаломиокардита. В дополнительном варианте осуществления IRES содержит последовательность примерно от нуклеотида номер 8327 до примерно нуклеотида 8876 последовательности SEQ ID NO: 19 или последовательность, по меньшей мере на 95, 98 или 99% идентичную ей. Еще в одном варианте осуществления гетерологичный полинуклеотид содержит полинуклеотид, имеющий последовательность SEQ ID NO: 3, 5, 11, 13, 15 или 17. В дополнительном варианте осуществления гетерологичная последовательность кодирует полипептид, содержащий последовательность SEQ ID NO: 4. Гетерологичная нуклеиновая кислота является оптимизированной для кодонов человека и кодирует полипептид последовательности SEQ ID NO: 4. В дополнительном варианте осуществления гетерологичная нуклеиновая кислота содержит последовательность SEQ ID NO: 19 примерно от нуклеотида номер 8877 примерно до 9353. В одном варианте осуществления 3' LTR получен из онкоретровируса. В дополнительном варианте осуществления 3' LTR содержит домен U3-R-U5. Еще в одном варианте осуществления 3' LTR содержит последовательность SEQ ID NO: 19 примерно от нуклеотида 9405 примерно до 9998 или последовательность, которая по меньшей мере на 95, 98 или 99,5% идентична ей.

Подробные описания одного или нескольких вариантов осуществления настоящего изобретения изложены в сопровождающих чертежах и в описании, приведенном ниже. Другие особенности, задачи и преимущества будут очевидны из описания и чертежей и формулы изобретения.

Краткое описание чертежей

На фиг. 1A-1D представлено (а) схематическое изображение рекомбинантного ретровирусного вектора по настоящему изобретению; (b) плазмидная карта полинуклеотида по настоящему изобретению; (с и d) последовательность полинуклеотида по настоящему изобретению (SEQ ID NO: 19).

На фиг. 2A-2D представлены схемы воспроизведения различных вариантов осуществления по настоящему изобретению, включая полипептиды с CD, OPRT и UPRT активностью.

На фиг. 3 показано, что наблюдаются более высокие уровни белка yCD2 по сравнению с белком yCD дикого типа в инфицированных клетках U-87.

На фиг. 4 показана стабильность вектора, содержащего полипептид CD по настоящему изобретению, и сравнение с другими векторами по настоящему изобретению.

На фиг. 5A-5B показано, что цитозиндезаминазная активность и вектор по настоящему изобретению обеспечивают сравнимую или лучшую экспрессию, и, следовательно, уничтожение инфицированных клеток крысы RG2 (5A) или U87 (5B) по сравнению с yCD активностью дикого типа (T5.0007), когда инфицированные клетки подвергаются воздействию повышенных уровней 5-FC.

На фиг. 6 показано, что специфическая активность в инфицированных клетках U87 T5.0002 (yCD2) выше, чем T5.0001 (частично оптимизированная yCD), которая выше, чем T5.0007 (wt yCD).

Одинаковые обозначения на различных чертежах означают одинаковые элементы.

Подробное описание

Используемые в контексте настоящего изобретения и в прилагаемой формуле изобретения формы единственного числа включают обозначения множественного числа, если из контекста явным образом не следует иное. Таким образом, например, ссылка на "клетку" означает множество таких клеток, а ссылка на "средство" включает ссылку на одно или несколько средств, известных специалистам в данной области, и так далее.

Также использование "или" означает "и/или", если не указано иное. Аналогичным образом, "содержат", "содержит", "содержащий", "включают", "включает" и "включающий" являются взаимозаменяемыми и не предназначены для ограничения.

Далее должно быть понятно, что в тех случаях, когда в описаниях различных вариантов осуществления используется термин "включающий", специалисту в данной области было бы понятно, что в некоторых особых случаях вариант осуществления альтернативно может быть описан с использованием формулировки "по существу состоящий из" или "состоящий из".

Если не определено иное, все технические и научные термины, используемые в настоящем описании, имеют те же значения, обычно понимаемые специалистом в области, к которой относится настоящее изобретение. Хотя способы и материалы, сходные или эквивалентные описанным в настоящем до-

кодирующей белок, такой как, но не только, регуляторные или промоторные последовательности. Мутация в гене может быть "молчащей", т.е. не отражающейся в изменении аминокислот при экспрессии, приводя к варианту гена с "консервативной последовательностью". В основном это происходит, когда одна аминокислота соответствует более чем одному кодону.

Неограничивающие примеры модифицированной аминокислоты включают гликозилированную аминокислоту, сульфатированную аминокислоту, пренилированную (например, фарнезилированную, геранилгеранилированную) аминокислоту, ацетилированную аминокислоту, ацилированную кислоту, пегилированную аминокислоту, биотинилированную аминокислоту, карбоксилированную аминокислоту, фосфорилированную аминокислоту и подобное. Ссылки, достаточные для инструктирования специалиста в отношении модификации аминокислот, в большом количестве встречаются повсюду в литературе. Примеры протоколов находятся у Walker (1998), Protein Protocols on CD-ROM (Humana Press, Towata, N.J.).

Для иллюстрации, последовательности нуклеиновой кислоты, кодирующие UPRT-азы *E.coli* (Anderson et al., 1992, Eur. J. Biochem 204, 51-56), *Lactococcus lactis* (Martinussen and Hammer, 1994, J. Bacteriol. 176, 6457-6463), *Mycobacterium bovis* (Kim et al., 1997, Biochem Mol. Biol. Int. 41, 1117-1124) и *Bacillus subtilis* (Martinussen et al., 1995, J. Bacteriol. 177, 271-274), могут быть использованы в контексте настоящего изобретения. Однако использование UPRT-азы дрожжей и, в частности, кодируемой геном *FUR1 S.cerevisiae*, последовательность которого описана у Kern et al. (1990, Gene 88, 149-157), включено в настоящее описание для ссылки. Чтобы обратить внимание, последовательности генов и последовательности соответствующих UPRT-аз могут быть найдены в литературе и специализированных базах данных (SWISSPROT, EMBL, Genbank Medline, etc.).

"Белок" или "полипептид" - термины, которые используются в настоящем описании взаимозаменяемо, содержат одну или несколько цепей химических структурных элементов, называемых аминокислотами, которые связаны вместе химическими связями, называемыми пептидными связями.

"Фермент" означает любое вещество, предпочтительно состоящее полностью или в значительной степени из белка, которое катализирует или активизирует более или менее специфично одну или несколько химических или биохимических реакций. Термин "фермент" также может относиться к каталитическому полинуклеотиду (например, РНК или ДНК). "Нативный" белок, фермент, полинуклеотид, ген или клетка, или белок, фермент, полинуклеотид, ген или клетка "дикого типа" означают белок, фермент, полинуклеотид, ген или клетку, которые встречаются в природе.

Термин "оперативно связанный" или "оперативно ассоциированный" относится к функциональной связи или ассоциации между регуляторной последовательностью и полинуклеотидом, регулируемым этой регуляторной последовательностью или между двумя различными полипептидами или полинуклеотидами, кодирующими такие полипептиды.

"Консервативная аминокислотная замена" или просто "консервативные варианты" конкретной последовательности относятся к замене одной аминокислоты или ряда аминокислот, по существу, на идентичные аминокислотные последовательности. Специалисту в данной области будет понятно, что индивидуальные замены, делеции или добавления, которые изменяют, добавляют или удаляют единичную аминокислоту или процентное содержание аминокислот в кодируемой последовательности, приводят к "консервативным вариантам", в которых изменения приводят к делеции аминокислоты, добавлению аминокислоты или замене аминокислоты химически сходной аминокислотой.

Таблицы консервативных замен, представляющие функционально сходные аминокислоты, хорошо известны из уровня техники. Например, одна группа консервативных замен включает аланин (А), серин (S) и треонин (Т). Другая группа консервативных замен включает аспарагиновую кислоту (D) и глутаминовую кислоту (E). Другая группа консервативных замен включает аспарагин (N) и глутамин (Q). Еще одна группа консервативных замен включает аргинин (R) и лизин (K). Другая группа консервативных замен включает изолейцин (I), лейцин (L), метионин (M) и валин (V). Другая группа консервативных замен включает фенилаланин (F), тирозин (Y) и триптофан (W).

Таким образом, "консервативные аминокислотные замены" перечисленных полипептидных последовательностей (например, SEQ ID NO: 2, 4, 6, 8, 10, 11, 12, 13 и т.д.) включают замены процентного содержания, обычно менее 10%, аминокислот полипептидной последовательности, консервативно выбранной аминокислотой той же группы консервативных замен.

Соответственно, консервативно замещенный вариант полипептида по настоящему изобретению может содержать 100, 75, 50, 25 или 10 замен консервативно замещенных вариантов той же группы консервативных замен.

"Активность" фермента представляет собой критерий его способности катализировать реакцию, т.е. "функционировать", и может быть выражена как скорость, с которой образуется продукт реакции. Например, активность фермента может быть представлена как количество продукта, образованного за единицу времени или на единицу фермента (например, концентрация или масса) или в значениях аффинности или констант диссоциации. Используемая взаимозаменяемо в контексте настоящего изобретения "цитозиндезаминная активность", "биологическая активность цитозиндезаминазы" или "функциональная активность цитозиндезаминазы" относится к активности, проявляемой белком или полипептидом

цитозиндезаминазы по настоящему изобретению, в отношении субстрата цитозиндезаминазы, как определено *in vivo* или *in vitro*, в соответствии со стандартными методиками. Анализы для измерения цитозиндезаминазной активности известны из уровня техники. Например, цитозиндезаминазную активность можно измерить путем определения скорости превращения 5-FC в 5-FU или цитозина в урацил. Определение 5-FC, 5-FU, цитозина и урацила может быть выполнено с помощью хроматографии и других способов, известных из уровня техники.

Специалисту в данной области будет понятно, что многие консервативные варианты конструкций нуклеиновой кислоты, которые описаны, дают на выходе функционально идентичную конструкцию. Например, как рассмотрено выше, вследствие вырожденности генетического кода "молчащие замены" (т.е. замены в последовательности нуклеиновой кислоты, которая в результате не приводит к изменению в кодируемом полипептиде) означали свойство каждой последовательности нуклеиновой кислоты, которая кодирует аминокислоту. Специалистам в данной области будет понятно, что вследствие вырожденности генетического кода может быть получено множество нуклеотидных последовательностей, кодирующих модифицированные цитозиндезаминазные полипептиды, некоторые из которых несут существенную идентичность с последовательностями нуклеиновых кислот, явным образом описанных в настоящем изобретении. Например, кодоны AGA, AGG, CGA, CGC, CGG и CGU - все кодируют аминокислоту аргинин. Таким образом, в каждом положении в нуклеиновых кислотах по настоящему изобретению, где аргинин определен кодоном, этот кодон может быть изменен в любой из соответствующих кодонов, описанных выше, без изменения кодируемого полипептида. Понятно, что U в последовательности РНК соответствует Т в последовательности ДНК.

Аналогично, "консервативные аминокислотные замены", в одной или нескольких аминокислотах в аминокислотной последовательности, замещенные другими аминокислотами с очень похожими свойствами, также легко идентифицируются как в высокой степени сходные с описанной конструкцией. Такие консервативные варианты каждой описанной последовательности являются особенностью полипептидов, представленных в настоящем описании.

"Консервативные варианты" представляют собой белки или ферменты, в которых заданный аминокислотный остаток был заменен без изменения в целом конформации и функции белка или фермента, включая, но не ограничиваясь этим, замену аминокислоты аминокислотой, обладающей похожими свойствами, включая полярный или неполярный характер, размер, форму и заряд. Аминокислоты, отличные от аминокислот, указанных в качестве консервативных, могут отличаться в белке или ферменте так, чтобы сходство процентного состава белка или аминокислотной последовательности между любыми двумя белками аналогичной функции могли варьировать и могли составлять, например, по меньшей мере 30%, по меньшей мере 50%, по меньшей мере 70%, по меньшей мере 80% или по меньшей мере 90%, как определено в соответствии со схемой выравнивания. Как изложено в настоящем описании, "сходство последовательностей" означает степень, до которой нуклеотидные или белковые последовательности являются родственными. Степень сходства между двумя последовательностями может быть основана на проценте идентичности последовательности и/или консервативности. "Идентичность последовательностей" в настоящем описании означает степень, до которой две нуклеотидные или аминокислотные последовательности являются инвариантными. "Выравнивание последовательностей" означает процесс выстраивания в одну линию двух или более последовательностей для достижения максимальных уровней идентичности (и, в случае аминокислотных последовательностей, консервативности) для целей определения степени сходства. Многочисленные способы выравнивания последовательностей и определения сходства/идентичности известны из уровня техники, например, такие как кластерный метод, в котором сходство основано на алгоритме MEGALIGN, а также BLASTN, BLASTP и FASTA (Lipman and Pearson, 1985; Pearson and Lipman, 1988). При использовании всех этих программ предпочтительными параметрами настройки являются те, которые в результате приводят к наибольшему сходству последовательностей. Например, "идентичность" или "процент идентичности" в отношении конкретной пары выравниваемых аминокислотных последовательностей может относиться к проценту идентичности аминокислотных последовательностей, полученному с помощью анализа ClustalW (версия W 1.8, доступная от Европейского института биоинформатики, Cambridge, UK), подсчета числа идентичных соответствий в выравнивании и деления этого числа идентичных соответствий на наибольшее из следующих показателей: (i) длина выравниваемых последовательностей и (ii) 96, и используя следующие параметры по умолчанию ClustalW для достижения медленных/точных попарных выравниваний-штраф за открытый разрыв: 10; штраф за удлинение разрыва: 0,10; матрица массы белка: Gonnet series; матрица массы ДНК: IUB; Toggle Slow/Fast попарные выравнивания = SLOW или FULL Alignment.

Две последовательности являются "оптимально выровненными", когда они выровнены по баллам сходства с использованием определенной матрицы аминокислотных замен (например, BLOSUM62), штрафа за разрыв и штрафа за удлинение разрыва, чтобы получить наиболее высокий балл, возможный для этой пары последовательностей. Матрицы аминокислотных замен и их использование в количественном определении сходства между двумя последовательностями хорошо известны из уровня техники и описаны, например, у Dayhoff et al. (1978), "A model of evolutionary change in proteins" в "Atlas of Protein Sequence and Structure", Vol. 5, Suppl. 3 (ed. M.O. Dayhoff), p. 345-352. Natl. Biomed. Res. Found., Washing-

ton, D.C. и Henikoff et al. (1992), *Proc. Nat'l. Acad. Sci. USA* 89: 10915-10919. Матрицу BLOSUM62 (фиг. 10) зачастую используют в качестве подстановочной матрицы по умолчанию для подсчета баллов в протоколах выравнивания последовательностей, таких как Gapped BLAST 2.0. Штраф за разрыв налагается на введение одного аминокислотного разрыва в одной из выровненных последовательностей, и штраф за удлинение разрыва налагается на каждое дополнительное пустое положение аминокислоты, включенное в уже открытый разрыв. Выравнивание определяется положениями аминокислот каждой последовательности, в которой выравнивание начинается и заканчивается, и необязательно вставкой разрыва или многочисленных разрывов в одной или обеих последовательностях для достижения наиболее возможного балла. Хотя оптимальное выравнивание и подсчет баллов может проводиться вручную, этот процесс облегчается использованием компьютерного алгоритма выравнивания, например, содержащий разрывы BLAST 2.0, описанный у Altschul et al. (1997) *Nucl. Acids Res.* 25: 3389-3402, и сделанный общедоступным на National Center for Biotechnology Information (NCBI) Website (www.ncbi.nlm.nih.gov). Оптимальные выравнивания, в том числе множественные выравнивания, могут быть получены с использованием, например, PSI-BLAST, доступного через веб-сайт NCBI и описанный Altschul et al. (1997), *Nucl. Acids Res.* 25:3389-3402.

В отношении аминокислотной последовательности, которая является оптимально выровненной с эталонной последовательностью, аминокислотный остаток "соответствует" положению в эталонной последовательности, с которой этот остаток образует пару при выравнивании. "Положение" обозначается числом, которое последовательно определяет каждую аминокислоту в эталонной последовательности на основании ее положения относительно N-конца. Вследствие делеций, вставок, усечений, слияний и т.д., которые должны быть приняты во внимание при определении оптимального выравнивания, вообще, номер аминокислотного остатка в тестируемой последовательности, определяемый простым подсчетом от N-конца, не обязательно будет таким же, как номер его соответствующего положения в эталонной последовательности. Например, в том случае, когда имеет место делеция в выравниваемой тестируемой последовательности, там не будет аминокислоты, которая соответствует положению в эталонной последовательности в месте делеций. В том случае, когда имеет место вставка в выравниваемой эталонной последовательности, эта вставка не будет соответствовать какому-либо положению аминокислот в эталонной последовательности. В случае усечений или слияний, могут быть участки аминокислот либо в эталонной, либо выравниваемой последовательности, которые не соответствуют какой-либо аминокислоте в соответствующей последовательности.

Неконсервативными модификациями конкретного полипептида являются модификации, которые заменяют любую аминокислоту, не охарактеризованные как консервативная замена, например любая замена, которая пересекает связи шести групп, изложенных выше. Эти замены включают замены основных или кислых аминокислот нейтральными аминокислотами (например, Asp, Glu, Asn или Gln на Val, Ile, Leu или Met), ароматической аминокислоты - основными или кислыми аминокислотами (например, Phe, Tyr или Trp на Asp, Asn, Glu или Gln) или любую другую замену, не заменяющую аминокислоту подобной аминокислотой. Основные боковые цепи включают лизин (K), аргинин (R), гистидин (H); кислые боковые цепи включают аспарагиновую кислоту (D), глутаминовую кислоту (E); незаряженные полярные боковые цепи включают глицин (G), аспарагин (N), глутамин (Q), серин (S), треонин (T), тирозин (Y), цистеин (C); неполярные боковые цепи включают аланин (A), валин (V), лейцин (L), изолейцин (I), пролин (P), фенилаланин (F), метионин (M), триптофан (W); бета-разветвленные боковые цепи включают треонин (T), валин (V), изолейцин (I); ароматические боковые цепи включают тирозин (Y), фенилаланин (F), триптофан (W), гистидин (H).

В соответствии с этим некоторые аминокислотные остатки в определенных положениях в полипептиде "исключены" из консервативных аминокислотных замен. Например, в настоящем описании представлен полипептид, содержащий последовательность SEQ ID NO: 4. В некоторых вариантах осуществления, полипептид может быть изменен консервативными аминокислотными заменами, как описано выше, однако некоторые остатки желательно оставлять незамещенными, например остатки в положениях 23, 108 и 140.

"Родительский" белок, фермент, полинуклеотид, ген или клетка представляют собой любой белок, фермент, полинуклеотид, ген или клетку, из которой любой другой белок, фермент, полинуклеотид, ген или клетка, происходят или получены с помощью любых способов, инструментов или методик, а сам родительский вариант может являться или может не являться природным или мутантным. Родительский полинуклеотид или ген кодирует родительский белок или фермент.

Полинуклеотид, полипептид или другой компонент является "выделенным", когда он частично или полностью отделен от компонентов, с которыми он обычно связан (других белков, нуклеиновых кислот, клеток, синтетических реагентов и т.д.). Нуклеиновая кислота или полипептид являются "рекомбинантными", когда они являются искусственными или сконструированными или получены из искусственного или сконструированного белка или нуклеиновой кислоты. Например, полинуклеотид, который встроен в вектор или любое другое гетерологичное положение, например в геном рекомбинантного организма, так, чтобы он не был связан с нуклеотидными последовательностями, которые обычно фланкируют полинуклеотид, в том виде, как он находится в природе, представляет собой рекомбинантный полинуклеотид.

Белок, экспрессируемый *in vitro* или *in vivo* из рекомбинантного полинуклеотида является примером рекомбинантного полипептида. Аналогично этому, полинуклеотидная последовательность, которая не встречается в природе, например вариант природного гена, является рекомбинантной.

Важно отметить, что некоторые специфические, важные и достоверные применения CD полинуклеотидов не требуют, чтобы этот полинуклеотид кодировал полипептид со значительной CD активностью или даже активностью варианта CD. Например, CD полинуклеотиды, которые не кодируют активные ферменты, могут быть ценными источниками родительских полинуклеотидов для применения в методиках диверсификации для получения вариантов CD полинуклеотида или не-CD полинуклеотидов, с желаемыми функциональными свойствами (например, высокой k_{cat} или k_{cat}/K_m , низкой K_m , высокой стабильностью в отношении нагревания или других факторов окружающей среды, высокой скоростью транскрипции или трансляции, устойчивости к протеолитическому расщеплению и т.д.).

CD полинуклеотиды, в том числе нуклеотидные последовательности, которые кодируют CD полипептиды и их варианты, фрагменты полипептидов CD, родственные слитые белки или их функциональные эквиваленты, используют в рекомбинантных молекулах ДНК, которые направляют экспрессию полипептидов CD в соответствующих клетках-хозяевах, таких как бактериальные клетки. Вследствие присутствия генетическому коду вырожденности другие последовательности нуклеиновой кислоты, которые кодируют, по существу, такую же или функционально эквивалентную аминокислотную последовательность, также могут быть использованы для клонирования и экспрессии полинуклеотидов CD.

Используемый в настоящем описании термин "клетка-хозяин" включает клетку любого типа, которая подвержена трансформации конструкцией нуклеиновой кислоты. Термин "трансформация" означает введение чужеродного (например, постороннего или внеклеточного) гена, последовательности ДНК или РНК в клетку-хозяина так, чтобы клетка-хозяин экспрессировала введенный ген или последовательность для продукции желаемого вещества, обычно белка или фермента, кодируемого введенным геном или последовательностью. Введенный ген или последовательность может включать регуляторные или контрольные последовательности, например стартовую, стоп, промоторную, сигнальную, секретлируемую или другие последовательности, используемые генетическим аппаратом клетки. Клетка-хозяин, которая получает и экспрессирует введенную ДНК или РНК, была "трансформирована" и является "трансформантом" или "клоном". ДНК или РНК, введенная в клетку-хозяина, может быть получена из любого источника, в том числе клеток того же рода или вида, как клетка-хозяин или клеток другого рода или вида.

Как будет понятно специалисту в данной области, может быть предпочтительным модифицировать кодирующую последовательность для усиления ее экспрессии в конкретной клетке-хозяине. Генетический код дублируется 64 возможными кодонами, но большинство организмов предпочтительно используют подгруппу этих кодонов. Кодоны, которые используются наиболее часто в видах, называют оптимальными кодонами, а те, которые не используются очень часто, классифицируют как редкие или редкоиспользуемые кодона (см., например, Zhang et al. (1991), Gene, 105:61-72). Кодоны могут быть заменены для отражения предпочтительной частоты использования кодонов клетки-хозяина, процесс - иногда называемый "оптимизацией кодона" или "контролем за отклонением разновидности кодонов".

Оптимизированные кодирующие последовательности, содержащие кодона, предпочтительные у конкретной прокариотической или эукариотической клетки-хозяина (см., также Muga et al. (1989), Nucl. Acids Res. 17:477-508) могут быть получены, например, для увеличения скорости трансляции или для получения рекомбинантных РНК транскриптов, обладающих желаемыми свойствами, такими как более длинный период полужизни, по сравнению с транскриптами, продуцируемыми из неоптимизированной последовательности. Стоп-кодона трансляции также могут быть модифицированы для отражения предпочтения клетки-хозяина. Например, предпочтительными стоп-кодонами для *S.cerevisiae* и млекопитающих являются UAA и UGA соответственно.

Предпочтительным стоп-кодоном для однодольных растений является UGA, тогда как насекомые и *E.coli* предпочитают использовать UAA в качестве стоп-кодона (Dalphin et al. (1996), Nucl. Acids Res. 24: 216-218).

Отклонения в частоте использования кодонов относятся к различиям среди организмов в частоте появления кодонов в последовательностях ДНК (генах), кодирующих белок. Кодон представляет собой серию трех нуклеотидов (триплетов), который кодирует определенный аминокислотный остаток в полипептидной цепи. Поскольку в ДНК существует четыре нуклеотида, аденин (А), гуанин (G), цитозин (C) и тимин (Т), существует 64 возможных триплета, кодирующих 20 аминокислот, и три кодона терминации трансляции (нонсенс). Вследствие этой избыточности, все кроме двух аминокислот кодируются более чем одним триплетом. Различные организмы зачастую демонстрируют конкретные предпочтения одного из нескольких кодонов, которые кодируют одну и ту же аминокислоту. То, как это предпочтение возникает, является очень обсуждаемой областью молекулярного развития.

В основном является общепризнанным то, что предпочтения кодонов отражают баланс между мутационными отклонениями и природным отбором оптимизации трансляции. Оптимальные кодона в быстрорастущих микроорганизмах, наподобие *Escherichia coli* или *Saccharomyces cerevisiae* (пекарских дрожжах), отражают состав их соответствующего геномного пула тРНК. Считается, что оптимальные кодона помогают достигать более высоких скоростей трансляции и высокой точности. В результате этих

факторов ожидается, что трансляционная селекция является более сильной в высокоэкспрессируемых генах, которая необходима в случае указанных выше микроорганизмов. В других организмах, которые не демонстрируют высокие скорости роста или которые представляют небольшие геномы, оптимизация частоты использования кодонов обычно отсутствует и предпочтения кодонов определяются характерными мутационными отклонениями, наблюдаемыми в этом конкретном геноме. Примерами этого являются *Homo sapiens* (человек) и *Helicobacter pylori*. Организмы, которые демонстрируют промежуточный уровень оптимизации использования кодонов, включают *Drosophila melanogaster* (плодовая мушка), *Caenorhabditis elegans* (нематодные черви) или *Arabidopsis thaliana* (резуха).

Термин "последовательности с оптимизированными кодонами" в основном относится к нуклеотидным последовательностям, которые были оптимизированы для конкретных видов клеток-хозяев путем замены любых кодонов, имеющих частоту использования примерно менее 20%. Нуклеотидные последовательности, которые были оптимизированы для экспрессии в заданных видах клеток-хозяев путем элиминации случайных последовательностей полиаденилирования, элиминации экзон/интронных сигналов сплайсинга, элиминации транспозон-подобных повторов и/или оптимизации содержания GC в дополнение к оптимизации кодонов, называется в настоящем описании "последовательностями усиленной экспрессии".

Таблица 1

Частота использования кодонов человека и предпочтение кодонов. Для каждого кодона в таблице показана частота использования каждого кодона (на тысячу) в кодирующих областях человека (первая колонка) и относительная частота каждого кодона среди синонимичных кодонов (вторая колонка)

Таблица частоты использования кодонов человека															
Gly	GGG	17,08	0,23	Arg	AGG	12,09	0,22	Trp	TGG	14,74	1,00	Arg	CGG	10,40	0,19
Gly	GGA	19,31	0,26	Arg	AGA	11,73	0,21	End	TGA	2,64	0,61	Arg	CGA	5,63	0,10
Gly	GGT	13,66	0,18	Ser	AGT	10,18	0,14	Cys	TGT	9,99	0,42	Arg	CGT	5,16	0,09
Gly	GGC	24,94	0,33	Ser	AGC	18,54	0,25	Cys	TGC	13,86	0,58	Arg	CGC	10,82	0,19
Glu	GAG	38,82	0,59	Lys	AAG	33,79	0,60	End	TAG	0,73	0,17	Gln	CAG	32,95	0,73
Glu	GAA	27,51	0,41	Lys	AAA	22,32	0,40	End	TAA	0,95	0,22	Gln	CAA	11,94	0,27
Asp	GAT	21,45	0,44	Asn	AAT	16,43	0,44	Tyr	TAT	11,80	0,42	His	CAT	9,56	0,41
Asp	GAC	27,06	0,56	Asn	AAC	21,30	0,56	Tyr	TAC	16,48	0,58	His	CAC	14,00	0,59
Val	GTG	28,60	0,48	Met	ATG	21,86	1,00	Leu	TTG	11,43	0,12	Leu	CTG	39,93	0,43
Val	GTA	6,09	0,10	Ile	ATA	6,05	0,14	Leu	TTA	5,55	0,06	Leu	CTA	6,42	0,07
Val	GTT	10,30	0,17	Ile	ATT	15,03	0,35	Phe	TTT	15,36	0,43	Leu	CTT	11,24	0,12
Val	GTC	15,01	0,25	Ile	ATC	22,47	0,52	Phe	TTC	20,72	0,57	Leu	CTC	19,14	0,20
Ala	GCG	7,27	0,10	Thr	ACG	6,80	0,12	Ser	TCG	4,38	0,06	Pro	CCG	7,02	0,11
Ala	GCA	15,50	0,22	Thr	ACA	15,04	0,27	Ser	TCA	10,96	0,15	Pro	CCA	17,11	0,27
Ala	GCT	20,23	0,28	Thr	ACT	13,24	0,23	Ser	TCT	13,51	0,18	Pro	CCT	18,03	0,29
Ala	GCC	28,43	0,40	Thr	ACC	21,52	0,38	Ser	TCC	17,37	0,23	Pro	CCC	20,51	0,33

Специалисту в данной области будет понятно, что изменения могут быть введены путем мутации в нуклеотидные последовательности любых полинуклеотидов, кодирующих полипептид, изложенный в любой из последовательностей: SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16 или 18 или имеющей нуклеотидную последовательность SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15 или 17, приводя, тем самым, к изменениям в аминокислотной последовательности кодируемых белков. В некоторых случаях изменение приведет к измененной функции этого полипептида. В других случаях это изменение не будет изменять функциональную способность кодируемого полипептида. В основном замены, которые не изменяют функцию полипептида, включают нуклеотидные замены, приводящие к аминокислотным заменам в "несущественных" аминокислотных остатках. "Несущественным" аминокислотным остатком является остаток, который может быть изменен из родительской последовательности без изменения биологической активности полученного в результате полипептида, например, катализирующего превращение 5-FC в 5-FU.

Также рассматриваются те ситуации, в которых желательно изменить активность родительского полинуклеотида, так, чтобы этот полипептид имел новую или повышенную активность в отношении конкретного субстрата или повышенную стабильность или сниженное разрушение. Понятно, что эти аминокислотные замены в основном не составляют "консервативные" замены. Вместо этого, эти замены составляют неконсервативные замены, введенные в последовательность для получения новой или улучшенной активности. Например, последовательность SEQ ID NO: 1 предоставляет родительскую последовательность нуклеиновой кислоты для цитозиндезаминазы *S.cervisiae* (SEQ ID NO: 2 представляет соответствующий полипептид). Последовательность SEQ ID NO: 3 представляет последовательность нуклеиновой кислоты мутантной последовательности, которая включает аминокислотные замены, которые придают повышенную стабильность этому полипептиду. Соответственно, молекула нуклеиновой кислоты, кодирующая аминокислотную последовательность SEQ ID NO: 2, представляет "родительскую" молекулу нуклеиновой кислоты, которая может быть мутирована для получения молекулы нуклеиновой

кислоты, которая кодирует модифицированный полипептид, который включает аминокислотные замены.

Также должно быть понятно, что модифицированный полипептид может составлять "родительский" полипептид из которого могут быть получены дополнительные замены. Соответственно, родительский полипептид и молекула нуклеиновой кислоты, которая кодирует родительский полипептид, включает модифицированные полипептиды и не только последовательности "дикого типа". Например, полинуклеотид последовательности SEQ ID NO: 5 представляет собой модифицированный полинуклеотид относительно последовательности SEQ ID NO: 1 (т.е. "родительский" полинуклеотид). Аналогично, полинуклеотид последовательности SEQ ID NO: 3 представляет собой модифицированный полинуклеотид относительно SEQ ID NO: 5. Соответственно, последовательность SEQ ID NO: 5 представляет собой родительскую последовательность SEQ ID NO: 3.

Мутационные способы создания разнообразия включают, например, сайт-направленный мутагенез (Ling et al. (1997), "Approaches to DNA mutagenesis: an overview", *Anal Biochem.* 254(2):157-178; Dale et al., (1996), "Oligonucleotide-directed random mutagenesis using the phosphorothioate method", *Methods Mol. Biol.* 57:369-374; Smith (1985), "In vitro mutagenesis", *Ann. Rev. Genet.* 19:423-462; Botstein & Shortle (1985), "Strategies and applications of in vitro mutagenesis", *Science* 229:1193-1201; Carter (1986), "Site-directed mutagenesis", *Biochem. J.* 237:1-7; и Kunkel (1987), "The efficiency of oligonucleotide directed mutagenesis" в *Nucleic Acids & Molecular Biology* (Eckstein, F. and Lilley, D.M. J. eds., Springer Verlag, Berlin)); мутагенез с использованием матриц, содержащих урацил (Kunkel (1985), "Rapid and efficient site-specific mutagenesis without phenotypic selection", *Proc. Natl. Acad. Sci. USA*, 82:488-492; Kunkel et al. (1987), "Rapid and efficient site-specific mutagenesis without phenotypic selection", *Methods в Enzymol.* 154, 367-382; и Bass et al. (1988), "Mutant Trp repressors with new DNA-binding specificities", *Science*, 242:240-245); олигонуклеотид-направленный мутагенез (*Methods in Enzymol.* 100: 468-500 (1983); *Methods in Enzymol.* 154: 329-350 (1987); Zoller & Smith (1982), "Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any DNA fragment", *Nucleic Acids Res.* 10:6487-6500; Zoller & Smith (1983), "Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors" *Methods in Enzymol.* 100:468-500; и Zoller & Smith (1987), "Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template", *Methods в Enzymol.* 154:329-350); фосфоротиаот-модифицированный ДНК мутагенез (Taylor et al. (1985), "The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA", *Nucl. Acids Res.* 13:8749-8764; Taylor et al. (1985), "The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA", *Nucl. Acids Res.* 13: 8765-8787; Nakamaye & Eckstein (1986), "Inhibition of restriction endonuclease Nci I cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis", *Nucl. Acids Res.* 14:9679-9698; Sayers et al. (1988), "Y-T Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis", *Nucl. Acids Res.* 16:791-802; и Sayers et al. (1988), "Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromide" *Nucl. Acids Res.* 16:803-814); мутагенез с использованием двухспиральной ДНК, содержащей разрывы (Kramer et al. (1984), "The gapped duplex DNA approach to oligonucleotide-directed mutation construction", *Nucl. Acids Res.* 12:9441-9456; Kramer & Fritz (1987), *Methods in Enzymol.* "Oligonucleotide-directed construction of mutations via. gapped duplex DNA", 154:350-367; Kramer et al. (1988), "Improved enzymatic in vitro reactions in the gapped duplex DNA approach to oligonucleotide-directed construction of mutations", *Nucl. Acids Res.* 16: 7207; и Fritz et al. (1988), "Oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions in vitro", *Nucl. Acids Res.* 16:6987-6999).

Дополнительные подходящие способы включают мочковое исправление ошибок спаривания (Kramer et al. (1984), "Point Mismatch Repair", *Cell* 38:879-887), мутагенез с использованием штаммов-хозяев с недостаточностью системы репарации (Carter et al. (1985), "Improved oligonucleotide site-directed mutagenesis using M13 vectors", *Nucl. Acids Res.* 13:4431-4443; и Carter (1987), "Improved oligonucleotide-directed mutagenesis using M13 vectors", *Methods в Enzymol.* 154:382-403), делеционный мутагенез (Eghtedarzadeh & Henikoff (1986), "Use of oligonucleotides to generate large deletions", *Nucl. Acids Res.* 14:5115), рестрикцию-селекцию и рестрикцию-очистку (Wells et al. (1986), "Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin", *Phil. Tirans. R. Soc. Lond. A* 317:415-423), мутагенез с помощью суммарного синтеза генов (Nambiar et al. (1984), "Total synthesis and cloning of a gene coding for the ribonuclease S protein", *Science* 223:1299-1301; Sakamar и Khorana (1988), "Total synthesis and expression of a gene for the α -subunit of bovine rod outer segment guanine nucleotide-binding protein (transducin)", *Nucl. Acids Res.* 14:6361-6372; Wells et al. (1985), "Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites", *Gene* 34:315-323; и Grundstrom et al. (1985), "Oligonucleotide-directed mutagenesis by microscale "shot-gun", gene synthesis", *Nucl. Acids Res.* 13:3305-3316); репарацию двухспиральных разрывов (Mandecki (1986); Arnold (1993), "Protein engineering for unusual environments", *Current Opinion в Biotechnology* 4:450-455; и "Oligonucleotide-directed double-strand break repair in plasmids of *Escherichia coli*: a method for site-specific mutagenesis", *Proc. Natl. Acad. Sci. USA*, 83:7177-7181). Дополнительные подробности, касающиеся многих из указанных выше способов, можно найти в *Methods in Enzymology Volume 154*, которые также описывают полезные способы контро-

ля для решения проблем поиска неисправностей с использованием различных способов мутагенеза.

Дополнительные подробности, касающиеся различных способов создания разнообразия, можно найти в следующих патентах США, публикациях PCT и публикациях EPO: патент США № 5605793 на имя Stemmer (25 февраля 1997 г.), "Methods for In vitro Recombination", патент США № 5811238 на имя Stemmer et al. (22 сентября 1998 г.), "Methods for Generating Polynucleotides having Desired Characteristics by Iterative Selection and Recombination", патент США № 5830721 на имя Stemmer et al. (3 ноября 1998 г.), "DNA Mutagenesis by Random Fragmentation and Reassembly", патент США № 5834252 на имя Stemmer, et al. (10 ноября 1998 г.), "End-Complementary Polymerase Reaction", патент США № 5837458 на имя Minshull, et al. (17 ноября 1998 г.), "Methods and Compositions for Cellular and Metabolic Engineering", WO 95/22625, Stemmer и Cramer, "Mutagenesis by Random Fragmentation and Reassembly", WO 96/33207, поданной Stemmer и Lipschutz "End Complementary Polymerase Chain Reaction", WO 97/20078, поданной Stemmer и Cramer "Methods for Generating Polynucleotides having Desired Characteristics by Iterative Selection and Recombination", WO 97/35966, поданной Minshull и Stemmer, "Methods and Compositions for Cellular and Metabolic Engineering", WO 99/41402, поданной Punnonen et al. "Targeting of Genetic Vaccine Vectors", WO 99/41383, поданной Punnonen et al. "Antigen Library Immunization", WO 99/41369, поданной Punnonen et al. "Genetic Vaccine Vector Engineering", WO 99/41368, поданной Punnonen et al. "Optimization of Immunomodulatory Properties of Genetic Vaccines", EP 752008, поданной Stemmer и Cramer, "DNA Mutagenesis by Random Fragmentation and Reassembly", EP 0932670, поданной Stemmer "Evolving Cellular DNA Uptake by Recursive Sequence Recombination", WO 99/23107, поданной Stemmer et al., "Modification of Virus Tropism and Host Range by Viral Genome Shuffling", WO 99/21979, поданной Apt et al., "Human Papillomavirus Vectors", WO 98/31837, поданной del Cardayre et al. "Evolution of Whole Cells and Organisms by Recursive Sequence Recombination", WO 98/27230, поданной Patten и Stemmer, "Methods and Compositions for Polypeptide Engineering", WO 98/13487, поданной Stemmer et al., "Methods for Optimization of Gene Therapy by Recursive Sequence Shuffling and Selection", WO 00/00632, "Methods for Generating Highly Diverse Libraries", WO 00/09679, "Methods for Obtaining in vitro Recombined Polynucleotide Sequence Banks and Resulting Sequences", WO 98/42832, поданной Arnold et al., "Recombination of Polynucleotide Sequences Using Random or Defined Primers", WO 99/29902, поданной Arnold et al., "Method for Creating Polynucleotide and Polypeptide Sequences", WO 98/41653, поданной Vind, "An in vitro Method for Construction of a DNA Library", WO 98/41622, поданной Borchert et al., "Method for Constructing a Library Using DNA Shuffling", WO 98/42727, поданной Pati and Zarling, "Sequence Alterations using Homologous Recombination", WO 00/18906, поданной Patten et al., "Shuffling of Codon-Altered Genes", WO 00/04190, поданной del Cardayre et al. "Evolution of Whole Cells and Organisms by Recursive Recombination", WO 00/42561, поданной Cramer et al., "Oligonucleotide Mediated Nucleic Acid Recombination", WO 00/42559, поданной Selifonov и Stemmer "Methods of Populating Data Structures for Use in Evolutionary Simulations", WO 00/42560, поданной Selifonov et al., "Methods for Making Character Strings, Polynucleotides & Polypeptides Having Desired Characteristics", WO 01/23401, поданной Welch et al., "Use of Codon-Variied Oligonucleotide Synthesis for Synthetic Shuffling" и WO 01/64864 "Single-Stranded Nucleic Acid Template-Mediated Recombination and Nucleic Acid Fragment Isolation", поданной Affholter.

Термины "вектор", "векторная конструкция" и "вектор экспрессии" означают носитель, с помощью которого последовательность ДНК или РНК (например, чужеродный ген) может быть встроена в клетку-хозяина так, чтобы трансформировать клетку-хозяина и активизировать экспрессию (например, транскрипцию и трансляцию) встроенной последовательности. Векторы обычно содержат ДНК трансмиссивного агента, в который чужеродную ДНК, кодирующую белок, встраивают с помощью технологии ферментов рестрикции. Обычным типом вектора является "плазмида", которая в основном представляет собой замкнутую молекулу двухспиральной ДНК, которая легко может принимать дополнительную (чужеродную) ДНК, и которую легко можно встраивать в подходящую клетку-хозяина. Большое число векторов, в том числе плазмид и грибных векторов, было описано для репликации и/или экспрессии в разнообразных эукариотических и прокариотических клетках-хозяевах. Неограничивающие примеры включают pKK плазмиды (Clontech), pUC плазмиды, pET плазмиды (Novagen, Inc., Madison, Wis.), pRSET или pREP плазмиды (Invitrogen, San Diego, Calif.) или pMAL плазмиды (New England Biolabs, Beverly, Mass.) и многие соответствующие клетки-хозяева, используя способы, описанные или процитированные в настоящем описании или иным образом известные специалистам в релевантной области. Рекомбинантные клонирующие векторы зачастую будут включать одну или несколько систем репликации для клонирования или экспрессии, один или несколько маркеров для селекции в клетке-хозяине, например резистентность к антибиотикам, и одну или несколько экспрессионных кассет.

Термины "экспрессируют" и "экспрессия" означают предоставление возможности или осуществление реализации информации в гене или последовательности ДНК, например, продукцию белка путем активации клеточных функций, вовлеченных в транскрипцию и трансляцию соответствующего гена или последовательности ДНК. Последовательность ДНК экспрессируется в клетке или клеткой для образования "продукта экспрессии", такого как белок. Также можно сказать, что сам продукт экспрессии, например полученный в результате белок, "экспрессирован" клеткой. Полинуклеотид или полипептид экспрессируется рекомбинантно, например, когда он экспрессируется или продуцируется в чужеродной клетке-

хозяине под контролем чужеродного или нативного промотора или в нативной клетке-хозяине под контролем чужеродного промотора.

В одном варианте осуществления вирусный вектор может быть репликационно-компетентным ретровирусным вектором, способным инфицировать только реплицирующиеся клетки млекопитающих.

Ретровирусы были классифицированы различными путями, но номенклатура была стандартизована в последнее десятилетие (см. ICTVdB - The Universal Virus Database, v. 4 на World Wide Web www.ncbi.nlm.nih.gov/ICTVdb/ICTVdb/ и пособие "Retroviruses" Eds Coffin, Hughs and Varmus, Cold Spring Harbor Press 1997; описание которых включено в настоящий документ путем ссылки). Репликационно-компетентный ретровирусный вектор может включать орторетровирусный или более характерно гамма-ретровирусный вектор. В одном аспекте репликационно-компетентный ретровирусный вектор содержит внутренний сайт связывания рибосомы (IRES) 5' с полинуклеотидом, кодирующим цитозиндезаминазу. В одном варианте осуществления полинуклеотид, кодирующий цитозиндезаминазу, находится 3' по отношению к ENV полинуклеотиду ретровирусного вектора.

Как рассмотрено ранее, общие сведения, в которых описываются методики молекулярной биологии, используемые в настоящем описании, в том числе использование векторов, промоторов и многие другие актуальные темы, включают Berger и Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology, Volume 152 (Academic Press, Inc., San Diego, Calif.) ("Berger"); Sambrook et al., Molecular Cloning A Laboratory Manual, 2nd ed., Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1989 ("Sambrook") и Current Protocols in Molecular Biology, F.M. Ausubel et al., eds., Current Protocols, совместное предприятие Greene Publishing Associates, Inc. и John Wiley & Sons, Inc., (дополнено на протяжении 1999) ("Ausubel"). Примеры протоколов, достаточные, чтобы проинструктировать специалистов в данной области по in vitro методам амплификации, в том числе полимеразной цепной реакции (PCR), лигазной цепной реакции (LCR), Q β -репликационной амплификации и другими методиками, опосредованных РНК полимеразой (например, NASBA), например, для продукции гомологичных нуклеиновых кислот по настоящему изобретению, можно найти у Berger, Sambrook и Ausubel, а также у Mullis et al. (1987) в патенте США № 4683202; Innis et al., eds. (1990), PCR Protocols: A Guide to Methods and Applications (Academic Press Inc. San Diego, Calif.) ("Innis"); Arnheim & Levinson (Oct. 1, 1990) C&EN 36-47; The Journal Of NIH Research (1991), 3:81-94; Kwoh et al. (1989), Proc. Natl. Acad. Sci. USA 86:1173; Guatelli et al. (1990), Proc. Natl. Acad. Sci. USA 87:1874; Lomell et al. (1989), J. Clin. Chem 35:1826; Landegren et al. (1988), Science, 241:1077-1080; Van Brunt (1990), Biotechnology, 8:291-294; Wu и Wallace (1989), Gene, 4:560; Barringer et al. (1990), Gene, 89:117 и Sooknanan and Malek (1995), Biotechnology, 13: 563-564. Усовершенствованные способы клонирования in vitro амплифицированных нуклеиновых кислот описаны у Wallace et al., в патенте США № 5426039. Усовершенствованные способы амплификации крупных нуклеиновых кислот с помощью ПЦР кратко изложены у Cheng et al. (1994), Nature, 369:684-685 и в ссылках, процитированных в этих публикациях, в которых получают ПЦР ампликоны вплоть до 40 тыс. пар оснований. Специалисту в данной области будет понятно, что, по существу, любая РНК может быть преобразована в двухспиральную ДНК, подходящую для рестрикционного расщепления, ПЦР удлинения и секвенирования с использованием обратной транскриптазы и полимеразы. См., например, Ausubel, Sambrook и Berger, все указано выше.

Также представлены сконструированные клетки-хозяева, которые трансдуцированы (трансформированы или трансфицированы) вектором, представленным в настоящем описании (например, клонирующим вектором или вектором экспрессии), а также получение полипептидов по настоящему изобретению с помощью рекомбинантных технологий. Вектором может быть, например, плаزمид, вирусная частица, фаг и т.д. Сконструированные клетки-хозяева могут быть культивированы в обычной питательной среде, модифицированной соответствующим образом для активации промоторов, отбора трансформантов или амплификации кодирующего полинуклеотида. Условиями культивирования, такими как температура, pH и подобные, являются условия, использованные ранее с клетками-хозяевами, выбранными для экспрессии, и будут очевидны специалистам в данной области и в ссылках, процитированных в настоящем описании, в том числе, например, Sambrook, Ausubel и Berger, а также, например, Freshney (1994), Culture of Animal Cells: A Manual of Basic Technique, 3rd ed. (Wiley-Liss, New York) и процитированных там ссылках.

Векторы могут быть использованы для трансформации соответствующего хозяина, чтобы обеспечить экспрессию белка или полипептида клеткой-хозяином. Примеры соответствующих экспрессионных клеток-хозяев включают клетки бактерий, такие как E.coli, B.subtilis, Streptomyces и Salmonella typhimurium; клетки грибов, такие как Saccharomyces cerevisiae, Pichia pastoris, и Neurospora crassa; клетки насекомых, такие как Drosophila и Spodoptera frugiperda; клетки млекопитающих, такие как CHO, COS, BHK, HEK 293 или Bowes меланомы; или клетки растений или экспланта и т.д.

В бактериальных системах ряд векторов экспрессии может быть выбран в зависимости от использования, предназначенного для полипептида цитозиндезаминазы. Например, в тех случаях, когда большие количества полипептида CD или его фрагментов необходимы для коммерческой продукции или для индукции антител, могут быть желательными векторы, которые направляют высокий уровень экспрессии слитых белков, которые легко очистить. Такие векторы включают, но не только, многофункциональные

E. coli клонирующие векторы и векторы экспрессии, такие как BLUESCRIPT (Stratagene), в которых последовательность, кодирующая полипептид CD, может быть лигирована в вектор в рамке считывания с последовательностями для аминоконцевого Met и последующими 7 остатками бета-галактозидазы, так, чтобы продуцировался гибридный белок; pIN векторы (Van Heeke & Schuster (1989), J. Biol. Chem. 264:5503-5509); pET векторы (Novagen, Madison Wis.) и подобные.

Аналогично, у дрожжей *Saccharomyces cerevisiae* ряд векторов, содержащих конститутивные или индуцибельные промоторы, такие как альфа фактор, алкоголь-оксидаза и PGH, могут быть использованы для продукции полипептидов CD по настоящему изобретению. Обзоры см. у Ausubel (supra) и Grant et al. (1987), *Methods in Enzymology*, 153:516-544.

Настоящее изобретение также относится к репликационно-компетентным ретровирусным векторам, имеющим повышенную стабильность относительно предшествующих ретровирусных векторов. Такая повышенная стабильность во время инфицирования и репликации является важной для лечения нарушений клеточной пролиферации. Сочетание эффективности трансдукции, трансгенной стабильности и специфичности к мишени обеспечивается репликационно-компетентными ретровирусами. Композиции и способы обеспечивают стабильность вставки и поддерживают транскрипционную активность трансгена и трансляционную эффективность кодируемого полипептида.

Ретровирусы могут передаваться горизонтально и вертикально. Эффективная инфекционная трансмиссия ретровирусов требует экспрессии на клетке-мишени рецепторов, которые специфически распознают белки оболочки вируса, хотя вирусы могут использовать рецептор-независимые, неспецифические пути вхождения с низкой эффективностью. Кроме того, тип клетки-мишени должен быть способен поддерживать все стадии цикла репликации после связывания и пенетрации вируса. Вертикальная трансмиссия происходит, когда вирусный геном становится интегрированным в зародышевую линию хозяина. Затем провирус будет передаваться от поколения к поколению как будто клеточный ген. Следовательно, эндогенные провирусы укореняются, которые зачастую лежат латентно, но которые могут активироваться, когда хозяин подвергается воздействию соответствующих агентов.

Как указано выше, интегрированный посредник ДНК называется провирусом. В предшествующей генной терапии и системах доставки генов используются способы и ретровирусы, которые требуют транскрипции провируса и сборки в инфицирующий вирус, либо в присутствии соответствующего вируса-помощника, либо в клеточной линии, содержащий соответствующие последовательности, способствующие инкапсулированию без совпадающей продукции контаминирующего вируса-помощника. Как описано ниже, вирус-помощник не требуется для продукции рекомбинантного ретровируса по настоящему изобретению, поскольку последовательности для инкапсулирования обеспечиваются в геноме, таким образом предоставляя репликационно-компетентный ретровирусный вектор для доставки генов или генной терапии.

Ретровирусный геном, применимый в способах и композициях по настоящему изобретению, содержит провирусную ДНК, содержащую по меньшей мере три гена: gag, pol, и env, эти гены могут быть фланкированы одним или двумя длинными концевыми повторами (LTR) или в провирусе фланкированы двумя длинными концевыми повторами (LTR) и последовательностями, содержащими действующие в цис-положении последовательности, такие как psi. Ген gag кодирует внутренние структурные белки (матрикс, капсид и нуклеокапсид); ген pol кодирует РНК-направленную ДНК полимеразу (обратную транскриптазу), протеазу и интегразу; и ген env кодирует гликопротеины вирусной оболочки. 5' и/или 3' LTR служат для активации транскрипции и полиаденилирования РНК вириона. LTR содержит все остальные последовательности, действующие в цис-положении, необходимые для репликации вируса. Lentiviruses имеют дополнительные гены, в том числе vif, vpr, tat, rev, vpu, nef, и vpx (в HIV-1, HIV-2 и/или SIV).

Расположенные рядом с 5' LTR последовательности необходимы для обратной транскрипции генома (участок связывания праймера тРНК) и для эффективного инкапсулирования вирусной РНК в частицы (участок Psi). Если последовательности, необходимые для инкапсулирования (или упаковки ретровирусной РНК в инфекционный вирион), отсутствуют в вирусном геноме, результатом является цис-дефект, который предотвращает инкапсулирование геномной вирусной РНК. Этот тип модифицированного вектора представляет собой то, что обычно использовалось в предшествующих системах доставки гена (т.е. системы, лишенные элементов, которые необходимы для инкапсулирования вириона).

В одном варианте осуществления настоящее изобретение относится к рекомбинантному ретровирусу, способному инфицировать неделяющуюся клетку, делящуюся клетку или клетку, у которой имеется нарушение клеточной пролиферации.

Рекомбинантные репликационно-компетентные ретровирусы по настоящему изобретению содержат полинуклеотидную последовательность, кодирующую вирусный GAG, вирусный POL, вирусный ENV, гетерологичный полинуклеотид, с предшествующим внутренним сайтом связывания рибосомы (IRES), инкапсулированным в вирионе.

Гетерологичная последовательность нуклеиновой кислоты оперативно связана с IRES. Используемый в настоящем описании термин "гетерологичная" последовательность нуклеиновой кислоты или трансген относится к (i) к последовательности, которая обычно не существует в ретровирусах дикого

типа, (ii) к последовательности, которая происходит из чужеродных видов, или (iii) если из того же вида, она может быть, по существу, модифицирована от ее первоначальной формы. Альтернативно, неизменная последовательность нуклеиновой кислоты, которая обычно не экспрессируется в клетке, является гетерологичной последовательностью нуклеиновой кислоты. В конкретном варианте осуществления гетерологичный полинуклеотид представляет собой полипептид по настоящему изобретению, обладающий цитозиндезаминазной активностью.

Выражение "неделяющаяся" клетка относится к клетке, которая не проходит стадию митоза. Неделящиеся клетки могут быть заблокированы в любой момент клеточного цикла (например, G_0/G_1 , $G_{1/S}$, $G_{2/M}$) при условии, что эта клетка не является активно делящейся. Для инфицирования *ex vivo* делящаяся клетка может быть обработана для блокировки клеточного деления с использованием стандартных методик, используемых специалистами в данной области, в том числе излучением, обработкой афидоколином, истощением сыворотки и контактным ингибированием. Однако должно быть понятно, что инфицирование *ex vivo* зачастую проводят без блокировки клеток, поскольку многие клетки уже остановлены (например, стволовые клетки). Например, рекомбинантный лентивирусный вектор способен инфицировать любую неделящуюся клетку, независимо от механизма, используемого для блокировки клеточного деления или момента в клеточном цикле, в который заблокирована клетка. Примеры уже существующих неделящихся клеток в организме включают нейрональные, мышечные клетки, клетки печени, кожи, сердца, легких и клетки костного мозга и их производные. Для делящихся клеток могут быть использованы онкоретровирусные векторы.

Под "делящейся" клеткой подразумевается клетка, которая подвергается активному митозу или мейозу. Такие делящиеся клетки включают стволовые клетки, клетки кожи (например, фибробласты и кератиноциты), гаметы и другие делящиеся клетки, известные из уровня техники. Представляющими особый интерес и охватываемые этим термином делящиеся клетки являются клетки, имеющие нарушения клеточной пролиферации, такие как неопластические клетки. Термин "нарушение клеточной пролиферации" относится к состоянию, характеризующемуся аномальным числом клеток. Это состояние может включать как гипертрофический (непрерывное умножение клеток, в результате приводящее к разрастанию клеточной популяции в пределах ткани), так и гипотрофический (отсутствие или недостаточность клеток в ткани) клеточный рост или чрезмерный приток или миграцию клеток в область организма. Клеточные популяции не являются обязательно трансформированными, образующими опухоль или злокачественными клетками, а могут включать нормальные клетки тоже. Нарушения клеточной пролиферации включают нарушения, связанные с разрастанием соединительных тканей, например, различные фиброзные состояния, в том числе, склеродермию, артрит и цирроз печени. Нарушения клеточной пролиферации включают неопластические нарушения, такие как карциномы тканей головы и шеи. Карциномы тканей головы и шеи включали бы, например, карциному ротовой полости, пищевода, горла, гортани, щитовидной железы, языка, губ, слюнных желез, носа, околоносовых пазух, носоглотки, верхнего свода носа и опухоли пазухи, обонятельную эстезионероэпителиому, сквамозно-клеточный рак, злокачественную меланому, синоназальную недифференцированную карциному (SNUC), неоплазию головного мозга (в том числе глиобластомы) или крови. Также включены карциномы региональных лимфатических узлов, в том числе шейных лимфатических узлов, предгортанных лимфатических узлов, легочных юкстапищеводных лимфатических узлов и поднижечелюстных лимфатических узлов (Harrison's Principles of Internal Medicine (eds., Isselbacher, et al., McGraw-Hill, Inc., 13th Edition, p.1850-1853, 1994). Другие типы злокачественных опухолей включают, но этим не ограничиваются, рак легких, рак толстой и прямой кишки, рак груди, рак предстательной железы, рак мочевыводящих путей, рак матки, лимфому, рак ротовой полости, рак поджелудочной железы, лейкоз, меланому, рак желудка, рак кожи и рак яичников.

В одном варианте осуществления гетерологичный полинуклеотид в векторе содержит цитозиндезаминазу, которая была оптимизирована для экспрессии в клетках человека. В дополнительном варианте осуществления цитозиндезаминаза содержит последовательность, которая была оптимизирована в отношении кодонов человека, и содержит мутации, которые повышают стабильность цитозиндезаминазы (например, уменьшенное разрушение или повышенную термостабильность) по сравнению с цитозиндезаминазой дикого типа. Еще в одном варианте осуществления гетерологичный полинуклеотид кодирует слитую конструкцию, содержащую цитозиндезаминазу (либо кодон-оптимизированную для экспрессии в клетках человека, либо неоптимизированную, либо мутированную, либо немутированную), оперативно связанную с полинуклеотидом, кодирующим полипептид, обладающий UPRT или OPRT активностью. В другом варианте осуществления гетерологичный полинуклеотид содержит CD полинуклеотид или слитую конструкцию по настоящему изобретению (например, SEQ ID NO: 3, 5, 11, 13, 15 или 17).

В другом варианте осуществления репликационно-компетентный ретровирусный вектор может содержать гетерологичный полинуклеотид, кодирующий полипептид, содержащий цитозиндезаминазу (как описано выше), и дополнительно может содержать полинуклеотид, содержащий молекулу *mi*-РНК или *si*-РНК, связанную с промотором, специфичным к клеточному типу или к типу ткани.

Термин "регуляторная последовательность нуклеиновой кислоты" относится обобщенно к промоторным последовательностям, сигналам полиаденилирования, последовательностям терминации транскрипции, регуляторным доменам, расположенным против хода транскрипции, точкам начала реплика-

ции, энхансерам и подобному, которые сообщают обеспечивают репликацию, транскрипцию и трансляцию кодирующей последовательности в клетке-реципиенте. Не всегда нужно присутствие всех этих контрольных последовательностей, при условии, что выбранная кодирующая последовательность способна реплицироваться, транскрибироваться и транслироваться в соответствующей клетке-хозяине. Специалист в данной области без труда может идентифицировать регуляторную последовательность нуклеиновой кислоты из общедоступных баз данных и материалов. Более того, специалист в данной области может идентифицировать регуляторную последовательность, которая применима для предназначенного использования, например, *in vivo*, *ex vivo* или *in vitro*.

Термин "промоторная область" используется в настоящем описании в его первоначальном смысле, относящемся к нуклеотидной области, содержащей регуляторную последовательность ДНК, где регуляторная последовательность получена из гена, который способен связываться с РНК полимеразой и инициировать транскрипцию кодирующей последовательности, расположенной по ходу транскрипции (в 3'-направлении). Регуляторная последовательность может быть гомологичной или гетерологичной желаемой генной последовательности. Например, может быть использован широкий ряд промоторов, в том числе вирусный промотор или промотор млекопитающих, как описано выше.

Внутренние сайты связывания рибосомы ("IRES") относятся к сегменту нуклеиновой кислоты, которые способствуют связыванию или удерживанию рибосомы во время трансляции кодирующей последовательности обычно в положении 3' по отношению IRES. В некоторых вариантах осуществления IRES может содержать акцепторный сайт сплайсинга/донорный сайт, однако предпочтительные IRES лишены акцепторного сайта сплайсинга/донорного сайта. Обычно связывание рибосом с матричной РНК имеет место через кэп, расположенный на 5'-конце всех эукариотических мРНК. Однако существуют исключения из этого универсального правила. Отсутствие кэпа в некоторых вирусных мРНК предполагает существование альтернативных структур, позволяющих осуществлять связывание рибосом на внутреннем сайте этих РНК. На сегодняшний день ряд таких структур, обозначенных IRES на основании их функции, был идентифицирован в 5' некодирующей области неэкспрессированных вирусных мРНК, например у пикоронавирусов, например вируса полиомиелита (Pelletier et al., 1988, Mol. Cell. Biol., 8, 1103-1112) и вируса EMCV (вирус энцефаломиокардита (Jang et al., J. Virol., 1988, 62, 2636-2643). Настоящее изобретение относится к использованию IRES применительно к репликативно-компетентному ретровирусному вектору.

Гетерологичная последовательность нуклеиновой кислоты обычно находится под контролем либо вирусных сигналов LTR промотор-энхансер, либо внутреннего промотора, и сигналов удерживания в ретровирусном LTR может все еще приводить к эффективной интеграции вектора в геном клетки-хозяина. В соответствии с этим рекомбинантные ретровирусные векторы по настоящему изобретению, желаемые последовательности, гены и/или генные фрагменты могут быть встроены в нескольких сайтах и под контролем различных регуляторных последовательностей. Например, сайтом для встраивания может быть вирусный проксимальный сайт энхансер/промотор (т.е. 5' LTR-контролируемый генный locus). Альтернативно, желаемые последовательности могут быть встроены в дистальный сайт (например, IRES последовательность 3' к гену env) или где присутствуют две или более гетерологичные последовательности, одна гетерологичная последовательность может быть под контролем первой регуляторной области, а вторая гетерологичная последовательность может быть под контролем второй регуляторной области. Другие дистальные сайты включают вирусные промоторные последовательности, где экспрессия желаемой последовательности или последовательностей происходит посредством сплайсинга цистрона проксимального промотора, может быть использован внутренний гетерологичный промотор, такой как SV40 или CMV или внутренний сайт связывания рибосомы (IRES).

В одном варианте осуществления ретровирусный геном по настоящему изобретению содержит IRES, содержащий клонирующий сайт для встраивания желаемой полинуклеотидной последовательности. В одном варианте осуществления IRES расположен 3' по отношению к гену env в ретровирусном векторе, но 5' по отношению к желаемой гетерологичной нуклеиновой кислоте. В соответствии с этим гетерологичная полинуклеотидная последовательность, кодирующая желаемый полипептид, может быть оперативно связана с IRES.

В другом варианте осуществления направляющая полинуклеотидная последовательность включена как часть рекомбинантного ретровирусного вектора по настоящему изобретению. Направляющая полинуклеотидная последовательность представляет собой направленный лиганд (например, пептидные гормоны, такие как герегулин, одноцепочечные антитела, рецептор или лиганд для рецептора), тканеспецифичный регуляторный элемент или регуляторный элемент, специфичный к клеточному типу (например, тканеспецифичный промотор или энхансер, или промотор или энхансер, специфичный к клеточному типу) или сочетание направленного лиганда и тканеспецифичного/специфичного к клеточному типу регуляторного элемента. Предпочтительно направленный лиганд оперативно связан с белком env ретровируса, создавая химерный ретровирусный белок env. Вирусные белки GAG, POL и ENV могут быть получены из любых подходящих ретровирусов (например, MLV или производных лентивирусов). В другом варианте осуществления вирусный белок ENV не является производным ретровируса (например, CMV или VSV).

Рекомбинантный ретровирус по настоящему изобретению, следовательно, является генетически модифицированным, таким образом, чтобы этот вирус был нацелен на конкретный тип клеток (например, гладкомышечные клетки, клетки печени, клетки почек, фибробласты, кератиноциты, мезенхимальные стволовые клетки, клетки костного мозга, хондроциты, эпителиальные клетки, кишечинальные клетки, неопластические клетки, клетки глиомы, нейрональные клетки и другие, известные из уровня техники), так, чтобы полинуклеотидный геном доставлялся в неделящуюся клетку-мишень, делящуюся клетку-мишень или клетку-мишень, имеющую нарушение клеточной пролиферации. Нацеливание может достигаться двумя путями. Первый путь направляет ретровирус к клетке-мишени путем связывания с клетками, имеющими молекулу на внешней поверхности клетки. В этом способе нацеливания ретровируса используется экспрессия направленного лиганда на оболочке ретровируса, чтобы облегчить нацеливание вируса на клетки или ткани, которые имеют рецептор или молекулу связывания, которая взаимодействует с направленным лигандом на поверхности ретровируса. После инфицирования клетки вирусом вирус вводит свою нуклеиновую кислоту в клетку и ретровирусный генетический материал может интегрироваться в геном клетки-хозяина. Во втором способе нацеливания используются регуляторные элементы, специфичные к клеткам или тканям для активации экспрессии и транскрипции вирусного генома в клетке-мишени, которая активно использует регуляторные элементы, как описано ниже более полно. Перенесенный ретровирусный генетический материал затем транскрибируется и транслируется в белки в клетке-хозяине. Направляющий регуляторный элемент обычно связан с 5' и/или 3' LTR, создавая химерный LTR.

Путем встраивания гетерологичной последовательности нуклеиновой кислоты, представляющей интерес, в вирусный вектор по настоящему изобретению, наряду с другим геном, который кодирует, например, лиганд для рецептора на специфической клетке-мишени, вектор теперь является специфичным к мишени. Вирусные векторы могут быть получены специфичными к мишени путем присоединения, например, сахара, гликолипида или белка. Нацеливание может осуществляться путем использования антитела для нацеливания на вирусный вектор. Специалистам в данной области будет известно или они без труда могут определить специфические полинуклеотидные последовательности, которые могут быть встроены в вирусный геном или белки, которые могут быть присоединены к оболочке вируса, чтобы позволить осуществлять нацеленную специфическую доставку вирусного вектора, содержащего последовательность нуклеиновой кислоты, представляющей интерес.

Ретровирус и способы по настоящему изобретению предоставляют репликационно-компетентный ретровирус, для которого не требуется вирус-помощник, или дополнительная последовательность нуклеиновой кислоты, или белки для размножения и продукции вириона. Например, последовательности нуклеиновой кислоты ретровируса по настоящему изобретению кодируют, например, группаспецифичный антиген и обратную транскриптазу (и интегразу, и протеазные ферменты, необходимые для созревания и обратной транскрипции) соответственно, как описано выше. Вирусные gag и pol могут быть получены из лентивируса, такого как HIV или онковируса, такой как MoMLV. Кроме того, полинуклеотидный геном ретровируса по настоящему изобретению включает последовательность, кодирующую белок оболочки вируса (ENV). Ген env может быть получен из любых ретровирусов. Белок env может представлять собой амфотропный оболочечный белок, которые позволяют осуществлять трансдукцию клеток человека или других видов или может представлять собой эктотрофный оболочечный белок, который способен трансдуцировать любые клетки мыши и крысы. Далее, может быть желательным нацелить рекомбинантный вирус путем связывания оболочечного белка с антителом или конкретным лигандом для нацеливания на рецептор конкретного клеточного типа. Как указано выше, ретровирусные векторы могут быть получены специфичным к мишени путем встраивания, например, гликолипида или белка. Нацеливание зачастую осуществляют путем использования антитела для нацеливания ретровирусного вектора на антиген на конкретном клеточном типе (например, клеточный тип, обнаруженный в определенной ткани или типе злокачественных клеток). Специалистам в данной области будут известны или могут быть легко установлены, без проведения лишних экспериментов, специфические способы для достижения доставки ретровирусного вектора к специфичной мишени. В одном варианте осуществления ген env доставляется не из ретровируса (например, CMV или VSV). Примеры ретровирусных env генов включают, но этим не ограничиваются, вирус мышинного лейкоза Молони (MoMuLV), вирус саркомы мышей Харви (HaMuSV), вирус опухоли молочной железы мышей (MuMTV), вирус лейкоза гиббонов (GaLV), вирус иммунодефицита человека (HIV) и вирус саркомы Рауса (RSV). Также могут быть использованы другие env гены, например, вируса везикулярного стоматита (VSV) (белок G), оболочки цитомегаловируса (CMV) или гемагглютинин вируса гриппа (HA).

В одном варианте осуществления ретровирусный геном является производным онкоретровируса или гамма-ретровируса, более конкретно, онкоретровируса или гамма-ретровируса млекопитающих. Под "производным" подразумевается, что родительская полинуклеотидная последовательность представляет собой онковирус дикого типа, который был модифицирован путем встраивания или удаления природных последовательностей (например, вставки IRES, вставки гетерологичного полинуклеотида, кодирующего, например, полипептид, обладающий цитозиндезаминазной активностью по настоящему изобретению и подобного).

В другом варианте осуществления настоящее изобретение относится к ретровирусным векторам, которые направляются с использованием регуляторных последовательностей. Клеточно-специфичные или тканеспецифичные регуляторные последовательности (например, промоторы) могут быть использованы для направления экспрессии генных последовательностей в конкретных популяциях клеток. Подходящие промоторы млекопитающих и вирусные промоторы для настоящего изобретения описаны в других местах настоящего документа. В соответствии с этим в одном варианте осуществления настоящее изобретение относится к ретровирусу, имеющему тканеспецифичные промоторные элементы на 5'-конце ретровирусного генома. Предпочтительно тканеспецифичные регуляторные элементы/последовательности находятся в области U3 LTR ретровирусного генома, включая, например, клеточно- или тканеспецифичные промоторы и энхансеры для неопластических клеток (например, энхансеры и промоторы, специфичные для опухолевых клеток) и индуцибельные промоторы (например, тетрациклин).

В некоторых обстоятельствах может быть желательным регулировать экспрессию. Например, различные вирусные промоторы с различной силой активности могут быть использованы в зависимости от желаемого уровня экспрессии. В клетках млекопитающих ранний промотор CMV зачастую используют для обеспечения сильной транскрипционной активации.

Модифицированные варианты промотора CMV, которые являются менее сильными, также были использованы, когда требовались сниженные уровни экспрессии трансгена. В тех случаях, когда желательна экспрессия трансгена в гематопозитических клетках, могут быть использованы ретровирусные промоторы, такие как LTR из MLV или MMTV. Другие вирусные промоторы, которые могут быть использованы, включают SV40, RSV LTR, HIV-1 и HIV-2 LTR, аденовирусные промоторы, например, из области E1A, E2A или MLP, AAV LTR, вируса мозаики цветной капусты, HSV-TK, и вируса саркомы птиц.

Аналогично, тканеспецифичные или селективные промоторы могут быть использованы для осуществления транскрипции в специфических тканях или клетках, чтобы снизить возможную токсичность или нежелательные эффекты на ткани, не являющиеся мишенями. Например, промоторы, такие как PSA, пробазин, простатическая кислая фосфатаза или простат-специфический glandулярный калликреин (hK2), могут быть использованы для направления генной экспрессии в предстательной железе. Другие промоторы/регуляторные домены, которые могут быть использованы, изложены в табл. 1.

При некоторых показаниях может быть желательным активировать транскрипцию в конкретное время после введения вектора для генной терапии. Это может быть сделано с такими промоторами, которые регулируются гормонами или цитокинами. Например, в терапевтических применениях, при которых показанием является ткань половых желез, в которой продуцируются или распространяются специфические стероидные гормоны, предпочтительным может быть использование андроген- или эстроген-регулируемых промоторов. Такие промоторы, которые являются гормонально регулируемыми, включают MMTV, MT-1, экдизон и RuBisco. Могут быть использованы другие гормон-регулируемые промоторы, например, чувствительные к гормонам щитовидной железы, гипофиза и надпочечников. Промоторы, чувствительные к цитокинам и белкам воспаления, которые могли быть использованы, включают кининоген K и T (Kageyama et al., 1987), c-fos, TNF-альфа, С-реактивный белок (Arcone et al., 1988), гаптоглобин (Oliviero et al., 1987), сывороточный амилоид A2, C/EBP альфа, IL-1, IL-6 (Poli и Cortese, 1989), комплемент C3 (Wilson et al., 1990), IL-8, альфа-1 кислый гликопротеин (Prowse и Baumann, 1988), альфа-1 антитрипсин, липопроотеинлипазу (Zechner et al., 1988), ангиотензиноген (Ron et al., 1990), фибриноген, c-jun (индуцируемый сложными эфирами форбола, TNF-альфа, УФ-излучением, ретиноевой кислотой и перекисью водорода), коллагеназу (индуцируемую сложными эфирами форбола и ретиноевой кислотой), металлотионеин (индуцируемый тяжелым металлом и глюкокортикоидом), стромелизин (индуцируемый сложным эфиром форбола, интерлейкином-1 и EGF), альфа-2 макроглобулин и альфа-1 антихимотрипсин. Опухольспецифичные промоторы, такие как остеокальцин, фактор, чувствительный к гипоксии (HRE), MAGE-4, CEA, альфа-фетопротейн, GRP78/BiP и тирозиназа, также могут быть использованы для регуляции генной экспрессии в опухолевых клетках.

Кроме того, этот перечень промоторов не следует рассматривать как исчерпывающий или ограничивающий, специалистам в данной области будут известны другие промоторы, которые могут быть использованы в сочетании с промоторами и способами, описанными в настоящем документе.

Таблица 2

Тканеспецифичные промоторы

Ткань	Промотор
Поджелудочная железа	Инсулинэластинамилаза pdr-1 pdx-1 глюкокиназа
Печень	Альбумин PERCK HBV энхансер α фетопропротеин аполипопротеин C α -1 антитрипсин вителлогенин, NF-AB транстиретин
Скелетные мышцы	Н цепь миозина мышечная креатинкиназа дистрофин калпаин p94 скелетный альфа- актин быстрый тропонин 1
Кожа	Кератин K6 Кератин K1
Легкие	CFTR цитокератин человека 18 (K18) белки сурфактанта легких A, B и C CC-10 P1
Гладкая мускулатура	sm22 α SM-альфа-актин
Эндотелий	Эндотелин-1 Е-селектин фактор фон Виллебранда TIE, KDR/flk-1 тирозиназа меланоцитов
Жировая ткань	липопротеинлипаза (Zechner et al., 1988) адипсин (Spiegelman et al., 1989) ацетил-CoA карбоксилаза (Pape и Kim, 1989) глицерофосфатдегидрогеназа (Dani et al., 1989) адипоцит P2 (Hunt et al., 1986)
Молочная железа	кислый белок молочной сыворотки (WAP) (Andres et al. PNAS 84:1299-1303 1987)
Кровь	β -глобин

"Тканеспецифичные регуляторные элементы" представляют собой регуляторные элементы (например, промоторы), которые способны направлять транскрипцию гена в одной ткани, оставаясь в основном "молчащими" в тканях других типов. Однако будет понятно, что тканеспецифичные промоторы могут иметь определяемое количество "фоновой" или "базовой" активности в тех тканях, в которых они являются молчащими. Степень, в которой промотор селективно активируется в ткани-мишени, может быть выражена как индекс селективности (активность в ткани-мишени/активность в контрольной ткани). В этом отношении тканеспецифичный промотор, используемый при осуществлении настоящего изобретения, обычно имеет индекс селективности примерно больше 5. Предпочтительно индекс селективности составляет примерно больше 15.

Далее будет понятно, что некоторые промоторы, хотя не ограниченные в активности в отношении одного типа ткани, могут, тем не менее, демонстрировать селективность в том, что они могут быть активными в одной группе тканей и менее активными или молчащими в другой группе. Такие промоторы также называют "тканеспецифичными", и их использование предусматривается настоящим изобретением. В соответствии с этим тканеспецифичные регуляторные элементы, используемые в настоящем изобретении, применимы для регуляции гетерологичных белков, таких как полипептиды, обладающие цитозиндезаминазной активностью по настоящему изобретению, а также применимые в качестве направляющей полинуклеотидной последовательности в ретровирусных векторах.

Ретровирусные векторы и полинуклеотиды CD и полипептиды по настоящему изобретению могут быть использованы для лечения широкого ряда заболеваний и нарушений, включающих ряд клеточно-пролиферативных заболеваний и нарушений (см., например, патенты США № 4405712 и 4650764; Friedmann, 1989, Science, 244:1275-1281; Mulligan, 1993, Science, 260:926-932, R. Crystal, 1995, Science 270:404-410, каждый из которых включен в настоящее описание посредством ссылки в полном объеме, см. также публикацию The Development of Human Gene Therapy, Theodore Friedmann, Ed., Cold Spring

Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1999. ISBN 0-87969-528-5, которая включена в настоящее описание путем ссылки в полном объеме).

Таким образом, настоящее изобретение относится к рекомбинантному ретровирусу, способному инфицировать неделящиеся клетки, делящиеся клетки или неопластические клетки, где рекомбинантный ретровирус содержит вирусный GAG; вирусный POL; вирусный ENV; гетерологичную нуклеиновую кислоту (например, содержащему полипептид по настоящему изобретению, обладающий цитозиндезаминазной активностью), оперативно связанную с IRES; и действующие в цис-положении последовательности нуклеиновой кислоты, необходимые для упаковки, обратной транскрипции и интеграции. Рекомбинантный ретровирус может представлять собой лентивирус, такой как HIV, или может представлять собой онковирус. Как описано выше для способа получения рекомбинантного ретровируса, рекомбинантный ретровирус по настоящему изобретению дополнительно может содержать по меньшей мере один из белков VPR, VIF, NEF, VPX, TAT, REV и VPU. Не желая связываться конкретной теорией, считается, что один или несколько из этих генных/белковых продуктов являются важными для повышения титра вируса продуцированного рекомбинантного ретровируса (например, NEF) или могут быть необходимыми для инфицирования и упаковки вириона.

Например, промотор может содержать CMV промотор, имеющий последовательность SEQ ID NO: 19, 20 или 22 от нуклеотида 1 до примерно нуклеотида 582, и может включать модификацию одного или нескольких (например, 2-5, 5-10, 10-20, 20-30, 30-50, 50-100 или более) оснований нуклеиновой кислоты при условии, что модифицированный промотор способен направлять и инициировать транскрипцию. В одном варианте осуществления промотор или регуляторная область содержит полинуклеотидный домен CMV-R-U5. Домен CMV-R-U5 содержит ранний промотор цитомегаловируса человека для области MLV R-U5. В одном варианте осуществления полинуклеотидный домен CMV-R-U5 содержит последовательность SEQ ID NO: 19, 20 или 22 примерно от нуклеотида 1 до примерно нуклеотида 1202 или последовательность, которая по меньшей мере на 95% идентична последовательности SEQ ID NO: 19, 20 или 22, где указанный полинуклеотид активизирует транскрипцию молекулы нуклеиновой кислоты, оперативно связанной с ним. Домен gag этого полинуклеотида может быть получен из любого числа ретровирусов, но обычно его получают из онкоретровируса и, более конкретно, из онкоретровируса млекопитающих. В одном варианте осуществления домен gag содержит последовательность примерно от нуклеотида номер 1203 до примерно нуклеотида 2819 или последовательность, по меньшей мере на 95, 98, 99 или 99,8% (округленно до ближайшего 10-го) идентичную ей. Домен pol этого полинуклеотида может быть получен из любого числа ретровирусов, но обычно его получают из онкоретровируса и, более конкретно, из онкоретровируса млекопитающих. В одном варианте осуществления домен pol содержит последовательность примерно от нуклеотида номер 2820 до примерно нуклеотида 6358 или последовательность, по меньшей мере на 95, 98, 99 или 99,9% (округленно до ближайшего 10-го) идентичную ей. Домен env этого полинуклеотида может быть получен из любого числа ретровирусов, но обычно его получают из онкоретровируса и, более конкретно, из онкоретровируса млекопитающих. В некоторых вариантах осуществления кодирующий домен env содержит амфотропный домен env. В одном варианте осуществления домен env содержит последовательность примерно от нуклеотида номер 6359 до примерно нуклеотида 8323 или последовательность, по меньшей мере на 95, 98, 99 или 99,8% (округленно до ближайшего 10-го) идентичную ей. Домен IRES домен этого полинуклеотида может быть получен из любого числа внутренних сайтов связывания рибосомы. В одном варианте осуществления IRES получен из вируса энцефаломиокардита. В одном варианте осуществления домен IRES содержит последовательность примерно от нуклеотида номер 8327 до примерно нуклеотида 8876 или последовательность, по меньшей мере на 95, 98 или 99% (округленно до ближайшего 10-го) идентичную ей, при условии, что этот домен позволяет осуществлять связывание рибосомы. Гетерологичный домен может содержать цитозиндезаминазу по настоящему изобретению. В одном варианте осуществления полинуклеотид CD содержит последовательность кодонов, оптимизированных для экспрессии в клетках человека. Еще в одном варианте осуществления полинуклеотид CD кодирует мутантный полипептид, обладающий цитозиндезаминазной активностью, которому мутации придают повышенную термостабильность, повышающей температуру плавления (T_m) на 10°C, позволяя задерживать кинетическую активность в широком температурном интервале и повышать накопленные уровни белка. В одном варианте осуществления цитозиндезаминаза содержит последовательность SEQ ID NO: 19 примерно от нуклеотида номер 8877 до примерно 9353. За гетерологичным доменом может следовать домен, обогащенный пурином. 3' LTR может быть получен из любого числа ретровирусов, обычно онкоретровируса и предпочтительно онкоретровируса млекопитающих. В одном варианте осуществления 3' LTR содержит домен U3-R-U5. Еще в одном варианте осуществления LTR содержит последовательность SEQ ID NO: 19 примерно от нуклеотида 9405 примерно до 9998 или последовательность, которая по меньшей мере на 95, 98 или 99,5% (округленно до ближайшего 10-го) идентичную ей.

Настоящее изобретение также относится к рекомбинантному ретровирусному вектору, содержащему от 5' к 3' CMV-R-U5, слияние раннего промотора цитомегаловируса человека с областью MLV R-U5; PBS, сайт связывания праймера с обратной транскриптазой; 5' сайт сплайсинга; сигнал упаковки Ψ; gag, ORF для MLV группоспецифичного антигена; pol, ORF для полипротеина MLV полимеразы; 3' сайт

сплайсинга; 4070A env, ORF для белка оболочки MLV штамма 4070A; IRES, внутренний сайт связывания рибосомы цитомегаловируса; модифицированную цитозиндезаминазу (термостабильную с оптимизированными кодонами); PPT, полипуриновый тракт; и U3-R-U5, MLV длинный концевой повтор. Эта структура дополнительно представлена на фиг. 1.

Настоящее изобретение также относится к ретровирусному вектору, содержащему последовательность SEQ ID NO: 19, 20 или 22.

Для примера, а не для ограничения, ретровирусный вектор, используемый при лечении нарушения клеточной пролиферации, будет включать амфотропный белок ENV, GAG и POL белки, промоторную последовательность в области U3 ретровирусного генома и любую последовательность, действующую в цис-положении, необходимую для репликации, упаковки и интеграции ретровирусного генома в клетку-мишень.

Следующие примеры предназначены для иллюстрации, а не для ограничения этого изобретения. Хотя такие примеры являются характерными из тех, которые могли быть использованы, другие процедуры, известные специалистам в данной области, могут быть использованы альтернативно.

Примеры

Пример 1. Конструирование модифицированных генов CD и встраивание в плазмидные векторы.

Была проведена генетическая оптимизация гена дикого типа цитозиндезаминазы дрожжей, чтобы он включал (1) три позиционные мутации, которые меняют три аминокислоты (A23L, I140L и V108I) для повышения термостабильности белка цитозиндезаминазы дрожжей; и (2) дополнительные модификации генной последовательности для повышения частоты использования кодонов человека для повышения эффективности трансляции белка в клетках человека без дополнительных изменений аминокислотной последовательности.

Конструкция последовательности для CD, включала CD оптимизированную, CD-UPRT (+/- линкер) и CD-OPRT-азу (+/-линкер). Конечная последовательность, кодирующая цитозиндезаминазу, может содержать на 5'-конце сайт PSII (полноразмерный) и на 3'-конце сайт NotI плюс поли A хвост для методики на основе кассеты PSII/NotI.

Следующая последовательность, содержащая цитозиндезаминазу дрожжей, была использована для клонирования, оптимизации и мутации (нуклеиновые кислоты в рамке содержат сайты рестрикции, используемые в последующих способах клонирования):

```

AACACGA[TTATAA]ATGGTGACAGGGGGAATGGCAAGCAAGTGGGATCAGAAGGGTATGGACAT
TGCCTATGAGGAGGCGCCTTAGGTTACAAAGAGGGTGGTGTTTCTATTGGCGGATGTCTTAT
CAATAACAAAGACGGAAGTGTTCTCGGTCTGGTGCACACATGAGATTTCAAAGGGATCCGC
CACACTACATGGTGAGATCTCCACTTTGGAACAACTGTGGGAGATTAGAGGGCAAAGTGACAA
AGATACCACTTTGTATACGACGCTGTCTCCATGCGACATGTGTACAGGTGCCATCATCATGTA
TGGTATTCCACGCTGTGTTGTCTGGTGAGAACGTTAATTTCAAAGTAAGGGCGAGAAATATTT
ACAAACTAGAGGTCACGAGGTTGTTGTTGTTGACGATGAGAGGTGTAAAAGATCATGAAACA
ATTTATCGATGAAAGACCTCAGGATTGGTTTGAAGATATTGGTGAGTAG[GCGGCCGC]GCCATA
GATAAAATAAAAGATTTTATTTAGTCTCCAGAAAAAGGGGGG (SEQ ID NO:31)

```

В табл. кратко изложены гены и полученные в результате плазмидные векторы, которые были получены, и их названия.

Таблица 3

Векторные конструкции и названия

Тосаген Код	Наименование	Первоначальное название	5'LTR пром	оболочка	Вектор	IRES	Трансген	3'LTR
T5.0000	pACE-yCD	pACE-CD (Tai et al. 2005)	CMV	Ampho (4070A)	pACE	EMCV	CD дрожжей дикого типа	MLV U3
T5.0001	pAC3-yCD1	CDopt последовательность	CMV	Ampho (4070A)	pAC3	EMCV	Модифицированная CD	MLV U3
T5.0002	pAC3-yCD2	CDopt+3pt	CMV	Ampho (4070A)	pAC3	EMCV	Модифицированная CD	MLV U3
T5.0003	pAC3-yCD2-U	CDopt+3pt-UPRT	CMV	Ampho (4070A)	pAC3	EMCV	CD2-UPRT	MLV U3
T5.0004	pAC3-yCD2-O	CDopt+3pt-OPRT	CMV	Ampho (4070A)	pAC3	EMCV	CD2-OPRT	MLV U3
T5.0005	pAC3-yCD2-LO	CDopt+3pt-LINK-OPRT	CMV	Ampho (4070A)	pAC3	EMCV	CD2-L-OPRT	MLV U3
T5.0006	pAC3-eGFP	pAC3-emd, pAC3GFP	CMV	Ampho (4070A)	pAC3	EMCV	Изумрудный GFP	MLV U3
T5.0007	pAC3-yCD	pAC3-yCD	CMV	Ampho (4070A)	pAC3	EMCV	CD дрожжей дикого типа	MLV U3

Репликационно-компетентный ретровирусный вектор, описанный Kasahara et al. pACE-CD (патент США № 6899871, описание которого включено в настоящий документ), был использован в качестве основы для дополнительных модификаций. Вектор (pAC3-yCD) был модифицирован для экспрессии модифицированного гена цитозиндезаминазы дрожжей, как описано в настоящей заявке, и был использован в этих конструкциях. См. 1А ниже для диаграммы векторной конструкции для исходного трансфицированного репликационно-компетентного ретровируса. CMV представляет собой ранний промотор CMV человека, U3, R и U5 представляют собой соответствующие области вирусного длинного концевой повтора (LTR). Gag, pol и env представляют собой области, кодирующие вирусные белки. 1B и 1D демонстрируют плазмидную структуру и последовательность по настоящему изобретению. После синтеза генов в контракторе (Bio Basic Inc., Markham, Ontario, Canada) их встраивали в сайт PstI-NotI основы вектора pAC3 (фиг. 1). Основу плазмиды обычно получают путем разрезания плазмиды pAC3-eGFP с использованием PstI и NotI и очистки крупного (около 11 тыс. пар оснований) фрагмента, кодирующего плазмиду и основу ретровируса).

А. Гуманизированный кодон-оптимизированный ген CD (CDopt, aka CD1, T5.0001).

Сравнение цитозиндезаминазы с кодонами, оптимизированными для экспрессии в клетках человека по Conrad et al. и PCT WO 99/60008 указывает на 91 суммарный кодон, оптимизированный у обоих, 36 кодонов идентичных, 47 кодонов имели изменения в третьей паре оснований (все кодируют одну и ту же аминокислоту) и 9 кодонов были различными (однако они кодировали одну и ту же аминокислоту). Из 9 кодонов, которые были различными:

AGC (Ser) на TCC (Ser)

CGT (Arg) на AGG (Arg)

CCA (Pro) на CCT (Pro)

все имели эквивалентное содержание GC и кодируют одну и ту же аминокислоту.

Нативная последовательность гена дрожжей была отдельно кодон-оптимизирована для получения следующего гена CD (CD1) и была названа T5.0001 при встраивании в плазмидный вектор pAC3, который кодирует репликационно-компетентный ретровирус (RCR) с IRES.

TTATAAATGGGTGACCGGCGGCATGGCCTCCAAGTGGGATCAAAAGGGCATGGATATCGCTTAC
GAGGAGGCCCGCCTGGGCTACAAGGAGGGCGGCGTGCCTATCGGCGGCTGTCTGATCAACAAC
AAGGACGGCAGTGTGCTGGGCAGGGGCCACAACATGAGGTTCCAGAAGGGCTCCGCCACCCTG
CACGGCGAGATCTCCACCCTGGAGAAGTGTGGCAGGCTGGAGGGCAAGGTGTACAAGGACACC
ACCCTGTACACCACCCTGTCCCTTGTGACATGTGTACCGGCGCTATCATCATGTACGGCATC
CCTAGGTGTGTGGTGGGCGAGAACGTGAAGTCCAAGGGCGAGAAGTACCTGCAAACC
AGGGGCCACGAGGTGGTGGTGTGTTGACGATGAGAGGTGTAAGAAGATCATGAAGCAGTTTCATC
GACGAGAGGCCTCAGGACTGGTTTCGAGGATATCGGCGAGTGATAAGCGGCGCGAGATAAAATA
AAAGATTTTATTATTAGTCTCCAGAAAAAGGGGGG. (SEQ ID NO:32)

В. Термостабилизированный ген CD.

Были проведены дополнительные модификации для повышения стабильности цитозиндезаминазы. Генетические модификации гена дикого типа цитозиндезаминазы дрожжей были проведены таким образом, чтобы этот ген содержал три позиционные мутации, которые меняют три аминокислоты (A23L, I140L и V108I) для повышения термостабильности белка цитозиндезаминазы дрожжей.

Следующие пары праймеров использовали в получении гена для полипептида цитозиндезаминазы по настоящему изобретению:

СМЫСЛОВОЙ: 5'-tcgaggatatcggcgagtgaaacccgttattctttttggc-3'

(SEQ ID NO:25)

АНТИСМЫСЛОВОЙ:

5'-

gccaaaaagaataacgggtttcactcgccgatatactcga-3' (SEQ ID NO:26)

СМЫСЛОВОЙ:

5'tcggcgagtgatccggcgcgcgccctccggcgcgcgccctccggcgcgcgcc

tccggcgcgcgcccaacccgttatt-3' (SEQ ID NO:27)

АНТИСМЫСЛОВОЙ:5'-

aataacgggttggcgccgcccggaggcgccgcccggaggcgccgcccgc

cggaggcgccgcccggatcactcgccga-3' (SEQ ID NO:28)

Для повышения стабильности нативной CD дрожжей три аминокислотные замены были встроены в этот белок. Эти замены были отдельными или в сочетании с оптимизацией кодонов для экспрессии в клетках человека.

Три аминокислотные замены представляли собой: A23L, V108I, I140L. Последовательность, кодирующая эти замены, показана ниже.

ATGGTGACAGGGGAATGGCAAGCAAGTGGGATCAGAAGGGTATGGACATTGCCTATG
AGGAGGCGC**TTA**TTAGGTTACAAAGAGGGTGGTGTTCCTATTGGCGGATGTCTTATCAATAACA
AAGACGGAAGTGTCTCGGTCTGGTCACAACATGAGATTTCAAAGGGATCCGCCACACTAC
ATGGTGAGATCTCCACTTTGGAAGCTGTGGGAGATTAGAGGGCAAAGTGTACAAAGATACCA
CTTTGTATACGACGCTGTCTCCATGCGACATGTGTACAGGTGCCATCATCATGTATGGTATTC
CACGCTGTGT**CAT**CGGTGAGAACGTTAATTTCAAAGTAAGGGCGAGAAATATTTACAACTA
GAGGTCACGAGGTTGTTGTTGACGATGAGAGGTGTAAAAAG**TTA**ATGAAACAATTTATCG
ATGAAAGACCTCAGGATTGGTTTGAAGATATTGGTGAGTAG**GCGCCGC**GCCATAGATAAAAT
AAAAAGATTTTATTTAGTCTCCAGAAAAAGGGGG (SEQ ID NO:33)

Кодируемый полипептид содержит следующую последовательность (замещенные аминокислоты выделены жирным шрифтом и подчеркнуты):

1 MVTGGMASKWDQKGMIDIAE**EAL**LG^YKEGGVPIGGCLINNKDGSVLGRGHNMRFQKGSAT
61 LHGEISTLENCGRLEGKVYKDTTLYTTLSPCDMCTGAIIMYGIPRCV**I**GENVNFKSKGEK
121 YLQTRGHEVVVDDERCK**L**MKQFIDERPDWFDIGE-

Создание конечной конструкции, которая включает 3 аминокислотные замены A23L/V108I/I140L, используя предпочтительные кодоны, и использует предпочтительную частоту использования кодонов человека для полной последовательности (этот ген называется CDopt+3pt [aka CD2] и T5.0002 при встраивании в плазмидный вектор pAC3, который кодирует RCR с IRES) (SEQ ID NO: 34):

1 ATGGTGACCGGCGGCATGGCCTCCAAGTGGGATCAAAAGGGCATGGATATCGCTTACGAG
61 GAGGCCCTGCTGGGCTACAAGGAGGGCGGCGTGCCTATCGGCGGCTGTCTGATCAACAAC
121
AAGGACGGCAGTGTGCTGGGCAGGGGCCACAACATGAGGTTCCAGAAGGGCTCCGCCACC
181
CTGCACGGCGAGATCTCCACCCTGGAGAACTGTGGCAGGCTGGAGGGCAAGGTGTACAAG
241
GACACCACCCTGTACACCACCCTGTCCCCTTGTGACATGTGTACCGGCGCTATCATCATG
301
TACGGCATCCCTAGGTGTGTG**AT**CGGCGAGAACGTGAACTTCAAGTCCAAGGGCGAGAAG
361
TACCTGCAAACCAGGGGCCACGAGGTGGTGGTTGTTGACGATGAGAGGTGTAAGAAG**CTG**
421
ATGAAGCAGTTCATCGACGAGAGGCCTCAGGACTGGTTCGAGGATATCGGCGAGTGATAA

Подчеркнутые кодоны обозначают предпочтительные кодоны для аминокислотных замен.

Конструкция CD оптимизированной последовательности (предпочтительность кодонов человека + 3 аминокислотные замены)

AACACGATTATAAATGGTGACCGGCGGCATGGCCTCCAAGTGGGATCAAAAGGGCATGGATAT
CGCTTACGAGGAGGCCCTGCTGGGCTACAAGAGGGCGGCGTGCCTATCGGCGGCTGTCTGAT
CAACAACAAGGACGGCAGTGTGCTGGGCGAGGGCCACAACATGAGGTTCCAGAAGGGCTCCGC
CACCTGACGCGGAGATCTCCACCCTGGAGAACTGTGGCAGGCTGGAGGGCAAGGTGTACAA
GGACACCACCCTGTACACCACCCTGTCCCCTTGTGACATGTGTACCGGCGCTATCATCATGTA
CGGCATCCCTAGGTGTGTGATCGGCGAGAACGTGAACTTCAAGTCCAAGGGCGAGAAGTACCT
GCAAAACAGGGGCCACGAGGTGGTGGTTGTGACGATGAGAGGTGTAAGAAGCTGATGAAGCA
GTTTCATCGACGAGAGGCCTCAGGACTGGTTCGAGGATATCGGCGAGTAAGCGGCCGGCCATA
GATAAAATAAAAGATTTTATTTAGTCTCCAGAAAAAGGGGGG (SEQ ID NO:35)

С. Конструкция слитого гена CD-UPRT (CDopt+3pt-UPRT, [aka CDopt-UPRT и CD2-UPRT], T5.0003 в pAC3 плазмидном RCR векторе).

Также была разработана слитая конструкция, содержащая полипептид CD, как описано выше, связанный с полипептидом UPRT для получения CD оптимизированной-UPRT последовательности с использованием схемы I, как изложено на фиг. 2А. Следующие праймеры были использованы для делеции стоп-старт между CD и UPRT.

Последовательности праймеров

Название праймера	Последовательность праймера (5'-3') (SEQ ID NO:)			
del118-123	5'-tcgaggatatcggcgagtgaacccggtattctttttggc-3' (36)			
del118-123-антисмысловой	5'-gccaaaaagaataacgggtttcactcgccgatatcctcga-3' (37)			
Название праймера	Длина (нуклеотидов)	Tm	Энергия дуплекса при 68°C	Затрата энергии ошибочно спаренных оснований
del118-123	40	79,06°C	-44,37 ккал/моль	21,1%
del118-123-антисмысловой	40	79,06°C	-47,95 ккал/моль	20,3%
Название праймера	Дуплекс праймер-матрица			
del118-123	5'-tcgaggatatcggcgagtga-----aaccggtattctttttggc-3' ccaagctcctatagccgctcactatctacttgggcaataagaaaaaccgaag			
del118-123-антисмысловой	gggttcgaggatatcggcgagtgatagatgaacccggtattctttttggcttc 3'-agctcctatagccgctcact-----ttgggcaataagaaaaaccg			

Полученный в результате слитый полинуклеотид содержит 1296 пар оснований и последовательность, изложенную непосредственно ниже:

AACACGA~~TTATAA~~ATGGTGACCGGCGGCATGGCCTCCAAGTGGGATCAAAAGGGCATGGATAT
 CGCTTACGAGGAGGCCCTGCTGGGCTACAAGGAGGGCGCGTGCCTATCGGCGGCTGTCTGAT
 CAACAACAAGGACGGCAGTGTGCTGGGCAGGGGCCACAACATGAGGTTCCAGAAGGGCTCCGC
 CACCCTGCACGGCGAGATCTCCACCCTGGAGAACTGTGGCAGGCTGGAGGGCAAGGTGTACAA
 GGACACCACCCTGTACACCACCCTGTCCCCTTGTGACATGTGTACCGGCGCTATCATCATGTA
 CGGCATCCCTAGGTGTGTGATCGGCGGAGAAGCTGAACTTCAAGTCCAAGGGCGAGAAGTACCT
 GCAAACCAGGGGCCACGAGGTGGTGGTTGTTGACGATGAGAGGTGTAAGAAGCTGATGAAGCA
 GTTCATCGACGAGAGGCCTCAGGACTGGTTCGAGGATATCGGCGGAGAACCCGTTATTCTTTTT
 GGCTTCTCCATTCTTGTACCTTACATATCTTATATATTATCCAAACAAAGGGTCTTTCGTTAG
 CAAACCTAGAAATCTGCAAAAATGTCTTCGGAACCATTTAAGAAGCTCTACTTGCTACCTCA
 AACAAACCAATTGCTGGGTTTGTACACCATCATCAGAAATAAGAATACAACCTAGACCTGATTT
 CATTTTCTACTCCGATAGAATCATCAGATTGTTGGTTGAAGAAGGTTTGAACCATCTACCTGT
 GCAAAAGCAAATTTGTGAAACTGACACCAACGAAACTTCGAAGGTGTCTCATTCATGGGTAA
 AATCTGTGGTGTTCATTTGTCAGAGCTGGTGAATCGATGGAGCAAGGATTAAGAGACTGTTG
 TAGGTCTGTGCGTATCGGTAATAATTTTAATTCAAAGGGACGAGGAGACTGCTTTACCAAAGTT
 ATTCTACGAAAAATTACCAGAGGATATATCTGAAAGGTATGTCTTCCTATTAGACCCAATGCT
 GGCCACCGGTGGTAGTGCTATCATGGCTACAGAAGTCTTGATTAAGAGAGGTGTTAAGCCAGA
 GAGAAATTTACTTCTTAAACCTAATCTGTAGTAAGGAAGGGATTGAAAAATACCATGCCGCCTT
 CCCAGAGGTCAGAATTGTTACTGGTGCCCTCGACAGAGGTCTAGATGAAAACAAGTATCTAGT
 TCCAGGGTTGGGTGACTTTGGTGACAGATACTACTGTGTTTAA~~GCGGCCGC~~GCCATAGATAAA
 ATAAAAGATTTTATTAGTCTCCAGAAAAAGGGGG (SEQ ID NO:38)

D. Конструирование слитого гена CD-линкер UPRT (CDopt+3pt-LINK-UPRT [aka CDopt-LINKER-UPRT и CD2-L-UPRT]).

Слитая конструкция также была получена путем клонирования линкерного (Ser-Gly-Gly-Gly-Gly)₄ домена между и в рамке считывания с CD полипептидом и UPRT полипептидом для получения CD оптимизированной-линкер-UPRT последовательности с использованием схемы II, как представлено в 2B. Для встраивания линкера были использованы следующие праймеры.

Название праймера	Последовательность праймера (5' - 3') (SEQ ID NO:)			
ins_60nt_after_477	5' - tcggcgagtgatccggcgggcggcctccggcgggcgccctccggcg gcgggcgccctccggcgggcgccaacccgttatt-3' (39)			
ins_60nt_after_477- антисмысловой	5' - aataacgggttgggcgccgcccggaggcgccgcccggaggcgcc gcccggaggcgccgcccggatcactcgccga-3' (40)			
Название праймера	Длина (нуклеотидов)	Tm	Энергия дуплекса при 68°C	Затрата энергии ошибочно спаренных оснований
ins_60nt_after_477	82	79,77°C	-30,19 ккал/моль	83,3%
ins_60nt_after_477- антисмысловой	82	79,77°C	-32,31 ккал/моль	82,2%

Полученная в результате конструкция имеет размер: 1356 пар оснований и последовательность, представленную непосредственно ниже:

AACACGA**TTATAA**ATGGTGACCGGCGGCATGGCCTCCAAGTGGGATCAAAAGGGCATGGATAT
CGCTTACGAGGAGGCCCTGCTGGGCTACAAGGAGGGCGGCGTGCCTATCGGCGGCTGTCTGAT
CAACAACAAGGACGGCAGTGTGCTGGGCAGGGGCCACAACATGAGGTTCCAGAAGGGCTCCGC
CACCCCTGCACGGCGAGATCTCCACCCTGGAGAACTGTGGCAGGCTGGAGGGCAAGCTGTACAA
GGACACCACCCTGTACACCACCCTGTCCCTTGTGACATGTGTACCGGCGCTATCATCATGTA
CGGCATCCCTAGGTGTGTGATCGGCGAGAACGTGAACTTCAAGTCCAAGGGCGAGAAGTACCT
GCAAACCAGGGGCCACGAGGTGGTGGTTGTTGACGATGAGAGGTGTAAGAAGCTGATGAAGCA
GTTTCATCGACGAGAGGCCTCAGGACTGGTTCGAGGATATCGGCGAGTCCGGCGGGCGGCGCCTC
CGGCGGGCGGCGCTCCGGCGGGCGGCGCCTCCGGCGGGCGGCCAACCCGTTATTCTTTTTGGC
TTCTCCATTCTTGTACCTTACATATCTTATATATTATCCAAACAAAGGGTCTTTTCGTTAGCAA
ACCTAGAAATCTGCAAAAATGTCTTCGGAACCATTTAAGAACGTCTACTTGCTACCTCAAAC
AAACCAATTGCTGGGTTTGTACACCATCATCAGAAATAAGAATACAACCTAGACCTGATTTTCAT
TTTCTACTCCGATAGAATCATCAGATTGTGGTTGAAGAAGGTTTGAACCATCTACCTGTGCA
AAAGCAAATTGTGGAACTGACACCAACGAAAACCTCGAAGGTGTCTCATTTCATGGGTAAAAT
CTGTGGTGTTCATTGTGAGAGCTGGTGAATCGATGGAGCAAGGATTAAGAGACTGTTGTAG
GTCTGTGCGTATCGGTAAAATTTAATTCAAAGGGACGAGGAGACTGCTTTACCAAAGTTATT
CTACGAAAAATTACCAGAGGATATATCTGAAAGGTATGTCTTCCTATTAGACCCAATGCTGGC
CACCGGTGGTAGTGCTATCATGGCTACAGAAGTCTTGATTAAGAGAGGTGTTAAGCCAGAGAG
AATTTACTTCTTAAACCTAATCTGTAGTAAGGAAGGGATTGAAAAATACCATGCCGCCTTCCC
AGAGGTGAGAATTGTTACTGGTGCCCTCGACAGAGGTCTAGATGAAAACAAGTATCTAGTTCC
AGGGTTGGGTGACTTTGGTGACAGATACTACTGTGTTTAA
GCGGCCGCGCCATAGATAAAATAAAAGATTTTATTTAGTCTCCAGAAAAGGGGGG (SEQ
ID NO:41)

Е. Конструирование слитого гена CD-OPRT (CDopt+3pt-OPRT [aka CDopt-OPRT и CD2-OPRT], T5.0004 при встраивании в pAC3 плазмидный RCR вектор).

Также была получена слитая конструкция, содержащая CD полипептид, как описано выше, соединенный с полипептидом OPRT для получения CD оптимизированной-OPRT-азы (CD гуманизированный + 3pt мутантный + OPRT-аза функциональный домен человека) с использованием схемы III, как показано на фиг. 2С.

Полученная в результате конструкция имеет размер 1269 пар нуклеотидов и последовательность, представленную непосредственно ниже:

```
AACACGATTATAAATGGTGACCGGCGGCATGGCCTCCAAGTGGGATCAAAAGGGCATGGATAT
CGCTTACGAGGAGGCCCTGCTGGGCTACAAGGAGGGCGGCGTGCCTATCGGCGGCTGTCTGAT
CAACAACAAGGACGGCAGTGTGCTGGGCAGGGGCCACAACATGAGGTTCAGAAGGGCTCCGC
CACCTGACGCGGAGATCTCCACCTGGAGAACTGTGGCAGGCTGGAGGGCAAGGTGTACAA
GGACACCACCTGTACACCACCTGTCCCCTGTGACATGTGTACCGGCGCTATCATCATGTA
CGGCATCCCTAGGTGTGTGATCGGCGAGAACGTGAACTTCAAGTCCAAGGGCGAGAAGTACCT
GCAAACCAGGGGCCACGAGGTGGTGGTTGTTGACGATGAGAGGTGTAAGAAGCTGATGAAGCA
GTTTCATCGACGAGAGGCCTCAGGACTGGTTCGAGGATATCGGCGAGGCGGTCGGTCGGTcagc
tttggggccattggtgacgggtctgtacgacgtgcaggctttcaagtttggggacttcgtgct
gaagagcgggctttcctcccccatctacatcgatctgcggggcatcgtgtctcgaccgctct
tctgagtcagggttgagatattttattccaaactgccccaaatgcaggcatcagttttgacac
cgtgtgtggagtgccttatacagctttgccattgggtacagttatctgttcaaccaatcaaat
tccaatgcttattagaaggaaagaacaaggattatggaactaagcgtcttgtagaaggaaac
tattaatccaggagaaacctgtttaatcattgaagatgttgccaccagtggtctagtggttt
ggaaactgttgaggttcttcagaaggagggttgaaaggtcactgatgccatagtgctgttgga
cagagagcagggaggcaaggacaagttgcaggcgacgggatccgcctccactcagtggtgtac
attgtccaaatgctggagattctcgagcagcagaaaaagttgatgctgagacagttgggag
agtgaagaggtttattcaggagaatgtctttgtggcagcgaatcataatggttctcccccttc
tataaaggaagcacccaaagaactcAGCTTCGGTGCACGTGCAGAGCTGCCCAGGATCCACCC
AGTTGCATCGAAGTAAGCGGCCGCGCCATAGATAAAATAAAAGATTTTATTTAGTCTCCAGAA
AAAGGGGGG (SEQ ID NO:42)
```

Ф. Конструирование слитого гена CD-линкер-OPRT (CDopt+3pt-LINK-OPRT, [aka CDopt-LINKER-OPRT и CD2-L-OPRT], T5.0005 в pAC3 плазмидном RCR векторе).

Слитая конструкция также была получена путем клонирования линкерного (Ser-Gly-Gly-Gly-Gly)₄ домена между и в рамке считывания с CD полипептидом и OPRT полипептидом для получения CD оптимизированной-линкер-OPRT последовательности с использованием схемы IV, как показано на фиг. 2D.

Полученная в результате конструкция имеет размер 1329 пар нуклеотидов и последовательность, представленную непосредственно ниже:

```
AACACGATTATAAATGGTGACCGGCGGCATGGCCTCCAAGTGGGATCAAAAGGGCATGGATAT
CGCTTACGAGGAGGCCCTGCTGGGCTACAAGGAGGGCGGCGTGCCTATCGGCGGCTGTCTGAT
CAACAACAAGGACGGCAGTGTGCTGGGCAGGGGCCACAACATGAGGTTCAGAAGGGCTCCGC
CACCTGACGCGGAGATCTCCACCTGGAGAACTGTGGCAGGCTGGAGGGCAAGGTGTACAA
GGACACCACCTGTACACCACCTGTCCCCTGTGACATGTGTACCGGCGCTATCATCATGTA
CGGCATCCCTAGGTGTGTGATCGGCGAGAACGTGAACTTCAAGTCCAAGGGCGAGAAGTACCT
GCAAACCAGGGGCCACGAGGTGGTGGTTGTTGACGATGAGAGGTGTAAGAAGCTGATGAAGCA
GTTTCATCGACGAGAGGCCTCAGGACTGGTTCGAGGATATCGGCGAGTCCGGCGGCGGCGCCTC
CGGCGGCGGCGCCTCCGGCGGCGGCGCCTCCGGCGGCGGCGCGGTGCTCGTgcagcttt
ggggccattggtgacgggtctgtacgacgtgcaggctttcaagtttggggacttcgtgctgaa
```

gagcgggctttctccccatctacatcgatctgcggggcatcgtgtctcgaccggtcttct
gagtcaggttgagatattttattccaaactgccccaaatgcaggcatcagttttgacaccgt
gtgtggagtgcttatacagctttgccattggctacagttatctgttcaaccaatcaaattcc
aatgcttattagaaggaaagaaacaaaggattatggaactaagcgtctttagaaggaaactat
taatccaggagaaacctgtttaatcattgaagatgttggtcaccagtggtctagtgttttggg
aactgttgaggttcttcagaaggagggttgaaaggtcactgatgccatagtgtgttgacag
agagcagggaggcaaggacaagttgcaggcgacgggatccgcctccactcagtggtgtacatt
gtccaaaatgctggagattctcgagcagcagaaaaaagttgatgctgagacagttgggagagt
gaagaggtttatttcaggagaatgtctttgtggcagcgaatcataatggttctccctttctat
aaaggaagcacccaaagaactcaGCTTCGGTGCACGTGCAGAGCTGCCAGGATCCACCCAGT
TGCATCGAAGTAA~~GCGGCCGC~~GCCATAGATAAAAAATAAAGATTTTATTAGTCTCCAGAAAAA
GGGGGG. (SEQ ID NO:43)

Пример 2. Получение инфицирующего вектора.

Вектор может быть получен целым рядом способов, но первым этапом является введение плазмидной ДНК вектора в клетки, для возможности продукции инфекционных частиц, которые затем могут быть собраны из клеточного супернатанта. После получения инфекционных частиц другие способы получения могут быть выполнены специалистами в данной области. Векторные частицы получали путем транзитной трансфекции клеток 293T (Pear et al. Proc. Natl. Acad. Sci. USA. 90:8392-8396 1993). Клетки 293T размораживали и помещали в культуру, затем дважды пересеивали во флаконы T-75, содержащие 15 мл среды DMEM, которая была получена путем перемешивания среды DMEM с высоким содержанием глюкозы (Hyclone# 30081, 500 мл) с FBS (Hyclone# SH30070, 50 мл), L-глутамином (Cellgro # 25-005-CI, 5 мл), NEAA (Hyclone #SH30238, 5 мл), и пенициллин-стреп (Cellgro # 30-002-CI, 5 мл). Эти флаконы инкубировали при 37°C и 5% CO₂. После 3-го пассажа клетки высевали в 6 флаконов T-25, содержащих каждый 5 мл среды, при плотности клеток 1,8×10⁶ клеток/T-25 (или 7,2×10⁴ клеток/см²). Через один день после посева T-25 клетки трансфицировали плазмидой T5.0002, которая экспрессировала вирусный вектор, с использованием набора для трансфекции фосфатом кальция (Calcium Phosphate Transfection Kit) от компании Promega (Cat# E1200). Через 18 ч после трансфекции среду в одном комплекте флаконов (3 флакона в каждом комплекте) заменяли свежей средой, содержащей 10 мМ NaB. Среду во 2-м комплекте флаконов не заменяли, который служил в качестве контроля (ноль NaB). Через 8 ч после обработки NaB среду во всех флаконах заменяли свежей средой, не содержащей NaB. В обоих комплектах флаконов допускали продолжение экспрессии до следующего дня (продолжительность 22 ч). Супернатанты из обоих комплектов флаконов собирали и исследовали титры с помощью qPCR, выраженные в единицах трансдукции (TU)/мл (смотри пример 3).

Результаты титрования показаны в следующей табл. 4.

Таблица 4

Условие	Первый титр	Второй титр (после хранения при -80°C в течение 68 дней)
Без NaB	1,5 (±0,05)×10 ⁶ TU/мл	1,2 (±0,2)×10 ⁶ TU/мл
10 мМ NaB	1,4 (±0,3)×10 ⁶ TU/мл	7,0 (±0,14)×10 ⁵ TU/мл

Последующие препараты вектора были получены таким же образом, без бутирата натрия. Другие векторные плазмиды были использованы таким же образом для получения векторных препаратов с титрами от 10⁵ до 10⁷ TU/мл. Такой материал дополнительно может быть очищен и концентрирован, при желании, как описано ниже, а также см. US 5792643; T. Rodriguez et al. J. Gene Med. 9:233 2007; заявка США 61218063. В некоторых вариантах осуществления настоящего изобретения дозу вычисляли по граммам массы головного мозга. В таких вариантах осуществления доза репликационно-компетентного ретровирусного вектора по настоящему изобретению, применимая в способах для лечения, может находиться в интервале от 10³ до 10⁷ TU на 1 г массы головного мозга.

Пример 3. Количественный ПЦР анализ титрования.

Функциональную концентрацию вектора, или титр, определяют с использованием количественного способа на основе ПЦР (qPCR). В этом способе вектор титруют путем инфицирования трансдуцируемой линии клеток-хозяев (например, клетки карциномы предстательной железы человека PC-3, ATCC Cat # CRL-1435) со стандартным объемом вектора и измерения результирующего количества провируса, находящегося в клетках-хозяевах после трансдукции. Клетки и вектор инкубируют в стандартных усло-

виях культивирования (37°C, 5% CO₂) в течение 24 ч, чтобы дать возможность завершить инфицирование до добавления антиретровирусного AZT для остановки репликации вектора.

Далее, клетки собирают с культуральной чашки и геномную ДНК (гДНК) очищают с использованием набора для очистки геномной ДНК Invitrogen Purelink и элюируют из колонки для очистки с использованием стерильной воды без содержания РНазы/ДНАзы. Соотношение поглощения A260/A280 измеряют на спектрофотометре для определения концентрации и относительной чистоты образца. Концентрации гДНК нормализуют с использованием дополнительного количества воды без содержания РНКазы-/ДНКазы до наиболее низкой концентрации любого заданного набора препаратов гДНК, так, чтобы входящая ДНК для qPCR оставалась неизменной для всех анализируемых образцов.

Чистоту геномной ДНК дополнительно определяют с помощью электрофореза аликвоты каждого образца на окрашенном этидия бромидом 0,8% агарозном геле. Если образец проходит диапазон поглощения при A260/A280 1,8-2,0 и демонстрирует одну полосу гДНК, тогда этот образец готов для проведения qPCR анализа провирусного числа копий вектора. Используя праймеры, которые задают область LTR провируса (обратно транскрибированная векторная ДНК и векторная ДНК, которая интегрируется в гДНК хозяина), qPCR проводят для оценки суммарного числа событий трансдукции, которые происходят, когда известный объем вектора использовали для трансдукции известного числа клеток. Количество событий трансдукции на реакцию вычисляют по стандартной кривой, в которой используется плазмида, несущая мишень известного числа копий, которая серийно разбавлена от 1E7 до 1E1 копий и измерена в идентичных условиях qPCR, как образцы. Зная, сколько геномных эквивалентов было использовано для каждой реакции qPCR (из концентрации, определенной ранее) и сколько событий трансдукции, которые произошли на реакцию, авторы изобретения определили суммарное число событий трансдукции, которые произошли, исходя из суммарного числа клеток, которые присутствовали во время трансдукции. Это значение является титром вектора после разбавления в среде, содержащей клетки во время исходной трансдукции. Для вычисления правильного значения титра, разбавление корректируется путем умножения на коэффициент, равный сумме объема культуры и объема титра, деленные на объем титра. Эти эксперименты проводят в реплицирующихся культурах и анализируют с помощью qPCR, используя измерения в трех повторах для каждого условия для определения среднего титра и связанного с ним стандартного отклонения и коэффициента вариативности.

Пример 4. Уровни экспрессии, измеренные с помощью вестерн-блоттинга.

На фиг. 3 показано, что наблюдаются более высокие уровни кодонов, оптимизированных для человека с тремя мутациями для более высокой стабильности по сравнению с белком уCD дикого типа в вестерн-блот анализе клеток U-87, инфицированных вирусом, кодирующим либо дикого типа (ACE-yCD), либо полностью оптимизированные (AC3-yCD2) гены цитозиндезаминазы.

Пример 5. Генетическая стабильность вирусных векторов.

Известно, что после обратной транскрипции и первого события интеграции в обработанные клетки, ДНК провируса и любое последующее потомство ретровируса имеет обычную LTR структуру из MLV на любом конце. Было показано, что эта конфигурация является стабильной после многократных циклов инфицирования (см. фиг. 4).

Приблизительно 10⁶ интактных клеток U-87 первоначально было инфицировано вирусным вектором при MOI 0,01, и их выращивали до полного инфицирования для завершения одного цикла инфекции. Затем супернатант снова возвращали на инфицированные клетки, а затем цикл повторяли. Геномную стабильность последовательности уCD2 оценивали с помощью ПЦР амплификации интегрированного провируса из инфицированных клеток, используя MLV специфичные праймеры, фланкирующие сайт встраивания трансгена. Для каждой группы инфекций, амплификацию векторной плазмиды (pAC3-yCD2 и Kasahara et al. вектор pACE-CD) также проводили, чтобы проследить расположение по размеру полно-размерного ампликона на геле. Появление любых полос, меньших полноразмерного ампликона, будет показателем нестабильности вектора. Такие эксперименты демонстрировали, что вектор по настоящему изобретению (T5.0002 - содержащий модифицированный вектор, и CDopt+3pt (CD2) гетерологичный полинуклеотид сохраняли стабильность на протяжении большего количества пассажей, чем pACE-CD или T5.007, каждый из которых несет (полинуклеотид) дрожжей дикого типа.

Пример 6. Эксперименты по цитолизу клеток.

В экспериментах на клеточных культурах *in vitro* показано, что цитозиндезаминаза в клетках, экспрессирующих белок уCD2, по меньшей мере, является такой же активной, как и цитозиндезаминаза из клеток, экспрессирующих белок уCD дикого типа, путем проведения титрований 5-FC с использованием клеток крыс RG2 (фиг. 5A) или клеток U-87 (фиг. 5B), инфицированных либо вирусом, полученным из pAC3-yCD2/T5.0002 [AC3-yCD2(V)] из pAC3-yCD/T5.0007 [AC3-yCD(V)], либо другими векторами. Вкратце, для клеток U-87 через 5 дней после инфицирования при множественности заражения 0,1 (т.е. 100% инфицированных) либо вектором AC3-yCD (CD дикого типа), либо вектором AC3-yCD2 (термостабильным и кодон-оптимизированным) подвергали воздействию повышенных количеств 5-FC или 0,1 mM 5-FU в качестве положительного контроля в течение 8 дней. На 8 день обработки 5-FC жизнеспособность клеточных культур оценивали с помощью анализа MTS (Promega CellTiter 96 AQUEOUS One Solution Proliferation Assay). Данные показывают сравнимый цитоллиз между двумя ретровирусными век-

торами при повышенных дозах обработки с использованием 5-FC. Культуры RG2 обрабатывали подобным образом, и также показано, если вообще что-либо, слабый сдвиг кривой цитолиза в направлении более низкой концентрации 5-FC для вируса из анализа экспрессии T5.0002CD. Клетки U87 трансдуцировали при множественности заражения (MOI) 0,1, культивировали в течение 5 дней, чтобы дать возможность для распространения вируса и собирали клетки с 5 дня после трансдукции. Затем клетки собирали центрифугированием при $800 \times g$ в течение 5 мин. Супернатант аспирировали из клеточного осадка и промывали 5 мл фосфатно-буферного солевого раствора (PBS) и снова центрифугировали при $800 \times g$ в течение 5 мин. Полученный в результате клеточный осадок переносили в 1,5 мл PBS, ресуспендировали пропуская через наконечник пипетки, и помещали в морозильную камеру при -20°C . Клетки лизировали методом замораживания/оттаивания. Ранее ресуспендированные клетки размораживали при комнатной температуре, пропускали через наконечник пипетки, перемешивали со смесью ингибиторов протеаз и снова замораживали при -20°C . Перед проведением ферментативного исследования, образцы снова размораживали при комнатной температуре и пропускали через носик пипетки. Затем суспензию центрифугировали при 14000 об/мин в настольной центрифуге в течение 5 мин. Супернатант отбрасывали из осадка и помещали в новую пробирку эппендорфа и помещали на лед. Ферментативную активность уCD оценивали, путем использования анализа HPLC.

Анализ HPLC проводили на Shimadzu устройстве LC20AT, соединенном последовательно с фотоматричным детектором и автоинжектором. Твердая фаза представляла собой Hypersil BDS C18 HPLC колонку с 5 мкм размером сфер и $4,0 \times 250$ мм размерами колонки. Подвижная фаза представляла собой 50 mM фосфат аммония, pH 2,1, содержащий 0,01% трет-бутиламмония перхлорат и 5% метанол; эта система была уравновешена при 22°C . Все реактивы были ACS степени чистоты, а растворители были степени чистоты "для ВЭЖХ". Получали реакционную смесь, состоящую из 800 мкл с конечной концентрацией 0,125 мг/мл 5FC (1 mM) в PBS и помещали в 1,5 мл флаконы для автоматического дозатора. Затем инициировали реакцию путем добавления 200 мкл каждого клеточного лизата. Реакционную смесь/флаконы для автоматического дозатора помещали в автодозатор и вводили 5 мкл реакционной смеси. Моменты времени были взяты периодически, путем извлечения 5 мкл аликвоты из каждого реакционного флакона и анализа на ВЭЖХ колонке. Скорости превращения 5FC в 5FU вычисляли путем сравнения площадей пика с известными количествами из ранее полученной стандартной кривой 5FU.

Скорость превращения 5FC в 5FU была получена путем нанесения на график количества 5FU (в нмолях) в зависимости от соответствующего интервала времени. Была получена концентрация белка для образца клеток и специфическую активность образцов клеточного лизата вычисляли путем деления скорости превращения (нмоль/мин) на количество белка, используемого в исследовании, в мг. На фиг. 6 показана специфическая активность векторов после 5 дней трансдукции при MOI 0,1. Данные демонстрируют, что $\text{pACE-uCD (T5.0000)} < \text{pAC3-uCD1 (T5.0001)} < \text{pAC3-CD2 (T5.0002)}$ в значении специфической активности цитозиндезаминазы в клетках культуры ткани.

Пример 7. Опухоли, обработанные полностью модифицированным геном CD (uCD2), элиминируются более эффективно, чем опухоли, обработанные немодифицированным геном CD дрожжей, и не рецидивируют (PR-01-08-001).

Для определения, какая векторная конструкция приводит к получению наиболее эффективного вектора в модели подкожного ксенотрансплантата мыши/человека. Оценивали три различные конструкции: T5.0001 (частично модифицированный (ген) CD); T5.0007 (немодифицированный ген CD дрожжей); T5.0002 (полностью модифицированный ген uCD2). Рост опухоли, выживание и регрессию опухоли после введения пролекарства 5-FC оценивали в модели подкожной глиомы человека (U87) у мышей с иммунодефицитом. Оценивали две различные концентрации 5-FC для определения ответа на дозу, наблюдаемого между 5-FC и конструкциями опухоль-вектор.

Всего было исследовано 12 групп, состоящих из 9-11 самок мышей на группу. Все мыши подвергались имплантации на спине с правого бока в 0 день либо: 98% смеси линии неинфицированных опухолевых клеток U-87 и 2% смеси линии клеток U-87, которые были инфицированы одной из трех векторных конструкций TOCA 511, T5.0002, T5.0001, T5.0007; или неинфицированной контрольной линии клеток U-87 (100%). Мышам инокулировали 2×10^6 клеток/мышь. Мыши в группах 1-3 не были заражены клетками U87, в группах 4-6 получали смесь клеток U87, трансдуцированных T5.0002 U87, группы 7-9 получали смесь клеток U87, трансдуцированных T5.0001, и группы 10-12 получали смесь клеток U87, трансдуцированных T5.0007. Опухоли выращивали в течение 6 дней, пока размер опухоли не достигал приблизительно 100 мм^3 . Каждая группа мышей, которым вводили вектор, была рандомизирована для введения одной из двух доз 5-FC (200 или 500 мг/кг/день), вводимых в виде однократной внутривенной инъекции начиная с 6 дня без введения 5-FC. Введение 5-FC продолжали ежедневно в течение 28 последующих дней. Выживание мышей после 29 дня оценивали по размеру опухоли, а затем скарифицировали в различные дни, извлекая опухоли для анализа, при их наличии. На 29 день мыши из групп 4-12 снова были рандомизированы и мышей подразделяли на 2 подгруппы (либо продолжали введение 5-FC для наблюдения за регрессией опухоли, либо прекращали воздействие 5-FC для наблюдения за возобновлением опухолевого роста).

Результаты: во всех опухолях, обработанных T5.0002, T5.0000 и T5.0007, демонстрировалась регрессия опухоли до неопределяемой при дозе 200 и 500 мг/кг на 29 день. У всех животных, получивших воздействие T5.0002, опухоли заново не образовывались, когда прекращалось воздействие 5FC, до 39 дней. В отличие от этого, некоторые опухоли, обработанные T5.0001 и T5.0007, рецидивировали к 39 дню. Эксперимент заканчивали на 39 день.

Заключение: T5.0002 плюс 5-FC представляет собой более эффективную противоопухолевую терапию, чем T5.0001 или T5.0007 плюс 5-FC.

Был описан ряд вариантов осуществления настоящего изобретения. Тем не менее, будет понятно, что различные модификации могут быть сделаны, не отклоняясь от сущности и объема настоящего изобретения. В соответствии с этим другие варианты осуществления входят в объем следующей формулы изобретения.

Список последовательностей

```

<110>  Tocagen Inc.
      Gruber, Harry E.
      Jolly, Douglas
      Perez, Omar

<120>  ВЕКТОРЫ ДЛЯ ГЕННОЙ ТЕРАПИИ И ЦИТОЗИНДЕЗАМИНАЗЫ

<130>  00014-007WO1

<140>  Еще не известно
<141>  2009-09-26

<150>  US 61/100,666
<151>  2008-09-26

<150>  US 61/120,618
<151>  2008-12-08

<150>  US 61/186,823
<151>  2009-06-13

<160>  22

<170>  PatentIn version 3.5

<210>  1
<211>  477
<212>  ДНК
<213>  Saccharomyces cerevisiae

<220>
<221>  CDS
<222>  (1)..(477)

<400>  1
atg gtg aca ggg gga atg gca agc aag tgg gat cag aag ggt atg gac      48
Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp
1          5          10          15

att gcc tat gag gag gcg gcc tta ggt tac aaa gag ggt ggt gtt cct      96
Ile Ala Tyr Glu Glu Ala Ala Leu Gly Tyr Lys Glu Gly Gly Val Pro
20         25         30

att ggc gga tgt ctt atc aat aac aaa gac gga agt gtt ctc ggt cgt      144
Ile Gly Gly Cys Leu Ile Asn Asn Lys Asp Gly Ser Val Leu Gly Arg
35         40         45

ggt cac aac atg aga ttt caa aag gga tcc gcc aca cta cat ggt gag      192
Gly His Asn Met Arg Phe Gln Lys Gly Ser Ala Thr Leu His Gly Glu
50         55         60

atc tcc act ttg gaa aac tgt ggg aga tta gag ggc aaa gtg tac aaa      240
Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys
65         70         75         80

gat acc act ttg tat acg acg ctg tct cca tgc gac atg tgt aca ggt      288
Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly
85         90         95

gcc atc atc atg tat ggt att cca cgc tgt gtt gtc ggt gag aac gtt      336
Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Val Gly Glu Asn Val

```

027693

100	105	110	
aat ttc aaa agt aag ggc gag aaa tat tta caa act aga ggt cac gag			384
Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu			
115	120	125	
ggt ggt ggt ggt gac gat gag agg tgt aaa aag atc atg aaa caa ttt			432
Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Ile Met Lys Gln Phe			
130	135	140	
atc gat gaa aga cct cag gat tgg ttt gaa gat att ggt gag tag			477
Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu			
145	150	155	
<210> 2			
<211> 158			
<212> Белок			
<213> Saccharomyces cerevisiae			
<400> 2			
Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp			
1	5	10	15
Ile Ala Tyr Glu Glu Ala Ala Leu Gly Tyr Lys Glu Gly Gly Val Pro			
20	25	30	
Ile Gly Gly Cys Leu Ile Asn Asn Lys Asp Gly Ser Val Leu Gly Arg			
35	40	45	
Gly His Asn Met Arg Phe Gln Lys Gly Ser Ala Thr Leu His Gly Glu			
50	55	60	
Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys			
65	70	75	80
Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly			
85	90	95	
Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Val Gly Glu Asn Val			
100	105	110	
Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu			
115	120	125	
Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Ile Met Lys Gln Phe			
130	135	140	
Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu			
145	150	155	
<210> 3			
<211> 477			

<212> ДНК
 <213> Искусственная последовательность

 <220>
 <223> Сконструированная цитозиндезаминаза

 <220>
 <221> CDS
 <222> (1)..(477)

 <400> 3
 atg gtg aca ggg gga atg gca agc aag tgg gat cag aag ggt atg gac 48
 Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp
 1 5 10 15

 att gcc tat gag gag gcg tta tta ggt tac aaa gag ggt ggt gtt cct 96
 Ile Ala Tyr Glu Glu Ala Leu Leu Gly Tyr Lys Glu Gly Gly Val Pro
 20 25 30

 att ggc gga tgt ctt atc aat aac aaa gac gga agt gtt ctc ggt cgt 144
 Ile Gly Gly Cys Leu Ile Asn Asn Lys Asp Gly Ser Val Leu Gly Arg
 35 40 45

 ggt cac aac atg aga ttt caa aag gga tcc gcc aca cta cat ggt gag 192
 Gly His Asn Met Arg Phe Gln Lys Gly Ser Ala Thr Leu His Gly Glu
 50 55 60

 atc tcc act ttg gaa aac tgt ggg aga tta gag ggc aaa gtg tac aaa 240
 Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys
 65 70 75 80

 gat acc act ttg tat acg acg ctg tct cca tgc gac atg tgt aca ggt 288
 Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly
 85 90 95

 gcc atc atc atg tat ggt att cca cgc tgt gtc atc ggt gag aac gtt 336
 Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Ile Gly Glu Asn Val
 100 105 110

 aat ttc aaa agt aag ggc gag aaa tat tta caa act aga ggt cac gag 384
 Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu
 115 120 125

 gtt gtt gtt gtt gac gat gag agg tgt aaa aag tta atg aaa caa ttt 432
 Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Leu Met Lys Gln Phe
 130 135 140

 atc gat gaa aga cct cag gat tgg ttt gaa gat att ggt gag tag 477
 Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu
 145 150 155

 <210> 4
 <211> 158
 <212> Белок
 <213> Искусственная последовательность

 <220>
 <223> Синтетическая конструкция

 <400> 4
 Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp

```

1             5             10             15

Ile Ala Tyr Glu Glu Ala Leu Leu Gly Tyr Lys Glu Gly Gly Val Pro
      20             25             30

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

Gly His Asn Met Arg Phe Gln Lys Gly Ser Ala Thr Leu His Gly Glu
      50             55             60

Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys
      65             70             75             80

Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly
      85             90             95

Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Ile Gly Glu Asn Val
      100             105             110

Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu
      115             120             125

Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Leu Met Lys Gln Phe
      130             135             140

Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu
      145             150             155

<210>  5
<211>  480
<212>  ДНК
<213>  Искусственная последовательность

<220>
<223>  Цитозиндезаминаза с оптимизированными кодонами человека

<220>
<221>  CDS
<222>  (1)..(480)

<400>  5
atg gtg acc ggc ggc atg gcc tcc aag tgg gat caa aag ggc atg gat      48
Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp
1             5             10             15

atc gct tac gag gag gcc gca ctg ggc tac aag gag ggc ggc gtg cct      96
Ile Ala Tyr Glu Glu Ala Ala Leu Gly Tyr Lys Glu Gly Gly Val Pro
      20             25             30

atc ggc ggc tgt ctg atc aac aac aag gac ggc agt gtg ctg ggc agg      144
Ile Gly Gly Cys Leu Ile Asn Asn Lys Asp Gly Ser Val Leu Gly Arg
      35             40             45

```

ggc cac aac atg agg ttc cag aag ggc tcc gcc acc ctg cac ggc gag 192
 Gly His Asn Met Arg Phe Gln Lys Gly Ser Ala Thr Leu His Gly Glu
 50 55 60

atc tcc acc ctg gag aac tgt ggc agg ctg gag ggc aag gtg tac aag 240
 Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys
 65 70 75 80

gac acc acc ctg tac acc acc ctg tcc cct tgt gac atg tgt acc ggc 288
 Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly
 85 90 95

gct atc atc atg tac ggc atc cct agg tgt gtg gtc ggc gag aac gtg 336
 Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Val Gly Glu Asn Val
 100 105 110

aac ttc aag tcc aag ggc gag aag tac ctg caa acc agg ggc cac gag 384
 Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu
 115 120 125

gtg gtg gtt gtt gac gat gag agg tgt aag aag atc atg aag cag ttc 432
 Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Ile Met Lys Gln Phe
 130 135 140

atc gac gag agg cct cag gac tgg ttc gag gat atc ggc gag tga taa 480
 Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu
 145 150 155

<210> 6
 <211> 158
 <212> Белок
 <213> Искусственная последовательность

<220>
 <223> Синтетическая конструкция

<400> 6

Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp
 1 5 10 15

Ile Ala Tyr Glu Glu Ala Ala Leu Gly Tyr Lys Glu Gly Gly Val Pro
 20 25 30

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

Gly His Asn Met Arg Phe Gln Lys Gly Ser Ala Thr Leu His Gly Glu
 50 55 60

Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys
 65 70 75 80

Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly
 85 90 95

Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Val Gly Glu Asn Val
100 105 110

Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu
115 120 125

Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Ile Met Lys Gln Phe
130 135 140

Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu
145 150 155

<210> 7
<211> 756
<212> ДНК
<213> *Saccharomyces cerevisiae*

<220>
<221> CDS
<222> (1)..(756)

<400> 7
atg aac ccg tta ttc ttt ttg gct tct cca ttc ttg tac ctt aca tat 48
Met Asn Pro Leu Phe Phe Leu Ala Ser Pro Phe Leu Tyr Leu Thr Tyr
1 5 10 15
ctt ata tat tat cca aac aaa ggg tct ttc gtt agc aaa cct aga aat 96
Leu Ile Tyr Tyr Pro Asn Lys Gly Ser Phe Val Ser Lys Pro Arg Asn
20 25 30
ctg caa aaa atg tct tcg gaa cca ttt aag aac gtc tac ttg cta cct 144
Leu Gln Lys Met Ser Ser Glu Pro Phe Lys Asn Val Tyr Leu Leu Pro
35 40 45
caa aca aac caa ttg ctg ggt ttg tac acc atc atc aga aat aag aat 192
Gln Thr Asn Gln Leu Leu Gly Leu Tyr Thr Ile Ile Arg Asn Lys Asn
50 55 60
aca act aga cct gat ttc att ttc tac tcc gat aga atc atc aga ttg 240
Thr Thr Arg Pro Asp Phe Ile Phe Tyr Ser Asp Arg Ile Ile Arg Leu
65 70 75 80
ttg gtt gaa gaa ggt ttg aac cat cta cct gtg caa aag caa att gtg 288
Leu Val Glu Glu Gly Leu Asn His Leu Pro Val Gln Lys Gln Ile Val
85 90 95
gaa act gac acc aac gaa aac ttc gaa ggt gtc tca ttc atg ggt aaa 336
Glu Thr Asp Thr Asn Glu Asn Phe Glu Gly Val Ser Phe Met Gly Lys
100 105 110
atc tgt ggt gtt tcc att gtc aga gct ggt gaa tcg atg gag caa gga 384
Ile Cys Gly Val Ser Ile Val Arg Ala Gly Glu Ser Met Glu Gln Gly
115 120 125
tta aga gac tgt tgt agg tct gtg cgt atc ggt aaa att tta att caa 432
Leu Arg Asp Cys Cys Arg Ser Val Arg Ile Gly Lys Ile Leu Ile Gln
130 135 140
agg gac gag gag act gct tta cca aag tta ttc tac gaa aaa tta cca 480

Arg Asp Glu Glu Thr Ala Leu Pro Lys Leu Phe Tyr Glu Lys Leu Pro
 145 150 155 160
 gag gat ata tct gaa agg tat gtc ttc cta tta gac cca atg ctg gcc 528
 Glu Asp Ile Ser Glu Arg Tyr Val Phe Leu Leu Asp Pro Met Leu Ala
 165 170 175
 acc ggt ggt agt gct atc atg gct aca gaa gtc ttg att aag aga ggt 576
 Thr Gly Gly Ser Ala Ile Met Ala Thr Glu Val Leu Ile Lys Arg Gly
 180 185 190
 gtt aag cca gag aga att tac ttc tta aac cta atc tgt agt aag gaa 624
 Val Lys Pro Glu Arg Ile Tyr Phe Leu Asn Leu Ile Cys Ser Lys Glu
 195 200 205
 ggg att gaa aaa tac cat gcc gcc ttc cca gag gtc aga att gtt act 672
 Gly Ile Glu Lys Tyr His Ala Ala Phe Pro Glu Val Arg Ile Val Thr
 210 215 220
 ggt gcc ctc gac aga ggt cta gat gaa aac aag tat cta gtt cca ggg 720
 Gly Ala Leu Asp Arg Gly Leu Asp Glu Asn Lys Tyr Leu Val Pro Gly
 225 230 235 240
 ttg ggt gac ttt ggt gac aga tac tac tgt gtt taa 756
 Leu Gly Asp Phe Gly Asp Arg Tyr Tyr Cys Val
 245 250
 <210> 8
 <211> 251
 <212> Белок
 <213> *Saccharomyces cerevisiae*
 <400> 8
 Met Asn Pro Leu Phe Phe Leu Ala Ser Pro Phe Leu Tyr Leu Thr Tyr
 1 5 10 15
 Leu Ile Tyr Tyr Pro Asn Lys Gly Ser Phe Val Ser Lys Pro Arg Asn
 20 25 30
 Leu Gln Lys Met Ser Ser Glu Pro Phe Lys Asn Val Tyr Leu Leu Pro
 35 40 45
 Gln Thr Asn Gln Leu Leu Gly Leu Tyr Thr Ile Ile Arg Asn Lys Asn
 50 55 60
 Thr Thr Arg Pro Asp Phe Ile Phe Tyr Ser Asp Arg Ile Ile Arg Leu
 65 70 75 80
 Leu Val Glu Glu Gly Leu Asn His Leu Pro Val Gln Lys Gln Ile Val
 85 90 95
 Glu Thr Asp Thr Asn Glu Asn Phe Glu Gly Val Ser Phe Met Gly Lys
 100 105 110
 Ile Cys Gly Val Ser Ile Val Arg Ala Gly Glu Ser Met Glu Gln Gly

115	120	125	
Leu Arg Asp Cys Cys Arg Ser Val Arg Ile Gly Lys Ile Leu Ile Gln 130 135 140			
Arg Asp Glu Glu Thr Ala Leu Pro Lys Leu Phe Tyr Glu Lys Leu Pro 145 150 155 160			
Glu Asp Ile Ser Glu Arg Tyr Val Phe Leu Leu Asp Pro Met Leu Ala 165 170 175			
Thr Gly Gly Ser Ala Ile Met Ala Thr Glu Val Leu Ile Lys Arg Gly 180 185 190			
Val Lys Pro Glu Arg Ile Tyr Phe Leu Asn Leu Ile Cys Ser Lys Glu 195 200 205			
Gly Ile Glu Lys Tyr His Ala Ala Phe Pro Glu Val Arg Ile Val Thr 210 215 220			
Gly Ala Leu Asp Arg Gly Leu Asp Glu Asn Lys Tyr Leu Val Pro Gly 225 230 235 240			
Leu Gly Asp Phe Gly Asp Arg Tyr Tyr Cys Val 245 250			
 <210> 9 <211> 1443 <212> ДНК <213> homo sapiens			
 <220> <221> CDS <222> (1)..(1443)			
 <400> 9			
atg gct gtt gct cgt gct gct ctt ggt cct ctt gtt act ggt ctt tat Met Ala Val Ala Arg Ala Ala Leu Gly Pro Leu Val Thr Gly Leu Tyr 1 5 10 15			48
gat gtt caa gct ttt aaa ttt ggt gat ttt gtt ctt aaa tct ggt ctt Asp Val Gln Ala Phe Lys Phe Gly Asp Phe Val Leu Lys Ser Gly Leu 20 25 30			96
tct tct cct att tat att gat ctt cgt ggt att gtt tct cgt cct cgt Ser Ser Pro Ile Tyr Ile Asp Leu Arg Gly Ile Val Ser Arg Pro Arg 35 40 45			144
ctt ctt tct caa gtt gct gat att ctt ttt caa act gct caa aat gct Leu Leu Ser Gln Val Ala Asp Ile Leu Phe Gln Thr Ala Gln Asn Ala 50 55 60			192
ggt att tct ttt gat act gtt tgt ggt gtt cct tat act gct ctt cct Gly Ile Ser Phe Asp Thr Val Cys Gly Val Pro Tyr Thr Ala Leu Pro			240

027693

65	70	75	80	
ctt gct act gtt att tgt tct act aat caa att cct atg ctt att cgt Leu Ala Thr Val Ile Cys Ser Thr Asn Gln Ile Pro Met Leu Ile Arg 85 90 95				288
cgt aaa gaa act aaa gat tat ggt act aaa cgt ctt gtt gaa ggt act Arg Lys Glu Thr Lys Asp Tyr Gly Thr Lys Arg Leu Val Glu Gly Thr 100 105 110				336
att aat cct ggt gaa act tgt ctt att att gaa gat gtt gtt act tct Ile Asn Ser Ser Gly Glu Thr Cys Leu Ile Ile Glu Asp Val Val Thr Ser 115 120 125				384
ggt tct tct gtt ctt gaa act gtt gaa gtt ctt caa aaa gaa ggt ctt Gly Ser Ser Val Leu Glu Thr Val Glu Val Leu Gln Lys Glu Gly Leu 130 135 140				432
aaa gtt act gat gct att gtt ctt ctt gat cgt gaa caa ggt ggt aaa Lys Val Thr Asp Ala Ile Val Leu Leu Asp Arg Glu Gln Gly Gly Lys 145 150 155 160				480
gat aaa ctt caa gct cat ggt att cgt ctt cat tct gtt tgt act ctt Asp Lys Leu Gln Ala His Gly Ile Arg Leu His Ser Val Cys Thr Leu 165 170 175				528
tct aaa atg ctt gaa att ctt gaa caa caa aaa aaa gtt gat gct gaa Ser Lys Met Leu Glu Ile Leu Glu Gln Gln Lys Lys Val Asp Ala Glu 180 185 190				576
act gtt ggt cgt gtt aaa cgt ttt att caa gaa aat gtt ttt gtt gct Thr Val Gly Arg Val Lys Arg Phe Ile Gln Glu Asn Val Phe Val Ala 195 200 205				624
gct aat cat aat ggt tct cct ctt tct att aaa gaa gct cct aaa gaa Ala Asn His Asn Gly Ser Pro Leu Ser Ile Lys Glu Ala Pro Lys Glu 210 215 220				672
ctt tct ttt ggt gct cgt gct gaa ctt cct cgt att cat cct gtt gct Leu Ser Phe Gly Ala Arg Ala Glu Leu Pro Arg Ile His Pro Val Ala 225 230 235 240				720
tct aaa ctt ctt cgt ctt atg caa aaa aaa gaa act aat ctt tgt ctt Ser Lys Leu Leu Arg Leu Met Gln Lys Lys Glu Thr Asn Leu Cys Leu 245 250 255				768
tct gct gat gtt tct ctt gct cgt gaa ctt ctt caa ctt gct gat gct Ser Ala Asp Val Ser Leu Ala Arg Glu Leu Leu Gln Leu Ala Asp Ala 260 265 270				816
ctt ggt cct tct att tgt atg ctt aaa act cat gtt gat att ctt aat Leu Gly Pro Ser Ile Cys Met Leu Lys Thr His Val Asp Ile Leu Asn 275 280 285				864
gat ttt act ctt gat gtt atg aaa gaa ctt att act ctt gct aaa tgt Asp Phe Thr Leu Asp Val Met Lys Glu Leu Ile Thr Leu Ala Lys Cys 290 295 300				912
cat gaa ttt ctt att ttt gaa gat cgt aaa ttt gct gat att ggt aat His Glu Phe Leu Ile Phe Glu Asp Arg Lys Phe Ala Asp Ile Gly Asn 305 310 315 320				960
act gtt aaa aaa caa tat gaa ggt ggt att ttt aaa att gct tct tgg Thr Val Lys Lys Gln Tyr Glu Gly Gly Ile Phe Lys Ile Ala Ser Trp				1008

027693

325	330	335	
gct gat ctt gtt aat gct cat gtt gtt cct ggt tct ggt gtt gtt aaa Ala Asp Leu Val Asn Ala His Val Val Pro Gly Ser Gly Val Val Lys 340 345 350			1056
ggt ctt caa gaa gtt ggt ctt cct ctt cat cgt ggt tgt ctt ctt att Gly Leu Gln Glu Val Gly Leu Pro Leu His Arg Gly Cys Leu Leu Ile 355 360 365			1104
gct gaa atg tct tct act ggt tct ctt gct act ggt gat tat act cgt Ala Glu Met Ser Ser Thr Gly Ser Leu Ala Thr Gly Asp Tyr Thr Arg 370 375 380			1152
gct gct gtt cgt atg gct gaa gaa cat tct gaa ttt gtt gtt ggt ttt Ala Ala Val Arg Met Ala Glu Glu His Ser Glu Phe Val Val Gly Phe 385 390 395 400			1200
att tct ggt tct cgt gtt tct atg aaa cct gaa ttt ctt cat ctt act Ile Ser Gly Ser Arg Val Ser Met Lys Pro Glu Phe Leu His Leu Thr 405 410 415			1248
cct ggt gtt caa ctt gaa gct ggt ggt gat aat ctt ggt caa caa tat Pro Gly Val Gln Leu Glu Ala Gly Gly Asp Asn Leu Gly Gln Gln Tyr 420 425 430			1296
aat tct cct caa gaa gtt att ggt aaa cgt ggt tct gat att att att Asn Ser Pro Gln Glu Val Ile Gly Lys Arg Gly Ser Asp Ile Ile Ile 435 440 445			1344
ggt ggt cgt ggt att att tct gct gct gat cgt ctt gaa gct gct gaa Val Gly Arg Gly Ile Ile Ser Ala Ala Asp Arg Leu Glu Ala Ala Glu 450 455 460			1392
atg tat cgt aaa gct gct tgg gaa gct tat ctt tct cgt ctt ggt gtt Met Tyr Arg Lys Ala Ala Trp Glu Ala Tyr Leu Ser Arg Leu Gly Val 465 470 475 480			1440
taa			1443
<210> 10			
<211> 480			
<212> Белок			
<213> homo sapiens			
<400> 10			
Met Ala Val Ala Arg Ala Ala Leu Gly Pro Leu Val Thr Gly Leu Tyr 1 5 10 15			
Asp Val Gln Ala Phe Lys Phe Gly Asp Phe Val Leu Lys Ser Gly Leu 20 25 30			
Ser Ser Pro Ile Tyr Ile Asp Leu Arg Gly Ile Val Ser Arg Pro Arg 35 40 45			
Leu Leu Ser Gln Val Ala Asp Ile Leu Phe Gln Thr Ala Gln Asn Ala 50 55 60			

027693

Gly Ile Ser Phe Asp Thr Val Cys Gly Val Pro Tyr Thr Ala Leu Pro
65 70 75 80

Leu Ala Thr Val Ile Cys Ser Thr Asn Gln Ile Pro Met Leu Ile Arg
85 90 95

Arg Lys Glu Thr Lys Asp Tyr Gly Thr Lys Arg Leu Val Glu Gly Thr
100 105 110

Ile Asn Pro Gly Glu Thr Cys Leu Ile Ile Glu Asp Val Val Thr Ser
115 120 125

Gly Ser Ser Val Leu Glu Thr Val Glu Val Leu Gln Lys Glu Gly Leu
130 135 140

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

Asp Lys Leu Gln Ala His Gly Ile Arg Leu His Ser Val Cys Thr Leu
165 170 175

Ser Lys Met Leu Glu Ile Leu Glu Gln Gln Lys Lys Val Asp Ala Glu
180 185 190

Thr Val Gly Arg Val Lys Arg Phe Ile Gln Glu Asn Val Phe Val Ala
195 200 205

Ala Asn His Asn Gly Ser Pro Leu Ser Ile Lys Glu Ala Pro Lys Glu
210 215 220

Leu Ser Phe Gly Ala Arg Ala Glu Leu Pro Arg Ile His Pro Val Ala
225 230 235 240

Ser Lys Leu Leu Arg Leu Met Gln Lys Lys Glu Thr Asn Leu Cys Leu
245 250 255

Ser Ala Asp Val Ser Leu Ala Arg Glu Leu Leu Gln Leu Ala Asp Ala
260 265 270

Leu Gly Pro Ser Ile Cys Met Leu Lys Thr His Val Asp Ile Leu Asn
275 280 285

Asp Phe Thr Leu Asp Val Met Lys Glu Leu Ile Thr Leu Ala Lys Cys
290 295 300

His Glu Phe Leu Ile Phe Glu Asp Arg Lys Phe Ala Asp Ile Gly Asn
305 310 315 320

Thr Val Lys Lys Gln Tyr Glu Gly Gly Ile Phe Lys Ile Ala Ser Trp
325 330 335

Ala Asp Leu Val Asn Ala His Val Val Pro Gly Ser Gly Val Val Lys
340 345 350

Gly Leu Gln Glu Val Gly Leu Pro Leu His Arg Gly Cys Leu Leu Ile
355 360 365

Ala Glu Met Ser Ser Thr Gly Ser Leu Ala Thr Gly Asp Tyr Thr Arg
370 375 380

Ala Ala Val Arg Met Ala Glu Glu His Ser Glu Phe Val Val Gly Phe
385 390 395 400

Ile Ser Gly Ser Arg Val Ser Met Lys Pro Glu Phe Leu His Leu Thr
405 410 415

Pro Gly Val Gln Leu Glu Ala Gly Gly Asp Asn Leu Gly Gln Gln Tyr
420 425 430

Asn Ser Pro Gln Glu Val Ile Gly Lys Arg Gly Ser Asp Ile Ile Ile
435 440 445

Val Gly Arg Gly Ile Ile Ser Ala Ala Asp Arg Leu Glu Ala Ala Glu
450 455 460

Met Tyr Arg Lys Ala Ala Trp Glu Ala Tyr Leu Ser Arg Leu Gly Val
465 470 475 480

<210> 11
<211> 1227
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Слитая конструкция CDopt-UPRT

<220>
<221> CDS
<222> (1)..(1227)

<400> 11
atg gtg acc ggc ggc atg gcc tcc aag tgg gat caa aag ggc atg gat 48
Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp
1 5 10 15
atc gct tac gag gag gcc ctg ctg ggc tac aag gag ggc ggc gtg cct 96
Ile Ala Tyr Glu Glu Ala Leu Leu Gly Tyr Lys Glu Gly Gly Val Pro
20 25 30
atc ggc ggc tgt ctg atc aac aac aag gac ggc agt gtg ctg ggc agg 144
Ile Gly Gly Cys Leu Ile Asn Asn Lys Asp Gly Ser Val Leu Gly Arg

027693

35	40	45	
ggc cac aac atg agg ttc cag aag ggc tcc gcc acc ctg cac ggc gag Gly His Asn Met Arg Phe Gln Lys Gly Ser Ala Thr Leu His Gly Glu 50 55 60			192
atc tcc acc ctg gag aac tgt ggc agg ctg gag ggc aag gtg tac aag Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys 65 70 75 80			240
gac acc acc ctg tac acc acc ctg tcc cct tgt gac atg tgt acc ggc Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly 85 90 95			288
gct atc atc atg tac ggc atc cct agg tgt gtg atc ggc gag aac gtg Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Ile Gly Glu Asn Val 100 105 110			336
aac ttc aag tcc aag ggc gag aag tac ctg caa acc agg ggc cac gag Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu 115 120 125			384
gtg gtg gtt gtt gac gat gag agg tgt aag aag ctg atg aag cag ttc Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Leu Met Lys Gln Phe 130 135 140			432
atc gac gag agg cct cag gac tgg ttc gag gat atc ggc gag aac ccg Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu Asn Pro 145 150 155 160			480
tta ttc ttt ttg gct tct cca ttc ttg tac ctt aca tat ctt ata tat Leu Phe Phe Leu Ala Ser Pro Phe Leu Tyr Leu Thr Tyr Leu Ile Tyr 165 170 175			528
tat cca aac aaa ggg tct ttc gtt agc aaa cct aga aat ctg caa aaa Tyr Pro Asn Lys Gly Ser Phe Val Ser Lys Pro Arg Asn Leu Gln Lys 180 185 190			576
atg tct tcg gaa cca ttt aag aac gtc tac ttg cta cct caa aca aac Met Ser Ser Glu Pro Phe Lys Asn Val Tyr Leu Leu Pro Gln Thr Asn 195 200 205			624
caa ttg ctg ggt ttg tac acc atc atc aga aat aag aat aca act aga Gln Leu Leu Gly Leu Tyr Thr Ile Ile Arg Asn Lys Asn Thr Thr Arg 210 215 220			672
cct gat ttc att ttc tac tcc gat aga atc atc aga ttg ttg gtt gaa Pro Asp Phe Ile Phe Tyr Ser Asp Arg Ile Ile Arg Leu Leu Val Glu 225 230 235 240			720
gaa ggt ttg aac cat cta cct gtg caa aag caa att gtg gaa act gac Glu Gly Leu Asn His Leu Pro Val Gln Lys Gln Ile Val Glu Thr Asp 245 250 255			768
acc aac gaa aac ttc gaa ggt gtc tca ttc atg ggt aaa atc tgt ggt Thr Asn Glu Asn Phe Glu Gly Val Ser Phe Met Gly Lys Ile Cys Gly 260 265 270			816
gtt tcc att gtc aga gct ggt gaa tcg atg gag caa gga tta aga gac Val Ser Ile Val Arg Ala Gly Glu Ser Met Glu Gln Gly Leu Arg Asp 275 280 285			864
tgt tgt agg tct gtg cgt atc ggt aaa att tta att caa agg gac gag Cys Cys Arg Ser Val Arg Ile Gly Lys Ile Leu Ile Gln Arg Asp Glu			912

290	295	300	
gag act gct tta cca aag tta ttc tac gaa aaa tta cca gag gat ata			960
Glu Thr Ala Leu Pro Lys Leu Phe Tyr Glu Lys Leu Pro Glu Asp Ile			
305	310	315	320
tct gaa agg tat gtc ttc cta tta gac cca atg ctg gcc acc ggt ggt			1008
Ser Glu Arg Tyr Val Phe Leu Leu Asp Pro Met Leu Ala Thr Gly Gly			
	325	330	335
agt gct atc atg gct aca gaa gtc ttg att aag aga ggt gtt aag cca			1056
Ser Ala Ile Met Ala Thr Glu Val Leu Ile Lys Arg Gly Val Lys Pro			
	340	345	350
gag aga att tac ttc tta aac cta atc tgt agt aag gaa ggg att gaa			1104
Glu Arg Ile Tyr Phe Leu Asn Leu Ile Cys Ser Lys Glu Gly Ile Glu			
	355	360	365
aaa tac cat gcc gcc ttc cca gag gtc aga att gtt act ggt gcc ctc			1152
Lys Tyr His Ala Ala Phe Pro Glu Val Arg Ile Val Thr Gly Ala Leu			
	370	375	380
gac aga ggt cta gat gaa aac aag tat cta gtt cca ggg ttg ggt gac			1200
Asp Arg Gly Leu Asp Glu Asn Lys Tyr Leu Val Pro Gly Leu Gly Asp			
385	390	395	400
ttt ggt gac aga tac tac tgt gtt taa			1227
Phe Gly Asp Arg Tyr Tyr Cys Val			
	405		
<210>	12		
<211>	408		
<212>	Белок		
<213>	Искусственная последовательность		
<220>			
<223>	Синтетическая конструкция		
<400>	12		
Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp			
1	5	10	15
Ile Ala Tyr Glu Glu Ala Leu Leu Gly Tyr Lys Glu Gly Gly Val Pro			
	20	25	30
Ile Gly Gly Cys Leu Ile Asn Asn Lys Asp Gly Ser Val Leu Gly Arg			
	35	40	45
Gly His Asn Met Arg Phe Gln Lys Gly Ser Ala Thr Leu His Gly Glu			
	50	55	60
Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys			
65	70	75	80
Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly			
	85	90	95

027693

Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Ile Gly Glu Asn Val
100 105 110

Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu
115 120 125

Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Leu Met Lys Gln Phe
130 135 140

Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu Asn Pro
145 150 155 160

Leu Phe Phe Leu Ala Ser Pro Phe Leu Tyr Leu Thr Tyr Leu Ile Tyr
165 170 175

Tyr Pro Asn Lys Gly Ser Phe Val Ser Lys Pro Arg Asn Leu Gln Lys
180 185 190

Met Ser Ser Glu Pro Phe Lys Asn Val Tyr Leu Leu Pro Gln Thr Asn
195 200 205

Gln Leu Leu Gly Leu Tyr Thr Ile Ile Arg Asn Lys Asn Thr Thr Arg
210 215 220

Pro Asp Phe Ile Phe Tyr Ser Asp Arg Ile Ile Arg Leu Leu Val Glu
225 230 235 240

Glu Gly Leu Asn His Leu Pro Val Gln Lys Gln Ile Val Glu Thr Asp
245 250 255

Thr Asn Glu Asn Phe Glu Gly Val Ser Phe Met Gly Lys Ile Cys Gly
260 265 270

Val Ser Ile Val Arg Ala Gly Glu Ser Met Glu Gln Gly Leu Arg Asp
275 280 285

Cys Cys Arg Ser Val Arg Ile Gly Lys Ile Leu Ile Gln Arg Asp Glu
290 295 300

Glu Thr Ala Leu Pro Lys Leu Phe Tyr Glu Lys Leu Pro Glu Asp Ile
305 310 315 320

Ser Glu Arg Tyr Val Phe Leu Leu Asp Pro Met Leu Ala Thr Gly Gly
325 330 335

Ser Ala Ile Met Ala Thr Glu Val Leu Ile Lys Arg Gly Val Lys Pro
340 345 350

Glu Arg Ile Tyr Phe Leu Asn Leu Ile Cys Ser Lys Glu Gly Ile Glu
355 360 365

Lys Tyr His Ala Ala Phe Pro Glu Val Arg Ile Val Thr Gly Ala Leu
370 375 380

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

Phe Gly Asp Arg Tyr Tyr Cys Val
405

<210> 13
<211> 1287
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Слитая конструкция - CDopt - линкер - UPRT

<220>
<221> CDS
<222> (1)..(1287)

<400> 13
atg gtg acc ggc ggc atg gcc tcc aag tgg gat caa aag ggc atg gat 48
Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp
1 5 10 15
atc gct tac gag gag gcc ctg ctg ggc tac aag gag ggc ggc gtg cct 96
Ile Ala Tyr Glu Glu Ala Leu Leu Gly Tyr Lys Glu Gly Gly Val Pro
20 25 30
atc ggc ggc tgt ctg atc aac aac aag gac ggc agt gtg ctg ggc agg 144
Ile Gly Gly Cys Leu Ile Asn Asn Lys Asp Gly Ser Val Leu Gly Arg
35 40 45
ggc cac aac atg agg ttc cag aag ggc tcc gcc acc ctg cac ggc gag 192
Gly His Asn Met Arg Phe Gln Lys Gly Ser Ala Thr Leu His Gly Glu
50 55 60
atc tcc acc ctg gag aac tgt ggc agg ctg gag ggc aag gtg tac aag 240
Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys
65 70 75 80
gac acc acc ctg tac acc acc ctg tcc cct tgt gac atg tgt acc ggc 288
Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly
85 90 95
gct atc atc atg tac ggc atc cct agg tgt gtg atc ggc gag aac gtg 336
Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Ile Gly Glu Asn Val
100 105 110
aac ttc aag tcc aag ggc gag aag tac ctg caa acc agg ggc cac gag 384
Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu
115 120 125
gtg gtg gtt gtt gac gat gag agg tgt aag aag ctg atg aag cag ttc 432

Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Leu Met Lys Gln Phe	
130 135 140	
atc gac gag agg cct cag gac tgg ttc gag gat atc ggc gag tcc ggc	480
Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu Ser Gly	
145 150 155 160	
ggc ggc gcc tcc ggc ggc ggc gcc tcc ggc ggc gcc tcc ggc ggc	528
Gly Gly Ala Ser Gly Gly Gly Ala Ser Gly Gly Gly Ala Ser Gly Gly	
165 170 175	
ggc gcc aac ccg tta ttc ttt ttg gct tct cca ttc ttg tac ctt aca	576
Gly Ala Asn Pro Leu Phe Phe Leu Ala Ser Pro Phe Leu Tyr Leu Thr	
180 185 190	
tat ctt ata tat tat cca aac aaa ggg tct ttc gtt agc aaa cct aga	624
Tyr Leu Ile Tyr Tyr Pro Asn Lys Gly Ser Phe Val Ser Lys Pro Arg	
195 200 205	
aat ctg caa aaa atg tct tcg gaa cca ttt aag aac gtc tac ttg cta	672
Asn Leu Gln Lys Met Ser Ser Glu Pro Phe Lys Asn Val Tyr Leu Leu	
210 215 220	
cct caa aca aac caa ttg ctg ggt ttg tac acc atc atc aga aat aag	720
Pro Gln Thr Asn Gln Leu Leu Gly Leu Tyr Thr Ile Ile Arg Asn Lys	
225 230 235 240	
aat aca act aga cct gat ttc att ttc tac tcc gat aga atc atc aga	768
Asn Thr Thr Arg Pro Asp Phe Ile Phe Tyr Ser Asp Arg Ile Ile Arg	
245 250 255	
ttg ttg gtt gaa gaa ggt ttg aac cat cta cct gtg caa aag caa att	816
Leu Leu Val Glu Glu Gly Leu Asn His Leu Pro Val Gln Lys Gln Ile	
260 265 270	
gtg gaa act gac acc aac gaa aac ttc gaa ggt gtc tca ttc atg ggt	864
Val Glu Thr Asp Thr Asn Glu Asn Phe Glu Gly Val Ser Phe Met Gly	
275 280 285	
aaa atc tgt ggt gtt tcc att gtc aga gct ggt gaa tcg atg gag caa	912
Lys Ile Cys Gly Val Ser Ile Val Arg Ala Gly Glu Ser Met Glu Gln	
290 295 300	
gga tta aga gac tgt tgt agg tct gtg cgt atc ggt aaa att tta att	960
Gly Leu Arg Asp Cys Cys Arg Ser Val Arg Ile Gly Lys Ile Leu Ile	
305 310 315 320	
caa agg gac gag gag act gct tta cca aag tta ttc tac gaa aaa tta	1008
Gln Arg Asp Glu Glu Thr Ala Leu Pro Lys Leu Phe Tyr Glu Lys Leu	
325 330 335	
cca gag gat ata tct gaa agg tat gtc ttc cta tta gac cca atg ctg	1056
Pro Glu Asp Ile Ser Glu Arg Tyr Val Phe Leu Leu Asp Pro Met Leu	
340 345 350	
gcc acc ggt ggt agt gct atc atg gct aca gaa gtc ttg att aag aga	1104
Ala Thr Gly Gly Ser Ala Ile Met Ala Thr Glu Val Leu Ile Lys Arg	
355 360 365	
ggt gtt aag cca gag aga att tac ttc tta aac cta atc tgt agt aag	1152
Gly Val Lys Pro Glu Arg Ile Tyr Phe Leu Asn Leu Ile Cys Ser Lys	
370 375 380	
gaa ggg att gaa aaa tac cat gcc gcc ttc cca gag gtc aga att gtt	1200

027693

Glu Gly Ile Glu Lys Tyr His Ala Ala Phe Pro Glu Val Arg Ile Val
385 390 395 400

act ggt gcc ctc gac aga ggt cta gat gaa aac aag tat cta gtt cca 1248
Thr Gly Ala Leu Asp Arg Gly Leu Asp Glu Asn Lys Tyr Leu Val Pro
405 410 415

ggg ttg ggt gac ttt ggt gac aga tac tac tgt gtt taa 1287
Gly Leu Gly Asp Phe Gly Asp Arg Tyr Tyr Cys Val
420 425

<210> 14
<211> 428
<212> Белок
<213> Искусственная последовательность

<220>
<223> Синтетическая конструкция

<400> 14

Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp
1 5 10 15

Ile Ala Tyr Glu Glu Ala Leu Leu Gly Tyr Lys Glu Gly Gly Val Pro
20 25 30

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

Gly His Asn Met Arg Phe Gln Lys Gly Ser Ala Thr Leu His Gly Glu
50 55 60

Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys
65 70 75 80

Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly
85 90 95

Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Ile Gly Glu Asn Val
100 105 110

Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu
115 120 125

Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Leu Met Lys Gln Phe
130 135 140

Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu Ser Gly
145 150 155 160

Gly Gly Ala Ser Gly Gly Gly Ala Ser Gly Gly Gly Ala Ser Gly Gly
165 170 175

027693

Gly Ala Asn Pro Leu Phe Phe Leu Ala Ser Pro Phe Leu Tyr Leu Thr
180 185 190

Tyr Leu Ile Tyr Tyr Pro Asn Lys Gly Ser Phe Val Ser Lys Pro Arg
195 200 205

Asn Leu Gln Lys Met Ser Ser Glu Pro Phe Lys Asn Val Tyr Leu Leu
210 215 220

Pro Gln Thr Asn Gln Leu Leu Gly Leu Tyr Thr Ile Ile Arg Asn Lys
225 230 235 240

Asn Thr Thr Arg Pro Asp Phe Ile Phe Tyr Ser Asp Arg Ile Ile Arg
245 250 255

Leu Leu Val Glu Glu Gly Leu Asn His Leu Pro Val Gln Lys Gln Ile
260 265 270

Val Glu Thr Asp Thr Asn Glu Asn Phe Glu Gly Val Ser Phe Met Gly
275 280 285

Lys Ile Cys Gly Val Ser Ile Val Arg Ala Gly Glu Ser Met Glu Gln
290 295 300

Gly Leu Arg Asp Cys Cys Arg Ser Val Arg Ile Gly Lys Ile Leu Ile
305 310 315 320

Gln Arg Asp Glu Glu Thr Ala Leu Pro Lys Leu Phe Tyr Glu Lys Leu
325 330 335

Pro Glu Asp Ile Ser Glu Arg Tyr Val Phe Leu Leu Asp Pro Met Leu
340 345 350

Ala Thr Gly Gly Ser Ala Ile Met Ala Thr Glu Val Leu Ile Lys Arg
355 360 365

Gly Val Lys Pro Glu Arg Ile Tyr Phe Leu Asn Leu Ile Cys Ser Lys
370 375 380

Glu Gly Ile Glu Lys Tyr His Ala Ala Phe Pro Glu Val Arg Ile Val
385 390 395 400

Thr Gly Ala Leu Asp Arg Gly Leu Asp Glu Asn Lys Tyr Leu Val Pro
405 410 415

Gly Leu Gly Asp Phe Gly Asp Arg Tyr Tyr Cys Val
420 425

```

<210> 15
<211> 1200
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Слитая конструкция - CDopt3 - OPRT

<220>
<221> CDS
<222> (1)..(1200)

<400> 15
atg gtg acc ggc ggc atg gcc tcc aag tgg gat caa aag ggc atg gat      48
Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp
1          5          10          15

atc gct tac gag gag gcc ctg ctg ggc tac aag gag ggc ggc gtg cct      96
Ile Ala Tyr Glu Ala Leu Leu Gly Tyr Lys Glu Gly Gly Val Pro
20          25          30

atc ggc ggc tgt ctg atc aac aac aag gac ggc agt gtg ctg ggc agg      144
Ile Gly Gly Cys Leu Ile Asn Asn Lys Asp Gly Ser Val Leu Gly Arg
35          40          45

ggc cac aac atg agg ttc cag aag ggc tcc gcc acc ctg cac ggc gag      192
Gly His Asn Met Arg Phe Gln Lys Gly Ser Ala Thr Leu His Gly Glu
50          55          60

atc tcc acc ctg gag aac tgt ggc agg ctg gag ggc aag gtg tac aag      240
Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys
65          70          75          80

gac acc acc ctg tac acc acc ctg tcc cct tgt gac atg tgt acc ggc      288
Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly
85          90          95

gct atc atc atg tac ggc atc cct agg tgt gtg atc ggc gag aac gtg      336
Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Ile Gly Glu Asn Val
100         105         110

aac ttc aag tcc aag ggc gag aag tac ctg caa acc agg ggc cac gag      384
Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu
115         120         125

gtg gtg gtt gtt gac gat gag agg tgt aag aag ctg atg aag cag ttc      432
Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Leu Met Lys Gln Phe
130         135         140

atc gac gag agg cct cag gac tgg ttc gag gat atc ggc gag gcg gtc      480
Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu Ala Val
145         150         155         160

gct cgt gca gct ttg ggg cca ttg gtg acg ggt ctg tac gac gtg cag      528
Ala Arg Ala Ala Leu Gly Pro Leu Val Thr Gly Leu Tyr Asp Val Gln
165         170         175

gct ttc aag ttt ggg gac ttc gtg ctg aag agc ggc ctt tcc tcc ccc      576
Ala Phe Lys Phe Gly Asp Phe Val Leu Lys Ser Gly Leu Ser Ser Pro
180         185         190

```

```

atc tac atc gat ctg cgg ggc atc gtg tct cga ccg cgt ctt ctg agt      624
Ile Tyr Ile Asp Leu Arg Gly Ile Val Ser Arg Pro Arg Leu Leu Ser
      195                200                205

cag gtt gca gat att tta ttc caa act gcc caa aat gca ggc atc agt      672
Gln Val Ala Asp Ile Leu Phe Gln Thr Ala Gln Asn Ala Gly Ile Ser
      210                215                220

ttt gac acc gtg tgt gga gtg cct tat aca gct ttg cca ttg gct aca      720
Phe Asp Thr Val Cys Gly Val Pro Tyr Thr Ala Leu Pro Leu Ala Thr
      225                230                235                240

gtt atc tgt tca acc aat caa att cca atg ctt att aga agg aaa gaa      768
Val Ile Cys Ser Thr Asn Gln Ile Pro Met Leu Ile Arg Arg Lys Glu
      245                250                255

aca aag gat tat gga act aag cgt ctt gta gaa gga act att aat cca      816
Thr Lys Asp Tyr Gly Thr Lys Arg Leu Val Glu Gly Thr Ile Asn Pro
      260                265                270

gga gaa acc tgt tta atc att gaa gat gtt gtc acc agt gga tct agt      864
Gly Glu Thr Cys Leu Ile Ile Glu Asp Val Val Thr Ser Gly Ser Ser
      275                280                285

gtt ttg gaa act gtt gag gtt ctt cag aag gag ggc ttg aag gtc act      912
Val Leu Glu Thr Val Glu Val Leu Gln Lys Glu Gly Leu Lys Val Thr
      290                295                300

gat gcc ata gtg ctg ttg gac aga gag cag gga ggc aag gac aag ttg      960
Asp Ala Ile Val Leu Leu Asp Arg Glu Gln Gly Gly Lys Asp Lys Leu
      305                310                315                320

cag gcg cac ggg atc cgc ctc cac tca gtg tgt aca ttg tcc aaa atg      1008
Gln Ala His Gly Ile Arg Leu His Ser Val Cys Thr Leu Ser Lys Met
      325                330                335

ctg gag att ctc gag cag cag aaa aaa gtt gat gct gag aca gtt ggg      1056
Leu Glu Ile Leu Glu Gln Gln Lys Lys Val Asp Ala Glu Thr Val Gly
      340                345                350

aga gtg aag agg ttt att cag gag aat gtc ttt gtg gca gcg aat cat      1104
Arg Val Lys Arg Phe Ile Gln Glu Asn Val Phe Val Ala Ala Asn His
      355                360                365

aat ggt tct ccc ctt tct ata aag gaa gca ccc aaa gaa ctc agc ttc      1152
Asn Gly Ser Pro Leu Ser Ile Lys Glu Ala Pro Lys Glu Leu Ser Phe
      370                375                380

ggt gca cgt gca gag ctg ccc agg atc cac cca gtt gca tcg aag taa      1200
Gly Ala Arg Ala Glu Leu Pro Arg Ile His Pro Val Ala Ser Lys
      385                390                395

<210> 16
<211> 399
<212> Белок
<213> Искусственная последовательность

<220>
<223> Синтетическая конструкция

<400> 16
Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp

```

027693

1	5	10	15
Ile Ala Tyr Glu Glu Ala Leu Leu Gly Tyr Lys Glu Gly Gly Val Pro	20	25	30
Ile Gly Gly Cys Leu Ile Asn Asn Lys Asp Gly Ser Val Leu Gly Arg	35	40	45
Gly His Asn Met Arg Phe Gln Lys Gly Ser Ala Thr Leu His Gly Glu	50	55	60
Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys	65	70	75
Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly	85	90	95
Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Ile Gly Glu Asn Val	100	105	110
Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu	115	120	125
Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Leu Met Lys Gln Phe	130	135	140
Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu Ala Val	145	150	155
Ala Arg Ala Ala Leu Gly Pro Leu Val Thr Gly Leu Tyr Asp Val Gln	165	170	175
Ala Phe Lys Phe Gly Asp Phe Val Leu Lys Ser Gly Leu Ser Ser Pro	180	185	190
Ile Tyr Ile Asp Leu Arg Gly Ile Val Ser Arg Pro Arg Leu Leu Ser	195	200	205
Gln Val Ala Asp Ile Leu Phe Gln Thr Ala Gln Asn Ala Gly Ile Ser	210	215	220
Phe Asp Thr Val Cys Gly Val Pro Tyr Thr Ala Leu Pro Leu Ala Thr	225	230	235
Val Ile Cys Ser Thr Asn Gln Ile Pro Met Leu Ile Arg Arg Lys Glu	245	250	255

027693

Thr Lys Asp Tyr Gly Thr Lys Arg Leu Val Glu Gly Thr Ile Asn Pro
260 265 270

Gly Glu Thr Cys Leu Ile Ile Glu Asp Val Val Thr Ser Gly Ser Ser
275 280 285

Val Leu Glu Thr Val Glu Val Leu Gln Lys Glu Gly Leu Lys Val Thr
290 295 300

Asp Ala Ile Val Leu Leu Asp Arg Glu Gln Gly Gly Lys Asp Lys Leu
305 310 315 320

Gln Ala His Gly Ile Arg Leu His Ser Val Cys Thr Leu Ser Lys Met
325 330 335

Leu Glu Ile Leu Glu Gln Gln Lys Lys Val Asp Ala Glu Thr Val Gly
340 345 350

Arg Val Lys Arg Phe Ile Gln Glu Asn Val Phe Val Ala Ala Asn His
355 360 365

Asn Gly Ser Pro Leu Ser Ile Lys Glu Ala Pro Lys Glu Leu Ser Phe
370 375 380

Gly Ala Arg Ala Glu Leu Pro Arg Ile His Pro Val Ala Ser Lys
385 390 395

<210> 17
<211> 1260
<212> ДНК
<213> Искусственная последовательность

<220>
<223> Слитая конструкция - CDopt3 - линкер - OPRT

<220>
<221> CDS
<222> (1)..(1260)

<400> 17
atg gtg acc ggc ggc atg gcc tcc aag tgg gat caa aag ggc atg gat 48
Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp
1 5 10 15
atc gct tac gag gag gcc ctg ctg ggc tac aag gag ggc ggc gtg cct 96
Ile Ala Tyr Glu Glu Ala Leu Leu Gly Tyr Lys Glu Gly Gly Val Pro
20 25 30
atc ggc ggc tgt ctg atc aac aac aag gac ggc agt gtg ctg ggc agg 144
Ile Gly Gly Cys Leu Ile Asn Asn Lys Asp Gly Ser Val Leu Gly Arg
35 40 45
ggc cac aac atg agg ttc cag aag ggc tcc gcc acc ctg cac ggc gag 192
Gly His Asn Met Arg Phe Gln Lys Gly Ser Ala Thr Leu His Gly Glu

50	55	60	
atc tcc acc ctg gag aac tgt ggc agg ctg gag ggc aag gtg tac aag Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys 65 70 75 80			240
gac acc acc ctg tac acc acc ctg tcc cct tgt gac atg tgt acc ggc Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly 85 90 95			288
gct atc atc atg tac ggc atc cct agg tgt gtg atc ggc gag aac gtg Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Ile Gly Glu Asn Val 100 105 110			336
aac ttc aag tcc aag ggc gag aag tac ctg caa acc agg ggc cac gag Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu 115 120 125			384
gtg gtg gtt gtt gac gat gag agg tgt aag aag ctg atg aag cag ttc Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Leu Met Lys Gln Phe 130 135 140			432
atc gac gag agg cct cag gac tgg ttc gag gat atc ggc gag tcc ggc Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu Ser Gly 145 150 155 160			480
ggc ggc gcc tcc ggc ggc ggc gcc tcc ggc ggc ggc gcc tcc ggc ggc Gly Gly Ala Ser Gly Gly Gly Ala Ser Gly Gly Gly Ala Ser Gly Gly 165 170 175			528
ggc gcc gcg gtc gct cgt gca gct ttg ggg cca ttg gtg acg ggt ctg Gly Ala Ala Val Ala Arg Ala Ala Leu Gly Pro Leu Val Thr Gly Leu 180 185 190			576
tac gac gtg cag gct ttc aag ttt ggg gac ttc gtg ctg aag agc ggg Tyr Asp Val Gln Ala Phe Lys Phe Gly Asp Phe Val Leu Lys Ser Gly 195 200 205			624
ctt tcc tcc ccc atc tac atc gat ctg cgg ggc atc gtg tct cga ccg Leu Ser Ser Pro Ile Tyr Ile Asp Leu Arg Gly Ile Val Ser Arg Pro 210 215 220			672
cgt ctt ctg agt cag gtt gca gat att tta ttc caa act gcc caa aat Arg Leu Leu Ser Gln Val Ala Asp Ile Leu Phe Gln Thr Ala Gln Asn 225 230 235 240			720
gca ggc atc agt ttt gac acc gtg tgt gga gtg cct tat aca gct ttg Ala Gly Ile Ser Phe Asp Thr Val Cys Gly Val Pro Tyr Thr Ala Leu 245 250 255			768
cca ttg gct aca gtt atc tgt tca acc aat caa att cca atg ctt att Pro Leu Ala Thr Val Ile Cys Ser Thr Asn Gln Ile Pro Met Leu Ile 260 265 270			816
aga agg aaa gaa aca aag gat tat gga act aag cgt ctt gta gaa gga Arg Arg Lys Glu Thr Lys Asp Tyr Gly Thr Lys Arg Leu Val Glu Gly 275 280 285			864
act att aat cca gga gaa acc tgt tta atc att gaa gat gtt gtc acc Thr Ile Asn Pro Gly Glu Thr Cys Leu Ile Ile Glu Asp Val Val Thr 290 295 300			912
agt gga tct agt gtt ttg gaa act gtt gag gtt ctt cag aag gag ggc Ser Gly Ser Ser Val Leu Glu Thr Val Glu Val Leu Gln Lys Glu Gly			960

027693

305	310	315	320	
ttg aag gtc act gat gcc ata gtg ctg ttg gac aga gag cag gga ggc				1008
Leu Lys Val Thr Asp Ala Ile Val Leu Leu Asp Arg Glu Gln Gly Gly	325	330	335	
aag gac aag ttg cag gcg cac ggg atc cgc ctc cac tca gtg tgt aca				1056
Lys Asp Lys Leu Gln Ala His Gly Ile Arg Leu His Ser Val Cys Thr	340	345	350	
ttg tcc aaa atg ctg gag att ctc gag cag cag aaa aaa gtt gat gct				1104
Leu Ser Lys Met Leu Glu Ile Leu Glu Gln Gln Lys Lys Val Asp Ala	355	360	365	
gag aca gtt ggg aga gtg aag agg ttt att cag gag aat gtc ttt gtg				1152
Glu Thr Val Gly Arg Val Lys Arg Phe Ile Gln Glu Asn Val Phe Val	370	375	380	
gca gcg aat cat aat ggt tct ccc ctt tct ata aag gaa gca ccc aaa				1200
Ala Ala Asn His Asn Gly Ser Pro Leu Ser Ile Lys Glu Ala Pro Lys	385	390	395	400
gaa ctc agc ttc ggt gca cgt gca gag ctg ccc agg atc cac cca gtt				1248
Glu Leu Ser Phe Gly Ala Arg Ala Glu Leu Pro Arg Ile His Pro Val	405	410	415	
gca tcg aag taa				1260
Ala Ser Lys				

<210> 18
 <211> 419
 <212> Блок
 <213> Искусственная последовательность

<220>
 <223> Синтетическая конструкция

<400> 18

Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp
1 5 10 15

Ile Ala Tyr Glu Glu Ala Leu Leu Gly Tyr Lys Glu Gly Gly Val Pro
20 25 30

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

Gly His Asn Met Arg Phe Gln Lys Gly Ser Ala Thr Leu His Gly Glu
50 55 60

Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys
65 70 75 80

Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly
85 90 95

027693

Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Ile Gly Glu Asn Val
100 105 110

Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu
115 120 125

Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Leu Met Lys Gln Phe
130 135 140

Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu Ser Gly
145 150 155 160

Gly Gly Ala Ser Gly Gly Gly Ala Ser Gly Gly Gly Ala Ser Gly Gly
165 170 175

Gly Ala Ala Val Ala Arg Ala Ala Leu Gly Pro Leu Val Thr Gly Leu
180 185 190

Tyr Asp Val Gln Ala Phe Lys Phe Gly Asp Phe Val Leu Lys Ser Gly
195 200 205

Leu Ser Ser Pro Ile Tyr Ile Asp Leu Arg Gly Ile Val Ser Arg Pro
210 215 220

Arg Leu Leu Ser Gln Val Ala Asp Ile Leu Phe Gln Thr Ala Gln Asn
225 230 235 240

Ala Gly Ile Ser Phe Asp Thr Val Cys Gly Val Pro Tyr Thr Ala Leu
245 250 255

Pro Leu Ala Thr Val Ile Cys Ser Thr Asn Gln Ile Pro Met Leu Ile
260 265 270

Arg Arg Lys Glu Thr Lys Asp Tyr Gly Thr Lys Arg Leu Val Glu Gly
275 280 285

Thr Ile Asn Pro Gly Glu Thr Cys Leu Ile Ile Glu Asp Val Val Thr
290 295 300

Ser Gly Ser Ser Val Leu Glu Thr Val Glu Val Leu Gln Lys Glu Gly
305 310 315 320

Leu Lys Val Thr Asp Ala Ile Val Leu Leu Asp Arg Glu Gln Gly Gly
325 330 335

Lys Asp Lys Leu Gln Ala His Gly Ile Arg Leu His Ser Val Cys Thr
340 345 350

Leu Ser Lys Met Leu Glu Ile Leu Glu Gln Gln Lys Lys Val Asp Ala
 355 360 365

Glu Thr Val Gly Arg Val Lys Arg Phe Ile Gln Glu Asn Val Phe Val
 370 375 380

Ala Ala Asn His Asn Gly Ser Pro Leu Ser Ile Lys Glu Ala Pro Lys
 385 390 395 400

Glu Leu Ser Phe Gly Ala Arg Ala Glu Leu Pro Arg Ile His Pro Val
 405 410 415

Ala Ser Lys

<210> 19
 <211> 11892
 <212> ДНК
 <213> Искусственная последовательность

<220>
 <223> вектор RCR - pAC3-yCD2

<400> 19
 tagttattaa tagtaatcaa ttacgggggtc attagttcat agcccatata tggagttccg 60
 cgttacataa cttacggtaa atggcccgcc tggtgaccg cccaacgacc ccgcccatt 120
 gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 180
 atgggtggag tatttacggt aaactgccc cttggcagta catcaagtgt atcatatgcc 240
 aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgccagta 300
 catgacotta tgggactttc ctacttgga gtacatctac gtattagtca tcgctattac 360
 catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 420
 atttccaagt ctccaccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 480
 ggactttcca aaatgtcgta acaactccgc ccattgacg caaatgggcg gtaggcgtgt 540
 acgggtgggag gtctatataa gcagagctgg ttagtgaac cggcgccagt cctccgattg 600
 actgagtcgc ccgggtaccc gtgtatcaa taaacctct tgcagttgca tccgacttgt 660
 ggtctcgctg ttccttggga gggctctctc tgagtgattg actacccgtc agcgggggtc 720
 tttcatttgg gggctcgtec gggatcggga gaccctgcc cagggaccac cgaccacca 780
 ccgggaggta agctggccag caacttatct gtgtctgtcc gattgtctag tgtctatgac 840
 tgattttatg cgctcgctc ggtactagt agctaactag ctctgtatct ggcggaccgc 900
 tggtggaact gacgagttcg gaacacccgg ccgaaccct gggagacgtc ccagggactt 960
 cgggggccgt tttgtggcc cgacctgagt ccaaaaatcc cgatcgtttt ggactctttg 1020

gtgcaccccc cttagaggag ggatatgtgg ttctggtagg agacgagaac ctaaaacagt	1080
tcccgcctcc gtctgaattt ttgttttcgg tttgggaccg aagccgcgcc gcgcgtcttg	1140
tctgctgcag catcgttctg tgttgtctct gtctgactgt gtttctgtat ttgtctgaga	1200
atatgggccca gactgttacc actcccttaa gtttgacctt aggtcactgg aaagatgtcg	1260
agcggatcgc tcacaaccag tcggtagatg tcaagaagag acgttggggtt accttctgct	1320
ctgcagaatg gccaaccttt aacgtcggat ggccgcgaga cggcaccttt aaccgagacc	1380
tcatacccca ggttaagatc aaggtctttt cacctggccc gcatggacac ccagaccagg	1440
tcccctacat cgtgacctgg gaagccttgg cttttgacct ccctccctgg gtcaagccct	1500
ttgtacaccc taagcctccg cctcctcttc ctccatccgc cccgtctctc ccccttgaac	1560
ctcctcgttc gaccccgctt cgatcctccc tttatccagc cctcactcct tctctaggcg	1620
ccaaacctaa acctcaagtt ctttctgaca gtggggggcc gctcatcgac ctacttacag	1680
aagacccccc gccttatagg gaccaagac caccctcttc cgacagggac ggaaatgggtg	1740
gagaagcgac cctgcggga gaggcaccgg acccctcccc aatggcatct cgcctacgtg	1800
ggagacggga gccccctgtg gccgactcca ctacctcgca ggcattcccc ctccgcgcag	1860
gaggaaacgg acagctcaa tactggccgt tctcctcttc tgacctttac aactggaaaa	1920
ataataaccc ttctttttct gaagatccag gtaaaactgac agctctgac gagtctgttc	1980
tcatacccca tcagcccacc tgggacgact gtcagcagct gttggggact ctgctgaccg	2040
gagaagaaaa acaacgggtg ctcttagagg ctagaaaggc ggtgcggggc gatgatgggc	2100
gccccactca actgccaat gaagtcgatg ccgcttttcc cctcgagcgc ccagactggg	2160
attacaccac ccaggcaggt aggaaccacc tagtccacta tcgccagttg ctcttagcgg	2220
gtctccaaaa cgcgggcaga agccccacca atttggccaa ggtaaaagga ataacacaag	2280
ggcccaatga gtctccctcg gccttcctag agagacttaa ggaagcctat cgcaggtaaa	2340
ctccttatga cctgaggac ccagggcaag aaactaatgt gtctatgtct ttcatttggc	2400
agtctgcccc agacattggg agaaagttag agaggttaga agatttaaaa aacaagacgc	2460
ttggagattht ggtagagag gcagaaaaga tctttaataa acgagaaacc ccggaagaaa	2520
gagaggaacg tatcaggaga gaaacagagg aaaaagaaga acgccgtagg acagaggatg	2580
agcagaaaga gaaagaaaga gatcgtagga gacatagaga gatgagcaag ctattggcca	2640
ctgtcgtagg tggacagaaa caggatagac agggaggaga acgaaggagg tcccaactcg	2700
atcgcgacca gtgtgcctac tgcaaagaaa aggggcactg ggctaaagat tgtccaaga	2760
aaccacgagg acctcgggga ccaagacccc agacctccct cctgacccta gatgactagg	2820
gaggtcaggg tcaggagccc cccctgaac ccaggataac cctcaaagtc ggggggcaac	2880
ccgtcacctt cctggtagat actggggccc aacctccgt gctgacccaa aatcctggac	2940

ccctaagtga taagtctgcc tgggtccaag gggctactgg aggaaagcgg tatcgctgga	3000
ccacggatcg caaagtacat ctagctaccg gtaaggtcac ccactctttc ctccatgtac	3060
cagactgtcc ctatcctctg ttaggaagag atttgctgac taaactaaaa gcccaaatcc	3120
actttgaggg atcaggagcc caggttatgg gaccaatggg gcagcccctg caagtgttga	3180
ccctaaatat agaagatgag catcggctac atgagacctc aaaagagcca gatgtttctc	3240
taggggtccac atggctgtct gattttctc aggcctgggc ggaaccggg ggcatgggac	3300
tggcagttcg ccaagctcct ctgatcatac ctctgaaagc aacctctacc cccgtgtcca	3360
taaaacaata ccccatgtca caagaagcca gactggggat caagccccac atacagagac	3420
tgttggaacca gggaatactg gtacctgcc agtccccctg gaacacgccc ctgctaccog	3480
ttaagaaacc agggactaat gattataggc ctgtccagga tctgagagaa gtcaacaagc	3540
gggtggaaga catccacccc accgtgcca acccttaca cctcttgagc gggctcccac	3600
cgtcccacca gtgggtacct gtgcttgatt taaaggatgc ctttttctgc ctgagactcc	3660
accccaccag tcagcctctc ttgcctttg agtggagaga tccagagatg ggaatctcag	3720
gacaattgac ctggaccaga ctcccacagg gtttcaaaaa cagtcccacc ctgtttgatg	3780
aggcactgca cagagaccta gcagacttcc ggatccagca ccagacttg atcctgctac	3840
agtacgtgga tgacttactg ctggccgcca cttctgagct agactgcca caaggtactc	3900
gggcctgtt acaaaccta gggaacctcg ggtatcgggc ctcgccaag aaagcccaa	3960
tttgccagaa acaggtcaag tatctgggt atcttctaaa agagggtcag agatggctga	4020
ctgaggccag aaaagagact gtgatgggc agcctactcc gaagaccct cgacaactaa	4080
gggagttcct agggacggca ggcttctgtc gcctctggat ccctgggtt gcagaaatgg	4140
cagccccctt gtacctctc accaaaacgg ggaactctgt taattgggc ccagaccaac	4200
aaaaggcta tcaagaaatc aagcaagctc ttctaactgc ccagccctg gggttgccag	4260
atttgactaa gccctttgaa ctctttgtcg acgagaagca gggctacgcc aaaggtgtcc	4320
taacgaaaa actgggacct tggcgtcggc cgggtgccta cctgtccaaa aagctagacc	4380
cagtagcagc tgggtggccc ccttgctac ggatggtagc agccattgcc gtactgacaa	4440
aggatgcagg caagctaacc atgggacagc cactagtcat tctggcccc catgcagtag	4500
aggcactagt caaacaaccc ccgaccgct ggctttccaa cggccggatg actcactatc	4560
aggccttgct ttggacacg gaccgggtcc agttcggacc ggtggtagcc ctgaaccgg	4620
ctacgtgct ccactgcct gaggaagggc tgcaacacaa ctgccttgat atcctggccg	4680
aagcccacgg aaccgaccc gacctaacg accagccgct ccagacgcc gaccacacct	4740
ggtacacgga tggaagcagt ctcttacaag agggacagc taaggcggga gctgcggtga	4800
ccaccgagac cgaggtaatc tgggctaaag ccctgccagc cgggacatcc gctcagcggg	4860

ctgaactgat agcactcacc caggccctaa agatggcaga aggtaagaag ctaaatgttt	4920
atactgatag cgttatgct tttgctactg cccatatcca tggagaaata tacagaaggc	4980
gtgggttgct cacatcagaa ggcaaagaga tcaaaaataa agacgagatc ttggccctac	5040
taaaagccct ctttctgccc aaaagactta gcataatcca ttgtccagga catcaaaagg	5100
gacacagcgc cgaggctaga ggcaaccgga tggtgacca agcggcccgga aaggcagcca	5160
tcacagagac tccagacacc tctaccctcc tcatagaaaa ttcacaccc tacacctcag	5220
aacattttca ttacacagt actgatataa aggacctaac caagttgggg gccatttatg	5280
ataaaacaaa gaagtattgg gtctaccaag gaaaacctgt gatgcctgac cagtttactt	5340
ttgaattatt agactttctt catcagctga ctcacctcag cttctcaaaa atgaaggctc	5400
tcctagagag aagccacagt ccctactaca tgctgaaccg ggatcgaaca ctcaaaaata	5460
tactgagac ctgcaaagct tgtgcacaag tcaacgccag caagtctgcc gttaaacagg	5520
gaactagggc cgcggggcat cggcccgga ctcattggga gatcgatttc accgagataa	5580
agcccggtt gtatggctat aaatatcttc tagtttttat agataccttt tctggctgga	5640
tagaagcctt cccaaccaag aaagaaaccg ccaaggtcgt aaccaagaag ctactagagg	5700
agatcttccc caggttcggc atgcctcagg tattgggaac tgacaatggg cctgccttcg	5760
tctccaaggt gagtcagaca gtggccgac tggtggggat tgattggaaa ttacattgtg	5820
catacagacc ccaaagctca ggccaggtag aaagaatgaa tagaaccatc aaggagactt	5880
taactaaatt aacgcttgca actggtctta gagactgggt gctcctactc cccttagccc	5940
tgtaccgagc cgcgaacacg cggggccccc atggcctcac cccatatgag atcttatatg	6000
gggcaccccc gcccttgta aacttcctg acctgacat gacaagagtt actaacagcc	6060
cctctctcca agctcactta caggtctctt acttagtcca gcacgaagtc tggagacctc	6120
tggcggcagc ctaccaagaa caactggacc gaccggtggt acctcaccct taccgagtcg	6180
gcgacacagt gtgggtccgc cgacaccaga ctaagaacct agaacctcgc tggaaaggac	6240
cttacacagt cctgtgacc acccccaccg ccctcaaagt agacggcatc gcagcttgga	6300
tacacgccgc ccacgtgaag gctgccgacc cgggggtgg accatcctct agactgacat	6360
ggcgcgttca acgctctcaa aacccctca agataagatt aaccctgga agcccttaat	6420
agtcatggga gtctgttag gagtaggat ggcagagagc ccccatcagg tctttaatgt	6480
aacctggaga gtcaccaacc tgatgactgg gcgtaccgcc aatgccacct ccctcctggg	6540
aactgtacaa gatgccttcc caaaattata ttttgatcta tgtgatctgg tcggagagga	6600
gtgggacctc tcagaccagg aaccgtatgt cgggtatggc tgcaagtacc ccgcaggag	6660
acagcggacc cggacttttg acttttacgt gtgcctggg cataccgtaa agtcgggggtg	6720
tgggggacca ggagagggt actgtggtaa atgggggtgt gaaaccaccg gacaggctta	6780

ctggaagccc acatcatcgt gggacctaat ctcccttaag cgcggttaaca ccccttggga	6840
cacgggatgc tctaaagttg cctgtggccc ctgctacgac ctctccaaag tatccaattc	6900
cttccaaggg gctactcgag ggggcagatg caaccctcta gtcctagaat tcaactgatgc	6960
aggaaaaaag gctaactggg acgggcccaa atcgtgggga ctgagactgt accggacagg	7020
aacagatcct attaccatgt tctccctgac ccggcaggtc cttaatgtgg gaccccagat	7080
ccccataggg cccaaccag tattaccga ccaaagactc ctttctcac caatagagat	7140
tgtaccggct ccacagccac ctagccccct caataccagt tccccctt ccaactaccag	7200
tacacctca acctccccta caagtccaag tgtcccacag ccaccccag gaactggaga	7260
tagactacta gctctagtca aaggagccta tcaggcgctt aacctacca atcccacaa	7320
gaccaagaa tgttggtgtg gcttagtgtc gggacctct tattacgaag gagtagcgg	7380
cgtgggcact tataccaatc attccaccgc tccggccaac tgtacggcca cttcccaaca	7440
taagcttacc ctatctgaag tgacaggaca gggcctatgc atgggggcag tacctaaaac	7500
tcaccaggcc ttatgtaaca ccaccaaag cgcggctca ggatcctact acctgcagc	7560
acccgccgga acaatgtggg cttgcagcac tggattgact ccctgcttgt ccaccacgg	7620
gtcaatcta accacagatt attgtgtatt agttgaactc tggcccagag taatttacca	7680
ctccccgat tatatgtatg gtcagcttga acagcgtaac aaatataaaa gagagccagt	7740
atcattgacc ctggcccttc tactaggagg attaaccatg ggagggattg cagctggaat	7800
agggacggg accactgcct taattaaaac ccagcagttt gagcagcttc atgccgctat	7860
ccagacagac ctcaacgaag tcgaaaagtc aattaccaac ctagaaaagt cactgacctc	7920
gttgtctgaa gtagtcctac agaaccgcag aggcctagat ttgtattcc taaaggagg	7980
aggctctgac gcagccctaa aagaagaatg ttgtttttat gcagaccaca cggggctagt	8040
gagagacagc atggccaaat taagagaaag gcttaatcag agacaaaaac tatttgagac	8100
aggccaagga tggttcgaag ggctgtttta tagatcccc tggtttacca ccttaatctc	8160
caccatcatg ggacctctaa tagtactctt actgatctta ctctttggac cttgcattct	8220
caatcgattg gtccaatttg ttaaagacag gatctcagtg gtccaggctc tggttttgac	8280
tcagcaatat caccagctaa aaccataga gtacgagcca tgaacgcgtt actggccgaa	8340
gccgcttgga ataaggccgg tgtgcgtttg tctatatgtt attttccacc atattgccgt	8400
cttttgga tgtgagggcc cggaacctg gccctgtctt cttgacgagc attcctaggg	8460
gtctttcccc tctcgccaaa ggaatgcaag gtctgttgaa tgtcgtgaag gaagcagttc	8520
ctctggaagc ttctgaaga caaacacgt ctgtacgac ctttgcagg cagcggaacc	8580
ccccacctgg cgacaggtgc ctctgcggcc aaaagccacg tgtataagat acacctgcaa	8640
aggcggcaca accccagtc cacgttgtga gttggatagt tgtggaaaga gtcaaatggc	8700

tctcctcaag cgtattcaac aaggggctga aggatgccca gaaggtaccc cattgtatgg	8760
gatctgatct ggggcctcgg tgcacatgct ttacatgtgt ttagtcgagg ttaaaaaaac	8820
gtctaggccc cccgaaccac ggggacgtgg ttttcctttg aaaaacacga ttataaatgg	8880
tgaccggcgg catggcctcc aagtgggatac aaaagggcat ggatatcgct tacgaggagg	8940
ccctgctggg ctacaaggag ggcggcgtgc ctatcggcgg ctgtctgatac aacaacaagg	9000
acggcagtggt gctgggcagg ggccacaaca tgaggttcca gaagggctcc gccaccctgc	9060
acggcgagat ctccaccctg gagaactgtg gcaggctgga gggcaagggtg tacaaggaca	9120
ccaccctgta caccaccctg tccccttggtg acatgtgtac cggcgctatc atcatgtacg	9180
gcataccctag gtgtgtgatac ggcgagaacg tgaacttcaa gtccaagggc gagaagtacc	9240
tgcaaaccag gggccacgag gtggtggttg ttgacgatga gaggtgtaag aagctgatga	9300
agcagttcat cgacgagagg cctcaggact ggttcgagga tatcggcgag taagcggccg	9360
cagataaaat aaaagatttt atttagtctc cagaaaaagg ggggaatgaa agaccccacc	9420
tgtaggtttg gcaagctagc ttaagtaacg ccattttgca aggcattgaa aaatacataa	9480
ctgagaatag agaagttcag atcaaggatc ggaacagatg gaacagctga atatgggcca	9540
aacaggatat ctgtggtgaa cagttcctgc cccggctcag ggccaagaac agatggaaca	9600
gctgaatatg ggccaacag gatattctgtg gtaagcagtt cctgccccgg ctgagggcca	9660
agaacagatg gtccccagat gcggtccagc cctcagcagt ttctagagaa ccatacagatg	9720
tttccagggt gcccacagga cctgaaatga ccctgtgcct tatttgaact aaccaatcag	9780
ttcgcttctc gcttctgttc gcgcgcttct gctccccgag ctcaataaaa gagcccacaa	9840
cccctcactc ggggcgccag tcctccgatt gactgagtcg cccgggtacc cgtgtatcca	9900
ataaacctc ttgcagttgc atccgacttg tggctctcgt gttccttggg agggctctct	9960
ctgagtgatt gactaccctg cagcgggggt ctttcattac atgtgagcaa aaggccagca	10020
aaaggccagg aaccgtaaaa aggccgcgtt gctggcggtt ttccataggc tccgcccccc	10080
tgacgagcat cacaaaaatc gacgctcaag tcagagggtg cgaaaccoga caggactata	10140
aagataccag gcgtttcccc ctggaagctc cctcgtgcgc tctcctgttc cgaccctgcc	10200
gcttaccgga tacctgtccg ctttctctcc ttccgggaagc gtggcgcttt ctcaatgctc	10260
acgctgtagg tatctcagtt cgggtgtaggt cgttcgctcc aagctgggct gtgtgcacga	10320
acccccgtt cagcccgacc gctgcgcctt atccggtaac tatcgtcttg agtccaaccc	10380
ggtaagacac gacttatcgc cactggcagc agccactggt aacaggatta gcagagcgag	10440
gtatgtaggc ggtgctacag agttcttgaa gtggtggcct aactacggct aactagaag	10500
gacagtatctt ggtatctgcg ctctgctgaa gccagttacc ttcggaaaaa gaggttgtag	10560
ctcttgatcc ggcaaacaaa ccaccgctgg tagcgggtgt tttttgttt gcaagcagca	10620

```

gattacgcgc agaaaaaag gatctcaaga agatcctttg atcttttcta cggggtctga 10680
cgctcagtg aacgaaaact cacgttaagg gatttttggtc atgagattat caaaaaggat 10740
cttcacctag atccttttaa attaaaaatg aagtttttaa tcaatctaaa gtatatatga 10800
gtaaacttgg tctgacagtt accaatgctt aatcagtgag gcacctatct cagcgatctg 10860
tctatttcgt tcatccatag ttgcctgact ccccgctgtg tagataacta cgatacggga 10920
gggcttacca tctggcccca gtgtgcaat gataccgga gaccacgct caccggctcc 10980
agatttatca gcaataaacc agccagccgg aagggccgag cgcagaagtg gtcctgcaac 11040
tttatccgcc tccatccagt ctattaattg ttgccgggaa gctagagtaa gtagttcgcc 11100
agttaatagt ttgcgcaacg ttgttgccat tgctgcaggc atcgtggtgt cagctcgtc 11160
gttttggtatg gcttcattca gctccggttc ccaacgatca aggcgagtta catgatcccc 11220
catgttgtgc aaaaaagcgg ttagctcctt cggctcctcg atcgttgtca gaagtaagtt 11280
ggccgcagtg ttatcactca tggttatggc agcactgcat aattctctta ctgtcatgcc 11340
atccgtaaga tgcttttctg tgactggtga gtactcaacc aagtcattct gagaatagtg 11400
tatggcgga cagagttgct cttgcccgcc gtcaacacgg gataataccg cgccacatag 11460
cagaacttta aaagtgtca tcattggaaa acgttcttcg gggcgaaaac tctcaaggat 11520
cttaccgctg ttgagatcca gttcgatgta acccactcgt gcacccaact gatcttcagc 11580
atcttttact ttcaccagcg tttctgggtg agcaaaaaca ggaaggcaaa atgccgcaaa 11640
aaaggaata agggcgacac ggaatgttg aatactcata ctcttccttt ttcaatatta 11700
ttgaagcatt tatcagggtt attgtctcat gagcgggatac atatttgaat gtatttagaa 11760
aaataaaca ataggggttc cgcgcacatt tccccgaaa gtgccacctg acgtctaaga 11820
aaccattatt atcatgacat taacctataa aaataggcgt atcacgaggc ctttcgtct 11880
tcaagaattc at 11892

```

```

<210> 20
<211> 11892
<212> ДНК
<213> Искусственная последовательность

```

```

<220>
<223> вектор RCR - pAC3-yCD

```

```

<400> 20
tagttattaa tagtaatcaa ttacgggggtc attagttcat agcccatata tggagttccg 60
cgttacataa cttacggtaa atggcccgcc tggtgaccg cccaacgacc cccgcccatt 120
gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 180
atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 240
aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgccagta 300

```

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac	360
catggtgatg cggtttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg	420
atttccaagt ctccaccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg	480
ggactttcca aaatgtcgtg acaactccgc cccattgacg caaatgggcg gtaggcgtgt	540
acgggtgggag gtctatataa gcagagctgg tttagtgaac cggcgccagt cctccgattg	600
actgagtcgc cgggttacc gtgtatccaa taaacctctc tgcagttgca tccgacttgt	660
ggtctcgtg ttccttggga gggctctctc tgagtgattg actaccgctc agcgggggtc	720
tttcatttgg gggctcgtcc gggatcggga gaccctgccc caggaccac cgaccacca	780
cggggaggta agctggccag caacttatct gtgtctgtcc gattgtctag tgtctatgac	840
tgattttatg cgctcgtc ggtactagt agctaactag ctctgtatct ggcggaacctg	900
tgttggaaact gacgagttcg gaacacccgg ccgcaaccct gggagacgtc ccagggaactt	960
cgggggacctg ttttgtggcc cgacctgagt ccaaaaatcc cgatcgtttt ggactctttg	1020
gtgcaccccc cttagaggag ggatagtgtg ttctggtagg agacgagaac ctaaaacagt	1080
tcccgcctcc gtctgaattt ttgctttcgg tttgggaccg aagccgcgcc gcgctctttg	1140
tctgctgcag catcgttctg tgtgtctctc gtctgactgt gttctctgat ttgtctgaga	1200
atatgggcca gactgttacc actcccttaa gtttgacctt aggtcactgg aaagatgtcg	1260
agcggatcgc tcacaaccag tcggtagatg tcaagaagag acgttgggtt acctctgct	1320
ctgcagaatg gccaaccttt aacgtcggat ggccgcgaga cggcaccttt aaccgagacc	1380
tcatacccca ggttaagatc aaggtctttt cacctggccc gcatggacac ccagaccagg	1440
tcccctacat cgtgacctgg gaagccttgg cttttgaccc ccctccctgg gtcaagccct	1500
ttgtacaccc taagcctcgg cctcctcttc ctccatccgc cccgtctctc ccccttgaa	1560
ctcctcgttc gaccccgct cgatcctccc tttatccagc cctcactcct tctctaggcg	1620
ccaaacctaa acctcaagtt ctttctgaca gtggggggcc gctcatcgac ctacttacag	1680
aagaccccc gccttatagg gacccaagac caccctctc cgacagggac ggaaatggtg	1740
gagaagcgac cctgcggga gaggcaccgg acccctcccc aatggcatct cgcctacgtg	1800
ggagacggga gccccctgtg gccgactcca ctacctgca ggcattcccc ctccgcgag	1860
gaggaaacgg acagcttcaa tactggcgt tctcctcttc tgacctttac aactggaaaa	1920
ataataaccc ttctttttct gaagatccag gtaaaactgac agctctgac gagtctgttc	1980
tcatacccca tcagcccacc tgggacgact gtcagcagct gttggggact ctgctgaccg	2040
gagaagaaaa acaacgggtg ctcttagagg ctagaaaggc ggtgcggggc gatgatgggc	2100
gccccactca actgccaat gaagtogatg ccgcttttcc cctcgagcgc ccagactggg	2160
attacaccac ccaggcaggt aggaaccacc tagtccacta tcgccagttg ctctagcgg	2220

gtctccaaaa cgcgggcaga agccccacca atttggccaa ggtaaaagga ataacacaag	2280
ggcccaatga gtctccctcg gccttcctag agagacttaa ggaagcctat cgaggtaca	2340
ctccttatga ccctgaggac ccagggcaag aaactaatgt gtctatgtct ttcatttggc	2400
agtctgcccc agacattggg agaaagttag agaggttaga agatttaaaa aacaagacgc	2460
ttggagattt ggttagagag gcagaaaaga tctttaataa acgagaaacc ccggaagaaa	2520
gagaggaacg tatcaggaga gaaacagagg aaaaagaaga acgccgtagg acagaggatg	2580
agcagaaaaga gaaagaaaga gatcgtagga gacatagaga gatgagcaag ctattggcca	2640
ctgtcgttag tggacagaaa caggatagac agggaggaga acgaaggagg tcccaactcg	2700
atcgcgacca gtgtgcctac tgcaaagaaa aggggcactg ggctaaagat tgtccaaga	2760
aaccacgagg acctcgggga ccaagacccc agacctccct cctgacccta gatgactagg	2820
gaggtcaggg tcaggagccc cccctgaac ccaggataac cctcaaagtc ggggggcaac	2880
ccgtcacctt cctgtagat actggggccc aacctccgt gctgaccaa aatcctggac	2940
ccctaagtga taagtctgcc tgggtccaag gggctactgg aggaaagcgg tatcgtgga	3000
ccacggatcg caaagtacat ctagtaccg gtaaggtcac ccactcttct ctccatgtac	3060
cagactgtcc ctatcctctg ttaggaagag atttgctgac taaactaaaa gcccaaatcc	3120
actttgaggg atcaggagcc caggttatgg gaccaatggg gcagcccctg caagtgttga	3180
ccctaaatat agaagatgag catcggtac atgagacctc aaaagagcca gatgtttctc	3240
taggggtccac atggctgtct gattttctct aggcctgggc ggaaaccggg ggcatgggac	3300
tggcagttcg ccaagctcct ctgatcatac ctctgaaagc aacctctacc cccgtgtcca	3360
taaaacaata ccccatgtca caagaagcca gactggggat caagcccac atacagagac	3420
tgttggaacca gggaatactg gtaccctgcc agtccccctg gaacacgccc ctgctacccg	3480
ttaagaaacc agggactaat gattataggc ctgtccagga tctgagagaa gtcaacaagc	3540
gggtggaaga catccacccc accgtgcccc acccttaca cctcttgagc gggctcccc	3600
cgtcccacca gtgttacct gtgcttgatt taaaggatgc ctttttctgc ctgagactcc	3660
accccaccag tcagcctctc ttgcctttg agtggagaga tccagagatg ggaatctcag	3720
gacaattgac ctggaccaga ctcccacagg gtttcaaaaa cagtcccacc ctgtttgatg	3780
aggcactgca cagagaccta gcagacttcc ggatccagca cccagacttg atcctgctac	3840
agtacgtgga tgacttactg ctggcgccca cttctgagct agactgcaa caaggactc	3900
gggacctgtt acaaacccta gggaacctcg ggtatcgggc ctcggccaag aaagcccaaa	3960
tttgccagaa acaggtcaag tatctggggt atcttctaaa agagggtcag agatggctga	4020
ctgaggccag aaaagagact gtgatggggc agcctactcc gaagaccct cgacaactaa	4080
gggagttcct agggacggca ggcttctgtc gcctctggat cctgggttt gcagaaatgg	4140

cagccccctt gtaccctctc accaaaacgg ggactctgtt taattggggc ccagaccaac	4200
aaaaggccta tcaagaaatc aagcaagctc ttctaactgc ccagccctg gggttgccag	4260
atttgactaa gccctttgaa ctctttgtcg acgagaagca gggctacgcc aaaggtgtcc	4320
taacgcaaaa actgggacct tggcgtcggc cggtaggccta cctgtccaaa aagctagacc	4380
cagtagcagc tgggtggccc ccttgccctac ggatggtagc agccattgcc gtactgacaa	4440
aggatgcagg caagetaacc atgggacagc cactagtcat tctggccccc catgcagtag	4500
aggcactagt caaacaaccc cccgaccgct ggctttccaa cgcccgatg actcactatc	4560
aggccttgct tttggacagc gaccgggtcc agttcggacc ggtggtagcc ctgaaccgg	4620
ctacgctgct cccactgcct gaggaagggc tgcaacacaa ctgccttgat atcctggccg	4680
aagcccacgg aacccgaccc gacctaacgg accagccgct ccagacgcc gaccacacct	4740
ggtacacgga tggaagcagt ctcttacaag agggacagcg taaggcggga gctgcggtga	4800
ccaccgagac cgaggtaatc tgggctaaag ccctgccagc cgggacatcc gctcagcggg	4860
ctgaactgat agcactcacc caggccctaa agatggcaga aggtaagaag ctaaattgtt	4920
atactgatag cgttatgct tttgtactg cccatatcca tggagaaata tacagaaggc	4980
gtgggttgct cacatcagaa ggcaaagaga tcaaaaataa agacgagatc ttggccctac	5040
taaaagccct ctttctgccc aaaagactta gcataatcca ttgtccagga catcaaaagg	5100
gacacagcgc cgaggctaga ggcaaccgga tggtgacca agcggcccca aaggcagcca	5160
tcacagagac tccagacacc tctaccctcc tcatagaaaa ttcacaccc tacacctcag	5220
aacattttca ttacacagt actgatataa aggacctaac caagttgggg gccatttatg	5280
ataaaaacaaa gaagtattgg gtctaccaag gaaaacctgt gatgcctgac cagtttactt	5340
ttgaattatt agactttctt catcagctga ctcacctcag cttctcaaaa atgaaggctc	5400
tcctagagag aagccacagt ccctactaca tgctgaaccg ggatcgaaca ctcaaaaata	5460
tcactgagac ctgcaaaact tgtgcacaag tcaacgccag caagtctgcc gttaaacagg	5520
gaactagggc cgcggggcat cggcccgga ctcattggga gatcgatttc accgagataa	5580
agcccggtat gtatggctat aaatatcttc tagtttttat agataccttt tctggctgga	5640
tagaagcctt cccaaccaag aaagaaaccg ccaaggtcgt aaccaagaag ctactagagg	5700
agatcttccc caggttcggc atgcctcagg tattgggaac tgacaatggg cctgccttcg	5760
tctccaaggt gagtcagaca gtggcgatc tgttggggat tgattggaaa ttacattgtg	5820
catacagacc ccaaagctca ggccaggtag aaagaatgaa tagaaccatc aaggagactt	5880
taactaaatt aacgcttgca actggtctca gagactgggt gtcctactc cccttagccc	5940
tgtaccgagc ccgcaacagc cggggccccc atggcctcac cccatatgag atcttatatg	6000
gggcaccccc gcccttgta aacttcctg accctgacat gacaagagtt actaacagcc	6060

cctctctcca agctcactta caggetctct acttagtcca gcacgaagtc tggagacctc	6120
tggcggcagc ctaccaagaa caactggacc gaccggtggt acctcaccct taccgagtcg	6180
gcgacacagt gtgggtccgc cgacaccaga ctaagaacct agaacctcgc tggaaaggac	6240
cttacacagt cctgctgacc acccccaccg ccctcaaagt agacggcatc gcagcttgga	6300
tacacgccgc ccacgtgaag gctgccgacc ccgggggtgg accatcctct agactgacat	6360
ggcgcgttca acgctctcaa aacccctca agataagatt aaccctgga agcccttaat	6420
agtcattgga gtcctgttag gagtaggat ggcagagagc ccccatcagg tctttaatgt	6480
aacctggaga gtcaccaacc tgatgactgg gcgtaccgcc aatgccacct ccctcctggg	6540
aactgtacaa gatgccttcc caaaattata ttttgatcta tgtgatctgg tcggagagga	6600
gtgggaccct tcagaccagg aaccgtatgt cgggtatggc tgcaagtacc ccgcagggag	6660
acagcggacc cggacttttg acttttacgt gtgccctggg cataccgtaa agtcggggtg	6720
tgggggacca ggagagggt actgtggtaa atgggggtgt gaaaccaccg gacaggctta	6780
ctggaagccc acatcatcgt gggacctaat ctcccttaag cgcggttaaca cccctggga	6840
cacgggatgc tctaaagttg cctgtggccc ctgctacgac ctctccaaag tatccaatc	6900
cttccaaggg gctactcgag ggggcagatg caaccctcta gtccatagaat tcaatgatgc	6960
aggaaaaaag gctaactggg acgggcccac atcgtgggga ctgagactgt accggacagg	7020
aacagatcct attaccatgt tctccctgac ccggcaggtc cttaatgtgg gaccccgagt	7080
ccccataggg cccaaccag tattaccga ccaagactc ccttctcac caatagagat	7140
tgtaccggct ccacagccac ctagccccct caataccagt tccccctt ccaactaccg	7200
tacaccctca acctccccta caagtccaag tgtcccacag ccaccccag gaactggaga	7260
tagactacta gctctagtca aaggagccta tcaggcgctt aacctacca atccgacaa	7320
gacccaagaa tgttggtgt gcttagtgtc gggacctcct tattacgaag gagtagcgt	7380
cgtgggcaact tataccaato attccaccgc tccggccaac tgtacggcca cttcccaaca	7440
taagcttacc ctatctgaag tgacaggaca gggcctatgc atgggggag tacctaaaac	7500
tcaccaggcc ttatgtaaca ccacccaaag cgcggetca ggatcctact acctgcagc	7560
acccgccgga acaatgtggg cttgcagcac tggattgact ccctgcttgt ccaccacgt	7620
gctcaatcta accacagatt attgtgtatt agttgaactc tggcccagag taatttacca	7680
ctccccgat tatatgtatg gtcagcttga acagcgtacc aaatataaaa gagagccagt	7740
atcattgacc ctggcccttc tactaggagg attaaccatg ggagggttg cagctggaat	7800
agggacgggg accactgcct taattaaaac ccagcagttt gagcagcttc atgcccgtat	7860
ccagacagac ctcaacgaag tcgaaaagtc aattaccaac ctagaaaagt cactgacctc	7920
gttgtctgaa gtagtcctac agaaccgcag aggcctagat ttgctattcc taaaggagg	7980

```

aggtctctgc gcagccctaa aagaagaatg ttgtttttat gcagaccaca cggggctagt      8040
gagagacagc atggccaaat taagagaaag gcttaatcag agacaaaaac tatttgagac      8100
aggccaagga tggttcgaag ggctgtttta tagatcccc tggtttacca ccttaatctc      8160
caccatcatg ggacctctaa tagtactctt actgatctta ctctttggac cttgcattct      8220
caatcgattg gtccaatttg ttaaagacag gatctcagtg gtccaggctc tggttttgac      8280
tcagcaatat caccagctaa aaccataga gtacgagcca tgaacgcgtt actggccgaa      8340
gccgcttga ataggccgg tgtgcgtttg tctatatgtt attttccacc atattgccgt      8400
cttttgcaa tgtgagggcc cggaacctg gcctgtctt cttgacgagc attcctaggg      8460
gtctttcccc tctcgccaaa ggaatgcaag gtctgttgaa tgtcgtgaag gaagcagttc      8520
ctctggaagc ttctgaaga caaacaacgt ctgtagcgac cctttgcagg cagcggaaac      8580
ccccacctg cgacaggtgc ctctcgggc aaaagccacg tgtataagat acacctgcaa      8640
aggcggcaca accccagtg cacgttgta gttggatagt tgtggaaaga gtcaaatggc      8700
tctcctcaag cgtattcaac aaggggctga aggatgcca gaaggtacc cattgtatgg      8760
gatctgatct ggggcctcgg tgcacatgct ttacatgtgt ttagtcgagg ttaaaaaaac      8820
gtctaggccc ccgaaccac ggggacgtg ttttcctttg aaaaacacga ttataaatgg      8880
tgacaggggg aatggcaagc aagtgggac agaagggtat ggacattgcc tatgaggagg      8940
cggccttagg ttacaaagag ggtggtgttc ctattggcgg atgtcttacc aataacaaag      9000
acggaagtgt tctcggtcgt ggtcacaaca tgagatttca aaagggtacc gccacactac      9060
atggtgagat ctccactttg gaaaactgtg ggagattaga gggcaaagtg tacaagata      9120
ccactttgta tacgacgtg tctccatgcg acatgtgtac aggtgccatc atcatgtatg      9180
gtattccacg ctgtgttgtc ggtgagaacg ttaatttcaa aagtaagggc gagaaatatt      9240
tacaactag aggtcacgag gttgttggtg ttgacgatga gaggtgtaaa aagatcatga      9300
aacaatttat cgatgaaaga cctcaggatt ggtttgaaga tattggtgag taggcggccg      9360
cagataaaat aaaagatttt atttagtctc cagaaaaagg ggggaatgaa agacccacc      9420
tgtaggtttg gcaagctagc ttaagtaacg ccattttgca aggcattgaa aaatacataa      9480
ctgagaatag agaagttcag atcaaggta ggaacagatg gaacagctga atatgggcca      9540
aacaggatat ctgtggtgta cagttcctgc ccgggtcag ggccaagaac agatggaaca      9600
gctgaatatg ggccaaacag gatatctgtg gtaagcagtt cctgccccgg ctgaggcca      9660
agaacagatg gtccccagat gcggtccagc cctcagcagt ttctagagaa ccatcagatg      9720
ttccagggt gcccaagga cctgaaatga ccctgtgcct tatttgaact aaccaatcag      9780
ttcgcttctc gcttctgttc gcgcgttct gctccccgag ctcaataaaa gagcccacaa      9840
cccctcactc ggggcgccag tcctccgatt gactgagtcg cccgggtacc cgtgtatcca      9900

```

ataaacccctc ttgcagttgc atccgacttg tggctctcgct gttccttggg agggctctcct	9960
ctgagtgatt gactaccogt cagcgggggt ctttcattac atgtgagcaa aaggccagca	10020
aaaggccagg aaccgtaaaa aggccgcgtt gctggcgttt ttccataggc tccgcccccc	10080
tgacgagcat cacaaaaatc gacgctcaag tcagaggtgg cgaaaccga caggactata	10140
aagataccag gcgtttcccc ctggaagctc cctcgtgcgc tctcctgttc cgaccctgcc	10200
gcttaccgga tacctgtccg cttttctccc ttcgggaagc gtggcgcttt ctcaatgtc	10260
acgctgtagg tatctcagtt cgggtgtagg cgttcgtcc aagctgggt gtgtgcacga	10320
acccccgtt cagcccgacc gctgcgcctt atccgtaac tatcgtcttg agtccaacct	10380
ggtaagacac gacttatcgc cactggcagc agccactggt aacaggatta gcagagcgag	10440
gtatgtaggc ggtgctacag agttcttgaa gtggtggcct aactacggct aactagaag	10500
gacagtattht ggtatctgcg ctctgctgaa gccagttacc ttcggaaaaa gagtggtag	10560
ctcttgatcc ggcaaaaaa ccaccgctgg tagcgggtgt tttttgttt gcaagcagca	10620
gattacgcgc agaaaaaag gatctcaaga agatcctttg atcttttcta cggggtctga	10680
cgctcagtg aacgaaaact cacgttaagg gattttgttc atgagattat caaaaaggat	10740
cttcacctag atccttttaa attaaaaatg aagttttaaa tcaatctaaa gtatatatga	10800
gtaaacttgg tctgacagtt accaatgctt aatcagtgag gcacctatct cagcgatctg	10860
tctatttcgt toatccatag ttgcctgact ccccgctgtg tagataacta cgatacggga	10920
gggcttacca tctggcccca gtgctgcaat gataccgcga gaccacgct caccggctcc	10980
agatttatca gcaataaacc agccagccgg aagggccgag cgcagaagtg gtctgcaac	11040
tttatccgcc tccatccagt ctattaattg ttgccgggaa gctagagtaa gtagttcgcc	11100
agttaatagt ttgcgaacg ttgttgccat tgctgcaggc atcgtggtgt cacgctcgtc	11160
gtttggtatg gcttcattca gctccggttc ccaacgatca aggcgagtta catgatcccc	11220
catgttgtgc aaaaaagcgg ttagctcctt cggctcctcg atcgttgtca gaagtaagtt	11280
ggccgcagtg ttatcactca tggttatggc agcactgcat aattctctta ctgtcatgcc	11340
atccgtaaga tgcttttctg tgactggtga gtactcaacc aagtcattct gagaatagtg	11400
tatgcggcga ccgagttgct cttgcccgcg gtcaacacgg gataataccg cgccacatag	11460
cagaacttta aaagtgtca tcattggaaa acgtttctcg gggcgaaaaa tctcaaggat	11520
cttaccgtg ttgagatcca gttcgatga acccactcgt gcacccaact gatcttcagc	11580
atcttttact ttcaccagcg tttctgggtg agcaaaaaa ggaaggcaaa atgccgcaaa	11640
aaagggaata agggcgacac ggaaatgttg aatactcata ctcttccttt ttcaatatta	11700
ttgaagcatt tatcagggtt attgtctcat gagcggatac atatttgaat gtatttagaa	11760
aaataaacia atagggttgc cgcgcacatt tccccgaaaa gtgccacctg acgtctaaga	11820

aaccattatt atcatgacat taacctataa aaataggcgt atcacgaggc cctttcgtct 11880
tcaagaattc at 11892

<210> 21
<211> 12007
<212> ДНК
<213> Искусственная последовательность

<220>
<223> вектор RCR - pACE-CD

<400> 21
tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg 60
cgttacataa cttacggtaa atggcccgcg tggtgacgcg cccaacgacc ccgcccatt 120
gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 180
atgggtggag tatttacggt aaactgccca cttggcagta catcaagtg atcatatgcc 240
aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgccagta 300
catgacctta tgggactttc ctacttgga gtacatctac gtattagtca tcgctattac 360
catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 420
atttccaagt ctccaccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 480
ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 540
acggtgggag gtctatataa gcagagctgg tttagtgaac cggcgccagt cctccgattg 600
actgagtcgc ccgggtaccc gtgtatccaa taaacctct tgcagttgca tccgacttgt 660
ggtctcgctg ttccttgga gggctctctc tgagtgttg actaccgctc agcgggggctc 720
tttcatttgg gggtcgtcc gggatcgga gacccctgcc cagggaccac cgaccacca 780
ccgggaggtg agctggccag caacttatct gtgtctgtcc gattgtctag tgtctatgac 840
tgattttatg cgctcgctc ggtactagt agctaactag ctctgtatct ggcggaaccg 900
tggtggaact gacgagttcg gaacaccgcg ccgcaaccct gggagacgct ccagggactt 960
cgggggccgt tttgtggcc cgacctgagt ccaaaaatcc cgatcgtttt ggactctttg 1020
gtgcaccccc cttagaggag ggatatgtgg ttctggtagg agacgagaac ctaaaacagt 1080
tcccgcctcc gtctgaattt ttgctttcgg ttggggaccg aagccgcgcc gcgctcttg 1140
tctgtgcag catcgttctg tgtgtctct gtctgactgt gtttctgtat ttgtctgaga 1200
atatgggcca gactgttacc actcccttaa gtttgacctt aggtcactgg aaagatgtcg 1260
agcggatcgc tcacaaccag tcggtagatg tcaagaagag acgttgggtt accttctgct 1320
ctgcagaatg gccaaccttt aacgtcggat ggccgcgaga cggcaccttt aaccgagacc 1380
tcatcaccca ggttaagatc aaggtctttt cacctggccc gcatggacac ccagaccagg 1440
tcccctacat cgtgacctgg gaagccttgg cttttgacct cctccctggt gtcaagccct 1500

ttgtacaccc taagcctccg cctcctcttc ctccatccgc cccgtctctc ccccttgaac	1560
ctcctcgttc gaccccgct cgatcctccc tttatccagc cctcactcct tctctaggcg	1620
ccaaacctaa acctcaagtt ctttctgaca gtggggggcc gctcatcgac ctacttacag	1680
aagaccccc gccttatagg gacccaagac ccccccttc cgacagggac ggaaatggtg	1740
gagaagcgac cctgcggga gaggcaccgg acccctcccc aatggcatct cgcctacgtg	1800
ggagacggga gccccctgtg gccgactcca ctacctgca ggcatcccc ctccgcgcag	1860
gaggaacgg acagcttcaa tactggccgt tctcctcttc tgacctttac aactggaaaa	1920
ataataaccc ttctttttct gaagatccag gtaaaactgac agctctgac gagtctgttc	1980
tcatacccca tcagcccacc tgggacgact gtcagcagct gttggggact ctgctgaccg	2040
gagaagaaaa acaacgggtg ctcttagagg ctagaaggc ggtgcggggc gatgatgggc	2100
gccccactca actgccaat gaagtcgatg ccgcttttcc cctcgagcgc ccagactggg	2160
attacaccac ccaggcaggt aggaaccacc tagtccacta tcgccagttg ctcttagcgg	2220
gtctccaaaa cgcgggcaga agccccacca atttgccaa ggtaaaagga ataacacaag	2280
ggcccaatga gtctccctcg gccttcttag agagacttaa ggaagcctat cgcaggtaca	2340
ctccttatga ccctgaggac ccagggcaag aaactaatgt gtctatgtct ttcatttggc	2400
agtctgcccc agacattggg agaaagttag agaggtaga agatttaaaa aacaagacgc	2460
ttggagattt ggttagagag gcagaaaaga tctttaataa acgagaaacc ccggaagaaa	2520
gagaggaacg tatcaggaga gaaacagagg aaaaagaaga acgccgtagg acagaggatg	2580
agcagaaaga gaaagaaaga gatcgtagga gacatagaga gatgagcaag ctattggcca	2640
ctgtcgttag tggacagaaa caggatagac agggaggaga acgaaggagg tcccaactcg	2700
atcgcgacca gtgtgcctac tgcaaagaaa aggggactg ggctaaagat tgtccaaga	2760
aaccacgagg acctcgggga ccaagacccc agacctccct cctgacccta gatgactagg	2820
gaggtcaggg tcaggagccc cccctgaac ccaggataac cctcaaagtc ggggggcaac	2880
ccgtcacctt cctggtagat actggggccc aacctccgt gctgacccaa aatcctggac	2940
ccctaagtga taagtctgcc tgggtccaag gggctactgg aggaaagcgg tatcgctgga	3000
ccacggatcg caaagtacat ctagtaccg gtaaggtcac ccactcttct ctccatgtac	3060
cagactgtcc ctatcctctg ttaggaagag atttgctgac taaactaaaa gcccaaatcc	3120
actttgaggg atcaggagcc caggttatgg gaccaatggg gcagcccctg caagtgttga	3180
ccctaaatat agaagatgag catcggtac atgagacctc aaaagagcca gatgtttctc	3240
tagggctcac atggctgtct gatcttctc aggcctgggc ggaaaccggg ggcatgggac	3300
tggcagttcg ccaagctcct ctgatcatc ctctgaaagc aacctctacc cccgtgtcca	3360
taaaacaata ccccatgtca caagaagcca gactggggat caagccccac atacagagac	3420

tgttggacca gggaatactg gtaccctgcc agtccccctg gaacacgccc ctgctaccgc	3480
ttaagaaacc agggactaat gattataggc ctgtccagga tctgagagaa gtcaacaagc	3540
gggtggaaga catccacccc accgtgccc acccttaca cctcttgagc gggctccac	3600
cgtoaccacca gtggtacact gtgcttgatt taaaggatgc ctttttctgc ctgagactcc	3660
acccaccag tcagcctctc ttgcctttg agtggagaga tccagagatg ggaatctcag	3720
gacaattgac ctggaccaga cttccacagg gtttcaaaaa cagtccacc ctgtttgatg	3780
aggcactgca cagagaccta gcagacttcc ggatccagca ccagacttg atcctgctac	3840
agtacgtgga tgacttactg ctggcgcca cttctgagct agactgcaa caaggtactc	3900
gggccctggt acaaaccta gggaacctcg ggatcgggc ctcgccaag aaagccaaa	3960
tttgccagaa acaggtcaag tatctgggt atcttctaaa agaggtcag agatggctga	4020
ctgaggccag aaaagagact gtgatgggc agcctactcc gaagaccct cgacaactaa	4080
gggagttcct agggacgca ggcttctgtc gcctctggat ccctgggtt gcagaaatgg	4140
cagccccctt gtaccctctc accaaaacgg ggactctgtt taattggggc ccagaccaac	4200
aaaaggccta tcaagaaac aagcaagctc ttctaactgc ccagccctg gggttgccag	4260
atttgactaa gccctttgaa ctctttgtcg acgagaagca gggctacgcc aaaggtgtcc	4320
taacgcaaaa actgggacct tggcgctggc cgggtggccta cctgtccaaa aagctagacc	4380
cagtagcagc tgggtggccc ccttgccctac ggatggtagc agccattgcc gtactgacaa	4440
aggatgcagg caagctaacc atgggacagc cactagtcac tctggcccc catgcagtag	4500
aggcactagt caaacaacc cccgaccgct ggctttccaa cgcccgatg actcactatc	4560
aggccttgct tttggacag gaccgggtcc agttcggacc ggtggtagcc ctgaacccg	4620
ctacgctgct cccactgcct gaggaagggc tgcaacacaa ctgccttgat atcctggccg	4680
aagcccacgg aacccgaccc gacctaacgg accagccgct ccagacgcc gaccacacct	4740
ggtacacgga tggaagcagt ctcttacaag agggacagcg taaggcgga gctgcggtga	4800
ccaccgagac cgaggtaac tgggctaaag ccctgccagc cgggacatcc gctcagcggg	4860
ctgaactgat agcactcacc caggccctaa agatggcaga aggtaagaag ctaaagtgtt	4920
atactgatag ccgttatgct tttgctactg cccatatcca tggagaaata tacagaaggc	4980
gtgggttgct cacatcagaa ggcaagaga tcaaaaataa agacgagatc ttggccctac	5040
taaaagccct ctttctgccc aaaagactta gcataatcca ttgtccagga catcaaaagg	5100
gacacagcgc cgaggctaga ggcaaccgga tggtgacca agcgcccgga aaggcagcca	5160
tcacagagac tccagacacc tctaccctcc tcatagaaaa tcatcacc tacacctcag	5220
aacattttca ttacacagt actgatataa aggacctaac caagttggg gccatttatg	5280
ataaaacaaa gaagtattgg gtctaccaag gaaaacctgt gatgcctgac cagtttactt	5340

ttgaattatt agactttctt catcagctga ctcacctcag cttctcaaaa atgaaggctc	5400
tcctagagag aagccacagt ccctactaca tgctgaaccg ggatcgaaca ctcaaaaata	5460
tcactgagac ctgcaaagct tgtgcacaag tcaacgccag caagtctgcc gttaaacagg	5520
gaactagggc cgcggggcat cggcccgga ctcattggga gatcgatttc accgagataa	5580
agccccgatt gtatggctat aaatatcttc tagtttttat agataccttt tctggctgga	5640
tagaagcctt cccaaccaag aaagaaaccg ccaaggctgt aaccaagaag ctactagagg	5700
agatcttccc caggttcggc atgcctcagg tattgggaac tgacaatggg cctgccttcg	5760
tctccaaggt gagtcagaca gtggccgacg tgttggggat tgattggaaa ttacattgtg	5820
catacagacc ccaaagctca ggccaggtag aaagaatgaa tagaaccatc aaggagactt	5880
taactaaatt aacgcttgca actggctcta gagactgggt gctcctactc cccttagccc	5940
tgtaccgagc cgcacaacg cggggcccc atggcctcac cccatatgag atcttatatg	6000
gggcaccccc gcccttgta aacttccctg accctgacat gacaagagtt actaacagcc	6060
cctctctcca agctcactta caggctctct acttagtcca gcacgaagtc tggagacctc	6120
tggcggcagc ctaccaagaa caactggacc gaccggtggt acctcaccct taccgagtcg	6180
gcgacacagt gtgggtccgc cgacaccaga ctaagaacct agaacctcgc tggaaaggac	6240
cttacacagt cctgtgacc acccccaccg ccctcaaagt agacggcatc gcagcttgga	6300
tacacgccgc ccacgtgaag gctgccgacc cgggggtggt accatcctct agactgacat	6360
ggcgcggtca acgctctcaa aacccccca agataagatt aaccctgga agcccttaat	6420
agtcattgga gtccgtttag gagtagggat ggcagagagc ccccatcagg tctttaatgt	6480
aacctggaga gtcaccaacc tgatgactgg gcgtaccgcc aatgccacct ccctcctggg	6540
aactgtacaa gatgccttcc caaaattata ttttgatcta tgtgatctgg tcggagagga	6600
gtgggaccct tcagaccagg aaccgtatgt cgggtatggc tgcaagtacc ccgagggag	6660
acagcggacc cggacttttg acttttacgt gtgccctggg cataccgtaa agtcgggggtg	6720
tgggggacca ggagagggct actgtggtaa atgggggtgt gaaaccaccg gacaggctta	6780
ctggaagccc acatcatcgt gggacctaat ctcccttaag cgcggtaaca cccctggga	6840
cacgggatgc tctaaagttg cctgtggccc ctgctacgac ctctccaaag tatccaattc	6900
cttccaaggg gctactcgag ggggcagatg caaccctcta gtctagaat tcactgatgc	6960
aggaaaaaag gctaactggg acgggcccaa atcgtgggga ctgagactgt accggacagg	7020
aacagatcct attaccatgt tctccctgac cggcagggtc cttaatgtgg gaccccgagt	7080
cccataggg cccaaccag tattaccga ccaaagactc cttcctcac caatagagat	7140
tgtaccggt ccacagccac ctagccccct caataccagt tccccccct ccaactaccg	7200
tacaccctca acctccccta caagtccaag tgtcccacag ccaccccag gaactggaga	7260

tagactacta gctctagtca aaggagccta tcaggcgctt aacctcacca atcccgacaa	7320
gacccaagaa tgttggtgtg gcttagtgtc gggacctcct tattacgaag gagtagcggt	7380
cgtgggcact tataccaatc attccaccgc tccggccaac tgtacggcca cttcccaaca	7440
taagcttacc ctatctgaag tgacaggaca gggcctatgc atgggggcag tacctaaaaac	7500
tcaccaggcc ttatgtaaca ccacccaaag cgcgggtcga ggatcctact accttgagc	7560
acccgccgga acaatgtggg cttgcagcac tggattgact ccctgcttgt ccaccacggt	7620
gctcaatcta accacagatt attgtgtatt agttgaactc tggcccagag taatttacca	7680
ctccccgat tataatgtatg gtcagcttga acagcgatcc aaatataaaa gagagccagt	7740
atcattgacc ctggcccttc tactaggagg attaaccatg ggagggtattg cagctggaat	7800
agggacgggg accactgcct taattaaaac ccagcagttt gagcagcttc atgccgctat	7860
ccagacagac ctcaacgaag tcgaaaagtc aattaccaac ctagaaaagt cactgacctc	7920
gttgtotgaa gtagtcctac agaaccgcag aggcctagat ttgctattcc taaaggaggg	7980
aggtctctgc gcagccctaa aagaagaatg ttgtttttat gcagaccaca cggggctagt	8040
gagagacagc atggccaaat taagagaaag gcttaatcag agacaaaaac tatttgagac	8100
aggccaagga tggttcgaag ggctgtttaa tagatcccc tggtttacca ccttaatctc	8160
caccatcatg ggacctctaa tagtactctt actgatctta ctctttggac cttgcattct	8220
caatcgatta gtccaatttg ttaaagacag gatatcagtg gtccaggctc tagttttgac	8280
tcaacaatat caccagctga agcctataga gtacgagcca tgacgtacgt tactggccga	8340
agccgcttg aataaggccg gtgtgcgttt gtctatatgt tattttccac catattgccg	8400
tcttttgga atgtgagggc ccggaaacct ggcctgtct tcttgacgag cattcctagg	8460
ggtctttccc ctctcgcaa aggaatgcaa ggtctgttga atgtcgtgaa ggaagcagtt	8520
cctctggaag cttcttgaag acaacaacg tctgtagcga ccctttgcag gcagcggaac	8580
ccccacctg gcgacagggt cctctcggc caaaagccac gtgtataaga tacacctgca	8640
aaggcggcac aaccccagtg ccacgttgtg agttggatag ttgtggaaag agtcaaatgg	8700
ctctcctcaa gcgtattcaa caaggggctg aaggatgccc agaaggatcc ccattgtatg	8760
ggatctgato tggggcctcg gtgcacatgc ttacatgtg tttagtcgag gttaaaaaaa	8820
cgtctaggcc ccccgaaaca cggggacgtg gttttccttt gaaaaacacg ataataccat	8880
ggtgacaggg ggaatggcaa gcaagtggga tcagaagggt atggacattg cctatgagga	8940
ggcggcctta gggtacaaag aggggtgtgt tcctattggc ggatgtctta tcaataacaa	9000
agacggaagt gttctcggtc gtggtcacia catgagattt caaaagggat ccgccacact	9060
acatgggtgag atctccactt tggaaaactg tgggagatta gagggcaaag tgtacaaaga	9120
taccactttg tatacgacgc tgtctccatg cgacatgtgt acaggtgccca tcatcatgta	9180

tggtattcca cgctgtgttg tcggtgagaa cgttaatttc aaaagtaagg gcgagaaata	9240
tttataaact agaggtcacg aggttgttgt tgttgacgat gagaggtgta aaaagatcat	9300
gaaacaattt atcgatgaaa gacctcagga ttggtttgaa gatattggtg agtaggcggc	9360
cgcgccatag ataaaaataa agattttatt tagtctccag aaaaaggggg gaatgaaaga	9420
ccccacctgt aggtttggca agctagctta agtaacgcca ttttgcaagg catggaaaaa	9480
tacataactg agaatagaga agttcagatc aaggtcagga acagatggaa cagctgaata	9540
tgggccaaac aggatatctg tggtaagcag ttcttgcccc ggctcagggc caagaacaga	9600
tggaacagct gaatatgggc caaacaggat atctgtggta agcagttcct gccccggctc	9660
agggccaaaga acagatggtc ccagatgcg gtccagccct cagcagtttc tagagaacca	9720
tcagatgttt ccaggggtgcc ccaaggacct gaaatgacct tgtgccttgt ttaactaac	9780
caatcagttc gcttctcgtc tctgttcgcg cgcttctgct ccccgagctc aataaaagag	9840
cccacaacc ctcactcggg gcgccagtc tccgattgac tgagtcgccc gggtagccgt	9900
gtatccaata aaccctcttg cagttgcacg cgacttgtag tctcgctgtt ccttgggagg	9960
gtctcctctg agtgattgac tacccgtcag cgggggtctt tcatttgggg gctcgtccgg	10020
gacgaggaga cccctgccca gggaccaccg acccaccacc gggaggtgag ctggctgcct	10080
cgcgcggttc ggtgatgacg gtgaaaacct ctgacatgtg agcaaaaggc cagcaaaagg	10140
ccaggaaccg taaaaaggcc gcgttgctgg cgtttttcca taggctccgc cccctgacg	10200
agcatcacia aaatcgacgc tcaagtcaga ggtggcgaaa cccgacagga ctataaagat	10260
accaggcgtt tccccctgga agctccctcg tgcgtctcc tgttccgacc ctgccgtta	10320
ccgataacct gtccgccttt ctcccttcgg gaagcgtggc gctttctcaa tgctcacgct	10380
gtaggatatc cagttcggtg taggtcgttc gctccaagct gggtgtgtg cacgaacccc	10440
ccgttcagcc cgaccgctgc gccttatccg gtaactatcg tcttgagtc aaccggtaa	10500
gacacgactt atcgccactg gcagcagcca ctggtaacag gattagcaga gcgaggtatg	10560
taggcggtgc tacagagttc ttgaagtggg ggcctaacta cggtacact agaaggacag	10620
tatttggtat ctgcgctctg ctgaagccag ttaccttcgg aaaaagagtt ggtagctctt	10680
gacgaggcaa acaaaccacc gctggtagcg gtgggttttt tgtttgcaag cagcagatta	10740
cgcgagaaaa aaaaggatct caagaagatc ctttgatctt ttctacgggg tctgacgctc	10800
agtggaaacg aaactcacgt taagggaatt tggatcatg attatcaaaa aggatcttca	10860
cctagatcct tttaaattaa aaatgaagtt ttaaataaat ctaaagtata tatgagtaaa	10920
cttggtctga cagttaccaa tgcttaatac gtgaggcacc tatctcagcg atctgtctat	10980
ttcgttcacg catagttgcc tgactccccg tcgtgtagat aactacgata cgggagggct	11040
taccatctgg cccagtgct gcaatgatac cgcgagacc acgctcaccg gctccagatt	11100

```

tatacagcaat aaaccagcca gccggaagg cgcagcgag aagtggctct gcaactttat 11160
ccgcctccat ccagtcctatt aattgttgcc gggaagctag agtaagtagt tcgccagtta 11220
atagtgttgc caacgttggt gccattgctg caggcatcgt ggtgtcacgc tcgtcgtttg 11280
gtatggcttc attcagctcc gggtcccaac gatcaaggcg agttacatga tcccccatgt 11340
tgtgcaaaaa agcgggttagc tccttcggtc ctccgatcgt tgtcagaagt aagttggccg 11400
cagtgttatc actcatgggt atggcagcac tgcataattc tcttactgtc atgccatccg 11460
taagatgctt ttctgtgact ggtgagtact caaccaagtc attctgagaa tagtgtatgc 11520
ggcgaccgag ttgctcttgc ccggcgtaaa cacgggataa taccgcgcca catagcagaa 11580
ctttaaaagt gctcatcatt ggaaaacgtt cttcggggcg aaaactctca aggatcttac 11640
cgctgttgag atccagttcg atgtaaccca ctctgtcacc caactgatct tcagcatctt 11700
ttactttcac cagcgtttct ggggtagcaa aaacagggaag gcaaaatgcc gcaaaaaagg 11760
gaataagggc gacacggaaa tgttgaatac tcatactctt cctttttcaa tattattgaa 11820
gcatttatca gggttattgt ctcatgagcg gatacatatt tgaatgtatt tagaaaaata 11880
aacaataagg ggttcgcgcg acatttcccc gaaaagtgcc acctgacgtc taagaaacca 11940
ttattatcat gacattaacc tataaaaata ggcgatcac gaggcccttt cgtcttcaag 12000
aattcat 12007

```

```

<210> 22
<211> 11893
<212> ДНК
<213> Искусственная последовательность

```

```

<220>
<223> вектор RCR - pAC3-γCD2

```

```

<400> 22
tagttattaa tagtaataaa ttacggggtc attagttcat agcccatata tggagttccg 60
cgttacataa cttacggtaa atggcccgcc tggtgaccg cccaacgacc ccgcccatt 120
gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 180
atgggtggag tatttacggt aaactgcccc cttggcagta catcaagtgt atcatatgcc 240
aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgccagta 300
catgacctta tgggactttc ctacttgcca gtacatctac gtattagtca tcgctattac 360
catggtgatg cggtttttgc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 420
atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 480
ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 540
acgggtggag gtctatataa gcagagctgg tttagtgaac cggcgccagt cctccgattg 600
actgagtcgc ccgggtaccc gtgtatccaa taaacctct tgcagttgca tccgacttgt 660

```

ggtctcgtg ttccttgga ggtctcctc tgagtgattg actaccgctc agcgggggtc	720
tttcatttgg gggctcgtcc gggatcgga gaccctgcc caggaccac cgaccacca	780
ccgggaggta agctggccag caacttatct gtgtctgtcc gattgtctag tgtctatgac	840
tgattttatg cgctcgtc ggtactagt agctaactag ctctgtatct ggcggaccg	900
tggtggaact gacgagttcg gaacaccgg ccgcaaccct gggagacgct ccagggactt	960
cgggggccgt ttttgtggcc cgacctgagt ccaaaaatcc cgatcgttt ggactctttg	1020
gtgcaccccc cttagaggag ggatatgtgg ttctggtagg agacgagaac ctaaaacagt	1080
tcccgctcc gtctgaattt ttgctttcgg ttgggaccg aagccgcgc gcgctcttg	1140
tctgtgcag catcgttctg tgttgtctct gtctgactgt gttctgtat ttgtctgaaa	1200
atatgggcca gactgttacc actcccttaa gtttgacctt aggtcactgg aaagatgtcg	1260
agcggtcgc tcacaaccag tcggtagatg tcaagaagag acgttgggtt acctctgct	1320
ctgcagaatg gccaaccttt aacgtcgat gccgcgaga cggcaccttt aaccgagacc	1380
tcataccca ggttaagatc aaggtcttt cacctggccc gcatggacac ccagaccagg	1440
tcccctacat cgtgacctgg gaagccttgg cttttgacct ccctccctgg gtcaagccct	1500
ttgtacacc taagcctcc cctctcttc ctccatccgc cccgtctctc cccctgaac	1560
ctcctcgttc gacccgcct cgatcctccc ttatccagc cctcactcct tctctaggcg	1620
ccaaacctaa acctcaagtt ctttctgaca gtggggggcc gctcatcgac ctacttacag	1680
aagaccccc gccttatagg gaccaagac cacccttc cgacaggga ggaaatggtg	1740
gagaagcgac cctgcggga gaggcaccg acccctccc aatggcatct cgcctacgtg	1800
ggagacggga gcccctgtg gccgactcca ctacctgca ggcattcccc ctccgcgag	1860
gaggaaacgg acagcttcaa tactggcgt tctcctctc tgacctttac aactggaaaa	1920
ataataaccc ttctttttct gaagatccag gtaaaactgac agctctgac gagtctgtcc	1980
tcataccca tcagccacc tgggacgact gtcagcagct gttggggact ctgctgaccg	2040
gagaagaaaa acaacgggtg ctcttagagg ctagaaaggc ggtgcggggc gatgatgggc	2100
gccccactca actgccaat gaagtcgat ccgcttttcc cctcgagcgc ccagactggg	2160
attacaccac ccaggcaggt aggaaccacc tagtccacta tcgccagttg ctcttagcgg	2220
gtctccaaaa cgcgggcaga agccccacca atttggcaa ggtaaaagga ataacacaag	2280
ggcccaatga gtctccctcg gccttctag agagacttaa ggaagcctat cgcaggta	2340
ctccttatga ccctgaggac ccagggcaag aaactaatgt gtctatgtct ttcatttggc	2400
agtctgcccc agacattggg agaaagttag agaggtaga agatttaaaa aacaagacgc	2460
ttggagattt ggttagagag gcagaaaaga tctttaataa acgagaaacc ccggaagaaa	2520
gagaggaacg tatcaggaga gaaacagagg aaaaagaaga acgccgtagg acagaggatg	2580

agcagaaaga gaaagaaaga gatcgtagga gacatagaga gatgagcaag ctattggcca	2640
ctgtcgttag tggacagaaa caggatagac agggaggaga acgaaggagg tcccaactcg	2700
atcgcgacca gtgtgcctac tgcaagaaa aggggactg ggctaaagat tgtccaaga	2760
aaccacgagg acctcgggga ccaagacccc agacctccct cctgacccta gatgactagg	2820
gaggtcaggg tcaggagccc cccctgaac ccaggataac cctcaaagtc ggggggcaac	2880
ccgtcacctt cctggtagat actggggccc aacactccgt gctgacccaa aatcctggac	2940
ccctaagtga taagtctgcc tgggtccaag gggtacttg aggaaagcgg tatcgtgga	3000
ccacggatcg caaagtacat ctagtaccg gtaaggtcac ccactcttcc ctccatgtac	3060
cagactgtcc ctatcctctg ttaggaagag atttgctgac taaactaaaa gcccaaatcc	3120
actttgaggg atcaggagcc caggttatgg gaccaatggg gcagcccctg caagtgttga	3180
ccctaaatat agaagatgag tatcggctac atgagacctc aaaagagcca gatgtttctc	3240
tagggtccac atggctgtct gattttctc aggcctgggc ggaaccggg ggcattggac	3300
tggcagttcg ccaagctcct ctgatcatac ctctgaaagc aacctctacc cccgtgtcca	3360
taaaacaata ccccatgtca caagaagcca gactggggat caagccccac atacagagac	3420
tgttggacca gggaatactg gtaccctgcc agtccccctg gaacacgccc ctgtaccgg	3480
ttaagaaacc agggactaat gattataggc ctgtccagga tctgagagaa gtcaacaagc	3540
gggtggaaga catccacccc accgtgccca acccttaca cctcttgagc gggctccac	3600
cgtcccacca gtgttacact gtgcttgatt taaaggatgc ctttttctgc ctgagactcc	3660
accccaccag tcagcctctc ttgcctttg agtggagaga tccagagatg ggaatctcag	3720
gacaattgac ctggaccaga ctcccacagg gtttcaaaaa cagtcccacc ctgtttgatg	3780
aggcactgca cagagacctc gcagacttcc ggatccagca ccagacttg atcctgtac	3840
agtacgtgga tgacttactg ctggccgcca cttctgagct agactgcca caaggtactc	3900
gggccctgtt acaaacccta gggaacctcg ggtatcgggc ctcgccaag aaagccaaa	3960
tttgccagaa acaggtcaag tatctgggtt atcttctaaa agagggtcag agatggctga	4020
ctgaggccag aaaagagact gtgatgggc agcctactcc gaagaccctc cgacaactaa	4080
gggagttcct agggacggca ggcttctgtc gcctctggat ccctgggttt gcagaaatgg	4140
cagccccctt gtaccctctc accaaaacgg ggactctgtt taattggggc ccagaccaac	4200
aaaaggccta tcaagaaatc aagcaagctc ttctaactgc ccagccctg gggttgccag	4260
atttgactaa gccctttgaa ctctttgtcg acgagaagca gggctacgcc aaaggtgtcc	4320
taacgcaaaa actgggacct tggcgctggc cgggtgccta cctgtccaaa aagctagacc	4380
cagtagcagc tgggtggccc ccttgcttac ggatggtagc agccattgcc gtactgacaa	4440
aggatgcagg caagctaacc atgggacagc cactagtcac tctggccccc catgcagtag	4500

aggcactagt caaacaaccc cccgaccgct ggctttccaa cgcccgatg actcactatc	4560
aggccttgct tttggacacg gaccgggtcc agttcggacc ggtggtagcc ctgaacccgg	4620
ctacgtgct cccactgcct gaggaagggc tgcaacacaa ctgccttgat atcctggccg	4680
aagccacagc aaccgaccc gacctaacgg accagccgct ccagacgcc gaccacacct	4740
ggtacacgga tggaagcagt ctcttacaag agggacacgc taaggcggga gctgcggtga	4800
ccaccgagac cgaggtaatc tgggctaaag ccctgccagc cgggacatcc gctcagcggg	4860
ctgaactgat agcactcacc caggccctaa agatggcaga aggtaagaag ctaaattgtt	4920
atactgatag ccgttatgct tttgctactg cccatatcca tggagaaata tacagaaggc	4980
gtgggttgct cacatcagaa ggcaaagaga tcaaaaataa agacgagatc ttggccctac	5040
taaaagccct ctttctgccc aaaagactta gcataatcca ttgtccagga catcaaaagg	5100
gacacagcgc cgaggctaga ggcaaccgga tggtgacca agcggcccca aaggcagcca	5160
tcacagagac tccagacacc tctaccctcc tcatagaaaa ttcacacccc tacacctcag	5220
aacattttca ttacacagt actgatataa aggacctaac caagttgggg gccatttatg	5280
ataaaacaaa gaagtattgg gtctaccaag gaaaacctgt gatgcctgac cagtttactt	5340
ttgaattatt agactttctt catcagctga ctcacctcag cttctcaaaa atgaaggctc	5400
tcctagagag aagccacagt ccctactaca tgctgaaccg ggatcgaaca ctcaaaaata	5460
tcactgagac ctgcaaaagt tgtgcacaag tcaacgccag caagtctgcc gttaaacagg	5520
gaactagggc ccgcgggcat cggcccgga ctcattggga gatcgatttc accgagataa	5580
agcccgatt gtatggctat aaatatcttc tagtttttat agataccttt tctggctgga	5640
tagaagcctt cccaaccaag aaagaaaccg ccaaggtcgt aaccaagaag ctactagagg	5700
agatcttccc caggttcggc atgcctcagg tattgggaac tgacaatggg cctgccttcg	5760
tctccaaggt gagtcagaca gtggccgac tgttggggat tgattggaaa ttacattgtg	5820
catacagacc ccaaagctca ggccaggtag aaagaatgaa tagaaccatc aaggagactt	5880
taactaaatt aacgcttgca actggctcta gagactgggt gctcctactc cccttagccc	5940
tgtaccgagc ccgcaacacg ccgggcccc atggcctcac cccatatgag atcttatatg	6000
gggaccccc gcccttgta aacttccctg accctgacat gacaagagtt actaacagcc	6060
cctctctcca agctcactta caggctctct acttagtcca gcacgaagtc tggagacctc	6120
tggcggcagc ctaccaagaa caactggacc gaccggtggt acctcacctt taccgagtcg	6180
gcgacacagt gtgggtccgc cgacaccaga ctaagaacct agaacctcgc tggaaaggac	6240
cttacacagt cctgtgacc acccccaccg ccctcaaagt agacggcatc gcagcttgga	6300
tacacgccgc ccacgtgaag gctgccgacc ccgggggtgg accatcctct agactgacat	6360
ggcgcggtta acgctctcaa aacccctca agataagatt aaccctgga agcccttaat	6420

agtcattgga gtcctgttag gagtaggat ggcagagagc ccccatcagg tctttaatgt	6480
aacctggaga gtcaccaacc tgatgactgg gcgtaccgcc aatgccacct cctcctggg	6540
aactgtacaa gatgccttcc caaaattata ttttgatcta tgtgatctgg tcggagagga	6600
gtgggaccct tcagaccagg aaccgtatgt cgggtatggc tgcaagtacc ccgcagggag	6660
acagcggacc cggacttttg acttttacgt gtgccctggg cataccgtaa agtcgggggtg	6720
tgggggacca ggagagggct actgtggtaa atgggggtgt gaaaccaccg gacaggctta	6780
ctggaagccc acatcatcgt gggacctaat ctcccttaag cgcggtaaca cccctggga	6840
cacgggatgc tctaaagttg cctgtggccc ctgctacgac ctctccaaag tatccaattc	6900
cttccaaggg gctactcgag ggggcagatg caaccctcta gtctagaat tcaactgatgc	6960
aggaaaaaag gctaactggg acgggcccac atcgtgggga ctgagactgt accggacagg	7020
aacagatcct attaccatgt tctccctgac ccggcaggtc cttaatgtgg gaccccgagt	7080
ccccataggg cccaaccagc tattaccoga ccaagactc ccttctcac caatagagat	7140
tgtaccggtc ccacagccac ctagcccct caataccagt taccgccctt ccactaccag	7200
tacaccctca acctccccta caagtccaag tgtcccacag ccaccccag gaactggaga	7260
tagactacta gctctagtca aaggagccta tcaggcgctt aacctacca atcccgacaa	7320
gacccaagaa tgttggtgtg gcttagtgtc gggacctcct tattacgaag gagtagcggt	7380
cgtgggcaact tataccaatc attccaccgc tccggccaac tgtacggcca cttcccaaca	7440
taagcttacc ctatctgaag tgacaggaca gggcctatgc atgggggcag tacctaaaac	7500
tcaccaggcc ttatgtaaca ccacccaaag cgcgggtca ggatcctact accttgacgc	7560
accgcggga acaatgtggg cttgcagcac tggattgact ccctgcttgt ccaccacggt	7620
gtcaaatcta accacagatt attgtgtatt agttgaactc tggcccagag taatttacca	7680
ctccccgat tatatgtatg gtcagcttga acagcgacc aaatataaaa gagagccagt	7740
atcattgacc ctggcccttc tactaggagg attaacctag ggagggattg cagctggaat	7800
agggacggg accactgcct taattaaaac ccagcagttt gagcagcttc atgccgctat	7860
ccagacagac ctcaacgaag tcgaaaagtc aattaccaac ctagaaaagt cactgacctc	7920
gttgtctgaa gtagtcctac agaaccgcag aggcctagat ttgctattcc taaaggaggg	7980
aggtctctgc gcagccctaa aagaagaatg ttgtttttat gcagaccaca cggggctagt	8040
gagagacagc atggccaaat taagagaaa gcttaatcag agacaaaaac tatttgagac	8100
aggccaagga tggttcgaag ggctgtttta tagatcccc tggtttacca ccttaatctc	8160
caccatcatg ggacctctaa tagtactctt actgatctta ctctttggac cttgcattct	8220
caatcgattg gtccaatttg ttaaagacag gatctcagtg gtccaggctc tggttttgac	8280
tcagcaatat caccagctaa aaccataga gtacgagcca tgaacgcgtt actggccgaa	8340

gccgcttgga ataaggccgg tgtgcgtttg tctatatgtt attttccacc atattgccgt	8400
cttttgga tgtgagggcc cggaacctg gccctgtctt cttgacgagc attcctaggg	8460
gtctttcccc tctcgccaaa ggaatgcaag gtctgttgaa tgtcgtgaag gaagcagttc	8520
ctctggaagc ttcttgaaga caaacaacgt ctgtagcgac cttttgcagg cagcggaaacc	8580
ccccacctgg cgacaggtgc ctctgcggcc aaaagccacg tgtataagat acacctgcaa	8640
aggcggcaca accccagtgc cacgttgtga gttggatagt tgtggaaaga gtcaaatggc	8700
tctcctcaag cgtattcaac aaggggctga aggatgccc gaaggtagcc cattgtatgg	8760
gatctgatct ggggcctcgg tgcacatgct ttacatgtgt ttagtcgagg ttaaaaaaac	8820
gtctagggccc ccgaaccac ggggacgtgg ttttctttt aaaaacacga ttataaatgg	8880
tgaccggcgg catggcctcc aagtgggacg aaaagggcat ggatatcgct tacgaggagg	8940
ccctgctggg ctacaaggag ggcgcggtgc ctatcgcgcg ctgtctgac aacaacaagg	9000
acggcagtggt gctgggcagg ggccacaaca tgaggttcca gaagggtccc gccaccctgc	9060
acggcgagat ctccaccctg gagaactgtg gcaggctgga gggcaagggtg tacaaggaca	9120
ccaccctgta caccaccctg tcccctgtg acatgtgtac cgcgctatc atcatgtacg	9180
gcatccctag gtgtgtgac ggcgagaacg tgaactcaa gtccaagggc gagaagtacc	9240
tgcaaaccag gggccacgag gtggtggtt ttgacgatga gaggtgtaag aagctgatga	9300
agcagttcat cgacgagagg cctcaggact ggttcgagga tatcgcgag taagcggcgg	9360
cagataaaat aaaagatttt atttagtctc cagaaaaagg ggggaatgaa agaccccacc	9420
tgtaggtttg gcaagctagc ttaagtaacg ccattttgca aggcattggaa aaatacataa	9480
ctgagaatag agaagttcag atcaaggta ggaacagatg gaacagctga atatgggcca	9540
aacaggatat ctgtggtgag cagttcctgc cccggctcag ggccaagaac agatggaaca	9600
gctgaatatg ggccaaacag gatatctgtg gtaagcagtt cctgccccgg ctcagggccca	9660
agaacagatg gtccccagat gcggtccagc cctcagcagt ttctagagaa ccacagatg	9720
tttccagggt gccccaagga cctgaaatga ccctgtgcct tatttgaact aaccaatcag	9780
ttcgtttctc gcttctgttc gcgcgcttct gctccccgag ctcaataaaa gagcccacaa	9840
cccctcactc ggggcgcag tctcctcgatt gactgagtcg cccgggtacc cgtgtatcca	9900
ataaaccttc ttgcagttgc atccgacttg tggctctcgt gttccttggg agggctctct	9960
ctgagtgatt gactaccctg cagcgggggt ctttcattac atgtgagcaa aaggccagca	10020
aaaggccagg aaccgtaaaa agggcgcgtt gctggcggtt ttccataggg tccgcccccc	10080
tgacgagcat caaaaaatc gacgctcaag tcagaggtgg cgaaaccga caggactata	10140
aagataccag gcgtttcccc ctggaagctc cctcgtgcgc tctcctgttc cgaccctgcc	10200
gcttaccgga tacctgtccg cctttctccc ttcggaagc gtggcgcttt ctcatagctc	10260

```

acgctgtagg tatctcagtt cgggtgtagt cgttcgctcc aagctgggct gtgtgcacga 10320
accccccggt cagcccgacc gctgcgcctt atccggtaac tatcgtcttg agtccaaccc 10380
ggtaagacac gacttatcgc cactggcagc agccactggt aacaggatta gcagagcgag 10440
gtatgtaggc ggtgctacag agttcttgaa gtggtggcct aactacggct acactagaag 10500
gacagtatth ggtatctcgc ctctgctgaa gccagttacc ttcggaaaaa gagttggtag 10560
ctcttgatcc ggcaaaaaa ccaccgctgg tagcgggtgt ttttttgtt gcaagcagca 10620
gattacgcgc agaaaaaaag gatctcaaga agatcctttg atcttttcta cggggtctga 10680
cgctcagtg aacgaaaact cacgttaagg gatttttggtc atgagattat caaaaaggat 10740
cttcacctag atccttttaa attaaaaatg aagtttttaa tcaatctaaa gtatatatga 10800
gtaaacttgg tctgacagtt accaatgctt aatcagttag gcacctatct cagcgatctg 10860
tctatttctg tcatccatag ttgcctgact ccccgctcgt tagataacta cgatacggga 10920
gggcttacca tctggcccca gtgctgcaat gataccgca gaccacgct caccgctcc 10980
agatttatca gcaataaacc agccagccgg aagggccgag cgcagaagtg gtctgcaac 11040
tttatccgcc tccatccagt ctattaattg ttgccgggaa gctagagtaa gtagttcgcc 11100
agttaatagt ttgcgcaacg ttgttgccat tgctgcaggc atcgtggtgt cagctcgtc 11160
gtttggtatg gcttcattca gctccggttc ccaacgatca aggcgagtta catgatcccc 11220
catgttgtgc aaaaaagcgg ttagctcctt cggctcctcg atcgttgtca gaagtaagtt 11280
ggccgcagtg ttatcactca tggttatggc agcactgcat aattctctta ctgtcatgcc 11340
atccgtaaga tgcttttctg tgactgggta gtactcaacc aagtcattct gagaatagtg 11400
tatgcggcga ccgagttgct cttgcccgcc gtcaacacgg gataataccg cgccacatag 11460
cagaacttta aaagtgtca tcattgaaa acgttcttgc gggcgaaaa tctcaaggat 11520
cttaccgctg ttgagatcca gttcgatgta acccactcgt gcacccaact gatcttcagc 11580
atcttttact ttcaccagcg tttctgggtg agcaaaaaca ggaaggcaaa atgccgcaaa 11640
aaaggggaata agggcgacac ggaaatgttg aatactcata ctcttccttt ttcaatatta 11700
ttgaagcatt tatcagggtt attgtctcat gagcggatac atatttgaat gtatttagaa 11760
aaataaacia ataggggttc cgcgcacatt tccccgaaa gtgccacctg acgtctaaga 11820
aaccattatt atcatgacat taacctataa aaataggcgt atcacgaggc cctttcgtct 11880
tcaagaattc cat 11893

```

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Рекомбинантный репликационно-компетентный ретровирус (RCR), способный трансформировать клетку-мишень человека полинуклеотидом, кодирующим полипептид с цитозиндезаминазной активностью и повышенной стабильностью в клетках человека, который катализирует превращение экзогенного 5-фторцитозина в 5-фторурацил, обеспечивая цитотоксическое действие у индивида с нарушением клеточной пролиферации, где ретровирус содержит

ретровирусный белок GAG;

ретровирусный белок POL;

оболочку ретровируса;

ретровирусный полинуклеотид, содержащий подходящий для экспрессии в клетке млекопитающего промотор на 5'-конце ретровирусного полинуклеотида, домены gag, pol и env ретровирусного полинуклеотида, последовательность длинных концевых повторов (LTR) на 3'-конце ретровирусного полинуклеотида;

кассету, содержащую внутренний сайт связывания рибосомы (IRES), функционально связанный с полинуклеотидом, кодирующим полипептид с цитозиндезаминазной активностью и содержащим нуклеотиды от приблизительно 8877 до приблизительно 9353 последовательности SEQ ID NO: 19, оптимизированной по кодонам для экспрессии в клетках человека, где указанная кассета расположена между доменом env и последовательностью LTR; и

последовательности, действующие в цис-положении, необходимые для обратной транскрипции, упаковки и интеграции RCR в клетку-мишень, где последовательности, действующие в цис-положении, расположены рядом с последовательностью LTR.

2. Ретровирус по п.1, где ретровирусный полинуклеотид получен из вируса лейкоза мышей (MLV), вируса мышинного лейкоза Молони (MoMLV), вируса лейкоза кошек или вируса лейкоза гиббонов (GALV).

3. Ретровирус по п.2, где MLV представляет собой амфотропный MLV.

4. Ретровирус по п.1, где ретровирус представляет собой гамма-ретровирус.

5. Ретровирус по п.1, где клетка-мишень представляет собой клетку с нарушением клеточной пролиферации.

6. Ретровирус по п.1, где клетка-мишень представляет собой неопластическую клетку.

7. Ретровирус по п.5, где нарушение клеточной пролиферации выбрано из рака легких, рака толстой и прямой кишки, рака груди, рака предстательной железы, рака мочевыводящих путей, рака матки, рака головного мозга, рака тканей головы и шеи, рака поджелудочной железы, меланомы, рака желудка и рака яичников.

8. Ретровирус по п.1, где промотор представляет собой промотор, регулируемый гормоном, или промотор, регулируемый цитокином.

9. Ретровирус по п.1, где промотор представляет собой тканеспецифичный промотор.

10. Ретровирус по п.9, где тканеспецифичный промотор содержит по меньшей мере один элемент андрогенного ответа (ARE).

11. Ретровирус по п.10, где элемент андрогенного ответа получен из промотора пробазина.

12. Ретровирус по п.9, где тканеспецифичный промотор включает промотор пробазина.

13. Ретровирус по п.1, где промотор представляет собой промотор CMV, содержащий нуклеотиды от приблизительно 1 до приблизительно 582 последовательности SEQ ID NO: 19, 20 или 22, причем промотор может включать модификацию одного или нескольких нуклеотидов без изменения функциональной активности.

14. Ретровирус по п.1, где промотор содержит нуклеотиды от приблизительно 1 до приблизительно 582 последовательности SEQ ID NO: 19 или 20.

15. Ретровирус по п.1, где промотор содержит полинуклеотидный домен CMV-R-U5.

16. Ретровирус по п.15, где домен CMV-R-U5 содержит ранний промотор цитомегаловируса человека, функционально связанный с областью R-U5 MLV.

17. Ретровирус по п.16, где домен CMV-R-U5 содержит нуклеотиды от приблизительно 1 до приблизительно 1202 последовательности SEQ ID NO: 19, 20 или 22 или последовательности, которая по меньшей мере на 95% идентична последовательности SEQ ID NO: 19, 20 или 22, где указанный полинуклеотидный домен активизирует транскрипцию молекулы нуклеиновой кислоты, функционально связанной с ним.

18. Ретровирус по п.1, где домен gag полинуклеотида получен из гамма-ретровируса.

19. Ретровирус по п.18, где домен gag содержит нуклеотиды от приблизительно 1203 до приблизительно 2819 последовательности SEQ ID NO: 19 или последовательности, которая по меньшей мере на 95, 98, 99 или 99,8% идентична ей.

20. Ретровирус по п.1, где домен pol полинуклеотида получен из гамма-ретровируса.

21. Ретровирус по п.20, где домен pol содержит нуклеотиды от приблизительно 2820 до прибли-

тельно 6358 последовательности SEQ ID NO: 19 или последовательности, которая по меньшей мере на 95, 98, 99 или 99,9% идентична ей.

22. Ретровирус по п.1, где домен env содержит нуклеотиды от приблизительно 6359 до приблизительно 8323 последовательности SEQ ID NO: 19 или последовательности, которая по меньшей мере на 95, 98, 99 или 99,8% идентична ей.

23. Ретровирус по п.1, где IRES получен из вируса энцефаломиокардита.

24. Ретровирус по п.23, где IRES содержит нуклеотиды от приблизительно 8327 до приблизительно 8876 последовательности SEQ ID NO: 19 или последовательности, которая по меньшей мере на 95, 98 или 99% идентична ей.

25. Ретровирус по п.1, где 3'-концевая последовательность LTR получена из гамма-ретровируса.

26. Ретровирус по п.25, где 3'-концевая последовательность LTR содержит домен U3-R-U5.

27. Ретровирус по п.26, где 3'-концевая последовательность LTR содержит нуклеотиды от приблизительно 9405 до приблизительно 9998 последовательности SEQ ID NO: 19 или последовательности, которая по меньшей мере на 95, 98 или 99,5% идентична ей.

28. Ретровирус по п.1, где ретровирусный полинуклеотид содержит последовательность SEQ ID NO: 19, 20 или 22.

29. Ретровирус по п.1, где ретровирусный полинуклеотид находится в проретровирусной форме и содержит в направлении от 5'-конца к 3'-концу

домен CMV-R-U5, содержащий ранний промотор цитомегаловируса человека, функционально связанный с областью R-U5 MLV;

последовательности, действующие в цис-положении, содержащие участок PBS для связывания праймера для обратной транскриптазы;

5'-концевой сайт сплайсинга и

Ψ сигнал упаковки;

домены gag и pol MLV, 3'-концевой сайт сплайсинга и домен env MLV штамма 4070A;

кассету, содержащую IRES из вируса энцефаломиокардита, полинуклеотид, кодирующий полипептид с цитозиндезаминазной активностью, содержащий нуклеотиды от приблизительно 8877 до приблизительно 9353 последовательности SEQ ID NO: 19, и полипуриновый тракт;

3'-концевую последовательность LTR, содержащую домен U3-R-U5 MLV.

30. Ретровирус по п.29, где ретровирусный полинуклеотид содержит последовательность SEQ ID NO: 19 или 22.

31. Ретровирус по п.29, где ретровирусный полинуклеотид получен из вируса лейкоза мышей (MLV), вируса мышиного лейкоза Молони (MoMLV), вируса лейкоза кошек или вируса лейкоза гиббонов (GALV).

32. Ретровирус по п.31, где MoMLV представляет собой амфотропный MoMLV.

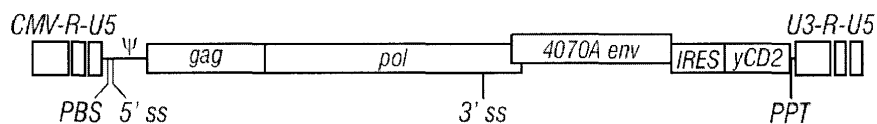
33. Применение ретровируса по п.29 для получения лекарственного средства, катализирующего превращение экзогенного 5-фторцитозина в 5-фторурацил в клетках индивида с нарушением клеточной пролиферации.

34. Применение по п.33, где ретровирусный полинуклеотид содержит последовательность SEQ ID NO: 19, 20 или 22.

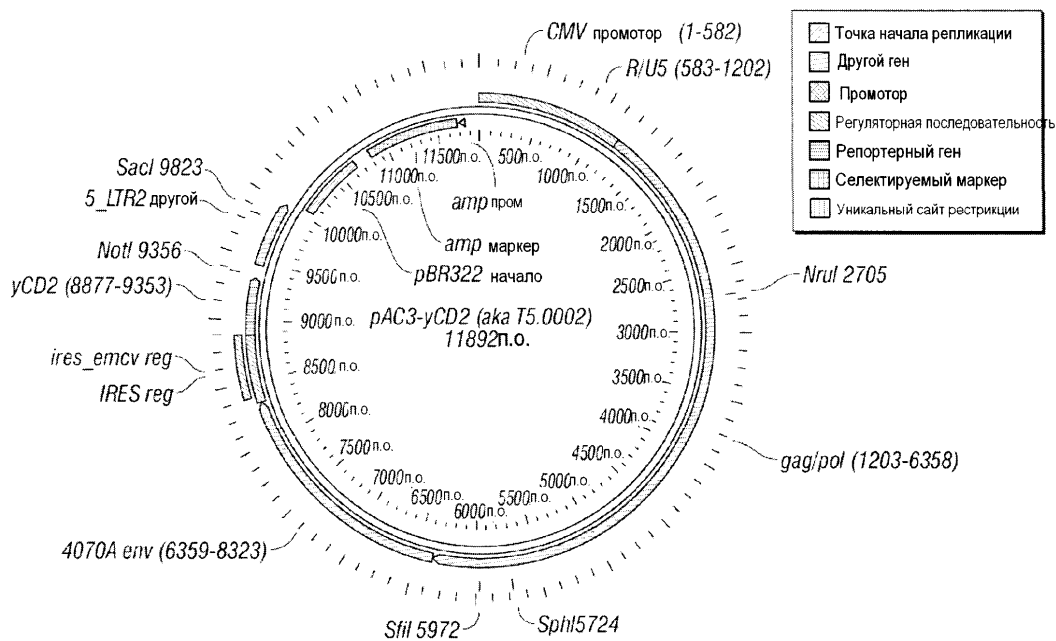
35. Применение по п.33, где нарушением клеточной пролиферации является мультиформная глиобластома.

36. Применение по п.33, где нарушение клеточной пролиферации выбрано из рака легких, рака толстой и прямой кишки, рака груди, рака предстательной железы, рака мочевыводящих путей, рака матки, рака головного мозга, рака тканей головы и шеи, рака поджелудочной железы, меланомы, рака желудка и рака яичников.

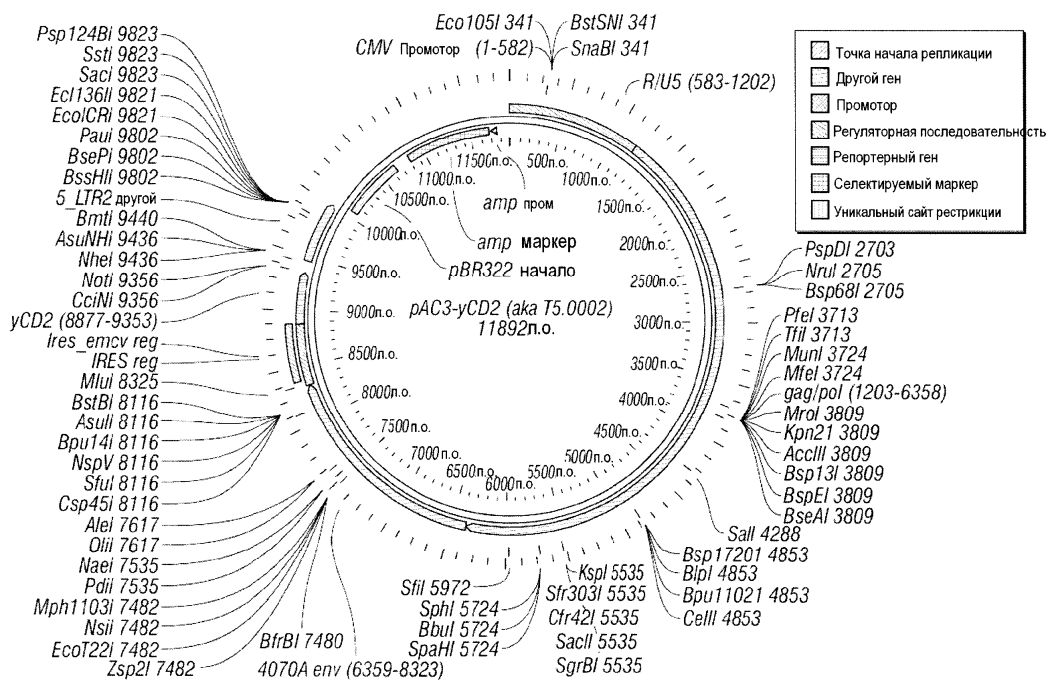
37. Применение ретровируса по п.1 для лечения нарушения клеточной пролиферации у индивида посредством каталитического превращения в его клетках экзогенного 5-фторцитозина в 5-фторурацил.



Фиг. 1А



Фиг. 1В



Фиг. 1С

CMV промотор

$$\begin{bmatrix} R \\ U5 \\ GT \end{bmatrix} = 5'ss$$

৫৫৫

Фиг. 1D-1

CCAGACCTCCCTCCTGACCTAGATGACTAGGGAGGTCAGGGTCAGGAGCCCCCTGAAACCAGGATAACCTCAAAGTC
 GGGGGGCAACCCGTCACCTTCCTGGTAGATACTGGGGCCCAACACTCCGTGCTGACCCAAATCCTGGACCCCTAAGTGATA
 AGTCGCGCTGGGTCCAAGGGGCTACTGGAGGAAAGCGGTATCGCTGGACCAGCGATCGCAAAGTACATCTAGCTACCGGTAA
 GGTCAACCCACTCTTTCTCCATGTACCAGACTGTCCCTATCCTCTGTAGGAAGAGATTTGCTGACTAAACTAAAAGCCCAA
 ATCCACTTTGAGGGATCAGGAGCCAGGTATGCGACCAATGGGCAGCCCTGCAAGTGTGACCCATAATATAGAAGATG
 AGCATCGCTACATGAGACCTCAAAGAGCCAGATGTTTCTCTAGGGTCCACATGGCTGTCTGATTTTCCTCAGGCCGTGGGC
 GGAACCCGGGGCATGGGACTGGCAGTTCGCCAAGCTCCTCTGATCATACCTCTGAAAGCAACCTCTACCCCGTGTCCATA
 AAACAATACCCCATCTCAACAAGAGCCAGACTGGGGATCAAGCCCCACATACAGAGACTGTTGCAACAGGGAATACCTGGTAC
 CCTGCCAGTCCCCCTGGAACACGCCCTGTACCCCTTAAGAAACCAGGACTAATGATTATAGGCCGTGTCAGGATCTGAG
 AGAAGTCAACAGCGGGTGAAGACATCCACCCACCGTGCCCAACCTTACAACTCTTGAGCGGGCTCCACCGCTCCACAC
 CAGTGGTACACTGTGCTTGATTTAAGGATGCGTTTCTGCTGAGACTCCACCCCAAGTCAGCCCTCTCTTCCGCTTTG
 AGTGGAGAGATCCAGAGATGGGAATCTCAGGACAATTGACCTGGACAGACTCCACAGGGTTTCAAAACAGTCCACCCCT
 GTTTGATGAGGCACTGCACAGAGACTAGCAGACTTCGGGATCCAGCACCAGACTTGATCTTGCTACAGTACGTGGATGAC
 TTACTGCTGGCCGCACTTCTGAGCTAGACTGCCAACAGGTACTCGGCCCTGTACAAACCTTAGGGAACCTCGGATAC
 GGGCCTCGGCCAAGAAAGCCCAATTTGCCAGAAACAGGTCAAGTATCTGGGGTATCTTCTAAAAGAGGTCAGAGATGGCT
 GACTGAGGCAGAAAGAGACTGTGATGGGCCAGCCTACTCCGAGACCCCTCGACAACTAAGGGAGTTCCTAGGACGGCA
 GGTCTCTGTGCGCTCTGGATCCCTGGGTTTGCAGAAATGGCAGCCCCCTTGACCTCTCACCAAAACGGGACTCTGTATA
 ATGGGGCCAGACCAACAAAAGGCTATCAAGAAATCAAGCAAGCTCTTCTAAGTGGCCAGCCCTGGGGTTGCCAGATT
 GACTAAGCCCTTTGAATCTTTGTGACGAGAAGCAGGGCTACGCCAAGGTGTCTAAGCGCAAAACGGGACCTTGGCGT
 CGGCCGGTGGCTACCTGTCCAAAAGCTAGACCCAGTAGCAGCTGGGTGGGCCCTTGCTTACGGATGTTAGCAGCCATTG
 CCGTACTGACAAAGGATGCAAGCAAGCTAACCATGGGACAGCCACTAGTCATTCTGGCCCCCATGCAGTAGAGGCACTAGT
 CAAACAACCCCGGACCGCTGGCTTTCCAACGCCCGGATGACTCACTATCAGGCCCTTGCTTTTGGACACGGACCGGCTCCAG
 TTCCGACCGGTGGTAGCCCTGAACCCGGCTACGCTGCTCCCACTGCTGAGGAAGGGTGCACACAACCTGCTTGATATCC
 TGGCCGAAGCCACGGAACCCGACCGCCTAACGGACAGCCGCTCCAGACGCCGACCAACCTGGTACACGGAAGGAAG
 CAGTCTCTTACAAGAGGGACAGCGTAAGGGGGAGCTGCGGTGACCACCGAGACCGAGGTAATCTGGGCTAAAGCCCTGCCA
 GCGGGGACATCCGCTCAGCGGCTGAATGATGCACTCAGCCAGGCCCTAAGAGTGGCAGAGGTAAAGACCTAAATGTTT
 ATACTGATAGCCGTATGCTTTTGTACTGCCATATCCATGGAGAAATATACAGAGGCGTGGGTTGCTCAGATCAGAAAG
 CAAGAGATCAAAATAAAGACGAGATCTTGGCCCTACTAAAAGCCCTCTTCTGCCAAAAGACTTAGCATTAATCCATTGT
 CCAGGACATCAAAAGGACACAGCCCGAGGCTAGAGGCAACCGGNTGGCTGACCAAGCGGCCGAAAGGCGCCATCACAG
 AGACTCCAGACCTCTACCTCTCTATAGAAATTCATCACCTTACCTCAGAACATTTTATTACACAGTGACTGATAT

[pol]

Фиг. 1D-2

AAAGGACCTAACCAAGTTGGGGCCATTTATGATAAAACAAAGAGTATTGGGTCTACCAAGGAAACCTGTGATGCCTGAC
 CAGTTTACTTTGAATATATAGACTTTCTTCATCAGCTGACTCACCTCAGCTTCTCAAAAATGAAGGCTCTCCTAGAGAGAA
 GCCACAGTCCCTACTACATGCTGAACCGGGATCGAACACTCAAAAATATCACTGAGACCTGCAAGGCTTGTGCACAGTCAA
 CGCCAGCAAGTCTGCCGTTAAACAGGGAAGTAGGGTCCGCGGGCATCGGCCCGCACTCATTGGGAGATCGATTTCACCGAG
 ATAAAGCCCGGATTTCTATGGCTATAAATATCTTCTAGTTTATATAGATACCTTTCTGGCTGGATAGAAGCTTCCCAACCA
 AGAAGAAACCGCAAGGTCGTAAACCAAGAGCTACTAGAGAGAGTCTTCCCAAGGTTGGCATGCCCTCAGGTATTGGGAAC
 TGACATGGGCTGCCCTGCTCTCCCAAGGTGAGTCAGACAGTGGCCGATCTGTGGGGATTGATTGGAAATTAACATTTGCA
 TACAGACCCCAAGGTCAGGCCAGGTAGAAGAATGAATAGAACCATCAAGGAGACTTTAACTAAATTAACGCTTGCAACTG
 GCTCTAGAGACTGGTGTCTTACTCTCCCTTAGCCCTGTACCGAGCCCGCAACACGCCGGGCCCATGGCTCACCCCATAT
 TGAGATCTTATATGGGGACCCCCGCCCTTTAAACTTCCCTGACCTGACATGACAAAGATTACTAACAGCCCTCTCTC
 CAGCTTCTACTTACAGGCTCTCTACTTAGTCCAGCAGGAAGTCTGGAGACCTCTGGCGGTAGGCTACCAAGAACAACTGGACC
 GACCGGTGTACTCTCACCCTTACCAGTGGCGACACAGTGTGGTCCGCCGACACCACTAAGAACCTAGAACCCTCGCTG
 GAAAGGACCTTACACAGTCTGCTGACCAACCCACCGCTCTCAAGTAGACGGCATCGCAGCTTGGATACAGCCCGCCAC
 GTGAAGGCTGCCGACCCCGGGGTGGACCATCTCTAGACTGACATGGCGGCTCAACGCTCTCAAAACCCCTCAAGATAA
 GATTAAACCTTGGAGGCTTAAATAGTCTAGGAGTCTCTTAGGAGTAGGGATGGCAGAGAGCCCCATCAGGCTTTAAT
 GTAACTGGAGAGTCAACAACCTGATGACTGGGCGTACCGCAATGCCACCTCCCTCTGGGAAGTGTACAAAGTGCCTCC
 CAAAATATATTTTGTATGTATGATCTGGTGGAGAGAGTGGGACCTTCAGACACAGGAACCGTATGTGGGTATGGCT
 CAAGATCCCGCAGGGAGACAGCGACCCGGACTTTGACTTTTACGTGTGCCCTGGGCATACCCTAAAGTCCGGGTGTGGG
 GAGCCAGGAGAGGCTACTGTGTAATGGGGTGTGAAACACCGGACAGGCTTACTGGAGCCACATCATCTGTGGGACC
 TAAATCTCCCTTAAAGCGGGTAAACCCCTTGGGACACGGATGCTTAAAGTGGCTGTGGCCCTGTACAGACTCTCTCAA
 AGTATCCAAATCTTCCAAAGGGCTACTCGAGGGGGAGATGCAACCTCTAGTCTTGAATTAACCTGATGCGAGAAAAAG
 GCTAACTGGGACGGGCCAAATCGTGGGAGTGGAGCTGTACCGGACAGGAACAGATCTTATACCATGTTCTCCCTGACCC
 GCGAGGCTCTTAAATGGGACCCGAGTCCCATAGGGCCCAACCCAGTATTACCGGACCAAGACTCCCTCTCTACCAAT
 AGAGATTGTACCGGCTCCACAGCCACTAGCCCTCTAATACAGTTACCCCTTCCACTACCACTACACCTCAACCTCC
 CCTACAGTCCCAAGTCTCCACAGCCACCCCGGAAGTGGAGTAGACTACTAGTCTAGTCAAGGAGCCTATCAGGCGC
 TTAAGCTCCGACCAATCCGACAAGACCCCAAGAACTGTGGCTGTGCTTAGTGTGGGACCTCTTATATCAGAGAGTAGGCT
 CGTGGGCACTTATACCAATCATTCACCGCTCCGGCCAACTGTACGGCCACTTCCCAACATAAGCTTACCTATCTGAAGTG
 ACAGGACAGGGCTATGATGGGGGAGTACCTAAACTCACAGGCTTATGTAAACCCACCAAGCGCGGCTCAGGAT
 CCTACTACCTTCAGGACCCCGCGGAACAATGTGGGCTTCAGGACTGGATTGACTCCCTGCTTGTCCACCAAGGCTGCTCAA
 TCTAACACAGATATTGTGTATTAGTGAACCTGGGCCAGAGTAATTTACCACTCCCCGATTTATATGTATGGTCAAGTT
 GAACAGGCTACCAAAATATAAAGAGAGCCAGTATCATTGACCTGGCCCTTCTACTAGGAGATTAAACATGGGAGGATTG
 CAGCTGGAAATAGGAGCGGGACACTGCCCTAATTAACCCAGCAGTTTGAGCAGCTTCATGCCGTATCCAGACAGACCT
 CAACGAAGTCAAAAGTCAATTACCAACCTAGAAAAGTCACTGACCTCGTTGTCTGAAGTAGTCTACAGAACCGCAGAGGC
 CTAGATTGTCTATTCTTAAAGAGGGAGGCTCTCTGCCAGCCCTAAAGAAAGATGTTGTTTATGCAGACACACGGGCT
 TAGTGAGAGACAGCATGGCCAAATTAAGAGAAAGGCTTAATCAGAGACAAAACATTTGAGACAGGCCAAGGATGTTTCA
 AGGGCTGTTAATAGATCCCTGCTTACCACTTAATCTCCACCATCATGGGACCTCTAATAGTACTCTTACTGATCTTA
 CTCTTTGGACCTTGCTATCTCAATCGATTGGTCCAAATTTGTAAAGACAGGATCTCAGTGTCCAGGCTCTGGTTTGA
 AGCAATATCACCAGCTAAACCCATAGAGTACGAGCATGAACCGGTACTGGCCGAAGCCGTTGGAAATAAGGCCGCTGTG
 CGTTGGCTATATCTTATTTCCACCATATTGCCGTCTTTGGCAATGTGAGGGCCCGGAAACCTGGCCCTGTCTCTTGA
 GAGCATCTTAGGGCTCTTCCCTCTCGCCAAAGGAATGCAAGGCTGTGTAATGTCTGAAGGAAGCAGTCTCTCTGGA
 GCTCTTGAAGACAAACAGCTCTGTAGCGACCTTTGACAGCAGCGGAACCCCCACCTGGCGACAGGTGCTCTGGGCTC
 AAAAGCCAGTGTATAAGATACACCTGCAAGGGCGGCACACCCAGTGCCAGCTGTGTAGTTGGATAGTTGTGGAAGAGT
 CAAATGGCTCTCTCAAGCTATTCAACAAGGGCTGAAGGATGCCAGAAAGTACCCCATTTGATGGGATCTGATCTGGGG
 CCTCGGTGCACATGCTTTACATGTGTTTAGTCGAGGTTAAAAAACGCTTAGGCCCCCGAACCCAGGGGAGCTGTTTCC
 TTTGAAAAACACGAATATAATGGGTGACCGGGGGATGGGCTCAAGTGGGATCAAAAGGGCATGGATATCGCTTACGAGGA
 GGCCTGCTGGCTACAGGAGGGCGCGGTGCTATCGCGGCTGTCTGATCAACAAAGGACGGCAGTGTGTGGGACAGG
 GGCACACATGAGGTTCCAGAAAGGCTCCGCCACCTGCACGGCGAGATCTCCACCTGAGAACTGTGGCAGGCTGGAGG
 GCAAGGTGTACAAAGACCAACCCCTGTACACACCCCTGTCCCTTGTGACATGTGTACCGCGCTATCATCATGTACGGCAT
 CCTAGTGTGTGATCGCGGAGAACGTGAACCTCAAGTCCAGGGCGAGAGTACCTGCAAAACAGGGGCCACGAGGTGGTG
 GTTGTGACGATGAGAGGTGTAAAGAGCTGATGAAGCAGTTCATCGACGAGAGGCTCAGGACTGTTTCGAGGATATCGGCG

[AG] = 3' ss

[4070A env]

[M1u1]

[EMCV IRES]

[Psi1]

yCD2

Фиг. 1D-3

```

AGTAAGCGGCGGAGATAAAATAAAGATTTTATTTAGTCTCCAGAAAAAGGGGGGAATGAAAGACCCCACTGTAGGTTTG
GCAAGCTAGCTTAAGTAACGCCATTTTGCAGGGCATGGAAAAATACATAACTGAGAATAGAGAAGTTCAGATCAAGGTCAGG
AACAGATGGAAACAGCTGAATATGGGGCAAAACAGGATATCTGTGGTAAGCAGTTCTTCCCGCGGCTCAGGGCCAAAGAACAGATGGTCCCC
GGAAACAGCTGAATATGGGGCAAAACAGGATATCTGTGGTAAGCAGTTCTTCCCGCGGCTCAGGGCCAAAGAACAGATGGTCCCC
AGATGCGGTCAGCGCTCAGCAGTTCTTAGAGAACCATCAGATGTTTCCAGGGTGCCCAAGGACCTGAAATGACCCGTGTGC
CTTATTTGAACTAACCAGTCAGTTCTGCTTCTCGCTTCTGTTTCCGCGCTTCTGCTCCCGGAGCTCAATAAAAGAGCCACAA
CCCCCTACTCGGGCGGCGAGTCTCTCGATTCAGTCAGTCCCGGGTACCCGTTATCCAAATAAACCCCTCTTGACGTTGCA
CCGACTTGTGGTCTCGCTGTTCTTGGGAGGGCTCTCTCTGAGTCATTGACTAGCCCTCAGCGGGGCTCTTTCATTACATCT
GAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGA
GCATCCAAAAATCGACGCTCAAGTCAGAGGTGGGAAACCCGACAGGACTATAAAGATACGAGGCTTTCCCGCTGGAAGC
TCCCTCGTGGCTCTCTCTGTTCCGACCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGSAGCGTGGCGCTTT
CTCAATGCTCAGCGCTGATGATCTCAGTTCGGTGTAGGTCGTTGCTTCCAGCTGGGCTGTGTGACGAAACCCCGTTCA
GCCCAGCGCTCGCGCTTATCCGTAACATATCTCTTGTGTCACCCGCTAAGACACGACTTATCGCCACTGGCAGCAGCC
ACTGTTAACAGGATAGCAGAGCGAGGTATGTAGCGGTGCTACAGAGTCTCTGAAGTGGTGGCCTAACACGGCTACACTA
GAAGGACAGATTTTGGTATCTCGGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTTTGATCCGGCAAAACA
AACCCACGCTGCTAGCGGTGGTTTTTTTGTGTCAGCAGAGATTACGGCAGAAAAAAGGATCTCAAGAAGATCTTTTG
ATCTTTTCTACGGGTCTGACGCTCAGTGGACGAAACACAGTTAAGGGATTTTGGTCAAGATATATCAAAAAGGATCT
TCACCTAGATCTTTTAAATTAATAATGAAGTTTTAAATCAATCTAAAGTATATATGATTAACCTGCTCAGAGTTACCA
ATGCTTAATCAGTGAAGCACCTATCTCAGCGATCTGTCTATTTCTGTTTCATCCATAGTTCCCTGACTCCCGCTGCTAGATA
ACTACGATACCGGAGGGCTTACCATCTGGCCCGAGTGTGCAATGATACCGGAGAGCCACGCTCACCGGCTCCAGATTTAT
CAGCAATAAACCCAGCCAGCGGAGGGCCGAGCGCAGAGTGGTCTGCACTTTATCCGCTCCATCCAGTCTATTAATTG
TTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTCCGCAACGTTTTCCTATTGCTGCAGGCATCGTGTGCA
CGCTCGTCTGTTTGGTATGGCTTCAATCAGCTCCGCTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAA
AAGCGGTAGCTCTTCCGTCCTCCGATCGTTGTCAGAAAGTAAAGTTGGCCGAGTCTTATCACTCATGGTTATGGCAGCACT
GCATAATCTCTTACTGTCTATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAAATAG
TGTATGCGGCGACCGAGTTGCTCTTCCCGGCGTCAACACGGGATAATACCGCGCCACATAGCAGAACTTTAAAGTGTCTCA
TCATTGGAACCGTTCTTCCGGGCGAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTTCGATGTAACCCACTCGTGC
ACCCAACTGATCTTCAGCATCTTTTACTTTTACCAGCGTTTCTGGGTGAGCAAAACAGGAAGGCAAAATGCCGCAAAAAG
GGAATAAGGGCGACACGGAATCTTGAATACCTTCTCTCTTTTCAATATTTATGAAGCATTATCAGGGTTATTGTCT
TCATGAGCGGATACATTTTGAATGATTTTGAATAAATAAACAAATAGGGGTTCCGCGCACATTTCCCGAAAAAGTGCCACC
TGACGCTCAAGAAACATTATTATCATGACATTAACTATAAAAAATAGCGTATCACGAGGCCCTTCTGCTTCAAGAAATTC
AT

```

Not I

[J3]

[R]

[J5]

AseI

Amp^r ORF

[ScaI]

Фиг. 1D-4

CMV Промотор (1-582)>>>

1 TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCG 60
ATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGC

61 CGTTACATAACTTACGGTAAATGGCCCGCTGGCTGACCGCCCAACGACCCCGCCATT 120
GCAATGTATTGAATGCCATTTACCGGGCGGACCGACTGGCGGGTTGCTGGGGGCGGGTAA

121 GACGTCATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCATTTGACGTCA 180
CTGCAGTTATTACTGCATACAAGGGTATCATTTGCGGTATCCCTGAAAGGTAACGTCAGT

181 ATGGGTGGAGTATTTACGGTAACTGCCCCACTTGGCAGTACATCAAGTGTATCATATGCC 240
TACCCACCTCATAAATGCCATTTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGG

241 AAGTACGCCCCCTATTGACGTCATGACGGTAAATGGCCCGCTGGCATTTATGCCCAGTA 300
TTCATGCGGGGGATAACTGCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTCAAT

Eco105I

SnaBI

BstSNI

301 CATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTAC 360
GTACTGGAATACCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGTAGCGATAATG

361 CATGGTGTATGCGGTTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTTGACTCACGGGG 420
GTACCACCTACGCCAAAACCGTCATGTAGTTACCCGCACCTATCGCCAAAACCTGAGTGGCCCC

421 ATTTCCAAGTCTCCACCCCATGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACG 480
TAAAGGTTCAAGAGGTGGGGTAACTGCAGTTACCCCTCAAACAAAACCGTGGTTTTAGTTGC

481 GGACTTTCCAAAATGTGTAACAACCTCCGCCCCATTGACGCAATGGGCGGTAGGCGTGT 540
CCTGAAAGGTTTTACAGCATTTGTTGAGGCGGGGTAACCTGCGTTTACCCGCCATCCGCACA

R область (583-650)>>>

541 ACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGGCGCCAGTCTCCGATTG 600
TGCCACCCCTCCAGATATATTGCTCTCGACCAAATCACTTGGCCGCGGTCAAGGAGGCTAAC

Фиг. 1D-4

U5 область (651-1202)>>>

601 ACTGAGTCGCCCGGGTACCCGTGTATCCAATAAACCCCTCTGCAGTTGCATCCGACTTGT 660
TGACTCAGCGGGCCCATGGGCACATAGGTTATTTGGGAGAACGTCAACGTAGGCTGAACA

661 GGTCTCGCTGTTCCCTGGGAGGGTCTCTCTGAGTGATTGACTACCCGTGAGCGGGGGTC 720
CCAGAGCGACAAGGAACCCCTCCAGAGGAGACTCACTAACTGATGGGCAGTCGCCCCAG

721 TTTCATTTGGGGGCTCGTCCGGGATCGGGAGACCCCTGCCAGGGACCCGACCCACCA 780
AAAGTAAACCCCGAGCAGGCCCTAGCCCTCTGGGGACGGGTCCCTGGTGGCTGGGTGGT

5'SS (788)

781 CCGGGAGGTAAGCTGGCCAGCAACTTATCTGTGCTGTCCGATTGTCTAGTGTCTATGAC 840
GGCCCTCCATTCGACCGGTGCTTGAATAGACACAGACAGGCTAACAGATCACAGATACTG

841 TGATTTTATGCGCCTGCGTCCGTACTAGTTAGCTAACTAGCTCTGTATCTGGCGGACCCG 900

Фиг. 1D-5

ACTAAAATACGCGGACGCGAGCCATGATCAATCGATTGATCGAGACATAGACCGCCTGGGC

901 TGGTGGAAGTACGAGTTCCGGAACACCCGGCCGCAACCCCTGGGAGACGTCCCAGGGACTT 960
ACCACCTTGACTGCTCAAGCCTTGTGGCCGGCGTTGGGACCCTCTGCAGGGTCCCTGAA

961 CGGGGGCCGTTTTTGTGGCCCGACCTGAGTCCAAAAATCCCGATCGTTTTGGACTCTTTG 1020
GCCCCCGGCAAAACACCGGGCTGGACTCAGGTTTTTAGGGCTAGCAAAACCTGAGAAAC

1021 GTGCACCCCCCTTAGAGGAGGATATGTGGTTCTGGTAGGAGACGAGAACCTAAAAACAGT 1080
CACGTGGGGGAATCTCCTCCTATACACCAAGACCATCCTCTGCTCTTGGATTTTGTCA

1081 TCCCGCCTCCGTCTGAATTTTGTCTTTCGGTTTGGGACC GAAGCCGCGCCGCGCTCTTG 1140
AGGGCGGAGGCAGACTTAAAAACGAAAGCCAAACCTGGCTTCGGCGCGGCGCGCAGAAC

1141 TCTGCTGCAGCATCGTTCTGTGTGTCTCTGTCTGACTGTGTTTCTGTATTGTCTGAGA 1200
AGACGACGTCTGTAGCAAGACACAACAGAGACAGACTGACACAAAGACATAAACAGACTCT

gag (1203, 2819) >>>

1201 ATATGGGCCAGACTGTTACCACTCCCTTAAGTTTGACCTTAGGTCACTGGAAAGATGTCTG 1260
TATACCCGGTCTGACAATGGTGAGGGAATTCAACTGGAATCCAGTGACCTTTCTACAGC

1261 AGCGGATCGCTCACAACCAAGTCGGTAGATGTCAAGAAGAGACGTTGGGTACCTTCTGCT 1320
TCGCCTAGCGAGTGTGGTCAGCCATCTACAGTTCTTCTCTGCAACCCAATGGAAGACGA

1321 CTGCAGAATGGCCAACCTTTAACGTCGGATGGCCGCGAGACGGCACCTTTAACCGAGACC 1380
GACGTCTTACCGGTTGGAATTCAGCCTACCGGCGCTCTGCCGTGGAATTTGGCTCTGG

1381 TCATCACCCAGGTTAAGATCAAGTCTTTTACCTGGCCCGCATGGACACCCAGACCAGG 1440
AGTAGTGGGTCCAATTCTAGTTCCAGAAAAGTGGACCGGGCGTACCTGTGGGTCTGGTCC

1441 TCCCTACATCGTGACCTGGGAAGCCTTGGCTTTTGACCCCCCTCCCTGGGTCAAGCCCT 1500
AGGGGATGTAGCACTGGACCCTTCGGAACCGAAAACCTGGGGGAGGGACCCAGTTCCGGA

1501 TTGTACACCCTAAGCCTCCGCCTCCTCTTCCATCCGCCCCGTCTCTCCCCCTTGAAC 1560
AACATGTGGGATTCGGAGGCGGAGGAGAAGGAGGTAGCCGGGCAGAGAGGGGGAACCTTG

1561 CTCCTCGTTCGACCCCGCCTCGATCCTCCCTTTATCCAGCCCTCACTCCTTCTTAGGCG 1620
GAGGAGCAAGCTGGGGCGGAGCTAGGAGGGAATAGGTCCGGGAGTGAGGAAGAGATCCGC

1621 CCAAACCTAAACCTCAAGTTCTTTCTGACAGTGGGGGGCCGCTCATCGACCTACTTACAG 1680
GGTTTGGATTGGAGTTCAAGAAAGACTGTCACCCCCCGGCGAGTAGCTGGATGAATGTC

1681 AAGACCCCCCGCCTTATAGGGACCCAAGACCACCCCTTCCGACAGGGACGGAAATGGTG 1740
TTCTGGGGGCGGAATATCCCTGGGTTCTGGTGGGGGAAGGCTGTCCCTGCCTTTACCAC

Фиг. 1D-6

1741 GAGAAGCGACCCCTGCCGGAGAGGCACCGGACCCCTCCCCAATGGCATCTCGCCTACGTG 1800
 CTCTTCGCTGGGGACGCCCTCTCCGTGGCTGGGGAGGGGTTACCGTAGAGCGGATGCAC
 1801 GGAGACGGGAGCCCCCTGTGGCCGACTCCACTACCTCGCAGGCATTCCCCCTCCGCGCAG 1860
 CCTCTGCCCTCGGGGGACACCGGCTGAGGTGATGGAGCGTCCGTAAGGGGGAGGCGCGTC
 1861 GAGGAAACGGACAGCTTCAATACTGGCCGTCTCCTCTTCTGACCTTTACAACCTGGAAAA 1920
 CTCCTTTGCCTGTGCAAGTTATGACCGGCAAGAGGAGAAGACTGGAAATGTTGACCTTTT
 1921 ATAATAACCCCTTCTTTTCTGAAGATCCAGGTAACTGACAGCTCTGATCGAGTCTGTTC 1980
 TATTATTGGGAAGAAAAGACTTCTAGGTCCATTTGACTGTCGAGACTAGCTCAGACAAG
 1981 TCATCACCCTATCAGCCACCTGGGACGACTGTCAGCAGCTGTTGGGGACTCTGCTGACCG 2040
 AGTAGTGGGTAGTCGGGTGGACCTGCTGACAGTCGTCGACAACCCTGAGACGACTGGC
 3121 ACTTTGAGGGATCAGGAGCCAGGTTATGGGACCAATGGGGCAGCCCTGCAAGTGTTGA 3180
 TGAACCTCCCTAGTCCCTCGGGTCCAATACCCTGGTTACCCCGTCGGGGACGTTCACT
 3181 CCCTAAATATAGAAGATGAGCATCGGCTACATGAGACCTCAAAAGAGCCAGATGTTTCTC 3240
 GGGATTTATATCTTCTACTCGTAGCCGATGTACTCTGGAGTTTTCTCGGTCTACAAAGAG
 3241 TAGGGTCCACATGGCTGTCTGATTTTCTCAGGCCTGGGCGGAAACCGGGGGCATGGGAC 3300
 ATCCCAGGTGTACCGACAGACTAAAAGGAGTCCGGACCCGCCTTTGGCCCCCGTACCCTG
 3' SS (3314)
 3301 TGGCAGTTCGCCAAGCTCCTCTGATCATACCTCTGAAAGCAACCTCTACCCCCGTGTCCA 3360
 ACGTCAAGCGGTTTCGAGGAGACTAGTATGGAGACTTTCGTTGGAGATGGGGGCACAGGT
 3361 TAAACAATACCCCATGTCAAGAAGCCAGACTGGGGATCAAGCCCCACATACAGAGAC 3420
 ATTTGTATGGGTACAGTGTCTTCGGTCTGACCCCTAGTTCGGGGTGTATGTCTCTG
 3421 TGTTGGACCAGGGAATACTGGTACCCTGCCAGTCCCCCTGGAACACGCCCCCTGCTACCCG 3480
 ACAACCTGGTCCCTTATGACCATGGGACGGTCAGGGGGACCTTGTGCGGGGACGATGGGC
 3481 TTAAGAAACCAGGGACTAATGATTATAGGCCTGTCCAGGATCTGAGAGAAGTCAACAAGC 3540
 AATTCCTTGGTCCCTGATTACTAATATCCGGACAGGTCCTAGACTCTCTTCAGTTGTTGG
 3541 GGGTGGAGACATCCACCCACCGTGCCCAACCCTTACAACCTCTTGAGCGGGCTCCAC 3600
 CCCACCTTCTGTAGGTGGGTGGCACGGGTGGGAATGTTGGAGAACTCGCCCGAGGGTG
 3601 CGTCCCACAGTGGTACACTGTGCTTGATTTAAAGGATGCCTTTTCTGCCTGAGACTCC 3660
 GCAGGGTGGTCACCATGTGACACGAACATAATTTCTACGAAAAAGACGGACTCTGAGG
 PfeI
 |
 TfiI
 |
 3661 ACCCCACAGTCAGCCTCTCTTCGCCTTTGAGTGGAGAGATCCAGAGATGGGAATCTCAG 3720
 TGGGGTGGTCAGTCGGAGAGAAGCGGAAACTCACCTCTCTAGGTCTCTACCCCTTAGAGTC

Фиг. 1D-7

MfeI
 |
 MunI
 |

3721 GACAATTGACCTGGACCAGACTCCACAGGGTTTCAAAAACAGTCCCACCCTGTTTGATG 3780
 CTGTTAACTGGACCTGGTCTGAGGTGTCCCAAAGTTTTTGTTCAGGGTGGGACAACTAC

MroI
 |
 BseAI
 |
 Bsp13I
 |
 BspEI
 |
 Kpn2I
 |
 AccIII
 |

3781 AGGCACTGCACAGAGACCTAGCAGACTTCCGGATCCAGCACCAGACTTGATCCTGCTAC 3840
 TCCGTGACGTGTCTCTGGATCGTCTGAAGGCCTAGGTCGTGGGTCTGAACTAGGACGATG

3841 AGTACGTGGATGACTTACTGCTGGCCGCCACTTCTGAGCTAGACTGCCAACAAGGTACTC 3900
 TCATGCACCTACTGAATGACGACCGGCGGTGAAGACTCGATCTGACGGTTGTTCCATGAG

3121 ACTTTGAGGGATCAGGAGCCCAGGTTATGGGACCAATGGGGCAGCCCCTGCAAGTGTTGA 3180
 TGAACCTCCCTAGTCCCTCGGGTCCAATACCCTGGTTACCCCGTCGGGGACGTTACAACT

3181 CCCTAAATATAGAAGATGAGCATCGGCTACATGAGACCTCAAAAGAGCCAGATGTTTCTC 3240
 GGGATTTATATCTTCTACTCGTAGCCGATGTACTCTGGAGTTTTCTCGGTCTACAAAGAG

3241 TAGGGTCCACATGGCTGTCTGATTTTCCTCAGGCCTGGGCGGAAACCGGGGGCATGGGAC 3300
 ATCCCAGGTGTACCGACAGACTAAAAGGAGTCCGACCCGCCCTTGGCCCCGTACCCTG

3' SS (3314)
 |

3301 TGGCAGTTCGCCAAGCTCCTCTGATCATACTCTGAAAGCAACCTCTACCCCGTGTCCA 3360
 ACCGTCAAGCGGTTCGAGGAGACTAGTATGGAGACTTTCGTTGGAGATGGGGGCACAGGT

3361 TAAAAACAATACCCCATGTCACAAGAAGCCAGACTGGGGATCAAGCCCCACATACAGAGAC 3420
 ATTTTGTATGGGGTACAGTGTCTTCGGTCTGACCCCTAGTTCGGGGTGATGTCTCTG

3421 TGTTGGACCAGGGAATACTGGTACCCTGCCAGTCCCCCTGGAACACGCCCCTGCTACCCG 3480
 ACAACCTGGTCCCTTATGACCATGGGACGGTCAGGGGGACCTTGTGCGGGGACGATGGGC

3481 TTAAGAAACCAGGGACTAATGATTATAGGCCTGTCCAGGATCTGAGAGAAGTCAACAAGC 3540
 AATTCTTTGGTCCCTGATTACTAATATCCGGACAGGTCCTAGACTCTCTTCAGTTGTTG

Фиг. 1D-8

```

3541 GGGTGGGAAGACATCCACCCACCGTGCCCAACCCTTACAACCTCTTGAGCGGGCTCCCAC 3600
      CCCACCTTCTGTAGGTGGGGTGGCACGGGTGGGAATGTTGGAGAACTCGCCCGAGGGTG

3601 CGTCCCACCAGTGGTACACTGTGCTTGATTTAAAGGATGCCTTTTCTGCCTGAGACTCC 3660
      GCAGGGTGGTCACCATGTGACACGAACTAAATTTCTACGAAAAAGACGGACTCTGAGG

                                     PfeI
                                     |
                                     TfiI
                                     |
3661 ACCCCACCAGTCAGCCTCTCTTCGCCTTTGAGTGGAGAGATCCAGAGATGGGAATCTCAG 3720
      TGGGGTGGTCAGTCGGAGAGAAGCGGAACTCACCTCTCTAGGTCTCTACCCTTAGAGTC

      MfeI
      |
      MunI
      |
3721 GACAATTGACCTGGACCAGACTCCACAGGGTTTCAAAAACAGTCCCACCCTGTTTGATG 3780
      CTGTTAACTGGACCTGGTCTGAGGGTGTCCCAAAGTTTTGTGAGGGTGGGACAAACTAC

                                     MroI
                                     |
                                     BseAI
                                     |
                                     Bsp13I
                                     |
                                     BspEI
                                     |
                                     Kpn2I
                                     |
                                     AccIII
                                     |
3781 AGGCACTGCACAGAGACCTAGCAGACTTCCGGATCCAGCACCAGACTTGATCCTGCTAC 3840
      TCCGTGACGTGTCTCTGGATCGTCTGAAGGCCTAGGTCTGGGTCTGAACTAGGACGATG

3841 AGTACGTGGATGACTTACTGCTGGCCGCCACTTCTGAGCTAGACTGCCAACAAGGTACTC 3900
      TCATGCACCTACTGAATGACGACCGGCGGTGAAGACTCGATCTGACGGTTGTTCCATGAG

```

Фиг. 1D-9

3901 GGGCCCTGTTACAAACCCCTAGGGAACCTCGGGTATCGGGCCTCGGCCAAGAAAGCCCAA 3960
 CCCGGGACAATGTTTGGGATCCCTTGGAGCCCATAGCCCCGGAGCCGGTTCTTTCGGGTTT
 3961 TTTGCCAGAAACAGGTCAAGTATCTGGGGTATCTTCTAAAAGAGGGTCAGAGATGGCTGA 4020
 AAACGGTCTTTGTCCAGTTCATAGACCCCATAGAAGATTTCTCCCACTCTCTACCGACT
 4021 CTGAGGCCAGAAAAGAGACTGTGATGGGGCAGCCTACTCCGAAGACCCCTCGACAACATA 4080
 GACTCCGGTCTTTTCTCTGACACTACCCCGTCGGATGAGGCTTCTGGGGAGCTGTTGATT
 4081 GGGAGTTCCTAGGGACGGCAGGCTTCTGTGCGCTCTGGATCCCTGGGTTTGCAGAAATGG 4140
 CCCTCAAGGATCCCTGCCGTCCGAAGACAGCGGAGACCTAGGGACCCAAACGCTTTTACC
 4141 CAGCCCCCTTGTACCCTCTCACAAAACGGGGACTCTGTTTAATTGGGGCCAGACCAAC 4200
 GTCGGGGGAACATGGGAGAGTGGTTTGGCCCTGAGACAAATTAACCCCGGCTCTGGTTG
 4201 AAAAGGCCTATCAAGAAATCAAGCAAGCTCTTCTAACTGCCCCAGCCCTGGGGTTGCCAG 4260
 TTTTCCGGATAGTTCCTTAGTTCGTTCGAGAAGATTGACGGGGTCGGGACCCCAACGGTC

 SalI
 |
 4261 ATTTGACTAAGCCCTTTGAACTCTTTGTGACGAGAGAAGCAGGGCTACGCCAAAGGTGTCC 4320
 TAAACTGATTCGGGAACTTGAGAAACAGCTGCTCTTCGTCCCGATGCGGTTTCCACAGG
 4321 TAACGCAAAAACCTGGGACCTTGGCGTCGGCCGGTGGCCTACCTGTCCAAAAGCTAGACC 4380
 ATTGCGTTTTTTGACCTGGAACCCGAGCCGGCCACCGGATGGACAGGTTTTTCGATCTGG
 4381 CAGTAGCAGCTGGGTGGCCCCCTTGCCCTACGGATGGTAGCAGCCATTGCCGTACTGACAA 4440
 GTCATCGTCGACCCACCGGGGAACGGATGCCTACCATCGTCGGTAACGGCATGACTGTT
 4441 AGGATGCAGGCAAGCTAACCATGGGACAGCCACTAGTCATTCTGGCCCCCATGCAGTAG 4500
 TCCTACGTCCGTTTCGATTGGTACCCTGTCGGTGATCAGTAAGACCGGGGGGTACGTCATC
 4501 AGGCACTAGTCAAACAACCCCCGACCGCTGGCTTTCCAACGCCCGGATGACTCACTATC 4560
 TCCGTGATCAGTTTGTGGGGGGCTGGCGACCGAAAGGTTGCGGGCCTACTGACTGATAG
 4561 AGGCCTTGCTTTTGGACACGGACCGGGTCCAGTTCGGACCGGTGGTAGCCCTGAACCCGG 4620
 TCCGGAACGAAAACCTGTGCTTGGCCAGGTCAAGCCTGGCCACCATCGGGACTTGGGCC
 4621 CTACGCTGCTCCCACTGCCTGAGGAAGGGCTGCAACACAACCTGCCTTGATATCCTGGCCG 4680
 GATGCGACGAGGGTGACGGACTCCTTCCCGACGTTGTGTGACGGAACATAGGACCGGC
 4681 AAGCCACGGAACCCGACCCGACCTAACGGACCGCCGCTCCAGACGCGGACACACCT 4740
 TTCGGGTGCCCTTGGGCTGGGCTGGATTGCCTGGTCGGCGAGGGTCTGCGGCTGGTGTGGA
 4741 GGTACACGGATGGAAGCAGTCTCTTACAAGAGGGACAGCGTAAGCGGGGAGCTGCGGTGA 4800
 CCATGTGCCTACCTTCGTGAGAGAATGTTCTCCCTGTCGCATTCCGCCCTCGACGCCACT

Фиг. 1D-10

BspI
 |
 CelIII
 |
 Bpu1102I
 |
 Bsp1720I
 |

```

4801 CCACCGAGACCGAGGTAATCTGGGCTAAAGCCCTGCCAGCCGGGACATCCGCTCAGCGGG 4860
    GGTGGCTCTGGCTCCATTAGACCCGATTTCCGGACGGTCGGCCCTGTAGGCGAGTCGCC
4861 CTGAACTGATAGCACTCACCAGGCCCTAAAGATGGCAGAAGGTAAGAAGCTAAATGTTT 4920
    GACTTGACTATCGTGAGTGGGTCCGGGATTTCTACCGTCTTCCATTCTTCGATTTACAAA
4921 ATACTGATAGCCGTTATGCTTTTGCTACTGCCCATATCCATGGAGAAATATACAGAAGGC 4980
    TATGACTATCGGCAATACGAAAACGATGACGGGTATAGGTACCTCTTTATATGTCTCCG
4981 GTGGGTTGCTCACATCAGAAAGCAAGAGATCAAAAATAAGACGAGATCTTGGCCCTAC 5040
    CACCCAACGAGTGAGTCTTCCGTTTCTCTAGTTTATTCTGCTCTAGAACCGGGATG
5041 TATAAGCCCTCTTTCTGCCAAAAGACTTATGCATAATCCATTGTCCAGGACATCAAAAGG 5100
    ATTTTCGGGAGAAAGACGGSTTTTCTGAATCGTATTAGGTAACAGGTCTGTAGTTTTCC
5101 GACACAGCGCCGAGGCTAGAGGCAACCGGATGGCTGACCAAGCGGCCCGAAAGGCAGCCA 5160
    CTGTGTCGCGGCTCCGATCTCCGTTGGCCTACCGACTGGTTCGCCGGGCTTTCCGTCGGT
5161 TCACAGAGACTCCAGACACCTCTACCTCCTCATAGAAAATTCATCACCTACACCTCAG 5220
    AGTGCTCTTGAGGTCTGTGAGATGGGAGGAGTATCTTTAAGTAGTGGGATGTGGAGTC
5221 AACATTTTCATTACACAGTGACTGATATAAAGGACCTAACCAAGTTGGGGGCCATTTATG 5280
    TTGTAAAAGTAATGTGTCAGTACTATATTTCCTGGATTGGTTCAACCCCGGTAAATAC
5281 ATAAACAAAGAAGTATTGGGTCTACCAAGGAAAACCTGTGATGCCTGACCAAGTTTACTT 5340
    TATTTTGTTCCTTATAACCCAGATGGTTCCTTTTGGACACTACGGACTGGTCAAAATGAA
5341 TTGAATTATTAGACTTTCTTCATCAGCTGACTCACCTCAGCTTCTCAAAAATGAAGGCTC 5400
    AACTTAATAATCTGAAAGAAGTAGTCGACTGAGTGGAGTCGAAGAGTTTTTACTTCCGAG
5401 TCCTAGAGAGAAGCCACAGTCCCTACTACATGCTGAACCGGGATCGAACACTCAAAAATA 5460
    AGGATCTCTCTTCGGTGTGAGGGATGATGTACGACTTGGCCCTAGCTTGTGAGTTTTTAT
  
```

Фиг. 1D-11

5461 TCACTGAGACCTGCAAAGCTTGTGCACAAGTCAACGCCAGCAAGTCTGCCGTTAAACAGG 5520
AGTGACTCTGGACGTTTCGAACACGTGTTCAAGTTGCGGTCGTTGACACGGCAATTTGTCC

SacII
|
SgrBI
|
Cfr42I
|
Sfr303I
|
KspI
|

5521 GAACTAGGGTCCGCGGGCATCGGCCCGGCACTCATTGGGAGATCGATTTACCGAGATAA 5580
CTTGATCCCAGGCGCCCGTAGCCGGCCGTGAGTAACCCCTCTAGCTAAAGTGGCTCTATT

5581 AGCCCCGATTGTATGGCTATAAATATCTTCTAGTTTTTATAGATACCTTTTCTGGCTGGA 5640
TCGGGCCTAACATACCGATATTTATAGAAGATCAAAAATATCTATGGAAAAGACCGACCT

5641 TAGAAGCCTTCCCAACCAAGAAAGAAACCGCCAAGGTGTAACCAAGAAAGCTACTAGAGG 5700
ATCTTCGGAAAGGTTGGTTCTTTCTTTGGCGGTCCAGCATTGGTTCTTCGATGATCTCC

PaeI
|
BbuI
|
SphI
|
SphI
|

5701 AGATCTTCCCCAGGTTCCGGCATGCCCTCAGGTATTGGGAAGTACAATGGGCCCTGCCTTCG 5760
TCTAGAAGGGGTCCAAGCCGTACGGAGTCCATAACCCCTTGACTGTTACCCGGACGGAAGC

6901 CTTCCAAGGGGCTACTCGAGGGGGCAGATGCAACCCCTCTAGTCCTAGAATTCAGTGATGC 6960
GAAGGTTCCCGATGAGCTCCCCGTCTACGTTGGGAGATCAGGATCTTAAGTGACTACC

6961 AGGAAAAAAGGCTAACTGGGACGGGCCCAATCGTGGGGACTGAGACTGTACCGGACAGG 7020
TCCTTTTTTCCGATTGACCCCTGCCCCGGTTTAGCACCCCTGACTCTGACATGGCCTGTCC

7021 AACAGATCCTATTACCATGTTCTCCCTGACCCGGCAGGTCCTTAATGTGGGACCCGAGT 7080
TTGTCTAGGATAATGGTACAAGAGGGACTGGGCCGTCCAGGAATTACACCCTGGGGCTCA

7081 CCCCATAGGGCCCAACCCAGTATTACCCGACCAAGACTCCCTTCCTCAACCAATAGAGAT 7140
GGGGTATCCCGGGTTGGGTCATAATGGGCTGGTTTCTGAGGGAAGGAGTGGTTATCTCTA

7141 TGTACCGGCTCCACAGCCACCTAGCCCCCTCAATACCAGTTACCCCCCTTCCACTACCAG 7200
ACATGGCCGAGGTGTCGGTGGATCGGGGAGTTATGGTCAATGGGGGAAGGTGATGCTC

7201 TACACCCTCAACCTCCCTACAAGTCCAAGTGTCCACAGCCACCCCAAGAACTGGAGA 7260
ATGTGGGAGTTGGAGGGGATGTTCAAGTTCACAGGGTGTGCGTGGGGTTCCTTGACCTCT

7261 TAGACTACTAGCTCTAGTCAAAGGAGCCTATCAGGCGCTTAACCTCACCAATCCCGACAA 7320
ATCTGATGATCGAGATCAGTTTCCTCGGATAGTCCGCGAATTGGAGTGGTTAGGGCTGTT

Фиг. 1D-12

```

7321  GACCCAAGAATGTTGGCTGTGCTTAGTGTCGGGACCTCCTTATTACGAAGGAGTAGCGGT 7380
      CTGGGTTCTTACAACCGACACGAATCACAGCCCTGGAGGAATAATGCTTCCTCATCGCCA

7381  CGTGGGCACTTATACCAATCATTCACCGCTCCGGCCAACTGTACGGCCACTTCCCAACA 7440
      GCACCCGTGAATATGGTTAGTAAGGTGGCGAGGCCGGTTGACATGCCGGTGAAGGGTTGT

                                     EcoT22I
                                     |
                                     NsiI
                                     |
                                     Mph1103I
                                     |
                                     Zsp2I
                                     |
                                     BfrBI
                                     | |
7441  TAAGCTTACCCTATCTGAAGTGACAGGACAGGGCCTATGCATGGGGGCAGTACCTAAAAC 7500
      ATTGAATGGGATAGACTTCACTGTCCTGTCCCGGATACGTACCCCGTCATGGATTTTG

                                     NaeI
                                     |
                                     PdiI
                                     |
                                     MroNI
                                     | |
                                     NgoMIV
                                     | |
7501  TCACCAGGCCTTATGTAACACCACCCAAAGCGCCGGCTCAGGATCCTACTACCTTGCAGC 7560
      AGTGGTCCGGAATACATTGTGGTGGGTTTCGCGGCCGAGTCCTAGGATGATGGAACGTCG

                                     AleI
                                     |
                                     OliI
                                     |
7561  ACCCGCCGGAACAATGTGGGCTTGCAGCACTGGATTGACTCCCTGCTTGTCCACCACGGT 7620
      TGGGCGGCCTTGTTACACCCGAACGTCGTGACCTAACTGAGGGACGAACAGGTGGTGCCA

7621  GCTCAATCTAACCACAGATTATTGTGTATTAGTTGAACTCTGGCCAGAGTAATTTACCA 7680
      CGAGTTAGATTGGTGTCTAATAACACATAATCAACTTGAGACCGGGTCTCATTAATGGT

```

Фиг. 1D-13

6901 CTTCCAAGGGGCTACTCGAGGGGGCAGATGCAACCCCTCTAGTCCTAGAATTCAGTGATGC 6960
 GAAGGTTCCCGGATGAGCTCCCCGCTACGTTGGGAGATCAGGATCTTAAGTGACTACG
 6961 AGGAAAAAAGGCTAACTGGGACGGGCCCCAAATCGTGGGGACTGAGACTGTACCGGACAGG 7020
 TCCTTTTTTCCGATTGACCTGCCCCGGTTTAGCACCCCTGACTCTGACATGGCCTGTCC
 7021 AACAGATCCTATTACCATGTTCTCCCTGACCCGGCAGGTCCTTAATGTGGGACCCCGAGT 7080
 TTGTCTAGGATAATGGTACAAGAGGGACTGGGCCGTCCAGGAATTACACCCTGGGGCTCA
 7081 CCCCATAGGGCCCCAACCCAGTATTACCCGACCAAGACTCCCTTCCTCACCAATAGAGAT 7140
 GGGGTATCCCGGGTTGGGTGATAATGGGCTGGTTTCTGAGGGAAGGAGTGGTTATCTCTA
 7141 TGTACCGGCTCCACAGCCACCTAGCCCCCTCAATACCAGTTACCCCCCTTCCACTACCAG 7200
 ACATGGCCGAGGTGTCGGTGGATCGGGGAGTTATGGTCAATGGGGGAAGGTGATGGTC
 7201 TACACCCTCAACCTCCCTACAAGTCCAAGTGTCCACAGCCACCCCGAGGAAGTGGAGA 7260
 ATGTGGGAGTTGGAGGGGATGTTCAAGTTCACAGGGTGTGGTGGGGGTCTTGACCTCT
 7261 TAGACTACTAGCTCTAGTCAAAGGAGCCTATCAGGCGCTTAACCTCACCAATCCCGACAA 7320
 ATCTGATGATCGAGATCAGTTTCCTCGGATAGTCCGCGAATTGGAGTGGTTAGGGCTGTT
 7321 GACCCAGAATGTTGGCTGTGCTTAGTGTGGGACCTCCTTATTACGAAGGAGTAGCGGT 7380
 CTGGGTTCCTACAACCGACACGAATCACAGCCCTGGAGGAATAATGCTTCCTCATCGCCA
 7381 CGTGGGCACTTATACCAATCATTCACCGCTCCGGCCAACTGTACGGCCACTTCCCAACA 7440
 GCACCCGTGAATATGGTTAGTAAGGTGGCGAGGCCGTTGACATGCCGGTGAAGGGTTGT

EcoT22I

|

NsiI

|

Mph1103I

|

Zsp2I

|

BfrBI

| |

7441 TAAGCTTACCCTATCTGAAGTGACAGGACAGGGCCTATGCATGGGGGACAGTACCTAAAAC 7500
 ATTCGAATGGGATAGACTTCACTGTCTGTCCCGGATACGTACCCCGTCATGGATTTTG

NaeI

|

PdiI

|

MroNI

| |

NgoMIV

| |

7501 TCACCAGGCCTTATGTAACACCAACCAAGCGCCGGCTCAGGATCCTACTACCTTGCAGC 7560
 AGTGGTCCGGAATACATTGTGGTGGGTTTCGCGGCCGAGTCCTAGGATGATGGAACGTCTG

Фиг. 1D-14

AleI
 |
 OliI
 |

7561 ACCCGCCGGAACAATGTGGGCTTGCAGCACTGGATTGACTCCCTGCTTGTCCACCACGGT 7620
 TGGCGGCGCCTTGTACACCGGAACGTCGTGACCTAACTGAGGGACGAACAGGTGGTGCCA
 7621 GCTCAATCTAACCACAGATTATTGTGTATTAGTTGAACTCTGGCCCAGAGTAATTTACCA 7680
 CGAGTTAGATTGGTGTCTAATAACACATAATCAACTTGAGACCGGGTCTCATTAAATGGT
 7681 CTCCCCCGATTATATGTATGGTCAGCTTGAACAGCGTACCAAATATAAAAGAGAGCCAGT 7740
 GAGGGGGCTAATATACATAACCAGTCGAACCTTGTCGCATGTTTATATTTCTCTCGGTCA
 7741 ATCATTGACCCTGGGCCCTTCTACTAGGAGGATTAACCATGGGAGGGATTGCAGCTGGAAT 7800
 TAGTAACTGGGACCGGGAAGATGATCCTCCTAATGGTACCCTCCCTAACGTCGACCTTA
 7801 AGGGACGGGGACCACTGCCTTAATTA AAAACCCAGCAGTTTGAGCAGCTTCATGCCGCTAT 7860
 TCCCTGCCCTGGTGACGGAATTAATTTTGGGTCGTCAAACCTCGTCGAAGTACGGCGATA
 7861 CCAGACAGACCTCAACGAAGTCGAAAAGTCAATTACCAACCTAGAAAAGTCACTGACCTC 7920
 GGTCTGTCTGGAGTTGCTTCAGCTTTTCAGTTAATGGTTGGATCTTTTCAGTGACTGGAG
 7921 GTTGTCTGAAGTAGTCCTACAGAACCGCAGAGGCTAGATTTGCTATTCTTAAAGGAGGG 7980
 CAACAGACTTCATCAGGATGTCTTGGCGTCTCCGGATCTAAACGATAAGGATTTCTCTCC
 7981 AGGTCTCTGCGCAGCCCTAAAAGAAGAATGTTGTTTTATGCAGACCACACGGGGCTAGT 8040
 TCCAGAGACGCGTCGGGATTTTCTTCTTACAACAAAAATACGTCTGGTGTGCCCCGATCA
 8041 GAGAGACAGCATGGCCAAATTAAGAGAAAGGCTTAATCAGAGACAAAAACTATTTGAGAC 8100
 CTCTCTGTCTGACCGGTTTAATTCTCTTTCCGAATTAGTCTCTGTTTTGATAAACTCTG

NspV
 |
 BstBI
 |
 Bsp119I
 |
 AsuII
 |
 Csp45I
 |
 SfuI
 |
 Bpu14I
 |
 BspT104I
 |

8101 AGGCCAAGGATGGTTCGAAGGGCTGTTTAAATAGATCCCCCTGGTTTACCACCTTAATCTC 8160
 TCCGGTTCCCTACCAAGCTTCCCGACAAATATCTAGGGGGACCAAATGGTGGAATTAGAG

Фиг. 1D-15

```

8161 CACCATCATGGGACCTCTAATAGTACTCTTACTGATCTTACTCTTTGGACCTTGCATTCT 8220
      GTGGTAGTACCCTGGAGATTATCATGAGAATGACTAGAAATGAGAAACCTGGAACGTAAGA

8221 CAATCGATTGGTCCAATTTGTAAAGACAGGATCTCAGTGGTCCAGGCTCTGGTTTTGAC 8280
      GTTAGCTAACCAGGTTAAACAATTTCTGTCTAGAGTACCAGGTCCGAGACCAAACTG
                                     IRES reg(8327,8876)>>>
                                     |
                                     MluI (8325)
                                     | |
8281 TCAGCAATATCACCAGCTAAAACCCATAGAGTACGAGCCATGAACGCGTTACTGGCCGAA 8340
      AGTCGTTATAGTGGTCGATTTTGGGTATCTCATGCTCGGTACTTGCGCAATGACCGGCTT

                                     ires_emcv reg(8378,8876)>>>
                                     |
8341 GCCGCTTGAATAAGGCCGGTGTGCGTTTGTCTATATGTTATTTCCACCATATTGCCGT 8400
      CGGCGAACCTTATTCGGCCACACGCAACAGATATACAATAAAAGGTGGTATAACGGCA

8401 CTTTTGGCAATGTGAGGGCCCGGAAACCTGGCCCTGTCTTCTTGACGAGCATTCCTAGGG 8460
      GAAAACCGTTACACTCCCGGGCCTTTGGACCGGGACAGAAGAACTGCTCGTAAGGATCCC
8461 GTCTTTCCCTCTCGCCAAAGGAATGCAAGGTCTCTTGAATGTCGTGAAGGAAGCAGTTC 8520
      CAGAAAGGGGAGAGCGGTTTCCTTACGTTCCAGACAACCTACAGCACTTCCTTCGTCAAG

8521 CTCTGGAAGCTTCTTGAAGACAAACAACGTCTGTAGCGACCCCTTTCAGGCAGCGGAACC 8580
      GAGACCTTCGAAGAACTTCTGTTTGTGACACATCGCTGGGAAACGTCCGTCGCCTTGG

8581 CCCACCTGGCGACAGGTGCCTCTGCGGCCAAAAGCCACGTGTATAAGATACACCTGCAA 8640
      GGGGTGGACCGCTGTCCACGGAGACGCCGGTTTTCGGTGCACATATTCTATGTGGACGTT

8641 AGGCGGCACAAACCCAGTGCCACGTTGTGAGTTGGATAGTTGTGGAAGAGTCAAATGGC 8700
      TCCGCCGTGTTGGGGTCACGGTGCAACACTCAACCTATCAACACCTTTCTCAGTTTACCG

8701 TCTCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCCAGAAGGTACCCCATTTGTATGG 8760
      AGAGGAGTTCCGATAAGTTGTTCCCGACTTCCTACGGGTCTTCCATGGGGTAACATACC

8761 GATCTGATCTGGGGCCTCGGTGCACATGCTTTACATGTGTTTAGTCGAGGTTAAAAAAC 8820
      CTAGACTAGACCCCGAGCCACGTGTACGAAATGTACACAAATCAGCTCCAATTTTTTTG

```

Фиг. 1D-16

yCD2 (8877, 9353)>>>
 |
 Psi I (8874)
 |

```

8821 GTCTAGCCCCCCCCGAACCAAGGGACCTGGTTCCTTTGAAAAACACGATTATATATGG 8880
      CACATCCGGGGGGCTTGGTCCCCCTGCACCAAAAGGAAACTTTTGTGCTAATATTTACC
8881 TGACCGCGCGGCATGGCCTCCAAGTGGGATCAAAGGGCATGGATATCGCTTACGAGGAGG 8940
      ACTGGCGCGCGTACCGGAGGTTCAACCTAGTTTCCCGTACCTATAGCGAATGCTCCTCC
8941 CCCTGCTGGGCTACAAGGAGGGCGGCTCCCTATCGCGGCTGCTCTCATCAACAACAAGG 9000
      GGGACGACCCGATGTTCTCCCGCGCAGGATAGCCGCGACAGACTAGTTGTTGTTCC
9001 ACCGCAGCTGTGCTGGGACGGGGCCACAACATGAGGTTCCAGAGGGCTCCGCCACCCCTCC 9060
      TCCCGTCACACGACCCGTCCCGGTGTTGTACTCCAAGGCTCTCCCGAGGCGGTGGGACC
9061 ACCGGCGAGATCTCCACCCCTGGAGAACTGTGGCAGGCTGGAGGGCAAGGCTCTACAAGGACA 9120
      TGGCGCTCTAGAGGTGGGACCTCTTGACACCGTCCGACCTCCCGTTCCACATGTTTCTCT
9121 CCACCCCTGTACACCAACCCCTGTCCCTTGTGACATGTGTACCGCGGCTATCATATGTAAC 9180
      GGTGGGACATGTGCTGGGACAGGGGAACACTGTACACATGGCGCGGATAGTAGTACATGC
9181 GCATCCCTAGGTGTGTGATCGGCGAGAACGTGAAGTTCAAGTCCAAGGGCGAGAACTACC 9240
      CGTAGGGATCCACACACTAGCCGCTCTTGCACTTGAAGTTGAGGTTCCCGCTCTTCATGG
9241 TGCAAACCAAGGGCCACGAGGTGTGTTGTTGACCATGAGAGGTTGAAGAAGCTCATGA 9300
      ACCTTTGGTCCCGGTGCTCCACCAACAACCTGCTACTCTCCACATTCTTCGACTACT

```

Not I (9356)
 |
 CciNI
 |

```

9301 AGCAGTTTCATCGACGAGAGGCTTCAGGACTGGTTCCAGGATATCGGCGAGTAAGCGGCGG 9360
      TCGTCAAGTAGCTGCTCTCCGGAGTCTGACCAAGCTCCTATAGCCGCTCATTCGCGGGC

```

УЗ область (9405, 9854)>>>
 |

```

9361 CAGATAAAATAAAAGATTTTATTTAGTCTCCAGAAAAAGGGGGGAATGAAAAGACCCCAAC 9420
      GTCTATTTTATTTCTAAAATAAATCAGAGTCTTTTCCCCCTTACTTTCTGGGCTGG

```

Фиг. 1D-17

		Bmt I		
	Nhe I			
	AsuNHI		5' LTR2	другой (9447, 9998)>>>>
9421	TGTAGGTTTGGCAAGCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGAAAAATACATAA			9480
	ACATCCAAACC GTTCGATCGAATT CATTGCGGTAAAACGTTCCGTACCTTTTTATGTATT			
9481	CTGAGAATAGAGAAGTTCAGATCAAGGTCAGGAACAGATGGAACAGCTGAATATGGGCCA			9540
	GACTCTTATCTCTTCAAGTCTAGTTCAGTCCTTGCTCTACCTTGTCGACTTATACCCGGT			
9541	AACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGAACA			9600
	TTGTCTATAGACACCATTCTGTC AAGGACGGGGCCGAGTCCC GGTTCTTGCTACCTTGT			
9601	GCTGAATATGGGCCAAACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCA			9660
	CGACTTATACCCGGTTTGTCTATAGACACCATTCTGTC AAGGACGGGGCCGAGTCCC GGTT			
9661	AGAACAGATGGTCCCCAGATGCGGTCCAGCCCTCAGCAGTTTCTAGAGAACCATCAGATG			9720
	TCTTGCTTACCAGGGGTCTACGCCAGGTCTGGGAGTCTGTC AAAGATCTCTTGCTAGTCTAC			
9721	TTTCCAGGGTGCCCCAAGGACCTGAAATGACCCCTGTGCCTTATTTGAACTAACCAATCAG			9780
	AAAGGTCCCACGGGGTTCCTGGACTTTACTGGGACACGGAATAAACTTGATTGGTTAGTC			
			SacI	
			SstI	
	PauI		Psp124BI	
	BsePI		EcoICRI	
	BssHII		Ecl136II	
9781	TTGCTTCTCGCTTCTGTTTCGCGCGCTTCTGCTCCCCGAGCTCAATAAAAGAGCCACAA			9840
	AAGCGAAGAGCGAAGACAAGCGCGCGAAGACGAGGGGCTCGAGTTATTTCTCGGGTGT			
		R область (9855, 9921)>>>>		
9841	CCCTCACTCGGGGCGCCAGCTCCTCCGATTGACTGAGTCGCCCCGGGTACCCGTGTATCCA			9900
	GGGGAGTGAGCCCCGCGGT CAGGAGGCTAACTGACTCAGCGGGCCCATGGGCACATAGGT			
		U5 область (9922, 9998)>>>>		
9901	ATAAACCTCTTG CAGTTGCATCCGACTTGTGGTCTCGCTGTTCTCTGGGAGGSTCTCCT			9960
	TATTTGGGAGAACGTCAACGTAGGCTGAACACCAGAGCGAC AAGGAACCTCC CAGAGGA			
9961	CTGAGTGATTGACTACCCGTCAGCGGGGGTCTTTTCATTACATGTGAGCAAAAGGCCAGCA			10020
	GACTCACTAACTGATGGGCAGTCGCCCCAGAAAGTAATGTACACTCGTTTTCCGGTCTG			

ФИГ. 1D-18

pBR322 начало (10045,10664)<<<
|

```

10021 AAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCC 10080
      TTTCCGGTCCTTGGCATTTTTCCGGCGCAACGACCGCAAAAAGGTATCCGAGGCGGGGGG

10081 TGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATA 10140
      ACTGCTCGTAGTGTTTTAGCTGCGAGTTCAGTCTCCACCGCTTTGGGCTGTCCTGATAT

10141 AAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCCTCTCCTGTTCCGACCCCTGCC 10200
      TTCTATGGTCCGCAAAAGGGGGACCTTCGAGGGAGCACGCGAGAGGACAAGGCTGGGACGG

10201 GCTTACCGGATACCTGTCCGCTTTCTCCCTCGGGGAAGCGTGGCGCTTTCTCAATGCTC 10260
      CGAATGGCCTATGGACAGGCGGAAAGAGGGAAGCCCTTCGCACCGCGAAAGAGTTACGAG

10261 ACGCTGTAGGTATCTCAGTTCGGGTGTAGGTGCTTCGCTCCAAGCTGGGCTGTGTGCACGA 10320
      TGCACATCCATAGATCAAGCCACATCCAGCAAGCGAGGTTGACCCGACACACGTGCT

10321 ACCCCCCGTTTCCGCGGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAAGTCCACCC 10380
      TGGGGGGCAAGTGGGCTGGCGACGCGAATAGGCCATTGATAGCAGAACTCAGGTTGGG

10381 GGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAG 10440
      CCATTCTGTGCTGAATAGCGGTGACCGTCGTCGGTGACCATGTCCCTAATCGTCTCGCTC

10441 GTATGTAGCGGCTGTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAG 10500
      CATAATCCGCCACGATGTCTCAAGAACTTCACCACCGGATTGATGCCGATGTGATCTTC

10501 GACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAG 10560
      CTGTCAATAAACCATAGACGCGAGACGACTTCGGTCAATGGAAGCCTTTTCTCAACCATC

10561 CTCTTGATCCGGCAAACAAACCCGCTGGTAGCGGTGGTTTTTTTGTGTTGCAAGCAGCA 10620
      GAGAACTAGGCCGTTTGTGTTGGTGGCGACCATCGCCACCAAAAAACAAACGTTCTGCTGT

10621 GATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGA 10680
      CTAATGCGCGTCTTTTTTTCCTAGAGTTCTTCTAGGAACTAGAAAAGATGCCCCAGACT

10681 CGCTCAGTGGAAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGAT 10740
      GCGAGTCACCTTGCTTTTGAAGTCAATTCCCTAAAACAGTACTCTAATAGTTTTTCTTA

10741 CTTACCTAGATCCTTTTAAATTAAAAATGAAGTTTAAATCAATCTAAAGTATATATGA 10800
      GAAGTGGATCTAGGAAATTTAATTTTTTACTTCAAAATTTAGTTAGATTTTCATATATACT

```

Фиг. 1D-19

amp маркер (10819,11679)<<<
|

10801 GTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTG 10860
CATTTGAACCAGACTGTCAATGGTTACGAATTAGTCACTCCGTGGATAGAGTCGCTAGAC

10861 TCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCGTCGTGTAGATAACTACGATACGGGA 10920
AGATAAAGCAAGTAGGTATCAACGGACTGAGGGGCAGCACATCTATTGATGCTATGCCCT

10921 GGGCTTACCATCTGGCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCC 10980
CCCGAATGGTAGACCGGGTCACGACGTTACTATGGCGCTCTGGGTGCGAGTGGCCGAGG

10981 AGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCTGCAAC 11040
TCTAAATAGTCGTTATTTGGTCGGTCGGCCTTCCCGGCTCGCGTCTTACCAGGACGTTG

11041 TTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCC 11100
AATAGGCGGAGGTAGGTCAGATAATTAACAACGGCCCTTCGATCTCATTTCATCAAGCGG

11101 AGTTAATAGTTTGGCGAACGTTGTTGCCATTGCTGCAGGCATCGTGGTGTACGCTCGTC 11160
TCAATTATCAAACGCGTTGCAACAACGGTAACGACGTCGCTAGCACCACAGTGCAGAGCAG

11161 GTTTGGTATGGCTTCATTGAGTCCGGTTCCTCAACGATCAAGGCGAGTTACATGATCCCC 11220
CAAAACATACCGAAGTAAGTCGAGGCCAAGGGTTGCTAGTTCGGCTCAATGTAAGGGG

11221 CATGTTGTGCAAAAAAGCGGTTAGTCTCCTTCGGTCTCCGATCGTTGTCAGAAAGTAAGTT 11280
GTACAACACGTTTTTTCGCCAATCGAGGAAGCCAGGAGGCTAGCAACAGTCTTCATTCAA

11281 GGCCGCGAGTGTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTATGCC 11340
CCGGCGTCACAATAGTGAGTACCAATACCGTCGTGACGTATTAAGAGAATGACAGTACGG

11341 ATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTG 11400
TAGGCATTCTACGAAAAGACACTGACCACTCATGAGTTGGTTCAGTAAGACTCTTATCAC

11401 TATGCGGCGACCGAGTTGCTCTTGGCCGGCGTCAACACGGGATAATACCGCGCCACATAG 11460
ATACGCGCGTGGCTCAACGAGAACGGGCGCAGTTGTGCCCTATTATGGCGCGGTGTATC

11461 CAGAACTTTAAAAGTGCTCATCATTTGGAACCGTTCTTCGGGGCGAAAACCTCAAGGAT 11520
GTCTTGAAATTTTCACGAGTAGTAACCTTTTGCAAGAAGCCCCGCTTTTGAGAGTTCCTA

11521 CTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGC 11580
GAATGGCGACAACCTCTAGGTCAAGCTACATTGGGTGAGCACGTGGGTTGACTAGAAGTCG

11581 ATCTTTTACTTTTACCAGCGTTTCTGGGTGAGCAAAACAGGAAGGCAAAATGCCGCAAA 11640
TAGAAAATGAAAGTGGTCGCAAGACCCACTCGTTTTTGTCTTCCGTTTTACGGCGTTT

11641 AAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTCAATATTA 11700
TTTCCCTTATTCCCGCTGTGCCTTTACAACCTATGAGTATGAGAAGGAAAAAGTTATAAT

Фиг. 1D-20

amp промотор (11721,11749)<<<
|

11701 TTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAA 11760
AACTTCGTAAATAGTCCCAATAACAGAGTACTCGCCTATGTATAAACTTACATAAATCTT

11761 AAATAAACAAATAGGGGTTCCGCGCACATTTCCCGGAAAAGTGCCACCTGACGTCTAAGA 11820
TTTATTTGTTTATCCCAAGGCGCGTGTAAAGGGGCTTTTACGGTGGAAGTGCAGATTCT

11821 AACCATTTATTATCATGACATTAACCTATAAAATAGGCGTATCACGAGGCCCTTTCGTCT 11880
TTGGTAATAATAGTACTGTAATTGGATATTTTATCCGCATAGTGCTCCGGGAAAGCAGA

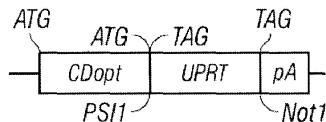
11881 TCAAGAATTCAT 11892
AGTTCTTAAGTA

Фиг. 1D-21

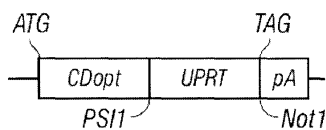
1. Отдельный синтез CDopt и UPRT генов с сайтами PSI1/Not1



2. Расщепление PSI1 и проведение лигирования слепых концов



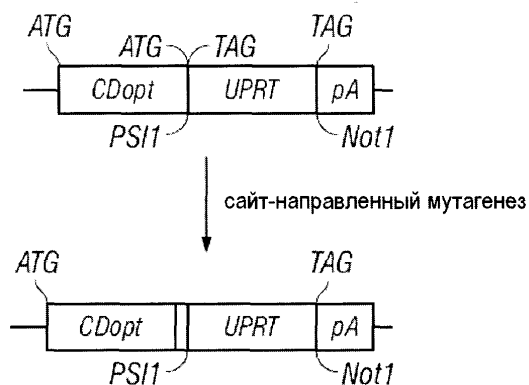
3. Делеция СТОП кодона гена CDopt и СТАРТ кодона UPRT путем сайт-направленного мутагенеза



4. Размер гена CDopt – UPRT: 1,3 тыс. осн.

Фиг. 2А

1. Синтез CDopt–ЛИНКЕР-UPRT: встраивание линкера сайт-направленным мутагенезом



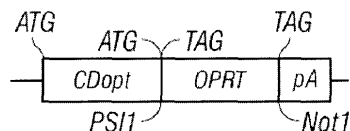
2. Размер гена CDopt–ЛИНКЕР-UPRT: 1,36 тыс. осн.

Фиг. 2В

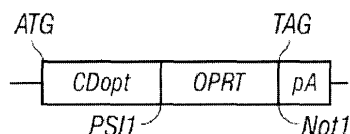
1. Отдельный синтез гена CDopt и домена OPRT с сайтами PSI1/Not1



2. Расщепление PSI1 и проведение лигирования слепых концов



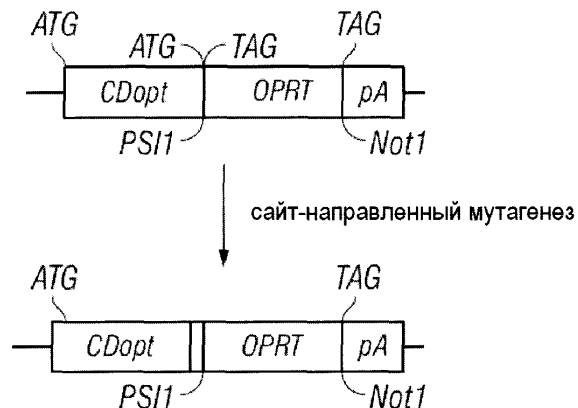
3. Делеция СТОП кодона гена CDopt и СТАРТ кодона OPRT путем сайт-направленного мутагенеза



4. Размер гена CDopt – OPRT: 1,3 тыс. осн.

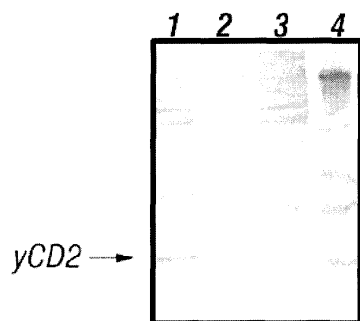
Фиг. 2С

1. Синтез CDopt–ЛИНКЕР-OPRT: встраивание линкера сайт-направленным мутагенезом



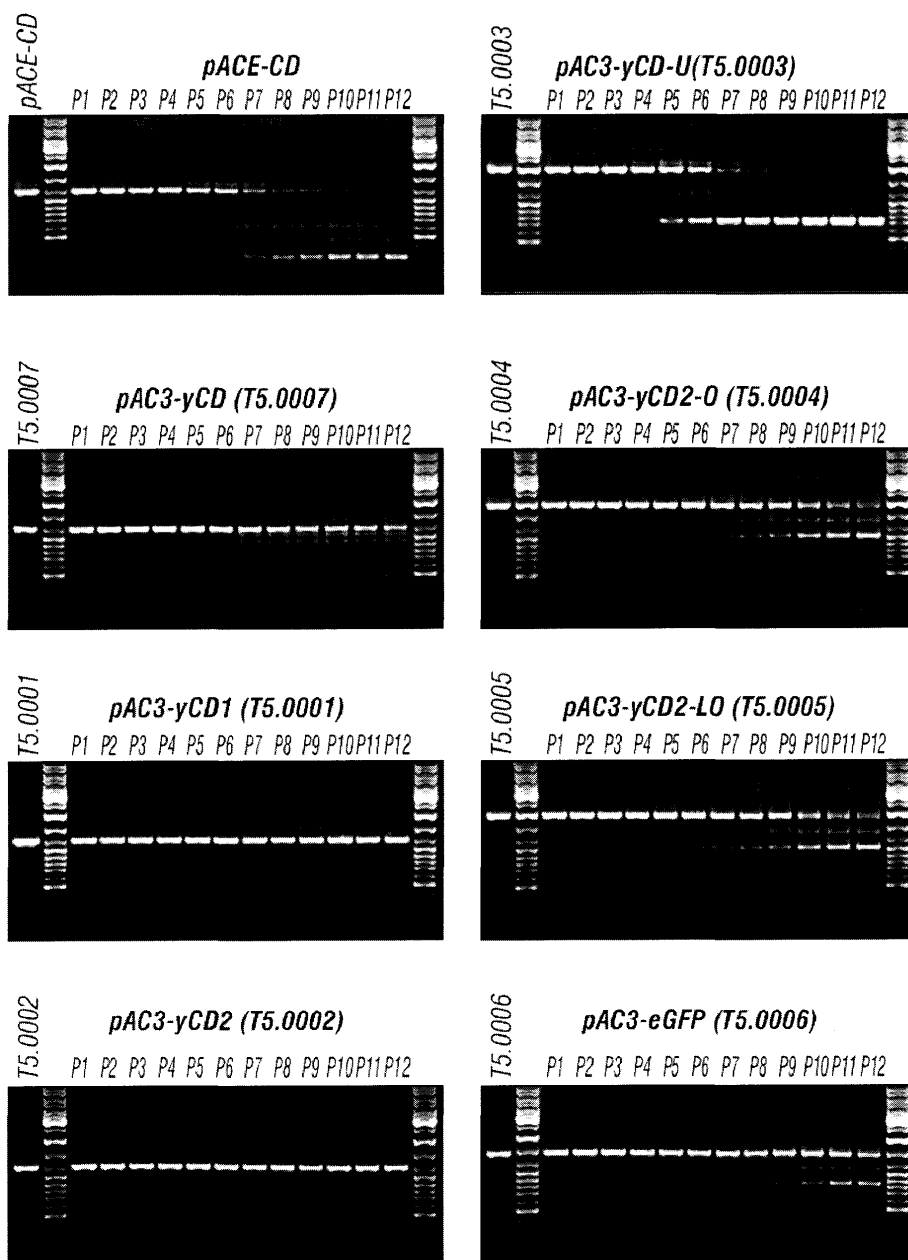
2. Размер гена CD–OPRT: 1,33 тыс. осн.

Фиг. 2D

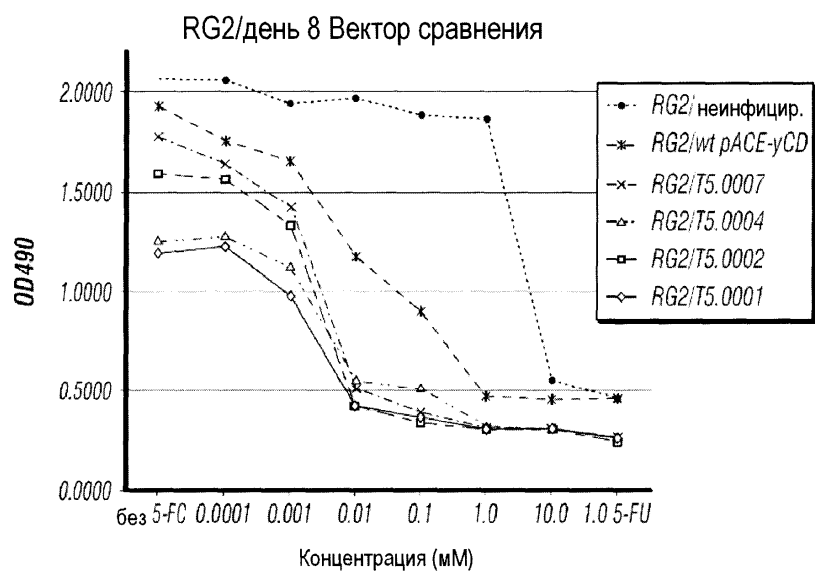


- 1 - U87 + AC3-yCD2 (V)
2 - U87 + ACE-yCD (V)
3 - U87 (неинфицированный)
4 - стандартная молекулярная масса

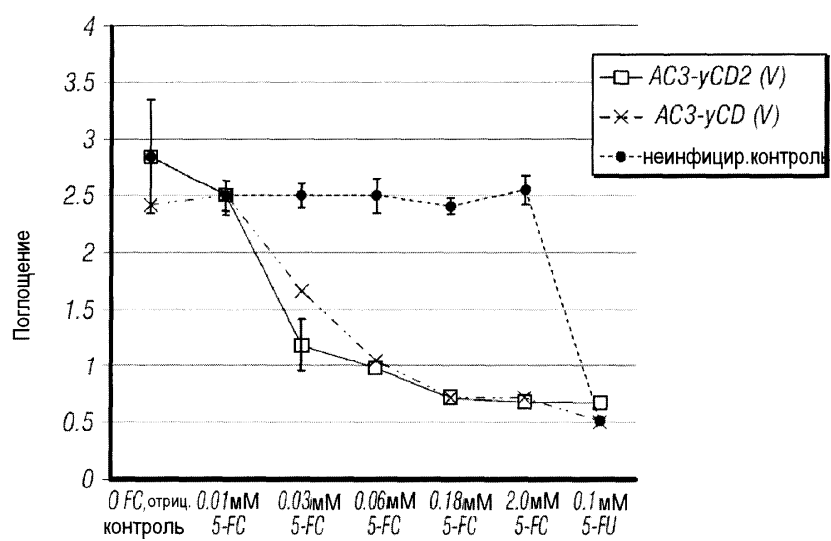
Фиг. 3



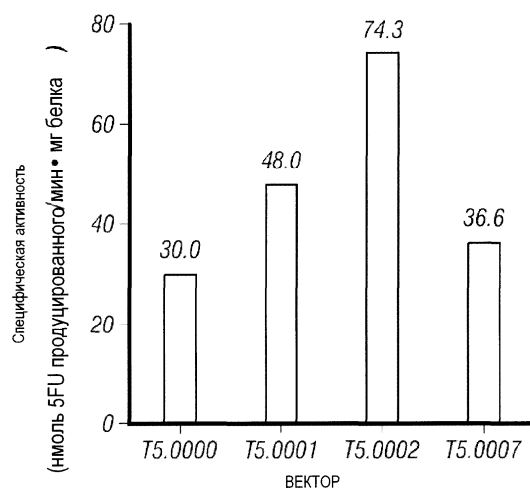
Фиг. 4



Фиг. 5А



Фиг. 5В



Фиг. 6

