

(19)



Евразийское
патентное
ведомство

(21) 201650071 (13) A1

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ

(43) Дата публикации заявки
2017.03.31

(51) Int. Cl. C12N 9/22 (2006.01)
C12N 15/90 (2006.01)

(22) Дата подачи заявки
2015.05.04

(54) TAL-ЭФФЕКТОРНАЯ НУКЛЕАЗА ДЛЯ ЦЕЛЕВОГО НОКАУТА ВИЧ-КОРЕЦЕПТОРА CCR5

(31) 10 2014 106 327.9

(32) 2014.05.07

(33) DE

(86) PCT/DE2015/200295

(87) WO 2015/169314 2015.11.12

(71) Заявитель:

ОБЩЕСТВО С ОГРАНИЧЕННОЙ
ОТВЕТСТВЕННОСТЬЮ
"ЭйДжиСиТи" (RU)

(72) Изобретатель:

Мок Ульрике (GB), Фесе Борис (DE)

(74) Представитель:

Пустовалова М.Л., Котлов Д.В.,
Черняев М.А., Яремчук А.А. (RU)

(57) Изобретение относится к новой TAL-эффекторной нуклеазе (TALEN) для прицельного нокаута ВИЧ-корцептора CCR5. В одном аспекте изобретение представляет пару TAL-эффекторных нуклеаз, включающую в себя первый и второй мономер TAL-эффекторной нуклеазы, причем каждый мономер TAL-эффекторной нуклеазы содержит эндонуклеазный домен со II типом эндонуклеазной активности и ДНК-связывающий домен TAL-эффектора, имеющий множество единиц повтора, каждая из которых имеет вариабельные последовательности из двух аминокислотных остатков (Repeat Variable Diresidue, RVD), и при этом а) ДНК-связывающий домен TAL-эффектора первого мономера TAL-эффекторной нуклеазы связывается с последовательностью-мишенью GCTGGTCATCCTCATCCTG (SEQ ID NO: 1) и/или содержит RVD-последовательность NH HD NG NH NH NG HD NI NG HD HD NG HD NI NG HD HD NG NN, и б) ДНК-связывающий домен TAL-эффектора второго мономера TAL-эффекторной нуклеазы связывается с последовательностью-мишенью AGATGTCAGTCATGCTCTT (SEQ ID NO: 2) и/или содержит RVD-последовательность NI NN NI NG NN NG HD NI NH NG HD NI NG NH HD NG HD NG NG.



A1

201650071

201650071

A1

TAL-эффекторная нуклеаза для целевого нокаута ВИЧ-корцептора CCR5

Изобретение относится к новой TAL-эффекторной нуклеазе (TALEN) для целевого нокаута ВИЧ-корцептора CCR5.

Помимо своей основной функции в клетке хемокиновый рецептор CCR5 играет важную роль при ВИЧ-инфекции. Он выступает в роли корцептора для так называемых CCR5-тропных штаммов ВИЧ, являясь посредником начальной ВИЧ-инфекции. Если на поверхности клетки Т-хелпера отсутствует CCR5, ВИЧ не может связаться с клеткой-хозяином, и инфицирование не происходит. Таким образом, гомозиготная делеция (CCR5 Δ 32) в гене CCR5, которая встречается, приблизительно, у 1% западноевропейцев и "белых" американцев ("европеиды"), практически полностью защищает от ВИЧ-инфицирования CCR5-тропными штаммами. Следовательно, CCR5 является очень интересной мишенью при лечении ВИЧ.

Прежние фармакологические подходы, направленные на блокаду CCR5, требуют пожизненного лечения в рамках комбинированной антиретровирусной терапии, АРТ. В долгосрочной перспективе это связано с возникновением потенциально тяжелых побочных эффектов, а также с отсутствием комплайэнса пациентов и развитием резистентности. Напротив, генетического разрушения («нокаута») CCR5 (в плане генной терапии), в идеальном случае, было бы достаточно в качестве однократного лечения, так как генетическая защита распространяется на все дочерние клетки. Это подтверждается не только естественной резистентностью CCR5 Δ 32-гомозиготных индивидов, но и описанными случаями успешного лечения ВИЧ-инфекции у так называемого "берлинского пациента" после аллогенной трансплантации стволовых клеток с CCR5 Δ 32-гомозиготными донорскими клетками (Hütter G et al. Long-term control of HIV by CCR Δ 32/ Δ 32 stem-cell transplantation. N Engl J Med. 2009, 360: 692-698; Allers K et al. Evidence for the cure of HIV infection by CCR5 Δ 32/ Δ 32 stem cell transplantation. Blood 2011; 117: 2791-2799).

На основе этих наблюдений были разработаны концепции генетического нокаута CCR5 у ВИЧ –инфицированных пациентов. В настоящее время наиболее перспективные стратегии основаны на так называемых «дизайнерских нуклеазах» (см. например Manjunath N. et al., Newer Gene Editing Technologies toward HIV Gene Therapy, Viruses 2013, 5, 2748-2766). Эти дизайнерские нуклеазы состоят из двух компонентов: распознающего домена, который определяет специфичность в геноме и может быть практически полностью разработан без ограничений, и нуклеазного домена, который индуцирует двухцепочечный разрыв в выбранном месте генома. Негомологичное восстановление концов данного разрыва с помощью внутриклеточных систем быстрой репарации ДНК приводит к смещению (нарушению) открытой рамки считывания целевого

гена и, таким образом, в идеальном случае, к нокауту. Первыми широко применяемыми дизайнерскими нуклеазами были нуклеазы «цинковые пальцы» (ZFN). Например, «Sangamo BioSciences, Inc.» в данное время тестирует разработанную ZFN ими CCR5-специфическую нуклеазу цинкового пальца под названием SB-728 (<http://www.sangamo.com/pipeline/sb-728.html>) для клинического применения (Tebas et al., Gene Editing of CCR5 in Autologous CD4 T Cells of Persons Infected with HIV. N Engl J Med 2014;370:901-10). Клинические исследования продемонстрировали осуществимость подхода, однако долгосрочный клинический эффект, выражаемый в виде вирусной нагрузки наблюдался только у испытуемого, который оказался гетерозиготным по естественной мутации CCR5Δ32.

TAL-эффекторные нуклеазы (эффекторные нуклеазы, подобные активаторам транскрипции, TALEN) являются следующим поколением дизайнерских нуклеаз

(см. например, Mussolino, C, Cathomen T. TALE nucleases: tailored genome engineering made easy, Curr Opin Biotechnol. 2012, 23(5): 644-50; WO 2011/072246 A2; EP 2510096 A2; WO 2011/154393 A1; WO 2011/159369 A1; WO 2012/093833 A2; WO 2013/182910 A2).

По сравнению с ZFN они отличаются в том числе более высокой специфичностью, что значительно снижает риск таких нецелевых эффектов, как закрепление мутации в другом, отличном от желаемого, месте генома (Händel E-M, Cathomen T. Zinc-finger nuclease based genome surgery: it's all about specificity. Curr Gene Ther 2011, 11: 28-37; Mussolino C et al. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic Acids Res. 2011; 39: 9283-9293).

CCR5-специфические TALEN уже известны (см., например, Mussolino C et al. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic Acids Res. 2011; 39: 9283-9293; WO 2011/146121 A1; WO 2012/093833 A2; US 2013/0217131 A1), клиническое исследование, однако пока еще не описано.

По-прежнему существует потребность в средствах для эффективного лечения ВИЧ-инфекции. Поэтому задачей изобретения является создание такого средства. В частности, задачей данного изобретения является создание препарата, при помощи которого может быть достигнут более эффективный нокаут ВИЧ-корцептора CCR5, чем созданный ранее.

Задача решается при осуществлении изобретения по п.1, а также других независимых пунктов формулы изобретения. Частные дополнительные варианты осуществления настоящего изобретения представлены в зависимых пунктах формулы.

В первом аспекте изобретение представляет пару TAL-эффекторных нуклеаз, которая включает в себя первый и второй мономер TAL-эффекторной нуклеазы, причем каждый мономер TAL-эффекторной нуклеазы содержит эндонуклеазный домен со II

типом эндонуклеазной активности и ДНК-связывающий домен TAL-эффектора, имеющий множество единиц повтора, каждая из которых имеет пару переменных аминокислот (Repeat Variable Diresidue, RVD), и при этом

а) ДНК-связывающий домен TAL-эффектора первого мономера TAL-эффекторной нуклеазы связывается с последовательностью-мишенью GCTGGTCATCCTCATCCTG (SEQ ID NO: 1) и/или содержит RVD-последовательность NH HD NG NH NH NG HD NI NG HD HD NG HD NI NG HD HD NG NN,

и

б) ДНК-связывающий домен TAL-эффектора второго мономера TAL-эффекторной нуклеазы связывается с последовательностью-мишенью AGATGTCAGTCATGCTCTT (SEQ ID NO: 2) и/или содержит RVD-последовательность NI NN NI NG NN NG HD NI NH NG HD NI NG NH HD NG HD NG NG.

Пара TAL-эффекторных нуклеаз в соответствии с настоящим изобретением способна вызвать нокаут CCR5 в первичных Т-лимфоцитах с не имеющей себе равных эффективностью >50%. К тому же, как ни удивительно, изобретение делает возможным последовательный биаллельный нокаут обеих CCR5-аллелей и, таким образом, обеспечивает полную защиту модифицированных клеток от проникновения ВИЧ, вопреки мнению ведущих экспертов в данной области техники, что это в настоящее время невозможно („Consistent nuclease-mediated biallelic knockdown is not yet tenable“, см. Kay, M.A. und Walker, B.D., 2014, Engineering Cellular Resistance to HIV, N Engl J Med 370:968-969). Кроме того, было установлено, что TALEN-пара, в соответствии с изобретением, идеально подходит для доставки в клетку, основанной на трансфекции в виде мРНК (щадящего, безопасного и GMP-совместимого метода). Таким образом, изобретение впервые представляет основанное на дизайнерской нуклеазе средство для лечения ВИЧ, которое сочетает в себе высокую нокаут-эффективность и нокаут-селективность с низким уровнем нецелевых эффектов и другие, полезные с фармакологической точки зрения, свойства.

Под "TAL-эффекторная нуклеаза" или "TALEN" (эффекторные нуклеазы, подобные активаторам транскрипции) подразумевается гибридный белок, который содержит ДНК-связывающий домен TAL-эффектора (TALE) и ДНК-расщепляющий домен эндонуклеазы рестрикции. TAL-эффекторы представляют собой ДНК-связывающие белки, которые получают из таких фитопатогенов, как *Xanthomonas* spp.

ДНК связывание TAL-эффекторов опосредовано через домен с переменным количеством (как правило, от 5 до 30) единиц повтора ("повторов"), которые, как правило, формируются из 33-35 аминокислот. Каждая единица повтора имеет два высоко переменных аминокислотных остатка (repeat-variable diresidue, RVD), расположенных, как правило, в позициях 12 и 13, и связывающихся только с одним основанием целевой

последовательности ДНК. Отношение между RVD и нуклеотидами целевой молекулы ДНК приводятся ниже.

RVD (однолитерный код)	RVD (трехлитерный код)	Нуклеотид(ы)
NH	Asn-His	G
HD	His-Asp	C
NG	Asn-Gly	T
NI	Asn-Ile	A
NN	Asn-Asn	R (G, A)
NK	Asn-Lys	G
NS	Asn-Ser	N (A, C, G, T)

Под "RVD-последовательностью" здесь подразумевается непрерывная последовательность RVD в одном ДНК-связывающем домене TAL-эффектора, в которой последовательность указана, если не указано иного, в направлении N-C, т.е. от N-конца к C-концу.

Очевидно, что в данном случае специалисту в данной области техники известно, что "RVD-последовательности" не следуют непосредственно друг за другом, поскольку они не связаны непосредственно друг с другом ковалентными связями, а непосредственно связаны между собой единицы повтора ("повторы"), в которых содержатся RVD, таким образом, соответствующие RVD отделены аминокислотами основной структуры единиц повтора.

Под "последовательностью-мишенью" здесь подразумевается нуклеотидная последовательность, как правило, последовательность ДНК, которая связана связывающим доменом TAL-эффектора.

Раскрытые в данном документе RVD-последовательности, каждая из которых состоит из 19 RVD, имеют следующие последовательности-мишени (в направлении 5'-3'; однолитерный код для аминокислот):

NH HD NG NH NH NG HD NI NG HD HD NG HD NI NG HD HD NG NN
GCTGGTCATCCTCATCCTG (SEQ ID NO: 1)

NI NN NI NG NN NG HD NI NH NG HD NI NG NH HD NG HD NG NG
AGATGTCAGTCATGCTCTT (SEQ ID NO: 2)

Под "единицей повтора" ("повтором") относительно связывающего домена TAL-эффектора здесь подразумевается непрерывная последовательность, как правило, из 33-35, в основном 34 аминокислот, которые, кроме высоко вариабельных аминокислот RVD в

позициях 12 и 13, имеют практически идентичные аминокислотные последовательности. Вполне возможно, что и в консервативной основной структуре единицы повтора, т.е. по сути, в однородной структуре, в которую включены высоко вариабельные аминокислоты RVD, варьируют единичные аминокислоты, например, в позициях 4, 10 и/или 32 в единице повтора из 34 аминокислот. Типичная единица повтора может иметь, например, следующую последовательность аминокислот (подстрочные цифры определяют позицию внутри единицы повтора):

LTPX₄QVVAIX₁₀SX₁₂X₁₃GGKQALETVQRLLPVLCQX₃₂HG (SEQ ID NO: 5)

X обозначает любую аминокислоту, при этом в позициях 12 и 13 стоят гипервариабельные аминокислоты RVD. В позиции 4 (X4) могут стоять, например, аминокислоты E, Q, D или A, в позиции 32 (X32) - аминокислоты A или D. В позиции 10 может стоять, например, A или V. Примеры единиц повтора приводятся ниже (XX обозначает гипервариабельные аминокислоты RVD, вариабельные аминокислоты подчеркнуты):

LTPEQVVAIASXXGGKQALETVQRLLPVLCQAHG (SEQ ID NO: 6)

LTPQQVVAIASXXGGKQALETVQRLLPVLCQAHG (SEQ ID NO: 7)

LTPDQVVAIASXXGGKQALETVQRLLPVLCQDHG (SEQ ID NO: 8)

LTPAQVVAIASXXGGKQALETVQRLLPVLCQDHG (SEQ ID NO: 9)

LTPEQVVAIVSXXGGKQALETVQRLLPVLCQAHG (SEQ ID NO: 10)

Связывающий домен TAL-эффектора может включать в себя один или несколько подобных вариантов единиц повтора, причем в расчет принимаются также комбинации различных вариантов.

Крайняя, ближайшая к нуклеазному домену единица повтора может содержать меньше, например, лишь первые 15,16,17,18,19 или 20 аминокислот остальных единиц повтора. Такая единица повтора называется "половинная единица повтора" или "полуповтор".

Под "ДНК-связывающим доменом" здесь подразумевается участок белка, который индуцирует связывание белка с молекулой ДНК. В случае ДНК-связывающего домена TAL-эффектора это происходит посредством описанных выше единиц повтора ("повторов").

Формулировку, согласно которой ДНК-связывающий домен TAL-эффектора связывается с последовательностью-мишенью, следует понимать так, что ДНК-связывающий домен TAL-эффектора за счет его RVD-последовательности специфично связывается с ДНК-последовательностью-мишенью. Вместе с тем не обязательно, хотя

предпочтительно, чтобы с каждым нуклеотидом последовательности-мишени был связан RVD в связывающем домене. Связь между RVD-последовательностью ДНК-связывающего домена TAL-эффектора и последовательностью-мишенью должна быть именно такой, чтобы произошло связывание с последовательностью-мишенью. "Специфично" в данном контексте означает, что связывание происходит в основном только с последовательностью-мишенью.

Под "мономером TAL-эффекторной нуклеазы" подразумевается TAL-эффекторная нуклеаза, которая состоит из одной единственной полипептидной цепи. Под "парой TAL-эффекторных нуклеаз" или "TALEN-парой" подразумевается TALEN, состоящая из двух мономеров TAL-эффекторной нуклеазы. Мономеры представляют собой левое или правое плечо TALEN, которые связывают противоположные нити ДНК и вместе вызывают разрыв ДНК в одном сайте.

Ссылка на TALEN-пару из "левой" или "правой" TALEN, или "левого" или "правого" TALEN-плеча отображает тот факт, что в одной TALEN-паре TALEN-мономеры используются попарно, т.е. вызывают разрыв нитей внутри двухцепочечной ДНК, так как один мономер связывается с последовательностью-мишенью смысловой нити, в то время как другой TALEN-мономер TALEN-пары связывается с последовательностью мишенью комплементарной антисмысловой нити, а именно таким образом, что нуклеазные домены направлены друг к другу, в одном общем участке ДНК, называемом "спейсер", располагаются между последовательностями-мишенями и каждый из нуклеазных доменов вызывает одноцепочечный разрыв. "Левый" и "правый" TALEN-мономеры являются частями определенной TALEN-пары, причем "левой" TALEN зачастую обозначают ту TALEN, которая связывается со смысловой нитью, в то время как "правая" TALEN связывается с комплементарной нитью. Тем не менее, ссылка на "левую" или "правую" TALEN, при этом, необязательно указывает на то, что "левая" TALEN связывается со смысловой нитью, а "правая" - с комплементарной нитью.

Таким образом, к данному изобретению относится также "TALEN-пара", т.е. пара из двух относящихся к изобретению мономеров, каждый из которых представляет левое или правое плечо TALEN. Также к данному изобретению относится пара TAL-эффекторных нуклеаз, содержащая мономер TAL-эффекторной нуклеазы, ДНК-связывающий домен TAL-эффектора которого связывается с последовательностью-мишенью GCTGGTCATCCTCATCCTG (SEQ ID NO: 1) и/или содержит RVD-последовательность NH HD NG NH NH NG HD NI NG HD HD NG HD NI NG HD HD NG NN; и мономер TAL-эффекторной нуклеазы, ДНК-связывающий домен TAL-эффектора которого связывается с последовательностью-мишенью AGATGTCAGTCATGCTCTT (SEQ ID NO: 2) и/или содержит RVD-последовательность NI NN NI NG NN NG HD NI NH NG HD NI NG NH HD NG HD NG NG.

Под "эндонуклеазным доменом со II типом эндонуклеазной активности" подразумевается полипептид, который обладает ДНК-расщепляющей активностью нуклеазы рестрикции и разрезает ДНК внутри или в непосредственной близости от последовательности узнавания, не требующий АТФ и не обладающий активностью метилтрансфераз. Под "эндонуклеазным доменом с IIS типом эндонуклеазной активности" подразумевается домен эндонуклеазы II типа, сайт расщепления которого находится в непосредственной близости от последовательности узнавания, но не внутри нее.

Под "CCR5" подразумевается человеческий C-C рецептор хемокина 5 (обозначаемый также CD195, CMKBR5 или CC-CKR5). Последовательность человеческого CCR5 показана в SEQ ID NO: 11 (см. номер доступа NC_018914.2 в базе данных NCBI).

Под "вектором" подразумевается система транспортировки для доставки в основном чужеродных нуклеиновых кислот в живую клетку-реципиента путем трансфекции или трансдукции. Под "вектором-переносчиком генов" подразумевается вектор, при помощи которого ген может быть введен в клетку. Векторы (-переносчики генов) хорошо известны специалистам. Примерами векторов-переносчиков генов являются плазмиды, вирусные векторы или мРНК.

Под "нуклеиновой кислотой" подразумевается полимер, мономерами которого являются нуклеотиды. Нуклеотид является соединением, состоящим из остатка сахара, азотсодержащего гетероциклического органического основания (нуклеотидного основания или нуклеооснования) и фосфатной группы. Остатком сахара, как правило, является пентоза, в случае ДНК это дезоксирибоза, в случае РНК это рибоза. Связь между нуклеотидами осуществляется через фосфатную группу при помощи фосфодиэфирных мостиков, как правило, между 3'-углеродным атомом сахара нуклеозида (соединения сахара и нуклеооснования) и 5'-углеродным атомом сахара следующего нуклеозида. Термин «нуклеиновая кислота» включает, например, ДНК, РНК и ДНК/РНК-гибриды. В том смысле, в котором он используется здесь, термин "нуклеиновая кислота" относится, прежде всего, к изолированной нуклеиновой кислоте. Под "изолированной нуклеиновой кислотой" подразумевается синтетически изготовленная или извлеченная из ее естественной или изначальной среды нуклеиновая кислота.

Термин "включающий в себя" используется таким образом, что он определяет не только объект, который проявляет признаки, соответствующие исключительно данному понятию, но и объект, который проявляет признаки, соответствующие данному определению, а также дополнительные признаки. Определение объекта благодаря тому, что он включает в себя определенные признаки, включает также определение данного объекта по окончательному списку этих признаков, т.е. по наличию исключительно этих признаков.

В предпочтительном варианте относящейся к изобретению пары TAL-эффекторных нуклеаз, эндонуклеазный домен в каждом мономере TAL-эффекторной нуклеазы является С-концевым по отношению к ДНК-связывающему домену TAL-эффектора. Предпочтительно, каждый повтор, за исключением того, который примыкает непосредственно к эндонуклеазному домену, включает в себя 33-35 аминокислот, преимущественно 34 аминокислоты, причем RVD находятся в каждой единице повтора в позициях 12 и 13. Особое предпочтение отдается всем единицам повтора, за исключением "полуповторов" аминокислотной последовательности соответствующей SEQ ID NO: 5, при чем в позиции 4 может стоять E, Q, D или A, в позиции 10 может стоять A или V и в позиции 32 может стоять A или D. Основная структура единиц повтора может быть одинаковой или различной для всех единиц повтора. Относительно аминокислот возможны изменения одной или нескольких единиц повтора в позициях внутри основной структуры, например, в позициях 4, 10 и/или 32. Единица повтора, непосредственно примыкающая к эндонуклеазному домену, может включать меньшее число аминокислот, например, 15, 16, 17, 18, 19 или 20, причем аминокислоты в этом случае соответствуют преимущественно первым 15, 16, 17, 18, 19 или 20 аминокислотам в других повторах. При этом, например, аминокислота в позиции 4 может отличаться, быть, например, E, Q, D или A, и/или может отличаться аминокислота в позиции 10, быть, например, V вместо A.

Особое предпочтение отдается эндонуклеазному домену мономера TAL-эффекторной нуклеазы эндонуклеазного домена IIS типа, особенно ДНК-расщепляющему домену эндонуклеазы FokI. Аминокислотная последовательность для подходящего расщепляющего домена FokI приведена в SEQ ID NO: 12. В расчет берется также другой расщепляющий домен эндонуклеазы II типа. Эндонуклеазы II типа известны специалистам, и подходящие расщепляющие домены могут быть определены при помощи рутинных исследований.

В особо предпочтительном варианте первый мономер TAL-эффекторной нуклеазы содержит аминокислотную последовательность, соответствующую SEQ ID NO: 3, а второй мономер TAL-эффекторной нуклеазы включает в себя аминокислотную последовательность, соответствующую SEQ ID NO: 4. В SEQ ID NO: 3 приведена левая TALEN (далее обозначаемая также CCR5-Uco-L или левое плечо CCR5-Uco), а в SEQ ID NO: 4 - правая TALEN (далее обозначаемая также CCR5-Uco-R или правое плечо CCR5-Uco) TALEN-пары, которые вместе вызывают двухцепочечный разрыв в ДНК-последовательности CCR5 внутри спейсера, расположенного между последовательностями-мишенями, соответствующими SEQ ID NO:1 и SEQ ID NO:2. Восстановление данного двухцепочечного разрыва происходит за счет клеточной репарационной системы (негомологичного соединения концов, НСК) с высокой вероятностью нарушения рамки считывания и, таким образом, к нокауту CCR5.

Во втором аспекте данное изобретение представляет также нуклеиновую кислоту, включающую в себя

а) первую нуклеиновую кислоту, которая кодирует первый мономер TAL-эффекторной нуклеазы, где первый мономер TAL-эффекторной нуклеазы содержит эндонуклеазный домен со II типом эндонуклеазной активности и ДНК-связывающий домен TAL-эффектора с множеством единиц повтора, каждая из которых имеет пару переменных аминокислот (RVD), и при этом ДНК-связывающий домен TAL-эффектора связывается с последовательностью-мишенью GCTGGTCATCCTCATCCTG (SEQ ID NO: 1) и/или содержит RVD-последовательность NH HD NG NH NH NG HD NI NG HD HD NG HD NI NG HD HD NG NN, и

б) вторую нуклеиновую кислоту, которая кодирует второй мономер TAL-эффекторной нуклеазы, где второй мономер TAL-эффекторной нуклеазы содержит эндонуклеазный домен со II типом эндонуклеазной активности и ДНК-связывающий домен TAL-эффектора с множеством единиц повтора, каждая из которых имеет пару переменных аминокислот (RVD), и при этом ДНК-связывающий домен TAL-эффектора связывается с последовательностью-мишенью AGATGTCAGTCATGCTCTT (SEQ ID NO: 2) и/или содержит RVD-последовательность NI NN NI NG NN NG HD NI NH NG HD NI NG NH HD NG HD NG NG.

TALEN-мономеры, составляющие относящуюся к изобретению TALEN-пару, в данном аспекте изобретения кодируются вместе в общей нуклеиновой кислоте. На примере нуклеиновой кислоты, речь может идти о плазмиде или другом подходящем (переносчике генов) векторе. Подходящие векторы и способы их получения и применения хорошо известны на современном техническом уровне. При известных условиях нуклеиновая кислота наряду с TALEN-кодом может содержать и другие элементы, например, один или несколько промоторов, сигналы полиаденилирования и др.

TALEN-мономеры, составляющие относящуюся к изобретению TALEN-пару, могут также кодироваться в двух отдельных нуклеиновых кислотах. В третьем аспекте данное изобретение представляет комбинацию нуклеиновых кислот, включающую в себя

а) первую нуклеиновую кислоту, которая кодирует первый мономер TAL-эффекторной нуклеазы, где первый мономер TAL-эффекторной нуклеазы содержит эндонуклеазный домен со II типом эндонуклеазной активности и ДНК-связывающий домен TAL-эффектора с множеством единиц повтора, каждая из которых имеет пару переменных аминокислот (RVD), и при этом ДНК-связывающий домен TAL-эффектора связывается с последовательностью-мишенью GCTGGTCATCCTCATCCTG (SEQ ID NO: 1) и/или содержит RVD-последовательность NH HD NG NH NH NG HD NI NG HD HD NG HD NI NG HD HD NG NN, и

б) вторую нуклеиновую кислоту, которая кодирует второй мономер TAL-эффекторной нуклеазы, где второй мономер TAL-эффекторной нуклеазы содержит

эндонуклеазный домен со II типом эндонуклеазной активности и ДНК-связывающий домен TAL-эффектора с множеством единиц повтора, каждая из которых имеет пару переменных аминокислот (RVD), и при этом ДНК-связывающий домен TAL-эффектора связывается с последовательностью-мишенью AGATGTCAGTCATGCTCTT (SEQ ID NO: 2) и/или содержит RVD-последовательность NI NN NI NG NN NG HD NI NH NG HD NI NG NH HD NG HD NG NG.

В случае и первой, и второй нуклеиновой кислоты, речь идет, предпочтительно о мРНК, особенно предпочтительно о консервированной мРНК (см. например, Kallen K.-J. et al., A novel, disruptive vaccination technology, *Hum Vaccin Immunother.* Oct 1, 2013; 9(10): 2263–2276, doi: 10.4161/hv.25181; Kallen K.-J. und Theß A., A development that may evolve into a revolution in medicine: mRNA as the basis for novel, nucleotide-based vaccines and drugs, *Ther Adv Vaccines.* Jan 2014; 2(1): 10–31, doi: 10.1177/2051013613508729). При известных условиях первая и вторая нуклеиновые кислоты наряду с TALEN-кодом могут содержать и другие элементы, например, один или несколько промоторов, сигналы полиаденилирования и др. мРНК, подходящие для левого и правого плеча, относящегося к изобретению TALEN, примерно указаны в SEQ ID NO: 17 и 18.

В случае мРНК, она транспортируется в клетку(и)-мишень(и), например, Т-лимфоциты, особое предпочтение отдается методу, описанному Berdien и др. (Berdien B et al., TALEN-mediated editing of endogenous T-cell receptors facilitates efficient reprogramming of T lymphocytes by lentiviral gene transfer, *Gene Therapy*, 2014, doi:10.1038/gt.2014.26). Особое предпочтение отдается одновременному введению обоих плеч TALEN (правому и левому плечу) в клетку.

Введение относящихся к изобретению TALEN-пар при помощи мРНК имеет ряд несомненных преимуществ для клинического применения. Можно отказаться от использования вектора-переносчика генов на основе ДНК, что несомненно упростит изготовление и практическое использование. мРНК-опосредованная экспрессия TALEN настолько коротка, что мРНК очень быстро деградирует в клетке-мишени. Это дополнительно уменьшает риск возникновения нецелевых эффектов. Кроме того, клетки-мишени должны культивироваться весьма короткое время *in vitro*. Соответствующая технология легко приспосабливается к требованиям GMP. Ожидается, что, в противоположность к вирусным или плазмидным векторам, будут отсутствовать побочные эффекты, возникающие вследствие самого переноса генов (например, инсерционный мутагенез) или, возникшей в результате нежелательного внедрения вектора, длительной TALEN-экспрессии (нецелевые эффекты, активация TALEN-специфического иммунного ответа).

В дополнительных аспектах данное изобретение представляет вектор, в частности вектор-переносчик генов, включающий в себя относящуюся к изобретению нуклеиновую кислоту и изолированную клетку-хозяина, включающую в себя относящийся к

изобретению вектор, относящуюся к изобретению нуклеиновую кислоту или комбинацию нуклеиновых кислот, причем в отношении изолированных клеток-хозяев речь не идет о зародышевых клетках живого организма человека, в частности о зародышевых клетках человека или эмбриональных зародышевых клетках человека, а также и об эмбриональных стволовых клетках, при получении которых был или будет уничтожен человеческий эмбрион.

Еще в одном дополнительном аспекте данное изобретение представляет фармацевтическую композицию, включающую в себя нуклеиновую кислоту, комбинацию нуклеиновых кислот или вектор, которые соответствуют данному изобретению. Фармацевтический состав может содержать вспомогательные вещества, например, такие как растворители, усилители растворимости, растворы-ускорители, солеобразующие агенты, соли, буферы, вещества, влияющие на вязкость и консистенцию, гелеобразователи, эмульгаторы, солюбилизаторы, увлажнители, сепараторы, антиоксиданты, консерванты, наполнители, носители и т.д.

Еще в одном дополнительном аспекте данное изобретение представляет лекарственное средство, включающее в себя нуклеиновую кислоту, комбинацию нуклеиновых кислот или фармацевтическую композицию, которые соответствуют данному изобретению.

Далее приводятся пояснения к изобретению при помощи примеров выполнения и прилагаемых для большей наглядности рисунков.

Рис. 1. Схематическое изображение ДНК-связывающего домена относящейся к изобретению CCR5-специфической TALEN-пары ("CCR5-Uco") и ее последовательности-мишени в гене CCR5. Каждая нижняя строка указывает последовательность-мишень для а) левого и б) правого TALEN-плеча, каждая верхняя строка - RVD (repeat variable diresidues) соответствующих TALEN-мономеров в рамочке (аминокислоты указаны однолитерным кодом); с) отрезок (нуклеотиды 135–221 последовательности SEQ ID NO:11) ДНК CCR5 с комплементарной цепочкой. Последовательности-мишени левого (верхняя) и правого (нижняя, на комплементарной цепочке) плеч CCR5-Uco-TALEN выделены рамками.

Рис. 2. Сравнение эффективности между относящейся к изобретению CCR5-TALEN („Uco“) и контрольной CCR5-TALEN („Mco“) из предшествующего уровня техники. Проводилось тестирование посредством трансфекции плазмид в CCR5-положительную клетку-репортер линии клеток 293Т. Для всех проверенных конструкций наблюдалась сравнимая эффективность трансфекции (на основе контртрансфекции eGFP). Нокаут CCR5 определялся на шестой день после трансфекции CCR5+ 293Т клетки клона при помощи специфических антител (анти-CD195-APC-Cy-7 антитела) (n=2). Для „mock“-

контроля проводилась трансфекция клеток с иррелевантной контрольной плазмидой (pUC), которая не кодирует TALEN.

Рис. 3. CCR5 нокаут в первичных Т-лимфоцитах с CCR5-Uco после мРНК-трансфекции. После активации ex-vivo приблизительно половина первичных Т-лимфоцитов здорового пробанда экспрессировали CCR5 (= клетки справа от пунктирной линии, см. "нетрансфицированные"): а) После трансфекции CCR5-специфической TALEN (Uco) уменьшилась доля CCR5-положительных клеток с нарастающим количеством трансфицированных мРНК (обратно пропорционально). б) Контрольная трансфекция единичных TALEN-плеч, напротив, не приводит к уменьшению доли CCR5-положительных клеток. Нокаут CCR5 определялся на шестой день после мРНК-трансфекции при помощи специфических антител (anti-CD195-PerCP-Cy5.5 антитела). с) Анализ сайта-мишени CCR5-Uco-TALEN показал генетический нокаут в 9 из 17 (>50%) проанализированных первичных Т-лимфоцитов.

Примеры

Была изготовлена и исследована относящаяся к изобретению CCR5-специфическая TALEN (далее "CCR5-Uco"). Относящаяся к изобретению TALEN отличается от описанной ранее TALEN в отношении последовательности-мишени, распознаваемой в гене CCR5 (см. рис.1).

По сравнению с кодон-оптимизированной CCR5-TALEN („Mco“; см. SEQ ID NO: 13, 14), основываясь на опубликованных работах лаборатории проф. Toni Cathomen (Фрайбург) (см. Mussolino C et al., A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity, Nucleic Acids Res. 2011, 39: 9283-929), относящаяся к изобретению CCR5-Uco-TALEN показывает более высокий уровень индукции CCR5-нокаута после трансфекции плазмиды в линии клеток-репортеров (см. рис.2). Последовательность нуклеиновых кислот кодон-оптимизированных TALEN-элементов для использования в клетках человека основывается на публикациях группы ученых во главе с Feng Zhang (Zhang et al., Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription, Nature Biotechnology, 2011, 29, 149-153; Sanjana NE et al., A TAL Effector Toolbox for Genome Engineering, Nature Protocols, 2012, 7: 171-192).

Ниже приведены RVD-последовательности CCR5-Mco-TALEN в соответствии предшествующим техническим уровнем:

Левое плечо (L) = NN NG NN NN NN HD NI NI HD NI NG NN HD NG NN NN NG HD;

Правое плечо (R) = HD NG NG HD NI NN HD HD NG NG NG NG NN HD NI NN NG NG.

Соответствующие последовательности распознавания на уровне ДНК:

L на смысловой нити = GTGGGCAACATGCTGGTC (SEQ ID NO: 15);

R на антисмысловой нити = CTTCAGCCTTTTGCAGTT (SEQ ID NO: 16).

Длина спейсера составляет 15 нуклеотидов. Производство TALEN-плазмид также основывается на публикациях группы ученых во главе с Zhang F, или Sanjana NE и др. (см. выше).

При помощи мРНК-трансфекции (см. Berdien B et al., TALEN-mediated editing of endogenous T-cell receptors facilitates efficient reprogramming of T lymphocytes by lentiviral gene transfer, *Gene Therapy*, 2014, doi:10.1038/gt.2014.26) с относящейся к изобретению CCR5-Uco производится нокаут в первичных Т-лимфоцитах. (см. рис. 3). Последовательности нуклеиновых кислот используемых здесь мРНК указаны в SEQ ID NO: 17 и 18. В SEQ ID NO: 17 указана мРНК левого TALEN-плеча, в SEQ ID NO: 18 указана мРНК правого TALEN-плеча. Нуклеотиды 10-3225 мРНК в SEQ ID NO:17 и 18 кодируют каждое из TALEN-плеч (мономеры), аминокислотные последовательности которых указаны в SEQ ID NO: 3, 4.

Трансфецированная мРНК была создана при помощи T7-промотора после AvrII-линеаризации вектора. Так как сайт расщепления AvrII 563bp находится за стоп-кодоном, указанная последовательность больше чем открытая рамка считывания. После линеаризации соответствующие Uco-TALEN-ДНК используются в качестве матрицы для синтеза мРНК при помощи T7 mScript™ Standard mRNA Production System от Cellscript (Madison, WI 53713 USA). мРНК была оснащена 5'Cap-концом и Poly-A-концом в соответствии с руководством производителя. Трансфекция мРНК производится путем электропорации первичных Т-клеток в течение 10 мс при 300V. При помощи Mco-CCR5-TALEN, однако, не удалось реализовать CCR5-нокаут в первичных Т-клетках или линиях Z-клеток (в то время как нокаут рецепторов Т-клеток был возможным, см. Berdien et al 2014, см. выше). Только с относящейся к изобретению CCR5-Uco-TALEN, при помощи мРНК-переноса, удалось выключить значительную часть >50% CCR5-аллелей. Это позволяет предположить, что только достаточно активные TALEN способны выполнять свою функцию в первичных Т-клетках после мРНК-трансфекции. Таким образом, относящаяся к изобретению CCR5-TALEN особенно хорошо подходит для использования с помощью мРНК-трансфекции, что делает ее крайне привлекательной с точки зрения клинического применения.

Обзор последовательностей:

SEQ ID NO:	Тип	Описание
01	ДНК	Последовательность-мишень TALEN CCR5-Uco L
02	ДНК	Последовательность-мишень TALEN CCR5-Uco R
03	ПРТ	TALEN CCR5-Uco L
04	ПРТ	TALEN CCR5-Uco R

05	ПРТ	Последовательность повтора (Консенсус)
06	ПРТ	Последовательность повтора
07	ПРТ	Последовательность повтора
08	ПРТ	Последовательность повтора
09	ПРТ	Последовательность повтора
10	ПРТ	Последовательность повтора
11	ДНК	hCCR5
12	ПРТ	FokI-домен расщепления
13	ПРТ	TALEN CCR5-Mco L
14	ПРТ	TALEN CCR5-Mco R
15	ДНК	Последовательность-мишень TALEN CCR5-Mco L
16	ДНК	Последовательность-мишень TALEN CCR5-Mco R
17	мРНК	мРНК CCR5-Uco L
18	мРНК	мРНК CCR5-Uco R

ПЕРЕЧЕНЬ ПОСЛЕДОВАТЕЛЬНОСТЕЙ - ПРОИЗВОЛЬНЫЙ ТЕКСТ

TALEN repeat = TALEN-единица повтора

Any amino acid, or E, Q, D or A = Любая аминокислота, или E, Q, D или A

Any amino acid, or A or V = Любая аминокислота, или A или V

Any amino acid, or A or D = Любая аминокислота, или A или D

Repeat variable diresidue (RVD) = Вариабельные последовательности из двух аминокислотных остатков (RVD)

FokI cleavage domain = FokI - домен расщепления

ПЕРЕЧЕНЬ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

<110> Общество с ограниченной ответственностью "ЭйДжиСиТи"

<120> TAL-ЭФФЕКТОРНАЯ НУКЛЕАЗА ДЛЯ ЦЕЛЕВОГО НОКАУТА ВИЧ-КОРЕЦЕПТОРА CCR5

<130> PAT 1473 PCT

<150> DE102014106327.9

<151> 2014-05-07

<160> 18

<170> BiSSAP 1.3

<210> 1

<211> 19

<212> ДНК

<213> Homo sapiens

<400> 1

gctggtcatc ctcatcctg

19

<210> 2

<211> 19

<212> ДНК

<213> Homo sapiens

<400> 2

agatgtcagt catgctctt

19

<210> 3

<211> 1071

<212> ПРТ

<213> Искусственная последовательность

<220>

<223> CCR5-Uco-L

<400> 3

Met	Asp	Tyr	Lys	Asp	His	Asp	Gly	Asp	Tyr	Lys	Asp	His	Asp	Ile	Asp
1				5					10					15	
Tyr	Lys	Asp	Asp	Asp	Asp	Lys	Met	Ala	Pro	Lys	Lys	Lys	Arg	Lys	Val
			20					25					30		
Gly	Ile	His	Gly	Val	Pro	Ala	Ala	Val	Asp	Leu	Arg	Thr	Leu	Gly	Tyr
		35				40						45			
Ser	Gln	Gln	Gln	Gln	Glu	Lys	Ile	Lys	Pro	Lys	Val	Arg	Ser	Thr	Val
		50				55					60				
Ala	Gln	His	His	Glu	Ala	Leu	Val	Gly	His	Gly	Phe	Thr	His	Ala	His
65					70					75					80
Ile	Val	Ala	Leu	Ser	Gln	His	Pro	Ala	Ala	Leu	Gly	Thr	Val	Ala	Val
			85						90					95	
Lys	Tyr	Gln	Asp	Met	Ile	Ala	Ala	Leu	Pro	Glu	Ala	Thr	His	Glu	Ala
			100					105					110		
Ile	Val	Gly	Val	Gly	Lys	Gln	Trp	Ser	Gly	Ala	Arg	Ala	Leu	Glu	Ala
		115				120					125				
Leu	Leu	Thr	Val	Ala	Gly	Glu	Leu	Arg	Gly	Pro	Pro	Leu	Gln	Leu	Asp
		130				135					140				
Thr	Gly	Gln	Leu	Leu	Lys	Ile	Ala	Lys	Arg	Gly	Gly	Val	Thr	Ala	Val
145					150					155					160
Glu	Ala	Val	His	Ala	Trp	Arg	Asn	Ala	Leu	Thr	Gly	Ala	Pro	Leu	Asn
				165					170						175

Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	His	Gly	Gly	Lys	180	185	190
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	195	200	205
His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	210	215	220
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	225	230	235
Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	245	250	255
Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	260	265	270
Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	275	280	285
Ser	Asn	His	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	290	295	300
Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	305	310	315
Ile	Ala	Ser	Asn	His	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	325	330	335
Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	340	345	350
Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	355	360	365
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	370	375	380
Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	385	390	395
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	405	410	415
Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	420	425	430
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	435	440	445
Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Val	Ser	Asn	Gly	Gly	Gly	Lys	450	455	460
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	465	470	475
His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	485	490	495
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	500	505	510
Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	His	515	520	525
Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	530	535	540
Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	545	550	555
Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	565	570	575
Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	580	585	590
Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	595	600	605
Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	610	615	620
Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	625	630	635
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	645	650	655
Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	660	665	670
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	675	680	685

Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala		
	690					695					700						
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly		
705					710					715					720		
Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys		
				725					730						735		
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala		
			740					745					750				
His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly		
	755						760					765					
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys		
	770					775						780					
Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn		
785					790					795					800		
Asn	Gly	Gly	Arg	Pro	Ala	Leu	Glu	Ser	Ile	Val	Ala	Gln	Leu	Ser	Arg		
				805					810						815		
Pro	Asp	Pro	Ala	Leu	Ala	Ala	Leu	Thr	Asn	Asp	His	Leu	Val	Ala	Leu		
			820					825					830				
Ala	Cys	Leu	Gly	Gly	Arg	Pro	Ala	Leu	Asp	Ala	Val	Lys	Lys	Gly	Leu		
	835						840					845					
Pro	His	Ala	Pro	Ala	Leu	Ile	Lys	Arg	Thr	Asn	Arg	Arg	Ile	Pro	Glu		
	850					855					860						
Arg	Thr	Ser	His	Arg	Val	Ala	Gly	Ser	Gln	Leu	Val	Lys	Ser	Glu	Leu		
865					870					875					880		
Glu	Glu	Lys	Lys	Ser	Glu	Leu	Arg	His	Lys	Leu	Lys	Tyr	Val	Pro	His		
				885					890						895		
Glu	Tyr	Ile	Glu	Leu	Ile	Glu	Ile	Ala	Arg	Asn	Ser	Thr	Gln	Asp	Arg		
			900					905						910			
Ile	Leu	Glu	Met	Lys	Val	Met	Glu	Phe	Phe	Met	Lys	Val	Tyr	Gly	Tyr		
	915						920						925				
Arg	Gly	Lys	His	Leu	Gly	Gly	Ser	Arg	Lys	Pro	Asp	Gly	Ala	Ile	Tyr		
	930					935					940						
Thr	Val	Gly	Ser	Pro	Ile	Asp	Tyr	Gly	Val	Ile	Val	Asp	Thr	Lys	Ala		
945					950					955					960		
Tyr	Ser	Gly	Gly	Tyr	Asn	Leu	Pro	Ile	Gly	Gln	Ala	Asp	Glu	Met	Gln		
				965					970						975		
Arg	Tyr	Val	Glu	Glu	Asn	Gln	Thr	Arg	Asn	Lys	His	Ile	Asn	Pro	Asn		
			980					985					990				
Glu	Trp	Trp	Lys	Val	Tyr	Pro	Ser	Val	Thr	Glu	Phe	Lys	Phe	Leu			
	995						1000					1005					
Phe	Val	Ser	Gly	His	Phe	Lys	Gly	Asn	Tyr	Lys	Ala	Gln	Leu	Thr	Arg		
	1010					1015					1020						
Leu	Asn	His	Ile	Thr	Asn	Cys	Asn	Gly	Ala	Val	Leu	Ser	Val	Glu	Glu		
1025						1030				1035					1040		
Leu	Leu	Ile	Gly	Gly	Glu	Met	Ile	Lys	Ala	Gly	Thr	Leu	Thr	Leu	Glu		
				1045					1050						1055		
Glu	Val	Arg	Arg	Lys	Phe	Asn	Asn	Gly	Glu	Ile	Asn	Phe	Arg	Ser			
			1060					1065					1070				

<210> 4

<211> 1071

<212> PRT

<213> Искусственная последовательность

<220>

<223> CCR5-Uco-R

<400> 4

Met	Asp	Tyr	Lys	Asp	His	Asp	Gly	Asp	Tyr	Lys	Asp	His	Asp	Ile	Asp		
1				5					10					15			
Tyr	Lys	Asp	Asp	Asp	Asp	Lys	Met	Ala	Pro	Lys	Lys	Lys	Arg	Lys	Val		
			20					25					30				
Gly	Ile	His	Gly	Val	Pro	Ala	Ala	Val	Asp	Leu	Arg	Thr	Leu	Gly	Tyr		
		35					40						45				

Ser	Gln	Gln	Gln	Gln	Glu	Lys	Ile	Lys	Pro	Lys	Val	Arg	Ser	Thr	Val
50						55					60				
Ala	Gln	His	His	Glu	Ala	Leu	Val	Gly	His	Gly	Phe	Thr	His	Ala	His
65					70					75					80
Ile	Val	Ala	Leu	Ser	Gln	His	Pro	Ala	Ala	Leu	Gly	Thr	Val	Ala	Val
				85					90					95	
Lys	Tyr	Gln	Asp	Met	Ile	Ala	Ala	Leu	Pro	Glu	Ala	Thr	His	Glu	Ala
			100					105					110		
Ile	Val	Gly	Val	Gly	Lys	Gln	Trp	Ser	Gly	Ala	Arg	Ala	Leu	Glu	Ala
		115					120					125			
Leu	Leu	Thr	Val	Ala	Gly	Glu	Leu	Arg	Gly	Pro	Pro	Leu	Gln	Leu	Asp
	130					135					140				
Thr	Gly	Gln	Leu	Leu	Lys	Ile	Ala	Lys	Arg	Gly	Gly	Val	Thr	Ala	Val
145					150					155					160
Glu	Ala	Val	His	Ala	Trp	Arg	Asn	Ala	Leu	Thr	Gly	Ala	Pro	Leu	Asn
				165				170						175	
Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys
			180					185					190		
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala
	195						200					205			
His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Asn	Gly
	210					215					220				
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys
225					230					235					240
Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn
				245				250						255	
Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val
			260					265					270		
Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala
	275						280					285			
Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu
	290					295					300				
Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala
305					310					315					320
Ile	Ala	Ser	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg
			325					330						335	
Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val
			340					345					350		
Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	
	355					360					365				
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu
	370					375					380				
Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu
385					390					395					400
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr
				405				410						415	
Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala
			420					425					430		
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly
	435					440						445			
Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	His	Gly	Gly	Lys
	450					455					460				
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala
465					470					475					480
His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly
				485				490						495	
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys
			500					505					510		
Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	His
	515						520					525			
Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val
	530					535					540				
Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala
545					550					555					560

Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	565	570	575
Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	580	585	590
Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	595	600	605
Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	610	615	620
Val	Ala	Ile	Ala	Ser	Asn	His	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	625	630	635
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	645	650	655
Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	660	665	670
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	675	680	685
Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	690	695	700
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	705	710	715
Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	725	730	735
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	740	745	750
His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	755	760	765
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	770	775	780
Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	785	790	795
Gly	Gly	Gly	Arg	Pro	Ala	Leu	Glu	Ser	Ile	Val	Ala	Gln	Leu	Ser	Arg	805	810	815
Pro	Asp	Pro	Ala	Leu	Ala	Ala	Leu	Thr	Asn	Asp	His	Leu	Val	Ala	Leu	820	825	830
Ala	Cys	Leu	Gly	Gly	Arg	Pro	Ala	Leu	Asp	Ala	Val	Lys	Lys	Gly	Leu	835	840	845
Pro	His	Ala	Pro	Ala	Leu	Ile	Lys	Arg	Thr	Asn	Arg	Arg	Ile	Pro	Glu	850	855	860
Arg	Thr	Ser	His	Arg	Val	Ala	Gly	Ser	Gln	Leu	Val	Lys	Ser	Glu	Leu	865	870	875
Glu	Glu	Lys	Lys	Ser	Glu	Leu	Arg	His	Lys	Leu	Lys	Tyr	Val	Pro	His	885	890	895
Glu	Tyr	Ile	Glu	Leu	Ile	Glu	Ile	Ala	Arg	Asn	Ser	Thr	Gln	Asp	Arg	900	905	910
Ile	Leu	Glu	Met	Lys	Val	Met	Glu	Phe	Phe	Met	Lys	Val	Tyr	Gly	Tyr	915	920	925
Arg	Gly	Lys	His	Leu	Gly	Gly	Ser	Arg	Lys	Pro	Asp	Gly	Ala	Ile	Tyr	930	935	940
Thr	Val	Gly	Ser	Pro	Ile	Asp	Tyr	Gly	Val	Ile	Val	Asp	Thr	Lys	Ala	945	950	955
Tyr	Ser	Gly	Gly	Tyr	Asn	Leu	Pro	Ile	Gly	Gln	Ala	Asp	Glu	Met	Gln	965	970	975
Arg	Tyr	Val	Glu	Glu	Asn	Gln	Thr	Arg	Asn	Lys	His	Ile	Asn	Pro	Asn	980	985	990
Glu	Trp	Trp	Lys	Val	Tyr	Pro	Ser	Ser	Val	Thr	Glu	Phe	Lys	Phe	Leu	995	1000	1005
Phe	Val	Ser	Gly	His	Phe	Lys	Gly	Asn	Tyr	Lys	Ala	Gln	Leu	Thr	Arg	1010	1015	1020
Leu	Asn	His	Ile	Thr	Asn	Cys	Asn	Gly	Ala	Val	Leu	Ser	Val	Glu	Glu	1025	1030	1035
Leu	Leu	Ile	Gly	Gly	Glu	Met	Ile	Lys	Ala	Gly	Thr	Leu	Thr	Leu	Glu	1045	1050	1055
Glu	Val	Arg	Arg	Lys	Phe	Asn	Asn	Gly	Glu	Ile	Asn	Phe	Arg	Ser		1060	1065	1070

<210> 5
<211> 34
<212> ПРТ
<213> Искусственная последовательность

<220>
<223> TALEN-единица повтора

<220>
<221> SITE
<222> 4
<223> Xaa = Любая аминокислота, или E, Q, D или A

<220>
<221> SITE
<222> 10
<223> Xaa = Любая аминокислота, или A или V

<220>
<221> SITE
<222> 12..13
<223> Вариабельные последовательности из двух аминокислотных остатков (RVD)

<220>
<221> SITE
<222> 32
<223> Xaa = Любая аминокислота, или A или D

<400> 5
Leu Thr Pro Xaa Gln Val Val Ala Ile Xaa Ser Xaa Xaa Gly Gly Lys
1 5 10 15
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Xaa
20 25 30
His Gly

<210> 6
<211> 34
<212> ПРТ
<213> Искусственная последовательность

<220>
<223> TALEN-единица повтора

<220>
<221> SITE
<222> 12..13
<223> Вариабельные последовательности из двух аминокислотных остатков (RVD)

<400> 6
Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser Xaa Xaa Gly Gly Lys
1 5 10 15
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala
20 25 30
His Gly

<210> 7
<211> 34
<212> ПРТ
<213> Искусственная последовательность

<220>

<223> TALEN-единица повтора

<220>

<221> SITE

<222> 12..13

<223> Вариабельные последовательности из двух аминокислотных остатков (RVD)

<400> 7

Leu	Thr	Pro	Gln	Gln	Val	Val	Ala	Ile	Ala	Ser	Xaa	Xaa	Gly	Gly	Lys
1			5				10						15		
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala
			20				25					30			
His	Gly														

<210> 8

<211> 34

<212> ПРТ

<213> Искусственная последовательность

<220>

<223> TALEN-единица повтора

<220>

<221> SITE

<222> 12..13

<223> Вариабельные последовательности из двух аминокислотных остатков (RVD)

<400> 8

Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Xaa	Xaa	Gly	Gly	Lys
1			5				10						15		
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp
			20				25					30			
His	Gly														

<210> 9

<211> 34

<212> ПРТ

<213> Искусственная последовательность

<220>

<223> TALEN-единица повтора

<220>

<221> SITE

<222> 12..13

<223> Вариабельные последовательности из двух аминокислотных остатков (RVD)

<400> 9

Leu	Thr	Pro	Ala	Gln	Val	Val	Ala	Ile	Ala	Ser	Xaa	Xaa	Gly	Gly	Lys
1			5				10						15		
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp
			20				25					30			
His	Gly														

<210> 10

<211> 34

<212> ПРТ

<213> Искусственная последовательность

<220>

<223> TALEN-единица повтора

<220>
 <221> SITE
 <222> 12..13
 <223> Вариабельные последовательности из двух аминокислотных остатков (RVD)

<400> 10
 Leu Thr Pro Glu Gln Val Val Ala Ile Val Ser Xaa Xaa Gly Gly Lys
 1 5 10 15
 Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala
 20 25 30
 His Gly

<210> 11
 <211> 1059
 <212> ДНК
 <213> Homo sapiens

<220>
 <223> hCCR5 ДНК

<400> 11
 atggattatc aagtgtcaag tccaatctat gacatcaatt attatacatc ggagccctgc 60
 caaaaaatca atgtgaagca aatcgcagcc cgcctcctgc ctccgctcta ctcaactggtg 120
 ttcatctttg gttttgtggg caacatgctg gtcactcctca tcctgataaa ctgcaaaagg 180
 ctgaagagca tgactgacat ctacctgctc aacctggcca tctctgacct gtttttcctt 240
 cttactgtcc ctttctgggc tcactatgct gccgccagc gggactttgg aaatacaatg 300
 tgtcaactct tgacagggct ctattttata ggcttcttct ctggaatctt cttcatcatc 360
 ctctgacaa tcgataggta cctggctgtc gtccatgctg tgtttgcttt aaaagccagg 420
 acggtcacct ttgggggtggg gacaagtgtg atcacttggg tgggtggctgt gtttgcgctc 480
 ctcccaggaa tcatctttac cagatctcaa aaagaaggtc ttcattacac ctgcagctct 540
 cattttccat acagtcagta tcaattctgg aagaatttcc agacattaaa gatagtcac 600
 ttgggggctgg tcctgccgct gcttgtcatg gtcactgtgt actcgggaat cctaaaaact 660
 ctgcttcggg gtcgaaatga gaagaagagg cacagggtgt tgaggcttat cttcaccatc 720
 atgattgttt attttctctt ctgggctccc tacaacattg tccttctcct gaacaccttc 780
 caggaattct ttggcctgaa taattgcagt agctctaaca ggttggacca agctatgcag 840
 gtgacagaga ctcttgggat gacgcactgc tgcacaaacc ccatcatcta tgcctttgtc 900
 ggggagaagt tcagaaacta cctcttagtc ttcttccaaa agcacattgc caaacgcttc 960
 tgcaaatgct gttctatctt ccagcaagag gctcccgagc gagcaagctc agtttacacc 1020
 cgatccactg gggagcagga aatatctgtg ggcttgtga 1059

<210> 12
 <211> 198
 <212> ПРТ
 <213> Planomicrobium okeanokoites

<220>

<223> FokI - домен расщепления

<400> 12

Gln	Leu	Val	Lys	Ser	Glu	Leu	Glu	Glu	Lys	Lys	Ser	Glu	Leu	Arg	His
1				5					10					15	
Lys	Leu	Lys	Tyr	Val	Pro	His	Glu	Tyr	Ile	Glu	Leu	Ile	Glu	Ile	Ala
			20					25					30		
Arg	Asn	Ser	Thr	Gln	Asp	Arg	Ile	Leu	Glu	Met	Lys	Val	Met	Glu	Phe
		35					40					45			
Phe	Met	Lys	Val	Tyr	Gly	Tyr	Arg	Gly	Lys	His	Leu	Gly	Gly	Ser	Arg
	50					55					60				
Lys	Pro	Asp	Gly	Ala	Ile	Tyr	Thr	Val	Gly	Ser	Pro	Ile	Asp	Tyr	Gly
65					70					75					80
Val	Ile	Val	Asp	Thr	Lys	Ala	Tyr	Ser	Gly	Gly	Tyr	Asn	Leu	Pro	Ile
				85					90					95	
Gly	Gln	Ala	Asp	Glu	Met	Gln	Arg	Tyr	Val	Glu	Glu	Asn	Gln	Thr	Arg
			100					105					110		
Asn	Lys	His	Ile	Asn	Pro	Asn	Glu	Trp	Trp	Lys	Val	Tyr	Pro	Ser	Ser
		115					120					125			
Val	Thr	Glu	Phe	Lys	Phe	Leu	Phe	Val	Ser	Gly	His	Phe	Lys	Gly	Asn
	130					135						140			
Tyr	Lys	Ala	Gln	Leu	Thr	Arg	Leu	Asn	His	Ile	Thr	Asn	Cys	Asn	Gly
145					150					155					160
Ala	Val	Leu	Ser	Val	Glu	Glu	Leu	Leu	Ile	Gly	Gly	Glu	Met	Ile	Lys
				165					170					175	
Ala	Gly	Thr	Leu	Thr	Leu	Glu	Glu	Val	Arg	Arg	Lys	Phe	Asn	Asn	Gly
			180					185					190		
Glu	Ile	Asn	Phe	Arg	Ser										
		195													

<210> 13

<211> 1037

<212> ППТ

<213> Искусственная последовательность

<220>

<223> CCR5-Mco-L

<400> 13

Met	Asp	Tyr	Lys	Asp	His	Asp	Gly	Asp	Tyr	Lys	Asp	His	Asp	Ile	Asp
1				5					10					15	
Tyr	Lys	Asp	Asp	Asp	Asp	Lys	Met	Ala	Pro	Lys	Lys	Lys	Arg	Lys	Val
			20					25					30		
Gly	Ile	His	Gly	Val	Pro	Ala	Ala	Val	Asp	Leu	Arg	Thr	Leu	Gly	Tyr
		35					40					45			
Ser	Gln	Gln	Gln	Gln	Glu	Lys	Ile	Lys	Pro	Lys	Val	Arg	Ser	Thr	Val
	50					55					60				
Ala	Gln	His	His	Glu	Ala	Leu	Val	Gly	His	Gly	Phe	Thr	His	Ala	His
65					70					75					80
Ile	Val	Ala	Leu	Ser	Gln	His	Pro	Ala	Ala	Leu	Gly	Thr	Val	Ala	Val
				85					90					95	
Lys	Tyr	Gln	Asp	Met	Ile	Ala	Ala	Leu	Pro	Glu	Ala	Thr	His	Glu	Ala
			100					105					110		
Ile	Val	Gly	Val	Gly	Lys	Gln	Trp	Ser	Gly	Ala	Arg	Ala	Leu	Glu	Ala
		115					120					125			
Leu	Leu	Thr	Val	Ala	Gly	Glu	Leu	Arg	Gly	Pro	Pro	Leu	Gln	Leu	Asp
	130					135					140				
Thr	Gly	Gln	Leu	Leu	Lys	Ile	Ala	Lys	Arg	Gly	Gly	Val	Thr	Ala	Val
145					150					155					160
Glu	Ala	Val	His	Ala	Trp	Arg	Asn	Ala	Leu	Thr	Gly	Ala	Pro	Leu	Asn
				165					170					175	
Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Asn	Gly	Gly	Lys

			180						185						190	
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	
		195				200						205				
His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	
		210				215					220					
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	
225					230					235						
Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	
				245					250			255				
Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	
		260						265					270			
Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	
		275				280						285				
Ser	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	
		290				295					300					
Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	
305					310					315						
Ile	Ala	Ser	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	
				325					330			335				
Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	
		340						345					350			
Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	
		355				360						365				
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	
		370				375					380					
Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	
385					390					395						
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	
				405					410			415				
Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	
		420						425					430			
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	
		435				440						445				
Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	
		450				455					460					
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	
465					470					475						
His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	
				485					490			495				
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	
		500						505					510			
Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	
		515				520						525				
Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	
		530				535					540					
Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	
545					550					555						
Ser	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val</					

690	695	700
Leu Glu Thr Val Gln Arg	Leu Leu Pro Val Leu Cys Gln Ala His Gly	
705	710	715
Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys		720
	725	730
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala		735
	740	745
His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser His Asp Gly		750
	755	760
Gly Arg Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro Asp		765
	770	775
Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys		780
	785	790
Leu Gly Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro His		795
	805	810
Ala Pro Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg Thr		815
	820	825
Ser His Arg Val Ala Gly Ser Gln Leu Val Lys Ser Glu Leu Glu Glu		830
	835	840
Lys Lys Ser Glu Leu Arg His Lys Leu Lys Tyr Val Pro His Glu Tyr		845
	850	855
Ile Glu Leu Ile Glu Ile Ala Arg Asn Ser Thr Gln Asp Arg Ile Leu		860
	865	870
Glu Met Lys Val Met Glu Phe Phe Met Lys Val Tyr Gly Tyr Arg Gly		875
	885	890
Lys His Leu Gly Gly Ser Arg Lys Pro Asp Gly Ala Ile Tyr Thr Val		895
	900	905
Gly Ser Pro Ile Asp Tyr Gly Val Ile Val Asp Thr Lys Ala Tyr Ser		910
	915	920
Gly Gly Tyr Asn Leu Pro Ile Gly Gln Ala Asp Glu Met Gln Arg Tyr		925
	930	935
Val Glu Glu Asn Gln Thr Arg Asn Lys His Ile Asn Pro Asn Glu Trp		940
	945	950
Trp Lys Val Tyr Pro Ser Ser Val Thr Glu Phe Lys Phe Leu Phe Val		955
	965	970
Ser Gly His Phe Lys Gly Asn Tyr Lys Ala Gln Leu Thr Arg Leu Asn		975
	980	985
His Ile Thr Asn Cys Asn Gly Ala Val Leu Ser Val Glu Glu Leu Leu		990
	995	1000
Ile Gly Gly Glu Met Ile Lys Ala Gly Thr Leu Thr Leu Glu Glu Val		1005
	1010	1015
Arg Arg Lys Phe Asn Asn Gly Glu Ile Asn Phe Arg Ser		1020
	1025	1030
		1035

<210> 14

<211> 1037

<212> ПРТ

<213> Искусственная последовательность

<220>

<223> CCR5-Mco-R

<400> 14

Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp	
1	5
Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val	
	20
Gly Ile His Gly Val Pro Ala Ala Val Asp Leu Arg Thr Leu Gly Tyr	
	35
Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg Ser Thr Val	
	50
Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr His Ala His	
	65
Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr Val Ala Val	
	70
	75
	80

				85				90				95				
Lys	Tyr	Gln	Asp	Met	Ile	Ala	Ala	Leu	Pro	Glu	Ala	Thr	His	Glu	Ala	
				100					105					110		
Ile	Val	Gly	Val	Gly	Lys	Gln	Trp	Ser	Gly	Ala	Arg	Ala	Leu	Glu	Ala	
				115					120					125		
Leu	Leu	Thr	Val	Ala	Gly	Glu	Leu	Arg	Gly	Pro	Pro	Leu	Gln	Leu	Asp	
				130					135					140		
Thr	Gly	Gln	Leu	Leu	Lys	Ile	Ala	Lys	Arg	Gly	Gly	Val	Thr	Ala	Val	
145					150					155					160	
Glu	Ala	Val	His	Ala	Trp	Arg	Asn	Ala	Leu	Thr	Gly	Ala	Pro	Leu	Asn	
				165					170					175		
Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	
				180					185					190		
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	
				195					200					205		
His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	
				210					215					220		
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	
225					230					235					240	
Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	
				245					250					255		
Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	
				260					265					270		
Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	
				275					280					285		
Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	
				290					295					300		
Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	
305					310					315					320	
Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	
				325					330					335		
Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	
				340					345					350		
Val	Ala	Ile	Ala	Ser	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	
				355					360					365		
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	
				370					375					380		
Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	
385					390					395					400	
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	
				405					410					415		
Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	
				420					425					430		
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	
				435					440					445		
Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	
				450					455					460		
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	
465					470					475					480	
His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	
				485					490					495		

<210> 16
<211> 18
<212> ПРТ
<213> Homo sapiens

<400> 16
Cys Thr Thr Cys Ala Gly Cys Cys Thr Thr Thr Thr Gly Cys Ala Gly
1 5 10 15
Thr Thr

<210> 17
<211> 3788
<212> ДНК
<213> Искусственная последовательность

<220>
<223> CCR5-Uco-L mRNA

<400> 17
ggggccacca uggасиаиаа ggaccaccgac ggagacuaаа aggaucauga uauugauuac 60
aaagacgaug acgauaagau ggccccaааg aagaagcgga aggucggauu ccacggaguc 120
ccagcagccg uagauiiugag aacuiiugga uaiiucacagc agcagcagga aaagaucaag 180
cccaaguga ggucgacagu cgcgcagcau caggaagcgc ugguggguca uggguuuaca 240
caugcccaca ucguagccuu gucgcagcac ccugcagccc uuggcacggu cgcggucaag 300
uaccaggaca ugauugcggc guugccggaа gccacacaug aggcgaucgu cggugugggg 360
aaacagugga gcggagcccc agcgcuiugag gccugiuuga cggucgcggg agagcugaga 420
gggccucccc uucagcugga cacgggccag uugcugaaga ucgcgaagcg gggaggaguc 480
acggcgguugc aggcggugca cgcguggcgc aaugcgcuca cgggagcacc ccucaaccug 540
accccagagc aggucguggc aauiugcagc aaccacgggg gaaagcaggc acucgaaacc 600
guccagaggu ugcugccugu gcugugccaa gcgcacggac uuacgccaga gcaggucgug 660
gcaauugcga gccaugacgg gggaaagcag gcacucgaaa ccguccagag guugcugccu 720
gugcugugcc aagcgcacgg acuaaccca gagcaggucg uggcaauugc gagcaacgga 780
gggggaaagc aggcacucga aaccguccag agguugcugc cugugcugug ccaagcgcac 840
ggguugaccc cagagcaggu cguggcauu gcgagcaacc acgggggaaa gcaggcacuc 900
gaaaccgucc agagguugcu gccugugcug ugccaagcgc acggccugac cccagagcag 960
gucguggcaa uugcgagcaa ccacggggga aagcaggcac ucgaaaccgu ccagagguug 1020
cugccugugc ugugccaagc gcacggacug acaccagagc aggucguggc aauiugcagc 1080
aacggagggg gaaagcaggc acucgaaacc guccagaggu ugcugccugu gcugugccaa 1140
gcgcacggac uuacacccga acaagucgug gcaauugcga gccaugacgg gggaaagcag 1200
gcacucgaaa ccguccagag guugcugccu gugcugugcc aagcgcacgg acuiacgcca 1260
gagcaggucg uggcaauugc gagcaacauc gggggaaagc aggcacucga aaccguccag 1320

agguugcugc	cugugcugug	ccaagcgcac	ggacuaaccc	cagagcaggu	cguggcaauu	1380
gugagcaacg	gagggggaaa	gcaggcacuc	gaaaccgucc	agagguugcu	gccugugcug	1440
ugccaagcgc	acggguugac	cccagagcag	gucguggcaa	uugcgagcca	ugacggggga	1500
aagcaggcac	ucgaaaccgu	ccagagguug	cugccugugc	ugugccaagc	gcacggccug	1560
acccagagc	aggucguggc	aauugcgagc	caugacgggg	gaaagcaggc	acucgaaacc	1620
guccagaggu	ugcugccugu	gcugugccaa	gcgcacggac	ugacaccaga	gcaggucgug	1680
gcaauugcga	gcaacggagg	gggaaagcag	gcacucgaaa	ccguccagag	guugcugccu	1740
gugcugugcc	aagcgcacgg	ccucacccca	gagcaggucg	uggcaauugc	gagccaugac	1800
gggggaaagc	aggcacucga	aaccguccag	agguugcugc	cugugcugug	ccaagcgcac	1860
ggacuacgc	cagagcaggu	cguggcaauu	gcgagcaaca	ucgggggaaa	gcaggcacuc	1920
gaaaccgucc	agagguugcu	gccugugcug	ugccaagcgc	acggacuaac	cccagagcag	1980
gucguggcaa	uugcgagcaa	cggaggggga	aagcaggcac	ucgaaaccgu	ccagagguug	2040
cugccugugc	ugugccaagc	gcacggguug	acccagagc	aggucguggc	aauugcgagc	2100
caugacgggg	gaaagcaggc	acucgaaacc	guccagaggu	ugcugccugu	gcugugccaa	2160
gcgcacggcc	ugaccccaga	gcaggucgug	gcaauugcga	gccaugacgg	gggaaagcag	2220
gcacucgaaa	ccguccagag	guugcugccu	gugcugugcc	aagcgcacgg	acugacacca	2280
gagcaggucg	uggcaauugc	gagcaacgga	gggggaaagc	aggcacucga	aaccguccag	2340
agguugcugc	cugugcugug	ccaagcgcac	ggacucacgc	cugagcaggu	aguggcuauu	2400
gcauccaaca	acgggggagc	acccgcacug	gagucaaucg	uggcccagcu	uucgaggccg	2460
gaccccgcg	uggccgcacu	cacuaaugau	caucuuguag	cgcuggccug	ccucggcgga	2520
cgacccgccu	uggaugcggu	gaagaagggg	cucccgcacg	cgccugcauu	gauuaagcgg	2580
accaacagaa	ggauucccga	gaggacauca	caucgagugg	cagguuccca	acucgugaag	2640
agugaacuug	aggagaaaaa	gucggagcug	cggcacaaau	ugaaauacgu	accgcaugaa	2700
uacaucgaac	uuaucgaaau	ugcuaggaac	ucgacucaag	acagaauccu	ugagaugaag	2760
guaauggagu	ucuuaugaa	gguuaugga	uaccgagggg	agcaucucgg	uggaucacga	2820
aaacccgacg	gagcaaucua	uacggugggg	agcccgauug	auuacggagu	gaucgucgac	2880
acgaaagccu	acagcggugg	guacaauuu	cccaucgggc	aggcagauga	gaugcaacgu	2940
uaugucgaag	aaaauacagc	caggaacaaa	cacaucaauc	caaaugagug	guggaaagug	3000
uauccuucan	cagugaccga	guuaaaguuu	uuguuugucu	cugggcuuuu	caaaggcaac	3060
uauaaggccc	agcucacacg	guugaaucc	auuacgaacu	gcaauggugc	gguuuugucc	3120
guagaggaac	ugcucauugg	uggagaaaug	aucaaagcgg	gaacucugac	acuggaagaa	3180
gucagacgca	aguuaaaca	uggcgagauc	aauuuccgcu	cauaaaaaau	cagccucgac	3240

ugugccuucu aguugccagc caucuguugu uugccccucc cccgugccuu ccuugacccu	3300
ggaaggugcc acucccacug uccuuuccua auaaaaugag gaaaauugcau cacaacacuc	3360
aaccsuaucu cggucuaauc uuuugauuuu uaagggauuu ugccgauuuc ggccuaauugg	3420
uuaaaaaauug agcugauuuu acaaaaauuu aacgcgaauu aaucuguggg aauguguguc	3480
aguuagggug uggaagucc ccaggcuccc cagcaggcag aaguaugcaa agcaugcauc	3540
ucauuuaguc agcaaccagg uguggaaagu cccaggcuc cccagcaggc agaaguaugc	3600
aaagcaugca ucuaauuag ucagcaacca uagucccgcc ccuaacuccg cccaucuccg	3660
cccuaacucc gcccaguucc gcccauucuc cgccccaugg cugacuaauu uuuuuuuuuu	3720
augcagaggc cgaggccgcc ucugccucug agcuauucca gaaguaguga ggaggcuuuu	3780
uuggaggc	3788

<210> 18

<211> 3788

<212> ДНК

<213> Искусственная последовательность

<220>

<223> CCR5-Uco-R mRNA

<400> 18

ggggccacca uggacuauaa ggaccacgac ggagacuaca aggaucauga uauugauuac	60
aaagacgaug acgaauagau ggccccaag aagaagcgga aggucggauu ccacggaguc	120
ccagcagccg uagauuugag aacuuuggga uauucacagc agcagcagga aaagaucaag	180
cccaaaguga ggucgacagu cgcgcagcau cacgaagcgc ugguggguca uggguuuaca	240
caugcccaca ucguagccuu gucgcagcac ccugcagccc uuggcacggu cgccgucaag	300
uaccaggaca ugaauugcggc guugccgga gccacacaug aggcgaucgu cggugugggg	360
aaacagugga gcggagcccc agcgcuuag gcccuguuga cggucgcggg agagcugaga	420
gggccucccc uucagcugga cacgggccag uugcugaaga ucgcgaagcg gggaggaguc	480
acggcgguuc aggcggugca cgcguggcgc aaugcgcuca cgggagcacc ccucaaccug	540
accccagagc aggucguggc aaugcgcagc aacaucgggg gaaagcaggc acucgaaacc	600
guccagaggu ugcugccugu gcugugccaa gcgcacggac uuacgccaga gcaggucgug	660
gcaauugcga gcaacaacgg gggaaagcag gcacucgaaa ccguccagag guugcugccu	720
gugcugugcc aagcgcacgg acuaacccca gagcaggucg uggcaauugc gagcaacauc	780
gggggaaagc aggcacucga aaccguccag agguugcugc cugugcugug ccaagcgcac	840
ggguugaccc cagagcaggu cguggcaau gcgagcaacg gagggggaaa gcaggcacuc	900
gaaaccgucc agagguugcu gccugugcug ugccaagcgc acggccugac cccagagcag	960
gucguggcaa uugcgagcaa caacggggga aagcaggcac ucgaaaccgu ccagagguug	1020

cugccugugc	ugugccaagc	gcacggacug	acaccagagc	aggucguggc	aaugcgagc	1080
aacggagggg	gaaagcaggc	acucgaaacc	guccagaggu	ugcugccugu	gcugugccaa	1140
gcgcacggac	uuacacccga	acaagucgug	gcaauugcga	gccaugacgg	gggaaagcag	1200
gcacucgaaa	ccguccagag	guugcugccu	gugcugugcc	aagcgcacgg	acuuacgcca	1260
gagcaggucg	uggcaauugc	gagcaacauc	gggggaaagc	aggcacucga	aaccguccag	1320
agguugcugc	cugugcugug	ccaagcgcac	ggacuaaccc	cagagcaggu	cguggcaauu	1380
gcgagcaacc	acgggggaaa	gcaggcacuc	gaaaccgucc	agagguugcu	gccugugcug	1440
ugccaagcgc	acggguugac	cccagagcag	gucguggcaa	uugcgagcaa	cggaggggga	1500
aagcaggcac	ucgaaaccgu	ccagagguug	cugccugugc	ugugccaagc	gcacggccug	1560
accccagagc	aggucguggc	aaugcgagc	caugacgggg	gaaagcaggc	acucgaaacc	1620
guccagaggu	ugcugccugu	gcugugccaa	gcgcacggac	ugacaccaga	gcaggucgug	1680
gcaauugcga	gcaacaucgg	gggaaagcag	gcacucgaaa	ccguccagag	guugcugccu	1740
gugcugugcc	aagcgcacgg	ccucacccca	gagcaggucg	uggcaauugc	gagcaacgga	1800
gggggaaagc	aggcacucga	aaccguccag	agguugcugc	cugugcugug	ccaagcgcac	1860
ggacuacgc	cagagcaggu	cguggcaauu	gcgagcaacc	acgggggaaa	gcaggcacuc	1920
gaaaccgucc	agagguugcu	gccugugcug	ugccaagcgc	acggacuaac	cccagagcag	1980
gucguggcaa	uugcgagcca	ugacggggga	aagcaggcac	ucgaaaccgu	ccagagguug	2040
cugccugugc	ugugccaagc	gcacggguug	accccagagc	aggucguggc	aaugcgagc	2100
aacggagggg	gaaagcaggc	acucgaaacc	guccagaggu	ugcugccugu	gcugugccaa	2160
gcgcacggcc	ugaccccaga	gcaggucgug	gcaauugcga	gccaugacgg	gggaaagcag	2220
gcacucgaaa	ccguccagag	guugcugccu	gugcugugcc	aagcgcacgg	acugacacca	2280
gagcaggucg	uggcaauugc	gagcaacgga	gggggaaagc	aggcacucga	aaccguccag	2340
agguugcugc	cugugcugug	ccaagcgcac	ggacucacgc	cugagcaggu	aguggcuauu	2400
gcauccaacg	gagggggcag	acccgcacug	gagucaaucg	uggcccagcu	uucgaggccg	2460
gaccccgcg	uggccgcacu	cacuaaugau	caucuuguag	cgcuggccug	ccucggcgga	2520
cgacccgccu	uggaugcggu	gaagaagggg	cucccgcacg	cgccugcauu	gauuaagcgg	2580
accaacagaa	ggauucccga	gaggacauca	caucgagugg	cagguuccca	acucgugaag	2640
agugaacuug	aggagaaaaa	gucggagcug	cggcacaaau	ugaaaauacgu	accgcaugaa	2700
uacaucgaac	uuaucgaaau	ugcuaggaac	ucgacucaag	acagaauccu	ugagaugaag	2760
guaauggagu	ucuuaugaa	gguuaugga	uaccgagggg	agcaucucgg	uggaucacga	2820
aaacccgacg	gagcaaucua	uacggugggg	agcccgauug	auuacggagu	gaucgucgac	2880
acgaaagccu	acagcggugg	guacaauuu	cccaucgggc	aggcagauga	gaugcaacgu	2940

uaugucgaag	aaaaucagac	caggaacaaa	cacaucaauc	caaaugagug	guggaaagug	3000
uauccuucan	cagugaccga	guuuuaguuu	uuguuugucu	cugggcuuuu	caaaggcaac	3060
uauaaggccc	agcucacacg	guugaaucac	auuacgaacu	gcaauggugc	gguuuugucc	3120
guagaggaac	ugcucauugg	uggagaaaug	aucaaagcgg	gaacucugac	acuggaagaa	3180
gucagacgca	aguuuuacaa	uggcgagauc	aauuuccgcu	cauaaaaaau	cagccucgac	3240
ugugccuucu	aguugccagc	caucuguugu	uugccccucc	cccuggccuu	ccuugacccu	3300
ggaaggugcc	acucccacug	uccuuuccua	auaaaaugag	gaaaauugcau	cacaacacuc	3360
aaccuauucu	cggucuauuc	uuuugauuuu	uaagggaauu	ugccgauuuc	ggccuauugg	3420
uuaaaaaaug	agcugauuuu	acaaaaauuu	aacgcgaauu	aaucucugug	aauguguguc	3480
aguuaaggug	uggaaagucc	ccaggcuccc	cagcaggcag	aaguaugcaa	agcaugcauc	3540
ucaauuaguc	agcaaccagg	uguggaaagu	ccccaggcuc	cccagcaggc	agaaguaugc	3600
aaagcaugca	ucucaauuag	ucagcaacca	uagucccgcc	ccuaacuccg	cccaucccg	3660
cccuaacucc	gcccaguucc	gcccauucuc	cgccccaugg	cugacuaauu	uuuuuuuuuu	3720
augcagaggc	cgaggccgcc	ucugccucug	agcuauucca	gaaguaguga	ggaggcuuuu	3780
uuggaggc						3788

Формула изобретения

1. Пара TAL-эффекторных нуклеаз, включающая в себя первый и второй мономер TAL-эффекторной нуклеазы, причем каждый мономер TAL-эффекторной нуклеазы включает в себя эндонуклеазный домен со II типом эндонуклеазной активности и ДНК-связывающий домен TAL-эффектора имеющий множество единиц повтора, каждая из которых имеет пару переменных аминокислот (Repeat Variable Diresidue, RVD), и при этом

а) ДНК-связывающий домен TAL-эффектора первого мономера TAL-эффекторной нуклеазы связывается с последовательностью-мишенью GCTGGTCATCCTCATCCTG (SEQ ID NO: 1) и/или содержит RVD-последовательность NH HD NG NH NH NG HD NI NG HD HD NG HD NI NG HD HD NG NN,

и

б) ДНК-связывающий домен TAL-эффектора второго мономера TAL-эффекторной нуклеазы связывается с последовательностью-мишенью AGATGTCAGTCATGCTCTT (SEQ ID NO: 2) и/или содержит RVD-последовательность NI NN NI NG NN NG HD NI NH NG HD NI NG NH HD NG HD NG NG.

2. Пара TAL-эффекторных нуклеаз по п.1, в которой эндонуклеазный домен в каждом мономере TAL-эффекторной нуклеазы является С-концевым по отношению к ДНК-связывающему домену TAL-эффектора и каждый повтор, за исключением того, который примыкает непосредственно к эндонуклеазному домену, содержит от 33 до 35 аминокислот, предпочтительно 34 аминокислоты, в которых RVD находятся в положениях 12 и 13 в каждом повторе

3. Пара TAL-эффекторных нуклеаз по пп.1 или 2, в которой эндонуклеазный домен мономеров TAL-эффекторных нуклеаз является ДНК-расщепляющим доменом эндонуклеазы FokI.

4. Пара TAL-эффекторных нуклеаз по пп.1-3, в которой первый мономер TAL-эффекторной нуклеазы содержит аминокислотную последовательность, соответствующую SEQ ID NO: 3, а второй мономер TAL-эффекторной нуклеазы содержит аминокислотную последовательность, соответствующую SEQ ID NO: 4.

5. Нуклеиновая кислота, включающая в себя:

а) первую нуклеиновую кислоту, которая кодирует первый мономер TAL-эффекторной нуклеазы, где первый мономер TAL-эффекторной нуклеазы содержит эндонуклеазный домен со II типом эндонуклеазной активности и ДНК-связывающий

домен TAL-эффектора с множеством единиц повтора, каждая из которых имеет пару переменных аминокислот (RVD), и при этом ДНК-связывающий домен TAL-эффектора связывается с последовательностью-мишенью GCTGGTCATCCTCATCCTG (SEQ ID NO: 1) и/или содержит RVD-последовательность NH HD NG NH NH NG HD NI NG HD HD NG HD NI NG HD HD NG NN, и

б) вторую нуклеиновую кислоту, которая кодирует второй мономер TAL-эффекторной нуклеазы, где второй мономер TAL-эффекторной нуклеазы содержит эндонуклеазный домен со II типом эндонуклеазной активности и ДНК-связывающий домен TAL-эффектора с множеством единиц повтора, каждая из которых имеет переменные последовательности из двух аминокислотных остатков (RVD), и при этом ДНК-связывающий домен TAL-эффектора связывается с последовательностью-мишенью AGATGTCAGTCATGCTCTT (SEQ ID NO: 2) и/или содержит RVD-последовательность NI NN NI NG NN NG HD NI NH NG HD NI NG NH HD NG HD NG NG.

6. Вектор, включающий в себя нуклеиновую кислоту по п.5.

7. Комбинация нуклеиновых кислот, включающая в себя

а) первую нуклеиновую кислоту, которая кодирует первый мономер TAL-эффекторной нуклеазы, где первый мономер TAL-эффекторной нуклеазы содержит эндонуклеазный домен со II типом эндонуклеазной активности и ДНК-связывающий домен TAL-эффектора со множеством единиц повтора, каждая из которых имеет пару переменных аминокислот (RVD), и при этом ДНК-связывающий домен TAL-эффектора связывается с последовательностью-мишенью GCTGGTCATCCTCATCCTG (SEQ ID NO: 1) и/или содержит RVD-последовательность NH HD NG NH NH NG HD NI NG HD HD NG HD NI NG HD HD NG NN, и

б) вторую нуклеиновую кислоту, которая кодирует второй мономер TAL-эффекторной нуклеазы, где второй мономер TAL-эффекторной нуклеазы содержит эндонуклеазный домен со II типом эндонуклеазной активности и ДНК-связывающий домен TAL-эффектора со множеством единиц повтора, каждая из которых имеет пару переменных аминокислот (RVD), и при этом ДНК-связывающий домен TAL-эффектора связывается с последовательностью-мишенью AGATGTCAGTCATGCTCTT (SEQ ID NO: 2) и/или содержит RVD-последовательность NI NN NI NG NN NG HD NI NH NG HD NI NG NH HD NG HD NG NG.

8. Комбинация нуклеиновых кислот по п.7, в которой первая нуклеиновая кислота является первой мРНК и вторая нуклеиновая кислота является второй мРНК.

9. Комбинация нуклеиновых кислот по п.8, в которой первая нуклеиновая кислота включает последовательность, соответствующую SEQ ID NO:17, а вторая нуклеиновая кислота включает последовательность, соответствующую SEQ ID NO:18.

10. Изолированная клетка-хозяин, включающая в себя нуклеиновую кислоту по п.5, комбинацию нуклеиновых кислот по пп.7–9 или вектор по п. 6, при условии, что клетка-хозяин не является зародышевой клеткой человека или эмбриональной зародышевой клеткой человека и не является эмбриональной стволовой клеткой, в процессе получения которых был или будет разрушен эмбрион человека

11. Фармацевтическая композиция, включающая в себя нуклеиновую кислоту по п.5, комбинацию нуклеиновых кислот по пп.7 - 9 или вектор по п.6.

12. Лекарственное средство, включающее в себя нуклеиновую кислоту по п. 5, комбинацию нуклеиновых кислот по любому из пп.7 - 9 или вектор по п.6 или фармацевтическую композицию п.11.

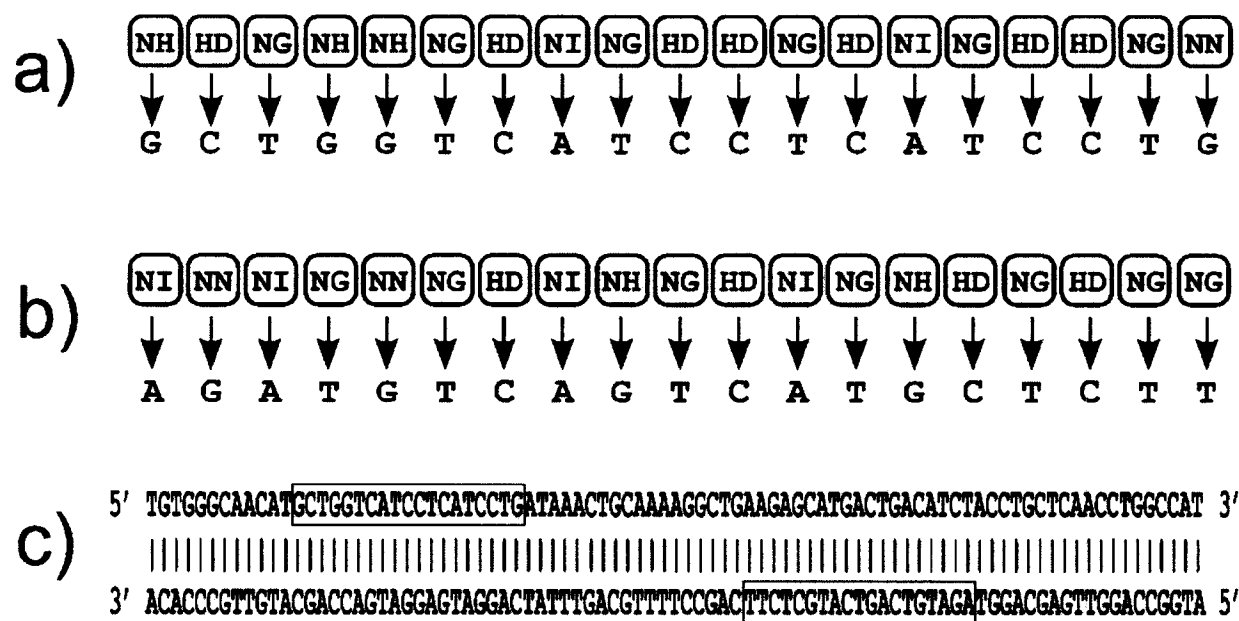


Рис. 1



Рис. 2

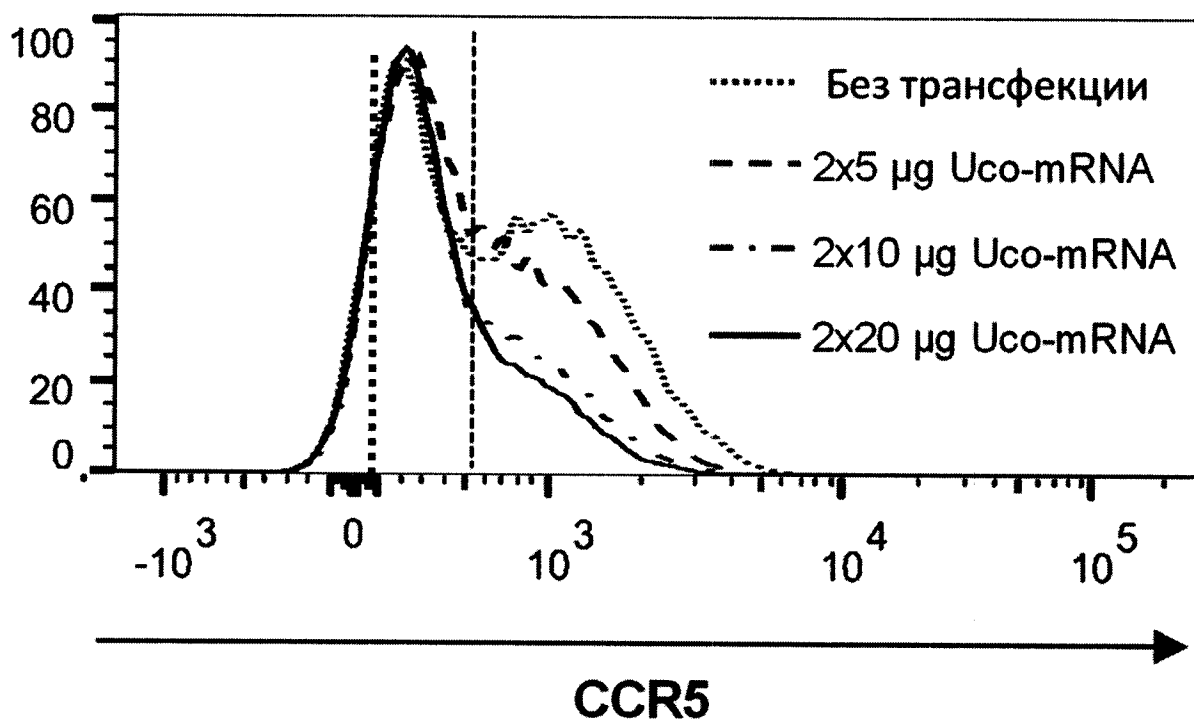


Рис. 3а

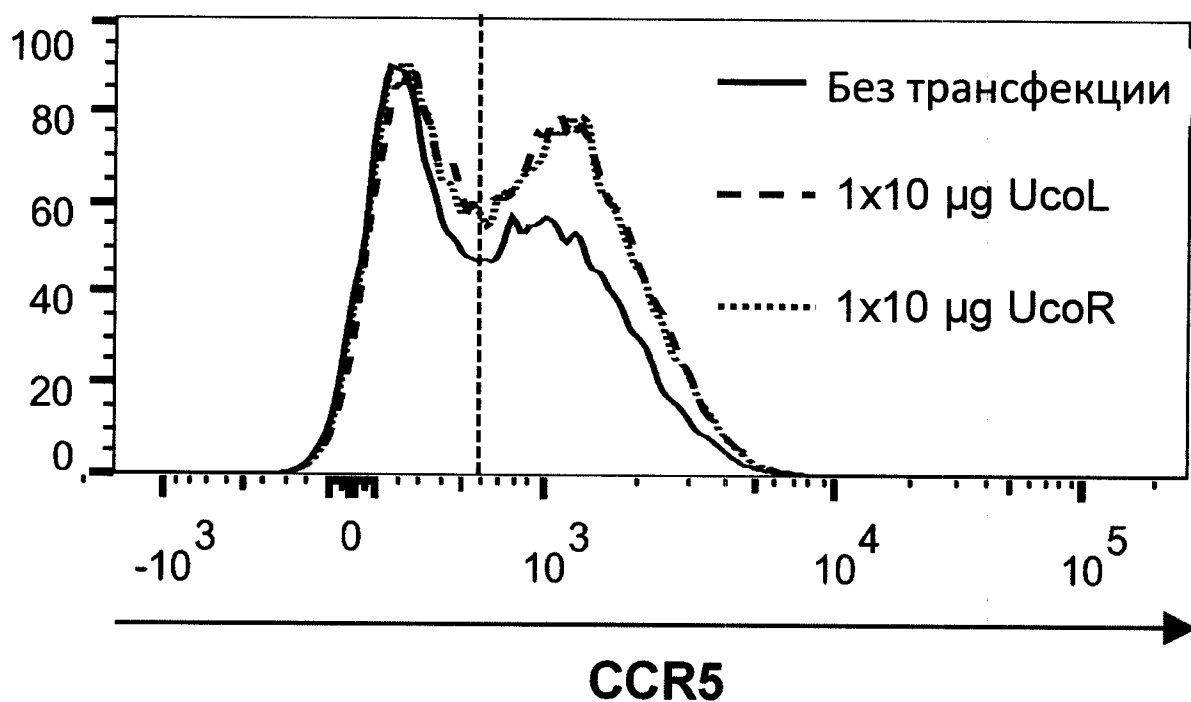


Рис. 3b

WT TTGGTTTTGTGGGCAACATGCTGGTCATCCTCATCCTGATAAACTGCAAAAGGCTGAAGAGCATGACTGACATCTACCTGCTCAACCTGGCCATCTCTGACCTG
 #1 TTGGTTTTGTGGGCAACATGCTGGTCATCCTCATCCTGATAAACTGCAAAAGGCTGAAGAGCATGACTGACATCTACCTGCTCAACCTGGCCATCTCTGACCTG
 #2 TTGGTTTTGTGGGCAACATGCTGGTCATCCTCATCCTGATAAACTGCAAAAGGCTGAAGAGCATGACTGACATCTACCTGCTCAACCTGGCCATCTCTGACCTG
 #3 TTGGTTTTGTGGGCAACATGCTGGTCATCCTCATCCAGATAAACTGCAAAAGGCTGAAGAGCATGACTGACATCTACCTGCTCAACCTGGCCATCTCTGACCTG
 #4 TTGGTTTTGTGGGCAACATGCTGGTCATCCTCATCCAGATAAACTGCAAAAGGCTGAAGAGCATGACTGACATCTACCTGCTCAACCTGGCCATCTCTGACCTG
 #5 TTGGTTTTGTGGGCAACATGCTGGTCATCCTCATCCTGATAAACTGCAAAAGGCTGAAGAGCATGACTGACATCTACCTGCTCAACCTGGCCATCTCTGACCTG
 #6 TTGGTTTTGTGGGCAACATGCTGGTCATCCTCATCCTGATAAACTGCAAAAGGCTGAAGAGCATGACTGACATCTACCTGCTCAACCTGGCCATCTCTGACCTG
 #7 TTGGTTTTGTGGGCAACATGCTGGTCATCCTCATCCAGATAAACTGCAAAAGGCTGAAGAGCATGACTGACATCTACCTGCTCAACCTGGCCATCTCTGACCTG
 #8 TTGGTTTTGTGGGCAACATGCTGGTCATCCTCATCCAGATAAACTGCAAAAGGCTGAAGAGCATGACTGACATCTACCTGCTCAACCTGGCCATCTCTGACCTG
 #9 TTGGTTTTGTGGGCAACATGCTGGTCATCCTCATCCTGATAAA-----GGCTGAAGAGCATGACTGACATCTACCTGCTCAACCTGGCCATCTCTGACCTG
 #10 TTGGTTTTGTGGGCAACATGCTGGTCATCCTCATCCTAA-----GGCTGAAGAGCATGACTGACATCTACCTGCTCAACCTGGCCATCTCTGACCTG
 #11 TTGGTTTTGTGGGCAACATGCTGGTCATCCTCATCCTGATAAA-----CTGAAGAGCATGACTGACATCTACCTGCTCAACCTGGCCATCTCTGACCTG
 #12 TTGGTTTTGTGGGCAACATGCTGGTCATCCTCATCCAGATAAAC-----GCATGACTGACATCTACCTGCTCAACCTGGCCATCTCTGACCTG
 #13 TTGGTTTTGTGGGCAACATGCTGGTCATCCTCATCCTGATAA-----ACTGACATCTACCTGCTCAACCTGGCCATCTCTGACCTG
 #14 TTGGTTTTGTGGGCAACATGCTGGTCATCCTCATCCAGATG-----ACTGACATCTACCTGCTCAACCTGGCCATCTCTGACCTG
 #15 TTGGTTTTGTGGGCAACATGCTGGTCATCCTCATGCATG-----ACTGACATCTACCTGCTCAACCTGGCCATCTCTGACCTG
 #16 TTGGTTTTGTGGGCAACATGCTGGTCATCCTCATCCA-----GATAAACT---GCCATCTCTGACCTG
 #17 TTGGTTTTGTGGGCAACANNCTGGTCGTCATCATG-----ACTGACATCTACCTGCTCAACCTGGCCATCTCTGACCTG

Pnc. 3c