



**(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ**

**(45)** Дата публикации и выдачи патента  
**2021.11.16**

**(21)** Номер заявки  
**201692361**

**(22)** Дата подачи заявки  
**2015.04.23**

**(51)** Int. Cl. **C07K 14/47** (2006.01)  
**C12N 15/86** (2006.01)  
**A61K 48/00** (2006.01)  
**C07K 14/755** (2006.01)

**(54) ЛЕНТИВИРУСНЫЙ ВЕКТОР ДЛЯ ГЕННОЙ ТЕРАПИИ, СПОСОБ ЕГО ПОЛУЧЕНИЯ И ЕГО ПРИМЕНЕНИЕ**

**(31) 1409089.8**

**(32) 2014.05.21**

**(33) GB**

**(43) 2017.05.31**

**(86) PCT/GB2015/051201**

**(87) WO 2015/177501 2015.11.26**

**(71)(73)** Заявитель и патентовладелец:  
**АЙПИ2АЙПИО ИННОВЕЙШНЗ  
ЛИМИТЕД (GB); АЙДИ ФАРМА КО.,  
ЛТД. (JP)**

**(72)** Изобретатель:  
**Элтон Эрик Уолтер Фредерик  
Вольфганг, Гризенбах Юта, Питель  
Камила Мальгожата, Пол-Смит  
Майкл Кристиан, Прингл Ян Эндрю,  
Хайд Стефен Чарльз, Джилл Дебора  
Ребекка, Дэвис Ли Адриан, Бойд Алан  
Кристофер, Маклэчлэн Джерард (GB),  
Иноуе Макото (JP)**

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

**(56)** EP-A1-1950307  
KATSUYUKI MITOMO ET AL.: "Toward  
Gene Therapy for Cystic Fibrosis Using a Lentivirus  
Pseudotyped With Sendai Virus Envelopes",  
MOLECULAR THERAPY, vol. 18, no. 6, 23 March  
2010 (2010-03-23), pages 1173-1182, XP055199531,  
ISSN: 1525-0016, DOI: 10.1038/mt.2010.13 page  
1178 - page 1179

NAKAJIMA T. ET AL.: "DEVELOPMENT  
OF NOVEL SIMIAN IMMUNODEFICIENCY  
VIRUS VECTORS CARRYING DUAL GENE  
EXPRESSION SYSTEM", HUMAN GENE

THERAPY, MARY ANN LIEBERT, NEW YORK,  
NY, US, vol. 11, no. 13, September 2000 (2000-09),  
pages 1863-1874, XP001033905, ISSN: 1043-0342,  
DOI: 10.1089/10430340050129486 cited in the  
application figure 2

WO-A2-2007110628

DANIELA ZYCHLINSKI ET AL.:  
"Physiological Promoters Reduce the Genotoxic  
Risk of Integrating Gene Vectors", MOLECULAR  
THERAPY, vol. 16, no. 4, April 2008 (2008-04),  
pages 718-725, XP055199418, ISSN: 1525-0016,  
DOI: 10.1038/mt.2008.5 page 719

JOHN F. ENGELHARDT: "The lung as a  
metabolic factory for gene therapy", JOURNAL  
OF CLINICAL INVESTIGATION, vol. 110,  
no. 4, 15 August 2002 (2002-08-15), pages  
429-432, XP055199495, ISSN: 0021-9738, DOI:  
10.1172/JCI200216443 page 429 - page 431

WO-A2-03029274

CHUAH M.K.L. ET AL.: "GENE  
THERAPY FOR HEMOPHILIA", JOURNAL  
OF GENE MEDICINE, JOHN WILEY &  
SONS, INC, US, vol. 3, no. 1, 2001,  
pages 3-20, XP009017044, ISSN: 1099-498X,  
DOI: 10.1002/1521-2254(200101/02)3:1<3 :: AID-  
JGM167>3.0.CO; 2-H page 7 - page 10

SCOTT R. WITTING ET AL.: "Efficient  
Large Volume Lentiviral Vector Production Using  
Flow Electroporation", HUMAN GENE THERAPY,  
vol. 23, no. 2, February 2012 (2012-02), pages  
243-249, XP055143670, ISSN: 1043-0342, DOI:  
10.1089/hum.2011.088 page 246 - page 248

JAMES CRONIN ET AL.: "Altering the  
Tropism of Lentiviral Vectors through Pseudotyping",  
CURRENT GENE THERAPY, vol. 5, no. 4, August  
2005 (2005-08), pages 387-398, XP055199682, ISSN:  
1566-5232, DOI: 10.2174/1566523054546224 page  
10 - page 12; table 1

**(57)** Изобретение относится к лентивирусным векторам для генной терапии, псевдотипированным гемагглютинин-нейраминидазой (HN) и гибридными (F) белками из респираторного парамиксовируса, содержащим промотор и трансген; и способам их получения. Настоящее изобретение также относится к применению указанных векторов в генной терапии, в частности для лечения заболеваний дыхательных путей, таких как муковисцидоз (КФ), способу лечения таких заболеваний, клетке-хозяину и композиции для применения в генной терапии.

Настоящее изобретение относится к лентивирусным векторам для переноса генов, псевдотипированным гемагглютинин-нейраминидазой (HN) и гибридными (F) белками из респираторного парамиксовируса, содержащим промотор и трансген; и способам их получения. Настоящее изобретение также относится к применению указанных векторов в генной терапии, в частности для лечения заболеваний дыхательных путей, таких как муковисцидоз (МВ).

#### **Предшествующий уровень техники**

Лентивирусы принадлежат к роду вирусов семейства *Retroviridae* и характеризуются длительным периодом инкубации. Лентивирусы могут доставлять значительное количество вирусной РНК в ДНК клетки-хозяина и имеют уникальную способность среди ретровирусов, обладая свойством инфицировать неделящиеся клетки, таким образом, они являются одним из наиболее эффективных методов создания вектора доставки гена.

Лентивирусные векторы, в особенности те, что получены из ВИЧ-1, являются широкоисследуемыми и часто применяемыми векторами. Эволюция остова лентивирусных векторов и способность вирусов доставлять молекулы рекомбинантных ДНК (трансгены) в клетки-мишени привели к их применению во многих направлениях практического использования. Два возможных направления использования вирусных векторов включают восстановление функциональных генов в генетической терапии и получение рекомбинантных белков *in vitro*.

Псевдотипирование представляет собой процесс получения вирусов или вирусных векторов в комбинации с чужеродными белками вирусной оболочки. Сами по себе чужеродные белки вирусной оболочки могут использоваться для изменения тропизма хозяина или увеличенной/уменьшенной стабильности вирусных частиц. Например, псевдотипирование позволяет устанавливать характер белков оболочки. Часто используемым белком является гликопротеин G вируса везикулярного стоматита (VSV), сокращенно VSV-G.

Понятно, что для эффективной и контролируемой ретровирусной экспрессии трансгена требуется присутствие интронных последовательностей. Однако введение таких интронов в ретровирусные векторы включает усложненные и требующие больших затрат времени способы вследствие используемых многостадийных процессов.

До настоящего времени вирусные агенты переноса генов были неприменимыми для лечения заболеваний, без трансдукции популяций стволовых клеток, вследствие адаптивного иммунного ответа хозяина, который препятствует успешному повторному введению.

Более того, перенос гена к эпителию дыхательных путей оказался более трудным, чем первоначально предполагали. Например, применение лентивирусных псевдотипов, для которого требуется разрушение эпителиальной целостности, чтобы осуществить трансдукцию в дыхательные пути, например, посредством использования детергентов, таких как лизофосфатидилхолин или этиленгликоль-бис(2-аминоэтиловый эфир)-N,N,N',N'-тетрауксусная кислота, было связано с повышенным риском сепсиса.

Одним примером клинической ситуации, которая могла бы получить благоприятный эффект от переноса генов в эпителий дыхательных путей является лечение муковисцидоза (МВ). МВ является летальным генетическим расстройством, вызываемым мутациями в гене регулятора трансмембранной проводимости при МВ (CFTR), который действует в качестве хлоридного канала в эпителиальных клетках дыхательных путей. МВ характеризуется рецидивирующими инфекциями дыхательных путей, увеличенными секретами дыхательных путей и, в конечном итоге, дыхательной недостаточностью. В Великобритании в настоящее время средний возраст умерших составляет ~25 лет. Для большинства генотипов не существует терапий, нацеленных на базовый дефект; применяемые в настоящее время терапии для облегчения симптомов требуют часов самостоятельно вводимого лечения ежедневно. Генная терапия в отличие от низкомолекулярных лекарственных средств является независимой от мутационного класса CFTR и является, таким образом, применимой ко всем пораженным МВ индивидуумам. Однако до настоящего времени ни один вирусный вектор не соответствовал требованиям для клинического применения, то же самое относится к другим заболеваниям, конкретно многим другим заболеваниям дыхательных путей.

В этой связи сталкиваются по меньшей мере с тремя основными проблемами. Эффективность переноса генов обычно является низкой, по меньшей мере отчасти, поскольку соответствующие рецепторы для многих вирусных векторов, по всей видимости, преобладающим образом локализованы на базолатеральной поверхности эпителия дыхательных путей. Во-вторых, проницаемость слизистого слоя дыхательных путей обычно является низкой. В конечном счете, способность к повторному введению вирусных векторов, обязательная для долговременной обработки самообновляющегося эпителия, является ограниченной.

Введение векторов для клинического применения является еще одним актуальным фактором. Следовательно, для эффективности лечения должна поддерживаться вирусная стабильность при использовании клинически значимых устройств (например, бронхоскопа и небулайзера).

Еще одним примером потенциальной мишени для генной терапии является недостаточность  $\alpha 1$ -антитрипсина (A1AT).

Недостаточность A1AT представляет собой наследственное нарушение, которое может вызывать

заболевание легких и заболевание печени. Симптомы включают одышку/свистящее дыхание, пониженную способность к физической нагрузке, потерю массы тела, рецидивирующие респираторные инфекции, утомляемость и учащенное сердцебиение при стоянии. У страдающих индивидуумов часто развивается эмфизема. Приблизительно у 10-15% пациентов с недостаточностью А1АТ развивается заболевание печени. Индивидуумы с недостаточностью А1АТ также имеют риск развития гепатоцеллюлярной карциномы.

А1АТ представляет собой секретируемый белок, вырабатываемый, в основном, в печени и затем переносимый в легкие, причем более мелкие количества также продуцируются в самих легких. Основной функцией А1АТ является связывать и нейтрализовать нейтрофил-эластазу. Генная терапия А1АТ, вероятно, будет иметь терапевтическое значение у пациентов с недостаточностью А1АТ, МВ и хроническим обструктивным заболеванием легких (ХОБЛ), где повышение или введение А1АТ может улучшить функцию легких.

Терапия А1АТ также является потенциально ценной для лечения нереспираторных/нелегочных заболеваний, таких как диабет типа 1 и типа 2, острый инфаркт миокарда, ревматоидный артрит, воспалительное заболевание кишечника, отторжение трансплантата, реакция трансплантат против хозяина (GvH), рассеянный склероз и инфекции, особенно вирусные инфекции, вследствие воздействия недостаточности А1АТ на другие ткани/органы, такие как печень и поджелудочная железа (см., например, Lewis Mol. Med. 2012; 18:957-970, которая включена в данное описание посредством ссылки).

Недостаточность А1АТ является привлекательным целевым заболеванием для генной терапии, поскольку терапевтические пороговые уровни являются хорошо определенными. Сравнение уровней А1АТ у субъектов с риском развития эмфиземы/ХОБЛ позволило определить защитный пороговый уровень, равный 11 мкМ в сыворотке, причем уровни ниже 11 мкМ используют в качестве пороговых значений для инициации стимулирующей белковой терапии, где это является доступным. Уровни А1АТ в лаважной жидкости дыхательных путей составляют только ~10% от сывороточного уровня, поскольку эпителий легких создает барьер, и терапевтическое пороговое значение в лаважной жидкости поверхности дыхательных путей, следовательно, считают равным 1,1 мкМ (см. Ferrara et al. Thorax. 2012 Aug; 67(8):669-74 и Abusriwil & Stockley 2006 Current Opinion in Pulmonary Medicine 12:125-131, каждая из которых включена в настоящее описание посредством ссылки).

Шесть разрешенных FDA к применению промышленно выпускаемых лекарственных форм белка А1АТ, выделенного из объединенной крови человека, находят клиническое применение в США для лечения пациентов с тяжелой недостаточностью А1АТ (посредством еженедельных внутривенных инъекций). Ферментозаместительная терапия (ERT) является дорогостоящей (~100000 долларов/год), и несмотря на то, что биохимическая эффективность действия для стимулирующей белковой терапии ERT была доказана, доказательство эффективности в клинической практике является значительно более трудным. ERT А1АТ в настоящее время недоступна во всех странах и в настоящее время является недоступной в Великобритании. Кроме того, используя принятые в настоящее время терапии, трудно достигнуть достаточных перманентных тканевых уровней, что может отчасти являться причиной ограниченной эффективности в клинической практике, наблюдаемой до настоящего времени.

Другие привлекательные мишени для генной терапии включают сердечно-сосудистые заболевания и заболевания крови, в частности нарушения, связанные с недостаточной свертываемостью крови, такие как гемофилия (А и В), болезнь фон Виллебранда и недостаточность фактора VII.

Гемофилия, в частности гемофилия А, является привлекательной мишенью для генной терапии. Гемофилия А представляет собой наследственное нарушение гемостаза, вызванное недостаточностью или мутацией фактора VIII (FVIII). Его наследование связано с полом, причем почти все пациенты являются мужчинами. Кровотечение обычно происходит в суставы. Кровотечение в мышцы, слизистую ткань и центральную нервную систему (ЦНС) является нетипичным, но может происходить. Тяжесть заболевания является обратно пропорциональной уровню FVIII: менее чем 1% (<0,01 МЕ/мл) приводит к тяжелому заболеванию, с кровотечением после минимального повреждения; между 1-5% (0,01-0,05 МЕ/мл) вызывает умеренное заболевание, с кровотечением после легкого повреждения; и более чем 5% (>0,05 МЕ/мл) вызывает легкое заболевание, с кровотечением только после значительной травмы или хирургического вмешательства.

Соответственно, существует потребность в векторе для генной терапии, который сделает возможным избежать одну или несколько проблем, описанных выше.

#### **Краткое описание сущности изобретения**

Авторы настоящего изобретения разработали лентивирусный вектор, который был псевдотипирован гемагглютинин-нейраминидазой (HN) и гибридными (F) белками из респираторного парамиксовируса, содержащий промотор и трансген. Обычно остов вектора состоит из вируса иммунодефицита обезьян (SIV), такого как SIV1 или вируса SIV иммунодефицита африканской зеленой мартышки (SIV-AGM). Предпочтительно остов вирусного вектора по изобретению берется из SIV-AGM. HN и F-белки функционируют, соответственно, чтобы присоединиться к сиаловым кислотам, и опосредуют слияние клеток для входа вектора в клетки-мишени. Авторы настоящего изобретения обнаружили, что этот специфически F/HN-псевдотипированный лентивирусный вектор может эффективно трансдуцироваться в эпи-

телей дыхательных путей, приводя в результате к экспрессии трансгена, перманентной в течение периодов, находящихся за пределами предполагаемой продолжительности жизни эпителиальных клеток дыхательных путей. Важно отметить, что авторы настоящего изобретения также обнаружили, что повторное введение не приводит к потере эффективности действия. Эти признаки делают векторы согласно настоящему изобретению привлекательными кандидатами для лечения заболеваний посредством их применения в экспрессии терапевтических белков: (i) внутри клеток дыхательных путей; (ii) секретлируемых в просвет дыхательных путей и (iii) секретлируемых в систему кровообращения.

Настоящее изобретение направлено на удовлетворение одной или нескольких из указанных выше потребностей и относится к лентивирусным векторам, псевдотипированным с использованием гемагглютинин-нейраминидазы (HN) и гибридных (F) белков из респираторного парамиксовируса, содержащие промотор и трансген. В одном варианте осуществления промотор предпочтительно представляет собой гибридный энхансер человеческого CMV/промотор EF1a (hCEF). Настоящее изобретение также относится к способам получения указанных векторов, композициям, содержащим указанные векторы, и их применению в терапии.

Векторы согласно настоящему изобретению обеспечивают более высокую и перманентную экспрессию гена посредством эффективного переноса генов. Идентифицированные выше проблемы решаются настоящим изобретением, которое предоставляет F/HN-псевдотипированные лентивирусные векторы, которые обладают способностью к (i) трансдукции в воздухоносные пути без разрушения эпителиальной целостности; (ii) персистирующей экспрессии гена; (iii) отсутствию хронической токсичности и (iv) эффективному повторному введению.

Долговременная/персистирующая стабильная экспрессия гена, предпочтительно на терапевтически эффективном уровне, может достигаться с использованием повторных доз вектора настоящего изобретения. Альтернативно, чтобы достичь желательной долговременной экспрессии, может применяться однократная доза.

В отличие от известных лентивирусных векторов, лентивирусные векторы по изобретению проявляют эффективное включение в клетки дыхательных путей, увеличенную экспрессию трансгена и не страдают потерей эффективности при повторном введении.

Таким образом, преимущественно, лентивирусные векторы согласно настоящему изобретению могут использоваться в генной терапии. Как пример, свойства эффективного включения векторов по изобретению в клетки дыхательных путей делают их в высокой степени подходящими для лечения заболеваний дыхательных путей. Лентивирусные векторы по изобретению могут также использоваться в методах генной терапии, чтобы стимулировать секрецию терапевтических белков. Как дополнительный пример, изобретение относится к секреции терапевтических белков в просвет дыхательных путей или системы кровообращения. Таким образом, введение вектора по изобретению и его включение в клетки дыхательных путей могут обеспечить использование легких (или носа или дыхательных путей) в качестве "фабрики" для производства терапевтического белка, который затем секретруется и попадает в общее кровообращение при терапевтических уровнях, где он может перемещаться к клеткам/тканям, представляющим интерес, чтобы оказать терапевтическое воздействие. В отличие от внутриклеточных или мембранных белков, выработка таких секретлируемых белков не основывается на специфичных для заболевания клетках-мишенях, являющихся трансдуцированными, что является значительным преимуществом, и достигает высоких уровней экспрессии белка. Таким образом, другие заболевания, которые не являются заболеваниями дыхательных путей, такие как сердечно-сосудистые заболевания и заболевания крови, в особенности нарушения, связанные с недостаточной свертываемостью крови, могут также подвергаться лечению векторами настоящего изобретения.

В качестве примера  $\alpha$ -1 антитрипсин (A1AT) представляет собой секретлируемую антипротеазу, которая продуцируется в основном в печени и затем перемещается в легкие, причем более мелкие количества также продуцируются в самом легком. Основной функцией A1AT является связывать и нейтрализовать/ингибировать нейтрофил-эластазу. Генная терапия с использованием A1AT, согласно настоящему изобретению, имеет отношение к пациенту с недостаточностью A1AT, а также пациенту с другими заболеваниями легких, таким как муковисцидоз или хроническое обструктивное заболевание легких (ХОБЛ), и дает возможность преодолеть некоторые проблемы, с которыми сталкивается ферментозаместительная терапия.

Авторы настоящего изобретения ранее продемонстрировали, что существует значительная корреляция между нейтрофил-эластазой (NE) и A1AT в образцах мокроты от пациентов с муковисцидозом, показывающая, что организм вырабатывает A1AT в ответ на стимуляцию NE. Авторы настоящего изобретения также показали, что существует статистически значимая корреляция между NE и показателем легочного очищения, маркером заболевания мелких дыхательных путей, означая, что увеличенная NE имеет отрицательное воздействие на функцию легких. Как показано в настоящем описании, авторы изобретения в настоящее время неожиданно продемонстрировали, что лентивирусные векторы по изобретению могут достигать высоких концентраций A1AT и долговременной (по меньшей мере 90 дней) экспрессии A1AT *in vivo*. Таким образом, генная терапия с использованием A1AT может нейтрализовать NE, улучшая функцию легких пациентов с муковисцидозом и/или ХОБЛ (и обладая терапевтическим



эффектом для других показаний, как описано в настоящем документе). Соответственно, настоящее изобретение относится к применению лентивирусного вектора, описанного в настоящем описании, для введения трансгена А1АТ и генной терапии состояний, включающих, но не ограниченных перечисленными, недостаточность А1АТ, муковисцидоз и/или ХОБЛ. Введение лентивирусного А1АТ непосредственно в назальный эпителий и/или легкое может преодолеть несколько ограничений, с которыми в настоящее время сталкивается ферментозаместительная терапия (А1АТ, выделенный из человеческой крови и вводимый внутривенно каждую неделю), обеспечивая стабильную, продолжительную экспрессию в ткани-мишени (легкое/назальный эпителий), простоту введения и неограниченную доступность.

В нескольких вариантах осуществления трансдукция с использованием лентивирусного вектора по изобретению приводит к секреции рекомбинантного белка в просвет легкого, а также в кровообращение. Одним благоприятным эффектом этого является то, что терапевтический белок достигает интерстициальной ткани. В случае недостаточности А1АТ этот эффект является преимущественным, поскольку ингибирование NE также требуется на данном участке. Генная терапия А1АТ может, следовательно, также быть благоприятной при показаниях для других заболеваний, неограничивающие примеры которых включают диабет типа 1 и типа 2, острый инфаркт миокарда, ишемическую болезнь сердца, ревматоидный артрит, воспалительное заболевание кишечника, отторжение трансплантата, реакцию трансплантат против хозяина (GvH), рассеянный склероз, заболевание печени, цирроз, васкулиты и инфекции, такие как бактериальные и/или вирусные инфекции.

А1АТ имеет многочисленные другие противовоспалительные и тканезащитные эффекты, например, на доклинических моделях диабета, реакции трансплантат против хозяина и воспалительного заболевания кишечника. Выработка А1АТ в легких и/или носу после трансдукции согласно настоящему изобретению может, следовательно, быть более широко применимой, с включением этих показаний.

Другие примеры заболеваний, которые могут подвергаться лечению с использованием генной терапии секретлируемым белком, согласно настоящему изобретению включают сердечно-сосудистые заболевания и заболевания крови, в особенности нарушения, связанные с недостаточной свертываемостью крови, такие как гемофилия (А и В), болезнь фон Виллебранда и недостаточность фактора VII.

В нескольких вариантах осуществления гемофилия А может подвергаться лечению согласно настоящему изобретению. Тяжесть заболевания является обратно пропорциональной уровню FVIII, и увеличение FVIII, равное 2-5% (0,02-0,05 МЕ/мл), является достаточным, чтобы быть терапевтически эффективным.

В нескольких вариантах осуществления нос является предпочтительным участком продуцирования для терапевтического белка с использованием вектора по изобретению для генной терапии вследствие по меньшей мере одной из следующих причин: (i) внеклеточные барьеры, такие как воспалительные клетки и мокрота, являются наименее выраженными в носу; (ii) простота введения вектора; (iii) более мелкие количества требуемого вектора и (iv) этические соображения. Таким образом, трансдукция назальных эпителиальных клеток с использованием лентивирусного вектора по изобретению может приводить к эффективной (на высоком уровне) и продолжительной экспрессии терапевтического трансгена, представляющего интерес.

Векторы согласно настоящему изобретению обеспечивают долговременную экспрессию гена, приводящую к долговременной экспрессии терапевтического белка. Как описано в настоящем описании, фразы "долговременная экспрессия", "перманентная экспрессия" и "персистирующая экспрессия" используют взаимозаменяемо. Долговременная экспрессия согласно настоящему изобретению означает экспрессию терапевтического гена и/или белка, предпочтительно на терапевтических уровнях, в течение по меньшей мере 45 дней, по меньшей мере 60 дней, по меньшей мере 90 дней, по меньшей мере 120 дней, по меньшей мере 180 дней, по меньшей мере 250 дней, по меньшей мере 360 дней, по меньшей мере 450 дней, по меньшей мере 730 дней или более. Предпочтительно долговременная экспрессия означает экспрессию в течение по меньшей мере 90 дней, по меньшей мере 120 дней, по меньшей мере 180 дней, по меньшей мере 250 дней, по меньшей мере 360 дней, по меньшей мере 450 дней, по меньшей мере 720 дней или более, более предпочтительно по меньшей мере 360 дней, по меньшей мере 450 дней, по меньшей мере 720 дней или более. Эта долговременная экспрессия может быть достигнута посредством повторных доз или однократной дозы.

Повторяющиеся дозы могут вводиться дважды в день, ежедневно, дважды в неделю, еженедельно, ежемесячно, каждые два месяца, каждые три месяца, каждые четыре месяца, каждые шесть месяцев, ежегодно, каждые два года или более. Дозирование может продолжаться настолько долго, насколько требуется, например в течение по меньшей мере шести месяцев, по меньшей мере одного года, двух лет, трех лет, четырех лет, пяти лет, десяти лет, пятнадцати лет, двадцати лет или более, вплоть до всего времени жизни пациента, подлежащего лечению.

Лентивирусные векторы, такие как векторы по изобретению, могут интегрироваться в геном трансдуцируемых клеток и приводить к долговременной экспрессии, что делает их подходящими для трансдукции стволовых клеток/клеток-предшественников. В легких некоторые типы клеток с регенеративной способностью были идентифицированы как ответственные за поддержание специфических последовательностей клеточных поколений в дыхательных путях и альвеолах. Такие типы включают базальные

клетки и подслизистые клетки протока железы в верхних дыхательных путях, клетки Клара и нейроэндокринные клетки в бронхиолярных воздухоносных путях, бронхоальвеолярные стволовые клетки в терминальных бронхиолах и пневмоциты типа II в альвеолах. Следовательно и без ограничения теорией, полагают, что векторы согласно настоящему изобретению вызывают долговременную генную экспрессию трансгена, представляющего интерес, посредством введения трансгена в одну или несколько долгоживущих эпителиальных клеток или один или несколько типов клеток дыхательных путей, таких как базальные клетки и подслизистые клетки протока железы верхних дыхательных путей, клетки Клара и нейроэндокринные клетки в бронхиолярных воздухоносных путях, бронхоальвеолярные стволовые клетки в терминальных бронхиолах и пневмоциты типа II в альвеолах.

Соответственно, лентивирусные векторы по изобретению могут трансдуцировать одну или несколько клеток или линий клеток с регенеративным потенциалом внутри легкого (включая дыхательные пути) для достижения долговременной экспрессии гена. В предпочтительном варианте осуществления лентивирусный вектор изобретения трансдуцирует базальные клетки, такие как клетки верхних дыхательных путей/дыхательных путей. Базальные клетки играют главную роль в процессах эпителиального поддержания и восстановления после повреждения. Кроме того, базальные клетки являются широко распределенными вдоль человеческого респираторного эпителия, с относительным распределением от 30% (крупные воздухоносные пути) до 6% (мелкие воздухоносные пути).

Лентивирусные векторы по изобретению могут использоваться для трансдукции изолированных и размноженных стволовых клеток/клеток-предшественников *ex vivo* до введения пациенту. Предпочтительно лентивирусные векторы по изобретению используют для трансдукции клеток в легких (или дыхательных путях) *in vivo*.

Векторы согласно настоящему изобретению обеспечивают высокие уровни экспрессии гена, приводящие в результате к высоким уровням (предпочтительно терапевтическим уровням) экспрессии терапевтического белка. Экспрессию можно измерить любым подходящим методом (качественным или количественным, предпочтительно количественным), и концентрации могут быть приведены в любых соответствующих единицах измерения, например нг/мл. Высокий уровень экспрессии согласно настоящему изобретению может означать экспрессию терапевтического гена и/или белка при концентрации, равной по меньшей мере 10 нг/мл, по меньшей мере 20 нг/мл, по меньшей мере 30 нг/мл, по меньшей мере 40 нг/мл, по меньшей мере 50 нг/мл, по меньшей мере 60 нг/мл, по меньшей мере 70 нг/мл, по меньшей мере 80 нг/мл, по меньшей мере 90 нг/мл, по меньшей мере 100 нг/мл, по меньшей мере 200 нг/мл, по меньшей мере 300 нг/мл, по меньшей мере 400 нг/мл, по меньшей мере 500 нг/мл, по меньшей мере 600 нг/мл, по меньшей мере 700 нг/мл, по меньшей мере 800 нг/мл, по меньшей мере 900 нг/мл, по меньшей мере 1000 нг/мл, по меньшей мере 2000 нг/мл, по меньшей мере 3000 нг/мл, по меньшей мере 4000 нг/мл, по меньшей мере 5000 нг/мл, по меньшей мере 10000, по меньшей мере 15000 нг/мл, по меньшей мере 20000 нг/мл или более. Терапевтическая экспрессия может быть задана с использованием таких же значений.

Лентивирусные векторы согласно настоящему изобретению обычно обеспечивают высокие уровни экспрессии трансгена при введении пациенту. Термины высокая экспрессия и терапевтическая экспрессия в настоящем описании используют взаимозаменяемо.

Высокий уровень экспрессии согласно настоящему изобретению может означать экспрессию терапевтического гена и/или белка при концентрации, равной по меньшей мере приблизительно 100 нМ, по меньшей мере приблизительно 200 нМ, по меньшей мере приблизительно 300 нМ, по меньшей мере приблизительно 400 нМ, по меньшей мере приблизительно 500 нМ, по меньшей мере приблизительно 600 нМ, по меньшей мере приблизительно 700 нМ, по меньшей мере приблизительно 800 нМ, по меньшей мере приблизительно 900 нМ, по меньшей мере приблизительно 1 мкМ, по меньшей мере приблизительно 1,1 мкМ, по меньшей мере приблизительно 1,2 мкМ, по меньшей мере приблизительно 1,3 мкМ, по меньшей мере приблизительно 1,4 мкМ, по меньшей мере приблизительно 1,5 мкМ, по меньшей мере приблизительно 2 мкМ, по меньшей мере приблизительно 3 мкМ, по меньшей мере приблизительно 4 мкМ, по меньшей мере приблизительно 5 мкМ, по меньшей мере приблизительно 6 мкМ, по меньшей мере приблизительно 7 мкМ, по меньшей мере приблизительно 8 мкМ, по меньшей мере приблизительно 9 мкМ, по меньшей мере приблизительно 10 мкМ, по меньшей мере приблизительно 11 мкМ, по меньшей мере приблизительно 12 мкМ, по меньшей мере приблизительно 13 мкМ, по меньшей мере приблизительно 14 мкМ, по меньшей мере приблизительно 15 мкМ, по меньшей мере приблизительно 20 мкМ, по меньшей мере приблизительно 25 мкМ, по меньшей мере приблизительно 30 мкМ, по меньшей мере приблизительно 40 мкМ, по меньшей мере приблизительно 50 мкМ, по меньшей мере приблизительно 75 мкМ или по меньшей мере приблизительно 100 мкМ или более. Терапевтическая экспрессия может быть задана с использованием этих же самых значений.

Высокий уровень экспрессии согласно настоящему изобретению может означать экспрессию терапевтического гена (обычно измеряемую по экспрессии мРНК), равную по меньшей мере приблизительно 1%, по меньшей мере приблизительно 2%, по меньшей мере приблизительно 3%, по меньшей мере приблизительно 4%, по меньшей мере приблизительно 5%, по меньшей мере приблизительно 6%, по меньшей мере приблизительно 7%, по меньшей мере приблизительно 8%, по меньшей мере приблизительно

9%, по меньшей мере приблизительно 10%, по меньшей мере приблизительно 15%, по меньшей мере приблизительно 20% или более в сравнении с уровнем экспрессии соответствующей эндогенной (дефектной) мРНК. Терапевтическая экспрессия может быть задана с использованием тех же самых значений. Например, типичный уровень экспрессии мРНК эндогенного CFTR может быть количественно оценен с позиций числа копий мРНК на клетку легкого, например одна копия мРНК эндогенного CFTR на клетку легкого, две копии мРНК эндогенного CFTR на клетку легкого, три копии мРНК эндогенного CFTR на клетку легкого, четыре копии мРНК эндогенного CFTR на клетку легкого, пять копий мРНК эндогенного CFTR на клетку легкого или более, предпочтительно две копии мРНК эндогенного CFTR на клетку легкого. Экспрессия терапевтического гена изобретения, такого как функциональный ген CFTR, может быть количественно определена относительно эндогенного гена, такого как эндогенные (дисфункциональные) гены CFTR с позиций числа копий мРНК на клетку или любой другой соответствующей единицы измерения.

Высокий уровень экспрессии согласно настоящему изобретению может означать экспрессию терапевтического гена и/или белка при концентрации, равной по меньшей мере приблизительно 0,5%, по меньшей мере приблизительно 1%, по меньшей мере приблизительно 2%, по меньшей мере приблизительно 3%, по меньшей мере приблизительно 4%, по меньшей мере приблизительно 5%, по меньшей мере приблизительно 6%, по меньшей мере приблизительно 7%, по меньшей мере приблизительно 8%, по меньшей мере приблизительно 9%, по меньшей мере приблизительно 10%, по меньшей мере приблизительно 15%, по меньшей мере приблизительно 20%, по меньшей мере приблизительно 25%, по меньшей мере приблизительно 30%, по меньшей мере приблизительно 40%, по меньшей мере приблизительно 50%, по меньшей мере приблизительно 60%, по меньшей мере приблизительно 70%, по меньшей мере приблизительно 80%, по меньшей мере приблизительно 90%, по меньшей мере 95%, по меньшей мере приблизительно 96%, по меньшей мере приблизительно 97%, по меньшей мере приблизительно 98%, по меньшей мере приблизительно 99% или более в сравнении с уровнем терапевтического гена и/или белка дикого типа, где уровень дикого типа представляет собой уровень у нормального индивидуума без заболевания. В нескольких вариантах осуществления экспрессию дикого типа принимают за 100%, причем любое улучшение экспрессии гена измеряют относительно этого значения. Как неограничивающий пример, если у нормального индивидуума без заболевания экспрессию функционального гена принимают за 100%, и у индивидуума с заболеванием, экспрессия функционального гена равна 0%, терапевтический уровень экспрессии гена или белка может составлять по меньшей мере приблизительно 0,5%, по меньшей мере приблизительно 1%, по меньшей мере приблизительно 2%, по меньшей мере приблизительно 3%, по меньшей мере приблизительно 4%, по меньшей мере приблизительно 5%, по меньшей мере приблизительно 6%, по меньшей мере приблизительно 7%, по меньшей мере приблизительно 8%, по меньшей мере приблизительно 9%, по меньшей мере приблизительно 10%, по меньшей мере приблизительно 15%, по меньшей мере приблизительно 20%, по меньшей мере приблизительно 25%, по меньшей мере приблизительно 30%, по меньшей мере приблизительно 40%, по меньшей мере приблизительно 50% или более в сравнении с уровнем для терапевтического гена и/или белка дикого типа. Как еще один неограничивающий пример, если у нормального индивидуума без заболевания экспрессию функционального гена принимают за 100%, и у индивидуума с заболеванием экспрессия функционального гена составляет 50%, терапевтический уровень экспрессии гена или белка может составлять по меньшей мере приблизительно 55%, по меньшей мере приблизительно 60%, по меньшей мере приблизительно 70%, по меньшей мере приблизительно 80%, по меньшей мере приблизительно 90%, по меньшей мере 95%, по меньшей мере приблизительно 96%, по меньшей мере приблизительно 97%, по меньшей мере приблизительно 98%, по меньшей мере приблизительно 99% или более в сравнении с уровнем для терапевтического гена и/или белка дикого типа.

Для секретируемых белков, таких как A1AT, обычно концентрация в легком или бронхоальвеолярной лаважной жидкости (измеренная с использованием BAL) является приблизительно десятикратной по отношению к концентрации в сыворотке. Как неограничивающий пример, если концентрация секретируемого белка в легком или бронхоальвеолярной лаважной жидкости находится в области 750 нг/мл, сывороточная концентрация белка находится в области 75 нг/мл.

Уровни экспрессии терапевтического гена и/или белка изобретения могут быть измерены в ткани легкого, бронхоальвеолярной лаважной жидкости и/или сыворотке/плазме соответствующим образом. Высокий и/или терапевтический уровень экспрессии может, следовательно, относиться к концентрации в легком, бронхоальвеолярной лаважной жидкости и/или сыворотке/плазме.

Как неограничивающий пример терапевтический уровень экспрессии CFTR обычно составляет 1-5% от терапевтического уровня мРНК CFTR в сравнении с уровнем экспрессии эндогенной (дефектной) мРНК CFTR.

Как еще один неограничивающий пример терапевтический уровень экспрессии A1AT обычно равен по меньшей мере приблизительно 1 мкМ в бронхоальвеолярной лаважной жидкости и/или по меньшей мере приблизительно 0,1 мкМ в сыворотке. В предпочтительном варианте осуществления терапевтический уровень экспрессии A1AT в бронхоальвеолярной лаважной жидкости равен по меньшей мере приблизительно 1,1 мкМ и/или терапевтический сывороточный уровень экспрессии A1AT согласно настоя-

шему изобретению составляет, по меньшей мере приблизительно 11 мкМ. Как еще один неограничивающий пример терапевтический уровень экспрессии A1AT в бронхоальвеолярной лаважной жидкости (ELF, т.е. промывной жидкости воздухоносных путей и воздушного пространства в легких) равен 70 мкг/мл (в сравнении с "нормальным" целевым уровнем АТТ (A1AT) в ELF, равным 200 мкг/мл).

Как еще один неограничивающий пример терапевтический уровень экспрессии белка FVIII обычно составляет по меньшей мере приблизительно 1-3% или по меньшей мере приблизительно 1-6% от уровня экспрессии у нормального индивидуума, не страдающего от гемофилии.

Терапевтический ген, включенный в вектор изобретения, может быть модифицирован, чтобы способствовать экспрессии. Например, последовательность гена может находиться в CpG-обедненной и/или кодон-оптимизированной форме, чтобы способствовать экспрессии гена. Стандартные методы для модификации генной последовательности, таким образом, являются известными в данной области.

Промотор, включенный в вектор по изобретению, может быть специфически выбран и/или модифицирован, чтобы в дальнейшем оптимизировать регуляцию экспрессии терапевтического гена. Также подходящие промоторы и стандартные методы для их модификации являются известными в данной области. Как неограничивающий пример, ряд подходящих (не содержащих CpG) промоторов, пригодных для применения в настоящем изобретении, описаны в Pringle et al. (J. Mol. Med. Berl. 2012, 90(12): 1487-96), которая полностью включена в данное описание посредством ссылки.

Вектор по изобретению может быть модифицирован для обеспечения прекращения экспрессии гена. Стандартные методы модификации вектора таким образом являются известными в данной области. Как неограничивающий пример широко используют Тет-чувствительные промоторы.

Векторы согласно настоящему изобретению также демонстрируют замечательную устойчивость к усилиям сдвига лишь с умеренным снижением способности к трансдукции при прохождении через клинически значимые устройства для доставки, такие как бронхоскопы, аэрозольные баллончики и ингаляторы.

В одном варианте осуществления, изобретение относится к лентивирусным векторам F/IN, содержащим промотор и транскрипционный элемент, не имеющий интрона расположенного между промотором и транскрипционным элементом. В одном варианте осуществления вектор согласно настоящему изобретению доставляют в клетки дыхательных путей. В варианте осуществления лентивирус представляет собой SIV. В одном варианте осуществления промотор является гибридом энхансера человеческого CMV/промотора EF1a (hCEF). Обычно в указанном промоторе изобретения отсутствует интрон, соответствующий нуклеотидам 570-709, и экзон, соответствующий нуклеотидам 728-733 промотора hCEF. Предпочтительный пример последовательности промотора hCEF изобретения представлен SEQ ID NO: 6. Промотор может являться промотором CMV. Пример последовательности промотора CMV представлен SEQ ID NO: 17. Другие промоторы для экспрессии транскрипционного элемента являются известными в данной области, и их пригодность для лентивирусных векторов по изобретению определяют с использованием установленных методов, известных в данной области. Неограничивающие примеры других промоторов включают UbC и UCCE. Как описано в настоящем документе, промотор может быть модифицирован, чтобы в дальнейшем регулировать экспрессию транскрипционного элемента изобретения.

В одном варианте осуществления транскрипционный элемент кодирует CFTR. Пример кДНК CFTR представлен SEQ ID NO: 7.

В одном варианте осуществления транскрипционный элемент кодирует A1AT. Пример транскрипционного элемента A1AT представлен SEQ ID NO: 15 или последовательностью, комплементарной SEQ ID NO: 26. SEQ ID NO: 15 является кодон-оптимизированным обедненным по CpG транскрипционным элементом A1AT, сконструированным авторами настоящего изобретения для усиления трансляции в человеческих клетках. Было показано, что такая оптимизация усиливает экспрессию гена вплоть до 15-кратного усиления. Таким образом, в одном варианте осуществления изобретение относится к полинуклеотиду, содержащему нуклеотидную последовательность или состоящему из нуклеотидной последовательности с SEQ ID NO: 15. Варианты такой же самой последовательности (определенной в настоящем описании), которые обладают таким же техническим эффектом усиления трансляции в сравнении с немодифицированной (дикого типа) последовательностью гена A1AT, также являются охваченными настоящим изобретением. Изобретение дополнительно относится к полипептиду, кодируемому указанным транскрипционным элементом A1AT, иллюстрированному полипептидом с SEQ ID NO: 27, плазмидам (в частности, плазмидам с геномом вектора, определенным в настоящем описании) и лентивирусным векторам, содержащим указанный транскрипционный элемент A1AT. В предпочтительном варианте осуществления в аспектах изобретения, относящихся к генной терапии A1AT, согласно настоящему изобретению, используют последовательность транскрипционного элемента A1AT с SEQ ID NO: 15.

В одном варианте осуществления транскрипционный элемент кодирует FVIII. Примеры транскрипционного элемента FVIII представлены посредством SEQ ID NO: 16 и 30 или посредством соответствующих последовательностей, комплементарных SEQ ID NO: 28 и 31.

Лентивирусные векторы, подходящие для использования в настоящем изобретении, включают вирус иммунодефицита человека (ВИЧ), вирус иммунодефицита обезьян (SIV), вирус иммунодефицита кошачьих (FIV), вирус инфекционной анемии лошадей (EIAV) и вирус висна/мазды. В одном варианте осуществления изобретения используют вектор SIV, предпочтительно SIV-AGM. В еще одном варианте

осуществления используют вектор ВИЧ.

Лентивирусные векторы согласно настоящему изобретению являются псевдотипированными гемагглютинин-нейраминидазой (HN) и гибридными (F) белками из респираторного парамиксовируса. В одном варианте осуществления респираторный парамиксовирус представляет собой вирус Сендай (вирус мышинного парагриппа типа 1).

В одном варианте осуществления изобретения лентивирусный вектор является интеграз-компетентным (IC). В альтернативном варианте осуществления лентивирусный вектор является интеграз-дефицитным (ID).

В еще одном другом варианте осуществления изобретения трансген по изобретению представляет собой любой один или несколько из DNAH5, DNAH11, DNAI1 и DNAI2 или другой известный родственный ген.

В одном варианте осуществления изобретения мишенью для доставки вектора является эпителий дыхательных путей. В этом варианте осуществления трансген кодирует  $\alpha$ -1 антитрипсин (A1AT), поверхностно-активный белок В (SFTPВ) или гранулоцитарно-макрофагальный колониестимулирующий фактор (GM-CSF). В еще одном другом варианте осуществления трансген кодирует моноклональное антитело (mAb) к возбудителю инфекции. В одном варианте осуществления трансген кодирует антитело к ФНО- $\alpha$ . В дополнительном варианте осуществления трансген кодирует терапевтический белок, участвующий в воспалительном, иммунном или метаболическом состоянии.

В одном варианте осуществления изобретения вектор доставляют в клетки дыхательных путей, чтобы обеспечить выработку белков, подлежащих секреции в систему кровообращения. В этом варианте осуществления трансген кодирует фактор VII, фактор VIII, фактор IX, фактор X, фактор XI и/или фактор фон Виллебранда. Такой вектор может использоваться в лечении заболеваний, в особенности сердечно-сосудистых заболеваний и заболеваний крови, предпочтительно расстройств, связанных с недостаточной свертываемостью крови, таких как гемофилия. В еще одном варианте осуществления трансген кодирует моноклональное антитело (mAb) против возбудителя инфекции. В одном варианте осуществления трансген кодирует белок, участвующий в воспалительном, иммунном или метаболическом состоянии, таком как болезнь лизосомального накопления.

Согласно изобретению оно относится к лентивирусному вектору F/HN-SIV, содержащему промотор hCEF и трансген CFTR, не имеющий интрона, расположенного между промотором и трансгеном. Аналогично, отсутствует интрон между промотором и трансгеном в геноме вектора (пДНК1) плазмиды (например, pGM326, описанной в настоящем документе, иллюстрированной на фиг. 1A и с последовательностью SEQ ID NO: 1).

Изобретение также относится к лентивирусному вектору F/HN-SIV, содержащему промотор hCEF и трансген A1AT, не имеющий интрона, расположенного между промотором и трансгеном. Такой лентивирусный вектор может быть получен посредством способа, описанного в настоящем описании, с использованием плазмиды, несущей трансген A1AT и промотор. Аналогично, отсутствует интрон между промотором и трансгеном A1AT в геноме вектора (пДНК1) плазмиды. Иллюстративная последовательность такой плазмиды приведена в SEQ ID NO: 9 (F/HN-SIV-hCEF-soA1AT, иллюстрированной на фиг. 15A).

Изобретение также относится к лентивирусному вектору F/HN-SIV, содержащему (i) промотор hCEF или промотор CMV и (ii) трансген FVIII; где отсутствует интрон между промотором и трансгеном. Такой лентивирусный вектор может быть получен посредством способа, описанного в настоящем описании, с использованием плазмиды, несущей трансген FVIII и промотор. Аналогично, отсутствует интрон между промотором и трансгеном FVIII в геноме вектора (пДНК1) плазмиды. Иллюстративные последовательности таких плазмид приведены в SEQ ID NO: 11-14 (иллюстрированных на фиг. 22A-E).

Лентивирусный вектор, описанный выше, содержит трансген. Трансген содержит последовательность нуклеиновой кислоты, кодирующую генный продукт, например белок.

Например, в одном варианте осуществления последовательность нуклеиновой кислоты, кодирующая CFTR, A1AT или FVIII, содержит последовательность нуклеиновой кислоты (или состоит из последовательности нуклеиновой кислоты), которая имеет по меньшей мере 90% (таким образом, как по меньшей мере 90, 92, 94, 95, 96, 97, 98, 99 или 100%) идентичности последовательности с последовательностью нуклеиновой кислоты CFTR, A1AT или FVIII соответственно. В дополнительном варианте осуществления последовательность нуклеиновой кислоты, кодирующая CFTR, A1AT или FVIII, содержит последовательность нуклеиновой кислоты (или состоит из последовательности нуклеиновой кислоты), которая имеет по меньшей мере 95% (таким образом, как по меньшей мере 95, 96, 97, 98, 99 или 100%) идентичности последовательности с последовательностью нуклеиновой кислоты CFTR, A1AT или FVIII соответственно. В одном варианте осуществления последовательность нуклеиновой кислоты, кодирующая CFTR, представлена SEQ ID NO: 7, последовательность нуклеиновой кислоты, кодирующая A1AT, представлена SEQ ID NO: 15 или последовательностью, комплементарной SEQ ID NO: 26, и/или последовательность нуклеиновой кислоты, кодирующая FVIII, представлена SEQ ID NO: 16 и 30 или посредством соответствующих последовательностей, комплементарных SEQ ID NO: 28 и 31 или их вариантов.

Термин "полипептид", используемый в настоящем описании, также охватывает варианты последовательности. Таким образом, полипептид, кодируемый трансгеном изобретения, может иметь по меньшей мере 90% (например, по меньшей мере 90, 92, 94, 95, 96, 97, 98, 99 или 100%) идентичности последовательности с функциональной полипептидной последовательностью CFTR, A1AT или FVIII соответственно. В дополнительном варианте осуществления аминокислотная последовательность CFTR, A1AT или трансгена FVIII содержит аминокислотную последовательность (или состоит из аминокислотной последовательности), которая имеет по меньшей мере 95% (например, по меньшей мере 95, 96, 97, 98, 99 или 100%) идентичности последовательности с функциональной полипептидной последовательностью CFTR, A1AT или FVIII соответственно. В одном варианте осуществления аминокислотная последовательность белка A1AT, кодируемого трансгеном изобретения, содержит аминокислотную последовательность (или состоит из аминокислотной последовательности) из SEQ ID NO: 27 или ее варианты. Предпочтительно указанный вариант белков A1AT по изобретению имеет по меньшей мере 90% (например, по меньшей мере 90, 92, 94, 95, 96, 97, 98, 99 или 100%), более предпочтительно по меньшей мере 95% или более идентичности последовательности с SEQ ID NO: 27.

В одном варианте осуществления последовательность нуклеиновой кислоты, кодирующей CFTR, A1AT или FVIII, содержит последовательность ДНК (или состоит из последовательности ДНК), которая является комплементарной последовательности ДНК CFTR, A1AT или FVII соответственно. В одном варианте осуществления CFTR, A1AT или трансген FVIII представляет CFTR, A1AT или FVIII с оптимизированной последовательностью (soCFTR2, soA1AT или FVIII). Пример представлен SEQ ID NO: 7, 15 и 16 соответственно. Иллюстративная последовательность, комплементарная последовательности A1AT с оптимизированной последовательностью, приведена в SEQ ID NO: 26. Иллюстративные последовательности, комплементарные оптимизированным последовательностям FVIII, приведены в SEQ ID NO: 28 и 31.

В одном варианте осуществления изобретения вектор F/HN экспрессии трансгена направляется промотором цитомегаловируса (CMV). В еще одном варианте осуществления вектор экспрессии трансгена направляется промотором фактора элонгации 1a (EF1a). В предпочтительном варианте осуществления вектор экспрессии трансгена направляется гибридом энхансера человеческого CMV/промотора EF1a (hCEF). В одном варианте осуществления промотор hCEF имеет замену всех динуклеотидов CG, на любой один из AG, TG или GT. Таким образом, в одном варианте осуществления промотор hCEF является не содержащим CpG.

В одном варианте осуществления лентивирусный вектор может быть получен с использованием плазмиды F/HN-SIV-hCEF-soCFTR2-IC. В этом варианте осуществления CFTR экспрессируется под управлением промотора hCEF. Этот лентивирусный вектор может быть описан, как содержащий F/HN-SIV-hCEF-soCFTR2-IC, так как он содержит F/HN элементы SIV, а также компетентную по интегразе кассету экспрессии, содержащую CFTR под управлением промотора hCEF. Этот лентивирусный вектор по изобретению способен производить долговременный, повторяемый, высокий уровень экспрессии в клетках дыхательных путей, без индуцирования чрезмерного иммунного ответа. Следовательно, изобретение относится к эффективному средству для генной терапии *in vivo*, например для переноса гена CFTR в легкое, пораженное муковисцидозом, для лечения муковисцидоза легких.

В предпочтительном варианте осуществления лентивирусный вектор может быть получен с использованием плазмиды F/HN-SIV-hCEF-soA1AT. В этом варианте осуществления A1AT экспрессируется под управлением промотора hCEF. Данный лентивирусный вектор может быть описан, как содержащий F/HN-SIV-hCEF-soA1AT, так как он содержит F/HN элементы SIV, а также кассету экспрессии, содержащую A1AT под управлением промотора hCEF. Этот лентивирусный вектор по изобретению способен производить долговременный, повторяемый, высокий уровень экспрессии в клетках дыхательных путей, без индуцирования чрезмерного иммунного ответа. Следовательно, изобретение относится к эффективному средству для генной терапии *in vivo*, например для переноса гена A1AT в легкое или нос пациента для выработки A1AT, который затем секретируется в систему кровообращения (как описано в настоящем документе). Таким образом, этот вектор и другие векторы по изобретению, содержащие трансген A1AT, могут использоваться в лечении недостаточности A1AT или при других показаниях, как описано в настоящем описании.

В еще одном предпочтительном варианте осуществления лентивирусные векторы могут быть получены с использованием F/HN-SIV-CMV-HFVIII-V3, F/HN-SIV-hCEF-HFVIII-V3, F/HN-SIV-CMV-HFVIII-N6-ко и/или F/HN-SIV-hCEF-HFVIII-N6-коплазмид. HFVIII относится к человеческому FVIII. В данном варианте осуществления FVIII экспрессируется под управлением промотора hCEF или промотора CMV. Эти лентивирусные векторы могут быть описаны как содержащие F/HN-SIV-CMV-HFVIII-V3, F/HN-SIV-hCEF-HFVIII-V3, F/HN-SIV-CMV-HFVIII-N6-ко и F/HN-SIV-hCEF-HFVIII-N6-ко соответственно, так как они включают F/HN элементы SIV, а также кассету экспрессии, содержащую FVIII под управлением промотора hCEF/промотора CMV. Вирусные векторные продукты, получаемые с использованием F/HN-SIV-CMV-HFVIII-V3, F/HN-SIV-hCEF-HFVIII-V3, F/HN-SIV-CMV-HFVIII-N6-ко и/или F/HN-SIV-hCEF-HFVIII-N6-коплазмиды, являются также известными, как VGM126, VGM127, VGM142 и VGM129 (см. фиг. 22). Эти лентивирусные векторы по изобретению способны производить долговременный, по-

вторяемый, высокий уровень экспрессии в клетках дыхательных путей, без индуцирования чрезмерного иммунного ответа. Следовательно, изобретение предоставляет эффективное средство генной терапии *in vivo*, например, для переноса гена FVIII в легкое или нос пациента для выработки FVIII, который затем секретируется в систему кровообращения (как описано в настоящем описании). Таким образом, эти векторы и другие векторы по изобретению, содержащие трансген FVIII, могут использоваться в лечении гемофилии или при других показаниях, как описано в настоящем описании.

Лентивирусные векторы по изобретению не содержат интрон между промотором и трансгеном. Аналогично, геном вектора плазмид по изобретению (используемый для генерации, указанных лентивирусных векторов, описанных в настоящем описании) также не содержит интрон между промотором и трансгеном. Изобретение, следовательно, относится в одном варианте осуществления к отсутствию интрона между промотором hCEF и кодирующими последовательностями, подлежащими экспрессии. В одном предпочтительном варианте осуществления кодирующая последовательность, подлежащая экспрессии, представляет собой последовательность нуклеиновой кислоты CFTR, A1AT и/или FVIII.

В одном варианте осуществления векторы по изобретению содержат центральный полипуриновый путь (сРРТ) и посттранскрипционные регуляторные элементы вируса гепатита сурков (WPRE). В одном варианте осуществления последовательность WPRE представлена SEQ ID NO: 8.

В одном варианте осуществления вектор по изобретению используют для генной терапии. В одном варианте осуществления заболевание, подлежащее лечению, представляет собой МВ. В еще одном другом варианте осуществления изобретения заболевание, подлежащее лечению, представляет собой первичную цилиарную дискинезию (PCD). В одном варианте осуществления вектор применяют для лечения острого повреждения легких. В одном варианте осуществления изобретения заболевание, подлежащее лечению, представляет собой недостаточность поверхностно-активного белка В (ПАБ-В), недостаточность  $\alpha$ -1-антитрипсина (A1AD), легочный альвеолярный протеиноз (РАР), хроническое обструктивное заболевание легких (ХОБЛ). В еще одном другом варианте осуществления заболевание представляет собой воспалительное, иммунное или метаболическое состояние.

Заболевание, подлежащее лечению, может являться сердечно-сосудистым заболеванием или болезнью крови, в особенности недостаточной свертываемостью крови. Таким образом, в нескольких вариантах осуществления заболевание, подлежащее лечению, представляет собой гемофилию А, гемофилию В или гемофилию С, недостаточность фактора VII и/или болезнь фон Виллебранда. В еще одном другом варианте осуществления заболевание, подлежащее лечению, представляет собой воспалительное заболевание, инфекционное заболевание или метаболическое состояние, такое как болезнь лизосомального накопления.

Неограничивающие примеры заболеваний, которые могут подвергаться лечению с использованием генной терапии A1AT согласно настоящему изобретению, включают диабет типа 1 и типа 2, острый инфаркт миокарда, ишемическую болезнь сердца, ревматоидный артрит, воспалительное заболевание кишечника, отторжение трансплантата, реакцию трансплантат против хозяина (GvH), рассеянный склероз, заболевание печени, цирроз, васкулиты и инфекции, такие как бактериальные и/или вирусные инфекции.

В одном аспекте изобретения с помощью вектора можно эффективно лечить заболевание посредством получения трансгена для коррекции заболевания. Например, посредством ввода функциональной копии гена CFTR для облегчения протекания или предотвращения заболевания легких у пациентов с муковисцидозом, независимо от лежащей в основе мутации.

В еще одном другом варианте осуществления изобретение относится к способу получения лентивируса. В этом варианте осуществления способ по изобретению представляет собой масштабируемый, соответствующий требованиям GMP способ. Таким образом, способ по изобретению обеспечивает генерацию очищенных F/НN векторов с высоким титром.

Способ по изобретению включает следующие стадии:

- (a) выращивание клеток в суспензии;
- (b) трансфекция клеток с использованием одной или нескольких плазмид;
- (c) добавление нуклеазы;
- (d) сбор лентивируса;
- (e) добавление трипсина и
- (f) очистка.

В одном варианте осуществления способа по изобретению одна или несколько плазмид предоставляют геном вектора, Gag-Pol, Rev, F и НN. Таким образом, могут использоваться пять плазмид для каждого из генома векторов, Gag-Pol, Rev, F и НN соответственно. В предпочтительном способе по изобретению с использованием 5 плазмид, плазида с геномом вектора кодирует весь генетический материал, который упакован в конечный лентивирусный вектор, включающий трансген. Обычно только часть генетического материала, обнаруженного в плазмиде с геномом вектора, заканчивается в вирусе. Плазида с геномом вектора может обозначаться в настоящем описании как "пДНК1". Другие четыре плазмиды являются производящими плазидами, кодирующими белки Gag-Pol, Rev, F и НN. Эти плазмиды могут обозначаться как "пДНК2а", "пДНК2b", "пДНК3а" и "пДНК3b" соответственно.

В одном варианте осуществления изобретения лентивирус представляют собой SIV, такой как SIV1,

предпочтительно SIV-AGM. В одном варианте осуществления белки F и HN являются производными парамиксовируса, такого как вирус Сендай. В одном варианте осуществления плаزمид с геномом вектора (пДНК1) содержит трансген и промотор трансгена.

В конкретном варианте осуществления, относящемся к CFTR, пять плазмид характеризуются фиг. 1A-1E, таким образом, пДНК1 является плазмидой pGM326 фиг. 1A, пДНК2а является плазмидой pGM299 фиг. 1B, пДНК2b является плазмидой pGM299 фиг. 1C, пДНК3а является плазмидой pGM301 фиг. 1D и пДНК3b является плазмидой pGM303 фиг. 1E. В данном варианте осуществления окончательный CFTR-содержащий лентивирусный вектор может обозначаться как VGM058 (см. примеры). Вектор VGM058 представляет собой предпочтительный вариант осуществления изобретения.

В варианте осуществления относящемся к A1AT, пять плазмид могут характеризоваться фиг. 15A (таким образом, плазида пДНК1 может являться pGM407) и 1B-E (как указано выше для конкретного варианта осуществления с CFTR).

В варианте осуществления, относящемся к FVIII, пять плазмид могут характеризоваться фиг. 22C-F (таким образом, плазида пДНК1 может представлять собой pGM411, pGM412, pGM413 или pGM414) и 1B-E.

В этих вариантах осуществления изобретения плазида, определенная на фиг. 1A, представлена посредством SEQ ID NO: 1; плазида, определенная на фиг. 1B, представлена посредством SEQ ID NO: 2; плазида, определенная на фиг. 1C, представлена посредством SEQ ID NO: 3; плазида, определенная на фиг. 1D, представлена посредством SEQ ID NO: 4; плазида, определенная на фиг. 1E, представлена посредством SEQ ID NO: 5; плазида, определенная на фиг. 15A, представлена посредством SEQ ID NO: 9 и F/HN-SIV-CMV-HFVIII-V3, F/HN-SIV-hCEF-HFVIII-V3, F/HN-SIV-CMV-HFVIII-N6-со и/или F/HN-SIV-hCEF-HFVIII-N6-коплазмиды, определенные на фиг. 22B, представлены посредством SEQ ID NO: 11-14 соответственно.

В 5-плазмидном способе по изобретению все пять плазмид способствуют образованию окончательного лентивирусного вектора. Во время получения лентивирусного вектора плазида с геномом вектора (пДНК1) обеспечивает энхансер/промотор, Psi, RRE, cPPT, mWPRE, SIN LTR, SV40 полиA (см. фиг. 1A), которые являются важными для производства вируса. При использовании pGM326 в качестве неограничивающего примера пДНК1 CMV энхансер/промотор, SV40 polyA, colE1 Ori и KanR участвуют в продуцировании лентивирусного вектора по изобретению (например, vGM058), но не обнаруживаются в окончательном вирусном векторе. RRE, cPPT (центральный полипуриновый путь), hCEF, soCFTR2 (трансген) и mWPRE из pGM326 обнаруживаются в окончательном вирусном векторе. SIN LTR (длинные концевые повторы, самоинактивирующийся SIN/IN) и Psi (сигнал упаковки) могут обнаруживаться в окончательном вирусном векторе.

Для других лентивирусных векторов по изобретению, соответствующие элементы из других плазмид с геномом вектора (пДНК1) требуются для получения (но не обнаруживаются в окончательном векторе), или присутствуют в окончательном вирусном векторе.

Белки F и HN из пДНК3а и пДНК3b (предпочтительно белки Sendai F и HN) являются важными для инфицирования целевых клеток конечным лентивирусным вектором, т.е. для входа в эпителиальные клетки пациентов (обычно клеток легких или назальные клетки, описанные в настоящем описании). Продукты плазмид пДНК2а и пДНК2b являются важными для трансдукции вируса, т.е. для введения лентивирусной ДНК в геном хозяина. Промотор, регуляторные элементы (такие как WPRE) и трансген являются важными для экспрессии трансгена клетке (клетках) мишенях.

В одном варианте осуществления стадии (a)-(f) выполняют последовательно. В одном варианте осуществления клетки представляют собой клетки HEK293 или клетки 293T/17. В одном варианте осуществления клетки выращивают в несодержащих компонентов из животных (безсывороточных) средах. В одном варианте осуществления трансфекцию выполняют посредством применения PEIPro™. В одном варианте осуществления нуклеаза представляет собой эндонуклеазу, например Бензоназу®. В одном варианте осуществления активность трипсина представлена не имеющим животного происхождения рекомбинантным ферментом, таким как TrypLE Select™.

В одном варианте осуществления изобретения добавление нуклеазы происходит на стадии до сбора. В альтернативном варианте осуществления добавление нуклеазы происходит на стадии после сбора. В еще одном другом варианте осуществления добавление трипсина происходит на стадии до сбора. В еще одном другом варианте осуществления добавление трипсина происходит на стадии после сбора.

В одном варианте осуществления стадия очистки включает стадию хроматографии. В этом варианте осуществления используют гель-проникающую хроматографию (SEC) с комбинированным режимом. В одном варианте осуществления используют анионообменную хроматографию. В этом варианте осуществления для стадии элюирования не используют никакого градиента соли.

В одном варианте осуществления этот метод используют для получения лентивирусных векторов по изобретению. В этом варианте осуществления вектор по изобретению содержит ген CFTR, A1AT и/или FVIII. В альтернативном варианте осуществления вектор по изобретению содержит любой из указанных выше генов или генов, кодирующих указанные выше белки.



В одном варианте осуществления способа по изобретению любую комбинацию из одной или нескольких специфичных конструкций плазмид, представленных на фиг. 1А-1Е, 15А и/или 22С-22F, используют для получения вектора по изобретению.

Изобретение дополнительно относится к способу лечения заболевания, включающему введение лентивирусного вектора по изобретению пациенту. В этом варианте осуществления изобретение относится к лентивирусному вектору по изобретению для применения в лечении заболевания легких. В одном варианте осуществления заболевание является хроническим заболеванием. В конкретном варианте осуществления изобретение относится к способу лечения МВ. В других вариантах осуществления изобретение относится к способу лечения первичной цилиарной дискинезии (PCD), недостаточности поверхностно-активного белка В (ПАБ-В), недостаточности  $\alpha$ -1-антитрипсина (A1AD), легочного альвеолярного протеиноза (РАР), хронического обструктивного заболевания легких (ХОБЛ). В еще одном другом варианте осуществления заболевание представляет собой воспалительное, иммунное или метаболическое состояние.

В еще одном другом варианте осуществления заболевание может являться сердечно-сосудистым заболеванием или болезнью крови, в особенности недостаточной свертываемостью крови, таким как гемофилия А, гемофилия В, гемофилия С, недостаточность фактора VII и/или болезнь фон Виллебранда, воспалительное заболевание, инфекционное заболевание или метаболическое состояние, такое как болезнь лизосомального накопления.

Заболевание может представлять собой диабет типа 1 и типа 2, острый инфаркт миокарда, ишемическую болезнь сердца, ревматоидный артрит, воспалительное заболевание кишечника, отторжение трансплантата, реакцию трансплантат против хозяина (GvH), рассеянный склероз, заболевание печени, цирроз, васкулиты и инфекции, такие как бактериальные и/или вирусные инфекции.

Лентивирусные векторы по изобретению могут вводиться в любой дозировке, подходящей для достижения желаемого терапевтического эффекта. Соответствующие дозировки могут определяться клиницистом или другим практикующим врачом с использованием стандартных методов и в ходе их обычной работы. Неограничивающие примеры подходящих дозировок включают  $1 \times 10^8$  единиц трансдукции (ЕТ),  $1 \times 10^9$  ЕТ,  $1 \times 10^{10}$  ЕТ,  $1 \times 10^{11}$  ЕТ или более.

Изобретение также относится к композициям, содержащим лентивирусные векторы, описанные выше, и фармацевтически приемлемый носитель. Неограничивающие примеры фармацевтически приемлемых носителей включают воду, физраствор и забуференный фосфатом физраствор. В нескольких вариантах осуществления однако композиция находится в лиофилизированной форме, и в этом случае она может включать стабилизатор, такой как альбумин бычьей сыворотки (BSA). В нескольких вариантах осуществления может быть желательно составлять композицию вместе с консервантом, таким как тиморосал или азид натрия, чтобы способствовать долговременному хранению.

Векторы по изобретению могут вводиться посредством любого соответствующего пути. Может быть желательным направлять композиции согласно настоящему изобретению (описанные выше) в дыхательную систему пациента. Эффективная передача терапевтической/профилактической композиции или лекарственного средства к участку инфекции в дыхательном пути может достигаться посредством перорального или интраназального введения, например, в виде аэрозолей (например, назальных спреев) или посредством катетеров. Обычно лентивирусные векторы по изобретению являются стабильными в клинически значимых ингаляторах, катетерах и аэрозолях и т.д.

Лекарственные формы для интраназального введения могут находиться в форме назальных капель или назального спрея. Интраназальная лекарственная форма может содержать капли, имеющие приблизительные диаметры в интервале 100-5000 мкм, такие как 500-4000 мкм, 1000-3000 мкм или 100-1000 мкм. Альтернативно, с учетом объема капли могут находиться в интервале приблизительно 0,001-100 мкл, таком как 0,1-50 мкл или 1,0-25 мкл или таком как 0,001-1 мкл.

Аэрозольная лекарственная форма может принимать форму порошка, суспензии или раствора. Размер частиц аэрозоля имеет отношение к способности аэрозоля к доставке. Более мелкие частицы могут перемещаться дальше вниз по дыхательным путям в направлении альвеол, чем могли бы более крупные частицы. В одном варианте осуществления аэрозольные частицы имеют такое распределение по диаметру, чтобы способствовать доставке на протяжении полной длины бронхов, бронхиол и альвеол. Альтернативно, распределение частиц по размерам может быть выбрано, чтобы целенаправленно действовать на конкретную часть дыхательного воздухоносного пути, например альвеолы. В случае аэрозольной доставки лекарственного средства частицы могут иметь диаметры в приблизительном интервале 0,1-50 мкм, предпочтительно 1-25 мкм, более предпочтительно 1-5 мкм.

Аэрозольные частицы могут использоваться для доставки с использованием ингалятора (например, через рот) или назального спрея. Аэрозольная лекарственная форма может необязательно содержать пропеллент и/или поверхностно-активное вещество.

Используемые в настоящем описании термины "последовательность нуклеиновой кислоты" и "полинуклеотид" используются взаимозаменяемо и не подразумевают какого-либо ограничения длины. Используемые в настоящем описании термины "нуклеиновая кислота" и "нуклеотид" используют взаимозаменяемо.

заменяемо. Термины "последовательность нуклеиновой кислоты" и "полинуклеотид" охватывают последовательности ДНК (включая кДНК) и РНК. Термины "трансен" и "ген" также используют взаимозаменяемо, и оба термина охватывают их фрагменты или варианты, кодирующие целевой белок.

Трансены согласно настоящему изобретению включают последовательности нуклеиновой кислоты, которые были удалены из их природного окружения, изоляты рекомбинантной или клонируемой ДНК и химически синтезированные аналоги или аналоги, биологически синтезируемые гетерологичными системами.

Полинуклеотиды согласно настоящему изобретению могут быть получены посредством любых средств, известных в данной области. Например, крупные количества полинуклеотидов могут быть получены посредством репликации в подходящей клетке-хозяине. Фрагменты природных или синтетических ДНК, кодирующие желательный фрагмент, будут включаться в конструкции рекомбинантных нуклеиновых кислот, обычно конструкции ДНК, способные к введению в прокариотическую или эукариотическую клетку и репликации в ней. Обычно конструкции ДНК будут подходящими для автономной репликации в одноклеточном хозяине, таком как дрожжи или бактерии, но также могут быть предназначены для введения в геном культивируемых клеточных линий насекомых, млекопитающих, растений или других эукариотических клеточных линий и интеграции в их геноме.

Полинуклеотиды согласно настоящему изобретению также могут быть получены химическим синтезом, например посредством фосфорамидитного метода или триэфирного метода, которые могут выполняться на промышленно выпускаемых автоматизированных олигонуклеотидных синтезаторах. Двухцепочечный фрагмент может быть получен из одноцепочечного продукта химического синтеза либо посредством синтеза комплементарной цепи и совместного отжига цепи при соответствующих условиях или посредством добавления комплементарной цепи с использованием ДНК-полимеразы с последовательностью соответствующего праймера.

Применительно к последовательности нуклеиновой кислоты термин "изолированная" в контексте настоящего изобретения означает, что полинуклеотидная последовательность была удалена из ее природного генетического окружения и, таким образом, не содержит других чужеродных или нежелательных кодирующих последовательностей (но может включать природные 5'- и 3'-нетранслируемые области, такие как промоторы и терминаторы) и находится в форме подходящих для применения в генетически сконструированных системах продуцирования белка. Такие изолированные молекулы являются молекулами, которые отделяют от их природного окружения.

С учетом вырожденности генетического кода среди полинуклеотидов согласно настоящему изобретению возможна значительная вариабельность последовательностей. Вырожденные кодоны, охватывающие все возможные кодоны для данной аминокислоты, приведены ниже.

Аминокислота	Кодоны	Вырожденный кодон
Cys	TGC TGT	TGY
Ser	AGC AGT TCA TCC TCG TCT	WSN
Thr	ACA ACC ACG ACT	ACN
Pro	CCA CCC CCG CCT	CCN
Ala	GCA GCC GCG GCT	GCN
Gly	GGA GGC GGG GGT	GGN
Asn	AAC AAT	AAV
Asp	GAC GAT	GAV
Glu	GAA GAG	GAR
Gln	CAA CAG	CAR
His	CAC CAT	CAY
Arg	AGA AGG CGA CGC CGG CGT	MGN
Lys	AAA AAG	AAR

Met	ATG	ATG
Ile	ATA ATC ATT	ATH
Leu	CTA CTC CTG CTT TTA TTG	YTN
Val	GTA GTC GTG GTT	GTN
Phe	TTC TTT	TTY
Tyr	TAC TAT	TAY
Trp	TGG	TGG
Ter	TAA TAG TGA	TRR
Asn/Asp		RAY
Glu/Gln		SAR
Любая		NNN

Рядовой специалист в области сможет понять, что при определении вырожденного кодона, представительного для всех возможных кодонов, кодирующих каждую аминокислоту, существует пластичность. Например, несколько полинуклеотидов, охватываемых вырожденной последовательностью, могут кодировать варианты аминокислотные последовательности, но рядовой специалист в данной области сможет легко идентифицировать такие варианты последовательности при сравнении с аминокислотными последовательностями настоящего изобретения.

"Вариантная" последовательность нуклеиновой кислоты имеет существенную гомологию или существенное сходство с эталонной последовательностью нуклеиновой кислоты (или ее фрагментом). Последовательность нуклеиновой кислоты или ее фрагмент являются "гомологичными по существу" (или "существенно идентичными") с эталонной последовательностью, если при оптимальном выравнивании (с соответствующими нуклеотидными вставками или делециями) с другой нуклеиновой кислотой (или ее комплементарной цепью) идентичность нуклеотидной последовательности составляет по меньшей мере приблизительно 70, 75, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98 или 99% нуклеотидных оснований. Методы определения гомологии последовательностей нуклеиновых кислот являются известными в данной области.

Альтернативно, "вариантная" последовательность нуклеиновой кислоты является, по существу, гомологичной (или, по существу, идентичной) эталонной последовательности (или ее фрагменту), если "вариантная" и эталонная последовательности способны к гибридизации при жестких (например, очень жестких) условиях гибридизации. На гибридизацию последовательности нуклеиновой кислоты будут воздействовать такие условия, как концентрация соли (например, NaCl), температура или органические растворители, в дополнение к составу оснований, длине комплементарных цепей и числу несовпадений нуклеотидных оснований между гибридизирующимися нуклеиновыми кислотами, что специалисты в данной области смогут легко понять. Предпочтительно используются жесткие температурные условия, и они включают температуры, как правило, превышающие 30°C, обычно превышающие 37°C и предпочтительно превышающие 45°C. Жесткие солевые условия будут, как правило, составлять меньше чем 1000 мМ, обычно меньше чем 500 мМ и предпочтительно меньше чем 200 мМ. pH обычно находится между 7,0 и 8,3. Комбинация параметров является гораздо более важной, чем любой одиночный параметр.

Методы определения процента идентичности последовательности нуклеиновой кислоты являются известными в данной области. Посредством примера при оценке идентичности последовательности нуклеиновой кислоты последовательность, имеющая заданное число смежных нуклеотидов, может быть выравнена с последовательностью нуклеиновой кислоты (имеющей такое же самое число смежных нуклеотидов) из соответствующей части последовательности нуклеиновой кислоты согласно настоящему изобретению. Программные средства, известные в данной области для определения процента идентичности последовательности нуклеиновой кислоты, включают Nucleotide BLAST.

Рядовой специалист в данной области поймет, что различные виды проявляют "преимущественную частоту использования кодона". Используемый в настоящем описании термин "преимущественная частота использования кодона" относится к кодонам, которые наиболее часто используются в клетках некоторых видов, таким образом, благоприятствуя одному или нескольким представителям возможных кодонов, кодирующих каждую аминокислоту. Например, аминокислота треонин (Thr) может кодироваться ACA, ACC, ACG или ACT, но в клетках-хозяевах млекопитающих ACC является наиболее часто используемым кодоном; у других видов, преимущественными могут быть другие кодоны. Преимущественные кодоны для конкретного вида клеток-хозяев могут вводиться в полинуклеотиды согласно настоящему изобретению посредством различных методов, известных в данной области. Введение последовательно-

стей преимущественного кодона в рекомбинантную ДНК может, например, увеличить выработку белка, делая трансляцию белка более эффективной в конкретном типе или виде клеток.

Таким образом, в одном варианте осуществления изобретения последовательность нуклеиновой кислоты является кодон-оптимизированной для экспрессии в клетке-хозяине.

"Фрагмент" полинуклеотида, представляющий интерес, содержит ряд последовательных нуклеотидов из последовательности указанного непроцессированного полинуклеотида. Посредством примера "фрагмент" полинуклеотида, представляющий интерес, может иметь (или состоять из) по меньшей мере 30 последовательных нуклеотидов из последовательности указанного полинуклеотида (например, по меньшей мере 35, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950 или 1000 последовательных остатков нуклеиновой кислоты указанного полинуклеотида). Фрагмент может включать по меньшей мере одну антигенную детерминанту и/или может кодировать по меньшей мере один антигенный эпитоп, соответствующий полипептиду, представляющему интерес. Обычно фрагмент, определенный в данном описании, сохраняет такую же функцию, как и непроцессированные полинуклеотид или полипептид.

### Описание чертежей

Настоящее изобретение теперь будет описано посредством примера только со ссылкой на сопроводительные чертежи, на которых

фиг. 1 показывает иллюстративные плазмиды, используемые в изобретении. Фиг. 1А-1Е показывают схематические чертежи плазмид, используемых для получения векторов по изобретению. В одном варианте осуществления изобретения фиг. 1А предоставляет инструмент изобретения;

фиг. 2 показывает продолжительность экспрессии трансгена F/HN-SIV в мышинной назальной ткани, перфузию которой проводили  $4 \times 10^8$  ЕТ F/HN-SIV-CMV-EGFP, и экспрессию EGFP, определенную через указанное число дней после лечения. Показан отрицательный контроль, где перфузию назальной ткани проводили разбавителем вектора (PBS);

фиг. 3 демонстрирует продолжительность экспрессии трансгена F/HN-SIV, показывая ее через 1 год после обработки. Экспрессию EGFP определяли у 10 независимых мышей через 360 дней после обработки;

фиг. 4 показывает клеточное распределение экспрессии трансгена F/HN-SIV. Экспрессию EGFP определяли в гистологических срезах носовой полости мышей (2 мм от кончика носа) через 30 дней после обработки. EGFP-положительные клетки производили белый точечный сигнал;

фиг. 5 показывает типы клеток, трансдуцированные с использованием обработки мышинного носа F/HN-SIV. 69% клеток, трансдуцированных в носовую полость мышей, представляли собой реснитчатые респираторные эпителиальные клетки. Другие трансдуцированные клеточные типы включали нейронные клетки в обонятельном эпителии (21%) и сквамозные клетки (7%);

фиг. 6 показывает повторное введение F/HN-SIV в нос мышей. Мышиная назальная ткань была трансдуцирована (как на фиг. 1) с использованием одной дозы F/HN-SIV-CMV-Lux или двух доз F/HN-SIV-CMV-EGFP, с последующим введением одной дозы F/HN-SIV-CMV-Lux при 28-дневных интервалах. Таким образом, повторное введение F/HN-SIV в нос мыши не изменяет уровней экспрессии гена. Экспрессию трансгена сравнивают с ведущей невирусной лекарственной формой для переноса генов (плаزمид CMV-Lux в комплексе с GL67A);

фиг. 7 отображает трансдукцию человеческих респираторных клеточных культур поверхности раздела воздух-жидкость (ALI). Человеческие культуры ALI трансдуцировали с использованием F/HN-SIV-Lux при указанной множественности заражения (MOI) и визуализировали по экспрессии Lux через 5 дней после обработки;

фиг. 8 демонстрирует что F/HN-SIV может направлять функциональную экспрессию CFTR. Клетки HEK293T трансдуцировали с использованием F/HN-SIV-CMV-EGFP-CFTR при 500 MOI и функциональную активность CFTR определяли по эффлюксу йодида. F/HN-SIV-CMV-EGFP служила в качестве отрицательного контроля;

фиг. 9 показывает, что F/HN-SIV эффективно трансдуцирует овечьи и человеческие первичные клетки легкого и легкой мышцы. Фиг. 9А показывает, что трансдукция человеческих клеток соскоба слизистой носа (MOI 250) и срезов человеческого и овечьего легкого, культивированных *ex vivo* (1Е7 ЕТ/срез) вместе с F/HN-SIV-CMV-Lux, приводит к значительной экспрессии трансгена люциферазы через 24-48 ч после трансдукции. Фиг. 9В и 9С отражают трансдукцию ( $\sim 1 \times 10^5$ ) первичных человеческих клеток легкого, пораженного МВ, культивированных на поверхности раздела воздух-жидкость (CF hALIs) с использованием векторов (3Е7 ЕТ) F/HN-SIV-SOCFTR2, содержащих CMV- и hCEF промоторы трансгена. (В) Число копий вектора (копий провирусной ДНК на копию эндогенной ДНК CFTR) через 6-8 дней после трансдукции. (С) Уровень экспрессии мРНК CFTR (%VE: копии мРНК CFTR на копию эндогенной мРНК CFTR  $\times 100$ ) через 6-8 дней после трансдукции. Горизонтальная линия на (С) представляет целевой уровень экспрессии, равный 5% VE, как полагают, представляющий терапевтический порог. После *in vivo* доставки векторов F/HN-SIV-EGFPLux, содержащих CMV-, EF1aS и промотор hCEFs в дефектной по интегразе (ID) или компетентной по интегразе форме (IC или без метки), экспрессию

трансгена в клетках дыхательных путей определяли в назальном эпителии (D) и эпителии легких (E) мышей (n=6-10/группу). Временную динамику экспрессии трансгена люциферазы регистрировали посредством визуализации биолюминесценции *in vivo* и нормализовали по доставляемой дозе. Фиг. 9F показывает представительные изображения биолюминесценции после трансдукции мышам *in vivo* на день 14 после трансдукции. Фиг. 9G отображает представительные изображения биолюминесценции после трансдукции *in vitro* не-MB hALI на день 5-6 после трансдукции. Фиг. 9H представляет экспрессию EGFP через 14 дней после трансдукции в мышинный назальный эпителий после доставки 1,6E8 ET F/HN-SIV-hCEF-EGFP<sub>Lux</sub> (vGM020). EGFP визуализировали посредством иммуногистохимии, ядра визуализировали посредством DAPI. Фиг. 9I показывает временную динамику экспрессии трансгена люциферазы у не-MB ALI, которую регистрировали посредством повторной визуализации биолюминесценции и нормализовали по доставляемой дозе;

фиг. 10 показывает, что F/HN-SIV эффективно трансдуцирует легкое овцы *in vivo*. Фиг. 10A показывает, что для моделирования вирусной доставки в легкое овцы авторы изобретения инстиллировали (аликвоты 3×100 мкл в течение ~5 мин) акрифлавин в проксимальные дыхательные пути при прямой бронхоскопической визуализации. Распределение акрифлавина можно оценить по оранжевому окрашиванию исеченных дыхательных путей после смерти. Следует отметить, что акрифлавин в значительной степени ограничен проводящими воздухоносными путями и отсутствует в альвеолярных областях. Стрелка указывает приблизительный участок инстилляции. Числа на линейке представляют собой см. Фиг. 10B является представлением в виде диаграммы легкого овцы (центр/верхняя часть трахеи). Зеленый круг представляет область на (A). На (B) линия указывает прохождение бронхоскопа для доставки аликвот 3×100 мкл 2,2E9 ET/мл (всего 6,6E8 ET) F/HN SIV CMV-EGFP<sub>Lux</sub> n=3 индивидуальным оцам (коды животных T121, T156 и T251). Через семь дней после доставки 5-6 блоков образцов ткани были взяты при вскрытии с интервалами ~1 см от участка инстилляции. Блоки были разделены на 2-3 приблизительно эквивалентных образца и проанализированы на предмет экспрессии трансгена посредством анализа люциферазы (C), нормализованного по содержанию белка; и (D) количественная РВ-ПЦР, нормализованная по уровням эндогенной мРНК CFTR. Горизонтальная линия на (C) представляет наивысшую активность люциферазы, отмеченную в каком-либо образце, обработанном с использованием невирусного вектора переноса генов, и (D) целевой уровень экспрессии, равный 5% VE, как полагают, представляющий терапевтический порог;

фиг. 11 изображает получение и очистку F/HN векторов SIV. F/HN векторы SIV получали посредством 5-плазмидной (пДНК) PEI-опосредованной временной трансфекции клеток 293T, выращенных в суспензии в масштабе 1 л в биореакторах WAVE с контролируемым pH (GE), с применением масштабированных способов изобретения. Векторы осветляли посредством глубоинной/конечной фильтрации (GE/Pall), загрязняющие нуклеиновые кислоты удаляли с использованием Бензоназы® (Merck), векторы активировали с использованием TрупLE Select™ (Life Technology), очищали и концентрировали посредством анионообменной мембранной хроматографии (Pall) и тангенциальной поточной фильтрации (Spectrum). Все емкости, контейнеры и колонки для процесса представляли собой одноразовые изделия в соответствии с требованиями cGMP. Все реагенты, за исключением плазмидной ДНК, не имели животного происхождения в соответствии с требованиями cGMP. Показаны данные для различных конфигураций вектора (промотор трансгена, трансген, статус интегразы). Физические и функциональные титры определяли, используя К-ПЦР. (A) Физический титр от первоначального осветленного сбора и конечный очищенный продукт. Средний выход процесса составляет ~44%. (B) Функциональный титр конечного продукта. Средний функциональный титр составляет ~2×10<sup>9</sup> ET/мл продукта и ~3×10<sup>9</sup> ET/л объема биореактора. Целевая производительность и выходы были превышены. В нижних выходящих векторах CFTR используют промотор CMV трансгена. В верхних выходящих векторах CFTR используют промоторы EF1aS и hCEF трансгена. (C) Частицы конечного продукта: соотношение инфекционности является плотно кластеризованным и аналогичным значениям от других производителей векторов с высоким качеством (Oxford BioMedica, BlueBirdBio). Медианная частица: соотношение инфекционности составляет ~300. Консистенции продукта достигали с использованием методов "Дизайна экспериментов", подтверждающего окончательный переход к процессу производства на основании разрешения контролирующего ведомства на основе QbD. (D) Функциональный титр окончательного продукта строго коррелирует с первоначальным физическим титром, указывающим на ненасыщающие условия с использованием концентрации вектора, не ограничивающей стадии процесса - предполагают, что масштабирование очистки будет эффективным;

фиг. 12 отображает профилирование участка ввода (IS) и выживаемость трансдуцированных мышей. Фиг. 12A-F представляет сравнение профилей IS от F/HN-SIV-трансдуцированного мышинного легкого и трансдуцированной вектором VSV-G-HIV мышинной сетчатки (Bartholomae et al., Mol. Ther (2011) 19:703-710). Профили IS для (A, C, E) были получены, исходя из результатов глубокого секвенирования ДНК легкого от двух мышей, трансдуцированных F/HN-SIV, и для (B, D, F) по результатам для последовательностей IS сетчатки с VSV-G-HIV (Schmidt laboratory). (A, B) Агрегированные участки IS (легкое, 2862; сетчатка, 262), нанесенные на кариограммы, генерируемые посредством геномного браузера En-

sembl. (C, D) Расстояния от IS до сайтов начала транскрипции (TSS). Число IS в каждом бине расстояний показаны полосками выше. Сумма превышает общее число проанализированных IS, поскольку типичный IS расположен рядом с несколькими TSS. Графики генерировали посредством применения геномного браузера UCSC и геномного анализатора GREAT (great.stanford.edu). (E, F) QuickMap (www.gtsg.org) и сравнения между случайными (ромб) и наблюдаемыми (квадрат) частотами введения на хромосому. (G) Выживаемость мышей, обработанных предшественником F/HN-вектор SIV (вирусный вектор, изготовленный согласно известному способу с использованием прилипающих клеток: черная линия, 24 месяца от данных) и вектор текущей генерации (GTC: темно-серая линия, данные для 8 месяцев) в сравнении с мышами, обработанными буфером (светло-серая линия, данные для 24 месяцев). Были объединены данные от различных экспериментов, включающих мышей, обработанных буфером или  $\sim 1E7$  ЕТ вируса посредством назального нюханья;

фиг. 13 показывает hCEF-опосредованную респираторную экспрессию трансгена - с использованием лентивирусного переноса гена с использованием обогащенного CpG трансгена - экспрессию трансгена (среднее свечение п/с/см<sup>2</sup>/ср) в зависимости от дней после дозирования. Высокие уровни экспрессии репортерного гена люциферазы Gaussia в сравнении с контролем (интактное легкое) наблюдали как в легком (квадрат), так и носу (круг) в течение по меньшей мере 168 дней после дозирования;

фиг. 14 показывает эффект трансдукции человеческих кишечных органоидов с использованием лентивирусного вектора CFTR (vGM058) изобретения. Левосторонняя панель показывает, что индуцированное форсколином набухание значительно ( $p < 0,001$ ) снижалось в трансдуцированных VGM058 органоидах. На правосторонней панели клетки A549 трансдуцировали VGM058 или контрольным вирусом, и функцию CFTR количественно оценивали с использованием анализа эффлюкса радиоактивного йодида. Значительные ( $p < 0,05$ ) уровни CFTR-опосредованного эффлюкса йодида обнаруживали в трансдуцированных VGM058 клетках;

фиг. 15A показывает схематические изображения плазмиды, применяемой для получения векторов A1AT изобретения. В одном варианте осуществления изобретения фиг. 15A представляет инструмент изобретения. Фиг. 15B показывает контрольную плазмиду, кодирующую ген-репортер люциферазы Gaussia;

фиг. 16 показывает, что F/HN-SIV эффективно трансдуцирует первичные клетки легкого человека в культуре ALI. В частности, трансдукция человеческих культур ALI приводит к значительной экспрессии трансгена люциферазы в течение по меньшей мере 80 дней после трансдукции. Каждая точка представляет среднее значение ОСЕ/мкл в среде для  $n=6$  ALI в показанной временной отметке. Вертикальные черточки представляют стандартную ошибку среднего;

фиг. 17 показывает экспрессию гена после трансдукции срезов человеческого легкого ( $n=6$  на группу) с использованием SIV1 hCEF-sogLux (фиг. 17A) и SIV1 hCEF-sohAAT (фиг. 17B). Наблюдали высокие уровни экспрессии. Каждая точка представляет среднее значение ОСЕ/мкл в среде для  $n=6$  срезов легкого в показанной временной отметке. Вертикальные черточки представляют стандартную ошибку среднего;

фиг. 18 показывает долговременную экспрессию ( $>12$  месяцев) люциферазы Gaussia после опосредованного лентивирусом переноса генов (SIV1 hCEF-soGLux) *in vivo*. А: гомогенат ткани легкого; В: жидкость бронхоальвеолярного лаважа (BAL); С: сыворотка. Каждая точка представляет среднее значение ОСЕ/мкл в одной группе животных ( $n=5$  или 6 на группу) при сборе в указанной временной отметке. Вертикальные черточки представляют стандартную ошибку среднего;

фиг. 19 показывает высокие уровни экспрессии A1AT после опосредованного лентивирусом переноса гена (SIV1 hCEF-sohAAT) *in vivo*. Каждая точка представляет одно животное. Горизонтальные черточки представляют медиану для каждой группы;

фиг. 20 показывает долговременную экспрессию A1AT после опосредованного лентивирусом переноса гена AAT (A1AT) *in vivo*. А: гомогенат легкого; В: BAL и С: сыворотка;

фиг. 21 показывает уровень A1A1 в бронхоальвеолярной лаважной жидкости после опосредованного лентивирусом переноса гена (SIV1 hCEF-sohAAT) *in vivo*. Опосредованный лентивирусом перенос гена (SIV1 hCEF-sohAAT) *in vivo*;

фиг. 22A показывает схематические изображения конструкций кДНК FVIII, используемых для получения векторов FVIII изобретения. Фиг. 22B показывает вирусные векторы по изобретению. В одном варианте осуществления изобретения фиг. 22B представляет инструмент изобретения. Фиг. 22C-F показывают схематические изображения плазмид пДНК1, используемых для получения векторов FVIII изобретения. В одном варианте осуществления изобретения фиг. 22C-F представляют инструменты по изобретению;

фиг. 23 показывает эффективность трансдукции HEK293T с использованием vGM142 (SIV-F/HN-FVTP-N6-co) для загрузки 1 (А) и загрузки 2 (В). Графики показывают активность FVIII для каждого МОИ через 48 и 72 ч после трансдукции. Каждый символ представляет независимый эксперимент ( $n=5-6$  экспериментов). Горизонтальная черта указывает на среднее для группы  $\pm$  SD. Анализ выполняли с использованием однофакторного дисперсионного анализа Anova (GraphPad Prism) с множеством сравнений с необработанным контролем \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ; \*\*\*\*  $p < 0,0001$ ;

фиг. 24 показывает оценку VGM142 в системе *in vivo* (мышинная модель). А: легкое; В: жидкость BAL; С: плазма. Графики показывают уровень hFVTII (в виде процентной доли от нормального уровня) в сравнении с различными обработками для групп 1-3 (группы 1 и 2 - 10 дней обработки, 3 дозы/в неделю, по 100 мкл вектора на мышь (общее количество доз было равно 3 на животное); группа 3 - 28 дней обработки, 3 дозы/в неделю, по 100 мкл вектора на мышь (общее количество доз было равно 12 на животное)). Каждый символ представляет индивидуальную мышь (группа 1 (n=4), обработанная общей дозой вектора, равной  $1,4 \times 10^6$  ЕТТ/мышь; группа 2 (n=3), обработанная общей дозой вектора, равной  $1,57 \times 10^8$  ЕТТ/мышь; и группа 3 (n=4), обработанная общей дозой вектора, равной  $3,36 \times 10^8$  ЕТТ/мышь). Горизонтальная черта показывает среднее значение FVIII: уровни  $Ag \pm SD$ . Анализ выполняли с использованием однофакторного дисперсионного анализа Anova (GraphPad Prism) с множественными сравнениями между группами обработки \*\*\*\*  $p < 0,0001$ .

#### Примеры

Изобретение теперь будет описано со ссылкой на примеры, приведенные ниже. Эти примеры не являются ограничивающими объем изобретения, и специалист в данной области смог бы понять, что могли бы использоваться подходящие эквиваленты в пределах объема настоящего изобретения. Таким образом, примеры могут рассматриваться как компонент частей изобретения, и индивидуальные аспекты, описанные в них, могут рассматриваться как раскрытые независимо или в любой комбинации.

Пример 1. Культура клеток.

Клетки HEK293T, Freestyle 293F (Life Technologies, Paisley, UK) и 293T/17 (CRL-11268; ATCC, Manassas, VA) поддерживали в минимальной питательной среде Игла в модификации Дульбекко (Invitrogen, Carlsbad, CA), содержащей 10% сыворотки крови эмбрионов коров и дополненной пенициллином (100 Ед/мл) и стрептомицином (100 мкг/мл) или среде для экспрессии Freestyle™ 293 (Life Technologies).

Пример 2. Построение плазмид.

pCAGGS-Fct4 и pCAGGS-SIVct+HN были построены следующим образом.

(i) Плазида SIVct/HN содержит ген, кодирующий цитоплазматический концевой сегмент SIVagm TMP (обратный), гибридный с внеклеточной частью вируса и трансмембранными областями белка SeV HN. Были синтезированы три олигонуклеотидные пары:

пара 1, 5' -  
TCGAGATGTGGTCTGAGTTAAAAATCAGGAGCAACGACGGAGGTGAAGGACCAGACGCCAACG  
ACCC-3' (SEQ ID NO: 18) и 5' -  
CCGGGGTTCGTTGGCGTCTGGTCCTTCACCTCCGTCGTTGCTCCTGATTTTTAACTCAGACCA  
CATC-3' (SEQ ID NO: 19); пара 2, 5' -  
CCGGGAAAGGGGGTGAACACATCCATATCCAGCCATCTCTACCTGTTTATGGACAGA-3'  
(SEQ ID NO: 20) и 5' -  
ACCTCTGTCCATAAACAGGTAGAGATGGCTGGATATGGATGTGTTGCACCCCTTTCC-3'  
(SEQ ID NO: 21); и пара 3, 5' -  
GGGTTAGGTGGTTGCTGATTCTCTCATTCACCCAGTGGG-3' (SEQ ID NO: 22) и  
5' -GATCCCCACTGGGTGAATGAGAGAATCAGCAACCACCTA-3' (SEQ ID NO: 23).

Эти олигонуклеотидные пары подвергали отжигу и клонировали в сайты XhoI и BamHI pBluescript KS+ (Stratagene) с получением pKS+SIVct. pCAGGS-SIVct/HN была построена посредством клонирования фрагмента XhoI-DraIII с 160 по из pKS+SIVct и фрагмента DraIII-Bsu36I с 1500 по из pCAGGS-HN, который несет ген HN дикого типа (HNwt) в сайте XhoI вектора pCAGGS, в сайты XhoI и Bsu36I pCAGGS. Эта плазида была построена таким образом, что цитоплазматический концевой сегмент белка HN был заменен на цитоплазматический концевой сегмент SIVagm TMP.

Для построения pCAGGS-SIVct+HN, гены, кодирующие цитоплазматический концевой сегмент SIVagm TMP и N-конец белка HN, сначала амплифицировали посредством ПЦР из pCAGGS-SIVct/HN с использованием пары праймеров

5'-GAGACTCGAGATGTGGTCTGAGTTAAAAATCAGG-3' (SEQ ID NO: 24) и  
5'-AGAGGTAGACCAGTACGAGTCACGTTTGGCCCTATCACCATCCCTAACCTCTGTCATAAAC-3'  
(SEQ ID NO: 25). Полученный в результате ПЦР фрагмент клонировали в сайты XhoI и AccI pKS+SIVct для генерации pKS+SIVct-H. Затем фрагмент XhoI-DraIII из pKS+SIVct-H и фрагмент DraIII-Bsu36I из pCAGGS-HN клонировали в сайты XhoI и Bsu36I pCAGGS с получением pCAGGS-SIVct+HN.

Последовательности cPPT и WPRE вводили в производную от SIV плазмиду для переноса генов. Пример использованной последовательности WPRE представлен в SEQ ID NO: 8.

Плазида pGM101 содержит точку начала репликации colE1, канамицин-устойчивый ген и промотор и была создана посредством синтеза синтетического гена (GeneArt, Regensburg, Germany; теперь Life

Technologies Ltd).

Гибрид CMV/SIV R U5 LTR, частичные последовательности Gag, RRE, cPPT, SIN U3 и R из pBS/CG2-Rc/s-CMV-D U (Nakajima et al. 2000 Human Gene Therapy 11:1863) амплифицировали посредством ПЦР и вводили наряду с последовательностью энхансер hCEF/промотор, амплифицированной посредством ПЦР из pGM169 (Hyde et al. Nature Biotechnology 26:549), и кДНК soCFTR2, выделенной из pGM169 на фрагмент фермента рестрикции NheI-ApaI в pGM101, для создания pGM326.

Энхансер CMV/промотор куриного  $\beta$ -актина наряду с ассоциированными последовательностями экзон/интрон, последовательностями SIV GagPol и RRE и поли A SV40/точка начала репликации амплифицировали посредством ПЦР из pCAGGS/Sagm-gtr (Nakajima et al. 2000 Human Gene Therapy 11:1863) для создания pGM297. Энхансер CMV/промотор наряду с ассоциированными последовательностями экзон/интрон и последовательностью полиА SV40 из pCI (Promega, Madison, WI, USA) изолировали на фрагмент фермента рестрикции BglIII-BamHI и последовательность SIV Rev, полученную из pCAGGS/Sagm-gtr, амплифицированную посредством ПЦР, вводили в pGM101 для создания pGM299.

Энхансер CMV/промотор куриного  $\beta$ -актина наряду с ассоциированными последовательностями экзон/интрон, кДНК Fct4 и полиА SV40/точка начала репликации из pCAGGS-Fct4 изолировали на фрагмент фермента рестрикции Sall-HindIII посредством комбинации генного синтеза, ПЦР и рекомбинации фрагмента фермента рестрикции и вводили в pGM101 для создания pGM301.

Энхансер CMV/промотор куриного  $\beta$ -актина наряду с ассоциированными последовательностями экзон/интрон, кДНК SIVct+HN и полиА SV40/точка начала репликации из pCAGGS-SIVct+HN изолировали на фрагмент фермента рестрикции Sall-HindIII посредством комбинации и вводили в pGM101 для создания pGM303.

Другие pGM-плазмиды изобретения получали, используя стандартные методы и согласно приведенному выше раскрытию.

На всем протяжении этих вариантов выполнения сборки плазмидных ДНК, ферменты рестрикции и ПЦР-полимеразы поставлялись New England Biolabs (Ipswich, MA, USA), а реагенты для очистки ДНК поставлялись Qiagen (Limburg, Netherlands).

Пример 3. Получение вектора SIV.

Система из четырех плазмид.

Дефектный по репликации самоактивирующийся вектор SIV был построен с минорными модификациями. Вкратце, SeV-F/HN-псевдотипированный вектор SIV получали посредством трансфекции клеток 293T/17 (чашки для культивирования с диаметром 15 см) с использованием четырех плазмид, в комплексе с реагентами Липофектамин/Плюс (Invitrogen) в соответствии с рекомендациями изготовителя [плазмида-1: 10 мкг полученной из SIV плазмиды переноса, несущей GFP, ген-репортер люциферазы (lux), или гибридная конструкция GFP-CFTR, плазмида-2: 3 мкг пакующей плазмиды, плазмида-3: 2 мкг pCAGGS-Fct4, плазмида 4: 2 мкг pCAGGS-SIVct+HN; фиг. 1A-E показывают схематические чертежи плазмид, используемых для получения векторов по изобретению]. Псевдотипированный VSV-G вектор SIV получали, используя аналогичный протокол, но плазмиду pVSV-G (2 мкг; Clontech, Mountain View, CA) использовали вместо pCAGGS-Fct4 и pCAGGS-SIVct+HN. Через 12 ч после трансфекции культуральную среду заменяли на 30 мл безсывороточной модифицированной Дульбекко среды Игла, содержащей 5 ммоль/л бутирата натрия. Бутират натрия стимулирует получение вектора, чтобы ингибировать деацетилазу гистона. Супернатант культуры, содержащий вектор SIV, собирали через 48 ч после трансфекции, фильтровали через фильтровальную мембрану 0,45 мкм, и дополнительно концентрировали посредством высокоскоростного центрифугирования (20000 g в течение 4 ч при 4°C, ротор Avanti JA18; Beckman Coulter, Brea, CA). Сгустки вектора суспендировали в PBS (Invitrogen) до 100-200-кратной концентрации и хранили при -80°C.

Система из пяти плазмид (предпочтительная):

SeV-F/HN-псевдотипированный вектор SIV получали посредством трансфекции клеток HEK293T или 293T/17, культивируемых в среде для экспрессии Freestyle™ 293 с использованием смеси из пяти плазмид, имеющих следующие характеристики: пДНК1 (например, pGM326; фиг. 1A) кодирует мРНК лентивирусного вектора; пДНК2а (например, pGM297; фиг. 1B) кодирует белки SIV Gag и Pol; пДНК2b (например, pGM299; фиг. 1C) кодирует белки SIV Rev; пДНК3а (например, pGM301; фиг. 1D) кодирует белок Fct4 из вируса Сендай [Kobayashi et al., 2003 J. Virol. 77:2607]; и пДНК3b (например, pGM303; фиг. 1E) кодирует SIVct+HN из вируса Сендай [Kobayashi et al., 2003 J. Virol. 77:2607] в комплексе с PEIpro (Polyplus, Illkirch, France).

Среду для культуры клеток дополняли через 12-24 ч после трансфекции бутиратом натрия. Бутират натрия стимулирует получение вектора посредством ингибирования деацетилазы гистона, что приводит к увеличению экспрессии компонентов гибридных белков SIV и вируса Сендай, кодируемых пятью плазмидами. Среду для культуры клеток дополняли через 44-52 ч и/или 68-76 ч после трансфекции с использованием 5 единиц/мл нуклеазы бензоазы (Merck Millipore, Nottingham, UK). Супернатант культуры, содержащий вектор SIV, собирали через 68-76,5 ч после трансфекции и осветляли посредством фильтрации через мембрану 0,45 мкм. Вектор SIV обрабатывали посредством обработки протеазой, содержащей



трипсиновую активность, например, не имеющим животного происхождения, рекомбинантным ферментом, таким как TrypLE Select™. Впоследствии, вектор SIV обычно дополнительно очищают и концентрируют посредством анионообменной хроматографии и/или ультрафильтрации/диафильтрации с фильтрацией тангенциальным потоком.

Такой же метод применяли для генерации лентивирусных векторов, содержащих трансгены A1AT и FVIII, с использованием плазмид фиг. 15A и 22B, при замене плазмид фиг. 1A для обеспечения соответствующего трансгена (см. примеры 15 и 20).

Пример 4. Титрование вектора.

Метод 1.

Титр частиц определяли, используя ПЦП с обратной транскриптазой в режиме реального времени. Вирусную РНК очищали, применяя мини-набор для вирусных РНК, QIAamp (QIAGEN, Strasse, Germany), и обратно транскрибировали, используя Superscript II (Invitrogen). Систему ПЦП с зондом QuantiTest (QIAGEN) и праймеры для амплификации 131 нуклеотидов (по) перекрывающих последовательности WPRE (прямой праймер: 5'-ggatacgctgcttaaatgcc-3', обратный праймер: 5'-acgccacgtgctgcaaac-3') применяли в соответствии с протоколом изготовителя на детекторной системе для последовательностей ABI PRISM 7700 (PE Applied Biosystems, Foster City, CA). ДНК плазмиды для переноса генов SIV ( $3 \times 10^4 \times 2 \times 10^6$  молекул) применяли в качестве стандарта.

Единицы трансдукции (ЕТ/мл) определяли посредством трансдуцирования клеток 293Т серийными разведениями исходного раствора вектора и количественной оценки трансдуцированных клеток посредством флуоресценции GFP (для F/HN-SIV-GFP и VSV-G-SIV-GFP) или окрашивания с использованием антител против люциферазы (для F/HN-SIV-lux).

Метод 2 (предпочтительный).

Титр частиц (ВЧ/мл) обычно определяли, используя ПЦП с обратной транскриптазой в режиме реального времени. Вирусную РНК очищали, используя мини-набор для вирусных РНК, QIAamp (QIAGEN, Strasse, Germany), и обратно транскрибировали, используя обратную транскриптазу (Life Technologies). Систему для количественной ПЦП, TaqMan (Life Technologies), с использованием праймеров, амплифицируя часть последовательности WPRE, применяли на детекторной системе для последовательностей ABI PRISM 7700 (Life Technologies). Транскрибированные *in vitro* молекулы РНК WPRE применяли в качестве количественных стандартов.

Единицы трансдукции (ЕТ/мл) определяли посредством трансдуцирования клеток 293Т/17 или Freestyle 293F серийными разведениями вектора SIV и количественной оценки WPRE, содержащей провирусную ДНК, с использованием системы для количественной ПЦП, TaqMan (Life Technologies), используя праймеры, амплифицируя часть последовательности WPRE на детекторной системе для последовательностей ABI PRISM 7700 (Life Technologies). Молекулы ДНК плазмид, содержащих последовательности WPRE, применяли в качестве количественных стандартов.

Пример 5. Генерация культур tEC, обогащенных базальными клетками.

Мышечные трахеальные эпителиальные клетки (tEC) выделяли следующим образом.

Мышей C57BL/6N умерщвляли и трахеи иссекали от гортани до основных бронхиальных ветвей, используя стерильные хирургические инструменты. Ткани помещали в пробирку, содержащую холодную среду Хэма F-12 с 100 Ед/мл пенициллина (P), 100 мкг/мл стрептомицина (S) и 2,5 мг/мл амфотерицина В (A) (среда Хэма F12/PSA) и выдерживали на льду. В стерильном вытяжном шкафу с ламинарным потоком воздуха трахеи очищали от прилегающих мышц и соединительной ткани, разрезали в продольном направлении, чтобы экспонировать внутренний респираторный эпителий, и помещали в 0,15% раствор проназы в среде F-12 (~5 мл в пробирке 15 мл). Расщепление ткани проводили в течение ночи (15-18 ч) при 4°C. Чтобы блокировать ферментативную реакцию, 10% сыворотку крови эмбрионов коров (FBS) добавляли к образцу обработанной ткани. После плавного переворачивания пробирки для большего отсоединения клеток трахеи помещали в новую пробирку, содержащую раствор 10% FBS/Хэма F-12/PS, и переворачивали, как прежде. Эту стадию повторяли еще два раза. Содержимое четырех пробирок объединяли вместе и центрифугировали при 500 g в течение 10 мин при 4°C. Сгусток ресуспендировали в растворе ДНКазы (0,5 мг/мл неочищенной панкреатической ДНКазы плюс 10 мг/мл BSA в растворе FBS/Хэма F-12/PS, приблизительно 200 мкл/трахею), инкубировали на льду в течение 5 мин и центрифугировали, как прежде.

После удаления супернатанта tEC ресуспендировали в целевой среде для клеток-предшественников (PCT), (CnT-17, CELLnTEC, Bern, Switzerland), в не содержащем антибиотиков и противогрибковых средств составе, конкретно разработанных для выделения и пролиферации клеток-предшественников человеческих и мышечных воздухоносных путей. tEC затем переносили в чашки для культур тканей Primaria (Becton Dickinson Labware, Franklin Lakes, NJ, USA) и инкубировали в течение 3-4 ч в 5% CO<sub>2</sub> при 37°C. Неприлипшие клетки собирали и центрифугировали при 500 g в течение 5 мин при 4°C и подсчитывали на гемоцитометре. Для генерации культур, обогащенных базальными клетками, tEC суспендировали в среде PCT и высевали на чашку Nunclon™Δ (Nunc A/S, Roskilde, Denmark), покрытую 50 мкг/мл коллагена типа I из сухожилий крысиного хвоста при рекомендуемой плотности посева, равной

$4 \times 10^3$  клеток/см<sup>2</sup>. tEC также культивировали в контрольной основной среде, содержащей DMEM/Хэма F12, дополненной L-глутамином (4 мМ), NEPES (15 мМ) и NaHCO<sub>3</sub> (3,4 мМ). Чашки инкубировали при 37°C с 5% CO<sub>2</sub>. Чтобы определить долю базальных клеток в популяции tEC до и после размножения в среде PCT, цитоспиновые препараты окрашивали антителом против KRT5 и соответствующим вторичным антителом.

Пул продуцированных tEC содержал как полностью дифференцированные клетки (клетки Клара и реснитчатые клетки), так и базальные клетки. Как только их высевали на проницаемые мембраны, tEC были способны генерировать систему культивирования на границе раздела воздух/жидкость (ALI) в результате пролиферации базальных клеток и редифференциации в секреторные и реснитчатые клетки. Для установления альтернативного и быстрого протокола для размножения базальных клеток, двухмерные (2D) культуры tEC с использованием доступной для приобретения патентованной целевой среды для клеток-предшественников (PCT), специфически составленные для поддержания пролиферации клеток предшественников дыхательных путей при поддержании их в недифференцированном статусе, подвергали оценке. В качестве отрицательного контроля tEC подвергали воздействию состава основной среды без добавления специфических факторов роста. tEC, посеянные на покрытые коллагеном пластиковые поверхности ( $4 \times 10^3$  клеток/см<sup>2</sup>) и подвезнутые воздействию среды PCT, были способны быстро расти и становились конфлюентными в пределах 5-8 дней, в то время как tEC, подвезнутые воздействию контрольной среды без факторов роста, были неспособны прорасти и размножиться. Чтобы установить, приводит ли применение среды PCT к значительному обогащению популяции базальных клеток, tEC собирали при приблизительно 80% слиянии (n=6 лунок), фиксировали и обрабатывали антителами против кератина 5 (Krt5) специфичным маркером базальных клеток. Свежевыделенные tEC применяли в качестве контролей (n=3 уникальных препаратов). Доля Krt5-положительных базальных клеток после размножения в среде PCT была выше ( $78 \pm 1,4\%$ ), чем в свежевыделенных пулах tEC ( $33 \pm 0,6\%$ ), демонстрируя, что базальные клетки мышиных дыхательных путей могут селективно и быстро размножаться из смешанного пула tEC с использованием промышленной среды.

Пример 6. Трансдукция культур tEC, обогащенных базальными клетками *ex vivo* с использованием F/HN-SIV-GFP.

Чтобы определить, может ли F/HN-вектор SIV эффективно трансдуцировать базальные клетки *ex vivo*, tEC, полученные в примере 5, выращивали до приблизительно 70% слияния в течение 7 дней в среде PCT и трансдуцировали с использованием F/HN-SIV, несущего ген-репортер зеленого флуоресцентного белка (F/HN-SIV-GFP) при MOI 100 и инкубировали при 37°C с 5% CO<sub>2</sub> в течение 3 дней. tEC, полученные из животных дикого типа, и GFP-трансгенных животных культивировали в тех же самых условиях и использовали в качестве отрицательной (вирусная трансдукция отсутствует) и положительной контрольных групп соответственно (n=3-6 лунок/группу).

Чтобы количественно оценить долю GFP-положительных клеток, культуры tEC, обогащенные базальными клетками, отделяли с помощью фермента аккутазы (CELLnTEC), ресуспендировали в PBS/1% BSA и подвергали анализу методом FACS, с подсчетом в среднем  $20,277 \pm 2,478$  клеток/группу. F/HN-вектор SIV трансдуцировал  $26 \pm 0,9\%$  tEC, обогащенных базальными клетками (p<0,0001 при сравнении с нетрансдуцированными контролями).

Для оценки того, являются ли трансдуцированные GFP-экспрессирующие клетки базальными клетками, через три дня после инфицирования клетки дважды окрашивали антителами против Krt5 и GFP. Иммунофлуоресцентное окрашивание культивируемых клеток показало, что приблизительно 40% Krt5-экспрессирующих клеток также экспрессировали GFP ген-репортер, показывая, что F/HN-вектор SIV может трансдуцировать предшественники базальных клеток *ex vivo*.

Пример 7. Введение в нос мыши *in vivo*.

Использовали мышей C57BL/6N (самки, 6-8 недель). Мышей анестезировали, помещали горизонтально на спину на нагретой доске и тонкий катетер (наружный диаметр <0,5 мм) вводили на глубину ~2,5 мм от кончика носа в левую ноздрю. Используя шприцевой насос (Cole-Parmer, Vernon Hills, IL), вектор (100 мкл) затем медленно перфузировали на поверхность назального эпителия (1,3 мкл/мин) в течение 75 мин. Несмотря на перфузию вируса в левую ноздрю, авторы изобретения обычно наблюдали трансфекцию как в левой, так и правой ноздрях, что было обусловлено дисперсией растворов через целую носовую полость. Мышей, трансдуцированных PBS и VSV-G-SIV, прекоинкубированных с использованием 1% лизофосфатидилхолина, как описано Limberis et al., 2002, применяли в качестве контролей. В указанные отметки времени (3-360 дней после трансдукции), мышей умерщвляли для визуализации экспрессии GFP. Как показано на фиг. 2, экспрессию GFP наблюдали в течение по меньшей мере 449 дней после трансдукции, в то время как отрицательный контроль не показывал экспрессии GFP. Как показано на фиг. 3, экспрессия трансгена с F/HN-вектором SIV была устойчивой, с наблюдаемой экспрессией GFP, в течение по меньшей мере 360 дней после трансдукции у 10 независимо тестированных мышей.

Аналогично, как показано на фиг. 13, трансдукция F/HN-вектором SIV изобретения, содержащим промотор hCEF, приводила к долговременной экспрессии CpG-обогащенного гена-репортера (люцифера-

зы). Высокие уровни экспрессии относительно контроля наблюдали как в легких, так и носу, в течение по меньшей мере 169 дней после трансдукции.

В экспериментах с повторным введением группы мышей трансдуцировали с использованием либо одной дозы F/HN-SIV-lux (группа с одной дозой) или двух доз of F/HN-SIV-GFP (день 0, день 28), с последующей обработкой F/HN-SIV-lux в день 56 (группа с повторной дозой). Важно отметить, что мыши, получавшие F/HN-SIV-lux (группа с одной дозой), и F/HN-SIV-lux в день 56 (группа с повторной дозой), имели сходный возраст и были трансдуцированы в одно и то же время. Экспрессию гена анализировали через 30 дней после введения F/HN-SIV-lux. Для сравнения мышей трансфицировали катионным липидом GL67A в комплексе с геном-репортером люциферазы, как описано ранее (Griesenbach, U. et al., *Methods Mol. Biol.* 2008; 433:229-42) и экспрессию люциферазы измеряли через 2 дня после трансфекции.

Как показано на фиг. 6, повторное введение F/HN-SIV в нос мыши не изменяло уровней экспрессии гена. Экспрессию трансгена сравнивают с лидирующей невирусной лекарственной формой для переноса генов (плазмиды CMV-Lux в комплексе с GL67A).

Профилирование участка ввода проводили на трансдуцированных мышах и исследовали время выживаемости, как изложено при описании фиг. 12 выше. Наблюдали, что трансдукция с использованием F/HN-вектора SIV изобретения не оказывает каких-либо неблагоприятных воздействий на выживаемость мышей в сравнении с существующим F/HN-вектором SIV или отрицательным контролем (только буфер) (см. фиг. 12G).

Пример 8. Индуцированная регенерация назальных эпителиальных клеток при обработке полидоканолом.

Назальные эпителиальные клетки очищали посредством обработки полидоканолом согласно описанному методу (Borthwick et al., *Am. J. Respir Cell. Mol. Biol.* 2001 Jun; 24 (6):662-70), с некоторой модификацией. Вкратце, мышей анестезировали и 10 мкл полидоканола (2%) вводили в нос в виде болуса посредством "назального нюханья". Для подтверждения очищения и регенерации назальных эпителиальных клеток назальную ткань перфузировали 10 мкл 2% (об./об. в PBS) полидоканола (монододецилового эфира нонаэтиленгликоля; SIGMA, St Louis, MO), и проводили гистологический анализ через 24 ч и 7 дней после обработки (n=3/группа). Для анализа трансдукции возможных клеток-предшественников или стволовых клеток авторы изобретения сначала вводили вектор F/HN-SIV-GFP ( $4 \times 10^8$  ЕТ/мышь) в назальный эпителий мыши. Через семь дней после трансдукции назальную ткань перфузировали 10 мкл 2% (об./об. в PBS) полидоканола, и данную обработку повторяли снова через 3 недели. Гистологические срезы анализировали через 58 дней после введения вектора (30 дней после последней обработки полидоканолом).

Пример 9. Биоломинесцентная визуализация.

Мышам посредством внутрибрюшинной инъекции вводили 150 мг/кг D-люциферина (Xenogen, Alameda, CA) за 10 мин до визуализации и анестезировали изофлураном. Биоломинесценцию (фотоны/с/см<sup>2</sup>/ср) от живых мышей измеряли, используя систему IVIS50 (Xenogen), при группировании по 4 в течение 10 мин, применяя программу из программного обеспечения Living Image (Xenogen). Для анатомической локализации генерировали псевдоцветное изображение, представляющее интенсивность света (синий: наименее интенсивный, красный: наиболее интенсивный), используя программное обеспечение Living Image, и наложенное на полутоновое стандартное изображение. Для количественного измерения биоломинесценции в носу эмиссию фотонов в заданной области (красная коробочка) измеряли посредством маркировки стандартизированной области для количественной оценки. Размер красной коробочки сохраняли постоянным, и ее помещали над головами животных, как указано на фигуре. Важно отметить, что области маркировали, используя полутоновое стандартное изображение, чтобы избежать смещения.

Пример 10. Препарат ткани для гистологической оценки экспрессии GFP и/или обнаружения базальных клеток.

Мышей умерщвляли и кожу удаляли. Головы отрезали на уровне глаз и кожу, челюсти, язык и мягкий кончик носа осторожно удаляли. Для визуализации *in situ* экспрессии GFP в носовой полости, флуоресценцию GFP обнаруживали, используя флуоресцентную стереоскопическую микроскопию (Leica, Ernst Leitz Optische Werke, Germany). Впоследствии, ткань фиксировали в 4% параформальдегиде (pH 7,4) в течение ночи при комнатной температуре и затем погружали в 20% ЭДТА (pH 7,5 в течение 5 дней) для декальцификации. Раствор ЭДТА заменяли по меньшей мере через два дня. После декальцификации ткань инкубировали в 15% сахарозе в течение ночи при комнатной температуре и затем погружали в Tissue Mount (Chiba Medical, Soka, Japan). Срезы толщиной десять микрометров были нарезаны в шести различных положениях в каждой мышинной голове (~0-6 мм от конца носовой кости). Экспрессию GFP наблюдали, применяя флуоресцентный микроскоп (Leica). Количественную оценку и идентификацию клеточных типов проводили на шести уровнях на мышь, используя объектив  $\times 40$  или  $\times 63$ . Пролонгированная экспозиция изображения была необходимой, чтобы зафиксировать структуру назального эпителия с использованием флуоресцентной микроскопии. Это привело к пиксельному насыщению GFP-положительных клеток и вызвало то, что GFP-положительные клетки кажутся почти белыми в большей степени, чем имеют обычный зеленый внешний вид, который авторы изобретения и другие, наблюдают

при более высоком увеличении.

Клеточное распределение экспрессии трансгена F/NH-SIV исследовали в гистологических срезах. Конкретно, экспрессию EGFP определяли в гистологических срезах полости носа мыши (2 мм от кончика носа) через 30 дней после обработки. Фиг. 4 показывает расположение экспрессии EGFP (фиг. 4, белый точечный сигнал).

Фиг. 5 показывает клеточные типы, трансдуцированные посредством обработки трансгеном/NH-SIV носа мыши. 69% клеток, трансдуцированных в полости носа мыши, представляли собой реснитчатые респираторные эпителиальные клетки. Другие трансдуцированные клеточные типы включали нейрональные клетки в обонятельном эпителии (21%) и сквамозные клетки (7%).

Чтобы обнаружить базальные клетки после обработки полидоканолом, проводили иммуноокрашивание на основе пероксидазы хрена (HRP), используя набор Envision (Dako, Glostrup, Denmark). Вкратце, предметные стекла обрабатывали 0,6% пероксидом водорода в метаноле в течение 15 мин, промывали в водопроводной воде и инкубировали с 1,5% нормальной козьей сывороткой (Abscam) в течение 30 мин. Предметные стекла затем инкубировали с кроличьими поликлональными антителами против цитокератина 5 (1:500) (Abscam) в течение 1 ч с последующей инкубацией с козьими антителами против кроличьих IgG, конъюгированными с HRP (предоставленными вместе с набором) в течение 30 мин. Предметные стекла затем промывали в PBS и инкубировали с субстратом пероксидазы, 3-амино-9-этилкарбазолом (AEC) (предоставленным вместе с набором) в течение 5 мин. Окончательно предметные стекла промывали в дистиллированной H<sub>2</sub>O, докрашивали водным гематоксилином Харриса (BDH) в течение 15 с, промывали в водопроводной воде и затем в дистиллированной H<sub>2</sub>O.

Иммунофлуоресцентное обнаружение GFP-положительных трансдуцированных назальных эпителиальных клеток и Krt5 положительных базальных клеток проводили, используя следующие первичные и вторичные антитела: козы поликлональные антитела против GFP (1:250) (Abscam), кроличьи моноклональные антитела против KRT5 (1:500) (Abscam), ослиные антитела Alexa Fluor 488 против козьих IgG (1:200) (Invitrogen, Paisley, UK) и козы антитела Alexa Fluor 594 против кроличьих IgG (1:200) (Invitrogen). Чтобы улучшить характеристики антител, срезы подвергали опосредованной нагревом демаскировке антигена в цитратном буфере (10 мМ лимонной кислоты, 0,05% Твин 20, pH 6,0) в течение 20 мин на водяной бане при 100°C. Окрашенные срезы устанавливали в реагенте, препятствующем выгоранию флуоресценции ProLong® Gold Antifade, вместе с DAPI (Invitrogen) и анализировали с использованием конфокального микроскопа, как прежде (все на Zeiss). GFP-положительные базальные клетки (идентификация основана на морфологии и расположении в эпителиальном слое) количественно оценивали всего на 13 срезах/мышь. Срезы, которые проявляли предполагаемые GFP-положительные базальные клетки, были отобраны для двойного окрашивания с использованием антител против KRT5 и против GFP, чтобы подтвердить фенотип базальных клеток.

Пример 11. Трансдукция культур ALI.

Полностью дифференцированные эпителиальные клетки дыхательных путей, выращенные как культуры ALI, приобретали от Epithelix (Geneva, Switzerland). ALI трансфицировали с использованием F/NH-SIV-lux при множественности заражения в интервале от ~25 до ~300 ЕТ/клетку. Через 6 ч вирус удаляли и ALI инкубировали в течение 10-26 дней. Базолатеральную среду заменяли каждые 48 ч во время этого периода инкубации. В установленные отметки времени ALI лизировали в 100 мкл буфера для лизиса репортера и экспрессию люциферазы количественно оценивали, используя систему для анализа люциферазы (Promega, Southampton, UK) в соответствии с инструкциями производителя. Содержание общего белка в культурах количественно оценивали, используя набор для анализа белка BioRad (BioRad, Hemel Hempstead, UK). Каждый образец анализировали в двух повторностях. Экспрессию люциферазы затем представляли в виде относительных световых единиц/мг общего белка. Для биолюминесцентной визуализации 100 мкг люциферина в PBS добавляли к апикальной мембране.

Как показано на фиг. 7, клетки в культурах ALI, были успешно трансдуцированы с использованием F/NH-SIV-Lux, о чем свидетельствует экспрессия люциферазы. Экспрессия люциферазы была выше при MOI 250 в сравнении с MOI 25.

Пример 12. Анализ эффлюкса йодида.

Клетки HEK293T трансфицировали с использованием F/NH-SIV-GFP-CFTR или контрольного вируса F/NH-SIV-GFP при множественности заражения, равной 500 ЕТ/клетку, и культивировали в течение 2 дней. Активность хлоридного канала CFTR анализировали посредством измерения степени эффлюкса <sup>125</sup>йодида, как описано ранее (Derand, R., et al., 2003). Скорости эффлюкса <sup>125</sup>йодида нормализовали по времени добавления смеси форсколин/IBMX (время 0). Кривые были построены посредством нанесения на график зависимости скорости эффлюкса <sup>125</sup>йодида от времени. Чтобы отразить кумулятивные уровни эффлюкса <sup>125</sup>йодида после стимуляции агонистом, все сравнения основаны на площадях под кривой зависимости время-эффлюкс <sup>125</sup>йодида. Площадь под кривой рассчитывали по правилу трапеции. Эксперименты выполняли в двух повторностях (n=6 лунок/группу/эксперимент).

Как показано на фиг. 8, F/NH-SIV может направлять функциональную экспрессию CFTR с относительным эффлюксом <sup>125</sup>йодида, равным 0,65. Напротив, плазмидный вектор GL67A позволял достичь более низкого значения эффлюкса <sup>125</sup>йодида, равного 0,33.

Пример 13. Трансдукция овечьих и человеческих первичных клеток легкого и легкого мыши.

F/HN-SIV эффективно трансдуцирует овечьи и человеческие первичные клетки легких и легкое мыши.

F/HN-SIV-CMV-Lux применяли, чтобы трансдуцировать человеческие клетки назальных соскобов (MOI 250) и клетки срезов человеческого и овечьего легкого, культивированные *ex vivo* ( $1 \times 10^7$  ET/slice). Как показано на фиг. 9A, трансдукция как человеческих клеток назальных соскобов, так и клеток срезов человеческого и овечьего легкого приводила к существенной экспрессии трансгена люциферазы (средние значения в области  $2 \times 10^2$  ОСЕ/мг белка для человеческих назальных соскобов,  $1 \times 10^7$  ОСЕ/мг белка для срезов человеческого легкого и  $2 \times 10^7$  ОСЕ/мг белка для срезов овечьего легкого) 24-48 ч после трансдукции.

Первичные клетки легкого человека с MB, культивированные на границе раздела фаз воздух-жидкость (CF hALI,  $\sim 1 \times 10^5$ ), трансдуцировали с использованием ( $3 \times 10^7$  ET) векторов F/HN-SIV-SOCFTR2, содержащих промоторы CMV- и hCEF-трансгена. Число копий вектора (копий провирусной ДНК на копию эндогенной ДНК CFTR) измеряли через 6-8 дней после трансдукции. Промоторы как CMV, так и hCEF были способны достигать числа копий вектора, равного по меньшей мере  $1 \times 10^1$  (фиг. 9B).

Также измеряли уровень экспрессии мРНК CFTR (%VE: копий мРНК CFTR на копию эндогенной мРНК CFTR  $\times 100$ ) через 6-8 дней после трансдукции. Горизонтальная пунктирная линия на фиг. 9C представляет целевой уровень экспрессии, равный 5% VE, который, как полагают, представляет терапевтический порог. Как F/HN-SIV-soCFTR-CMV, так и F/HN-SIV-soCFTR2-hCEF индуцировали экспрессию, которая была существенно выше этой целевой отметки (в области  $40583 \pm 10687$  и  $18509 \pm 13588$  соответственно, среднее  $\pm$  SD,  $n=4$ ).

После доставки *in vivo* векторов F/HN-SIV-EGFPLux, содержащих CMV, EF1a и промоторы hCEF в дефицитной по интегразе (ID) или компетентной по интегразе форме (IC или без метки), в клетки дыхательных путей экспрессию трансгена определяли в назальном эпителии (фиг. 9D) и эпителии легких (фиг. 9E) мышей ( $n=6-10$ /группа). Временную динамику экспрессии трансгена люциферазы регистрировали посредством повторной биолюминесцентной визуализации *in vivo* и нормализовали по доставляемой дозе. Четыре из пяти тестируемых векторов (vGM012 CMV, vGM014 EF1a, vGM020 hCEF и vGM076 hCEF ID) позволяли достичь экспрессии в носу выше целевого уровня в течение всего времени протекания эксперимента. Пятый вектор, vGM074 CMV ID, позволял достичь экспрессии в носу выше принятого уровня экспрессии в течение всего времени протекания эксперимента.

Два из пяти тестируемых векторов (vGM014 EF1a и vGM020 hCEF) позволяли достичь экспрессии в легких выше целевого уровня в течение всего времени протекания эксперимента. Один вектор, vGM012 CMV, позволял достичь экспрессии в легких выше принятого уровня экспрессии в течение всего времени протекания эксперимента.

Биолюминесценцию обнаруживали после трансдукции мышам *in vivo* на день 14 после трансдукции. Представительные изображения показаны на фиг. 9F. Из пяти тестируемых векторов вектор vGM020 hCEF позволял достичь наивысшего уровня экспрессии *in vivo*.

Биолюминесценцию также обнаруживали после трансдукции *in vitro* не-MB hALI на день 5-6 после трансдукции. Представительные изображения показаны на фиг. 9G. Из пяти тестируемых векторов снова vGM020 hCEF позволял достичь наивысшего уровня экспрессии.

Фиг. 9H показывает экспрессию EGFP через 14 дней после трансдукции в мышинном назальном эпителии после доставки  $1,6 \times 10^8$  ET F/HN-SIV-hCEF-EGFPLux (vGM020), как визуализировано посредством иммуногистохимии (ядра, окрашенные DAPI).

Временную динамику экспрессии трансгена люциферазы в не-MB ALI регистрировали посредством повторной биолюминесцентной визуализации и нормализовали по доставляемой дозе. Как показано на фиг. 9I, векторы vGM014 EF1a и vGM020 hCEF позволяли достичь наивысшего уровня экспрессии.

F/HN-SIV также эффективно трансдуцирует овечье легкое *in vivo*. Акрифлавин инстиллировали (аликвоты  $3 \times 100$  мкл в течение  $\sim 5$  мин) в проксимальные дыхательные пути при прямой бронхоскопической визуализации. Распределение акрифлавина можно было оценить по оранжевому окрашиванию исеченных дыхательных путей после вскрытия (фиг. 10A).

Акрифлавин был в значительной степени ограничен проводящими воздухоносными путями и отсутствовал в альвеолярных областях. Стрелка на фиг. 10A указывает приблизительный участок инстилляции. Числа на линейке приведены в см.

Фиг. 10B является представлением в виде диаграммы овечьего легкого (центр/верхушка трахеи). Круг представляет область на фиг. 10A. На фиг. 10B стрелка указывает прохождение бронхоскопа для доставки аликвот  $3 \times 100$  мкл  $2,2 \times 10^9$  ET/мл ( $6,6 \times 10^8$  ET всего) F/HN SIV CMV- EGFPLux  $n=3$  отдельным овцам (коды животных T121, T156 и T251). Через семь дней после доставки 5-6 блоков образцов тканей были взяты после вскрытия при  $\sim 1$  см интервалах от участка инстилляции.

Блоки образцов были разделены на 2-3 приблизительно эквивалентных образца и подвергнуты анализу на предмет экспрессии трансгена, результаты которых показаны на фиг. 10C в виде анализов лю-

циферазы, нормализованных по содержанию белка; и на фиг. 10D в виде количественной РВ-ПЦР, нормализованной по уровням эндогенной мРНК CFTR. Горизонтальная линия на фиг. 10C представляет наивысшую активность люциферазы, отмеченную у какого-либо образца, обработанного невирусным вектором переноса генов, и на фиг. 10D целевой уровень экспрессии, равный 5% VE (как полагают, представляет терапевтический порог, см. выше). Для каждой группы обработки среднее значение было выше, чем сравнительное значение для невирусного вектора (фиг. 10C), и также достигало экспрессии выше целевого 5% порога VE.

Пример 14. Экспрессия и функция CFTR, измеренные с использованием человеческих кишечных органоидов при CF.

Кишечные органоиды человека с MB генерировали, как описано Dekkers J.F. et al. (Nature Medicine 2013, 19(7): 939-945). Вкратце, кишечные биопсии промывали в растворах, содержащих ЭДТА, для диссоциации клеток крипты. Клетки крипты затем трансдуцировали с использованием VGM058 (приблизительно  $1 \times 10^7$  единиц трансдукции) или контрольного вируса ( $n=3$  лунок/состояние), заливали в Matrigel и давали возможность образовать органоиды в течение 3-4 дней. Функцию CFTR оценивали посредством подвергания органоидов воздействию форсколина (приблизительно 5 мкМ форсколина), который увеличивает внутриклеточные уровни цАМФ и таким образом активирует CFTR. В ответ на активацию CFTR органоиды увеличивают транспорт хлорида, что приводит к поглощению воды и набуханию. Органоиды (минимально 10/лунку) непосредственно анализировали посредством конфокальной микроскопии живых клеток (LSM710, Zeiss,  $\times 5$  объектив). Стимулированное форсколином набухание органоидов автоматически количественно определяли с использованием программного обеспечения Volocity для визуализации (Improvision). Увеличение общей площади органоидов (плоскость  $xu$ ) относительно площади при  $t=0$  обработки форсколином рассчитывали и усредняли. Индуцированное форсколином набухание значительно ( $p < 0,001$ ) увеличивалось в трансдуцированных VGM058 органоидах в сравнении с контролями (см. фиг. 14A). Значительные уровни CFTR-опосредованного эффлюкса йодида ( $p < 0,05$ ) были также обнаружены в трансдуцированных VGM058 клетках (фиг. 14B) с использованием анализа эффлюкса йодида, раскрытого в данном описании (см. пример 12).

Пример 15. Генерация лентивирусных векторов для A1AT.

Лентивирусные векторы получали, используя остов SIV и 5-плазмидный метод, описанный выше в примерах 2 и 3 (для лентивируса CFTR), и с использованием промотора hCEF, как описано в настоящем описании. Генерировали две отдельных лентивирусных конструкции: одну с трансгеном человеческого  $\alpha$ -1-антитрипсина (hAAT); одну с трансгеном *Gaussia luciferase* (Glux) (см. фиг. 15A и B, SEQ ID NO: 9 и 10 соответственно). кДНК, содержащиеся в этих векторах, были кодон-оптимизированными и CpG-обедненными.

Пример 16. Культивирование на границе раздела фаз (ALI) с использованием лентивирусного вектора гена-репортера lux.

Полностью дифференцированные дикого типа человеческие ALI культуры (MucilAir) приобретали от Epihelix SARL (Geneva, CH). ALI культивировали при 37°C и 5% CO<sub>2</sub>, и базолатеральную культуральную среду заменяли каждые 2-3 дня. Культуральную среду хранили при -20°C до дополнительного анализа.

ALI трансдуцировали (в день 0) посредством пипетирования 100 мкл вируса ( $1 \times 10^7$  единиц трансфекции Tagman (ETT) на апикальную поверхность. Вирус удаляли через 4 ч инкубации при 37°C и базолатеральную среду заменяли.

В указанные отметки времени после трансдукции апикальную поверхность ALI промывали посредством инкубации со стерильным PBS в течение 1 ч. Промывки удаляли и хранили при -20°C до дальнейшего анализа.

Анализ люциферазы *Gaussia* (New England Biolabs, Ipswich, USA) проводили в соответствии с рекомендациями изготовителя. 15 мкл образца анализировали в двух повторностях и люминесценцию определяли на считывающем устройстве для планшетов Appliskan. Экспрессию Glux выражали в виде ОСЕ/мкл жидкости (ОСЕ=относительные световые единицы).

Фиг. 16 представляет результаты, причем каждая точка представляет среднее значение ОСЕ/мкл в среде для  $n=6$  ALI при показанной временной отметке (стандартная ошибка указана посредством планок погрешностей). Опосредованный лентивирусом перенос генов на поверхностях раздела фаз воздух-жидкость у человека приводил к продолжительной экспрессии секретируемого репортерного белка люциферазы *Gaussia*.

Пример 17. Экспрессия трансгена в культурах срезов легкого с использованием A1AT и лентивирусных векторов гена-репортера и lux.

Нарезанные с высокой точностью срезы легкого человека получали, как описано у Moreno L. et al., Respir Res 2006 Aug 21; 7:111. Срезы легкого помещали в 12-луночные планшеты для культур тканей (1 срез на лунку) в 1 мл среды и инкубировали при 37°C и 5% CO<sub>2</sub>. Среду меняли ежедневно и хранили при -20°C до дальнейшего анализа.

В день 0 срезы легкого ( $n=6$  на группу) трансдуцировали с использованием вируса SIV hCEF-

sogLux ( $1 \times 10^6$  ЕТТ) или вируса SIV1 hCEF-sohAAT ( $2 \times 10^6$  ЕТТ), разбавленного в среде до конечного объема, равного 1000 мкл, и инкубировали в течение 4 ч. После инкубации среду заменяли и хранили при  $-20^\circ\text{C}$  до дальнейшего анализа.

Экспрессию люциферазы *Gaussia* определяли, как описано выше (пример 16). Как показано на фиг. 17А, высокие уровни экспрессии секретируемого репортерного белка люциферазы *Gaussia* сопровождали опосредованный лентивирусом перенос генов в срезы легкого человека. Экспрессию ААТ (также именуемого в данном описании как А1АТ) определяли, используя сэндвич-ELISA (Abcam, Cambridge, UK), выполняемый в соответствии с рекомендациями изготовителя. 50 мкл образца анализировали в двух повторностях и измеряли на считывающем устройстве для микропланшетов при 450 нм. Как показано на фиг. 17В, высокие уровни экспрессии  $\alpha$ -1-антитрипсина (ААТ/А1АТ) сопровождали опосредованный лентивирусом перенос генов в срезы легкого человека.

Пример 18. *In vivo* введение А1АТ и лентивирусных векторов гена-репортера lux в нос мыши.

Трансдукция мышинного легкого.

Самок мышей C57BL/6 (Charles River, UK) анестезировали изофлураном и им давали 100 мкл вируса посредством назальной инстилляции, как описано у Xenariou S. et al., *Gene Ther* 2007 May;14(9): 768-75. Животным давали между 1 и 5 доз и наблюдали ежедневно на предмет признаков токсичности.

В случае люциферазы *Gaussia* самок мышей C57BL/6 анестезировали в день 0, используя изофлуран, и им давали однократную дозу 100 мкл вируса SIV1 hCEF-soGLux ( $1 \times 10^6$  ЕТТ) посредством назальной инстилляции. Контрольных животных обрабатывали DMEM (среда для культур тканей), основным компонентом препарата вируса, используемого в исследовании.

В случае А1АТ самок мышей C57bl/6 ( $n=5$  на группу) обрабатывали 3 дозами SIV1 hCEF-sohAAT при 10-дневных интервалах (100 мкл на дозу,  $6,8 \times 10^7$  ЕТТ; общая доза  $2,4 \times 10^8$  ЕТТ). Контрольных животных инстиллировали с использованием 100 мкл стерилизованного PBS (основного компонента загрузки полученного лентивируса, используемого в исследовании) в каждой отметке дозирования.

Через 10 дней после третьей дозы животных умерщвляли и гомогенат ткани легкого, жидкость бронхоальвеолярного лаважа и сыворотку анализировали на предмет экспрессии ААТ.

Дополнительно исследовали долговременную экспрессию А1АТ. В дни 1-5 эксперимента, мышей C57bl/6 обрабатывали 100 мкл SIV1 hCEF-sohAAT посредством назальной инстилляции (5 доз по  $4 \times 10^5$  ЕТТ, т.е. всего  $2 \times 10^6$  ЕТТ на животное). Контрольных животных инстиллировали 100 мкл DMEM (среда для культуры ткани), основного компонента загрузки полученного лентивируса, используемого в исследовании. Животных умерщвляли при различных отметках времени после трансдукции и гомогенат ткани легкого, жидкость бронхоальвеолярного лаважа, жидкость и сыворотку анализировали на предмет экспрессии ААТ.

Сбор мышинной ткани.

Мышей умерщвляли при указанных временных отметках после трансдукции. Кровь собирали посредством пункции левого желудочка и центрифугировали при 760 g в течение 10 мин для получения сыворотки. Сыворотку последовательно замораживали при  $-80^\circ\text{C}$ .

Бронхоальвеолярный лаваж (BAL) проводили иссечением шеи, вводом канюли в трахею и закреплением ее на месте с помощью проводника шовной нити. 500 мкл PBS инстиллировали в легкое и аспирировали три раза для получения тщательной промывки эпителиальной выстилки. Образец немедленно быстро замораживали в жидком азоте и хранили при  $-80^\circ\text{C}$  для дальнейшего анализа.

Легкие затем иссекали и быстро замораживали в жидком азоте и последовательно гомогенизировали в пробирках с лизирующим матриксом D (MP Biomedicals), центрифугировали в машине FastPrep (ThermoFisher Scientific, Waltham, MA, USA) при 4 м/с в течение 45 с и хранили при  $-80^\circ\text{C}$  для дальнейшего анализа.

Экспрессию ААТ (А1АТ) и люциферазы *Gaussia* (Glux) определяли, как описано в примерах 16 и 17 выше. *In vivo* трансдукция клеток мышинных дыхательных путей с использованием однократной дозы лентивирусного вектора гена-репортера lux примера 15 приводила к долговременной экспрессии (по меньшей мере 12 месяцев) секретируемого репортерного белка люциферазы *Gaussia*, в гомогенате легких (фиг. 18А), жидкости бронхоальвеолярного лаважа (BAL, фиг. 18В) и сыворотке (фиг. 18С).

Высокие уровни экспрессии А1АТ наблюдали в гомогенате легкого, BAL и сыворотке после опосредованного лентивирусом переноса гена ААТ (А1АТ) *in vivo* (фиг. 19) с наблюдаемым более чем 100-кратным увеличением экспрессии АТТ(А1АТ) в гомогенате легкого и BAL в сравнении с соответствующими отрицательными (PBS) контролями. Значительное увеличение (по меньшей мере на один порядок величины) экспрессии АТТ (А1АТ) также наблюдали в сыворотке.

Дополнительно длительную экспрессию (по меньшей мере в течение 90 дней)  $\alpha$ -1-антитрипсина наблюдали в гомогенате легкого (фиг. 20А), BAL (фиг. 20В) и сыворотке (фиг. 20С) после опосредованного лентивирусом переноса гена ААТ (А1АТ) *in vivo*.

Пример 19. Анализ мочи.

Мышей C57bl/6 ( $n=5$  на группу) обрабатывали 3 дозами SIV1 hCEF-sohAAT при 10-дневных интервалах (100 мкл на дозу,  $6,8 \times 10^7$  ЕТТ; суммарная доза  $2,4 \times 10^8$  ЕТТ). Контрольных животных инстиллиро-

вали с использованием 100 мкл стерилизованного PBS (основного компонента полученной загрузки лентивируса, используемой в исследовании) при каждой отметке дозирования.

Через 10 дней после третьей дозы животных умерщвляли и гомогенат ткани легкого, жидкость бронхоальвеолярного лаважа и сыворотку анализировали на предмет экспрессии A1AT.

Анализ мочи (Abscam, Cambridge, UK) проводили в соответствии с инструкциями изготовителя.

Во-первых, получали серийные разведения образцов мышинной сыворотки и жидкости BAL и анализировали их, чтобы определить соответствующие разведения для применения в дальнейших экспериментах.

Во-вторых, анализировали соответствующие образцы сыворотки и жидкости BAL от отдельных мышей (n=14), чтобы рассчитать кратность различия между концентрацией мочи в сыворотке и жидкости BAL, эквивалентную эффекту разведения BAL на бронхоальвеолярную лаважную жидкость (по Renard S.I. et al., J. Appl. Physiol. (1985), 1986 Feb; 60(2):532-8). Среднее разведение BAL было 41-кратным (интервал 24-88).

Принимая во внимание этот эффект разведения, рассчитывали концентрацию АТТ (A1AT) в бронхоальвеолярной лаважной жидкости. Конкретно, концентрацию ААТ в жидкости бронхоальвеолярного лаважа умножали на фактор разведения, чтобы получить оценку "истинной" концентрации ААТ в бронхоальвеолярной лаважной жидкости.

"Защитный" целевой уровень АТТ (A1AT) в бронхоальвеолярной лаважной жидкости (ELF, т.е. жидкости выстилающей воздухоносные пути и воздушное пространство в легких) составлял 70 мкг/мл (в сравнении с "нормальным" целевым уровнем АТТ (A1AT) в ELF, равным 200 мкг/мл). Как показано на фиг. 21, терапевтические уровни  $\alpha$ -1-антитрипсина в бронхоальвеолярной лаважной жидкости после опосредованного лентивирусом переноса генов АТТ (A1AT) *in vivo*.

Пример 20. Генерация лентивирусных векторов для FVIII.

Четыре различных лентивирусных вектора FVIII получали, используя остов SIV и 5-плазмидный метод, описанный выше в примерах 2 и 3 (для лентивируса CFTR) и 15 (для лентивируса A1AT). Промотор-трансгенные плазмиды имели SEQ ID NO: 11-14 соответственно.

Последовательность SIV была идентичной последовательностям конструкций CFTR (примеры 2 и 3) за исключением промотора и кДНК. Промотор человеческого цитомегаловируса (CMV) или тканеспецифический промотор/энхансер hCEFI применяли, как указано (фиг. 22), для направления экспрессии трансгенов FVIII.

SIV-F/HN-FVIII-N6-со содержал человеческую кДНК FVIII дикого типа, из которой домен BDD был удален посредством делеции и заменен на кодон-оптимизированный 226-аминокислотный фрагмент 6N-гликозилирования.

SIV-FVIII-V3 содержит человеческую кДНК FVIII дикого типа, из которой 226-аминокислотный сайт гликозилирования был удален посредством делеции и заменен на 17-аминокислотный пептид, который экспрессирует триады 6N-гликозилирования внутри В-домена (McIntosh et al., Blood 2013 121(17); 3335-3344).

Пример 21. Количественная оценка антигена hFVIII и уровней активности на моделях *in vivo* и *in vitro*.

Уровни человеческого антигена FVIII у мышинной модели количественно определяли посредством фермент-связанного иммуносорбентного анализа (ELISA) в соответствии с протоколом изготовителя. Вкратце, плазму, BAL и легкое анализировали на предмет присутствия антигена FVIII, используя As-serachrom (FVIII:Ag) Elisa (Stago Diagnostics, France).

Образцы разбавляли 1:2 и инкубировали в 96-луночной планшете, покрытом мышинными моноклональными антителами против фрагмента фактора VIII, в течение 2 ч при комнатной температуре. После промывки в планшет добавляли вторичные против мышинных антитела, связанные с пероксидазой, и инкубацию проводили в течение 2 ч при 37°C. hFVIII:уровни Ag определяли спектрофотометрически при 450 нм, используя субстрат TMB (данные не показаны).

Еще один анализ ELISA применяли для оценки активности FVIII на модели HEK293T *in vitro* (FVIII:C, Affinity Biological, Canada). Супернатанты собирали через 48 и 72 ч после трансдукции HEK293T с использованием SIV-F/HN-FVIII-N6 или SIV-F/HN-FVIII-N3. Активность FVIII оценивали в соответствии со следующими инструкциями изготовителя, используя 50 мкл супернатантов, анализируемые в двух повторностях. В качестве отрицательного контроля тестировали супернатант от необработанных клеток HEK293T. Активность hFVIII рассчитывали по стандартной кривой, генерируемой с использованием ряда разведений нормальной человеческой объединенной плазмы (13-й Британский Стандарт для концентрата фактора VIII коагуляции крови, человеческий; NIBSC).

Клетки HEK293T трансдуцировали с использованием двух различных загрузок VGM142 (загрузка 1- $5,9 \times 10^8$  ЕТТ/мл и загрузка 2- $2,8 \times 10^8$  ЕТТ/мл). Клетки HEK293T трансдуцировали с использованием загрузки 1 вектора VGM142 (фиг. 23А) и загрузки 2 вектора VGM142 (фиг. 23В) при 3 различных МОИ (МОИ; 10; 100) и собирали через 48 и 72 ч после трансдукции.

Как ясно видно на фиг. 23, увеличение активности FVIII наблюдали с увеличением МОИ как для за-



грузки 1, так и загрузки 2 VGM142 как через 48 ч, так и через 72 ч после трансдукции. Кроме того, активность FVТII увеличивалась в интервале от 48 до 72 ч для каждого тестируемого MOI.

Пример 22. Введение лентивирусных векторов FVIII и гена-репортера lux in vivo в нос мыши.

Трансдукция в мышинное легкое.

Все методики с участием животных выполняли согласно условиям и ограничению UK Home Office Project и персональным лицензируемым правилам в соответствии с Animal Scientific Procedure Act (1986).

Самок мышей C57BL/6 дикого типа в возрасте 6-8 недель (Charles River, UK) анестезировали, используя изофлюоран, и им давали 100 мкл вируса в забуференном фосфатом физрастворе по Дульбекко (D-PBS), как описано прежде (Griesenbach et al., 2012), и оценивали присутствие антигена FVIII.

В двух экспериментах (группа 1 и 2) мыши получали 3 дозы (через день) SIV-F/HN-FVIII-N6 (vGM142) и их умерщвляли через 10 дней после первой дозы. Группа 1 (n=4) была обработана общей дозой вектора, равной  $1,4 \times 10^6$  ЕТТ/мышь. Группу 2 (n=3) обрабатывали общей дозой вектора, равной  $1,57 \times 10^8$  ЕТТ/мышь.

В одном эксперименте (группа 3) мышей обрабатывали 12 дозами (через день) SIV-F/HN-FVIII-N6 (vGM142) и умерщвляли через 28 дней после первой дозы. Группу 3 (n=4) обрабатывали общей дозой вектора, равной  $3,36 \times 10^8$  ЕТТ/мышь).

Плазму, жидкость BAL и легкие собирали (как описано в примере 14). Вкратце, мышей умерщвляли в указанные временные отметки после трансдукции. Кровь затем собирали из сердца в пробирки для сбора с антикоагулянтом, 3,2 цитрата натрия, перед центрифугированием при 2000-2500 ×g для получения плазмы. Жидкость BAL собирали, применяя 3 последовательных инстиляции PBS (500 мл) в мышинные легкие при комнатной температуре. Супернатанты хранили при - 80°C. Легкие собирали и хранили при - 80°C перед гомогенизацией ткани.

Затем оценивали наличие экспрессии FVIII. Уровни FVIII оценивали в гомогенатах тканей легких (фиг. 24А). Жидкость BAL (фиг. 24В) и плазму (фиг. 24С) собирали отдельно в 3 независимых экспериментах через 10 и/или 28 дней после обработки SIV-F/HN-FVIII-N6. Анализ выполняли с использованием однофакторного дисперсионного анализа Anova (GraphPad Prism) с множеством сравнений между группами обработки (\*\*\*\* p<0,0001).

Как ясно видно на фиг. 24А, все три группы обработки производили наблюдаемое увеличение уровней hFVIII в ткани легкого в сравнении с соответствующим контролем (D-PBS). 28-Дневная обработка группы 3 приводила к значительному увеличению экспрессии hFVIII в сравнении с 10-дневными обработками групп 1 и 2. Аналогичные результаты наблюдали для образцов BAL жидкости (фиг. 24В), несмотря на то, что в этих образцах также имелось значительное увеличение уровней hFVIII у группы 2 в сравнении с группой 1. Обработка группы 3 приводила к значительному увеличению уровней hFVIII в плазме (фиг. 24С).

Пример 23. Способ производства лентивирусных векторов согласно изобретению.

Клетки НЕК293 выращивают в суспензии, в среде для экспрессии Freestyle (химически определенной, не содержащей компонентов животного происхождения и белков) и проводят подсчет клеток. Определяют концентрацию глюкозы и проводят титрование до ~35 мМ. Готовят смесь для трансфекции пДНК/PEIPro™ и клетки трансфицируют при 0,33 мг пДНК/1Е9 клеток.

Подсчет клеток проводят снова и добавляют дополнительную среду для экспрессии Freestyle. Повторно определяют концентрацию глюкозы и проводят титрование до ~35 мМ. Можно добавить 5 мкл/мл Бензоназы® и проводят трехстадийную линейную очистку вируса. Добавляют Бензоназу® с последующим добавлением TрупLE Select™. Вирус охлаждают до 0°C и хранят на влажном льду для всех последующих стадий. После отфильтровывания какого-либо невирусного вещества в форме частиц (фильтр mPES 0,45 мкм), вирус загружают на Mustang® Q XT (3 мл мембраны/л очищенного вируса) с последующей промывкой 0,15М NaCl Трис рН 7,5 и элюированием 1,0М NaCl Трис рН 7,5. Фракцию вируса собирают и разбавляют до 0,1-0,2 от первоначального объема средой Freestyle. TрупLE Select™ может быть добавлен теперь, если его не добавляли ранее, и Бензоназа® также может быть добавлена на этой стадии в дополнение к или вместо добавленной ранее.

Проводят круговую фильтрацию с тангенциальным потоком (TFF) (УФ до ~0,1-0,05 первоначального объема=HV; ДФ ультраконцентрата ×5 HV против буфера для лекарственной формы; УФ до ~0,001-0,002 от первоначального объема) и ультраконцентрат собирают. Вторая стадия TFF может выполняться, и может применяться более мелкий по размеру блок TFF для ДФ и/или конечной УФ. Дополнительные стадии могут включать смешанный режим/SEC и стерильную фильтрацию через 0,45 мкм или 0,2 мкм.

Фиг. 11 изображает получение и очистку F/HN-векторов SIV. F/HN-векторы SIV получали посредством 5-плазмидной (пДНК) PEI-опосредованной транзientной трансфекции клеток 293Т, выращенных в суспензии в масштабе 1 л в биореакторах WAVE с контролируемым рН (GE), применяя масштабируемые способы изобретения. Векторы очищали посредством глубоинной/окончательной фильтрации (GE/Pall), загрязняющие нуклеиновые кислоты удаляли с использованием Бензоназы® (Merck), векторы активировали с помощью TрупLE Select™ (Life Technology), очищали и концентрировали посредством анионообменной мембранной хроматографии (Pall) и тангенциальной поточной фильтрации (Spectrum).

Все емкости, контейнеры и колонки для процессов соответствовали требованиям cGMP для одноразового применения. Все реагенты, за исключением плазмидной ДНК, соответствовали требованиям cGMP к продуктам, не содержащим компонентов животного происхождения. Показаны данные для различных конфигураций вектора (промотор трансгена, трансген, статус интегразы). Физические и функциональные титры определяли, используя К-ПЦР.

См. результаты данного иллюстративного способа по изобретению, обсуждаемые при описании фиг. 11, приведенном выше.

Ключ к SEQ ID NO.

- SEQ ID NO: 1. Плазмида, определенная на фиг. 1A (пДНК1 PGM326).
- SEQ ID NO: 2. Плазмида, определенная на фиг. 1B (пДНК2a PGM297).
- SEQ ID NO: 3. Плазмида, определенная на фиг. 1C (пДНК2b PGM299).
- SEQ ID NO: 4. Плазмида, определенная на фиг. 1D (пДНК3a PGM301).
- SEQ ID NO: 5. Плазмида, определенная на фиг. 1E (пДНК3b PGM303).
- SEQ ID NO: 6. Иллюстрированный промотор hCEF.
- SEQ ID NO: 7. Иллюстрированный трансген CFTR [soCFTR2).
- SEQ ID NO: 8. Иллюстрированный компонент WPRE (mWPRE).
- SEQ ID NO: 9. Плазмида F/HN-SIV-hCEF-soA1AT, определенная на фиг. 15 (пДНК1 pGM407).
- SEQ ID NO: 10. Плазмида F/HN-SIV-hCEF-sogLux, определенная на фиг. 15 (пДНК1 pGM358).
- SEQ ID NO: 11. Плазмида F/HN-SIV-CMV-HFVIII-V3, определенная на фиг. 22C (пДНК1 pGM411).
- SEQ ID NO: 12. Плазмида F/HN-SIV-hCEF-HFVIII-V3, определенная на фиг. 22D (пДНК1 pGM413).
- SEQ ID NO: 13. F/HN-SIV-CMV-HFVIII-N6-коплазмида, определенная на фиг. 22E (пДНК1 pGM412).
- SEQ ID NO: 14. F/HN-SIV-hCEF-HFVIII-N6-коплазмида, определенная на фиг. 22F (пДНК1 pGM414).
- SEQ ID NO: 15. Иллюстрированный трансген A1AT.
- SEQ ID NO: 16. Иллюстрированный трансген FVIII (N6).
- SEQ ID NO: 17. Иллюстрированный промотор CMV.
- SEQ ID NO: 18. Праймер для построения pCAGGS-Fct4.
- SEQ ID NO: 19. Праймер для построения pCAGGS-Fct4.
- SEQ ID NO: 20. Праймер для построения pCAGGS-Fct4.
- SEQ ID NO: 21. Праймер для построения pCAGGS-Fct4.
- SEQ ID NO: 22. Праймер для построения pCAGGS-Fct4.
- SEQ ID NO: 23. Праймер для построения pCAGGS-Fct4.
- SEQ ID NO: 24. Праймер для построения pCAGGS-SIVct+HN.
- SEQ ID NO: 25. Праймер для построения pCAGGS-SIVct+HN.
- SEQ ID NO: 26. Цепь, комплементарная к иллюстрированному трансгену A1AT.
- SEQ ID NO: 27. Иллюстрированный пептид A1A1.
- SEQ ID NO: 28. Цепь, комплементарная к иллюстрированному трансгену FVIII (N6).
- SEQ ID NO: 29. Иллюстрированный пептид FVIII (N6).
- SEQ ID NO: 30. Иллюстрированный трансген FVIII (V3).
- SEQ ID NO: 31. Цепь, комплементарная к иллюстрированному трансгену FVIII (V3).
- SEQ ID NO: 32. Иллюстрированный пептид FVIII (V3).
- SEQ ID NO: 33. Цепь, комплементарная к иллюстрированному промотору CMV.

## Последовательности

SEQ ID NO: 1

1 GGTACCTCAA TATTGGCCAT TAGCCATATT ATTCATTGGT TATATAGCAT  
 AAATCAATAT  
 61 TGGCTATTGG CCATTGCATA CGTTGTATCT ATATCATAAT ATGTACATTT  
 ATATTGGCTC  
 121 ATGTCCAATA TGACCGCCAT GTTGGCATTG ATTATTGACT AGTTATTAAT  
 AGTAATCAAT  
 181 TACGGGGTCA TTAGTTCATA GCCCATATAT GGAGTTCGCG GTTACATAAC  
 TTACGGTAAA  
 241 TGGCCCCCTT GGCTGACCGC CCAACGACCC CCGCCCATTTG ACGTCAATAA  
 TGACGTATGT  
 301 TCCCATAGTA ACGCCAATAG GGACTTTCCA TTGACGTCAA TGGGTGGAGT  
 ATTTACGGTA  
 361 AACTGCCAC TTGGCAGTAC ATCAAGTGTA TCATATGCCA AGTCCGCCCC  
 CTATTGACGT  
 421 CAATGACGGT AAATGGCCCC CCTGGCATTG TGCCAGTAC ATGACCTTAC  
 GGGACTTTCC  
 481 TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC  
 GGTTTTGGCA  
 541 GTACACCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA TTTCCAAGTC  
 TCCACCCCAT  
 601 TGACGTCAAT GGGAGTTTGT TTTGGCACCA AAATCAACGG GACTTTCCAA  
 AATGTCGTAA  
 661 CAACTGCGAT CGCCCGCCCC GTTGACGCAA ATGGGCGGTA GCGGTGTACG  
 GTGGGAGGTC  
 721 TATATAAGCA GAGCTCGCTG GCTTGTAACT CAGTCTCTTA CTAGGAGACC  
 AGCTTGAGCC  
 781 TGGGTGTTCG CTGGTTAGCC TAACCTGGTT GGCCACCAGG GGTAAGGACT  
 CCTTGGCTTA  
 841 GAAAGCTAAT AACTTGCCT GCATTAGAGC TTATCTGAGT CAAGTGTCTT  
 CATTGACGCC  
 901 TCACTCTCTT GAACGGGAAT CTTCTTACT GGGTTCTCTC TCTGACCCAG  
 GCGAGAGAAA  
 961 CTCCAGCAGT GCGCCCCGAA CAGGGACTTG AGTGAGAGTG TAGGCACGTA  
 CAGCTGAGAA  
 1021 GCGTCGGAC GCGAAGGAAG CGCGGGGTGC GACGCGACCA  
 AGAAGGAGAC TTGGTGAGTA  
 1081 GGCTTCTCGA GTGCCGGGAA AAAGCTCGAG CCTAGTTAGA  
 GGACTAGGAG AGGCCGTAGC  
 1141 CGTAACTACT CTTGGCAAG TAGGGCAGGC GGTGGGTACG  
 CAATGGGGGC GGCTACCTCA  
 1201 GCACTAAATA GGAGACAATT AGACCAATTT GAGAAAATAC  
 GACTTCGCCC GAACGAAAG  
 1261 AAAAAAGTACC AAATTAAACA TTTAATATGG GCAGGCAAGG  
 AGATGGAGCG CTTCCGCCCTC  
 1321 CATGAGAGGT TGTGGAGAC AGAGGAGGGG TGTAAGAGAA  
 TCATAGAAGT CCTTACCCC  
 1381 CTAGAACCAA CAGGATCGGA GGGCTTAAAA AGTCTGTTCA  
 ATCTTGTGTG CGTGCTATAT  
 1441 TGCTTGCACA AGGAACAGAA AGTGAAAGAC ACAGAGGAAG  
 CAGTAGCAAC AGTAAGACAA  
 1501 CACTGCCATC TAGTGGAAAA AGAAAAAAGT GCAACAGAGA  
 CATCTAGTGG ACAAAGAAA  
 1561 AATGACAAGG GAATAGCAGC GCCACCTGGT GGCAGTCAGA  
 ATTTCCAGC GCAACAACAA

## 038966

1621	GGAAATGCCT	GGGTACATGT	ACCCTTGTCA	CCGCGCACCT
TAAATGCGTG	GGTAAAAGCA			
1681	GTAGAGGAGA	AAAAATTTGG	AGCAGAAATA	GTACCCATTT
TTTTGTTTCA	AGCCCTATCG			
1741	AATTCCCGTT	TGTGCTAGGG	TTCTTAGGCT	TCTTGGGGGC
TGCTGGAACT	GCAATGGGAG			
1801	CAGCGGCGAC	AGCCCTGACG	GTCCAGTCTC	AGCATTTGCT
TGCTGGGATA	CTGCAGCAGC			
1861	AGAAGAATCT	GCTGGCGGCT	GTGGAGGCTC	AACAGCAGAT
GTTGAAGCTG	ACCATTGCGG			
1921	GTGTTAAAAA	CCTCAATGCC	CGCGTCACAG	CCCTTGAGAA
GTACCTAGAG	GATCAGGCAC			
1981	GACTAAACTC	CTGGGGTGC	GCATGAAAC	AAGTATGTCA
TACCACAGTG	GAGTGGCCCT			
2041	GGACAAATCG	GACTCCGGAT	TGGCAAAATA	TGACTTGGTT
GGAGTGGGAA	AGACAAATAG			
2101	CTGATTTGGA	AAGCAACATT	ACGAGACAAT	TAGTGAAGGC
TAGAGAACAA	GAGGAAAAGA			
2161	ATCTAGATGC	CTATCAGAAG	TTAACTAGTT	GGTCAGATTT
CTGGTCTTGG	TTCGATTTC			
2221	CAAAATGGCT	TAACATTTTA	AAAATGGGAT	TTTTAGTAAT
AGTAGGAATA	ATAGGGTTAA			
2281	GATTACTTTA	CACAGTATAT	GGATGTATAG	TGAGGGTTAG
GCAGGGATAT	GTTCCCTAT			
2341	CTCCACAGAT	CCATATCCGC	GGCAATTTTA	AAAGAAAGGG
AGGAATAGGG	GGACAGACTT			
2401	CAGCAGAGAG	ACTAATTAAT	ATAATAACAA	CACAATTAGA
AATACAACAT	TTACAAACCA			
2461	AAATTCAAAA	AATTTTAAAT	TTTAGAGCCG	CGGAGATCTG
TTACATAACT	TATGGTAAAT			
2521	GGCCTGCCTG	GCTGACTGCC	CAATGACCCC	TGCCCAATGA
TGTCAATAAT	GATGTATGTT			
2581	CCCATGTAAT	GCCAATAGGG	ACTTTCCATT	GATGTCAATG
GGTGGAGTAT	TTATGGTAAC			
2641	TGCCCACTTG	GCAGTACATC	AAGTGTATCA	TATGCCAAGT
ATGCCCCCTA	TTGATGTCAA			

## 038966

2701	TGATGGTAAA	TGGCCTGCCT	GGCATTATGC	CCAGTACATG
	ACCTTATGGG	ACTTTCCTAC		
2761	TTGGCAGTAC	ATCTATGTAT	TAGTCATTGC	TATTACCATG
	GGAATTCACT	AGTGGAGAAG		
2821	AGCATGCTTG	AGGGCTGAGT	GCCCCCTCAGT	GGGCAGAGAG
	CACATGGCCC	ACAGTCCCTG		
2881	AGAAGTTGGG	GGGAGGGGTG	GGCAATTGAA	CTGGTGCCTA
	GAGAAGGTGG	GGCTTGGGTA		
2941	AACTGGGAAA	GTGATGTGGT	GTACTGGCTC	CACCTTTTTC
	CCCAGGGTGG	GGGAGAACCA		
3001	TATATAAGTG	CAGTAGTCTC	TGTGAACATT	CAAGCTTCTG
	CCTTCTCCCT	CCTGTGAGTT		
3061	TGCTAGCCAC	CATGCAGAGA	AGCCCTCTGG	AGAAGGCCTC
	TGTGGTGAGC	AAGCTGTTCT		
3121	TCAGCTGGAC	CAGGCCCATC	CTGAGGAAGG	GCTACAGGCA
	GAGACTGGAG	CTGTCTGACA		
3181	TCTACCAGAT	CCCCTCTGTG	GA CTCTGCTG	ACAACCTGTC
	TGAGAAGCTG	GAGAGGGAGT		
3241	GGGATAGAGA	GCTGGCCAGC	AAGAAGAACC	CCAAGCTGAT
	CAATGCCCTG	AGGAGATGCT		
3301	TCTTCTGGAG	ATTCATGTTT	TATGGCATCT	TCCTGTACCT
	GGGGGAAGTG	ACCAAGGCTG		
3361	TGCAGCCTCT	GCTGCTGGGC	AGAATCATTG	CCAGCTATGA
	CCCTGACAAC	AAGGAGGAGA		
3421	GGAGCATTGC	CATCTACCTG	GGCATTGGCC	TGTGCCTGCT
	GTTTCATTGTG	AGGACCCTGC		
3481	TGCTGCACCC	TGCCATCTTT	GGCCTGCACC	ACATTGGCAT
	GCAGATGAGG	ATTGCCATGT		
3541	TCAGCCTGAT	CTACAAGAAA	ACCCTGAAGC	TGTCCAGCAG
	AGTGCTGGAC	AAGATCAGCA		
3601	TTGGCCAGCT	GGTGAGCCTG	CTGAGCAACA	ACCTGAACAA
	GTTTGATGAG	GGCCTGGCCC		
3661	TGGCCCACTT	TGTGTGGATT	GCCCCCTGTC	AGGTGGCCCT
	GCTGATGGGC	CTGATTTGGG		
3721	AGCTGCTGCA	GGCCTCTGCC	TTTTGTGGCC	TGGGCTTCCT
	GATTGTGCTG	GCCCTGTTTC		

## 038966

3781	AGGCTGGCCT	GGCAGGATG	ATGATGAAGT	ACAGGGACCA
	GAGGGCAGGC	AAGATCAGTG		
3841	AGAGGCTGGT	GATCACCTCT	GAGATGATTG	AGAACATCCA
	GTCTGTGAAG	GCCTACTGTT		
3901	GGGAGGAAGC	TATGGAGAAG	ATGATTGAAA	ACCTGAGGCA
	GACAGAGCTG	AAGCTGACCA		
3961	GGAAGGCTGC	CTATGTGAGA	TACTTCAACA	GCTCTGCCTT
	CTTCTTCTCT	GGCTTCTTTG		
4021	TGGTGTTCCT	GTCTGTGCTG	CCCTATGCCC	TGATCAAGGG
	GATCATCCTG	AGAAAGATTT		
4081	TCACCACCAT	CAGCTTCTGC	ATTGTGCTGA	GGATGGCTGT
	GACCAGACAG	TTCCCTGGG		
4141	CTGTGCAGAC	CTGGTATGAC	AGCCTGGGGG	CCATCAACAA
	GATCCAGGAC	TTCTGCAGA		
4201	AGCAGGAGTA	CAAGACCCTG	GAGTACAACC	TGACCACCAC
	AGAAGTGGTG	ATGGAGAATG		
4261	TGACAGCCTT	CTGGGAGGAG	GGCTTTGGGG	AGCTGTTTGA
	GAAGGCCAAG	CAGAACAACA		
4321	ACAACAGAAA	GACCAGCAAT	GGGGATGACT	CCCTGTTCTT
	CTCCAACTTC	TCCCTGCTGG		
4381	GCACACCTGT	GCTGAAGGAC	ATCAACTTCA	AGATTGAGAG
	GGGGCAGCTG	CTGGCTGTGG		
4441	CTGGATCTAC	AGGGGCTGGC	AAGACCAGCC	TGCTGATGAT
	GATCATGGGG	GAGCTGGAGC		
4501	CTTCTGAGGG	CAAGATCAAG	CACTCTGGCA	GGATCAGCTT
	TTGCAGCCAG	TTCAGCTGGA		
4561	TCATGCCTGG	CACCATCAAG	GAGAACATCA	TCTTTGGAGT
	GAGCTATGAT	GAGTACAGAT		
4621	ACAGGAGTGT	GATCAAGGCC	TGCCAGCTGG	AGGAGGACAT
	CAGCAAGTTT	GCTGAGAAGG		
4681	ACAACATTGT	GCTGGGGGAG	GGAGGCATTA	CACTGTCTGG
	GGGCCAGAGA	GCCAGAATCA		
4741	GCCTGGCCAG	GGCTGTGTAC	AAGGATGCTG	ACCTGTACCT
	GCTGGACTCC	CCCTTTGGCT		
4801	ACCTGGATGT	GCTGACAGAG	AAGGAGATTT	TTGAGAGCTG
	TGTGTGCAAG	CTGATGGCCA		

## 038966

4861	ACAAGACCAG	AATCCTGGTG	ACCAGCAAGA	TGGAGCACCT
GAAGAAGGCT	GACAAGATCC			
4921	TGATCCTGCA	TGAGGGCAGC	AGCTACTTCT	ATGGGACCTT
CTCTGAGCTG	CAGAACCTGC			
4981	AGCCTGACTT	CAGCTCTAAG	CTGATGGGCT	GTGACAGCTT
TGACCAGTTC	TCTGCTGAGA			
5041	GGAGGAACAG	CATCCTGACA	GAGACCCTGC	ACAGATTGAG
CCTGGAGGGA	GATGCCCTTG			
5101	TGAGCTGGAC	AGAGACCAAG	AAGCAGAGCT	TCAAGCAGAC
AGGGGAGTTT	GGGGAGAAGA			
5161	GGAAGAACTC	CATCCTGAAC	CCCATCAACA	GCATCAGGAA
GTTCAGCATT	GTGCAGAAAA			
5221	CCCCCTGCA	GATGAATGGC	ATTGAGGAAG	ATTCTGATGA
GCCCCTGGAG	AGGAGACTGA			
5281	GCCTGGTGCC	TGATTCTGAG	CAGGGAGAGG	CCATCCTGCC
TAGGATCTCT	GTGATCAGCA			
5341	CAGGCCCTAC	ACTGCAGGCC	AGAAGGAGGC	AGTCTGTGCT
GAACCTGATG	ACCCACTCTG			
5401	TGAACCAGGG	CCAGAACATC	CACAGGAAAA	CCACAGCCTC
CACCAGGAAA	GTGAGCCTGG			
5461	CCCCTCAGGC	CAATCTGACA	GAGCTGGACA	TCTACAGCAG
GAGGCTGTCT	CAGGAGACAG			
5521	GCCTGGAGAT	TTCTGAGGAG	ATCAATGAGG	AGGACCTGAA
AGAGTGCTTC	TTTGATGACA			
5581	TGGAGAGCAT	CCCTGCTGTG	ACCACCTGGA	ACACCTACCT
GAGATACATC	ACAGTGACACA			
5641	AGAGCCTGAT	CTTTGTGCTG	ATCTGGTGCC	TGGTGATCTT
CCTGGCTGAA	GTGGCTGCCT			
5701	CTCTGGTGGT	GCTGTGGCTG	CTGGGAAACA	CCCCACTGCA
GGACAAGGGC	AACAGCACCC			
5761	ACAGCAGGAA	CAACAGCTAT	GCTGTGATCA	TCACCTCCAC
CTCCAGCTAC	TATGTGTTCT			
5821	ACATCTATGT	GGGAGTGGCT	GATACCCTGC	TGGCTATGGG
CTTCTTTAGA	GGCCTGCCCC			
5881	TGGTGACAC	ACTGATCACA	GTGAGCAAGA	TCCTCCACCA
CAAGATGCTG	CACTCTGTGC			

## 038966

5941	TGCAGGCTCC	TATGAGCACC	CTGAATACCC	TGAAGGCTGG
GGGCATCCTG	AACAGATTCT			
6001	CCAAGGATAT	TGCCATCCTG	GATGACCTGC	TGCCCTCAC
CATCTTTGAC	TTCATCCAGC			
6061	TGCTGCTGAT	TGTGATTGGG	GCCATTGCTG	TGGTGGCAGT
GCTGCAGCCC	TACATCTTTG			
6121	TGGCCACAGT	GCCTGTGATT	GTGGCCTTCA	TCATGCTGAG
GGCCTACTTT	CTGCAGACCT			
6181	CCCAGCAGCT	GAAGCAGCTG	GAGTCTGAGG	GCAGAAGCCC
CATCTTCACC	CACCTGGTGA			
6241	CAAGCCTGAA	GGCCCTGTGG	ACCCTGAGAG	CCTTTGGCAG
GCAGCCCTAC	TTGAGACCC			
6301	TGTCCACAA	GGCCCTGAAC	CTGCACACAG	CCAACTGGTT
CCTCTACCTG	TCCACCCTGA			
6361	GATGGTTCCA	GATGAGAATT	GAGATGATCT	TTGTCATCTT
CTTCATTGCT	GTGACCTTCA			
6421	TCAGCATTCT	GACCACAGGA	GAGGGAGAGG	GCAGAGTGGG
CATTATCCTG	ACCCTGGCCA			
6481	TGAACATCAT	GAGCACACTG	CAGTGGGCAG	TGAACAGCAG
CATTGATGTG	GACAGCCTGA			
6541	TGAGGAGTGT	GAGCAGAGTG	TTCAAGTTCA	TTGATATGCC
CACAGAGGGC	AAGCCTACCA			
6601	AGAGCACCAA	GCCCTACAAG	AATGGCCAGC	TGAGCAAAGT
GATGATCATT	GAGAACAGCC			
6661	ATGTGAAGAA	GGATGATATC	TGGCCCAGTG	GAGGCCAGAT
GACAGTGAAG	GACCTGACAG			
6721	CCAAGTACAC	AGAGGGGGGC	AATGCTATCC	TGGAGAACAT
CTCCTTCAGC	ATCTCCCCTG			
6781	GCCAGAGAGT	GGGACTGCTG	GGAAGAACAG	GCTCTGGCAA
GTCTACCCTG	CTGTCTGCCT			
6841	TCCTGAGGCT	GCTGAACACA	GAGGGAGAGA	TCCAGATTGA
TGGAGTGTCC	TGGGACAGCA			
6901	TCACACTGCA	GCAGTGGAGG	AAGGCCTTTG	GTGTGATCCC
CCAGAAAGTG	TTCATCTTCA			
6961	GTGGCACCTT	CAGGAAGAAC	CTGGACCCCT	ATGAGCAGTG
GTCTGACCAG	GAGATTTGGA			



## 038966

7021	AAGTGGCTGA	TGAAGTGGGC	CTGAGAAGTG	TGATTGAGCA
	GTTCCCTGGC	AAGCTGGACT		
7081	TTGTCCTGGT	GGATGGGGGC	TGTGTGCTGA	GCCATGGCCA
	CAAGCAGCTG	ATGTGCCTGG		
7141	CCAGATCAGT	GCTGAGCAAG	GCCAAGATCC	TGCTGTGGA
	TGAGCCTTCT	GCCACCTGG		
7201	ATCCTGTGAC	CTACCAGATC	ATCAGGAGGA	CCCTCAAGCA
	GGCCTTTGCT	GACTGCACAG		
7261	TCATCCTGTG	TGAGCACAGG	ATTGAGGCCA	TGCTGGAGTG
	CCAGCAGTTC	CTGGTGATTG		
7321	AGGAGAACAA	AGTGAGGCAG	TATGACAGCA	TCCAGAAGCT
	GCTGAATGAG	AGGAGCCTGT		
7381	TCAGGCAGGC	CATCAGCCCC	TCTGATAGAG	TGAAGCTGTT
	CCCCCACAGG	AACAGCTCCA		
7441	AGTGCAAGAG	CAAGCCCCAG	ATTGCTGCCC	TGAAGGAGGA
	GACAGAGGAG	GAAGTCAGG		
7501	ACACCAGGCT	GTGAGGGCCC	AATCAACCTC	TGGATTACAA
	AATTTGTGAA	AGATTGACTG		
7561	GTATTCTTAA	CTATGTTGCT	CCTTTTACGC	TATGTGGATA
	CGCTGCTTTA	ATGCCTTTGT		
7621	ATCATGTCTAT	TGCTTCCCGT	ATGGCTTTCA	TTTTCTCCTC
	CTTGTATAAA	TCCTGGTTGC		
7681	TGTCTCTTTA	TGAGGAGTTG	TGGCCCGTTG	TCAGGCAACG
	TGGCGTGGTG	TGCACTGTGT		
7741	TTGCTGACGC	AACCCCCACT	GGTTGGGGCA	TTGCCACCAC
	CTGTCAGCTC	CTTTCCGGGA		
7801	CTTTCGCTTT	CCCCCTCCCT	ATTGCCACGG	CGGAACACTAT
	CGCCGCCTGC	CTTGCCCGCT		
7861	GCTGGACAGG	GGCTCGGCTG	TTGGGCACTG	ACAATTCGGT
	GGTGTGTCG	GGGAAATCAT		
7921	CGTCCTTTCC	TTGGCTGCTC	GCCTGTGTTG	CCACCTGGAT
	TCTGCGCGGG	ACGTCCTTCT		
7981	GCTACGTCCC	TTCGGCCCTC	AATCCAGCGG	ACCTTCCTTC
	CCGCGGCCTG	CTGCCGGCTC		
8041	TGCGGCCTCT	TCCGCGTCTT	CGCCTTCGCC	CTCAGACGAG
	TCGGATCTCC	CTTTGGGCCG		

## 038966

8101	CCTCCCGCA	AGCTTCGCAC	TTTTTAAAAG	AAAAGGGAGG
ACTGGATGGG	ATTTATTACT			
8161	CCGATAGGAC	GCTGGCTTGT	AACTCAGTCT	CTTACTAGGA
GACCAGCTTG	AGCCTGGGTG			
8221	TTCGCTGGTT	AGCCTAACCT	GGTTGGCCAC	CAGGGGTAAG
GACTCCTTGG	CTTAGAAAGC			
8281	TAATAAACTT	GCCTGCATTA	GAGCTCTTAC	GCCTCCCGGG
CTCGAGATCC	GCATCTCAAT			
8341	TAGTCAGCAA	CCATAGTCCC	GCCCTAACT	CCGCCCATCC
CGCCCCTAAC	TCCGCCAGT			
8401	TCCGCCCATF	CTCCGCCCA	TGGCTGACTA	ATTTTTTTTA
TTTATGCAGA	GGCCGAGGCC			
8461	GCCTCGGCCT	CTGAGCTATT	CCAGAAGTAG	TGAGGAGGCT
TTTTTGGAGG	CCTAGGCTTT			
8521	TGCAAAAAGC	TAAC TTGTTT	ATTGCAGCTT	ATAATGGTTA
CAAATAAAAGC	AATAGCATCA			
8581	CAAATTTTAC	AAATAAAGCA	TTTTTTTAC	TGCATTCTAG
TTGTGGTTTG	TCCAAACTCA			
8641	TCAATGTATC	TTATCATGTC	TGTCGGCTTC	CTCGCTCACT
GACTCGCTGC	GCTCGGTCGT			
8701	TCGGCTGCGG	CGAGCGGTAT	CAGCTCACTC	AAAGGCGGTA
ATACGGTTAT	CCACAGAATC			
8761	AGGGGATAAC	GCAGGAAAGA	ACATGTGAGC	AAAAGGCCAG
CAAAGGCCA	GGAACCGTAA			
8821	AAAGGCCGCG	TTGCTGGCGT	TTTTCCATAG	GCTCCGCCCC
CCTGACGAGC	ATCACAAAAA			
8881	TCGACGCTCA	AGTCAGAGGT	GGCGAAACCC	GACAGGACTA
TAAAGATACC	AGGCGTTTCC			
8941	CCCTGGAAGC	TCCCTCGTGC	GCTCTCCTGT	TCCGACCCTG
CCGCTTACCG	GATACCTGTC			
9001	CGCCTTTCTC	CCTTCGGGAA	GCGTGGCGCT	TTCTCATAGC
TCACGCTGTA	GGTATCTCAG			
9061	TTCGGTGTAG	GTCGTTCGCT	CCAAGCTGGG	CTGTGTGCAC
GAACCCCCCG	TTCAGCCCGA			
9121	CCGCTGCGCC	TTATCCGGTA	ACTATCGTCT	TGAGTCCAAC
CCGGTAAGAC	ACGACTTATC			

## 038966

9181	GCCACTGGCA	GCAGCCACTG	GTAACAGGAT	TAGCAGAGCG
AGGTATGTAG	GCGGTGCTAC			
9241	AGAGTTCTTG	AAGTGGTGGC	CTAACTACGG	CTACACTAGA
AGAACAGTAT	TTGGTATCTG			
9301	CGCTCTGCTG	AAGCCAGTTA	CCTTCGGAAA	AAGAGTTGGT
AGCTCTTGAT	CGGCAAACA			
9361	AACCACCGCT	GGTAGCGGTG	GTTTTTTTGT	TTGCAAGCAG
CAGATTACGC	GCAGAAAAA			
9421	AGGATCTCAA	GAAGATCCTT	TGATCTTTTC	TACGGGGTCT
GACGCTCAGT	GGAACGAAAA			
9481	CTCACGTTAA	GGGATTTTGG	TCATGAGATT	ATCAAAAAGG
ATCTTCACCT	AGATCCTTTT			
9541	AAATTAATAA	TGAAGTTTTA	AATCAATCTA	AAGTATATAT
GAGTAAACTT	GGTCTGACAG			
9601	TTAGAAAAAC	TCATCGAGCA	TCAAATGAAA	CTGCAATTTA
TTCATATCAG	GATTATCAAT			
9661	ACCATATTTT	TGAAAAAGCC	GTTTCTGTAA	TGAAGGAGAA
AACTCACCGA	GGCAGTTCCA			
9721	TAGGATGGCA	AGATCCTGGT	ATCGGTCTGC	GATTCCGACT
CGTCCAACAT	CAATACAACC			
9781	TATTAATTTT	CCCTCGTCAA	AAATAAGGTT	ATCAAGTGAG
AAATCACCAT	GAGTGACGAC			
9841	TGAATCCGGT	GAGAATGGCA	ACAGCTTATG	CATTTCTTTC
CAGACTTGTT	CAACAGGCCA			
9901	GCCATTACGC	TCGTCAATCA	AATCACTCGC	ATCAACCAAA
CCGTTATTCA	TTCGTGATTG			
9961	CGCCTGAGCG	AGACGAAATA	CGCGATCGCT	GTTAAAAGGA
CAATTACAAA	CAGGAATCGA			
10021	ATGCAACCGG	CGCAGGAACA	CTGCCAGCGC	ATCAACAATA
TTTTACCTG	AATCAGGATA			
10081	TTCTTCTAAT	ACCTGGAATG	CTGTTTTTCC	GGGGATCGCA
GTGGTGAGTA	ACCATGCATC			
10141	ATCAGGAGTA	CGGATAAAAT	GCTTGATGGT	CGGAAGAGGC
ATAAATTCG	TCAGCCAGTT			
10201	TAGTCTGACC	ATCTCATCTG	TAACATCATT	GGCAACGCTA
CCTTTGCCAT	GTTTCAGAAA			

038966

10261 CAACTCTGGC GCATCGGGCT TCCCATACAA TCGATAGATT  
GTCGCACCTG ATTGCCCGAC  
10321 ATTATCGCGA GCCCATTAT ACCCATATAA ATCAGCATCC  
ATGTTGGAAT TTAATCGCGG  
10381 CCTAGAGCAA GACGTTTCCC GTTGAATATG GCTCATAACA  
CCCCTGTAT TACTGTTTAT  
10441 GTAAGCAGAC AGTTTTATTG TTCATGATGA TATATTTTTA  
TCTTGTGCAA TGTAACATCA  
10501 GAGATTTTGA GACACAACAA TTGGTCGACG GATCC  
**SEQ ID NO:2**  
1 GCTCGAGACT AGTGACTIONG TGAGTAGGCT TCGAGCCTAG TTAGAGGACT  
AGGAGAGGCC  
61 GTAGCCGTAA CTACTIONTGG CAAGTAGGGC AGGCGGTGGG TACGCAATGG  
GGCGGCTAC  
121 CTCAGACTA AATAGGAGAC AATTAGACCA ATTTGAGAAA ATACGACTTC  
GCCCCAACGG  
181 AAAGAAAAAG TACCAAATTA AACATTTAAT ATGGGCAGGC AAGGAGATGG  
AGCGCTTCGG  
241 CCTCCATGAG AGGTTGTTGG AGACAGAGGA GGGGTGTAAA AGAATCATAG  
AAGTCCTCTA  
301 CCCCTAGAA CCAACAGGAT CGGAGGGCTT AAAAAGTCTG TTCAATCTTG  
TGTGCGTACT  
361 ATATTGCTTG CACAAGGAAC AGAAAGTGAA AGACACAGAG GAAGCAGTAG  
CAACAGTAAG  
421 ACAACACTGC CATCTAGTGG AAAAAGAAAA AAGTGCAACA GAGACATCTA  
GTGGACAAAA  
481 GAAAAATGAC AAGGGAATAG CAGCGCCACC TGGTGGCAGT CAGAATTTTC  
CAGCGCAACA  
541 ACAAGGAAAT GCCTGGGTAC ATGTACCCTT GTCACCGCGC ACCTTAAATG  
CGTGGGTAAA  
601 AGCAGTAGAG GAGAAAAAAT TTGGAGCAGA AATAGTACCC ATGTTTCAAG  
CCCTATCAGA  
661 AGGCTGCACA CCCTATGACA TTAATCAGAT GCTTAATGTG CTAGGAGATC  
ATCAAGGGGC  
721 ATTACAAATA GTGAAAGAGA TCATTAATGA AGAAGCAGCC CAGTGGGATG  
TAACACACCC

## 038966

781 ACTACCCGCA GGACCCCTAC CAGCAGGACA GCTCAGGGAC CCTCGCGGCT  
 CAGATATAGC  
 841 AGGGACCACC AGCTCAGTAC AAGAACAGTT AGAATGGATC TATACTGCTA  
 ACCCCCGGGT  
 901 AGATGTAGGT GCCATCTACC GGAGATGGAT TATTCTAGGA CTTCAAAGT  
 GTGTCAAAT  
 961 GTACAACCCA GTATCAGTCC TAGACATTAG GCAGGGACCT AAAGAGCCCT  
 TCAAGGATTA  
 1021 TGTGGACAGA TTTTACAAGG CAATTAGAGC AGAACAAGCC  
 TCAGGGGAAG TGAAACAATG  
 1081 GATGACAGAA TCATTACTCA TTCAAAATGC TAATCCAGAT  
 TGTAAGGTCA TCCTGAAGGG  
 1141 CCTAGGAATG CACCCACCC TTGAAGAAAT GTTAACGGCT  
 TGTCAGGGGG TAGGAGGCC  
 1201 AAGCTACAAA GCAAAAGTAA TGGCAGAAAT GATGCAGACC  
 ATGCAAAATC AAAACATGGT  
 1261 GCAGCAGGGA GTTCCAAAAA GACAAAGACC CCCACTAAGA  
 TGTATAAATT GTGAAAATT  
 1321 TGGCCATATG CAAAGACAAT GTCCGGAACC AAGGAAAACA  
 AAATGTCTAA AGTGTGAAA  
 1381 ATTGGGACAC CTAGCAAAAG ACTGCAGGGG ACAGGTGAAT  
 TTTTtagggt ATGGACGGTG  
 1441 GATGGGGGCA AAACCGAGAA ATTTTCCCGC CGCTACTCTT  
 GGAGCGGAAC CGAGTGCGCC  
 1501 TCCTCCACCG AGCGGCACCA CCCCATACGA CCCAGCAAAG  
 AAGCTCCTGC AGCAATATGC  
 1561 AGAGAAAGGG AAACAACCTGA GGGAGCAAAA GAGGAATCCA  
 CCGGCAATGA ATCCGGATTG  
 1621 GACCGAGGGA TATTCTTTGA ACTCCCTCTT TGGAGAAGAC  
 CAATAAAGAC AGTGTATATA  
 1681 GAAGGGGTCC CCATTAAGGC ACTGCTAGAC ACAGGGGCAG  
 ATGACACCAT AATTAAAGAA  
 1741 AATGATTTAC AATTATCAGG TCCATGGAGA CCCAAAATTA  
 TAGGGGGCAT AGGAGGAGGC  
 1801 CTTAATGTAA AAGAATATAA CGACAGGGAA GTAAAAATAG  
 AAGATAAAAT TTTGAGAGGA

## 038966

1861	ACAATATTGT	TAGGAGCAAC	TCCCATTAAT	ATAATAGGTA
	GAAATTTGCT	GGCCCCGGCA		
1921	GGTGCCCGGT	TAGTAATGGG	ACAATTATCA	GAAAAAATTC
	CTGTCACACC	TGTCAAATTG		
1981	AAGGAAGGGG	CTCGGGGACC	CTGTGTAAGA	CAATGGCCTC
	TCTCTAAAGA	GAAGATTGAA		
2041	GCTTTACAGG	AAATATGTTC	CCAATTAGAG	CAGGAAGGAA
	AAATCAGTAG	AGTAGGAGGA		
2101	GAAAAATGCAT	ACAATACCCC	AATATTTTGC	ATAAAGAAGA
	AGGACAAATC	CCAGTGGAGG		
2161	ATGCTAGTAG	ACTTTAGAGA	GTTAAATAAG	GCAACCCAAG
	ATTTCTTTGA	AGTGCAATTA		
2221	GGGATACCCC	ACCCAGCAGG	ATTAAGAAAAG	ATGAGACAGA
	TAACAGTTTT	AGATGTAGGA		
2281	GACGCCATT	ATTCCATACC	ATTGGATCCA	AATTTTAGGA
	AATATACTGC	TTTACTATT		
2341	CCCACAGTGA	ATAATCAGGG	ACCCGGGATT	AGGTATCAAT
	TCAACTGTCT	CCCACAAGGG		
2401	TGGAAAGGAT	CTCCTACAAT	CTTCCAAAAT	ACAGCAGCAT
	CCATTTTGA	GGAGATAAAA		
2461	AGAAACTTGC	CAGCACTAAC	CATTGTACAA	TACATGGATG
	ATTTATGGGT	AGGTTCTCAA		
2521	GAAAAATGAAC	ACACCCATGA	CAAATTAGTA	GAACAGTTAA
	GAACAAAATT	ACAAGCCTGG		
2581	GGCTTAGAAA	CCCCAGAAAA	GAAGGTGCAA	AAAGAACCAC
	CTTATGAGTG	GATGGGATAC		
2641	AAACTTTGGC	CTCACAAATG	GGAACTAAGC	AGAATACAAC
	TGGAGGAAAA	AGATGAATGG		
2701	ACTGTCAATG	ACATCCAGAA	GTTAGTTGGG	AAACTAAATT
	GGGCAGCACA	ATTGTATCCA		
2761	GGTCTTAGGA	CCAAGAATAT	ATGCAAGTTA	ATTAGAGGAA
	AGAAAAATCT	GTTAGAGCTA		
2821	GTGACTTGGG	CACCTGAGGC	AGAAGCTGAA	TATGCAGAAA
	ATGCAGAGAT	TCTTAAACA		
2881	GAACAGGAAG	GAACCTATTA	CAAACCAGGA	ATACCTATTA
	GGGCAGCAGT	ACAGAAATTG		

## 038966

2941	GAAGGAGGAC	AGTGGAGTTA	CCAATTCAAA	CAAGAAGGAC
AAGTCTTGAA	AGTAGGAAAA			
3001	TACACCAAGC	AAAAGAACAC	CCATACAAAT	GAACTTCGCA
CATTAGCTGG	TTTAGTGCA			
3061	AAGATTGCA	AAGAAGCTCT	AGTTATTTGG	GGGATATTAC
CAGTTCTAGA	ACTCCCATA			
3121	GAAAGAGAGG	TATGGGAACA	ATGGTGGGCG	GATTACTGGC
AGGTAAGCTG	GATTCCCGAA			
3181	TGGGATTTTG	TCAGCACCCC	ACCTTTGCTC	AAACTATGGT
ACACATTAAC	AAAAGAACCC			
3241	ATACCCAAGG	AGGACGTTTA	CTATGTAGAT	GGAGCATGCA
ACAGAAATTC	AAAAGAAGGA			
3301	AAAGCAGGAT	ACATCTCACA	ATACGGAAAA	CAGAGAGTAG
AAACATTAGA	AAACACTACC			
3361	AATCAGCAAG	CAGAATTAAC	AGCTATAAAA	ATGGCTTTGG
AAGACAGTGG	GCCTAATGTG			
3421	AACATAGTAA	CAGACTCTCA	ATATGCAATG	GGAAATTTGA
CAGCACAACC	CACACAAAGT			
3481	GATTCACCAT	TAGTAGAGCA	AATTATAGCC	TTAATGATAC
AAAAGCAACA	AATATATTTG			
3541	CAGTGGGTAC	CAGCACATAA	AGGAATAGGA	GGAAATGAGG
AGATAGATAA	ATTAGTGAGT			
3601	AAAGGCATTA	GAAGAGTTTT	ATTCTTAGAA	AAAATAGAAG
AAGCTCAAGA	AGAGCATGAA			
3661	AGATATCATA	ATAATTGGAA	AAACCTAGCA	GATACATATG
GGCTTCCACA	AATAGTAGCA			
3721	AAAGAGATAG	TGGCCATGTG	TCCAAAATGT	CAGATAAAGG
GAGAACCAGT	GCATGGACAA			
3781	GTGGATGCCT	CACCTGGAAC	ATGGCAGATG	GATTGTACTC
ATCTAGAAGG	AAAAGTAGTC			
3841	ATAGTTGCGG	TCCATGTAGC	CAGTGGATTC	ATAGAAGCAG
AAGTCATACC	TAGGGAAACA			
3901	GGAAAAGAAA	CGGCAAAGTT	TCTATTAAAA	ATACTGAGTA
GATGGCCTAT	AACACAGTTA			
3961	CACACAGACA	ATGGGCCTAA	CTTTACCTCC	CAAGAAGTGG
CAGCAATATG	TTGGTGGGGA			

## 038966

4021	AAAATTGAAC	ATACAACAGG	TATACCATAT	AACCCCAAT
CTCAAGGATC	AATAGAAAGC			
4081	ATGAACAAAC	AATTAAGA	GATAATTGGG	AAAATAAGAG
ATGATTGCCA	ATATACAGAG			
4141	ACAGCAGTAC	TGATGGCTTG	CCATATTCAC	AATTTTAAAA
GAAAGGGAGG	AATAGGGGGA			
4201	CAGACTTCAG	CAGAGAGACT	AATTAATATA	ATAACAACAC
AATTAGAAAT	ACAACATTTA			
4261	CAAACAAAA	TTCAAAAAT	TTTAAATTTT	AGAGTCTACT
ACAGAGAAGG	GAGAGACCCT			
4321	GTGTGAAAG	GACCAGCACA	ATTAATCTGG	AAAGGGGAG
GAGCAGTGGT	CCTCAAGGAC			
4381	GGAAGTGACC	TAAAGTTGT	ACCAAGAAGG	AAAGCTAAAA
TTATTAAGGA	TTATGAACCC			
4441	AAACAAAGAG	TGGTAATGA	GGGTGACGTG	GAAGGTACCA
GGGGATCTGA	TAACATAATG			
4501	GCAGGAATA	GTCAGATATT	GGATGAGACA	AAGAAATTTG
AAATGGAAT	ATTATATGCA			
4561	TCAGCTGGCG	GCCGCAATT	CACTAGTGAT	TCCCGTTTGT
GCTAGGGTTC	TTAGGCTTCT			
4621	TGGGGCTGC	TGGAATGCA	ATGGGAGCAG	CGGCGACAGC
CCTGACGGTC	CAGTCTCAGC			
4681	ATTTGCTTGC	TGGGATACTG	CAGCAGCAGA	AGAATCTGCT
GGCGGCTGTG	GAGGCTCAAC			
4741	AGCAGATGTT	GAAGCTGACC	ATTTGGGGTG	TTAAAACTT
CAATGCCCGC	GTCACAGCCC			
4801	TTGAGAAGTA	CCTAGAGGAT	CAGGCACGAC	TAAACTCCTG
GGGTGCGCA	TGAAACAAG			
4861	TATGTCATAC	CACAGTGGAG	TGGCCCTGGA	CAAATCGGAC
TCCGGATTGG	CAAAATATGA			
4921	CTTGGTTGGA	GTGGGAAAGA	CAAATAGCTG	ATTTGGAAAG
CAACATTACG	AGACAATTAG			
4981	TGAAGGCTAG	AGAACAAGAG	GAAAAGAATC	TAGATGCCTA
TCAGAAGTTA	ACTAGTTGGT			
5041	CAGATTCTG	GTCTTGGTTC	GATTTCTCAA	AATGGCTTAA
CATTTTAAAA	ATGGGATTTT			



## 038966

5101	TAGTAATAGT	AGGAATAATA	GGGTTAAGAT	TACTTTACAC
	AGTATATGGA TGTATAGTGA			
5161	GGGTTAGGCA	GGGATATGTT	CCTCTATCTC	CACAGATCCA
	TATCCAATCG AATTCCCGCG			
5221	GCCGCAATTC	ACTCCTCAGG	TGCAGGCTGC	CTATCAGAAG
	GTGGTGGCTG GTGTGGCCAA			
5281	TGCCCTGGCT	CACAAATACC	ACTGAGATCT	TTTTCCCTCT
	GCCAAAAATT ATGGGGACAT			
5341	CATGAAGCCC	CTTGAGCATC	TGACTTCTGG	CTAATAAAGG
	AAATTTATTT TCATTGCAAT			
5401	AGTGTGTTGG	AATTTTTTGT	GTCTCTCACT	CGGAAGGACA
	TATGGGAGGG CAAATCATT			
5461	AAAACATCAG	AATGAGTATT	TGGTTTAGAG	TTTGCCAACA
	TATGCCCATA TGCTGGCTGC			
5521	CATGAACAAA	GGTTGGCTAT	AAAGAGGTCA	TCAGTATATG
	AAACAGCCCC CTGCTGTCCA			
5581	TTCCTTATTC	CATAGAAAAG	CCTTGACTTG	AGGTTAGATT
	TTTTTTATAT TTTGTTTTGT			
5641	GTTATTTTTT	TCTTTAACAT	CCCTAAAATT	TTCCTTACAT
	GTTTTACTAG CCAGATTTTT			
5701	CCTCCTCTCC	TGACTACTCC	CAGTCATAGC	TGTCCCTCTT
	CTCTTATGGA GATCCCTCGA			
5761	CCTGCAGCCC	AAGCTTGCGG	TAATCATGGT	CATAGCTGTT
	TCCTGTGTGA AATTGTTATC			
5821	CGCTCACAAT	TCCACACAAC	ATACGAGCCG	GAAGCATAAA
	GTGTAAGGCC TGGGGTGCC			
5881	AATGAGTGAG	CTAACTCACA	TTAATTGCGT	TGCGCTCACT
	GCCCCGTTTC CAGTCGGGAA			
5941	ACCTGTCGTG	CCAGCGGATC	CGCATCTCAA	TTAGTCAGCA
	ACCATAGTCC CGCCCCAAC			
6001	TCCGCCCATC	CCGCCCTAA	CTCCGCCCAG	TTCCGCCCAT
	TCTCCGCCCC ATGGCTGACT			
6061	AATTTTTTTT	ATTTATGCAG	AGGCCGAGGC	CGCCTCGGCC
	TCTGAGCTAT TCCAGAAGTA			
6121	GTGAGGAGGC	TTTTTTGGAG	GCCTAGGCTT	TTGCAAAAAG
	CTAACTTGTT TATTGCAGCT			

## 038966

6181	TATAATGGTT	ACAAATAAAG	CAATAGCATC	ACAAATTTCA
CAAATAAAGC	ATTTTTTTCA			
6241	CTGCATTCTA	GTTGTGGTTT	GTCCAAACTC	ATCAATGTAT
CTTATCATGT	CTGTCCGCTT			
6301	CCTCGCTCAC	TGACTCGCTG	CGCTCGGTCG	TTCGGCTGCG
GCGAGCGGTA	TCAGTCACT			
6361	CAAAGGCGGT	AATACGGTTA	TCCACAGAAT	CAGGGGATAA
CGCAGGAAAG	AACATGTGAG			
6421	CAAAAGGCCA	GCAAAAGGCC	AGGAACCGTA	AAAAGGCCGC
GTTGCTGGCG	TTTTTCCATA			
6481	GGCTCCGCCC	CCCTGACGAG	CATCACAAAA	ATCGACGCTC
AAGTCAGAGG	TGGCGAAACC			
6541	CGACAGGACT	ATAAAGATAC	CAGGCGTTTC	CCCCTGGAAG
CTCCCTCGTG	CGCTCTCCTG			
6601	TTCCGACCCT	GCCGCTTACC	GGATACCTGT	CCGCCTTTCT
CCCTTCGGGA	AGCGTGGCGC			
6661	TTTCTCATAG	CTCACGCTGT	AGGTATCTCA	GTTGCGTGTA
GGTCGTTCGC	TCCAAGCTGG			
6721	GCTGTGTGCA	CGAACCCCCC	GTTAGCCCG	ACCGCTGCGC
CTTATCCGGT	AACTATCGTC			
6781	TTGAGTCCAA	CCCGGTAAGA	CACGACTTAT	CGCCACTGGC
AGCAGCCACT	GGTAACAGGA			
6841	TTAGCAGAGC	GAGGTATGTA	GGCGGTGCTA	CAGAGTTCTT
GAAGTGGTGG	CCTAACTACG			
6901	GCTACACTAG	AAGAACAGTA	TTTGGTATCT	GCGCTCTGCT
GAAGCCAGTT	ACCTTCGGAA			
6961	AAAGAGTTGG	TAGCTCTTGA	TCCGGCAAAC	AAACCACCGC
TGGTAGCGGT	GGTTTTTTTG			
7021	TTTGCAAGCA	GCAGATTACG	CGCAGAAAAA	AAGGATCTCA
AGAAGATCCT	TTGATCTTTT			
7081	CTACGGGGTC	TGACGCTCAG	TGGAACGAAA	ACTCACGTTA
AGGGATTTTG	GTCATGAGAT			
7141	TATCAAAAAG	GATCTTCACC	TAGATCCTTT	TAAATTAATA
ATGAAGTTTT	AAATCAATCT			
7201	AAAGTATATA	TGAGTAAACT	TGGTCTGACA	GTTAGAAAAA
CTCATCGAGC	ATCAAATGAA			

## 038966

7261	ACTGCAATTT	ATTCATATCA	GGATTATCAA	TACCATATTT
	TTGAAAAGC CGTTTCTGTA			
7321	ATGAAGGAGA	AAACTCACCG	AGGCAGTTCC	ATAGGATGGC
	AAGATCCTGG TATCGGTCTG			
7381	CGATTCCGAC	TCGTCCAACA	TCAATACAAC	CTATTAATTT
	CCCCTCGTCA AAAATAAGGT			
7441	TATCAAGTGA	GAAATCACCA	TGAGTGACGA	CTGAATCCGG
	TGAGAATGGC AACAGCTTAT			
7501	GCATTTCTTT	CCAGACTTGT	TCAACAGGCC	AGCCATTACG
	CTCGTCATCA AAATCACTCG			
7561	CATCAACCAA	ACCGTTATTC	ATTCGTGATT	GCGCCTGAGC
	GAGACGAAAT ACGCGATCGC			
7621	TGTTAAAAGG	ACAATTACAA	ACAGGAATCG	AATGCAACCG
	GCGCAGGAAC ACTGCCAGCG			
7681	CATCAACAAT	ATTTTCACCT	GAATCAGGAT	ATTCTTCTAA
	TACCTGGAAT GCTGTTTTTC			
7741	CGGGGATCGC	AGTGGTGAGT	AACCATGCAT	CATCAGGAGT
	ACGGATAAAA TGCTTGATGG			
7801	TCGGAAGAGG	CATAAATTCC	GTCAGCCAGT	TTAGTCTGAC
	CATCTCATCT GTAACATCAT			
7861	TGGCAACGCT	ACCTTTGCCA	TGTTTCAGAA	ACAACCTCTGG
	CGCATCGGGC TTCCCATACA			
7921	ATCGATAGAT	TGTCGCACCT	GATTGCCCGA	CATTATCGCG
	AGCCCATTTA TACCCATATA			
7981	AATCAGCATC	CATGTTGGAA	TTAATCGCG	GCCTAGAGCA
	AGACGTTTCC CGTTGAATAT			
8041	GGCTCATAAC	ACCCCTTGTA	TTACTGTTTA	TGTAAGCAGA
	CAGTTTTATT GTTCATGATG			
8101	ATATATTTTT	ATCTTGTGCA	ATGTAACATC	AGAGATTTTG
	AGACACAACA ATTGTCGACA			
8161	TTGATTATTG	ACTAGTTATT	AATAGTAATC	AATTACGGGG
	TCATTAGTTC ATAGCCATA			
8221	TATGGAGTTC	CGCGTTACAT	AACTTACGGT	AAATGGCCCG
	CCTGGCTGAC CGCCCAACGA			
8281	CCCCCGCCA	TTGACGTCAA	TAATGACGTA	TGTTCCCATATA
	GTAACGCCAA TAGGGACTTT			

## 038966

8341	CCATTGACGT	CAATGGGTGG	AGTATTTACG	GTAAACTGCC
	CACTTGGCAG	TACATCAAGT		
8401	GTATCATATG	CCAAGTACGC	CCCCTATTGA	CGTCAATGAC
	GGTAAATGGC	CCGCCTGGCA		
8461	TTATGCCCAG	TACATGACCT	TATGGGACTT	TCCTACTTGG
	CAGTACATCT	ACGTATTAGT		
8521	CATCGCTATT	ACCATGGTCG	AGGTGAGCCC	CACGTTCTGC
	TTCACTCTCC	CCATCTCCCC		
8581	CCCCTCCCCA	CCCCAATTT	TGTATTTATT	TATTTTTTAA
	TTATTTTGTG	CAGCGATGGG		
8641	GGCGGGGGGG	GGGGGGGGGC	GCGCGCCAGG	CGGGGCGGGG
	CGGGGCGAGG	GGCGGGGCGG		
8701	GGCGAGGCGG	AGAGGTGCGG	CGGCAGCCAA	TCAGAGCGGC
	GCGCTCCGAA	AGTTTCCTTT		
8761	TATGGCGAGG	CGGCGGCGGC	GGCGGCCCTA	TAAAAAGCGA
	AGCGCGCGGC	GGGCGGGAGT		
8821	CGCTGCGCGC	TGCCTTCGCC	CCGTGCCCCG	CTCCGCCCGC
	GCCTCGCGCC	GCCCCGCCCG		
8881	GCTCTGACTG	ACCGCGTTAC	TCCCACAGGT	GAGCGGGCGG
	GACGGCCCTT	CTCCTCCGGG		
8941	CTGTAATTAG	CGCTTGTTT	AATGACGGCT	TGTTTCTTTT
	CTGTGGCTGC	GTGAAAGCCT		
9001	TGAGGGGCTC	CGGGAGGGCC	CTTTGTGCGG	GGGAGCGGG
	TCGGGGGGTG	CGTGCGTGTG		
9061	TGTGTGCGTG	GGGAGCGCCG	CGTGCGGCTC	CGCGCTGCCC
	GGCGGCTGTG	AGCGCTGCGG		
9121	GCGCGGCGCG	GGGCTTTGTG	CGCTCCGCAG	TGTGCGCGAG
	GGGAGCGCGG	CCGGGGGCGG		
9181	TGCCCCGCGG	TGCGGGGGGG	GCTGCGAGGG	GAACAAAGGC
	TGCGTGCGGG	GTGTGTGCGT		
9241	GGGGGGGTGA	GCAGGGGGTG	TGGGCGCGTC	GGTCGGGCTG
	CAACCCCCCC	TGCACCCCCC		
9301	TCCCCGAGTT	GCTGAGCACG	GCCCCGGCTT	GGGTGCGGGG
	CTCCGTACGG	GGCGTGGCGC		
9361	GGGGCTCGCC	GTGCCGGGCG	GGGGTGGCG	GCAGGTGGGG
	GTGCCGGGCG	GGGCGGGGCG		

038966

9421 GCCTCGGGCC GGGGAGGGCT CGGGGAGGG GCGCGGCGGC  
 CCCCGGAGCG CCGGCGGCTG  
 9481 TCGAGGCGCG GCGAGCCGCA GCCATTGCCT TTTATGGTAA  
 TCGTGCGAGA GGGCGCAGGG  
 9541 ACTTCCTTTG TCCCAAATCT GTGCGGAGCC GAAATCTGGG  
 AGGCGCCGCC GCACCCCTC  
 9601 TAGCGGGCGC GGGGCGAAGC GGTGCGGCGC CGGCAGGAAG  
 GAAATGGGCG GGGAGGGCCT  
 9661 TCGTGCGTCG CCGCGCCGCC GTCCCCTTCT CCCTCTCCAG  
 CCTCGGGGCT GTCCGCGGGG  
 9721 GGACGGCTGC CTTCGGGGGG GACGGGGCAG GCGGGGGTTC  
 GGCTTCTGGC GTGTGACCGG  
 9781 CGGCTCTAGA GCCTCTGCTA ACCATGTTCA TGCCTTCTTC  
 TTTTCTTAC AGCTCCTGGG  
 9841 CAACGTGCTG GTTATTGTGC TGTCTCATCA TTTGGCAA GAATT  
**SEQ ID NO:3**  
 1 TCAATATTGG CCATTAGCCA TATTATTCAT TGGTTATATA GCATAAATCA  
 ATATTGGCTA  
 61 TTGGCCATG CATACTTGT ATCTATATCA TAATATGTAC ATTTATATTG  
 GCTCATGTCC  
 121 AATATGACCG CCATGTTGGC ATTGATTATT GACTAGTTAT TAATAGTAAT  
 CAATTACGGG  
 181 GTCATTAGTT CATAGCCCAT ATATGGAGTT CCGCGTTACA TAACTTACGG  
 TAAATGGCCC  
 241 GCCTGGCTGA CCGCCCAACG ACCCCCGCCC ATTGACGTCA ATAATGACGT  
 ATGTTCCCAT  
 301 AGTAACGCCA ATAGGGACTT TCCATTGACG TCAATGGGTG GAGTATTTAC  
 GGTAACACTGC  
 361 CCACTTGCCA GTACATCAAG TGTATCATAT GCCAAGTCCG CCCCTATTG  
 ACGTCAATGA  
 421 CGGTAAATGG CCCGCTGGC ATTATGCCCA GTACATGACC TTACGGGACT  
 TTCCTACTTG  
 481 GCAGTACATC TACGTATTAG TCATCGCTAT TACCATGGTG ATGCGGTTTT  
 GGCAGTACAC  
 541 CAATGGGCGT GGATAGCGGT TTGACTCAG GGGATTTCCA AGTCTCCACC  
 CCATTGACGT

601 CAATGGGAGT TTGTTTGGC ACCAAAATCA ACGGGACTTT CAAAATGTC  
 GTAATAACCC  
 661 CGCCCCGTTG ACGCAAATGG GCGGTAGGCG TGTACGGTGG GAGGTCTATA  
 TAAGCAGAGC  
 721 TCGTTTAGTG AACCGTCAGA TCACTAGAAG CTTTATTGCG GTAGTTTATC  
 ACAGTTAAAT  
 781 TGCTAACGCA GTCAGTGCTT CTGACACAAC AGTCTCGAAC TTAAGCTGCA  
 GAAGTTGGTC  
 841 GTGAGGCACT GGGCAGGTAA GTATCAAGGT TACAAGACAG GTTTAAGGAG  
 ACCAATAGAA  
 901 ACTGGGCTTG TCGAGACAGA GAAGACTCTT GCGTTTCTGA TAGGCACCTA  
 TTGGTCTTAC  
 961 TGACATCCAC TTTGCCTTTC TCTCCACAGG TGTCCACTCC CAGTTCAATT  
 ACAGCTCTTA  
 1021 AGGCTAGAGT ACTTAATACG ACTCACTATA GGCTAGCCTC  
 GAGAATTCGA TTATGCCCTT  
 1081 AGGACCAGAA GAAAGAAGAT TGCTTCGCTT GATTGGCTC  
 CTTTACAGCA CCAATCCATA  
 1141 TCCACCAAGT GGGGAAGGGA CGGCCAGACA ACGCCGACGA  
 GCCAGGAGAA GGTGGAGACA  
 1201 ACAGCAGGAT CAAATTAGAG TCTTGGTAGA AAGACTCCAA  
 GAGCAGGTGT ATGCAGTTGA  
 1261 CCGCCTGGCT GACGAGGCTC AACACTTGCC TATACAACAG  
 TTGCCTGACC CTCTCATTC  
 1321 AGCTTAGAAT CACTAGTGAA TTCACGCGTG GTACCTCTAG  
 AGTCGACCCG GGCGGCCGCT  
 1381 TCGAGCAGAC ATGATAAGAT ACATTGATGA GTTTGGACAA  
 ACCACAATA GAATGCAGTG  
 1441 AAAAAAATGC TTTATTGTG AAATTTGTGA TGCTATTGCT  
 TTATTTGTAA CCATTATAAG  
 1501 CTGCAATAAA CAAGTTAACA ACAACAATTG CATTCAATTT  
 ATGTTTCAGG TTCAGGGGGA  
 1561 GATGTGGGAG GTTTTTAAA GCAAGTAAAA CCTCTACAAA  
 TGTGGTAAAA TCGATAAGGA  
 1621 TCCGTCGACC AATTGTTGTG TCTCAAATC TCTGATGTTA  
 CATTGCACAA GATAAAAATA

## 038966

1681	TATCATCATG	AACAATAAAA	CTGTCTGCTT	ACATAAACAG
	TAATACAAGG	GGTGTATGA		
1741	GCCATATTCA	ACGGGAAACG	TCTTGCTCTA	GGCCGCGATT
	AAATCCAAC	ATGGATGCTG		
1801	ATTTATATGG	GTATAAATGG	GCTCGCGATA	ATGTCGGGCA
	ATCAGGTGCG	ACAATCTATC		
1861	GATTGTATGG	GAAGCCCGAT	GCGCCAGAGT	TGTTTCTGAA
	ACATGGCAAA	GGTAGCGTTG		
1921	CCAATGATGT	TACAGATGAG	ATGGTCAGAC	TAAACTGGCT
	GACGGAATTT	ATGCCTCTTC		
1981	CGACCATCAA	GCATTTTATC	CGTACTCCTG	ATGATGCATG
	GTTACTCACC	ACTGCGATCC		
2041	CCGGAAAAAC	AGCATTCCAG	GTATTAGAAG	AATATCCTGA
	TTCAGGTGAA	AATATTGTTG		
2101	ATGCGCTGGC	AGTGTTCCTG	CGCCGGTTGC	ATTGCGATTCC
	TGTTTGTAAT	TGTCCTTTTA		
2161	ACAGCGATCG	CGTATTTCTG	CTCGCTCAGG	CGCAATCACG
	AATGAATAAC	GGTTTGGTTG		
2221	ATGCGAGTGA	TTTTGATGAC	GAGCGTAATG	GCTGGCCTGT
	TGAACAAGTC	TGAAAGAAA		
2281	TGCATAAGCT	GTTGCCATTC	TCACCGGATT	CAGTCGTCAC
	TCATGGTGAT	TTCTCACTTG		
2341	ATAACCTTAT	TTTTGACGAG	GGGAAATTAA	TAGGTTGTAT
	TGATGTTGGA	CGAGTCGGAA		
2401	TCGCAGACCG	ATACCAGGAT	CTTGCCATCC	TATGGAAGCTG
	CCTCGGTGAG	TTTTCTCCTT		
2461	CATTACAGAA	ACGGCTTTTT	CAAAAATATG	GTATTGATAA
	TCCTGATATG	AATAAATTGC		
2521	AGTTTCATTT	GATGCTCGAT	GAGTTTTTCT	AACTGTCAGA
	CCAAGTTTAC	TCATATATAC		
2581	TTTAGATTGA	TTTTAAACTT	CATTTTTAAT	TTAAAAGGAT
	CTAGGTGAAG	ATCCTTTTTG		
2641	ATAATCTCAT	GACCAAATC	CCTTAACGTG	AGTTTTCGTT
	CCACTGAGCG	TCAGACCCCG		
2701	TAGAAAAGAT	CAAAGGATCT	TCTTGAGATC	CTTTTTTCT
	GCGCGTAATC	TGCTGCTTGC		

038966

2761 AAACAAAAA ACCACCGCTA CCAGCGGTGG TTTGTTTGCC  
GGATCAAGAG CTACCAACTC  
2821 TTTTCCGAA GGTAAGTGGC TTCAGCAGAG CGCAGATACC  
AAATACTGTT CTTCTAGTGT  
2881 AGCCGTAGTT AGGCCACCAC TTCAAGAACT CTGTAGCACC  
GCCTACATAC CTCGCTCTGC  
2941 TAATCCTGTT ACCAGTGGCT GCTGCCAGTG GCGATAAGTC  
GTGTCTTACC GGGTTGGACT  
3001 CAAGACGATA GTTACCGGAT AAGGCGCAGC GGTCCGGCTG  
AACGGGGGGT TCGTGACAC  
3061 AGCCAGCTT GGAGCGAACG ACCTACACCG AACTGAGATA  
CCTACAGCGT GAGCTATGAG  
3121 AAAGCGCCAC GCTTCCGAA GGGAGAAAGG CGGACAGGTA  
TCCGGTAAGC GGCAGGGTCG  
3181 GAACAGGAGA GCGCACGAGG GAGCTTCCAG GGGGAAACGC  
CTGGTATCTT TATAGTCTTG  
3241 TCGGGTTTCG CCACCTCTGA CTTGAGCGTC GATTTTTGTG  
ATGCTCGTCA GGGGGGCGGA  
3301 GCCTATGGAA AAACGCCAGC AACGCGGCCT TTTTACGGTT  
CCTGGCCTTT TGCTGGCCTT  
3361 TTGCTCACAT GGCTCGACAG ATCT

**SEQ ID NO:4**

1 ATTGATTATT GACTAGTTAT TAATAGTAAT CAATTACGGG GTCATTAGTT  
CATAGCCCAT  
61 ATATGGAGTT CCGCGTTACA TAACTTACGG TAAATGGCCC GCCTGGCTGA  
CCGCCCAACG  
121 ACCCCCGCCC ATTGACGTCA ATAATGACGT ATGTTCCCAT AGTAACGCCA  
ATAGGGACTT  
181 TCCATTGACG TCAATGGGTG GAGTATTTAC GGTAAGTGC CCACTTGGCA  
GTACATCAAG  
241 TGTATCATAT GCCAAGTACG CCCCCTATTG ACGTCAATGA CGGTAAATGG  
CCCGCCTGGC  
301 ATTATGCCCA GTACATGACC TTATGGGACT TTCCTACTTG GCAGTACATC  
TACGTATTAG  
361 TCATCGCTAT TACCATGGTC GAGGTGAGCC CCACGTTCTG CTTACTCTC  
CCCATCTCCC



038966

421 CCCCTCCCC ACCCCAATT TTGTATTAT TTATTTTTTA ATTATTTTGT  
 GCAGCGATGG  
 481 GGGCGGGGGG GGGGGGGGGG CGCGGCCAG GCGGGGCGGG GCGGGGCAAG  
 GGGCGGGGGG  
 541 GGGCGAGGCG GAAAGGTGCG GCGGCAGCCA ATCAAAGCGG CGCGCTCCGA  
 AAGTTTCCTT  
 601 TTATGGCGAG GCGGCGGGG CGGCGGCCCT ATAAAAAGCG AAGCGCGCGG  
 CGGGCGGGAG  
 661 TCGCTGCGCG CTGCCTTCGC CCCGTGCCCC GCTCCGCCGC CGCCTCGCGC  
 CGCCCGCCCC  
 721 GGCTCTGACT GACCGCGTTA CTCCCACAGG TGAGCGGGCG GGACGGCCCT  
 TCTCCTCCGG  
 781 GCTGTAATTA GCGCTTGGTT TAATGACGGC TTGTTTCTTT TCTGTGGCTG  
 CGTGAAGCC  
 841 TTGAGGGGCT CCGGAGGGC CTTTGTGCG GGGGAGCGG CTCGGGGGGT  
 GCGTGCGTGT  
 901 GTGTGTGCGT GGGGAGCGCC GCGTGCGGCT CCGCGCTGCC CGGCGGCTGT  
 GAGCGCTGCG  
 961 GGGCGGGCG GGGGCTTTGT GCGCTCCGCA GTGTGCGCGA GGGGAGCGCG  
 GCCGGGGGGC  
 1021 GTGCCCGCG GTGCGGGGGG GGCTGCGAGG GGAACAAAGG  
 CTGCGTGCGG GGTGTGTGCG  
 1081 TGGGGGGGTG AGCAGGGGGT GTGGGCGCGT CGGTGCGGCT  
 GCAACCCCC CTGCACCCCC  
 1141 CTCCCCGAGT TGCTGAGCAC GGCCCGGCTT CGGGTGCGGG  
 GCTCCGTACG GGGCGTGGCG  
 1201 CGGGGCTCGC CGTGCCGGGC GGGGGTGGC GGCAGGTGGG  
 GGTGCCGGGC GGGGCGGGC  
 1261 CGCCTCGGGC CGGGGAGGGC TCGGGGAGG GCGCGGGCGG  
 CCCCCGAGC GCCGGCGGCT  
 1321 GTCGAGGCGC GCGGAGCCGC AGCCATTGCC TTTTATGGTA  
 ATCGTGCGAG AGGGCGCAGG  
 1381 GACTTCCTTT GTCCCAAATC TGTGCGGAGC CGAAATCTGG  
 GAGGCGCCGC CGCACCCCT  
 1441 CTAGCGGGCG CGGGGCGAAG CCGTGCGGCG CCGGCAGGAA  
 GGAAATGGGC GGGGAGGGCC

## 038966

1501	TTCGTGCGTC	GCCGCGCCGC	CGTCCCCTTC	TCCCTCTCCA
	GCCTCGGGGC	TGTCCGCGGG		
1561	GGGACGGCTG	CCTTCGGGGG	GGACGGGGCA	GGGCGGGGTT
	CGGCTTCTGG	CGTGTGACCG		
1621	GCGGCTCTAG	AGCCTCTGCT	AACCATGTTC	ATGCCTTCTT
	CTTTTCCTA	CAGCTCCTGG		
1681	GCAACGTGCT	GGTTATTGTG	CTGTCTCATC	ATTTTGCCAA
	AGAATTCGAT	TGCCATGGCA		
1741	ACATATATCC	AGAGAGTACA	GTGCATCTCA	ACATCACTAC
	TGGTTGTTCT	CACCACATTG		
1801	GTCTCGTGTC	AGATTCCCAG	GGATAGGCTC	TCTAACATAG
	GGTCATAGT	CGATGAAGGG		
1861	AAATCACTGA	AGATAGCTGG	ATCCCACGAA	TCGAGGTACA
	TAGTACTGAG	TCTAGTTCCG		
1921	GGGGTAGACT	TTGAGAATGG	GTGCGGAACA	GCCCAGGTTA
	TCCAGTACAA	GAGCCTACTG		
1981	AACAGGCTGT	TAATCCCATT	GAGGGATGCC	TTAGATCTTC
	AGGAGGCTCT	GATAACTGTC		
2041	ACCAATGATA	CGACACAAAA	TGCCGGTGCT	CCCCAGTCGA
	GATTCTTCGG	TGCTGTGATT		
2101	GGTACTATCG	CACTTGGAGT	GGCGACATCA	GCACAAATCA
	CCGCAGGGAT	TGCACTAGCC		
2161	GAAGCGAGGG	AGGCCAAAAG	AGACATAGCG	CTCATCAAAG
	AATCGATGAC	AAAAACACAC		
2221	AAGTCTATAG	AACTGCTGCA	AAACGCTGTG	GGGGAACAAA
	TTCTTGCTCT	AAAGACACTC		
2281	CAGGATTTCCG	TGAATGATGA	GATCAAACCC	GCAATAAGCG
	AATTAGGCTG	TGAGACTGCT		
2341	GCCTTAAGAC	TGGGTATAAA	ATTGACACAG	CATTACTCCG
	AGCTGTTAAC	TGCGTTCGGC		
2401	TCGAATTTCCG	GAACCATCGG	AGAGAAGAGC	CTCACGCTGC
	AGGCGCTGTC	TTCACTTTAC		
2461	TCTGCTAACA	TTACTGAGAT	TATGACCACA	ATCAGGACAG
	GGCAGTCTAA	CATCTATGAT		
2521	GTCATTTATA	CAGAACAGAT	CAAAGGAACG	GTGATAGATG
	TGGATCTAGA	GAGATACATG		

## 038966

2581	GTCACCCTGT	CTGTGAAGAT	CCCTATTCTT	TCTGAAGTCC
	CAGGTGTGCT	CATACACAAG		
2641	GCATCATCTA	TTTCTTACAA	CATAGACGGG	GAGGAATGGT
	ATGTGACTGT	CCCCAGCCAT		
2701	ATACTCAGTC	GTGCTTCTTT	CTTAGGGGGT	GCAGACATAA
	CCGATTGTGT	TGAGTCCAGA		
2761	TTGACCTATA	TATGCCCCAG	GGATCCCGCA	CAACTGATAC
	CTGACAGCCA	GCAAAAGTGT		
2821	ATCCTGGGGG	ACACAACAAG	GTGCCTGTCT	ACAAAAGTTG
	TGGACAGCCT	TATCCCCAAG		
2881	TTTGTCTTTG	TGAATGGGGG	CGTTGTTGCT	AACTGCATAG
	CATCCACATG	TACCTGCGGG		
2941	ACAGGCCGAA	GACCAATCAG	TCAGGATCGC	TCTAAAGGTG
	TAGTATTCTT	AACCCATGAC		
3001	AACTGTGGTC	TTATAGGTGT	CAATGGGGTA	GAATTGTATG
	CTAACCGGAG	AGGGCAGCAT		
3061	GCCACTTGGG	GGGTCCAGAA	CTTGACAGTC	GGTCTGCAA
	TTGCTATCAG	ACCCGTTGAT		
3121	ATTTCTCTCA	ACCTTGCTGA	TGCTACGAAT	TTCTTGCAAG
	ACTCTAAGGC	TGAGCTTGAG		
3181	AAAGCACGGA	AAATCCTCTC	GGAGGTAGGT	AGATGGTACA
	ACTCAAGAGA	GACTGTGATT		
3241	ACGATCATAG	TAGTTATGGT	CGTAATATTG	GTGGTCATTA
	TAGTGATCAT	CATCGTGCTT		
3301	TATAGACTCA	GAAGGTGAAA	TCACTAGTGA	ATTCACTCCT
	CAGGTGCAGG	CTGCCTATCA		
3361	GAAGGTGGTG	GCTGGTGTGG	CCAATGCCCT	GGCTCACAAA
	TACCACTGAG	ATCTTTTCC		
3421	CTCTGCCAAA	AATTATGGGG	ACATCATGAA	GCCCCTTGAG
	CATCTGACTT	CTGGCTAATA		
3481	AAGGAAATTT	ATTTTCATTG	CAATAGTGTG	TTGGAATTTT
	TTGTGTCTCT	CACTCGGAAG		
3541	GACATATGGG	AGGGCAAATC	ATTTAAAACA	TCAGAATGAG
	TATTTGGTTT	AGAGTTTGGC		
3601	AACATATGCC	CATATGCTGG	CTGCCATGAA	CAAAGGTTGG
	CTATAAAGAG	GTCATCAGTA		

## 038966

3661	TATGAAACAG	CCCCCTGCTG	TCCATTCCTT	ATTCCATAGA
	AAAGCCTTGA	CTTGAGGTTA		
3721	GATTTTTTTT	ATATTTTGTT	TTGTGTTATT	TTTTTCTTTA
	ACATCCCTAA	AATTTTCCTT		
3781	ACATGTTTTA	CTAGCCAGAT	TTTTCTCCT	CTCCTGACTA
	CTCCCAGTCA	TAGCTGTCCC		
3841	TCTTCTCTTA	TGGAGATCCC	TCGACCTGCA	GCCCAAGCTT
	GCGGTAATCA	TGGTCATAGC		
3901	TGTTTCTGT	GTGAAATTGT	TATCCGCTCA	CAATTCACA
	CAACATACGA	GCCGGAAGCA		
3961	TAAAGTGTA	AGCCTGGGGT	GCCTAATGAG	TGAGCTAACT
	CACATTAATT	GCGTTGCGCT		
4021	CACTGCCCCG	TTTCCAGTCG	GGAAACCTGT	CGTGCCAGCG
	GATCCGCATC	TCAATTAGTC		
4081	AGCAACCATA	GTCCCGCCCC	TAACTCCGCC	CATCCCGCCC
	CTAACTCCGC	CCAGTTCCGC		
4141	CCATTCTCCG	CCCCATGGCT	GACTAATTTT	TTTTATTTAT
	GCAGAGGCCG	AGGCCGCCTC		
4201	GGCCTCTGAG	CTATTCCAGA	AGTAGTGAGG	AGGCTTTTTT
	GGAGGCCTAG	GCTTTTGCAA		
4261	AAAGCTAACT	TGTTTATTGC	AGCTTATAAT	GGTTACAAAT
	AAAGCAATAG	CATCACAAAT		
4321	TTCACAAATA	AAGCATTTTT	TTCACTGCAT	TCTAGTTGTG
	GTTTGTCCAA	ACTCATCAAT		
4381	GTATCTTATC	ATGTCTGTCC	GCTTCCTCGC	TCACTGACTC
	GCTGCGCTCG	GTCGTTCCGC		
4441	TGCGGCGAGC	GGTATCAGCT	CACTCAAAGG	CGGTAATACG
	GTTATCCACA	GAATCAGGGG		
4501	ATAACGCAGG	AAAGAACATG	TGAGCAAAG	GCCAGCAAAA
	GGCCAGGAAC	CGTAAAAAGG		
4561	CCGCGTTGCT	GGCGTTTTTC	CATAGGCTCC	GCCCCCTGA
	CGAGCATCAC	AAAAATCGAC		
4621	GCTCAAGTCA	GAGGTGGCGA	AACCCGACAG	GACTATAAAG
	ATACCAGGCG	TTTCCCCCTG		
4681	GAAGCTCCCT	CGTGCGCTCT	CCTGTTCCGA	CCCTGCCGCT
	TACCGGATAC	CTGTCCGCCT		

## 038966

4741	TTCTCCCTTC	GGGAAGCGTG	GCGCTTTCTC	ATAGCTCAGC
CTGTAGGTAT	CTCAGTTCGG			
4801	TGTAGGTCGT	TCGCTCCAAG	CTGGGCTGTG	TGCACGAACC
CCCGGTTTCAG	CCCGACCGCT			
4861	GCGCCTTATC	CGGTAACTAT	CGTCTTGAGT	CCAACCCGGT
AAGACACGAC	TTATCGCCAC			
4921	TGGCAGCAGC	CACTGGTAAC	AGGATTAGCA	GAGCGAGGTA
TGTAGGCGGT	GCTACAGAGT			
4981	TCTTGAAGTG	GTGGCCTAAC	TACGGCTACA	CTAGAAGAAC
AGTATTTGGT	ATCTGCGCTC			
5041	TGCTGAAGCC	AGTTACCTTC	GGAAAAAGAG	TTGGTAGCTC
TTGATCCGGC	AAACAAACCA			
5101	CCGCTGGTAG	CGGTGGTTTT	TTTGTTTGCA	AGCAGCAGAT
TACGCGCAGA	AAAAAAGGAT			
5161	CTCAAGAAGA	TCCTTTGATC	TTTTCTACGG	GGTCTGACGC
TCAGTGAAC	GAAAACTCAC			
5221	GTTAAGGGAT	TTTGGTCATG	AGATTATCAA	AAAGGATCTT
CACCTAGATC	CTTTTAAATT			
5281	AAAAATGAAG	TTTTAAATCA	ATCTAAAGTA	TATATGAGTA
AACTTGGTCT	GACAGTTAGA			
5341	AAAACTCATC	GAGCATCAAA	TGAAACTGCA	ATTTATTCAT
ATCAGGATTA	TCAATACCAT			
5401	ATTTTTGAAA	AAGCCGTTTC	TGTAATGAAG	GAGAAAACTC
ACCGAGGCAG	TTCCATAGGA			
5461	TGGCAAGATC	CTGGTATCGG	TCTGCGATTC	CGACTCGTCC
AACATCAATA	CAACCTATTA			
5521	ATTTCCCTC	GTCAAAAATA	AGGTTATCAA	GTGAGAAATC
ACCATGAGTG	ACGACTGAAT			
5581	CCGGTGAGAA	TGGCAACAGC	TTATGCATTT	CTTTCCAGAC
TTGTTCAACA	GGCCAGCCAT			
5641	TACGCTCGTC	ATCAAAATCA	CTCGCATCAA	CCAAACCGTT
ATTCATTTCGT	GATTGCGCCT			
5701	GAGCGAGACG	AAATACGCGA	TCGCTGTAA	AAGGACAATT
ACAAACAGGA	ATCGAATGCA			
5761	ACCGGCGCAG	GAACACTGCC	AGCGCATCAA	CAATATTTTC
ACCTGAATCA	GGATATTCTT			

038966

5821 CTAATACCTG GAATGCTGTT TTCCCGGGA TCGCAGTGGT  
GAGTAACCAT GCATCATCAG  
5881 GAGTACGGAT AAAATGCTTG ATGGTCGGAA GAGGCATAAA  
TTCCGTCAGC CAGTTTAGTC  
5941 TGACCATCTC ATCTGTAACA TCATTGGCAA CGCTACCTTT  
GCCATGTTTC AGAAACAAC  
6001 CTGGCGCATC GGGCTTCCCA TACAATCGAT AGATTGTCGC  
ACCTGATTGC CCGACATTAT  
6061 CGCGAGCCCA TTTATACCCA TATAAATCAG CATCCATGTT  
GGAATTTAAT CGCGCCTAG  
6121 AGCAAGACGT TTCCCGTTGA ATATGGCTCA TAACACCCCT  
TGTATTACTG TTTATGTAAG  
6181 CAGACAGTTT TATTGTTCAT GATGATATAT TTTTATCTTG  
TGCAATGTAA CATCAGAGAT

6241 TTTGAGACAC AACAATTGGT CGAC

SEQ ID NO:5

1 ATTGATTATT GACTAGTTAT TAATAGTAAT CAATTACGGG GTCATTAGTT  
CATAGCCCAT  
61 ATATGGAGTT CCGCGTTACA TAACTTACGG TAAATGGCCC GCCTGGCTGA  
CCGCCCAACG  
121 ACCCCCGCCC ATTGACGTCA ATAATGACGT ATGTTCCCAT AGTAACGCCA  
ATAGGGACTT  
181 TCCATTGACG TCAATGGGTG GAGTATTTAC GGTAACACTGC CCACTTGGCA  
GTACATCAAG  
241 TGTATCATAT GCCAAGTACG CCCCCTATTG ACGTCAATGA CGGTAAATGG  
CCCGCCTGGC  
301 ATTATGCCCA GTACATGACC TTATGGGACT TTCCTACTTG GCAGTACATC  
TACGTATTAG  
361 TCATCGCTAT TACCATGGTC GAGGTGAGCC CCACGTTCTG CTTCACCTCTC  
CCCATCTCCC  
421 CCCCCTCCCC ACCCCCAATT TTGTATTAT TTATTTTTTA ATTATTTTGT  
GCAGCGATGG  
481 GGGCGGGGGG GGGGGGGGG CGCGGCCAG GCGGGGCGG GCGGGGCGAG  
GGGCGGGGCG  
541 GGGCGAGGCG GAAAGGTGCG GCGGCAGCCA ATCAAAGCGG CGCGCTCCGA  
AAGTTTCCTT

601 TTATGGCGAG GCGGCGGCGG CGGCGGCCCT ATAAAAAGCG AAGCGGCGGG  
 CGGGCGGGAG  
 661 TCGCTGCGCG CTGCCTTCGC CCCGTGCCCC GCTCCGCCGC CGCCTCGCGC  
 CGCCCGCCCC  
 721 GGCTCTGACT GACCGCGTTA CTCCACAGG TGAGCGGGCG GGACGGCCCT  
 TCTCCTCCGG  
 781 GCTGTAATTA GCGCTTGGTT TAATGACGGC TTGTTTCTTT TCTGTGGCTG  
 CGTGAAAGCC  
 841 TTGAGGGGCT CCGGGAGGGC CCTTTGTGCG GGGGGAGCGG CTCGGGGGGT  
 GCGTGCGTGT  
 901 GTGTGTGCGT GGGGAGCGCC GCGTGCGGCT CCGCGCTGCC CGGCGGCTGT  
 GAGCGCTGCG  
 961 GGC GCGCGCGC GGGGCTTTGT GCGCTCCGCA GTGTGCGCGA GGGGAGCGCG  
 GCCGGGGGCG  
 1021 GTGCCCCGCG GTGCGGGGGG GGCTGCGAGG GGAACAAAGG  
 CTGCGTGCGG GGTGTGTGCG  
 1081 TGGGGGGGTG AGCAGGGGGT GTGGGCGCGT CGGTGCGGGT  
 GCAACCCCC CTGCACCCCC  
 1141 CTCCCCGAGT TGCTGAGCAC GGCCCGCCTT CGGGTGCGGG  
 GCTCCGTACG GGGCGTGGCG  
 1201 CGGGGCTCGC CGTGCCGGGC GGGGGGTGGC GGCAGGTGGG  
 GGTGCCGGGC GGGGCGGGGC  
 1261 CGCCTCGGGC CGGGGAGGGC TCGGGGAGG GGC GCGGCGG  
 CCCCCGAGC GCCGGCGGCT  
 1321 GTCGAGGCGC GCGGAGCCGC AGCCATTGCC TTTTATGGTA  
 ATCGTGCGAG AGGGCGCAGG  
 1381 GACTTCCTTT GTCCCAAATC TGTGCGGAGC CGAAATCTGG  
 GAGGCGCCGC CGCACCCCTT  
 1441 CTAGCGGGCG CGGGGCGAAG CCGTGCGGCG CCGGCAGGAA  
 GGAAATGGGC GGGGAGGGCC  
 1501 TTCGTGCGTC GCCGCGCCGC CGTCCCCTTC TCCCTCTCCA  
 GCCTCGGGGC TGTCCGCGGG  
 1561 GGGACGGGGC AGGGCGGGGT TCGGCTTCTG GCGTGTGACC  
 GCGGCTCTA GAGCCTCTGC  
 1621 TAACCATGTT CATGCCTTCT TCTTTTCTCT ACAGCTCCTG  
 GGCAACGTGC TGGTTATTGT

## 038966

1681	GCTGTCTCAT	CATTTTGGCA	AAGAATTCCT	CGAGCATGTG
	GTCTGAGTTA AAAATCAGGA			
1741	GCAACGACGG	AGGTGAAGGA	CCAGAGGACG	CCAACGACCC
	CCGGGAAAAG GGGGTGCAAC			
1801	ACATCCATAT	CCAGCCATCT	CTACCTGTTT	ATGGACAGAG
	GGTTAGGGAT GGTGATAGGG			
1861	GCAAACGTGA	CTCGTACTGG	TCTACTTCTC	CTAGTGGTAG
	CACCACAAAA CCAGCATCAG			
1921	GTTGGGAGAG	GTCAGTAAA	GCCGACACAT	GGTTGCTGAT
	TCTCTCATTC ACCCAGTGGG			
1981	CTTTGTCAAT	TGCCACAGTG	ATCATCTGTA	TCATAATTC
	TGCTAGACAA GGGTATAGTA			
2041	TGAAAGAGTA	CTCAATGACT	GTAGAGGCAT	TGAACATGAG
	CAGCAGGGAG GTGAAAGAGT			
2101	CACTTACCAG	TCTAATAAGG	CAAGAGGTTA	TAGCAAGGGC
	TGTCAACATT CAGAGCTCTG			
2161	TGCAAACCGG	AATCCCAGTC	TTGTTGAACA	AAAACAGCAG
	GGATGTCATC CAGATGATTG			
2221	ATAAGTCGTG	CAGCAGACAA	GAGCTCACTC	AGCACTGTGA
	GAGTACGATC GCAGTCCACC			
2281	ATGCCGATGG	AATTGCCCCA	CTTGAGCCAC	ATAGTTTCTG
	GAGATGCCCT GTCGGAGAAC			
2341	CGTATCTTAG	CTCAGATCCT	GAAATTCAT	TGCTGCCTGG
	TCCGAGCTTG TTATCTGGTT			
2401	CTACAACGAT	CTCTGGATGT	GTTAGGCTCC	CTTCACTCTC
	AATTGGCGAG GCAATCTATG			
2461	CCTATTCATC	AAATCTCATT	ACACAAGGTT	GTGCTGACAT
	AGGGAAATCA TATCAGGTCC			
2521	TGCAGCTAGG	GTACATATCA	CTCAATTCAG	ATATGTTCCC
	TGATCTTAAC CCCGTAGTGT			
2581	CCCACACTTA	TGACATCAAC	GACAATCGGA	AATCATGCTC
	TGTGGTGGCA ACCGGGACTA			
2641	GGGGTTATCA	GCTTTGCTCC	ATGCCGACTG	TAGACGAAAG
	AACCGACTAC TCTAGTGATG			
2701	GTATTGAGGA	TCTGGTCCTT	GATGTCCTGG	ATCTCAAAGG
	GAGAACTAAG TCTACCCGGT			



## 038966

2761	ATCGCAACAG	CGAGGTAGAT	CTTGATCACC	CGTTCTCTGC
ACTATACCCC	AGTGTAGGCA			
2821	ACGGCATTGC	AACAGAAGGC	TCATTGATAT	TTCTTGGGTA
TGGTGGACTA	ACCACCCTC			
2881	TGCAGGGTGA	TACAAAATGT	AGGACCCAAG	GATGCCAACA
GGTGTGCGAA	GACACATGCA			
2941	ATGAGGCTCT	GAAAATTACA	TGGCTAGGAG	GGAAACAGGT
GGTCAGCGTG	ATCATCCAGG			
3001	TCAATGACTA	TCTCTCAGAG	AGGCCAAAGA	TAAGAGTCAC
AACCATTCCA	ATCACTCAA			
3061	ACTATCTCGG	GGCGGAAGGT	AGATTATTAA	AATTGGGTGA
TCGGGTGTAC	ATCTATACAA			
3121	GATCATCAGG	CTGGCACTCT	CAACTGCAGA	TAGGAGTACT
TGATGTCAGC	CACCCTTGA			
3181	CTATCAACTG	GACACCTCAT	GAAGCCTTGT	CTAGACCAGG
AAATAAAGAG	TGCAATTGGT			
3241	ACAATAAGTG	TCCGAAGGAA	TGCATATCAG	GCGTATACAC
TGATGCTTAT	CCATTGTCCC			
3301	CTGATGCAGC	TAACGTCGCT	ACCGTCACGC	TATATGCCAA
TACATCGCGT	GTCAACCCAA			
3361	CAATCATGTA	TTCTAACACT	ACTAACATTA	TAAATATGTT
AAGGATAAAG	GATGTTCAAT			
3421	TAGAGGCTGC	ATATACCACG	ACATCGTGTA	TCACGCATTT
TGGTAAAGGC	TACTGCTTTC			
3481	ACATCATCGA	GATCAATCAG	AAGAGCCTGA	ATACCTTACA
GCCGATGCTC	TTTAAGACTA			
3541	GCATCCCTAA	ATTATGCAAG	GCCGAGTCTT	AAGCGGCCGC
GCATGCCAAT	TCACTCCTCA			
3601	GGTGCAGGCT	GCCTATCAGA	AGGTGGTGGC	TGGTGTGGCC
AATGCCCTGG	CTCACAAATA			
3661	CCACTGAGAT	CTTTTTCCCT	CTGCCAAAAA	TTATGGGGAC
ATCATGAAGC	CCCTTGAGCA			
3721	TCTGACTTCT	GGCTAATAAA	GGAAATTTAT	TTTCATTGCA
ATAGTGTGTT	GGAATTTTTT			
3781	GTGTCTCTCA	CTCGGAAGGA	CATATGGGAG	GGCAAATCAT
TTAAAACATC	AGAATGAGTA			

## 038966

3841	TTTGGTTTAG	AGTTTGCAA	CATATGCCCA	TATGCTGGCT
	GCCATGAACA	AAGGTTGGCT		
3901	ATAAAGAGGT	CATCAGTATA	TGAAACAGCC	CCCTGCTGTC
	TATTCCTTAT	TCCATAGAAA		
3961	AGCCTTGACT	TGAGGTTAGA	TTTTTTTTAT	ATTTTGTTTT
	GTGTTATTTT	TTTCTTTAAC		
4021	ATCCCTAAAA	TTTTCTTAC	ATGTTTTACT	AGCCAGATTT
	TTCTCTCTCT	CCTGACTACT		
4081	CCCAGTCATA	GCTGTCCCTC	TTCTCTTATG	GAGATCCCTC
	GACCTGCAGC	CCAAGCTTGG		
4141	CGTAATCATG	GTCATAGCTG	TTCTCTGTGT	GAAATTGTTA
	TCCGCTCACA	ATTCCACACA		
4201	ACATACGAGC	CGGAAGCATA	AAGTGTAAG	CCTGGGGTGC
	CTAATGAGTG	AGCTAACTCA		
4261	CATTAATTGC	GTTGCGCTCA	CTGCCCGCTT	TCCAGTCGGG
	AAACCTGTGC	TGCCAGCGGA		
4321	TCCGCATCTC	AATTAGTCAG	CAACCATAGT	CCCGCCCTA
	ACTCCGCCCA	TCCCGCCCTT		
4381	AACTCCGCC	AGTTCCGCC	ATTCTCCGCC	CCATGGCTGA
	CTAATTTTTT	TTATTATATG		
4441	AGAGGCCGAG	GCCGCCTCGG	CCTCTGAGCT	ATTCCAGAAG
	TAGTGAGGAG	GCTTTTTTGG		
4501	AGGCCTAGGC	TTTTGCAAAA	AGCTAACTTG	TTTATGTCAG
	CTTATAATGG	TTACAAATAA		
4561	AGCAATAGCA	TCACAAATTT	CACAAATAAA	GCATTTTTTT
	CACTGCATTC	TAGTTGTGGT		
4621	TTGTCCAAC	TCATCAATGT	ATCTTATCAT	GTCTGTCCGC
	TTCTCTCGCTC	ACTGACTCGC		
4681	TGCGCTCGGT	CGTTCGGCTG	CGGCGAGCGG	TATCAGCTCA
	CTCAAAGGCG	GTAATACGGT		
4741	TATCCACAGA	ATCAGGGGAT	AACGCAGGAA	AGAACATGTG
	AGCAAAAGGC	CAGCAAAAGG		
4801	CCAGGAACCG	TAAAAAGGCC	GCGTTGCTGG	CGTTTTTCCA
	TAGGCTCCGC	CCCCCTGACG		
4861	AGCATCACAA	AAATCGACGC	TCAAGTCAGA	GGTGCGAAAA
	CCCGACAGGA	CTATAAAGAT		

## 038966

4921	ACCAGGCGTT	TCCCCCTGGA	AGCTCCCTCG	TGCGCTCTCC
TGTTCCGACC	CTGCCGCTTA			
4981	CCGATACCT	GTCCGCCTTT	CTCCCTTCGG	GAAGCGTGGC
GCTTTCTCAT	AGCTCACGCT			
5041	GTAGGTATCT	CAGTTCGGTG	TAGGTCGTTT	GCTCCAAGCT
GGGCTGTGTG	CACGAACCCC			
5101	CCGTCAGCC	CGACCGCTGC	GCCTTATCCG	GTAACCTATCG
TCTTGAGTCC	AACCCGGTAA			
5161	GACACGACTT	ATCGCCACTG	GCAGCAGCCA	CTGGTAAACAG
GATTAGCAGA	GCGAGGTATG			
5221	TAGGCGGTGC	TACAGAGTTC	TTGAAGTGGT	GGCCTAACTA
CGGCTACACT	AGAAGAACAG			
5281	TATTTGGTAT	CTGCGCTCTG	CTGAAGCCAG	TTACCTTCGG
AAAAAGAGTT	GGTAGCTCTT			
5341	GATCCGGCAA	ACAAACCACC	GCTGGTAGCG	GTGGTTTTTTT
TGTTTGCAAG	CAGCAGATTA			
5401	CGCGCAGAAA	AAAAGGATCT	CAAGAAGATC	CTTTGATCTT
TTCTACGGGG	TCTGACGCTC			
5461	AGTGAACGA	AAACTCACGT	TAAGGGATTT	TGGTCATGAG
ATTATCAAAA	AGGATCTTCA			
5521	CCTAGATCCT	TTTAAATTA	AAATGAAGTT	TTAAATCAAT
CTAAAGTATA	TATGAGTAAA			
5581	CTTGGTCTGA	CAGTTAGAAA	AACTCATCGA	GCATCAAAATG
AAACTGCAAT	TTATTCATAT			
5641	CAGGATTATC	AATACCATAT	TTTTGAAAAA	GCCGTTTCTG
TAATGAAGGA	GAAAACCTAC			
5701	CGAGGCAGTT	CCATAGGATG	GCAAGATCCT	GGTATCGGTC
TGCGATTCCG	ACTCGTCCAA			
5761	CATCAATACA	ACCTATTAAT	TTCCCCTCGT	CAAAAATAAG
GTTATCAAGT	GAGAAATCAC			
5821	CATGAGTGAC	GAATGAATCC	GGTGAGAATG	GCAACAGCTT
ATGCATTTCT	TTCCAGACTT			
5881	GTTCAACAGG	CCAGCCATTA	CGCTCGTCAT	CAAAATCACT
CGCATCAACC	AAACCGTTAT			
5941	TCATTCGTGA	TTGCGCCTGA	GCGAGACGAA	ATACGCGATC
GCTGTTAAAA	GGACAATTAC			

038966

6001 AAACAGGAAT CGAATGCAAC CGGCGCAGGA AACTGCCAG  
CGCATCAACA ATATTTTCAC  
6061 CTGAATCAGG ATATTCTTCT AATACCTGGA ATGCTGTTTT  
TCCGGGGATC GCAGTGGTGA  
6121 GTAACCATGC ATCATCAGGA GTACGGATAA AATGCTTGAT  
GGTCGGAAGA GGCATAAATT  
6181 CCGTCAGCCA GTTTAGTCTG ACCATCTCAT CTGTAACATC  
ATTGGCAACG CTACCTTTGC  
6241 CATGTTTCAG AAACAACCTCT GGCGCATCGG GCTTCCATA  
CAATCGATAG ATTGTGCGAC  
6301 CTGATTGCCC GACATTATCG CGAGCCCATT TATACCCATA  
TAAATCAGCA TCCATGTTGG  
6361 AATTTAATCG CGGCC TAGAG CAAGACGTTT CCCGTTGAAT  
ATGGCTCATA ACACCCCTTG  
6421 TATTACTGTT TATGTAAGCA GACAGTTTTA TTGTTTATGA  
TGATATATTT TTATCTTGTG  
6481 CAATGTAACA TCAGAGATTT TGAGACACAA CAATTGGTGC AC

SEQ ID NO: 6

1 AGATCTGTTA CATAACTTAT GGTAATGGC CTGCCTGGCT GACTGCCCAA  
TGACCCCTGC  
61 CCAATGATGT CAATAATGAT GTATGTTCCC ATGTAATGCC AATAGGGACT  
TTCCATTGAT  
121 GTCAATGGGT GGAGTATTTA TGGTAACTGC CCACTTGCCA GTACATCAAG  
TGTATCATAT  
181 GCCAAGTATG CCCCTATTG ATGTCAATGA TGGTAAATGG CCTGCCTGGC  
ATTATGCCCA  
241 GTACATGACC TTATGGGACT TTCCTACTTG GCAGTACATC TATGTATTAG  
TCATTGCTAT  
301 TACCATGGGA ATTCACTAGT GGAGAAGAGC ATGCTTGAGG GCTGAGTGCC  
CCTCAGTGGG  
361 CAGAGAGCAC ATGGCCACA GTCCCTGAGA AGTTGGGGGG AGGGGTGGGC  
AATTGAACTG  
421 GTGCCTAGAG AAGGTGGGGC TTGGGTAAAC TGGGAAAGTG ATGTGGTGTA  
CTGGCTCCAC  
481 CTTTTTCCCC AGGGTGGGGG AGAACCATAT ATAAGTGCAG TAGTCTCTGT  
GAACATTCAA

541 GCTTCTGCCT TCTCCCTCCT GTGAGTTTGC TAGC

**SEQ ID NO:7**

1 GCTAGCCACC ATGCAGAGAA GCCCTCTGGA GAAGGCCTCT GTGGTGAACA  
AGCTGTTCTT

61 CAGCTGGACC AGGCCCATCC TGAGGAAGGG CTACAGGCAG AACTGGAGC  
TGTCTGACAT

121 CTACCAGATC CCCTCTGTGG ACTCTGCTGA CAACCTGTCT GAGAAGCTGG  
AGAGGGAGTG

181 GGATAGAGAG CTGGCCAGCA AGAAGAACCC CAAGCTGATC AATGCCCTGA  
GGAGATGCTT

241 CTTCTGGAGA TTCATGTTCT ATGGCATCTT CCTGTACCTG GGGGAAGTGA  
CCAAGGCTGT

301 GCAGCCTCTG CTGCTGGGCA GAATCATTGC CAGCTATGAC CCTGACAACA  
AGGAGGAGAG

361 GAGCATTGCC ATCTACCTGG GCATTGGCCT GTGCCTGCTG TTCATTGTGA  
GGACCCTGCT

421 GCTGCACCCT GCCATCTTTG GCCTGCACCA CATTGGCATG CAGATGAGGA  
TTGCCATGTT

481 CAGCCTGATC TACAAGAAAA CCCTGAAGCT GTCCAGCAGA GTGCTGGACA  
AGATCAGCAT

541 TGGCCAGCTG GTGAGCCTGC TGAGCAACAA CCTGAACAAG TTTGATGAGG  
GCCTGGCCCT

601 GGCCCACTTT GTGTGGATTG CCCCTCTGCA GGTGGCCCTG CTGATGGGCC  
TGATTTGGGA

661 GCTGCTGCAG GCCTCTGCCT TTTGTGGCCT GGGCTTCCTG ATTGTGCTGG  
CCCTGTTTCA

721 GGCTGGCCTG GGCAGGATGA TGATGAAGTA CAGGGACCAG AGGGCAGGCA  
AGATCAGTGA

781 GAGGCTGGTG ATCACCTCTG AGATGATTGA GAACATCCAG TCTGTGAAGG  
CCTACTGTTG

841 GGAGGAAGCT ATGGAGAAGA TGATTGAAAA CCTGAGGCAG ACAGAGCTGA  
AGCTGACCAG

901 GAAGGCTGCC TATGTGAGAT ACTTCAACAG CTCTGCCTTC TTCTTCTCTG  
GCTTCTTTGT

961 GGTGTTCTCTG TCTGTGCTGC CCTATGCCCT GATCAAGGGG ATCATCCTGA  
GAAAGATTTT

## 038966

1021	CACCACCATC	AGCTTCTGCA	TTGTGCTGAG	GATGGCTGTG
	ACCAGACAGT	TCCCCTGGGC		
1081	TGTGCAGACC	TGGTATGACA	GCCTGGGGGC	CATCAACAAG
	ATCCAGGACT	TCCTGCAGAA		
1141	GCAGGAGTAC	AAGACCCTGG	AGTACAACCT	GACCACCACA
	GAAGTGGTGA	TGGAGAATGT		
1201	GACAGCCTTC	TGGGAGGAGG	GCTTTGGGGA	GCTGTTTGTG
	AAGGCCAAGC	AGAACAACAA		
1261	CAACAGAAAG	ACCAGCAATG	GGGATGACTC	CCTGTTCTTC
	TCCAACTTCT	CCCTGCTGGG		
1321	CACACCTGTG	CTGAAGGACA	TCAACTCAA	GATTGAGAGG
	GGGCAGCTGC	TGGCTGTGGC		
1381	TGGATCTACA	GGGGCTGGCA	AGACCAGCCT	GCTGATGATG
	ATCATGGGGG	AGCTGGAGCC		
1441	TTCTGAGGGC	AAGATCAAGC	ACTCTGGCAG	GATCAGCTTT
	TGCAGCCAGT	TCAGCTGGAT		
1501	CATGCCTGGC	ACCATCAAGG	AGAACATCAT	CTTTGGAGTG
	AGCTATGATG	AGTACAGATA		
1561	CAGGAGTGTG	ATCAAGGCCT	GCCAGCTGGA	GGAGGACATC
	AGCAAGTTTG	CTGAGAAGGA		
1621	CAACATTGTG	CTGGGGGAGG	GAGGCATTAC	ACTGTCTGGG
	GGCCAGAGAG	CCAGAATCAG		
1681	CCTGGCCAGG	GCTGTGTACA	AGGATGCTGA	CCTGTACCTG
	CTGGACTCCC	CCTTTGGCTA		
1741	CCTGGATGTG	CTGACAGAGA	AGGAGATTTT	TGAGAGCTGT
	GTGTGCAAGC	TGATGGCCAA		
1801	CAAGACCAGA	ATCCTGGTGA	CCAGCAAGAT	GGAGCACCTG
	AAGAAGGCTG	ACAAGATCCT		
1861	GATCCTGCAT	GAGGGCAGCA	GCTACTTCTA	TGGGACCTTC
	TCTGAGCTGC	AGAACCTGCA		
1921	GCCTGACTTC	AGCTCTAAGC	TGATGGGCTG	TGACAGCTTT
	GACCAGTTCT	CTGCTGAGAG		
1981	GAGGAACAGC	ATCCTGACAG	AGACCCTGCA	CAGATTCAGC
	CTGGAGGGAG	ATGCCCTGT		
2041	GAGCTGGACA	GAGACCAAGA	AGCAGAGCTT	CAAGCAGACA
	GGGGAGTTTG	GGGAGAAGAG		

## 038966

2101	GAAGAACTCC	ATCCTGAACC	CCATCAACAG	CATCAGGAAG
TTCAGCATTG	TGCAGAAAAC			
2161	CCCCCTGCAG	ATGAATGGCA	TTGAGGAAGA	TTCTGATGAG
CCCCTGGAGA	GGAGACTGAG			
2221	CCTGGTGCCT	GATTCTGAGC	AGGGAGAGGC	CATCCTGCCT
AGGATCTCTG	TGATCAGCAC			
2281	AGGCCCTACA	CTGCAGGCCA	GAAGGAGGCA	GTCTGTGCTG
AACCTGATGA	CCCACTCTGT			
2341	GAACCAGGGC	CAGAACATCC	ACAGGAAAAC	CACAGCCTCC
ACCAGGAAAG	TGAGCCTGGC			
2401	CCCTCAGGCC	AATCTGACAG	AGCTGGACAT	CTACAGCAGG
AGGCTGTCTC	AGGAGACAGG			
2461	CCTGGAGATT	TCTGAGGAGA	TCAATGAGGA	GGACCTGAAA
GAGTGCTTCT	TTGATGACAT			
2521	GGAGAGCATC	CCTGCTGTGA	CCACCTGGAA	CACCTACCTG
AGATACATCA	CAGTGCACAA			
2581	GAGCCTGATC	TTTGTGCTGA	TCTGGTGCCT	GGTGATCTTC
CTGGCTGAAG	TGGCTGCCTC			
2641	TCTGGTGGTG	CTGTGGCTGC	TGGGAAACAC	CCCACTGCAG
GACAAGGGCA	ACAGCACCCA			
2701	CAGCAGGAAC	AACAGCTATG	CTGTGATCAT	CACCTCCACC
TCCAGCTACT	ATGTGTTCTA			
2761	CATCTATGTG	GGAGTGGCTG	ATACCCTGCT	GGCTATGGGC
TTCTTTAGAG	GCCTGCCCCT			
2821	GGTGCACACA	CTGATCACAG	TGAGCAAGAT	CCTCCACCAC
AAGATGCTGC	ACTCTGTGCT			
2881	GCAGGCTCCT	ATGAGCACCC	TGAATACCCT	GAAGGCTGGG
GGCATCCTGA	ACAGATTCTC			
2941	CAAGGATATT	GCCATCCTGG	ATGACCTGCT	GCCTCTCACC
ATCTTTGACT	TCATCCAGCT			
3001	GCTGCTGATT	GTGATTGGGG	CCATTGCTGT	GGTGGCAGTG
CTGCAGCCCT	ACATCTTTGT			
3061	GGCCACAGTG	CCTGTGATTG	TGGCCTTCAT	CATGCTGAGG
GCCTACTTTC	TGCAGACCTC			
3121	CCAGCAGCTG	AAGCAGCTGG	AGTCTGAGGG	CAGAAGCCCC
ATCTTCACCC	ACCTGGTGAC			

## 038966

3181	AAGCCTGAAG	GGCCTGTGGA	CCCTGAGAGC	CTTTGGCAGG
CAGCCCTACT	TTGAGACCCT			
3241	GTTCCACAAG	GCCCTGAACC	TGCACACAGC	CAACTGGTTC
CTCTACCTGT	CCACCCTGAG			
3301	ATGGTTCCAG	ATGAGAATTG	AGATGATCTT	TGTCATCTTC
TTCATTGCTG	TGACCTTCAT			
3361	CAGCATTCTG	ACCACAGGAG	AGGGAGAGGG	CAGAGTGGGC
ATTATCCTGA	CCCTGGCCAT			
3421	GAACATCATG	AGCACACTGC	AGTGGGCAGT	GAACAGCAGC
ATTGATGTGG	ACAGCCTGAT			
3481	GAGGAGTGTG	AGCAGAGTGT	TCAAGTTCAT	TGATATGCCC
ACAGAGGGCA	AGCCTACCAA			
3541	GAGCACCAAG	CCCTACAAGA	ATGGCCAGCT	GAGCAAAGTG
ATGATCATTG	AGAACAGCCA			
3601	TGTGAAGAAG	GATGATATCT	GGCCCAGTGG	AGGCCAGATG
ACAGTGAAGG	ACCTGACAGC			
3661	CAAGTACACA	GAGGGGGGCA	ATGCTATCCT	GGAGAACATC
TCCTTCAGCA	TCTCCCCTGG			
3721	CCAGAGAGTG	GGA CTGCTGG	GAAGAACAGG	CTCTGGCAAG
TCTACCCTGC	TGTCTGCCTT			
3781	CCTGAGGCTG	CTGAACACAG	AGGGAGAGAT	CCAGATTGAT
GGAGTGTCTT	GGGACAGCAT			
3841	CACACTGCAG	CAGTGGAGGA	AGGCCTTTGG	TGTGATCCCC
CAGAAAGTGT	TCATCTTCAG			
3901	TGGCACCTTC	AGGAAGAACC	TGGACCCCTA	TGAGCAGTGG
TCTGACCAGG	AGATTGGAA			
3961	AGTGGCTGAT	GAAGTGGGCC	TGAGAAGTGT	GATTGAGCAG
TTCCTGGCA	AGCTGGACTT			
4021	TGTCCTGGTG	GATGGGGGCT	GTGTGCTGAG	CCATGGCCAC
AAGCAGCTGA	TGTGCCTGGC			
4081	CAGATCAGTG	CTGAGCAAGG	CCAAGATCCT	GCTGCTGGAT
GAGCCTTCTG	CCCACCTGGA			
4141	TCCTGTGACC	TACCAGATCA	TCAGGAGGAC	CCTCAAGCAG
GCCTTTGCTG	ACTGCACAGT			
4201	CATCCTGTGT	GAGCACAGGA	TTGAGGCCAT	GCTGGAGTGC
CAGCAGTTCC	TGGTGATTGA			



038966

4261 GGAGAACAAA GTGAGGCAGT ATGACAGCAT CCAGAAGCTG  
CTGAATGAGA GGAGCCTGTT  
4321 CAGGCAGGCC ATCAGCCCCT CTGATAGAGT GAAGCTGTTC  
CCCCACAGGA ACAGCTCCAA  
4381 GTGCAAGAGC AAGCCCAGA TTGCTGCCCT GAAGGAGGAG  
ACAGAGGAGG AAGTGCAGGA  
4441 CACCAGGCTG TGAGGGCCC

SEQ ID NO:8

1 GGGCCCAATC AACCTCTGGA TTACAAAATT TGTGAAAGAT TGACTGGTAT  
TCTTAACTAT  
61 GTTGCTCCTT TTACGCTATG TGGATACGCT GCTTTAATGC CTTTGTATCA  
TGCTATTGCT  
121 TCCCGTATGG CTTTCATTTT CTCCTCCTTG TATAAATCCT GGTGTGCTGC  
TCTTTATGAG  
181 GAGTTGTGGC CCGTTGTGAG GCAACGTGGC GTGGTGTGCA CTGTGTTTGC  
TGACGCAACC  
241 CCCACTGGTT GGGGCATTGC CACCACCTGT CAGCTCCTTT CCGGGACTTT  
CGCTTTCCCC  
301 CTCCTATTG CCACGGCGGA ACTCATCGCC GCCTGCCTTG CCCGCTGCTG  
GACAGGGGCT  
361 CGGCTGTGG GCACTGACAA TTCCGTGGTG TTGTCGGGGA AATCATCGTC  
CTTTCCTTGG  
421 CTGCTCGCCT GTGTGCCAC CTGGATTCTG CGCGGGACGT CCTTCTGCTA  
CGTCCCTTCG  
481 GCCCTCAATC CAGCGGACCT TCCTTCCCGC GGCCTGCTGC CGGCTCTGCG  
GCCTCTCCG  
541 CGTCTTCGCC TTCGCCCTCA GACGAGTCGG ATCTCCCTTT GGGCCGCCCTC  
CCCGCAAGCT

SEQ ID NO:9

1 GGTACCTCAA TATTGGCCAT TAGCCATATT ATTCATTGGT TATATAGCAT  
AAATCAATAT  
61 TGGCTATTGG CCATTGCATA CGTTGTATCT ATATCATAAT ATGTACATTT  
ATATTGGCTC  
121 ATGTCCAATA TGACCGCCAT GTTGGCATTG ATTATTGACT AGTTATTAAT  
AGTAATCAAT  
181 TACGGGGTCA TTAGTTCATA GCCCATATAT GGAGTTCCGC GTTACATAAC

TTACGGTAAA  
 241 TGGCCCGCCT GGCTGACCGC CCAACGACCC CCGCCCATG ACGTCAATAA  
 TGACGTATGT  
 301 TCCCATAGTA ACGCCAATAG GGACTTTCCA TTGACGTCAA TGGGTGGAGT  
 ATTTACGGTA  
 361 AACTGCCAC TTGGCAGTAC ATCAAGTGTA TCATATGCCA AGTCCGCCCC  
 CTATTGACGT  
 421 CAATGACGGT AAATGGCCCG CCTGGCATTG TGCCAGTAC ATGACCTTAC  
 GGGACTTTCC  
 481 TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC  
 GGTTTTGGCA  
 541 GTACACCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA TTTCCAAGTC  
 TCCACCCCAT  
 601 TGACGTCAAT GGGAGTTTGT TTTGGCACCA AAATCAACGG GACTTTCCAA  
 AATGTCGTAA  
 661 CAACTGCGAT CGCCCGCCCC GTTGACGCAA ATGGGCGGTA GGC GTGTACG  
 GTGGGAGGTC  
 721 TATATAAGCA GAGCTCGCTG GCTTGTAACT CAGTCTCTTA CTAGGAGACC  
 AGCTTGAGCC  
 781 TGGGTGTTTCG CTGGTTAGCC TAACCTGGTT GGCCACCAGG GGTAAGGACT  
 CCTTGGCTTA  
 841 GAAAGCTAAT AAATTGCCT GCATTAGAGC TTATCTGAGT CAAGTGTCCCT  
 CATTGACGCC  
 901 TCACTCTCTT GAACGGGAAT CTTCTTACT GGGTTCTCTC TCTGACCCAG  
 GCGAGAGAAA  
 961 CTCCAGCAGT GGC GCCCGAA CAGGGACTTG AGTGAGAGTG TAGGCACGTA  
 CAGCTGAGAA  
 1021 GCGCTCGGAC GCGAAGGAAG CGCGGGGTGC GACGCGACCA  
 AGAAGGAGAC TTGGTGAGTA  
 1081 GGCTTCTCGA GTGCCGGGAA AAAGCTCGAG CCTAGTTAGA  
 GGACTAGGAG AGGCCGTAGC  
 1141 CGTAACTACT CTTGGGCAAG TAGGGCAGGC GGTGGGTACG  
 CAATGGGGGC GGCTACCTCA  
 1201 GCACTAAATA GGAGACAATT AGACCAATTT GAGAAAATAC  
 GACTTCGCCC GAACGGAAAAG  
 1261 AAAAAGTACC AAATTAAACA TTTAATATGG GCAGGCAAGG

## 038966

AGATGGAGCG	CTTCGGCCTC			
1321	CATGAGAGGT	TGTTGGAGAC	AGAGGAGGGG	TGTAAAAGAA
TCATAGAAGT	CCTCTACCCC			
1381	CTAGAACCAA	CAGGATCGGA	GGCCTTAAAA	AGTCTGTTCA
ATCTTGTGTG	CGTGTATAT			
1441	TGCTTGCACA	AGGAACAGAA	AGTGAAAGAC	ACAGAGGAAG
CAGTAGCAAC	AGTAAGACAA			
1501	CACTGCCATC	TAGTGGAAAA	AGAAAAAAGT	GCAACAGAGA
CATCTAGTGG	ACAAAAGAAA			
1561	AATGACAAGG	GAATAGCAGC	GCCACCTGGT	GGCAGTCAGA
ATTTTCCAGC	GCAACAACAA			
1621	GGAAATGCCT	GGGTACATGT	ACCCTTGTCA	CCGCGCACCT
TAAATGCGTG	GGTAAAAGCA			
1681	GTAGAGGAGA	AAAAATTTGG	AGCAGAAATA	GTACCCATTT
TTTTGTTTCA	AGCCCTATCG			
1741	AATTCCCCTT	TGTGCTAGGG	TTCTTAGGCT	TCTTGGGGGC
TGCTGGAACT	GCAATGGGAG			
1801	CAGCGGCGAC	AGCCCTGACG	GTCCAGTCTC	AGCATTGCT
TGCTGGGATA	CTGCAGCAGC			
1861	AGAAGAATCT	GCTGGCGGCT	GTGGAGGCTC	AACAGCAGAT
GTTGAAGCTG	ACCATTTGGG			
1921	GTGTTAAAAA	CCTCAATGCC	CGCGTCACAG	CCCTTGAGAA
GTACCTAGAG	GATCAGGCAC			
1981	GACTAAACTC	CTGGGGGTGC	GCATGGAAAC	AAGTATGTCA
TACCACAGTG	GAGTGGCCCT			
2041	GGACAAATCG	GACTCCGGAT	TGGCAAATA	TGACTTGGTT
GGAGTGGGAA	AGACAAATAG			
2101	CTGATTTGGA	AAGCAACATT	ACGAGACAAT	TAGTGAAGGC
TAGAGAACAA	GAGGAAAAGA			
2161	ATCTAGATGC	CTATCAGAAG	TTAACTAGTT	GGTCAGATTT
CTGGTCTTGG	TTCGATTTCT			
2221	CAAAATGGCT	TAACATTTTA	AAAATGGGAT	TTTTAGTAAT
AGTAGGAATA	ATAGGGTTAA			
2281	GATTACTTTA	CACAGTATAT	GGATGTATAG	TGAGGGTTAG
GCAGGGATAT	GTTCTCTAT			
2341	CTCCACAGAT	CCATATCCGC	GGCAATTTTA	AAAGAAAGGG

## 038966

AGGAATAGGG	GGACAGACTT			
2401	CAGCAGAGAG	ACTAATTAAT	ATAATAACAA	CACAATTAGA
AATACAACAT	TTACAAACCA			
2461	AAATTCAAAA	AATTTTAAAT	TTTAGAGCCG	CGGAGATCTG
TTACATAACT	TATGGTAAAT			
2521	GGCCTGCCTG	GCTGACTGCC	CAATGACCCC	TGCCCAATGA
TGTCAATAAT	GATGTATGTT			
2581	CCCATGTAAT	GCCAATAGGG	ACTTTCCATT	GATGTCAATG
GGTGGAGTAT	TTATGGTAAC			
2641	TGCCCACTTG	GCAGTACATC	AAGTGTATCA	TATGCCAAGT
ATGCCCCCTA	TTGATGTCAA			
2701	TGATGGTAAA	TGGCCTGCCT	GGCATTATGC	CCAGTACATG
ACCTTATGGG	ACTTTCCTAC			
2761	TTGGCAGTAC	ATCTATGTAT	TAGTCATTGC	TATTACCATG
GGAATTCACT	AGTGGAGAAG			
2821	AGCATGCTTG	AGGGCTGAGT	GCCCCTCAGT	GGGCAGAGAG
CACATGGCCC	ACAGTCCCTG			
2881	AGAAGTTGGG	GGGAGGGGTG	GGCAATTGAA	CTGGTGCCTA
GAGAAGGTGG	GGCTTGGGTA			
2941	AACTGGGAAA	GTGATGTGGT	GTA CTGGCTC	CACCTTTTTC
CCCAGGGTGG	GGGAGAACCA			
3001	TATATAAGTG	CAGTAGTCTC	TGTGAACATT	CAAGCTTCTG
CCTTCTCCCT	CCTGTGAGTT			
3061	TGCTAGCCAC	CATGCCCAGC	TCTGTGTCCT	GGGGCATTCT
GCTGCTGGCT	GGCCTGTGCT			
3121	GTCTGGTGCC	TGTGTCCCTG	GCTGAGGACC	CTCAGGGGGA
TGCTGCCCAG	AAAACAGACA			
3181	CCTCCCACCA	TGACCAGGAC	CACCCACCTT	TCAACAAGAT
CACCCCAAC	CTGGCAGAGT			
3241	TTGCCTTCAG	CCTGTACAGA	CAGCTGGCCC	ACCAGAGCAA
CAGCACCAAC	ATCTTTTTCA			
3301	GCCCTGTGTC	CATTGCCACA	GCCTTTGCCA	TGCTGAGCCT
GGGCACCAAG	GCTGACACCC			
3361	ATGATGAGAT	CCTGGAAGGC	CTGAACTTCA	ACCTGACAGA
GATCCCTGAG	GCCCAGATCC			
3421	ATGAGGGCTT	CCAGGAACTG	CTGAGAACCC	TGAACCAGCC

## 038966

AGACAGCCAG	CTGCAGCTGA			
3481	CAACAGGCAA	TGGGCTGTTT	CTGTCTGAGG	GCCTGAAGCT
GGTGGACAAG	TTTCTGGAAG			
3541	ATGTGAAGAA	GCTGTACCAC	TCTGAGGCCT	TCACAGTGAA
CTTTGGGGAC	ACAGAAGAGG			
3601	CCAAGAAACA	GATCAATGAC	TATGTGAAA	AGGGCACCCA
GGCAAGATT	GTGGACCTTG			
3661	TGAAAGAGCT	GGACAGGGAC	ACTGTGTTTG	CCCTTGTGAA
CTACATCTTC	TTCAAGGGCA			
3721	AGTGGGAGAG	GCCCTTTGAA	GTGAAGGACA	CTGAGGAAGA
GGAATTCCAT	GTGGACCAAG			
3781	TGACCACAGT	GAAGGTGCCA	ATGATGAAGA	GACTGGGGAT
GTTCAATATC	CAGCACTGCA			
3841	AGAACTGAG	CAGCTGGGTG	CTGCTGATGA	AGTACCTGGG
CAATGCTACA	GCCATATTCT			
3901	TTCTGCCTGA	TGAGGGCAAG	CTGCAGCACC	TGGAAAATGA
GCTGACCCAT	GACATCATCA			
3961	CCAAATTTCT	GGAAAATGAG	GACAGAAGAT	CTGCCAGCCT
GCATCTGCCC	AAGCTGAGCA			
4021	TCACAGGCAC	ATATGACCTG	AAGTCTGTGC	TGGGACAGCT
GGGAATCACC	AAGGTGTTCA			
4081	GCAATGGGGC	AGACCTGAGT	GGAGTGACAG	AGGAAGCCCC
TCTGAAGCTG	TCCAAGGCTG			
4141	TGCACAAGGC	AGTGCTGACC	ATTGATGAGA	AGGGCACAGA
GGCTGCTGGG	GCCATGTTTC			
4201	TGGAAGCCAT	CCCCATGTCC	ATCCCCCAG	AAGTGAAGTT
CAACAAGCCC	TTTGTGTTCC			
4261	TGATGATTGA	GCAGAACACC	AAGAGCCCC	TGTTTATGGG
CAAGGTTGTG	AACCCACCC			
4321	AGAAATGAGG	GCCCAATCAA	CCTCTGGATT	ACAAAATTTG
TGAAAGATTG	ACTGGTATTC			
4381	TTAACTATGT	TGCTCCTTTT	ACGCTATGTG	GATACGCTGC
TTTAATGCCT	TTGTATCATG			
4441	CTATTGCTTC	CCGTATGGCT	TTCATTTTCT	CCTCCTTGTA
TAAATCCTGG	TTGCTGTCTC			
4501	TTTATGAGGA	GTTGTGGCCC	GTTGTCAGGC	AACGTGGCGT

## 038966

GGTGTGCACT GTGTTTGCTG  
4561 ACGCAACCCC CACTGGTTGG GGCATTGCCA CCACCTGTCA  
GCTCCTTTCC GGGACTTTCC  
4621 CTTTCCCCTT CCTATTGCC ACGGCGGAAC TCATCGCCGC  
CTGCCTTGCC CGCTGCTGGA  
4681 CAGGGGCTCG GCTGTTGGGC ACTGACAATT CCGTGGTGTT  
GTCGGGGAAA TCATCGTCTT  
4741 TTCCTTGGCT GCTCGCCTGT GTTGCCACCT GGATTCTGCG  
CGGGACGTCC TTCTGCTACG  
4801 TCCCTTCGGC CCTCAATCCA GCGGACCTTC CTTCCC GCGG  
CCTGCTGCCG GCTCTGCGGC  
4861 CTCTTCCGCG TCTTCGCCTT CGCCCTCAGA CGAGTCGGAT  
CTCCCTTTGG GCCGCCTCCC  
4921 CGCAAGCTTC GCACTTTTTA AAAGAAAAGG GAGGACTGGA  
TGGGATTTAT TACTCCGATA  
4981 GGACGCTGGC TTGTAACTCA GTCTCTTACT AGGAGACCAG  
CTTGAGCCTG GGTGTTCGCT  
5041 GGTTAGCCTA ACCTGGTTGG CCACCAGGGG TAAGGACTCC  
TTGGCTTAGA AAGCTAATAA  
5101 ACTTGCTGTC ATTAGAGCTC TTACGCGTCC CGGGCTCGAG  
ATCCGCATCT CAATTAGTCA  
5161 GCAACCATAG TCCCGCCCCT AACTCCGCCC ATCCCGCCCC  
TAACTCCGCC CAGTTCGGCC  
5221 CATTCTCCGC CCCATGGCTG ACTAATTTTT TTTATTTATG  
CAGAGGCCGA GGCCGCTCG  
5281 GCCTCTGAGC TATTCAGAA GTAGTGAGGA GGCTTTTTTG  
GAGGCCTAGG CTTTTGCAAA  
5341 AAGCTAACTT GTTTATTGCA GCTTATAATG GTTACAAATA  
AAGCAATAGC ATCACA AATT  
5401 TCACAAATAA AGCATTTTTT TCACTGCATT CTAGTTGTGG  
TTTGTCCAAA CTCATCAATG  
5461 TATCTTATCA TGTCTGTCCG CTTCCCTCGCT CACTGACTCG  
CTGCGCTCGG TCGTTCGGCT  
5521 GCGGCGAGCG GTATCAGCTC ACTCAAAGGC GGTAATACGG  
TTATCCACAG AATCAGGGGA  
5581 TAACGCAGGA AAGAACATGT GAGCAAAAGG CCAGCAAAAG

## 038966

GCCAGGAACC	GTAAAAAGGC			
5641	CGCGTTGCTG	GCGTTTTTCC	ATAGGCTCCG	CCCCCCTGAC
GAGCATCACA	AAAATCGACG			
5701	CTCAAGTCAG	AGGTGGCGAA	ACCCGACAGG	ACTATAAAGA
TACCAGGCGT	TTCCCCCTGG			
5761	AAGCTCCCTC	GTGCGCTCTC	CTGTTCCGAC	CCTGCCGCTT
ACCGGATACC	TGTCCGCCTT			
5821	TCTCCCTTCG	GGAAGCGTGG	CGCTTTCTCA	TAGCTCACGC
TGTAGGTATC	TCAGTTCGGT			
5881	GTAGGTCGTT	CGCTCCAAGC	TGGGCTGTGT	GCACGAACCC
CCCGTTCAGC	CCGACCGCTG			
5941	CGCCTTATCC	GGTAACTATC	GTCTTGAGTC	CAACCCGGTA
AGACACGACT	TATCGCCACT			
6001	GGCAGCAGCC	ACTGGTAACA	GGATTAGCAG	AGCGAGGTAT
GTAGGCGGTG	CTACAGAGTT			
6061	CTTGAAGTGG	TGGCCTAACT	ACGGCTACAC	TAGAAGAACA
GTATTTGGTA	TCTGCGCTCT			
6121	GCTGAAGCCA	GTTACCTTCG	GAAAAAGAGT	TGGTAGCTCT
TGATCCGGCA	AACAAACCAC			
6181	CGCTGGTAGC	GGTGGTTTTT	TTGTTTGCAA	GCAGCAGATT
ACGCGCAGAA	AAAAAGGATC			
6241	TCAAGAAGAT	CCTTTGATCT	TTTCTACGGG	GTCTGACGCT
CAGTGAACG	AAAATCAGC			
6301	TTAAGGGATT	TTGGTCATGA	GATTATCAAA	AAGGATCTTC
ACCTAGATCC	TTTTAAATTA			
6361	AAAATGAAGT	TTTAAATCAA	TCTAAAGTAT	ATATGAGTAA
ACTTGGTCTG	ACAGTTAGAA			
6421	AAACTCATCG	AGCATCAAAT	GAAACTGCAA	TTTATTGATA
TCAGGATTAT	CAATACCATA			
6481	TTTTTGAAAA	AGCCGTTTCT	GTAATGAAGG	AGAAACTCA
CCGAGGCAGT	TCCATAGGAT			
6541	GGCAAGATCC	TGGTATCGGT	CTGCGATTCC	GACTCGTCCA
ACATCAATAC	AACCTATTAA			
6601	TTTCCCCTCG	TCAAAAATAA	GGTTATCAAG	TGAGAAATCA
CCATGAGTGA	CFACTGAATC			
6661	CGGTGAGAAT	GGCAACAGCT	TATGCATTTT	TTTCCAGACT

## 038966

TGTTCAACAG GCCAGCCATT  
 6721 ACGCTCGTCA TCAAATCAC TCGCATCAAC CAAACCGTTA  
 TTCATTTCGTG ATTGGCCCTG  
 6781 AGCGAGACGA AATACGCGAT CGCTGTTAAA AGGACAATTA  
 CAAACAGGAA TCGAATGCAA  
 6841 CCGGCGCAGG AACACTGCCA GCGCATCAAC AATATTTTCA  
 CCTGAATCAG GATATTCCTC  
 6901 TAATACCTGG AATGCTGTTT TTCCGGGGAT CGCAGTGGTG  
 AGTAACCATG CATCATCAGG  
 6961 AGTACGGATA AAATGCTTGA TGGTCGGAAG AGGCATAAAT  
 TCCGTCAGCC AGTTTAGTCT  
 7021 GACCATCTCA TCTGTAACAT CATTGGCAAC GCTACCTTTG  
 CCATGTTTCA GAAACAACCTC  
 7081 TGGCGCATCG GGCTTCCCAT ACAATCGATA GATTGTCGCA  
 CCTGATTGCC CGACATTATC  
 7141 GCGAGCCCAT TTATACCCAT ATAAATCAGC ATCCATGTTG  
 GAATTTAATC GCGGCCTAGA  
 7201 GCAAGACGTT TCCCGTTGAA TATGGCTCAT AACACCCCTT  
 GTATTACTGT TTATGTAAGC  
 7261 AGACAGTTTT ATTGTTTCATG ATGATATATT TTTATCTTGT  
 GCAATGTAAC ATCAGAGATT  
 7321 TTGAGACACA ACAATTGGTC GACGGATCC  
**SEQ ID NO:10**  
 1 GGTACCTCAA TATTGGCCAT TAGCCATATT ATTCATTGGT TATATAGCAT  
 AAATCAATAT  
 61 TGGCTATTGG CCATTGCATA CGTTGTATCT ATATCATAAT ATGTACATTT  
 ATATTGGCTC  
 121 ATGTCCAATA TGACCGCCAT GTTGGCATTG ATTATTGACT AGTTATTAAT  
 AGTAATCAAT  
 181 TACGGGGTCA TTAGTTCATA GCCCATATAT GGAGTTCCGC GTTACATAAC  
 TTACGGTAAA  
 241 TGGCCCGCCT GGCTGACCGC CCAACGACCC CCGCCCATG ACGTCAATAA  
 TGACGTATGT  
 301 TCCCATAGTA ACGCCAATAG GGACTTTCCA TTGACGTCAA TGGGTGGAGT  
 ATTTACGGTA  
 361 AACTGCCAC TTGGCAGTAC ATCAAGTGTA TCATATGCCA AGTCCGCCCC



038966

CTATTGACGT  
 421 CAATGACGGT AAATGGCCCG CCTGGCATTG TGCCAGTAC ATGACCTTAC  
 GGGACTTTCC  
 481 TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC  
 GGTTTGGCA  
 541 GTACACCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA TTTCCAAGTC  
 TCCACCCCAT  
 601 TGACGTCAAT GGGAGTTTGT TTTGGCACCA AAATCAACGG GACTTTCCAA  
 AATGTCGTAA  
 661 CAACTGCGAT CGCCCGCCCC GTTGACGCAA ATGGGCGGTA GGC GTGTACG  
 GTGGGAGGTC  
 721 TATATAAGCA GAGCTCGCTG GCTTGTAACT CAGTCTCTTA CTAGGAGACC  
 AGCTTGAGCC  
 781 TGGGTGTTTCG CTGGTTAGCC TAACCTGGTT GGCCACCAGG GGTAAGGACT  
 CCTTGGCTTA  
 841 GAAAGCTAAT AAACCTGCCT GCATTAGAGC TTATCTGAGT CAAGTGTCTC  
 CATTGACGCC  
 901 TCACTCTCTT GAACGGGAAT CTTCTTACT GGGTCTCTC TCTGACCCAG  
 GCGAGAGAAA  
 961 CTCCAGCAGT GGCGCCGAA CAGGGACTTG AGTGAGAGTG TAGGCACGTA  
 CAGCTGAGAA  
 1021 GCGTTCGGAC GCGAAGGAAG CGCGGGTGC GACGCGACCA  
 AGAAGGAGAC TTGGTGAGTA  
 1081 GGCTTCTCGA GTGCCGGGAA AAAGCTCGAG CCTAGTTAGA  
 GGACTAGGAG AGGCCGTAGC  
 1141 CGTAACTACT CTTGGGCAAG TAGGGCAGGC GGTGGGTACG  
 CAATGGGGGC GGCTACCTCA  
 1201 GCACTAAATA GGAGACAATT AGACCAATTT GAGAAAATAC  
 GACTTCGCCC GAACGGAAAG  
 1261 AAAAAGTACC AAATTAACA TTTAATATGG GCAGGCAAGG  
 AGATGGAGCG CTTCCGCCCTC  
 1321 CATGAGAGGT TGTGGGAGAC AGAGGAGGGG TGTAAGAGAA  
 TCATAGAAGT CCTCTACCCC  
 1381 CTAGAACCAA CAGGATCGGA GGGCTTAAAA AGTCTGTTCA  
 ATCTTGTGTG CGTGCTATAT  
 1441 TGCTTGCACA AGGAACAGAA AGTGAAAGAC ACAGAGGAAG

CAGTAGCAAC	AGTAAGACAA			
1501	CACTGCCATC	TAGTGAAAA	AGAAAAAAGT	GCAACAGAGA
CATCTAGTGG	ACAAAAGAAA			
1561	AATGACAAGG	GAATAGCAGC	GCCACCTGGT	GGCAGTCAGA
ATTTTCCAGC	GCAACAACAA			
1621	GGAAATGCCT	GGGTACATGT	ACCCTTGTCA	CCGCGCACCT
TAAATGCGTG	GGTAAAAGCA			
1681	GTAGAGGAGA	AAAAATTTGG	AGCAGAAATA	GTACCCATTT
TTTTGTTTCA	AGCCCTATCG			
1741	AATTCCTGTT	TGTGCTAGGG	TTCTTAGGCT	TCTTGGGGGC
TGCTGGAACT	GCAATGGGAG			
1801	CAGCGGCGAC	AGCCCTGACG	GTCCAGTCTC	AGCATTTGCT
TGCTGGGATA	CTGCAGCAGC			
1861	AGAAGAATCT	GCTGGCGGCT	GTGGAGGCTC	AACAGCAGAT
GTTGAAGCTG	ACCATTGGG			
1921	GTGTTAAAAA	CCTCAATGCC	CGCGTCACAG	CCCTTGAGAA
GTACCTAGAG	GATCAGGCAC			
1981	GACTAAACTC	CTGGGGGTGC	GCATGGAAAC	AAGTATGTCA
TACCACAGTG	GAGTGGCCCT			
2041	GGACAAATCG	GACTCCGGAT	TGGCAAAATA	TGACTTGGTT
GGAGTGGGAA	AGACAAATAG			
2101	CTGATTGGGA	AAGCAACATT	ACGAGACAAT	TAGTGAAGGC
TAGAGAACAA	GAGGAAAAGA			
2161	ATCTAGATGC	CTATCAGAAG	TTAACTAGTT	GGTCAGATTT
CTGGTCTTGG	TTCGATTTCT			
2221	CAAAATGGCT	TAACATTTTA	AAAATGGGAT	TTTTAGTAAT
AGTAGGAATA	ATAGGGTTAA			
2281	GATTACTTTA	CACAGTATAT	GGATGTATAG	TGAGGGTTAG
GCAGGGATAT	GTTCTCTAT			
2341	CTCCACAGAT	CCATATCCGC	GGCAATTTTA	AAAGAAAGGG
AGGAATAGGG	GGACAGACTT			
2401	CAGCAGAGAG	ACTAATTAAT	ATAATAACAA	CACAATTAGA
AATACAACAT	TTACAAACCA			
2461	AAATTCAAAA	AATTTTAAAT	TTTAGAGCCG	CGGAGATCTG
TTACATAACT	TATGGTAAAT			
2521	GGCCTGCCTG	GCTGACTGCC	CAATGACCCC	TGCCCAATGA

## 038966

TGCAATAAT	GATGTATGTT			
2581	CCCATGTAAT	GCCAATAGGG	ACTTTCATT	GATGTCAATG
GGTGGAGTAT	TTATGGTAAC			
2641	TGCCCACTTG	GCAGTACATC	AAGTGTATCA	TATGCCAAGT
ATGCCCCCTA	TTGATGTCAA			
2701	TGATGGTAAA	TGGCCTGCCT	GGCATTATGC	CCAGTACATG
ACCTTATGGG	ACTTTCCTAC			
2761	TTGGCAGTAC	ATCTATGTAT	TAGTCATTGC	TATTACCATG
GGAATTCACT	AGTGGAGAAG			
2821	AGCATGCTTG	AGGGCTGAGT	GCCCCTCAGT	GGGCAGAGAG
CACATGGCCC	ACAGTCCCTG			
2881	AGAAGTTGGG	GGGAGGGGTG	GGCAATTGAA	CTGGTGCCTA
GAGAAGGTGG	GGCTTGGGTA			
2941	AACTGGGAAA	GTGATGTGGT	GTACTGGCTC	CACCTTTTTC
CCCAGGGTGG	GGGAGAACCA			
3001	TATATAAGTG	CAGTAGTCTC	TGTGAACATT	CAAGCTTCTG
CCTTCTCCCT	CCTGTGAGTT			
3061	TGCTAGCCAC	CATGGGAGTG	AAGGTGCTGT	TTGCCCTGAT
CTGCATTGCT	GTGGCTGAGG			
3121	CCAAGCCAC	AGAGAACAAT	GAGGACTTCA	ACATTGTGGC
TGTGGCCAGC	AACTTTGCCA			
3181	CCACAGACCT	GGATGCTGAC	AGGGGCAAGC	TGCCTGGCAA
GAAGCTGCCC	CTGGAAGTCC			
3241	TGAAAGAGAT	GGAAGCCAAT	GCCAGGAAGG	CTGGCTGCAC
AAGAGGCTGT	CTGATCTGCC			
3301	TGAGCCACAT	CAAGTGCACC	CCCAAGATGA	AGAAGTTCAT
CCCTGGCAGG	TGCCACACCT			
3361	ATGAAGGGGA	CAAAGAGTCT	GCCCAGGGGG	GAATTGGAGA
GGCCATTGTG	GACATCCCTG			
3421	AGATCCCTGG	CTTCAAGGAC	CTGGAACCCA	TGGAACAGTT
CATTGCCCAG	GTGGACCTGT			
3481	GTGTGGACTG	CACTACAGGC	TGTCTCAAGG	GCCTGGCCAA
TGTGCAGTGC	TCTGACCTGC			
3541	TGAAGAAGTG	GCTGCCCCAG	AGATGTGCCA	CCTTTGCCAG
CAAGATCCAG	GGCCAGGTGG			
3601	ACAAGATCAA	GGGAGCTGGG	GGAGATTGAT	GAGGGCCCAA

## 038966

TCAACCTCTG	GATTACAAAA			
3661	TTTGTGAAAG	ATTGACTGGT	ATTCTTAACT	ATGTTGCTCC
TTTTACGCTA	TGTGGATACG			
3721	CTGCTTTAAT	GCCTTTGTAT	CATGCTATTG	CTTCCCGTAT
GGCTTTCATT	TTCTCCTCCT			
3781	TGTATAAATC	CTGGTTGCTG	TCTCTTTATG	AGGAGTTGTG
GCCCGTTGTC	AGGCAACGTG			
3841	GCGTGGTGTG	CACTGTGTTT	GCTGACGCAA	CCCCACTGG
TTGGGGCATT	GCCACCACCT			
3901	GTCAGCTCCT	TTCCGGGACT	TTCGCTTTCC	CCCTCCCTAT
TGCCACGGCG	GAATCATCG			
3961	CCGCCTGCCT	TGCCCGCTGC	TGGACAGGGG	CTCGGCTGTT
GGGCACTGAC	AATTCCGTGG			
4021	TGTTGTCGGG	GAAATCATCG	TCCTTTCCTT	GGCTGCTCGC
CTGTGTTGCC	ACCTGGATTG			
4081	TGCGCGGGAC	GTCCTTCTGC	TACGTCCCTT	CGGCCCTCAA
TCCAGCGGAC	CTTCCTTCCC			
4141	GCGGCCTGCT	GCCGGCTCTG	CGGCCTCTT	CGCGTCTTCG
CCTTCGCCCT	CAGACGAGTC			
4201	GGATCTCCCT	TTGGGCCGCC	TCCCCGCAAG	CTTCGCACTT
TTTAAAAGAA	AAGGGAGGAC			
4261	TGGATGGGAT	TTATTACTCC	GATAGGACGC	TGGCTTGTA
CTCAGTCTCT	TACTAGGAGA			
4321	CCAGCTTGAG	CCTGGGTGTT	CGCTGGTTAG	CCTAACCTGG
TTGGCCACCA	GGGGTAAGGA			
4381	CTCCTTGGCT	TAGAAAGCTA	ATAAACTTGC	CTGCATTAGA
GCTCTTACGC	GTCCCGGGCT			
4441	CGAGATCCGC	ATCTCAATTA	GTCAGCAACC	ATAGTCCCGC
CCCTAACTCC	GCCCATCCCG			
4501	CCCCTAACTC	CGCCCAGTTC	CGCCCATTCT	CGCCCCATG
GCTGACTAAT	TTTTTTTATT			
4561	TATGCAGAGG	CCGAGGCCGC	CTCGGCCTCT	GAGCTATTCC
AGAAGTAGTG	AGGAGGCTTT			
4621	TTTGGAGGCC	TAGGCTTTTG	CAAAAAGCTA	ACTTGTTTAT
TGCAGCTTAT	AATGGTTACA			
4681	AATAAAGCAA	TAGCATCACA	AATTTCAACA	ATAAAGCATT

## 038966

TTTTTCACTG	CATTCTAGTT				
4741	GTGGTTTGTC	CAAACATCATC	AATGTATCTT	ATCATGTCTG	
TCCGCTTCCT	CGCTCACTGA				
4801	CTCGCTGCGC	TCGGTCGTTC	GGCTGCGGCG	AGCGGTATCA	
GCTCACTCAA	AGGCGGTAAT				
4861	ACGGTTATCC	ACAGAATCAG	GGGATAACGC	AGGAAAGAAC	
ATGTGAGCAA	AAGGCCAGCA				
4921	AAAGGCCAGG	AACCGTAAAA	AGGCCGCGTT	GCTGGCGTTT	
TCCCATAGGC	TCCGCCCCC				
4981	TGACGAGCAT	CACAAAAATC	GACGCTCAAG	TGACAGGTGG	
CGAAACCCGA	CAGGACTATA				
5041	AAGATAACCAG	GCGTTTCCCC	CTGGAAGCTC	CCTCGTGCGC	
TCTCCTGTTC	CGACCCTGCC				
5101	GCTTACCGGA	TACCTGTCCG	CCTTTCTCCC	TTCGGGAAGC	
GTGGCGCTTT	CTCATAGCTC				
5161	ACGCTGTAGG	TATCTCAGTT	CGGTGTAGGT	CGTTCGCTCC	
AAGCTGGGCT	GTGTGCACGA				
5221	ACCCCCGTT	CAGCCCGACC	GCTGCGCCTT	ATCCGTAAC	
TATCGTCTTG	AGTCCAACCC				
5281	GGTAAGACAC	GACTTATCGC	CACTGGCAGC	AGCCACTGGT	
AACAGGATTA	GCAGAGCGAG				
5341	GTATGTAGGC	GGTGCTACAG	AGTTCTTGAA	GTGGTGGCCT	
AACTACGGCT	ACACTAGAAG				
5401	AACAGTATTT	GGTATCTGCG	CTCTGCTGAA	GCCAGTTACC	
TTCGGAAAAA	GAGTTGGTAG				
5461	CTCTTGATCC	GGCAAACAAA	CCACCGCTGG	TAGCGGTGGT	
TTTTTTGTTT	GCAAGCAGCA				
5521	GATTACGCGC	AGAAAAAAG	GATCTCAAGA	AGATCCTTTG	
ATCTTTTCTA	CGGGGTCTGA				
5581	CGCTCAGTGG	AACGAAAACT	CACGTTAAGG	GATTTTGGTC	
ATGAGATTAT	CAAAAAGGAT				
5641	CTTACCTAG	ATCCTTTTAA	ATTAAAAATG	AAGTTTTAAA	
TCAATCTAAA	GTATATATGA				
5701	GTAAACTTGG	TCTGACAGTT	AGAAAAACTC	ATCGAGCATC	
AAATGAAACT	GCAATTTATT				
5761	CATATCAGGA	TTATCAATAC	CATATTTTTG	AAAAAGCCGT	

## 038966

TTCTGTAATG AAGGAGAAAA  
 5821 CTCACCGAGG CAGTTCCATA GGATGGCAAG ATCCTGGTAT  
 CGGTCTGCGA TTCCGACTCG  
 5881 TCCAACATCA ATACAACCTA TTAATTTCCC CTCGTCAAAA  
 ATAAGGTTAT CAAGTGAGAA  
 5941 ATCACCATGA GTGACGACTG AATCCGGTGA GAATGGCAAC  
 AGCTTATGCA TTTCTTTCCA  
 6001 GACTTGTTCA ACAGGCCAGC CATTACGCTC GTCATCAAAA  
 TCACTCGCAT CAACCAAACC  
 6061 GTTATTCATT CGTGATTGCG CCTGAGCGAG ACGAAAATACG  
 CGATCGCTGT TAAAAGGACA  
 6121 ATTACAAACA GGAATCGAAT GCAACCGGCG CAGGAACACT  
 GCCAGCGCAT CAACAATATT  
 6181 TTCACCTGAA TCAGGATATT CTCTAATAC CTGGAATGCT  
 GTTTTTCCGG GGATCGCAGT  
 6241 GGTGAGTAAC CATGCATCAT CAGGAGTACG GATAAAATGC  
 TTGATGGTCG GAAGAGGCAT  
 6301 AAATCCGTC AGCCAGTTTA GTCTGACCAT CTCATCTGTA  
 ACATCATTGG CAACGCTACC  
 6361 TTTGCCATGT TTCAGAAACA ACTCTGGCGC ATCGGGCTTC  
 CCATACAATC GATAGATTGT  
 6421 CGCACCTGAT TGCCCGACAT TATCGCGAGC CCATTTATAC  
 CCATATAAAT CAGCATCCAT  
 6481 GTTGAATTT AATCGCGGCC TAGAGCAAGA CGTTTCCCGT  
 TGAATATGGC TCATAACACC  
 6541 CCTTGTATTA CTGTTTATGT AAGCAGACAG TTTTATTGTT  
 CATGATGATA TATTTTATC  
 6601 TTGTGCAATG TAACATCAGA GATTTTGAGA CACAACAATT  
 GGTCGACGGA TCC

## SEQ ID NO:11

1 GGTACCTCAA TATTGCCAT TAGCCATATT ATTCATTGGT TATATAGCAT  
 AAATCAATAT  
 61 TGGCTATTGG CCATTGCATA CGTTGTATCT ATATCATAAT ATGTACATTT  
 ATATTGGCTC  
 121 ATGTCCAATA TGACCGCCAT GTTGGCATTG ATTATTGACT AGTTATTAAT  
 AGTAATCAAT

038966

181 TACGGGGTCA TTAGTTCATA GCCCATATAT GGAGTTCGCG GTTACATAAC  
 TTACGGTAAA  
 241 TGGCCCCGCT GGCTGACCGC CCAACGACCC CCGCCCATTG ACGTCAATAA  
 TGACGTATGT  
 301 TCCCATAGTA ACGCCAATAG GGACTTTCCA TTGACGTCAA TGGGTGGAGT  
 APTTACGGTA  
 361 AACTGCCAC TTGGCAGTAC ATCAAGTGTA TCATATGCCA AGTCCGCCCC  
 CTATTGACGT  
 421 CAATGACGGT AAATGGCCCG CCTGGCATTG TGCCAGTAC ATGACCTTAC  
 GGGACTTTCC  
 481 TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC  
 GGTTTTGGCA  
 541 GTACACCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA TTTCCAAGTC  
 TCCACCCCAT  
 601 TGACGTCAAT GGGAGTTTGT TTTGGCACCA AAATCAACGG GACTTTCCAA  
 AATGTCGTAA  
 661 CAACTGCGAT CGCCGCCCC GTTGACGCAA ATGGGCGGTA GGCGTGTACG  
 GTGGGAGGTC  
 721 TATATAAGCA GAGCTCGCTG GCTTGTAAC TACTCTCTTA CTAGGAGACC  
 AGCTTGAGCC  
 781 TGGGTGTTTCG CTGGTTAGCC TAACCTGGTT GGCCACCAGG GGTAAGGACT  
 CCTTGGCTTA  
 841 GAAAGCTAAT AAACCTGCCT GCATTAGAGC TTATCTGAGT CAAGTGCCT  
 CATTGACGCC  
 901 TCACTCTCTT GAACGGGAAT CTTCCCTACT GGGTTCTCTC TCTGACCCAG  
 GCGAGAGAAA  
 961 CTCCAGCAGT GGCGCCCGAA CAGGGACTTG AGTGAGAGTG TAGGCACGTA  
 CAGCTGAGAA  
 1021 GGCGTCGGAC GCGAAGGAAG CGCGGGTGC GACGCGACCA  
 AGAAGGAGAC TTGGTGAGTA  
 1081 GGCTTCTCGA GTGCCGGGAA AAAGCTCGAG CCTAGTTAGA  
 GGACTAGGAG AGGCCGTAGC  
 1141 CGTAACTACT CTGGGCAAGT AGGGCAGGCG GTGGGTACGC  
 AATGGGGGCG GCTACCTCAG  
 1201 CACTAAATAG GAGACAATTA GACCAATTTG AGAAAATACG  
 ACTTCGCCCG AACGGAAAGA

## 038966

1261	AAAAGTACCA	AATTAACAT	TTAATATGGG	CAGGCAAGGA
GATGGAGCGC	TTCGGCCTCC			
1321	ATGAGAGGTT	GTTGGAGACA	GAGGAGGGGT	GTAAAAGAAT
CATAGAAGTC	CTCTACCCCC			
1381	TAGAACCAAC	AGGATCGGAG	GGCTTAAAAA	GTCTGTTCAA
TCTTGTGTGC	GTGCTATATT			
1441	GCTTGCACAA	GGAACAGAAA	GTGAAAGACA	CAGAGGAAGC
AGTAGCAACA	GTAAGACAAC			
1501	ACTGCCATCT	AGTGGAAAAA	GAAAAAAGTG	CAACAGAGAC
ATCTAGTGGA	CAAAGAAAAA			
1561	ATGACAAGGG	AATAGCAGCG	CCACCTGGTG	GCAGTCAGAA
TTTTCCAGCG	CAACAACAAG			
1621	GAAATGCCTG	GGTACATGTA	CCCTTGTAC	CGCGCACCTT
AAATGCGTGG	GTAAAAGCAG			
1681	TAGAGGAGAA	AAAATTGGA	GCAGAAATAG	TACCCATGTT
TCAAGCCCTA	TCGAATTCCC			
1741	GTTTGTGCTA	GGGTTCTTAG	GCTTCTGGG	GGCTGCTGGA
ACTGCAATGG	GAGCAGCGGC			
1801	GACAGCCCTG	ACGGTCCAGT	CTCAGCATTT	GCTTGTGGG
ATACTGCAGC	AGCAGAAGAA			
1861	TCTGTGGCG	GCTGTGGAGG	CTCAACAGCA	GATGTTGAAG
CTGACCATTT	GGGGTGTAA			
1921	AAACCTCAAT	GCCCCGCTCA	CAGCCCTTGA	GAAGTACCTA
GAGGATCAGG	CACGACTAAA			
1981	CTCCTGGGG	TGCGCATGGA	AACAAGTATG	TCATACCACA
GTGGAGTGGC	CCTGGACAAA			
2041	TCGGACTCCG	GATTGGCAAA	ATATGACTTG	GTTGGAGTGG
GAAAGACAAA	TAGCTGATTT			
2101	GGAAAGCAAC	ATTACGAGAC	AATTAGTGAA	GGCTAGAGAA
CAAGAGGAAA	AGAATCTAGA			
2161	TGCCTATCAG	AAGTTAACTA	GTTGGTCAGA	TTTCTGGTCT
TGGTTCGATT	TCTCAAAATG			
2221	GCTTAACATT	TTAAAAATGG	GATTTTTAGT	AATAGTAGGA
ATAATAGGGT	TAAGATTACT			
2281	TTACACAGTA	TATGGATGTA	TAGTGAGGGT	TAGGCAGGGA
TATGTTCCCTC	TATCTCCACA			



## 038966

2341	GATCCATATC	CGCGGCAATT	TTAAAAGAAA	GGGAGGAATA
	GGGGGACAGA	CTTCAGCAGA		
2401	GAGACTAATT	AATATAATAA	CAACACAATT	AGAAATACAA
	CATTTACAAA	CCAAAATTCA		
2461	AAAAATTTTA	AATTTTAGAG	CCGCGGAGAT	CTCAATATTG
	GCCATTAGCC	ATATTATTCA		
2521	TTGGTTATAT	AGCATAAATC	AATATTGGCT	ATTGGCCATT
	GCATACGTTG	TATCTATATC		
2581	ATAATATGTA	CATTTATATT	GGCTCATGTC	CAATATGACC
	GCCATGTTGG	CATTGATTAT		
2641	TGACTAGTTA	TTAATAGTAA	TCAATTACGG	GGTCATTAGT
	TCATAGCCCA	TATATGGAGT		
2701	TCCGCGTTAC	ATAACTTACG	GTAATGGCC	CGCCTGGCTG
	ACCGCCCAAC	GACCCCGCC		
2761	CATTGACGTC	AATAATGACG	TATGTTCCCA	TAGTAACGCC
	AATAGGGACT	TTCCATTGAC		
2821	GTCAATGGGT	GGAGTATTTA	CGGTAAACTG	CCCCTGGCC
	AGTACATCAA	GTGTATCATA		
2881	TGCCAAGTCC	GCCCCCTATT	GACGTCAATG	ACGGTAAATG
	GCCCCCTGG	CATTATGCC		
2941	AGTACATGAC	CTTACGGGAC	TTTCCTACTT	GGCAGTACAT
	CTACGTATTA	GTCATCGCTA		
3001	TTACCATGGT	GATGCGGTTT	TGGCAGTACA	CCAATGGGCG
	TGGATAGCGG	TTTACTCAC		
3061	GGGGATTTCC	AAGTCTCCAC	CCCATTGACG	TCAATGGGAG
	TTTGTTTTGG	CACCAAAATC		
3121	AACGGGACTT	TCCAAAATGT	CGTAATAACC	CCGCCCCGTT
	GACGCAAATG	GGCGGTAGGC		
3181	GTGTACGGTG	GGAGGTCTAT	ATAAGCAGAG	CTCGTTTAGT
	GAACCGTCAG	ATCACTAGAA		
3241	GCTTTATTGC	GGTAGTTTAT	CACAGTTAAA	TTGCTAACGC
	AGTCAGTGCT	TCTGACACAA		
3301	CAGTCTCGAA	CTTAAGCTGC	AGAAGTTGGT	CGTGAGGCAC
	TGGGCAGGCT	AGCCACCAAT		
3361	GCAGATTGAG	CTGAGCACCT	GCTTCTTCTT	GTGCCTGCTG
	AGGTTCTGCT	TCTCTGCCAC		

## 038966

3421	CAGGAGATAC	TACCTGGGGG	CTGTGGAGCT	GAGCTGGGAC
TACATGCAGT	CTGACCTGGG			
3481	GGAGCTGCCT	GTGGATGCCA	GGTTCCCCCC	CAGAGTGCCC
AAGAGCTTCC	CCTTCAACAC			
3541	CTCTGTGGTG	TACAAGAAGA	CCCTGTTTGT	GGAGTTCACT
GACCACCTGT	TCAACATTGC			
3601	CAAGCCCAGG	CCCCCTGGA	TGGGCCTGCT	GGGCCCCACC
ATCCAGGCTG	AGGTGTATGA			
3661	CACTGTGGTG	ATCACCCCTGA	AGAACATGGC	CAGCCACCCT
GTGAGCCTGC	ATGCTGTGGG			
3721	GGTGAGCTAC	TGGAAGGCCT	CTGAGGGGGC	TGAGTATGAT
GACCAGACCA	GCCAGAGGGA			
3781	GAAGGAGGAT	GACAAGGTGT	TCCCTGGGGG	CAGCCACACC
TATGTGTGGC	AGGTGCTGAA			
3841	GGAGAATGGC	CCCATGGCCT	CTGACCCCT	GTGCCTGACC
TACAGCTACC	TGAGCCATGT			
3901	GGACCTGGTG	AAGGACCTGA	ACTCTGGCCT	GATTGGGGCC
CTGCTGGTGT	GCAGGGAGGG			
3961	CAGCCTGGCC	AAGGAGAAGA	CCCAGACCCT	GCACAAGTTC
ATCCTGCTGT	TTGCTGTGTT			
4021	TGATGAGGGC	AAGAGCTGGC	ACTCTGAAAC	CAAGAACAGC
CTGATGCAGG	ACAGGGATGC			
4081	TGCCTCTGCC	AGGGCCTGGC	CCAAGATGCA	CACTGTGAAT
GGCTATGTGA	ACAGGAGCCT			
4141	GCCTGGCCTG	ATTGGCTGCC	ACAGGAAGTC	TGTGTACTGG
CATGTGATTG	GCATGGGCAC			
4201	CACCCCTGAG	GTGCACAGCA	TCTTCTGGA	GGGCCACACC
TTCCTGGTCA	GGAACCACAG			
4261	GCAGGCCAGC	CTGGAGATCA	GCCCCATCAC	CTTCTGACT
GCCCAGACCC	TGCTGATGGA			
4321	CCTGGGCCAG	TTCCTGCTGT	TCTGCCACAT	CAGCAGCCAC
CAGCATGATG	GCATGGAGGC			
4381	CTATGTGAAG	GTGGACAGCT	GCCCTGAGGA	GCCCCAGCTG
AGGATGAAGA	ACAATGAGGA			
4441	GGCTGAGGAC	TATGATGATG	ACCTGACTGA	CTCTGAGATG
GATGTGGTGA	GGTTTATGA			

## 038966

4501	TGACAACAGC	CCCAGCTTCA	TCCAGATCAG	GTCTGTGGCC
AAGAAGCACC	CCAAGACCTG			
4561	GGTGCCTACT	ATTGCTGCTG	AGGAGGAGGA	CTGGGACTAT
GCCCCCTGG	TGCTGGCCCC			
4621	TGATGACAGG	AGCTACAAGA	GCCAGTACCT	GAACAATGGC
CCCCAGAGGA	TTGGCAGGAA			
4681	GTACAAGAAG	GTCAGGTTCA	TGGCCTACAC	TGATGAAACC
TTCAAGACCA	GGGAGGCCAT			
4741	CCAGCATGAG	TCTGGCATCC	TGGGCCCCCT	GCTGTATGGG
GAGGTGGGGG	ACACCCTGCT			
4801	GATCATCTTC	AAGAACCAGG	CCAGCAGGCC	CTACAACATC
TACCCCATG	GCATCACTGA			
4861	TGTGAGGCC	CTGTACAGCA	GGAGGCTGCC	CAAGGGGGTG
AAGCACCTGA	AGGACTTCCC			
4921	CATCCTGCCT	GGGAGATCT	TCAAGTACAA	GTGGACTGTG
ACTGTGGAGG	ATGGCCCCAC			
4981	CAAGTCTGAC	CCCAGGTGCC	TGACCAGATA	CTACAGCAGC
TTTGTGAACA	TGGAGAGGGA			
5041	CCTGGCCTCT	GGCCTGATTG	GCCCCCTGCT	GATCTGCTAC
AAGGAGTCTG	TGGACCAGAG			
5101	GGGCAACCAG	ATCATGTCTG	ACAAGAGGAA	TGTGATCCTG
TTCTCTGTGT	TTGATGAGAA			
5161	CAGGAGCTGG	TACCTGACTG	AGAACATCCA	GAGGTTCCCTG
CCCAACCCTG	CTGGGGTGCA			
5221	GCTGGAGGAC	CCTGAGTTCC	AGGCCAGCAA	CATCATGCAC
AGCATCAATG	GCTATGTGTT			
5281	TGACAGCCTG	CAGCTGTCTG	TGTGCCTGCA	TGAGGTGGCC
TACTGGTACA	TCCTGAGCAT			
5341	TGGGGCCCAG	ACTGACTTCC	TGTCTGTGTT	CTTCTCTGGC
TACACCTTCA	AGCACAAAGAT			
5401	GGTGTATGAG	GACACCCTGA	CCCTGTTCCT	CTTCTCTGGG
GAGACTGTGT	TCATGAGCAT			
5461	GGAGAACCCT	GGCCTGTGGA	TTCTGGGCTG	CCACAACCTCT
GACTTCAGGA	ACAGGGGCAT			
5521	GACTGCCCTG	CTGAAAGTCT	CCAGCTGTGA	CAAGAACACT
GGGGACTACT	ATGAGGACAG			

## 038966

5581	CTATGAGGAC	ATCTCTGCCT	ACCTGCTGAG	CAAGAACAAT
	GCCATTGAGC	CCAGGAGCTT		
5641	CAGCCAGAAT	GCCACTAATG	TGTCTAACAA	CAGCAACACC
	AGCAATGACA	GCAATGTGTC		
5701	TCCCCAGTG	CTGAAGAGGC	ACCAGAGGGA	GATCACCAGG
	ACCACCCTGC	AGTCTGACCA		
5761	GGAGGAGATT	GACTATGATG	ACACCATCTC	TGTGGAGATG
	AAGAAGGAGG	ACTTTGACAT		
5821	CTACGACGAG	GACGAGAACC	AGAGCCCCAG	GAGCTTCCAG
	AAGAAGACCA	GGCACTACTT		
5881	CATTGCTGCT	GTGGAGAGGC	TGTGGGACTA	TGGCATGAGC
	AGCAGCCCCC	ATGTGCTGAG		
5941	GAACAGGGCC	CAGTCTGGCT	CTGTGCCCCA	GTTCAAGAAG
	GTGGTGTTC	AGGAGTTCAC		
6001	TGATGGCAGC	TTCACCCAGC	CCCTGTACAG	AGGGGAGCTG
	AATGAGCACC	TGGGCCTGCT		
6061	GGGCCCTTAC	ATCAGGGCTG	AGGTGGAGGA	CAACATCATG
	GTGACCTTCA	GGAACCAGGC		
6121	CAGCAGGCC	TACAGCTTCT	ACAGCAGCCT	GATCAGCTAT
	GAGGAGGACC	AGAGGCAGGG		
6181	GGCTGAGCCC	AGGAAGAACT	TTGTGAAGCC	CAATGAAACC
	AAGACCTACT	TCTGGAAGGT		
6241	GCAGCACCAC	ATGGCCCCCA	CCAAGGATGA	GTTTGACTGC
	AAGGCCTGGG	CCTACTTCTC		
6301	TGATGTGGAC	CTGGAGAAGG	ATGTGCACTC	TGGCCTGATT
	GGCCCCCTGC	TGGTGTGCCA		
6361	CACCAACACC	CTGAACCCTG	CCCATGGCAG	GCAGGTGACT
	GTGCAGGAGT	TTGCCCTGTT		
6421	CTTACCATC	TTTGATGAAA	CCAAGAGCTG	GTACTTCACT
	GAGAACATGG	AGAGGAACTG		
6481	CAGGGCCCCC	TGCAACATCC	AGATGGAGGA	CCCCACCTTC
	AAGGAGAACT	ACAGGTTCCTA		
6541	TGCCATCAAT	GGCTACATCA	TGGACACCCT	GCCTGGCCTG
	GTGATGGCCC	AGGACCAGAG		
6601	GATCAGGTGG	TACCTGCTGA	GCATGGGCAG	CAATGAGAAC
	ATCCACAGCA	TCCACTTCTC		

## 038966

6661	TGGCCATGTG	TTCACTGTGA	GGAAGAAGGA	GGAGTACAAG
ATGGCCCTGT	ACAACCTGTA			
6721	CCCTGGGGTG	TTTGAGACTG	TGGAGATGCT	GCCCAGCAAG
GCTGGCATCT	GGAGGTGGA			
6781	GTGCCTGATT	GGGAGCACC	TGCATGCTGG	CATGAGCACC
CTGTTCTTGG	TGTACAGCAA			
6841	CAAGTGCCAG	ACCCCCCTGG	GCATGGCCTC	TGGCCACATC
AGGGACTTCC	AGATCACTGC			
6901	CTCTGGCCAG	TATGGCCAGT	GGGCCCCCAA	GCTGGCCAGG
CTGCACTACT	CTGGCAGCAT			
6961	CAATGCCTGG	AGCACCAAGG	AGCCCTTCAG	CTGGATCAAG
GTGGACCTGC	TGGCCCCCAT			
7021	GATCATCCAT	GGCATCAAGA	CCCAGGGGGC	CAGGCAGAAG
TTCAGCAGCC	TGTACATCAG			
7081	CCAGTTCATC	ATCATGTACA	GCCTGGATGG	CAAGAAGTGG
CAGACCTACA	GGGGCAACAG			
7141	CACTGGCACC	CTGATGGTGT	TCTTTGGCAA	TGTGGACAGC
TCTGGCATCA	AGCACAAACAT			
7201	CTTCAACCCC	CCCATCATTG	CCAGATACAT	CAGGCTGCAC
CCCACCCACT	ACAGCATCAG			
7261	GAGCACCCCTG	AGGATGGAGC	TGATGGGCTG	TGACCTGAAC
AGCTGCAGCA	TGCCCCTGGG			
7321	CATGGAGAGC	AAGGCCATCT	CTGATGCCCA	GATCACTGCC
AGCAGCTACT	TCACCAACAT			
7381	GTTTGCCACC	TGGAGCCCCA	GCAAGGCCAG	GCTGCACCTG
CAGGGCAGGA	GCAATGCCTG			
7441	GAGGCCCCAG	GTCAACAACC	CCAAGGAGTG	GCTGCAGGTG
GACTTCCAGA	AGACCATGAA			
7501	GGTGACTGGG	GTGACCACCC	AGGGGGTGAA	GAGCCTGCTG
ACCAGCATGT	ATGTGAAGGA			
7561	GTTCTGATC	AGCAGCAGCC	AGGATGGCCA	CCAGTGGACC
CTGTTCTTCC	AGAATGGCAA			
7621	GGTGAAGGTG	TTCCAGGGCA	ACCAGGACAG	CTTCACCCCT
GTGGTGAACA	GCCTGGACCC			
7681	CCCCCTGCTG	ACCAGATACC	TGAGATTCA	CCCCCAGAGC
TGGGTGCACC	AGATTGCCCT			

## 038966

7741	GAGGATGGAG	GTGCTGGGCT	GTGAGGCCCA	GGACCTGTAC
	TGAGCGGCCG	CGGGCCCAAT		
7801	CAACCTCTGG	ATTACAAAAT	TTGTGAAAGA	TTGACTGGTA
	TTCTTAACTA	TGTTGCTCCT		
7861	TTTACGCTAT	GTGGATACGC	TGCTTTAATG	CCTTTGTATC
	ATGCTATTGC	TTCCCGTATG		
7921	GCTTTCATTT	TCTCCTCCTT	GTATAAATCC	TGGTTGCTGT
	CTCTTTATGA	GGAGTTGTGG		
7981	CCCGTTGTCA	GGCAACGTGG	CGTGGTGTGC	ACTGTGTTTG
	CTGACGCAAC	CCCCACTGGT		
8041	TGGGGCATTG	CCACCACCTG	TCAGCTCCTT	TCCGGGACTT
	TCGCTTTCCC	CCTCCCTATT		
8101	GCCACGGCGG	AACTCATCGC	CGCCTGCCTT	GCCCGCTGCT
	GGACAGGGGC	TCGGCTGTTG		
8161	GGCACTGACA	ATTCCGTGGT	GTTGTCGGGG	AAATCATCGT
	CCTTTCCTTG	GCTGCTCGCC		
8221	TGTGTTGCCA	CCTGGATTCT	GCGCGGGACG	TCCTTCTGCT
	ACGTCCCTTC	GGCCCTCAAT		
8281	CCAGCGGACC	TTCCTTCCCG	CGGCCTGCTG	CCGGCTCTGC
	GGCCTCTTCC	GCGTCTTCGC		
8341	CTTCGCCCTC	AGACGAGTCG	GATCTCCCTT	TGGGCCGCCT
	CCCCGCAAGC	TTCGCACTTT		
8401	TTAAAAGAAA	AGGGAGGACT	GGATGGGATT	TATTACTCCG
	ATAGGACGCT	GGCTTGTAAC		
8461	TCAGTCTCTT	ACTAGGAGAC	CAGCTTGAGC	CTGGGTGTTT
	GCTGGTTAGC	CTAACCTGGT		
8521	TGGCCACCAG	GGGTAAGGAC	TCCTTGGCTT	AGAAAAGCTAA
	TAAACTTGCC	TGCATTAGAG		
8581	CTCTTACGCG	TCCCAGGCTC	GAGATCCGCA	TCTCAATTAG
	TCAGCAACCA	TAGTCCCGCC		
8641	CCTAACTCCG	CCCATCCCGC	CCCTAACTCC	GCCCAGTTCC
	GCCCATTCTC	CGCCCCATGG		
8701	CTGACTAATT	TTTTTTATTT	ATGCAGAGGC	CGAGGCCGCC
	TCGGCCTCTG	AGCTATTCCA		
8761	GAAGTAGTGA	GGAGGCTTTT	TTGGAGGCCT	AGGCTTTTGC
	AAAAAGCTAA	CTTGTTTATT		

## 038966

8821	GCAGCTTATA	ATGGTTACAA	ATAAAGCAAT	AGCATCACAA
	ATTTACAAA TAAAGCATT			
8881	TTTCACTGC	ATTCTAGTTG	TGGTTTGTCC	AAACTCATCA
	ATGTATCTTA TCATGTCTGT			
8941	CCGCTTCCTC	GCTCACTGAC	TCGCTGCGCT	CGGTCGTTCCG
	GCTGCGGCGA GCGGTATCAG			
9001	CTCACTCAAA	GGCGGTAATA	CGGTTATCCA	CAGAATCAGG
	GGATAACGCA GGAAAGAACA			
9061	TGTGAGCAAA	AGGCCAGCAA	AAGGCCAGGA	ACCGTAAAAA
	GGCCGCGTTG CTGGCGTTTT			
9121	TCCATAGGCT	CGCCCCCCT	GACGAGCATC	ACAAAAATCG
	ACGCTCAAGT CAGAGGTGGC			
9181	GAAACCCGAC	AGGACTATAA	AGATACCAGG	CGTTTCCCCC
	TGGAAGCTCC CTCGTGCGCT			
9241	CTCCTGTTCC	GACCCTGCCG	CTTACCGGAT	ACCTGTCCGC
	CTTCTCCCT TCGGGAAGCG			
9301	TGGCGCTTTC	TCATAGCTCA	CGCTGTAGGT	ATCTCAGTTC
	GGTGTAGGTC GTTCGCTCCA			
9361	AGCTGGGCTG	TGTGCACGAA	CCCCCGTTC	AGCCCGACCG
	CTGCGCCTTA TCCGTAAC			
9421	ATCGTCTTGA	GTCCAACCCG	GTAAGACACG	ACTTATCGCC
	ACTGGCAGCA GCCACTGGTA			
9481	ACAGGATTAG	CAGAGCGAGG	TATGTAGGCG	GTGCTACAGA
	GTTCTTGAAG TGGTGGCCTA			
9541	ACTACGGCTA	CACTAGAAGA	ACAGTATTTG	GTATCTGCGC
	TCTGCTGAAG CCAGTTACCT			
9601	TCGGAAAAAG	AGTTGGTAGC	TCTTGATCCG	GCAAACAAC
	CACCGCTGGT AGCGGTGGTT			
9661	TTTTTGTTTG	CAAGCAGCAG	ATTACGCGCA	GAAAAAAGG
	ATCTCAAGAA GATCCTTTGA			
9721	TCTTTTCTAC	GGGGTCTGAC	GCTCAGTGG	ACGAAAACTC
	ACGTTAAGGG ATTTTGGTCA			
9781	TGAGATTATC	AAAAAGGATC	TTCACCTAGA	TCCTTTTAAA
	TTAAAAATGA AGTTTTAAAT			
9841	CAATCTAAAG	TATATATGAG	TAAACTTGGT	CTGACAGTTA
	GAAAAACTCA TCGAGCATCA			

038966

9901 AATGAACTG CAATTTATTC ATATCAGGAT TATCAATACC  
 ATATTTTTGA AAAAGCCGTT  
 9961 TCTGTAATGA AGGAGAAAAC TCACCGAGGC AGTTCCATAG  
 GATGGCAAGA TCCTGGTATC  
 10021 GGTCTGCGAT TCCGACTCGT CCAACATCAA TACAACCTAT  
 TAATTTCCCC TCGTCAAAAA  
 10081 TAAGGTTATC AAGTGAGAAA TCACCATGAG TGACGACTGA  
 ATCCGGTGAG AATGGCAACA  
 10141 GCTTATGCAT TTCTTTCCAG ACTTGTTCAA CAGGCCAGCC  
 ATTACGCTCG TCATCAAAAT  
 10201 CACTCGCATC AACCAAACCG TTATTCATTC GTGATTGCGC  
 CTGAGCGAGA CGAAATACGC  
 10261 GATCGCTGTT AAAAGGACAA TTACAAACAG GAATCGAATG  
 CAACCGGCGC AGGAACACTG  
 10321 CCAGCGCATC AACAATATTT TCACCTGAAT CAGGATATTC  
 TTCTAATACC TGGAAATGCTG  
 10381 TTTTCCGGG GATCGCAGTG GTGAGTAACC ATGCATCATC  
 AGGAGTACGG ATAAAATGCT  
 10441 TGATGGTCGG AAGAGGCATA AATTCGTCA GCCAGTTTAG  
 TCTGACCATC TCATCTGTAA  
 10501 CATCATTGGC AACGCTACCT TTGCCATGTT TCAGAAACAA  
 CTCTGGCGCA TCGGGCTTCC  
 10561 CATACAATCG ATAGATTGTC GCACCTGATT GCCCGACATT  
 ATCGCGAGCC CATTTATACC  
 10621 CATATAAATC AGCATCCATG TTGGAATTTA ATCGCGGCCT  
 AGAGCAAGAC GTTTCCCGTT  
 10681 GAATATGGCT CATAACACCC CTTGTATTAC TGTTTATGTA  
 AGCAGACAGT TTTATTGTTT  
 10741 ATGATGATAT ATTTTTATCT TGTGCAATGT AACATCAGAG  
 ATTTTGAGAC ACAACAATTG  
 10801 GTCGACGGAT CC  
**SEQ ID NO:12**  
 1 GGTACCTCAA TATTGGCCAT TAGCCATATT ATTCATTGGT TATATAGCAT  
 AAATCAATAT  
 61 TGGCTATTGG CCATTGCATA CGTTGTATCT ATATCATAAT ATGTACATTT  
 ATATTGGCTC



038966

121 ATGTCCAATA TGACCGCCAT GTTGGCATTG ATTATTGACT AGTTATTAAT  
 AGTAATCAAT  
 181 TACGGGGTCA TTAGTTCATA GCCCATATAT GGAGTTCGC GTTACATAAC  
 TTACGGTAAA  
 241 TGGCCCGCCT GGCTGACCGC CCAACGACCC CCGCCCATG ACGTCAATAA  
 TGACGTATGT  
 301 TCCCATAGTA ACGCCAATAG GGACTTTCCA TTGACGTCAA TGGGTGGAGT  
 ATTTACGGTA  
 361 AACTGCCCAC TTGGCAGTAC ATCAAGTGTA TCATATGCCA AGTCCGCCCC  
 CTATTGACGT  
 421 CAATGACGGT AAATGGCCCG CCTGGCATTG TGCCAGTAC ATGACCTTAC  
 GGGACTTTCC  
 481 TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC  
 GGTTTTGGCA  
 541 GTACACCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA TTTCCAAGTC  
 TCCACCCCAT  
 601 TGACGTCAAT GGGAGTTTGT TTTGGCACCA AAATCAACGG GACTTTCCAA  
 AATGTCGTAA  
 661 CAACTGCGAT CGCCCGCCCC GTTGACGCAA ATGGGCGGTA GGCCTGTACG  
 GTGGGAGGTC  
 721 TATATAAGCA GAGCTCGCTG GCTTGTA ACT CAGTCTCTTA CTAGGAGACC  
 AGCTTGAGCC  
 781 TGGGTGTTTCG CTGGTTAGCC TAACCTGGTT GGCCACCAGG GGTAAGGACT  
 CCTTGGCTTA  
 841 GAAAGCTAAT AAAGTTGCCT GCATTAGAGC TTATCTGAGT CAAGTGTCCCT  
 CATTGACGCC  
 901 TCACTCTCTT GAACGGGAAT CTTCCCTACT GGGTCTCTC TCTGACCCAG  
 GCGAGAGAAA  
 961 CTCCAGCAGT GGCGCCCGAA CAGGGACTTG AGTGAGAGTG TAGGCACGTA  
 CAGCTGAGAA  
 1021 GCGTCCGGAC GCGAAGGAAG CGCGGGGTGC GACGCGACCA  
 AGAAGGAGAC TTGGTGAGTA  
 1081 GGCTTCTCGA GTGCCGGGAA AAAGCTCGAG CCTAGTTAGA  
 GGACTAGGAG AGGCCGTAGC  
 1141 CGTAACTACT CTGGGCAAGT AGGGCAGGCG GTGGGTACGC  
 AATGGGGCG GCTACCTCAG

## 038966

1201	CACTAAATAG	GAGACAATTA	GACCAATTTG	AGAAAATACG
	ACTTCGCCCC AACGGAAAGA			
1261	AAAAGTACCA	AATTAACAT	TTAATATGGG	CAGGCAAGGA
	GATGGAGCGC TTCGGCCTCC			
1321	ATGAGAGGTT	GTTGGAGACA	GAGGAGGGGT	GTAAAAGAAT
	CATAGAAGTC CTCTACCCC			
1381	TAGAACCAAC	AGGATCGGAG	GGCTTAAAAA	GTCTGTTCAA
	TCTTGTGTGC GTGCTATATT			
1441	GCTTGCACAA	GGAACAGAAA	GTGAAAGACA	CAGAGGAAGC
	AGTAGCAACA GTAAGACAAC			
1501	ACTGCCATCT	AGTGGAAAAA	GAAAAAAGTG	CAACAGAGAC
	ATCTAGTGGG CAAAAGAAAA			
1561	ATGACAAGGG	AATAGCAGCG	CCACCTGGTG	GCAGTCAGAA
	TTTTCCAGCG CAACAACAAG			
1621	GAAATGCCTG	GGTACATGTA	CCCTTGTAC	CGCGCACCTT
	AAATGCGTGG GTAAAAGCAG			
1681	TAGAGGAGAA	AAAATTTGGA	GCAGAAATAG	TACCCATGTT
	TCAAGCCCTA TCGAATTCCC			
1741	GTTTGTGCTA	GGGTCTTAG	GCTTCTGGG	GGCTGCTGGA
	ACTGCAATGG GAGCAGCGGC			
1801	GACAGCCCTG	ACGGTCCAGT	CTCAGCATT	GCTTGCTGGG
	ATACTGCAGC AGCAGAAGAA			
1861	TCTGCTGGCG	GCTGTGGAGG	CTCAACAGCA	GATGTTGAAG
	CTGACCATT	GGGGTGTAA		
1921	AAACCTCAAT	GCCCCGCTCA	CAGCCCTTGA	GAAGTACCTA
	GAGGATCAGG CACGACTAAA			
1981	CTCCTGGGGG	TGCGCATGGA	AACAAGTATG	TCATACCACA
	GTGGAGTGGC CCTGGACAAA			
2041	TCGGACTCCG	GATTGGCAAA	ATATGACTTG	GTTGGAGTGG
	GAAAGACAAA TAGCTGATTT			
2101	GGAAAGCAAC	ATTACGAGAC	AATTAGTGAA	GGCTAGAGAA
	CAAGAGGAAA AGAATCTAGA			
2161	TGCCTATCAG	AAGTAACTA	GTTGGTCAGA	TTTCTGGTCT
	TGGTTCGATT TCTCAAATG			
2221	GCTTAACATT	TTAAAAATGG	GATTTT	AGT
	ATAATAGGGT TAAGATTACT			

## 038966

2281	TTACACAGTA	TATGGATGTA	TAGTGAGGGT	TAGGCAGGGA
	TATGTTCCCTC	TATCTCCACA		
2341	GATCCATATC	CGCGGCAATT	TTAAAAGAAA	GGGAGGAATA
	GGGGGACAGA	CTTCAGCAGA		
2401	GAGACTAATT	AATATAATAA	CAACACAATT	AGAAATACAA
	CATTTACAAA	CCAAAATTCA		
2461	AAAAATTTTA	AATTTTAGAG	CCGCGGAGAT	CTGTTACATA
	ACTTATGGTA	AATGGCCTGC		
2521	CTGGCTGACT	GCCCAATGAC	CCCTGCCCAA	TGATGTCAAT
	AATGATGTAT	GTTCCCATGT		
2581	AATGCCAATA	GGGACTTTCC	ATTGATGTCA	ATGGGTGGAG
	TATTTATGGT	AAGTCCAC		
2641	TTGGCAGTAC	ATCAAGTGTA	TCATATGCCA	AGTATGCCCC
	CTATTGATGT	CAATGATGGT		
2701	AAATGGCCTG	CCTGGCATT	TGCCCAGTAC	ATGACCTTAT
	GGGACTTTCC	TACTTGGCAG		
2761	TACATCTATG	TATTAGTCAT	TGCTATTACC	ATGGGAATTC
	ACTAGTGGAG	AAGAGCATGC		
2821	TTGAGGGCTG	AGTGCCCCTC	AGTGGCAGA	GAGCACATGG
	CCCACAGTCC	CTGAGAAGTT		
2881	GGGGGAGGG	GTGGGCAATT	GAACTGGTGC	CTAGAGAAGG
	TGGGGCTTGG	GTAAACTGGG		
2941	AAAGTGATGT	GGTGTACTGG	CTCCACCTTT	TTCCCCAGGG
	TGGGGGAGAA	CCATATATAA		
3001	GTGCAGTAGT	CTCTGTGAAC	ATTCAAGCTT	CTGCCTTCTC
	CCTCCTGTGA	GTTTGCTAGC		
3061	CACCAATGCA	GATTGAGCTG	AGCACCTGCT	TCTTCCTGTG
	CCTGCTGAGG	TTCTGCTTCT		
3121	CTGCCACCAG	GAGATACTAC	CTGGGGGCTG	TGGAGCTGAG
	CTGGGACTAC	ATGCAGTCTG		
3181	ACCTGGGGGA	GCTGCCTGTG	GATGCCAGGT	TCCCCCCCAG
	AGTGCCCAAG	AGCTTCCCCT		
3241	TCAACACCTC	TGTGGTGTAC	AAGAAGACCC	TGTTTGTGGA
	GTTCACTGAC	CACCTGTTCA		
3301	ACATTGCCAA	GCCCAGGCC	CCCTGGATGG	GCCTGTGGG
	CCCCACCATC	CAGGCTGAGG		

## 038966

3361	TGTATGACAC	TGTGGTGATC	ACCCTGAAGA	ACATGGCCAG
	CCACCCGTGTG	AGCCTGCATG		
3421	CTGTGGGGGT	GAGCTACTGG	AAGGCCTCTG	AGGGGGCTGA
	GTATGATGAC	CAGACCAGCC		
3481	AGAGGGAGAA	GGAGGATGAC	AAGGTGTTCC	CTGGGGGCAG
	CCACACCTAT	GTGTGGCAGG		
3541	TGCTGAAGGA	GAATGGCCCC	ATGGCCTCTG	ACCCCTGTG
	CCTGACCTAC	AGCTACCTGA		
3601	GCCATGTGGA	CCTGGTGAAG	GACCTGAACT	CTGGCCTGAT
	TGGGGCCCTG	CTGGTGTGCA		
3661	GGGAGGGCAG	CCTGGCCAAG	GAGAAGACCC	AGACCCTGCA
	CAAGTTCATC	CTGCTGTTTG		
3721	CTGTGTTTGA	TGAGGGCAAG	AGCTGGCACT	CTGAAACCAA
	GAACAGCCTG	ATGCAGGACA		
3781	GGGATGTGTC	CTCTGCCAGG	GCCTGGCCCA	AGATGCACAC
	TGTGAATGGC	TATGTGAACA		
3841	GGAGCCTGCC	TGGCCTGATT	GGCTGCCACA	GGAAGTCTGT
	GTACTGGCAT	GTGATTGGCA		
3901	TGGGCACCAC	CCCTGAGGTG	CACAGCATCT	TCCTGGAGGG
	CCACACCTTC	CTGGTCAGGA		
3961	ACCACAGGCA	GGCCAGCCTG	GAGATCAGCC	CCATCACCTT
	CCTGACTGCC	CAGACCCTGC		
4021	TGATGGACCT	GGGCCAGTTC	CTGCTGTTCT	GCCACATCAG
	CAGCCACCAG	CATGATGGCA		
4081	TGGAGGCCTA	TGTGAAGGTG	GACAGCTGCC	CTGAGGAGCC
	CCAGCTGAGG	ATGAAGAACA		
4141	ATGAGGAGGC	TGAGGACTAT	GATGATGACC	TGACTGACTC
	TGAGATGGAT	GTGGTGAGGT		
4201	TTGATGATGA	CAACAGCCCC	AGCTTCATCC	AGATCAGGTC
	TGTGGCCAAG	AAGCACCCCA		
4261	AGACCTGGGT	GCACTACATT	GCTGCTGAGG	AGGAGGACTG
	GGACTATGCC	CCCCTGGTGC		
4321	TGGCCCCTGA	TGACAGGAGC	TACAAGAGCC	AGTACCTGAA
	CAATGGCCCC	CAGAGGATTG		
4381	GCAGGAAGTA	CAAGAAGGTC	AGGTTCATGG	CCTACACTGA
	TGAAACCTTC	AAGACCAGGG		

## 038966

4441	AGGCCATCCA	GCATGAGTCT	GGCATCCTGG	GCCCCCTGCT
	GTATGGGGAG	GTGGGGGACA		
4501	CCCTGCTGAT	CATCTTCAAG	AACCAGGCCA	GCAGGCCCTA
	CAACATCTAC	CCCCATGGCA		
4561	TCACTGATGT	GAGGCCCTG	TACAGCAGGA	GGCTGCCCAA
	GGGGTGAAG	CACCTGAAGG		
4621	ACTTCCCCAT	CCTGCCTGGG	GAGATCTTCA	AGTACAAGTG
	GACTGTGACT	GTGGAGGATG		
4681	GCCCCACCAA	GTCTGACCCC	AGGTGCCTGA	CCAGATACTA
	CAGCAGCTTT	GTGAACATGG		
4741	AGAGGGACCT	GGCCTCTGGC	CTGATTGGCC	CCCTGCTGAT
	CTGCTACAAG	GAGTCTGTGG		
4801	ACCAGAGGGG	CAACCAGATC	ATGTCTGACA	AGAGGAATGT
	GATCCTGTTC	TCTGTGTTTG		
4861	ATGAGAACAG	GAGCTGGTAC	CTGACTGAGA	ACATCCAGAG
	GTTCCTGCCC	AACCCTGCTG		
4921	GGGTGCAGCT	GGAGGACCCT	GAGTTCAGG	CCAGCAACAT
	CATGCACAGC	ATCAATGGCT		
4981	ATGTGTTTGA	CAGCCTGCAG	CTGTCTGTGT	GCCTGCATGA
	GGTGGCCTAC	TGGTACATCC		
5041	TGAGCATTGG	GGCCCAGACT	GACTTCCTGT	CTGTGTTCTT
	CTCTGGCTAC	ACCTTCAAGC		
5101	ACAAGATGGT	GTATGAGGAC	ACCCTGACCC	TGTTCCCCTT
	CTCTGGGGAG	ACTGTGTTCA		
5161	TGAGCATGGA	GAACCCTGGC	CTGTGGATTC	TGGGCTGCCA
	CAACTCTGAC	TTCAGGAACA		
5221	GGGGCATGAC	TGCCCTGCTG	AAAGTCTCCA	GCTGTGACAA
	GAACACTGGG	GACTACTATG		
5281	AGGACAGCTA	TGAGGACATC	TCTGCCTACC	TGCTGAGCAA
	GAACAATGCC	ATTGAGCCCA		
5341	GGAGCTTCAG	CCAGAATGCC	ACTAATGTGT	CTAACAACAG
	CAACACCAGC	AATGACAGCA		
5401	ATGTGTCTCC	CCCAGTGCTG	AAGAGGCACC	AGAGGGAGAT
	CACCAGGACC	ACCCTGCAGT		
5461	CTGACCAGGA	GGAGATTGAC	TATGATGACA	CCATCTCTGT
	GGAGATGAAG	AAGGAGGACT		

## 038966

5521	TTGACATCTA	CGACGAGGAC	GAGAACCAGA	GCCCCAGGAG
CTTCCAGAAG	AAGACCAGGC			
5581	ACTACTTCAT	TGCTGCTGTG	GAGAGGCTGT	GGGACTATGG
CATGAGCAGC	AGCCCCCATG			
5641	TGCTGAGGAA	CAGGGCCCAG	TCTGGCTCTG	TGCCCCAGTT
CAAGAAGGTG	GTGTTCCAGG			
5701	AGTTCACTGA	TGGCAGCTTC	ACCCAGCCCC	TGTACAGAGG
GGAGCTGAAT	GAGCACCTGG			
5761	GCCTGCTGGG	CCCCTACATC	AGGGCTGAGG	TGGAGGACAA
CATCATGGTG	ACCTTCAGGA			
5821	ACCAGGCCAG	CAGGCCCTAC	AGCTTCTACA	GCAGCCTGAT
CAGCTATGAG	GAGGACCAGA			
5881	GGCAGGGGGC	TGAGCCCAGG	AAGAACTTTG	TGAAGCCCAA
TGAAACCAAG	ACCTACTTCT			
5941	GGAAGGTGCA	GCACCACATG	GCCCCACCA	AGGATGAGTT
TGACTGCAAG	GCCTGGGCCT			
6001	ACTTCTCTGA	TGTGGACCTG	GAGAAGGATG	TGCACTCTGG
CCTGATTGGC	CCCCTGCTGG			
6061	TGTGCCACAC	CAACACCCTG	AACCCTGCC	ATGGCAGGCA
GGTGACTGTG	CAGGAGTTTG			
6121	CCCTGTTCTT	CACCATCTTT	GATGAAACCA	AGAGCTGGTA
CTTCACTGAG	AACATGGAGA			
6181	GGAAGTGCAG	GGCCCCCTGC	AACATCCAGA	TGGAGGACCC
CACCTTCAAG	GAGAACTACA			
6241	GGTTCCATGC	CATCAATGGC	TACATCATGG	ACACCCTGCC
TGGCCTGGTG	ATGGCCCAGG			
6301	ACCAGAGGAT	CAGGTGGTAC	CTGCTGAGCA	TGGGCAGCAA
TGAGAACATC	CACAGCATCC			
6361	ACTTCTCTGG	CCATGTGTTC	ACTGTGAGGA	AGAAGGAGGA
GTACAAGATG	GCCCTGTACA			
6421	ACCTGTACCC	TGGGGTGTTT	GAGACTGTGG	AGATGTGCC
CAGCAAGGCT	GGCATCTGGA			
6481	GGGTGGAGTG	CCTGATTGGG	GAGCACCTGC	ATGCTGGCAT
GAGCACCCCTG	TTCCTGGTGT			
6541	ACAGCAACAA	GTGCCAGACC	CCCCTGGGCA	TGGCCTCTGG
CCACATCAGG	GACTTCCAGA			

## 038966

6601	TCACTGCCTC	TGGCCAGTAT	GGCCAGTGGG	CCCCCAAGCT
	GGCCAGGCTG	CACTACTCTG		
6661	GCAGCATCAA	TGCCTGGAGC	ACCAAGGAGC	CCTTCAGCTG
	GATCAAGGTG	GACCTGCTGG		
6721	CCCCATGAT	CATCCATGGC	ATCAAGACCC	AGGGGGCCAG
	GCAGAAGTTC	AGCAGCCTGT		
6781	ACATCAGCCA	GTTCATCATC	ATGTACAGCC	TGGATGGCAA
	GAAAGTGGCAG	ACCTACAGGG		
6841	GCAACAGCAC	TGGCACCCCTG	ATGGTGTTC	TTGGCAATGT
	GGACAGCTCT	GGCATCAAGC		
6901	ACAACATCTT	CAACCCCCC	ATCATTGCCA	GATACATCAG
	GCTGCACCCC	ACCCACTACA		
6961	GCATCAGGAG	CACCCTGAGG	ATGGAGCTGA	TGGGCTGTGA
	CCTGAACAGC	TGCAGCATGC		
7021	CCCTGGGCAT	GGAGAGCAAG	GCCATCTCTG	ATGCCAGAT
	CACTGCCAGC	AGCTACTTCA		
7081	CCAACATGTT	TGCCACCTGG	AGCCCCAGCA	AGGCCAGGCT
	GCACCTGCAG	GGCAGGAGCA		
7141	ATGCCTGGAG	GCCCCAGGTC	AACAACCCCA	AGGAGTGGCT
	GCAGGTGGAC	TTCCAGAAGA		
7201	CCATGAAGGT	GACTGGGGTG	ACCACCCAGG	GGGTGAAGAG
	CCTGCTGACC	AGCATGTATG		
7261	TGAAGGAGTT	CCTGATCAGC	AGCAGCCAGG	ATGGCCACCA
	GTGGACCCTG	TTCTTCCAGA		
7321	ATGGCAAGGT	GAAGGTGTTT	CAGGGCAACC	AGGACAGCTT
	CACCCCTGTG	GTGAACAGCC		
7381	TGGACCCCC	CCTGCTGACC	AGATACCTGA	GGATTCACCC
	CCAGAGCTGG	GTGCACCAGA		
7441	TTGCCCTGAG	GATGGAGGTG	CTGGGCTGTG	AGGCCAGGA
	CCTGTACTGA	GCGGCCGCGG		
7501	GCCCAATCAA	CCTCTGGATT	ACAAAATTTG	TGAAAGATTG
	ACTGGTATTC	TTAACATATG		
7561	TGCTCCTTTT	ACGCTATGTG	GATACGCTGC	TTTAATGCCT
	TTGTATCATG	CTATTGCTTC		
7621	CCGTATGGCT	TTCATTTTCT	CCTCCTTGTA	TAAATCCTGG
	TTGCTGTCTC	TTTATGAGGA		

## 038966

7681	GTTGTGGCCC	GTTGTCAGGC	AACGTGGCGT	GGTGTGCACT
	GTGTTTGCTG	ACGCAACCCC		
7741	CACTGGTTGG	GGCATTGCCA	CCACCTGTCA	GCTCCTTTCC
	GGGACTTTCG	CTTTCCCGCT		
7801	CCCTATTGCC	ACGGCGGAAC	TCATCGCCGC	CTGCCTTGCC
	CGCTGCTGGA	CAGGGGCTCG		
7861	GCTGTTGGGC	ACTGACAATT	CCGTGGTGTT	GTCGGGAAA
	TCATCGTCCT	TTCCTTGCTT		
7921	GCTCGCCTGT	GTTGCCACCT	GGATTCTGCG	CGGGACGTCC
	TTCTGCTACG	TCCCTTCGGC		
7981	CCTCAATCCA	GCGGACCTTC	CTTCCCGCGG	CCTGCTGCCG
	GCTCTGCGGC	CTCTTCCGCG		
8041	TCTTCGCCTT	CGCCCTCAGA	CGAGTCGGAT	CTCCCTTTGG
	GCCGCCTCCC	CGCAAGCTTC		
8101	GCACTTTTTA	AAAGAAAAGG	GAGGACTGGA	TGGGATTTAT
	TACTCCGATA	GGACGCTGGC		
8161	TTGTAACTCA	GTCTCTTACT	AGGAGACCAG	CTTGAGCCTG
	GGTGTTTCGCT	GGTTAGCCTA		
8221	ACCTGGTTGG	CCACCAGGGG	TAAGGACTCC	TTGGCTTAGA
	AAGCTAATAA	ACTTGCCTGC		
8281	ATTAGAGCTC	TTACGCGTCC	CGGGCTCGAG	ATCCGCATCT
	CAATTAGTCA	GCAACCATAG		
8341	TCCCGCCCTT	AACTCCGCC	ATCCCGCCCC	TAACTCCGCC
	CAGTTCCGCC	CATTCTCCGC		
8401	CCCATGGCTG	ACTAATTTTT	TTTATTTATG	CAGAGGCCGA
	GGCCGCTCG	GCCTCTGAGC		
8461	TATCCAGAA	GTAGTGAGGA	GGCTTTTTTG	GAGGCCTAGG
	CTTTTGCAA	AAGCTAACTT		
8521	GTTTATTGCA	GCTTATAATG	GTTACAAATA	AAGCAATAGC
	ATCACAAATT	TCACAAATAA		
8581	AGCATTTTTT	TCACTGCATT	CTAGTTGTGG	TTTGTCCAAA
	CTCATCAATG	TATCTTATCA		
8641	TGTCTGTCCG	CTTCTCGCT	CACTGACTCG	CTGCGCTCGG
	TCGTTCCGGCT	GCGGCGAGCG		
8701	GTATCAGCTC	ACTCAAAGGC	GGTAATACGG	TTATCCACAG
	AATCAGGGGA	TAACGCAGGA		



## 038966

8761	AAGAACATGT	GAGCAAAAGG	CCAGCAAAAG	GCCAGGAACC
GTAAAAAGGC	CGCGTTGCTG			
8821	CGGTTTTTCC	ATAGGCTCCG	CCCCCTGAC	GAGCATCACA
AAAAATCGACG	CTCAAGTCAG			
8881	AGGTGGCGAA	ACCCGACAGG	ACTATAAAGA	TACCAGGCGT
TTCCCCCTGG	AAGCTCCCTC			
8941	GTGCGCTCTC	CTGTTCCGAC	CCTGCCGCTT	ACCGGATACC
TGTCCGCCTT	TCTCCCTTCG			
9001	GGAAGCGTGG	CGCTTCTCA	TAGCTCAGC	TGTAGGTATC
TCAGTTCGGT	GTAGGTCGTT			
9061	CGCTCCAAGC	TGGGCTGTGT	GCACGAACCC	CCCCTTCAGC
CCGACCGCTG	CGCCTTATCC			
9121	GGTAACTATC	GTCTTGAGTC	CAACCCGGTA	AGACACGACT
TATCGCCACT	GGCAGCAGCC			
9181	ACTGGTAACA	GGATTAGCAG	AGCGAGGTAT	GTAGCGGGTG
CTACAGAGTT	CTTGAAGTGG			
9241	TGGCCTAACT	ACGGCTACAC	TAGAAGAACA	GTATTTGGTA
TCTGCGCTCT	GCTGAAGCCA			
9301	GTTACCTTCG	GAAAAAGAGT	TGGTAGCTCT	TGATCCGGCA
AACAAACCAC	CGCTGGTAGC			
9361	GGTGGTTTTT	TTGTTTGCAA	GCAGCAGATT	ACGCGCAGAA
AAAAAGGATC	TCAAGAAGAT			
9421	CCTTTGATCT	TTTCTACGGG	GTCTGACGCT	CAGTGGAAAC
AAAACCTACG	TTAAGGGATT			
9481	TTGGTCATGA	GATTATCAAA	AAGGATCTTC	ACCTAGATCC
TTTTAAATTA	AAAATGAAGT			
9541	TTTAAATCAA	TCTAAAGTAT	ATATGAGTAA	ACTTGGTCTG
ACAGTTAGAA	AAACTCATCG			
9601	AGCATCAAAT	GAAACTGCAA	TTTATTCATA	TCAGGATTAT
CAATACCATA	TTTTTGAAAA			
9661	AGCCGTTTCT	GTAATGAAGG	AGAAAACCTCA	CCGAGGCAGT
TCCATAGGAT	GGCAAGATCC			
9721	TGGTATCGGT	CTGCGATTCC	GACTCGTCCA	ACATCAATAC
AACCTATTAA	TTTCCCTCG			
9781	TCAAAAATAA	GGTTATCAAG	TGAGAAATCA	CCATGAGTGA
CGACTGAATC	CGGTGAGAAT			

## 038966

9841 GGCAACAGCT TATGCATTTT TTCCAGACT TGTTCACAG  
 GCCAGCCATT ACGCTCGTCA  
 9901 TCAAAATCAC TCGCATCAAC CAAACCGTTA TTCATTCGTG  
 ATTGCGCCTG AGCGAGACGA  
 9961 AATACGCGAT CGCTGTTAAA AGGACAATTA CAAACAGGAA  
 TCGAATGCAA CCGGCGCAGG  
 10021 AACACTGCCA GCGCATCAAC AATATTTTCA CCTGAATCAG  
 GATATTCTTC TAATACCTGG  
 10081 AATGCTGTTT TTCCGGGGAT CGCAGTGGTG AGTAACCATG  
 CATCATCAGG AGTACGGATA  
 10141 AAATGCTTGA TGGTCGGAAG AGGCATAAAT TCCGTCAGCC  
 AGTTTAGTCT GACCATCTCA  
 10201 TCTGTAACAT CATTGGCAAC GCTACCTTTG CCATGTTTCA  
 GAAACAACCTC TGGCGCATCG  
 10261 GGCTTCCCAT ACAATCGATA GATTGTCGCA CCTGATTGCC  
 CGACATTATC GCGAGCCCAT  
 10321 TTATACCCAT ATAAATCAGC ATCCATGTTG GAATTTAATC  
 GCGGCCTAGA GCAAGACGTT  
 10381 TCCCGTTGAA TATGGCTCAT AACACCCCTT GTATTACTGT  
 TTATGTAAGC AGACAGTTTT  
 10441 ATTGTTTCATG ATGATATATT TTTATCTTGT GCAATGTAAC  
 ATCAGAGATT TTGAGACACA  
 10501 ACAATTGGTC GACGGATCC  
**SEQ ID NO:13**  
 1 GGTACCTCAA TATTGGCCAT TAGCCATATT ATTCATTGGT TATATAGCAT  
 AAATCAATAT  
 61 TGGCTATTGG CCATTGCATA CGTTGTATCT ATATCATAAT ATGTACATTT  
 ATATTGGCTC  
 121 ATGTCCAATA TGACCGCCAT GTTGGCATTG ATTATTGACT AGTTATTAAT  
 AGTAATCAAT  
 181 TACGGGGTCA TTAGTTCATA GCCCATATAT GGAGTCCGC GTTACATAAC  
 TTACGGTAAA  
 241 TGGCCCGCCT GGCTGACCGC CCAACGACCC CCGCCATTG ACGTCAATAA  
 TGACGTATGT  
 301 TCCCATAGTA ACGCCAATAG GGACTTTCCA TTGACGTCAA TGGGTGGAGT  
 ATTTACGGTA

361 AACTGCCCAC TTGGCAGTAC ATCAAGTGTA TCATATGCCA AGTCCGCCCC  
 CTATTGACGT  
 421 CAATGACGGT AAATGGCCCG CCTGGCATTG TGCCAGTAC ATGACCTTAC  
 GGGACTTTCC  
 481 TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC  
 GGTTTGGCA  
 541 GTACACCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA TTTCCAAGTC  
 TCCACCCCAT  
 601 TGACGTCAAT GGGAGTTTGT TTTGGCACCA AAATCAACGG GACTTTCCAA  
 AATGTCGTAA  
 661 CAACTGCGAT CGCCCGCCCC GTTGACGCAA ATGGGCGGTA GGC GTGACG  
 GTGGGAGGTC  
 721 TATATAAGCA GAGCTCGCTG GCTTGTAAC T CAGTCTCTTA CTAGGAGACC  
 AGCTTGAGCC  
 781 TGGGTGTTTCG CTGGTTAGCC TAACCTGGTT GGCCACCAGG GGTAAGGACT  
 CCTTGGCTTA  
 841 GAAAGCTAAT AAACCTGCCT GCATTAGAGC TTATCTGAGT CAAGTGTCCCT  
 CATTGACGCC  
 901 TCACTCTCTT GAACGGGAAT CTTCCCTACT GGGTTCTCTC TCTGACCCAG  
 GCGAGAGAAA  
 961 CTCCAGCAGT GGC GCCCGAA CAGGGACTTG AGTGAGAGTG TAGGCACGTA  
 CAGCTGAGAA  
 1021 GGCGTCGGAC GCGAAGGAAG CGCGGGTGC GACGCGACCA  
 AGAAGGAGAC TTGGTGAGTA  
 1081 GGCTTCTCGA GTGCCGGGAA AAAGCTCGAG CCTAGTTAGA  
 GGACTAGGAG AGGCCGTAGC  
 1141 CGTAACTACT CTGGGCAAGT AGGGCAGGCG GTGGGTACGC  
 AATGGGGGCG GCTACCTCAG  
 1201 CACTAAATAG GAGACAATTA GACCAATTTG AGAAAATACG  
 ACTTCGCCCC AACGGAAAAGA  
 1261 AAAAGTACCA AATTAACAT TTAATATGGG CAGGCAAGGA  
 GATGGAGCGC TTCGGCCTCC  
 1321 ATGAGAGGTT GTTGGAGACA GAGGAGGGGT GTAAAAGAAT  
 CATAGAAGTC CTCTACCCCC  
 1381 TAGAACCAAC AGGATCGGAG GGCTTAAAAA GTCTGTTCAA  
 TCTTGTGTGC GTGCTATATT

## 038966

1441	GCTTGCACAA	GGAACAGAAA	GTGAAAGACA	CAGAGGAAGC
AGTAGCAACA	GTAAGACAAC			
1501	ACTGCCATCT	AGTGGAAAAA	GAAAAAAGTG	CAACAGAGAC
ATCTAGTGGA	CAAAGAAAAA			
1561	ATGACAAGGG	AATAGCAGCG	CCACCTGGTG	GCAGTCAGAA
TTTTCCAGCG	CAACAACAAG			
1621	GAAATGCCTG	GGTACATGTA	CCCTTGTCAC	CGCGCACCTT
AAATGCGTGG	GTAAGAGCAG			
1681	TAGAGGAGAA	AAAATTTGGA	GCAGAAATAG	TACCCATGTT
TCAAGCCCTA	TCGAATTCCC			
1741	GTTTGTGCTA	GGGTTCTTAG	GCTTCTTGGG	GGCTGCTGGA
ACTGCAATGG	GAGCAGCGGC			
1801	GACAGCCCTG	ACGGTCCAGT	CTCAGCATTT	GCTTGCTGGG
ATACTGCAGC	AGCAGAAGAA			
1861	TCTGCTGGCG	GCTGTGGAGG	CTCAACAGCA	GATGTTGAAG
CTGACCATTT	GGGGTGTAA			
1921	AAACCTCAAT	GCCCGCGTCA	CAGCCCTTGA	GAAGTACCTA
GAGGATCAGG	CACGACTAAA			
1981	CTCCTGGGGG	TGCGCATGGA	AACAAGTATG	TCATACCACA
GTGGAGTGGC	CCTGGACAAA			
2041	TCGGACTCCG	GATTGGCAAA	ATATGACTTG	GTTGGAGTGG
GAAAGACAAA	TAGCTGATTT			
2101	GGAAAGCAAC	ATTACGAGAC	AATTAGTGAA	GGCTAGAGAA
CAAGAGGAAA	AGAATCTAGA			
2161	TGCCTATCAG	AAGTTAACTA	GTTGGTCAGA	TTTCTGGTCT
TGGTTCGATT	TCTCAAAATG			
2221	GCTTAACATT	TTAAAAATGG	GATTTTTAGT	AATAGTAGGA
ATAATAGGGT	TAAGATTACT			
2281	TTACACAGTA	TATGGATGTA	TAGTGAGGGT	TAGGCAGGGA
TATGTTCCCTC	TATCTCCACA			
2341	GATCCATATC	CGCGGCAATT	TTAAAAGAAA	GGGAGGAATA
GGGGGACAGA	CTTCAGCAGA			
2401	GAGACTAATT	AATATAATAA	CAACACAATT	AGAAATACAA
CATTTACAAA	CCAAAATTCA			
2461	AAAAATTTTA	AATTTTAGAG	CCGCGGAGAT	CTCAATATTG
GCCATTAGCC	ATATTATTCA			

## 038966

2521	TTGGTTATAT	AGCATAAATC	AATATTGGCT	ATTGGCCATT
	GCATACGTTG	TATCTATATC		
2581	ATAATATGTA	CATTTATATT	GGCTCATGTC	CAATATGACC
	GCCATGTTGG	CATTGATTAT		
2641	TGACTAGTTA	TTAATAGTAA	TCAATTACGG	GGTCATTAGT
	TCATAGCCCA	TATATGGAGT		
2701	TCCGCGTTAC	ATAACTTACG	GTAAATGGCC	CGCCTGGCTG
	ACCGCCCAAC	GACCCCGGCC		
2761	CATTGACGTC	AATAATGACG	TATGTTCCCA	TAGTAACGCC
	AATAGGGACT	TTCCATTGAC		
2821	GTCAATGGGT	GGAGTATTTA	CGGTAAACTG	CCCCTGGCC
	AGTACATCAA	GTGTATCATA		
2881	TGCCAAGTCC	GCCCCCTATT	GACGTCAATG	ACGGTAAATG
	GCCCCCTGG	CATTATGCCC		
2941	AGTACATGAC	CTTACGGGAC	TTTCCTACTT	GGCAGTACAT
	CTACGTATTA	GTCATCGCTA		
3001	TTACCATGGT	GATGCGGTTT	TGGCAGTACA	CCAATGGGCG
	TGGATAGCGG	TTTACTCAC		
3061	GGGATTTCC	AAGTCTCCAC	CCCATTGACG	TCAATGGGAG
	TTTGTTTTGG	CACCAAATC		
3121	AACGGGACTT	TCCAAAATGT	CGTAATAACC	CCGCCCGGTT
	GACGCAAATG	GGCGGTAGGC		
3181	GTGTACGGTG	GGAGGTCTAT	ATAAGCAGAG	CTCGTTTAGT
	GAACCGTCAG	ATCACTAGAA		
3241	GCTTTATTGC	GGTAGTTTAT	CACAGTTAAA	TTGCTAACGC
	AGTCAGTGCT	TCTGACACAA		
3301	CAGTCTCGAA	CTTAAGCTGC	AGAAGTTGGT	CGTGAGGCAC
	TGGGCAGGCT	AGCCACCAAT		
3361	GCAGATTGAG	CTGAGCACCT	GCTTCTTCCT	GTGCCTGCTG
	AGGTTCTGCT	TCTCTGCCAC		
3421	CAGGAGATAC	TACCTGGGGG	CTGTGGAGCT	GAGCTGGGAC
	TACATGCAGT	CTGACCTGGG		
3481	GGAGCTGCCT	GTGGATGCCA	GGTTCCCCCC	CAGAGTGCCC
	AAGAGCTTCC	CCTTCAACAC		
3541	CTCTGTGGTG	TACAAGAAGA	CCCTGTTTGT	GGAGTTCACT
	GACCACCTGT	TCAACATTGC		

## 038966

3601	CAAGCCCAGG	CCCCCCTGGA	TGGGCCTGCT	GGGCCCCACC
ATCCAGGCTG	AGGTGTATGA			
3661	CACTGTGGTG	ATCACCCCTGA	AGAACATGGC	CAGCCACCCT
GTGAGCCTGC	ATGCTGTGGG			
3721	GGTGAGCTAC	TGGAAGGCCT	CTGAGGGGGC	TGAGTATGAT
GACCAGACCA	GCCAGAGGGA			
3781	GAAGGAGGAT	GACAAGGTGT	TCCCTGGGGG	CAGCCACACC
TATGTGTGGC	AGGTGCTGAA			
3841	GGAGAATGGC	CCCATGGCCT	CTGACCCCTT	GTGCCTGACC
TACAGCTACC	TGAGCCATGT			
3901	GGACCTGGTG	AAGGACCTGA	ACTCTGGCCT	GATTGGGGCC
CTGCTGGTGT	GCAGGGAGGG			
3961	CAGCCTGGCC	AAGGAGAAGA	CCCAGACCCT	GCACAAGTTC
ATCCTGCTGT	TTGCTGTGTT			
4021	TGATGAGGGC	AAGAGCTGGC	ACTCTGAAAC	CAAGAACAGC
CTGATGCAGG	ACAGGGATGC			
4081	TGCCTCTGCC	AGGGCCTGGC	CCAAGATGCA	CACTGTGAAT
GGCTATGTGA	ACAGGAGCCT			
4141	GCCTGGCCTG	ATTGGCTGCC	ACAGGAAGTC	TGTGTACTGG
CATGTGATTG	GCATGGGCAC			
4201	CACCCCTGAG	GTGCACAGCA	TCTTCCTGGA	GGGCCACACC
TTCTTGGTCA	GGAACCACAG			
4261	GCAGGCCAGC	CTGGAGATCA	GCCCCATCAC	CTTCCTGACT
GCCCAGACCC	TGCTGATGGA			
4321	CCTGGGCCAG	TTCTGCTGT	TCTGCCACAT	CAGCAGCCAC
CAGCATGATG	GCATGGAGGC			
4381	CTATGTGAAG	GTGGACAGCT	GCCCTGAGGA	GCCCCAGCTG
AGGATGAAGA	ACAATGAGGA			
4441	GGCTGAGGAC	TATGATGATG	ACCTGACTGA	CTCTGAGATG
GATGTGGTGA	GGTTTGATGA			
4501	TGACAACAGC	CCCAGCTTCA	TCCAGATCAG	GTCTGTGGCC
AAGAAGCACC	CCAAGACCTG			
4561	GGTGCACTAC	ATTGCTGCTG	AGGAGGAGGA	CTGGGACTAT
GCCCCCTGG	TGCTGGCCCC			
4621	TGATGACAGG	AGCTACAAGA	GCCAGTACCT	GAACAATGGC
CCCCAGAGGA	TTGGCAGGAA			

## 038966

4681	GTACAAGAAG	GTCAGGTTCA	TGGCCTACAC	TGATGAAACC
	TTCAAGACCA GGGAGGCCAT			
4741	CCAGCATGAG	TCTGGCATCC	TGGGCCCCCT	GCTGTATGGG
	GAGGTGGGGG ACACCCTGCT			
4801	GATCATCTTC	AAGAACCAGG	CCAGCAGGCC	CTACAACATC
	TACCCCATG GCATCACTGA			
4861	TGTGAGGCC	CTGTACAGCA	GGAGGCTGCC	CAAGGGGGTG
	AAGCACCTGA AGGACTTCCC			
4921	CATCCTGCCT	GGGGAGATCT	TCAAGTACAA	GTGGACTGTG
	ACTGTGGAGG ATGGCCCCAC			
4981	CAAGTCTGAC	CCCAGGTGCC	TGACCAGATA	CTACAGCAGC
	TTTGTGAACA TGAGAGGGGA			
5041	CCTGGCCTCT	GGCCTGATTG	GCCCCCTGCT	GATCTGCTAC
	AAGGAGTCTG TGGACCAGAG			
5101	GGGCAACCAG	ATCATGTCTG	ACAAGAGGAA	TGTGATCCTG
	TTCTCTGTGT TTGATGAGAA			
5161	CAGGAGCTGG	TACCTGACTG	AGAACATCCA	GAGGTTCCCTG
	CCCAACCCCTG CTGGGGTGCA			
5221	GCTGGAGGAC	CCTGAGTTCC	AGGCCAGCAA	CATCATGCAC
	AGCATCAATG GCTATGTGTT			
5281	TGACAGCCTG	CAGCTGTCTG	TGTGCCTGCA	TGAGGTGGCC
	TACTGGTACA TCCTGAGCAT			
5341	TGGGGCCCAG	ACTGACTTCC	TGTCTGTGTT	CTTCTCTGGC
	TACACCTTCA AGCACAAGAT			
5401	GGTGTATGAG	GACACCCTGA	CCCTGTTCCC	CTTCTCTGGG
	GAGACTGTGT TCATGAGCAT			
5461	GGAGAACCCT	GGCCTGTGGA	TTCTGGGCTG	CCACAACCTCT
	GACTTCAGGA ACAGGGGCAT			
5521	GACTGCCCTG	CTGAAAGTCT	CCAGCTGTGA	CAAGAACACT
	GGGGACTACT ATGAGGACAG			
5581	CTATGAGGAC	ATCTCTGCCT	ACCTGCTGAG	CAAGAACAAT
	GCCATTGAGC CCAGGAGCTT			
5641	CAGCCAGAAC	AGCAGGCACC	CCAGCACCAG	GCAGAAGCAG
	TTCAATGCCA CCACCATCCC			
5701	TGAGAATGAC	ATAGAGAAGA	CAGACCCATG	GTTTGCCCAC
	CGGACCCCCA TGCCCAAGAT			

## 038966

5761	CCAGAATGTG	AGCAGCTCTG	ACCTGCTGAT	GCTGCTGAGG
CAGAGCCCCA	CCCCCATGG			
5821	CCTGAGCCTG	TCTGACCTGC	AGGAGGCCAA	GTATGAAACC
TTCTCTGATG	ACCCAGCCCC			
5881	TGGGGCCATT	GACAGCAACA	ACAGCCTGTC	TGAGATGACC
CACTTCAGGC	CCCAGCTGCA			
5941	CCACTCTGGG	GACATGGTGT	TCACCCCTGA	GTCTGGCCTG
CAGCTGAGGC	TGAATGAGAA			
6001	GCTGGGCACC	ACTGCTGCCA	CTGAGCTGAA	GAAGCTGGAC
TTCAAAGTCT	CCAGCACCAG			
6061	CAACAACCTG	ATCAGCACCA	TCCCTCTGA	CAACCTGGCT
GCTGGCACTG	ACAACACCAG			
6121	CAGCCTGGGC	CCCCCAGCA	TGCCTGTGCA	CTATGACAGC
CAGCTGGACA	CCACCCTGTT			
6181	TGGCAAGAAG	AGCAGCCCCC	TGACTGAGTC	TGGGGCCCCC
CTGAGCCTGT	CTGAGGAGAA			
6241	CAATGACAGC	AAGCTGCTGG	AGTCTGGCCT	GATGAACAGC
CAGGAGAGCA	GCTGGGGCAA			
6301	GAATGTGAGC	AGCAGGGAGA	TCACCAGGAC	CACCCTGCAG
TCTGACCAGG	AGGAGATTGA			
6361	CTATGATGAC	ACCATCTCTG	TGGAGATGAA	GAAGGAGGAC
TTTGACATCT	ACGACGAGGA			
6421	CGAGAACCAG	AGCCCCAGGA	GCTTCCAGAA	GAAGACCAGG
CACTACTTCA	TTGCTGCTGT			
6481	GGAGAGGCTG	TGGGACTATG	GCATGAGCAG	CAGCCCCCAT
GTGCTGAGGA	ACAGGGCCCA			
6541	GTCTGGCTCT	GTGCCCCAGT	TCAAGAAGGT	GGTGTTCAG
GAGTTCACTG	ATGGCAGCTT			
6601	CACCCAGCCC	CTGTACAGAG	GGGAGCTGAA	TGAGCACCTG
GGCCTGCTGG	GCCCCTACAT			
6661	CAGGGCTGAG	GTGGAGGACA	ACATCATGGT	GACCTTCAGG
AACCAGGCCA	GCAGGCCCTA			
6721	CAGCTTCTAC	AGCAGCCTGA	TCAGCTATGA	GGAGGACCAG
AGGCAGGGGG	CTGAGCCCAG			
6781	GAAGAACTTT	GTGAAGCCCA	ATGAAACCAA	GACCTACTTC
TGGAAGGTGC	AGCACCACAT			



## 038966

6841	GGCCCCACC	AAGGATGAGT	TTGACTGCAA	GGCCTGGGCC
TACTTCTCTG	ATGTGGACCT			
6901	GGAGAAGGAT	GTGCACTCTG	GCCTGATTGG	CCCCCTGCTG
GTGTGCCACA	CCAACACCCT			
6961	GAACCCTGCC	CATGGCAGGC	AGGTGACTGT	GCAGGAGTTT
GCCCTGTTCT	TCACCATCTT			
7021	TGATGAAACC	AAGAGCTGGT	ACTTCACTGA	GAACATGGAG
AGGAACTGCA	GGGCCCCCTG			
7081	CAACATCCAG	ATGGAGGACC	CCACCTTCAA	GGAGAACTAC
AGGTTCCATG	CCATCAATGG			
7141	CTACATCATG	GACACCCTGC	CTGGCCTGGT	GATGGCCCAG
GACCAGAGGA	TCAGGTGGTA			
7201	CCTGCTGAGC	ATGGGCAGCA	ATGAGAACAT	CCACAGCATC
CACTTCTCTG	GCCATGTGTT			
7261	CACTGTGAGG	AAGAAGGAGG	AGTACAAGAT	GGCCCTGTAC
AACCTGTACC	CTGGGGTGTT			
7321	TGAGACTGTG	GAGATGCTGC	CCAGCAAGGC	TGGCATCTGG
AGGGTGGAGT	GCCTGATTGG			
7381	GGAGCACCTG	CATGCTGGCA	TGAGCACCCCT	GTTCTTGGTG
TACAGCAACA	AGTGCCAGAC			
7441	CCCCCTGGGC	ATGGCCTCTG	GCCACATCAG	GGACTTCCAG
ATCACTGCCT	CTGGCCAGTA			
7501	TGGCCAGTGG	GCCCCAAGC	TGGCCAGGCT	GCACTACTCT
GGCAGCATCA	ATGCCTGGAG			
7561	CACCAAGGAG	CCCTTCAGCT	GGATCAAGGT	GGACCTGCTG
GCCCCCATGA	TCATCCATGG			
7621	CATCAAGACC	CAGGGGGCCA	GGCAGAAGTT	CAGCAGCCTG
TACATCAGCC	AGTTCATCAT			
7681	CATGTACAGC	CTGGATGGCA	AGAAGTGGCA	GACCTACAGG
GGCAACAGCA	CTGGCACCCCT			
7741	GATGGTGTTT	TTTGGCAATG	TGGACAGCTC	TGGCATCAAG
CACAACATCT	TCAACCCCC			
7801	CATCATTGCC	AGATACATCA	GGCTGCACCC	CACCCACTAC
AGCATCAGGA	GCACCCTGAG			
7861	GATGGAGCTG	ATGGGCTGTG	ACCTGAACAG	CTGCAGCATG
CCCCTGGGCA	TGGAGAGCAA			

## 038966

7921	GGCCATCTCT	GATGCCCAGA	TCACTGCCAG	CAGCTACTTC
ACCAACATGT	TTGCCACCTG			
7981	GAGCCCCAGC	AAGGCCAGGC	TGCACCTGCA	GGGCAGGAGC
AATGCCTGGA	GGCCCCAGGT			
8041	CAACAACCCC	AAGGAGTGGC	TGCAGGTGGA	CTTCCAGAAG
ACCATGAAGG	TGACTGGGGT			
8101	GACCACCCAG	GGGGTGAAGA	GCCTGCTGAC	CAGCATGTAT
GTGAAGGAGT	TCCTGATCAG			
8161	CAGCAGCCAG	GATGGCCACC	AGTGGACCCT	GTTCTTCCAG
AATGGCAAGG	TGAAGGTGTT			
8221	CCAGGGCAAC	CAGGACAGCT	TCACCCCTGT	GGTGAACAGC
CTGGACCCCC	CCCTGCTGAC			
8281	CAGATACCTG	AGGATTCACC	CCCAGAGCTG	GGTGCACCAG
ATTGCCCTGA	GGATGGAGGT			
8341	GCTGGGCTGT	GAGGCCCAGG	ACCTGTACTG	AGCGGCCGCG
GGCCCAATCA	ACCTCTGGAT			
8401	TACAAAATTT	GTGAAAGATT	GACTGGTATT	CTTAACTATG
TTGCTCCTTT	TACGCTATGT			
8461	GGATACGCTG	CTTTAATGCC	TTTGTATCAT	GCTATTGCTT
CCCGTATGGC	TTTCATTTTC			
8521	TCCTCCTTGT	ATAAATCCTG	GTTGCTGTCT	CTTTATGAGG
AGTTGTGGCC	CGTTGTCAGG			
8581	CAACGTGGCG	TGGTGTGCAC	TGTGTTTGCT	GACGCAACCC
CCACTGGTTG	GGGCATTGCC			
8641	ACCACCTGTC	AGCTCCTTTC	CGGGACTTTC	GCTTTCCCCC
TCCCTATTGC	CACGGCGGAA			
8701	CTCATCGCCG	CCTGCCTTGC	CCGCTGCTGG	ACAGGGGCTC
GGCTGTTGGG	CACTGACAAT			
8761	TCCGTGGTGT	TGTCGGGGAA	ATCATCGTCC	TTTCCTTGGC
TGCTCGCCTG	TGTTGCCACC			
8821	TGGATTCTGC	GCGGGACGTC	CTTCTGCTAC	GTCCCTTCGG
CCCTCAATCC	AGCGGACCTT			
8881	CCTTCCCXCG	GCCTGCTGCC	GGCTCTGCGG	CCTCTTCCGC
GTCTTCGCCT	TCGCCCTCAG			
8941	ACGAGTCGGA	TCTCCCTTTG	GGCCGCCTCC	CCGCAAGCTT
CGCACTTTTT	AAAAGAAAAG			

## 038966

9001	GGAGGACTGG	ATGGGATTTA	TTACTCCGAT	AGGACGCTGG
CTTGTA	AGTCTCTTAC			
9061	TAGGAGACCA	GCTTGAGCCT	GGGTGTTCCG	TGGTTAGCCT
AACCTGG	GCCACCAGGG			
9121	GTAAGGACTC	CTTGGCTTAG	AAAGCTAATA	AACTTGCCTG
CATTAGAG	CTTACGCGTC			
9181	CCGGGCTCGA	GATCCGCATC	TCAATTAGTC	AGCAACCATA
GTCCCGCCC	TA	ACTCCGCC		
9241	CATCCCGCCC	CTAACTCCGC	CCAGTCCGC	CCATTCTCCG
CCCCATGG	GACTAATTTT			
9301	TTTTATTTAT	GCAGAGGCCG	AGGCCGCCCTC	GGCCTCTGAG
CTATTCCAG	AGTAGTGAGG			
9361	AGGCTTTTTT	GGAGGCCTAG	GCTTTTGCAA	AAAGCTAACT
TGTTTATTG	AGCTTATAAT			
9421	GGTTACAAAT	AAAGCAATAG	CATCACAAT	TTCACAAATA
AAGCATTTTT	TTC	ACTGCAT		
9481	TCTAGTTGTG	GTTTGTCCAA	ACTCATCAAT	GTATCTTATC
ATGCTGTG	GCTTCTCTCG			
9541	TCACTGACTC	GCTGCGCTCG	GTCGTTCCGC	TGCGCGGAGC
GGTATCAG	CACTCAAAGG			
9601	CGGTAATACG	GTTATCCACA	GAATCAGGGG	ATAACGCAGG
AAAGAACATG	TGAGCAAAA			
9661	GCCAGCAAAA	GGCCAGGAAC	CGTAAAAAGG	CCGCGTTGCT
GGCGTTTTTC	CATAGGCTCC			
9721	GCCCCCTGA	CGAGCATCAC	AAAAATCGAC	GCTCAAGTCA
GAGGTGGCG	AACCCGACAG			
9781	GACTATAAAG	ATACCAGGCG	TTTCCCCCTG	GAAGCTCCCT
CGTGCGCTCT	CCTGTTCCGA			
9841	CCCTGCCGCT	TACCGGATAC	CTGTCCGCCT	TTCTCCCTTC
GGGAAGCGTG	GCGCTTTCTC			
9901	ATAGCTCACG	CTGTAGGTAT	CTCAGTTCGG	TGTAGGTCGT
TCGCTCCAAG	CTGGGCTGTG			
9961	TGCACGAACC	CCCCGTTACG	CCCACCGCT	GCGCCTTATC
CGGTA	ACTATCGTCTT	GAGT		
10021	CCAACCCGGT	AAGACACGAC	TTATCGCCAC	TGGCAGCAGC
CACTGGTA	AGGATTAGCA			

## 038966

10081	GAGCGAGGTA	TGTAGGCGGT	GCTACAGAGT	TCTTGAAGTG
	GTGGCCTAAC	TACGGCTACA		
10141	CTAGAAGAAC	AGTATTTGGT	ATCTGCGCTC	TGCTGAAGCC
	AGTTACCTTC	GGAAAAAGAG		
10201	TTGGTAGCTC	TTGATCCGGC	AAACAAACCA	CCGCTGGTAG
	CGGTGGTTTT	TTTGTTCGCA		
10261	AGCAGCAGAT	TACGCGCAGA	AAAAAAGGAT	CTCAAGAAGA
	TCCTTTGATC	TTTCTACGG		
10321	GGTCTGACGC	TCAGTGGAAC	GAAAACTCAC	GTTAAGGGAT
	TTTGGTCATG	AGATTATCAA		
10381	AAAGGATCTT	CACCTAGATC	CTTTTAAATT	AAAAATGAAG
	TTTTAAATCA	ATCTAAAGTA		
10441	TATATGAGTA	AACTTGGTCT	GACAGTTAGA	AAAACATCAT
	GAGCATCAA	TGAAACTGCA		
10501	ATTTATTCAT	ATCAGGATTA	TCAATACCAT	ATTTTTGAAA
	AAGCCGTTTC	TGTAATGAAG		
10561	GAGAAAACTC	ACCGAGGCAG	TTCCATAGGA	TGGCAAGATC
	CTGGTATCGG	TCTGCGATTC		
10621	CGACTCGTCC	AACATCAATA	CAACCTATTA	ATTTCCCTC
	GTCAAAAATA	AGGTTATCAA		
10681	GTGAGAAATC	ACCATGAGTG	ACGACTGAAT	CCGGTGAGAA
	TGGCAACAGC	TTATGCATTT		
10741	CTTCCAGAC	TTGTTCAACA	GGCCAGCCAT	TACGCTCGTC
	ATCAAAATCA	CTCGCATCAA		
10801	CCAAACCGTT	ATTCATTCGT	GATTGCGCCT	GAGCGAGACG
	AAATACGCGA	TCGCTGTTAA		
10861	AAGGACAATT	ACAAACAGGA	ATCGAATGCA	ACCGGCGCAG
	GAACACTGCC	AGCGCATCAA		
10921	CAATATTTTC	ACCTGAATCA	GGATATTCTT	CTAATACCTG
	GAATGCTGTT	TTTCCGGGGA		
10981	TCGCAGTGGT	GAGTAACCAT	GCATCATCAG	GAGTACGGAT
	AAAATGCTTG	ATGGTCGGAA		
11041	GAGGCATAAA	TTCCGTCAGC	CAGTTTAGTC	TGACCATCTC
	ATCTGTAACA	TCATTGGCAA		
11101	CGCTACCTTT	GCCATGTTTC	AGAAACAACCT	CTGGCGCATC
	GGGCTTCCCA	TACAATCGAT		

## 038966

11161 AGATTGTCGC ACCTGATTGC CCGACATTAT CGCGAGCCCA  
 TTTATACCCA TATAAATCAG  
 11221 CATCCATGTT GGAATTTAAT CGCGGCCTAG AGCAAGACGT  
 TTCCCGTTGA ATATGGCTCA  
 11281 TAACACCCCT TGTATTACTG TTTATGTAAG CAGACAGTTT  
 TATTGTTTCAT GATGATATAT  
 11341 TTTTATCTTG TGCAATGTAA CATCAGAGAT TTTGAGACAC  
 AACAATGGT CGACGGATCC

## SEQ ID NO:14

1 GGTACCTCAA TATTGGCCAT TAGCCATATT ATTCATTGGT TATATAGCAT  
 AAATCAATAT  
 61 TGGCTATTGG CCATTGCATA CGTTGTATCT ATATCATAAT ATGTACATTT  
 ATATTGGCTC  
 121 ATGTCCAATA TGACCGCCAT GTTGGCATTG ATTATTGACT AGTTATTAAT  
 AGTAATCAAT  
 181 TACGGGGTCA TTAGTTCATA GCCCATATAT GGAGTCCGC GTTACATAAC  
 TTACGGTAAA  
 241 TGGCCCGCCT GGCTGACCGC CCAACGACCC CCGCCCATG ACGTCAATAA  
 TGACGTATGT  
 301 TCCCATAGTA ACGCCAATAG GGACTTTCCA TTGACGTCAA TGGGTGGAGT  
 ATTTACGGTA  
 361 AACTGCCCAC TTGGCAGTAC ATCAAGTGTA TCATATGCCA AGTCCGCCCC  
 CTATTGACGT  
 421 CAATGACGGT AAATGGCCCG CCTGGCATTG TGCCAGTAC ATGACCTTAC  
 GGGACTTTCC  
 481 TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC  
 GGTTTTGGCA  
 541 GTACACCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA TTTCCAAGTC  
 TCCACCCCAT  
 601 TGACGTCAA GGGAGTTTGT TTTGGCACCA AAATCAACGG GACTTTCCAA  
 AATGTCGTAA  
 661 CAACTGCGAT CGCCGCCCC GTTGACGCAA ATGGGCGGTA GCGTGTACG  
 GTGGGAGGTC  
 721 TATATAAGCA GAGCTCGCTG GCTTGTAAC CAGTCTCTTA CTAGGAGACC  
 AGCTTGAGCC  
 781 TGGGTGTTTC CTGGTTAGCC TAACCTGGTT GGCCACCAGG GGTAAAGGACT

CCTTGGCTTA

841 GAAAGCTAAT AAACCTGCCT GCATTAGAGC TTATCTGAGT CAAGTGTCTT  
CATTGACGCC

901 TCACTCTCTT GAACGGGAAT CTTCTTACT GGGTTCTCTC TCTGACCCAG  
GCGAGAGAAA

961 CTCCAGCAGT GCGCCCCGAA CAGGGACTTG AGTGAGAGTG TAGGCACGTA  
CAGCTGAGAA

1021 GCGTTCGGAC GCGAAGGAAG CGCGGGGTGC GACGCGACCA  
AGAAGGAGAC TTGGTGAGTA

1081 GGCTTCTCGA GTGCCGGGAA AAAGCTCGAG CCTAGTTAGA  
GGACTAGGAG AGGCCGTAGC

1141 CGTAACTACT CTTGGGCAAG TAGGGCAGGC GGTGGGTACG  
CAATGGGGGC GGCTACCTCA

1201 GCACATAATA GGAGACAATT AGACCAATTT GAGAAAATAC  
GACTTCGCCC GAACGGAAAG

1261 AAAAAGTACC AAATTAACA TTTAATATGG GCAGGCAAGG  
AGATGGAGCG CTTCGGCCTC

1321 CATGAGAGGT TGTTGGAGAC AGAGGAGGGG TGTAAAAGAA  
TCATAGAAGT CCTCTACCCC

1381 CTAGAACCAA CAGGATCGGA GGGCTTAAAA AGTCTGTTCA  
ATCTTGTGTG CGTGCTATAT

1441 TGCTTGACACA AGGAACAGAA AGTGAAAGAC ACAGAGGAAG  
CAGTAGCAAC AGTAAGACAA

1501 CACTGCCATC TAGTGGAAAA AGAAAAAAGT GCAACAGAGA  
CATCTAGTGG AAAAAAGAAA

1561 AATGACAAGG GAATAGCAGC GCCACCTGGT GGCAGTCAGA  
ATTTTCCAGC GCAACAACAA

1621 GGAAATGCCT GGGTACATGT ACCCTTGTCA CCGCGCACCT  
TAAATGCGTG GGTAAGGCA

1681 GTAGAGGAGA AAAAATTTGG AGCAGAAATA GTACCCATGT  
TTCAAGCCCT ATCGAATTCC

1741 CGTTTGTGCT AGGGTTCTTA GGCTTCTTGG GGGCTGCTGG  
AACTGCAATG GGAGCAGCGG

1801 CGACAGCCCT GACGGTCCAG TCTCAGCATT TGCTTGCTGG  
GATACTGCAG CAGCAGAAGA

1861 ATCTGCTGGC GGCTGTGGAG GCTCAACAGC AGATGTTGAA

## 038966

GCTGACCATT	TGGGGTGTTA			
1921	AAAACCTCAA	TGCCCCGCTC	ACAGCCCTTG	AGAAGTACCT
AGAGGATCAG	GCACGACTAA			
1981	ACTCCTGGGG	GTGCGCATGG	AAACAAGTAT	GTCATACCAC
AGTGGAGTGG	CCCTGGACAA			
2041	ATCGGACTCC	GGATTGGCAA	AATATGACTT	GGTTGGAGTG
GGAAAGACAA	ATAGCTGATT			
2101	TGGAAAGCAA	CATTACGAGA	CAATTAGTGA	AGGCTAGAGA
ACAAGAGGAA	AAGAATCTAG			
2161	ATGCCATCA	GAAGTAACT	AGTTGGTCAG	ATTTCTGGTC
TTGGTTCGAT	TTCTCAAAAT			
2221	GGCTTAACAT	TTTAAAAATG	GGATTTTTAG	TAATAGTAGG
AATAATAGGG	TTAAGATTAC			
2281	TTTACACAGT	ATATGGATGT	ATAGTGAGGG	TTAGGCAGGG
ATATGTTCCCT	CTATCTCCAC			
2341	AGATCCATAT	CCGCGGCAAT	TTTAAAAGAA	AGGGAGGAAT
AGGGGGACAG	ACTTCAGCAG			
2401	AGAGACTAAT	TAATATAATA	ACAACACAAT	TAGAAAATACA
ACATTTACAA	ACCAAAAATC			
2461	AAAAAATTTT	AAATTTTAGA	GCCGCGGAGA	TCTGTTACAT
AAC TTATGGT	AAATGGCCTG			
2521	CCTGGCTGAC	TGCCCAATGA	CCCCTGCCCA	ATGATGTCAA
TAATGATGTA	TGTTCCCATG			
2581	TAATGCCAAT	AGGGACTTTC	CATTGATGTC	AATGGGTGGA
GTATTTATGG	TAAGTCCCA			
2641	CTTGGCAGTA	CATCAAGTGT	ATCATATGCC	AAGTATGCCC
CCTATTGATG	TCAATGATGG			
2701	TAAATGGCCT	GCCTGGCATT	ATGCCCAGTA	CATGACCTTA
TGGGACTTTC	CTACTTGGCA			
2761	GTACATCTAT	GTATTAGTCA	TTGCTATTAC	CATGGGAATT
CACTAGTGGA	GAAGAGCATG			
2821	CTTGAGGGCT	GAGTGCCCT	CAGTGGGCAG	AGAGCACATG
GCCCACAGTC	CCTGAGAAGT			
2881	TGGGGGGAGG	GGTGGGCAAT	TGAACTGGTG	CCTAGAGAAG
GTGGGGCTTG	GGTAAACTGG			
2941	GAAAGTGATG	TGGTGTACTG	GCTCCACCTT	TTTCCCCAGG

## 038966

GTGGGGGAGA ACCATATATA  
 3001 AGTGCAGTAG TCTCTGTGAA CATTCAAGCT TCTGCCTTCT  
 CCCTCCTGTG AGTTTGCTAG  
 3061 CCACCAATGC AGATTGAGCT GAGCACCTGC TTCTTCCTGT  
 GCCTGCTGAG GTTCTGCTTC  
 3121 TCTGCCACCA GGAGATACTA CCTGGGGGCT GTGGAGCTGA  
 GCTGGGACTA CATGCAGTCT  
 3181 GACCTGGGGG AGCTGCCTGT GGATGCCAGG TTCCCCCCA  
 GAGTGCCCAA GAGCTTCCCC  
 3241 TTCAACACCT CTGTGGTGTA CAAGAAGACC CTGTTTGTGG  
 AGTTCACTGA CCACCTGTTC  
 3301 AACATTGCCA AGCCCAGGCC CCCCTGGATG GGCCTGCTGG  
 GCCCCACCAT CCAGGCTGAG  
 3361 GTGTATGACA CTGTGGTGAT CACCCTGAAG AACATGGCCA  
 GCCACCCTGT GAGCCTGCAT  
 3421 GCTGTGGGGG TGAGCTACTG GAAGGCCTCT GAGGGGGCTG  
 AGTATGATGA CCAGACCAGC  
 3481 CAGAGGGAGA AGGAGGATGA CAAGGTGTTT CCTGGGGGCA  
 GCCACACCTA TGTGTGGCAG  
 3541 GTGCTGAAGG AGAATGGCCC CATGGCCTCT GACCCCTGT  
 GCCTGACCTA CAGCTACCTG  
 3601 AGCCATGTGG ACCTGGTGAA GGACCTGAAC TCTGGCCTGA  
 TTGGGGCCCT GCTGGTGTGC  
 3661 AGGGAGGGCA GCCTGGCCAA GGAGAAGACC CAGACCCTGC  
 ACAAGTTCAT CCTGCTGTTT  
 3721 GCTGTGTTTG ATGAGGGCAA GAGCTGGCAC TCTGAAACCA  
 AGAACAGCCT GATGCAGGAC  
 3781 AGGGATGCTG CCTCTGCCAG GCCTGGCCC AAGATGCACA  
 CTGTGAATGG CTATGTGAAC  
 3841 AGGAGCCTGC CTGGCCTGAT TGGCTGCCAC AGGAAGTCTG  
 TGTA CTGGCA TGTGATTGGC  
 3901 ATGGGCACCA CCCCTGAGGT GCACAGCATC TTCCTGGAGG  
 GCCACACCTT CCTGGTCAGG  
 3961 AACCACAGGC AGGCCAGCCT GGAGATCAGC CCCATCACCT  
 TCCTGACTGC CCAGACCCTG  
 4021 CTGATGGACC TGGCCAGTT CCTGCTGTTT TGCCACATCA



## 038966

GCAGCCACCA GCATGATGGC  
 4081 ATGGAGGCCT ATGTGAAGGT GGACAGCTGC CCTGAGGAGC  
 CCCAGCTGAG GATGAAGAAC  
 4141 AATGAGGAGG CTGAGGACTA TGATGATGAC CTGACTGACT  
 CTGAGATGGA TGTGGTGAGG  
 4201 TTTGATGATG ACAACAGCCC CAGCTTCATC CAGATCAGGT  
 CTGTGGCCAA GAAGCACCCC  
 4261 AAGACCTGGG TGCACTACAT TGCTGCTGAG GAGGAGGACT  
 GGGACTATGC CCCCCTGGTG  
 4321 CTGGCCCTG ATGACAGGAG CTACAAGAGC CAGTACCTGA  
 ACAATGGCCC CCAGAGGATT  
 4381 GGCAGGAAGT ACAAGAAGGT CAGGTTCATG GCCTACACTG  
 ATGAAACCTT CAAGACCAGG  
 4441 GAGGCCATCC AGCATGAGTC TGGCATCCTG GGGCCCTGCG  
 TGTATGGGGA GGTGGGGGAC  
 4501 ACCCTGCTGA TCATCTTCAA GAACCAGGCC AGCAGGCCCT  
 ACAACATCTA CCCCATGGC  
 4561 ATCACTGATG TGAGGCCCTT GTACAGCAGG AGGCTGCCCCA  
 AGGGGGTGAA GCACCTGAAG  
 4621 GACTTCCCCA TCCTGCCTGG GGAGATCTTC AAGTACAAGT  
 GGAATGTGAC TGTGGAGGAT  
 4681 GGCCCCACCA AGTCTGACCC CAGGTGCCTG ACCAGATACT  
 ACAGCAGCTT TGTGAACATG  
 4741 GAGAGGGACC TGGCCTCTGG CCTGATTGGC CCCCTGCTGA  
 TCTGCTACAA GGAGTCTGTG  
 4801 GACCAGAGGG GCAACCAGAT CATGTCTGAC AAGAGGAATG  
 TGATCCTGTT CTCTGTGTTT  
 4861 GATGAGAACA GGAGCTGGTA CCTGACTGAG AACATCCAGA  
 GGTTCCTGCC CAACCCTGCT  
 4921 GGGGTGCAGC TGGAGGACCC TGAGTTCAG GCCAGCAACA  
 TCATGCACAG CATCAATGGC  
 4981 TATGTGTTTG ACAGCCTGCA GCTGTCTGTG TGCCTGCATG  
 AGGTGGCCTA CTGGTACATC  
 5041 CTGAGCATTG GGGCCAGAC TGACTTCCTG TCTGTGTTCT  
 TCTCTGGCTA CACCTTCAAG  
 5101 CACAAGATGG TGTATGAGGA CACCCTGACC CTGTTCCCTT

## 038966

TCTCTGGGGA	GACTGTGTTC				
5161	ATGAGCATGG	AGAACCCTGG	CCTGTGGATT	CTGGGCTGCC	
ACAACTCTGA	CTTCAGGAAC				
5221	AGGGGCATGA	CTGCCCTGCT	GAAAGTCTCC	AGCTGTGACA	
AGAACACTGG	GGACTACTAT				
5281	GAGGACAGCT	ATGAGGACAT	CTCTGCCCTAC	CTGCTGAGCA	
AGAACAATGC	CATTGAGCCC				
5341	AGGAGCTTCA	GCCAGAACAG	CAGGCACCCC	AGCACCAGGC	
AGAAGCAGTT	CAATGCCACC				
5401	ACCATCCCTG	AGAATGACAT	AGAGAAGACA	GACCCATGGT	
TTGCCACCG	GACCCCATG				
5461	CCCAAGATCC	AGAATGTGAG	CAGCTCTGAC	CTGCTGATGC	
TGCTGAGGCA	GAGCCCCACC				
5521	CCCCATGGCC	TGAGCCTGTC	TGACCTGCAG	GAGGCCAAGT	
ATGAAACCTT	CTCTGATGAC				
5581	CCCAGCCCTG	GGGCCATTGA	CAGCAACAAC	AGCCTGTCTG	
AGATGACCCA	CTTCAGGCC				
5641	CAGCTGCACC	ACTCTGGGGA	CATGGTGTTC	ACCCCTGAGT	
CTGGCCTGCA	GCTGAGGCTG				
5701	AATGAGAAGC	TGGGCACCAC	TGCTGCCACT	GAGCTGAAGA	
AGCTGGACTT	CAAAGTCTCC				
5761	AGCACCAGCA	ACAACCTGAT	CAGCACCATC	CCCTCTGACA	
ACCTGGCTGC	TGGCACTGAC				
5821	AACACCAGCA	GCCTGGGCCC	CCCCAGCATG	CCTGTGCACT	
ATGACAGCCA	GCTGGACACC				
5881	ACCCTGTTTG	GCAAGAAGAG	CAGCCCCCTG	ACTGAGTCTG	
GGGGCCCCCT	GAGCCTGTCT				
5941	GAGGAGAACA	ATGACAGCAA	GCTGCTGGAG	TCTGGCCTGA	
TGAACAGCCA	GGAGAGCAGC				
6001	TGGGGCAAGA	ATGTGAGCAG	CAGGGAGATC	ACCAGGACCA	
CCCTGCAGTC	TGACCAGGAG				
6061	GAGATTGACT	ATGATGACAC	CATCTCTGTG	GAGATGAAGA	
AGGAGGACTT	TGACATCTAC				
6121	GACGAGGACG	AGAACCAGAG	CCCCAGGAGC	TTCCAGAAGA	
AGACCAGGCA	CTACTTCATT				
6181	GCTGCTGTGG	AGAGGCTGTG	GGACTATGGC	ATGAGCAGCA	

## 038966

GCCCCATGT	GCTGAGGAAC				
6241	AGGGCCCAGT	CTGGCTCTGT	GCCCCAGTTC	AAGAAGGTGG	
TGTTCCAGGA	GTTCACTGAT				
6301	GGCAGCTTCA	CCCAGCCCCT	GTACAGAGGG	GAGCTGAATG	
AGCACCTGGG	CCTGTGGGC				
6361	CCCTACATCA	GGGCTGAGGT	GGAGGACAAC	ATCATGGTGA	
CCTTCAGGAA	CCAGGCCAGC				
6421	AGGCCCTACA	GCTTCTACAG	CAGCCTGATC	AGCTATGAGG	
AGGACCAGAG	GCAGGGGGCT				
6481	GAGCCCAGGA	AGAACTTTGT	GAAGCCCAAT	GAAACCAAGA	
CCTACTTCTG	GAAGGTGCAG				
6541	CACCACATGG	CCCCACCAA	GGATGAGTTT	GACTGCAAGG	
CCTGGGCCTA	CTTCTCTGAT				
6601	GTGGACCTGG	AGAAGGATGT	GCACTCTGGC	CTGATTGGCC	
CCCTGCTGGT	GTGCCACACC				
6661	AACACCCTGA	ACCCTGCCCA	TGGCAGGCAG	GTGACTGTGC	
AGGAGTTTGC	CCTGTTCTTC				
6721	ACCATCTTTG	ATGAAACCAA	GAGCTGGTAC	TTCACTGAGA	
ACATGGAGAG	GAATGCAGG				
6781	GCCCCCTGCA	ACATCCAGAT	GGAGGACCCC	ACCTTCAAGG	
AGAACTACAG	GTTCCATGCC				
6841	ATCAATGGCT	ACATCATGGA	CACCCTGCCT	GGCCTGGTGA	
TGGCCCAGGA	CCAGAGGATC				
6901	AGGTGGTACC	TGCTGAGCAT	GGGCAGCAAT	GAGAACATCC	
ACAGCATCCA	CTTCTCTGGC				
6961	CATGTGTTCA	CTGTGAGGAA	GAAGGAGGAG	TACAAGATGG	
CCCTGTACAA	CCTGTACCCT				
7021	GGGGTGTTTG	AGACTGTGGA	GATGCTGCCC	AGCAAGGCTG	
GCATCTGGAG	GGTGGAGTGC				
7081	CTGATTGGGG	AGCACCTGCA	TGCTGGCATG	AGCACCTGT	
TCCTGGTGTA	CAGCAACAAG				
7141	TGCCAGACCC	CCCTGGGCAT	GGCCTCTGGC	CACATCAGGG	
ACTTCCAGAT	CACTGCCTCT				
7201	GGCCAGTATG	GCCAGTGGGC	CCCCAAGCTG	GCCAGGCTGC	
ACTACTCTGG	CAGCATCAAT				
7261	GCCTGGAGCA	CCAAGGAGCC	CTTCAGCTGG	ATCAAGGTGG	

## 038966

ACCTGCTGGC	CCCCATGATC				
7321	ATCCATGGCA	TCAAGACCCA	GGGGCCAGG	CAGAAGTTCA	
GCAGCCTGTA	CATCAGCCAG				
7381	TTCATCATCA	TGTACAGCCT	GGATGGCAAG	AAGTGGCAGA	
CCTACAGGGG	CAACAGCACT				
7441	GGCACCTGA	TGGTGTCTT	TGGCAATGTG	GACAGCTCTG	
GCATCAAGCA	CAACATCTTC				
7501	AACCCCCCA	TCATTGCCAG	ATACATCAGG	CTGCACCCCA	
CCCCTACAG	CATCAGGAGC				
7561	ACCCTGAGGA	TGGAGCTGAT	GGGCTGTGAC	CTGAACAGCT	
GCAGCATGCC	CCTGGGCATG				
7621	GAGAGCAAGG	CCATCTCTGA	TGCCCAGATC	ACTGCCAGCA	
GCTACTTCAC	CAACATGTTT				
7681	GCCACCTGGA	GCCCCAGCAA	GGCCAGGCTG	CACCTGCAGG	
GCAGGAGCAA	TGCCTGGAGG				
7741	CCCCAGGTCA	ACAACCCCAA	GGAGTGGCTG	CAGGTGGACT	
TCCAGAAGAC	CATGAAGGTG				
7801	ACTGGGGTGA	CCACCCAGGG	GGTGAAGAGC	CTGCTGACCA	
GCATGTATGT	GAAGGAGTTC				
7861	CTGATCAGCA	GCAGCCAGGA	TGGCCACCAG	TGGACCTGT	
TCTTCCAGAA	TGGCAAGGTG				
7921	AAGGTGTTC	AGGCAACCA	GGACAGCTTC	ACCCCTGTGG	
TGAACAGCCT	GGACCCCCC				
7981	CTGCTGACCA	GATACCTGAG	GATTCACCCC	CAGAGCTGGG	
TGCACCAGAT	TGCCCTGAGG				
8041	ATGGAGGTGC	TGGGCTGTGA	GGCCCAGGAC	CTGTACTGAG	
CGGCCGCGGG	CCCAATCAAC				
8101	CTCTGGATTA	CAAAATTTGT	GAAAGATTGA	CTGGTATTCT	
TAACTATGTT	GCTCCTTTTA				
8161	CGCTATGTGG	ATACGCTGCT	TTAATGCCTT	TGTATCATGC	
TATTGCTTCC	CGTATGGCTT				
8221	TCATTTTCTC	CTCCTTGAT	AAATCCTGGT	TGCTGTCTCT	
TTATGAGGAG	TTGTGGCCCG				
8281	TTGTCAGGCA	ACGTGGCGTG	GTGTGCACTG	TGTTTGCTGA	
CGCAACCCCC	ACTGGTTGGG				
8341	GCATTGCCAC	CACCTGTCAG	CTCCTTCCG	GGACTTTCGC	

## 038966

TTTCCCCCTC	CCTATTGCCA				
8401	CGGCGGAACT	CATCGCCGCC	TGCCTTGCCC	GCTGCTGGAC	
AGGGGCTCGG	CTGTTGGGCA				
8461	CTGACAATTC	CGTGGTGTG	TCGGGGAAAT	CATCGTCCTT	
TCCTTGCTG	CTCGCCTGTG				
8521	TTGCCACCTG	GATTCTGCGC	GGGACGTCCT	TCTGCTACGT	
CCCTTCGGCC	CTCAATCCAG				
8581	CGGACCTTCC	TTCCCGCGGC	CTGCTGCCGG	CTCTGCCGGC	
TCTTCCGCGT	CTTCGCCTTC				
8641	GCCCTCAGAC	GAGTCGGATC	TCCCTTTGGG	CCGCCTCCCC	
GCAAGCTTCG	CACTTTTTAA				
8701	AAGAAAAGGG	AGGACTGGAT	GGGATTTATT	ACTCCGATAG	
GACGCTGGCT	TGTAACTCAG				
8761	TCTCTTACTA	GGAGACCAGC	TTGAGCCTGG	GTGTTCTGCTG	
GTTAGCCTAA	CCTGGTTGGC				
8821	CACCAGGGGT	AAGGACTCCT	TGGCTTAGAA	AGCTAATAAA	
CTTGCCCTGCA	TTAGAGCTCT				
8881	TACGCGTCCC	GGGCTCGAGA	TCCGCATCTC	AATTAGTCAG	
CAACCATAGT	CCCGCCCCTA				
8941	ACTCCGCCCA	TCCCGCCCCT	AACTCCGCCC	AGTTCGCCCC	
ATTCTCCGCC	CCATGGCTGA				
9001	CTAATTTTTT	TTATTTATGC	AGAGGCCGAG	GCCGCCTCGG	
CCTCTGAGCT	ATTCCAGAAG				
9061	TAGTGAGGAG	GCTTTTTTGG	AGGCCTAGGC	TTTTGCAAAA	
AGCTAACTTG	TTTATTGCAG				
9121	CTTATAATGG	TTACAATAAA	AGCAATAGCA	TCACAAATTT	
CACAAATAAA	GCATTTTTTT				
9181	CACTGCATTC	TAGTTGTGGT	TTGTCCAAAC	TCATCAATGT	
ATCTTATCAT	GTCTGTCCGC				
9241	TTCTCTGCTC	ACTGACTCGC	TGCGCTCGGT	CGTTCGGCTG	
CGGCGAGCGG	TATCAGCTCA				
9301	CTCAAAGGCG	GTAATACGGT	TATCCACAGA	ATCAGGGGAT	
AACGCAGGAA	AGAACATGTG				
9361	AGCAAAAAGGC	CAGCAAAAAGG	CCAGGAACCG	TAAAAAGGCC	
GCGTTGCTGG	CGTTTTTCCA				
9421	TAGGCTCCGC	CCCCCTGACG	AGCATCACAA	AAATCGACGC	

## 038966

TCAAGTCAGA	GGTGGCGAAA			
9481	CCCGACAGGA	CTATAAAGAT	ACCAGGCGTT	TCCCCCTGGA
AGCTCCCTCG	TGCGCTCTCC			
9541	TGTTCCGACC	CTGCCGCTTA	CCGGATACCT	GTCCGCCTTT
CTCCCTTCGG	GAAGCGTGGC			
9601	GCTTTCTCAT	AGCTCACGCT	GTAGGTATCT	CAGTTCGGTG
TAGGTCGTTC	GCTCCAAGCT			
9661	GGGCTGTGTG	CACGAACCCC	CCGTTCAGCC	CGACCGCTGC
GCCTTATCCG	GTAACATATCG			
9721	TCTTGAGTCC	AACCCGGTAA	GACACGACTT	ATCGCCACTG
GCAGCAGCCA	CTGGTAACAG			
9781	GATTAGCAGA	GCGAGGTATG	TAGGCGGTGC	TACAGAGTTC
TTGAAGTGGT	GGCCTAACTA			
9841	CGGCTACACT	AGAAGAACAG	TATTTGGTAT	CTGCGCTCTG
CTGAAGCCAG	TTACCTTCGG			
9901	AAAAAGAGTT	GGTAGCTCTT	GATCCGGCAA	ACAAACCACC
GCTGGTAGCG	GTGGTTTTTT			
9961	TGTTTGCAAG	CAGCAGATTA	CGCGCAGAAA	AAAAGGATCT
CAAGAAGATC	CTTTGATCTT			
10021	TTCTACGGGG	TCTGACGCTC	AGTGGAACGA	AAACTCACGT
TAAGGGATTT	TGGTCATGAG			
10081	ATTATCAAAA	AGGATCTTCA	CCTAGATCCT	TTTAAATTAA
AAATGAAGTT	TTAAATCAAT			
10141	CTAAAGTATA	TATGAGTAAA	CTTGGTCTGA	CAGTTAGAAA
AACTCATCGA	GCATCAAATG			
10201	AAACTGCAAT	TTATTCATAT	CAGGATTATC	AATACCATAT
TTTTGAAAAA	GCCGTTTCTG			
10261	TAATGAAGGA	GAAAACAC	CGAGGCAGTT	CCATAGGATG
GCAAGATCCT	GGTATCGGTC			
10321	TGCGATTCCG	ACTCGTCCAA	CATCAATACA	ACCTATTAAT
TTCCCTTCGT	CAAAAATAAG			
10381	GTTATCAAGT	GAGAAACAC	CATGAGTGAC	GACTGAATCC
GGTGAGAATG	GCAACAGCTT			
10441	ATGCATTTCT	TTCCAGACTT	GTTCAACAGG	CCAGCCATTA
CGCTCGTCAT	CAAAATCACT			
10501	CGCATCAACC	AAACCGTTAT	TCATTCGTGA	TTGCGCCTGA

GCGAGACGAA ATACGCGATC  
 10561 GCTGTTAAAA GGACAATTAC AACAGGAAT CGAATGCAAC  
 CGGCGCAGGA AACTGCCCAG  
 10621 CGCATCAACA ATATTTTCAC CTGAATCAGG ATATTTCTCT  
 AATACCTGGA ATGCTGTTTT  
 10681 TCCGGGGATC GCAGTGGTGA GTAACCATGC ATCATCAGGA  
 GTACGGATAA AATGCTTGAT  
 10741 GGTGGAAGA GGCATAAATT CCGTCAGCCA GTTTAGTCTG  
 ACCATCTCAT CTGTAACATC  
 10801 ATTGGCAACG CTACCTTTGC CATGTTTCAG AAACAACCTCT  
 GCGCATCGG GCTTCCCATA  
 10861 CAATCGATAG ATTGTCGCAC CTGATTGCC GACATTATCG  
 CGAGCCCATT TATACCCATA  
 10921 TAAATCAGCA TCCATGTTGG AATTTAATCG CGGCCTAGAG  
 CAAGACGTTT CCCGTTGAAT  
 10981 ATGGCTCATA ACACCCCTTG TATTACTGTT TATGTAAGCA  
 GACAGTTTTA TTGTTTCATGA  
 11041 TGATATATTT TTATCTTGTG CAATGTAACA TCAGAGATTT  
 TGAGACACAA CAATTGGTTCG  
 11101 ACGGATCC

**SEQ ID NO: 15**

ATGCCAGCTCTGTGTCCTGGGGCATTCTGCTGCTGGCTGGCCTGTGCTGTCTGGTGC  
 CTGTGTCCCTGGCTGAGGACCCCTCAGGGGATGCTGCCAGAAAACAGACACCTCCCACCATG  
 ACCAGGACCACCCACCTTCAACAAGATCACCCCAACCTGGCAGAGTTTGCCTTCAGCCTGT  
 ACAGACAGCTGGCCCACCAGAGCAACAGCACCAACATCTTTTCAGCCCTGTGTCCATTGCCA  
 CAGCCTTTGCCATGCTGAGCCTGGGCACCAAGGCTGACACCCATGATGAGATCCTGGAAGGCC  
 TGAACTTCAACCTGACAGAGATCCCTGAGGCCAGATCCATGAGGGCTTCCAGGAAGTCTGA  
 GAACCTGAACCAGCCAGACAGCCAGCTGCAGCTGACAACAGGCAATGGGCTGTTCCTGTCTG  
 AGGGCCTGAAGCTGGTGGACAAGTTTCTGGAAGATGTGAAGAAGCTGTACCCTCTGAGGCCT  
 TCACAGTGAACCTTGGGGACACAGAAGAGGCCAAGAAACAGATCAATGACTATGTGAAAAGG  
 GCACCCAGGGCAAGATTGTGGACCTTGTGAAAGAGCTGGACAGGGACACTGTGTTGCCCTTG  
 TGAATACATCTTCTTCAAGGGCAAGTGGGAGAGGCCCTTGAAGTGAAGGACACTGAGGAAG  
 AGGACTTCCATGTGGACCAAGTGACCACAGTGAAGGTGCCAATGATGAAGAGACTGGGGATGT  
 TCAATATCCAGCACTGCAAGAACTGAGCAGCTGGGTGCTGCTGATGAAGTACCTGGGCAATG  
 CTACAGCCATATCTTCTGCCTGATGAGGGCAAGCTGCAGCACCTGGAAAATGAGCTGACCC  
 ATGACATCATACCAAATTTCTGGAAAATGAGGACAGAAGATCTGCCAGCCTGCATCTGCCCA

AGCTGAGCATCACAGGCACATATGACCTGAAGTCTGTGCTGGGACAGCTGGGAATCACCAAGG  
TGTTTCAGCAATGGGGCAGACCTGAGTGGAGTGACAGAGGAAGCCCTCTGAAGCTGTCCAAGG  
CTGTGCACAAGGCAGTGTGACCATTGATGAGAAGGGCACAGAGGCTGCTGGGGCCATGTTTC  
TGGAAGCCATCCCATGTCCATCCCCCAGAAGTGAAGTTCAACAAGCCCTTGTGTTCTTGA  
TGATTGAGCAGAACCAAGAGCCCTGTTTCATGGGCAAGTTGTGAACCCACCCAGAAAT  
GA

## SEQ ID NO: 16

ATGCAGATTGAGCTGAGCACCTGCTTCTTCTGTGCCTGCTGAGGTTCTGCTTCTCTG  
CCACCAGGAGATACTACCTGGGGGCTGTGGAGCTGAGCTGGGACTACATGCAGTCTGACCTGG  
GGGAGCTGCCTGTGGATGCCAGGTTCCCCCAGAGTGCCCAAGAGCTTCCCTTCAACACCT  
CTGTGGTGTACAAGAAGACCTGTTTGTGGAGTTCAGTACCACCTGTTCAACATTGCCAAGC  
CCAGGCCCCCTGGATGGGCTGCTGGGCCCCACCATCCAGGCTGAGGTGTATGACACTGTGG  
TGATCACCTGAAGAACATGGCCAGCCACCTGTGAGCCTGCATGCTGTGGGGGTGAGCTACT  
GGAAGCCCTCTGAGGGGGCTGAGTATGATGACCAGACCAGCCAGAGGGAGAAGGAGGATGACA  
AGGTGTTCCCTGGGGGCAGCCACACCTATGTGTGGCAGGTGCTGAAGGAGAATGGCCCCATGG  
CCTCTGACCCCTGTGCCTGACCTACAGCTACCTGAGCCATGTGGACCTGGTGAAGGACCTGA  
ACTCTGGCCTGATTGGGGCCCTGCTGGTGTGCAGGGAGGGCAGCCTGGCCAAGGAGAAGACCC  
AGACCCTGCACAAGTTCATCTGCTGTTTGTGTTGATGAGGGCAAGAGCTGGCACCTTG  
AAACCAAGAACAGCCTGATGCAGGACAGGATGCTGCCTTGCCAGGGCTGGCCAAGATGC  
ACACTGTGAATGGTATGTGAACAGGAGCCTGCCTGGCCTGATTGGCTGCCACAGGAAGTCTG  
TGTACTGGCATGTGATTGGCATGGGCACCACCCCTGAGGTGCACAGCATCTTCTGGAGGGCC  
ACACCTTCTGGTCAGGAACCACAGGCAGGCCAGCTGGAGATCAGCCCCATCACCTTCTTGA  
CTGCCAGACCCTGCTGATGGACCTGGGCCAGTTCTGCTGTTCTGCCACATCAGCAGCCACC  
AGCATGATGGCATGGAGGCCATGTGAAGGTGGACAGCTGCCCTGAGGAGCCCCAGCTGAGGA  
TGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATGGATGTGG  
TGAGGTTTGTGATGATGACAACAGCCCCAGCTTCATCCAGATCAGGTCTGTGGCCAAGAAGCACC  
CCAAGACCTGGGTGCACTACATTGCTGCTGAGGAGGAGGACTGGGACTATGCCCCCTGGTGC  
TGCCCCCTGATGACAGGAGCTACAAGAGCCAGTACCTGAACAATGGCCCCAGAGGATTGGCA  
GGAAGTACAAGAAGGTCAGGTTTCATGGCCTACACTGATGAAACCTTCAAGACCAGGGAGGCCA  
TCCAGCATGAGTCTGGCATCCTGGGCCCCCTGCTGTATGGGGAGGTGGGGACACCTGTGTA  
TCATCTTCAAGAACCAGGCCAGCAGGCCCTACAACATCTACCCCATGGCATCACTGATGTGA  
GGCCCCGTGACAGCAGGAGGCTGCCAAGGGGGTGAAGCACCTGAAGGACTTCCCCATCTGTC  
CTGGGGAGATCTTCAAGTACAAGTGGACTGTGACTGTGGAGGATGGCCCCACCAAGTCTGACC  
CCAGGTGCTGACCAGATACTACAGCAGCTTTGTGAACATGGAGAGGGACCTGGCCTCTGGCC  
TGATTGGCCCCCTGCTGATCTGCTACAAGGAGTCTGTGGACCAGAGGGCAACCAGATCATGT  
CTGACAAGAGGAATGTGATCTGTTCTGTGTTTGTGATGAGAACAGGAGCTGGTACCTGACTG



AGAACATCCAGAGTTCTCGCCCAACCCTGCTGGGGTGCAGCTGGAGGACCCTGAGTTCAGG  
CCAGCAACATCATGCACAGCATCAATGGCTATGTGTTTGACAGCCTGCAGCTGTCTGTGTGCC  
TGCATGAGGTGGCTACTGGTACATCCTGAGCATTGGGGCCAGACTGACTTCCTGTCTGTGT  
TCTTCTCTGGCTACACCTTCAAGCACAAGATGGTGTATGAGGACACCCTGACCCTGTCCCT  
TCTCTGGGGAGACTGTGTTTCATGAGCATGGAGAACCCTGGCCTGTGGATTCTGGGCTGCCACA  
ACTCTGACTTCAGGAACAGGGGCATGACTGCCCTGCTGAAAGTCTCCAGCTGTGACAAGAACA  
CTGGGGACTACTATGAGGACAGCTATGAGGACATCTCTGCCTACCTGCTGAGCAAGAACAATG  
CCATTGAGCCCAGGAGCTTCAGCCAGAACAGCAGGCACCCAGCACCAGGCAGAAGCAGTTCA  
ATGCCACCACCATCCCTGAGAATGACATAGAGAAGACAGACCCATGGTTTGCCACCAGGACCC  
CCATGCCCAAGATCCAGAATGTGAGCAGCTCTGACCTGCTGATGCTGCTGAGGCAGAGCCCCA  
CCCCCATGGCCTGAGCCTGTCTGACCTGCAGGAGGCCAAGTATGAAACCTTCTCTGATGACC  
CCAGCCCTGGGGCCATTGACAGCAACAACAGCCTGTCTGAGATGACCACCTCAGGCCCCAGC  
TGCACCACCTCTGGGGACATGGTGTTCACCCCTGAGTCTGGCCTGCAGCTGAGGCTGAATGAGA  
AGCTGGGCACCCTGCTGCCACTGAGCTGAAGAAGCTGGACTTCAAAGTCTCCAGCACCAGCA  
ACAACCTGATCAGCACCATCCCTCTGACAACCTGGCTGCTGGCACTGACAACACCAGCAGCC  
TGGGCCCCCAGCATGCCTGTGCACTATGACAGCCAGCTGGACACCACCTGTTTGCAAGA  
AGAGCAGCCCCCTGACTGAGTCTGGGGCCCCCTGAGCCTGCTGAGGAGAACAATGACAGCA  
AGCTGCTGGAGTCTGGCCTGATGAACAGCCAGGAGCAGCTGGGGCAAGAATGTGAGCAGCA  
GGGAGATCACCAGGACCCTGCAGTCTGACCAGGAGGAGATTGACTATGATGACACCATCT  
CTGTGGAGATGAAGAAGGAGGACTTTGACATCTACGACGAGGACGAGAACCAGAGCCCCAGGA  
GCTTCCAGAAGAAGACCAGGCACTACTTTCATTGCTGCTGTGGAGAGGCTGTGGGACTATGGCA  
TGAGCAGCAGCCCCATGTGCTGAGGAACAGGGCCAGTCTGGCTCTGTGCCCCAGTTCAAGA  
AGGTGGTGTTCAGGAGTTCACTGATGGCAGCTTCACCCAGCCCTGTACAGAGGGGAGCTGA  
ATGAGCACCTGGGCCTGCTGGGCCCTACATCAGGGCTGAGGTGGAGGACAACATCATGGTGA  
CCTTCAGGAACCAGGCCAGCAGGCCCTACAGCTTCTACAGCAGCCTGATCAGCTATGAGGAGG  
ACCAGAGGCAGGGGGCTGAGCCCAGGAAGAACTTTGTGAAGCCCAATGAAACCAAGACCTACT  
TCTGGAAGGTGCAGCACCACATGGCCCCACCAAGGATGAGTTTGACTGCAAGGCTGGGCCT  
ACTTCTCTGATGTGGACCTGGAGAAGGATGTGCACCTCTGGCCTGATTGGCCCCCTGCTGGTGT  
GCCACACCAACACCTGAACCTGCCCATGGCAGGAGGTGACTGTGCAGGAGTTTGCCTGT  
TCTTCACCATCTTTGATGAAACCAAGAGCTGGTACTTCACTGAGAACATGGAGAGGAATGCA  
GGGCCCCCTGCAACATCCAGATGGAGGACCCACCTCAAGGAGAACTACAGGTTCCATGCCA  
TCAATGGCTACATCATGGACACCCTGCCTGGCCTGGTGTGCCCCAGGACCAGAGGATCAGGT  
GGTACCTGCTGAGCATGGGCAGCAATGAGAACATCCACAGCATCCACTTCTCTGGCCATGTGT  
TCACTGTGAGGAAGAAGGAGGAGTACAAGATGGCCCTGTACAACCTGTACCCTGGGGTGTG  
AGACTGTGGAGATGCTGCCAGCAAGGCTGGCATCTGGAGGGTGGAGTGCCCTGATGGGGAGC  
ACCTGCATGCTGGCATGAGCACCCCTGTTCTGGTGTACAGCAACAAGTGCCAGACCCCCCTGG

GCATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCTGGCCAGTATGGCCAGTGGG  
 CCCCCAAGCTGGCCAGGCTGCACTACTCTGGCAGCATCAATGCCTGGAGCACCAGGAGCCCT  
 TCAGCTGGATCAAGGTGGACCTGTGGCCCCATGATCATCCATGGCATCAAGACCCAGGGGG  
 CCAGGCAGAAGTTCAGCAGCCTGTACATCAGCCAGTTCATCATCATGTACAGCCTGGATGGCA  
 AGAAGTGGCAGACCTACAGGGGCAACAGCAGTGGCACCCCTGATGGTGTCTTTGGCAATGTGG  
 ACAGTCTGGCATCAAGCACAACATCTTCAACCCCCCATCATGGCAGATACATCAGGCTGC  
 ACCCCACCCACTACAGCATCAGGAGCACCCCTGAGGATGGAGCTGATGGGCTGTGACCTGAACA  
 GCTGCAGCATGCCCTGGGCATGGAGAGCAAGGCCATCTCTGATGCCAGATCACTGCCAGCA  
 GCTACTTCACCAACATGTTTGGCCCTGGAGCCCCAGCAAGGCCAGGCTGCACCTGCAGGGCA  
 GGAGCAATGCCTGGAGGCCCAAGTCAACAACCCCAAGGAGTGGCTGCAGGTGGACTTCCAGA  
 AGACCATGAAGGTGACTGGGGTGACCACCCAGGGGGTGAAGAGCCTGCTGACCAGCATGTATG  
 TGAAGGAGTTCCTGATCAGCAGCAGCCAGGATGGCCACCAGTGGACCCCTGTTCTTCCAGAATG  
 GCAAGGTGAAGGTGTTCCAGGGCAACCAGGACAGCTTCAACCCCTGTGGTGAACAGCCTGGACC  
 CCCCCCTGCTGACCAGATACCTGAGGATTCACCCCAGAGCTGGGTGCACCAGATTGCCCTGA  
 GGATGGAGGTGCTGGGCTGTGAGGCCAGGACCTGTACTGA

**SEQ ID NO: 17**

CCGCGGAGATCTCAATATTGGCCATTAGCCATATTATTCATTGGTTATATAGCATAAA  
 TCAATATTGGCTATTGGCCATTGCATACGTTGTATCTATATCATAATATGTACATTTATATTG  
 GCTCATGTCCAATATGACCGCATGTTGGCATTGATTATTGACTAGTTATTAATAGTAATCAA  
 TTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCGCGTTACATAACTTACGGTAAATG  
 GCCCCCTGGCTGACCGCCCAACGACCCCGCCCATTGACGTCAATAATGACGTATGTTCCCA  
 TAGTAACGCCAATAGGACTTTCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCC  
 ACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTA  
 AATGGCCCGCCTGGCATTATGCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACA  
 TCTACGTATTAGTCATCGCTATTACCATGGTGTGCGGTTTTGGCAGTACACCAATGGCGTG  
 GATAGCGTTTTGACTCACGGGATTTCCAAGTCTCCACCCATTGACGTCAATGGGAGTTTTGT  
 TTTGGCACCAAAATCAACGGGACTTTCAAAATGTCGTAATAACCCGCCCCGTTGACGCAAA  
 TGGCCGGTAGGCGGTGACGGTGGGAGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGA  
 TCACTAGAAGCTTTATTGGCGTAGTTTATCACAGTTAAATGCTAACGCAGTCAGTGCTTCTG  
 ACACAACAGTCTCGAACTTAAGCTGCAGAAGTTGGTCGTGAGGCACCTGGGCAGGCTAGC

**SEQ ID NO: 18**

TCGAGATGTGGTCTGAGTTAAAAATCAGGAGCAACGACGGAGGTGAAGGACCAGACGC  
 CAACGACCC

**SEQ ID NO: 19**

CCGGGGTTCGTTGGCGTCTGGTCCCTCACCTCCGTCGTTGCTCCTGATTTTTAACTCA  
 GACCACATC

SEQ ID NO: 20

CCGGGAAAGGGGTGCAACACATCCATATCCAGCCATCTCTACCTGTTTATGGACA

SEQ ID NO: 21

ACCCTCTGTCCATAAACAGGTAGAGATGGCTGGATATGGATGTGTTGCACCCCTTTCC

SEQ ID NO: 22

GGGTTAGGTGGTTGCTGATTCTCTCATTACCCAGTGGG

SEQ ID NO: 23

GATCCCCACTGGGTGAATGAGAGAATCAGCAACCACCTA

SEQ ID NO: 24

GAGACTCGAGATGTGGTCTGAGTTAAAAATCAGG

SEQ ID NO: 25

AGAGGTAGACCAGTACGAGTCACGTTTGCCCTATCACCATCCCTAACCCCTGTGCAT  
AAAC

SEQ ID NO: 26

TACGGGTCGAGACACAGGACCCCGTAAGACGACGACCGACCCGGACACGACAGACCAG  
GACACAGGACCGACTCCTGGGAGTCCCCCTACGACGGGTCTTTTGTCTGTGGAGGTGGTAC  
TGGTCTGTGGGGTGGAAAGTTGTCTAGTGGGGTTGGACCGTCTCAAACGGAAGTCGGACA  
TGTCTGTGACCGGGTGGTCTCGTTGTCGTGGTTGTAGAAAAAGTCGGGACACAGGTAACGGT  
GTCGAAAACGGTACGACTCGGACCCGTGGTCCGACTGTGGGTACTACTCTAGGACCTCCGG  
ACTTGAAGTTGGACTGTCTCTAGGGACTCCGGGTCTAGGTACTCCCGAAGGTCCTTGACGACT  
CTTGGGACTTGGTCGGTCTGTGCGTCGACGTCGACTGTTGTCGGTTACCCGACAAGGACAGAC  
TCCCGGACTTCGACCACCTGTTCAAAGACCTTCTACACTTCTTCGACATGGTGAGACTCCGGA  
AGTGTCACTTGAACCCCTGTGTCTTCTCCGGTCTTTGTCTAGTTACTGATACACCTTTTCC  
CGTGGGTCCCGTTCTAACACCTGGAACACTTTCTCGACCTGTCCCTGTGACACAAACGGGAAC  
ACTTGATGTAGAAGAAGTTCCCGTTCACCCCTCCTCCGGAAACTTCACTTCCTGTGACTCCTTC  
TCCTGAAGGTACACCTGGTTCACTGGTGTCACTCCACGGTTACTACTTCTCTGACCCCTACA  
AGTTATAGGTTCGTGACGTTCTTTGACTCGTCGACCCACGACGACTACTTCATGGACCCGTTAC  
GATGTCGGTATAAGAAAGACGGACTACTCCCGTTCGACGTCGTGGACCTTTTACTCGACTGGG  
TACTGTAGTAGTGGTTTAAAGACCTTTTACTCCTGTCTTCTAGACGGTCGGACGTAGACGGT  
TCGACTCGTAGTTCGGTGTACTGGACTTCAGACACGACCCGTTCGACCCTTAGTGGTTCC  
ACAAGTCGTTACCCCGTCTGGACTCACCTCACTGTCTCCTTCGGGGAGACTTCGACAGGTTCC  
GACACGTGTTCCGTCACGACTGGTAACTACTCTTCCCGTGTCTCCGACGACCCCGGTACAAAG  
ACCTTCGGTAGGGGTACAGGTAGGGGGTCTTCACTTCAAGTTGTTCCGGAAACACAAGGACT  
ACTAACTCGTCTTGTGGTTCTCGGGGACAAGTACCCGTTCCAACACTGGGGTGGGTCTTTA  
CT

SEQ ID NO: 27

AEDPQGDAAQKTDTSHHDDHPTFAEDPQGDAAQKTDTSHHDDHPTFNKITPNLAEF  
 AFSLYRQLAHQSNSTNIFFSPVSIATAFAMLSLGTKADTHDEILEGLNFNLTEIPEAQIHEGF  
 QELLRTLNPDSQLQLTGTGNLFLSEGLKLVDFLEDVKKLYHSEAFVNFQDTEEAQKQIND  
 YVEKGTQGKIVDLVKELEDRDVFALVNYIFFKQWERPFVVKDTEEDDFHVDQVTTVKVPMK  
 RLGMFNIQHCKKLSWVLLMKYLGNATAIFFLPDEGKQLHLENELTHDIIITKFLENEDRRSAS  
 LHLPKLSITGTLDKSVLGLGITKVFNSGADLSGVTEEAPLKLKAVHKAVLTIIDEKGTAA  
 GAMFLEAIPMSIPPEVKFNKPFVFLMIEQNTKSPLFMGKVVNPTQK

**SEQ ID NO: 28**

TACGTCTAACTCGACTCGTGGACGAAGAAGACACGGACGACTCCAAGACGAAGAGAC  
 GGTGGTCTCTA  
 TGATGGACCCCCGACACCTCGACTCGACCTGATGTACGTGAGACTGGACCCCCCTCGA  
 CCGACACCTACG  
 TGCCAAGGGGGGTCTCACGGTTCTCGAAGGGGAAGTTGTGGAGACACCACATGTTC  
 TTCTGGGACAAA  
 CACCTCAAGTGACTGGTGGACAAGTTGTAAACGGTTCGGGTCCGGGGGACCTACCCGG  
 ACGACCCGGGGT  
 GGTAGGTCCGACTCCACATACTGTGACACCACTAGTGGGACTTCTTGTACCGGTCCGGT  
 GGGACACTCGGA  
 CGTACGACACCCCACTCGATGACCTTCCGGAGACTCCCCGACTCATACTACTGGTC  
 TGGTCCGTCTCC  
 CTCTTCTCTACTGTTCCACAAGGGACCCCGTCCGGTGTGGATACACACCCGTCCACG  
 ACTTCTCTTAC  
 CGGGTACCGGAGACTGGGGGACACGGACTGGATGTCGATGGACTCGGTACACCTGGA  
 CCACTTCCTGGA  
 CTTGAGACCGGACTAACCCGGGACGACCACACGTCCTCCCGTCCGGACCGGTCCCTC  
 TTCTGGGTCTGG  
 GACGTGTTCAAGTAGGACGACAAACGACACAACTACTCCCGTTCFCGACCGTGAGAC  
 TTTGGTCTTGT  
 CGGACTACGTCCTGTCCCTACGACGGAGACGGTCCCGACCGGGTCTACGTGTGACA  
 CTTACCGATAACA  
 CTTGTCTCCGGACCGGACTAACCAGCGGTGTCCTTCAGACACATGACCGTACAC  
 TAACCGTACCCG  
 TGGTGGGACTCCACGTGTCGTAGAAGGACCTCCCGGTGTGGAAGGACCGTCTTGG  
 TGTCCGTCCGGT  
 CGGACCTCTAGTCGGGGTAGTGGGAAGGACTGACGGGTCTGGGACGACTACCTGGACCC  
 GGTC AAGGACGA

CAAGACGGTGTAGTCGTCGGTGGTCTACTACCGTACCTCCGGATACACTTCCACCTG  
TCGACGGGACTC  
CTCGGGGTCGACTCCTACTTCTTGTACTCCTCCGACTCCTGATACTACTACTGGACT  
GACTGAGACTCT  
ACCTACCCACTCCAAACTACTACTGTTGTCGGGGTCGAAGTAGGTCTAGTCCAGACA  
CCGGTTCCTCGT  
GGGGTTCGGACCCACGTGATGTAACGACGACTCCTCCTCCTGACCCTGATACGGGGG  
GACCACGACCGG  
GGACTACTGTCTCGATGTTCTCGGTCATGGACTTGTACCGGGGGTCTCCTAACCGT  
CCTTCATGTTCT  
TCCAGTCCAAGTACCGGATGTGACTACTTTGGAAGTTCGGTCCCTCCGGTAGGTCGT  
ACTCAGACCGTA  
GGACCCGGGGGACGACATACCCCTCCACCCCTGTGGGACGACTAGTAGAAGTTCTTG  
GTCCGGTCGTCC  
GGGATGTTGTAGATGGGGGTACCGTAGTGACTACACTCCGGGGACATGTCGTCCCTCCG  
ACGGGTTCCTCC  
ACTTCGTGGACTTCTGAAGGGGTAGGACGGACCCCTCTAGAAGTTCATGTTACCTG  
ACACTGACACCT  
CCTACCGGGTGGTTCAGACTGGGGTCCACGGACTGGTCTATGATGTCGTGAAACAC  
TTGTACCTCTCC  
CTGGACCGGAGACCGGACTAACCGGGGACGACTAGACGATGTTCTCAGACACCTGG  
TCTCCCGTTGG  
TCTAGTACAGACTGTTCTCCTTACACTAGGACAAGAGACACAACTACTCTTGCTCTC  
GACCATGGACTG  
ACTCTTGTAGGTCTCCAAGGACGGGTGGGACGACCCACGTCGACCTCCTGGGACTC  
AAGGTCCGGTCG  
TTGTAGTACGTGTCGTAGTTACCGATACACAACTGTCGGACGTCGACAGACACCGG  
ACGTACTCCACC  
GGATGACCATGTAGGACTCGTAACCCCGGGTCTGACTGAAGGACAGACACAAGAAGAG  
ACCGATGTGGAA  
GTTTCGTGTTCTACCACATACTCCTGTGGGACTGGGACAAGGGGAAGAGACCCCTCTGA  
CACAAGTACTCG  
TACCTCTTGGGACCGGACACCTAAGACCCGACGGTGTGAGACTGAAGTCTTGTCCC  
CGTACTGACGGG  
ACGACTTTCAGAGGTCGACACTGTTCTTGTGACCCCTGATGATACTCCTGTGATACT  
CCTGTAGAGACG

GATGGACGACTCGTTCTTGTACGGTAACTCGGGTCCCGAAGTCGGTCTTGTCTGCC  
GTGGGGTCGTGG  
TCCGTCTTCGTCAAGTTACGGTGGTGGTAGGGACTCTTACTGTATCTTCTGTCTGG  
GTACCAAACGGG  
TGGCCTGGGGTACGGGTCTAGGTCTTACACTCGTCGAGACTGGACGACTACGACGA  
CTCCGTCTCGGG  
GTGGGGGTACCGGACTCGGACAGACTGGACGTCTCCGGTTCATACTTTGGAAGAGA  
CTACTGGGGTCG  
GGACCCGGTAACTGTCGTTGTTGTCGGACAGACTCTACTGGGTGAAGTCCGGGGTCG  
ACGTGGTGAGAC  
CCCTGTACCACAAGTGGGGACTCAGACCGGACGTGACTCCGACTTACTCTTCGACCC  
GTGGTGACGACG  
GTGACTCGACTTCTTCGACCTGAAGTTTCAGAGGTCGTGGTCGTTGTTGGACTAGTCG  
TGGTAGGGGAGA  
CTGTTGGACCGACGACCGTACTGTTGTTGGTCGTCGGACCCGGGGGGTTCGTACGGAC  
ACGTGATACTGT  
CGGTCGACCTGTGGTGGGACAAACCGTTCTTCTCGTCGGGGACTGACTCAGACCCCC  
GGGGGACTCGGA  
CAGACTCCTTGTACTGTCGTTTCGACGACCTCAGACCGGACTACTTGTCCGGTCCCT  
TCGTCGACCCCG  
TTCTTACACTCGTCGTCCTCTAGTGGTCCTGGTGGGACGTCAGACTGGTCCCTCTCT  
AACTGATACTAC  
TGTGGTAGAGACACCTTACTTCTTCTCCTGAAACTGTAGATGCTGCTCCTGTCTT  
GGTCTCGGGTC  
CTCGAAGGTCTTCTTGTGGTCCGTGATGAAGTACGACGACACCTCTCCGACACCCGT  
ATACCGTACTCG  
TCGTCGGGGTACACGACTCCTTGTCCCGGTGACACCGAGACACGGGGTCAAGTTCT  
TCCACCACAAGG  
TCCTCAAGTGACTACCGTCAAGTGGGTCGGGACATGTCTCCCTCGACTTACTCGT  
GGACCCGGACGA  
CCCGGGATGTAGTCCCGACTCCACCTCCTGTTGTAGTACACTGGGAAGTCTTGGTC  
CGGTCTCGGGG  
ATGTCGAAGATGTCGTCGGACTAGTCGATACTCCTCCTGGTCTCCGTCCCCGACTCG  
GGTCTTCTTGA  
AACACTTCGGGTACTTTGGTTCGGATGAAGACCTTCCACGTCGTGGTGTACCGGGG  
GTGGTTCCTACT

CAAACCTGACGTTCCGGACCCGGATGAAGAGACTACACCTGGACCTTCTCTACACGTG  
AGACCCGGACTAA  
CCGGGGGACGACCACACGGTGTGGTTGTGGGACTTGGGACGGGTACCGTCCGTCCACT  
GACACGTCTCTCA  
AACGGGACAAGAAGTGGTAGAACTACTTTGGTTCTCGACCATGAAGTGACTCTTGTA  
CCTCTCCTTGAC  
GTCCCGGGGACGTTGTAGGTCTACCTCCTGGGGTGAAGTTCCTCTTGATGTCCAAG  
GTACGGTAGTTA  
CCGATGTAGTACCTGTGGGACGGACCGGACCACTACCGGGTCTGGTCTCTAGTCCA  
CCATGGACGACT  
CGTACCCGTCGTTACTCTTGTAGGTGTCGTAGGTGAAGAGACCGGTACACAAGTGACA  
CTCCTTCTTCTCT  
CCTCATGTTCTACCGGGACATGTTGGACATGGGACCCCAAACTCTGACACCTCTAC  
GACGGGTCGTTT  
CGACCGTAGACCTCCACCTCACGGACTAACCCTCGTGGACGTACGACCGTACTCGT  
GGGACAAGGACC  
ACATGTCGTTGTTACGGTCTGGGGGACCCGTACCGGAGACCGGTGTAGTCCCTGAA  
GGTCTAGTGACG  
GAGACCGGTCATACCGGTCACCCGGGGTTCGACCGGTCCGACGTGATGAGACCGTCG  
TAGTTACGGACC  
TCGTGGTTCTCGGGAAGTCGACCTAGTTCACCTGGACGACCGGGGTACTAGTAGG  
TACCGTAGTTCT  
GGGTCCCCCGGTCCGTCTTCAAGTCGTCGGACATGTAGTCGGTCAAGTAGTAGTACAT  
GTCGGACCTACC  
GTTCTTCACCGTCTGGATGTCCCGTTGTCGTGACCGTGGGACTACCACAAGAAACCG  
TTACACCTGTCG  
AGACCGTAGTTTCGTGTTGTAGAAGTTGGGGGGTAGTAACGGTCTATGTAGTCCGACG  
TGGGGTGGGTGA  
TGTGCTAGTCCCTCGTGGGACTCCTACCTCGACTACCCGACACTGGACTTGTGACGTC  
GTACGGGGACCC  
GTACCTCTCGTTCGGGTAGAGACTACGGGTCTAGTGACGGTTCGATGAAGTGGTTG  
TACAAACGGTGG  
ACCTCGGGGTCGTTCCGGTCCGACGTGGACGTCCCGTCTCGTTACGGACCTCCGGGG  
TCCAGTTGTTGG  
GGTTCCTACCGACGTCCACCTGAAGGTCTTCTGGTACTTCCACTGACCCCACTGGTG  
GGTCCCCACTT

CTCGGACGACTGGTCGTACATACACTTCTCAAGGACTAGTCGTCGTCGGTCTACCG  
GTGGTCACCTGG

GACAAGAAGGTCTTACCGTCCACTTCCACAAGGTCCCCTGGTCTGTGCGAAGTGGG  
GACACCACTGT

CGGACCTGGGGGGGACGACTGGTCTATGGACTCCTAAGTGGGGTCTCGACCCACGT  
GGTCTAACGGGA

CTCTACCTCCAGACCCGACACTCCGGGTCTGGACATGACT

**SEQ ID NO: 29**

MQIELSTCFFLCLRFCFSATRRYYLGAVELSWDYMQSDLGELPVDARFPFRVPSFP  
FNTSVVYKKTFLVFEFTDHLFNI AKPRPPWMGLGPTIQAEVYDTVVITLKNMASHPVS LHAVG  
VSYWKASEGAEYDDQTSQREKEDDKVFP GGSHTYVWQVLKENGPMASDPLCLTYSYLSHVDLV  
KDLNSGLIGALLVCREGLAKEKTQTLHKFILLFAVFDEGKSWHSETKNSLMQDRDAASARAW  
PKMHTVNGYVNRSLPGLIGCHRKS VYWHVIGMGTTP EVHSIFLEGHTFLVRNHRQASLEISPI  
TFLTAQTLMLDGLQFLFCHISSHQHDGMEAYVKVDSCEPEPQLRMKNNEEAEDYDDDLTDSE  
MDVVRFDNNSPSFIQIRSVAKKHPKTWVHYIAAEEEDWDYAPLV LAPDDR SYKSQYLNNGPQ  
RIGRKYKVRFMAYTDETFKTREAIQHESGILGPLYGEVGDLLIIFKNQASRPYNIYPHGI  
TDVRLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRYSSFVNMRDL  
ASGLIGPLLCYKESVDQRGNQIMSDKRNVI LFSVFDENRSWYL TENIQRFLPNPAGVQLEDP  
EFQASNIMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKHKMVYEDTLT  
LFFFSGETVFMSENPGLWILGCHNSDFRNRGMTALLKVSSCDKNTGDYEDSYEDISAYLLS  
KNAIEPRSFQNSRHPSTRKQKFNATTIPENDIEKTD PWF AHRT PMPKI QNVSSDLLMLLR  
QSPTPHGLSLSDLQEAKEYETFSDDPSPGAIDSNNLSSEMTHFRPQLHHS GDMVFT PESGLQLR  
LNEKLGTTAATELKKLDFKVSSTSNLI STIPSDNLAAGTDNTSSLGPPSPVHYDSQLDTTL  
FGKKSPLTESGGPLSLSEENNSKLLSEGLMNSQESSWGKNVSSREITR TTLQSDQEEIDYD  
DTISVEMKKEDFDIYDEDENQSPRSFQKKTRHYFIAAVERLWDYGMSSSPHVLNRNRAQSGSVP  
QFKKVVFEFTDGSFTQPLYRGELNEHLG L L G P Y I R A E V E D N I M V T F R N Q A S R P Y S F Y S S L I S  
YEEDQRQGAEPKRFVKNETKTYFWKVQHMAPTKDEFCKAWAYFSDVDLEKDVHSGLIGP  
LLVCHTNTLNPAHGRQVTQVEFALFFTI FDETKSWYFTENMERNCRAPCNI QMEDPTFKENYR  
FHAINGYIMDTLPGLVMAQDQRIRWYLLSMGSNENIHS IHFSGHVFTVRKKEEYKMALYNLYP  
GVFETVEMLPSKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTP LGMASGHIRD FQITASGQY  
GQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMI IHGIKTQGARQKFSSLYISQFI IMYS  
LDGKKWQTYRGNSTGTLMVFFGNVDSSGIKHNI FNPP I IARYIRLHPHYSIRSTLRMELMGC  
DLNSCSMPLGMESKAISDAQITASSYFTNM FATWSPSKARLHLQGRSNAWRPQVNNPKEWLQV  
DFQKTMKVTGVTQGVKSLTSMYVKEFLISSQDGHQWTLFFQNGKVKVFGNQDSFTPVVN  
SLDPPLLTRYLRIHQPQSWVHQIALRMEVLGCEAQDLY

**SEQ ID NO: 30**



ATGCAGATTGAGCTGAGCACCTGCTTCTCCTGTGCCTGCTGAGGTTCTGCTTCTCTG  
CCACCAGGAGATACTACCTGGGGCTGTGGAGCTGAGCTGGGACTACATGCAGTCTGACCTGG  
GGGAGCTGCCTGTGGATGCCAGGTTCCCCCAGAGTGCCCAAGAGCTTCCCCTTCAACACCT  
CTGTGGTGTACAAGAAGACCCTGTTTGTGGAGTTCACTGACCACCTGTCAACATTGCCAAGC  
CCAGGCCCCCTGGATGGGCTGCTGGGCCCCACCATCCAGGCTGAGGTGTATGACACTGTGG  
TGATCACCTGAAGAACATGGCCAGCCACCCTGTGAGCCTGCATGCTGTGGGGGTGAGCTACT  
GGAAGGCCTCTGAGGGGGCTGAGTATGATGACCAGACCAGCCAGAGGAGAAGGAGGATGACA  
AGGTGTTCCCCTGGGGCAGCCACACCTATGTGTGGCAGGTGCTGAAGGAGAATGGCCCCATGG  
CCTCTGACCCCCCTGTGCCTGACCTACAGCTACCTGAGCCATGTGGACCTGGTGAAGGACCTGA  
ACTCTGGCCTGATTGGGGCCTGCTGGTGTGCAGGAGGGCAGCCTGGCCAAGGAGAAGACCC  
AGACCCTGCACAAGTTCATCCTGTGTGTTGTGTGTTGATGAGGGCAAGAGCTGGCACCTG  
AAACCAAGAACAGCCTGATGCAGGACAGGGATGCTGCCTCTGCCAGGGCCTGGCCCAAGATGC  
ACACTGTGAATGGCTATGTGAACAGGAGCCTGCCTGGCCTGATTGGCTGCCACAGGAAGTCTG  
TGTACTGGCATGTGATTGGCATGGGCACCACCCTGAGGTGCACAGCATCTTCTGGAGGGCC  
ACACCTTCTGGTCAAGAACACAGCCAGCCAGCCTGGAGATCAGCCCCATCACCTTCTGTA  
CTGCCCAGACCCTGCTGATGGACCTGGGCCAGTTCCTGCTGTTCTGCCACATCAGCAGCCACC  
AGCATGATGGCATGGAGGCCTATGTGAAGGTGGACAGCTGCCCTGAGGAGCCCCAGCTGAGGA  
TGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATGGATGTGG  
TGAGGTTTGTATGATGACAACAGCCCCAGCTTCATCCAGATCAGGTCTGTGGCCAAGAAGCACC  
CCAAGACCTGGGTGCACTACATTGCTGCTGAGGAGGAGGACTGGGACTATGCCCCCTGGTGC  
TGGCCCCTGATGACAGGAGCTACAAGAGCCAGTACCTGAACAATGGCCCCAGAGGATTGGCA  
GGAAGTACAAGAAGGTCAGGTTTCATGGCCTACACTGATGAAACCTTCAAGACCAGGGAGGCCA  
TCCAGCATGAGTCTGGCATCCTGGGCCCCCTGCTGTATGGGGAGGTGGGGACACCCTGCTGA  
TCATCTTCAAGAACCAGGCCAGCAGGCCCTACAACATCTACCCCCATGGCATCACTGATGTGA  
GGCCCCGTACAGCAGGAGGCTGCCAAGGGGGTGAAGCACCTGAAGGACTTCCCCATCTGTC  
CTGGGGAGATCTTCAAGTACAAGTGGACTGTGACTGTGGAGGATGGCCCCACCAAGTCTGACC  
CCAGGTGCCTGACCAGATACTACAGCAGCTTTGTGAACATGGAGAGGACCTGGCCTCTGGCC  
TGATTGGCCCCCTGCTGATCTGCTACAAGGAGTCTGTGGACCAGAGGGGCAACCAGATCATGT  
CTGACAAGAGGAATGTGATCCTGTCTCTGTGTTTGTGATGAGAACAGGAGCTGGTACCTGACTG  
AGAACATCCAGAGGTTCCCTGCCAACCCTGCTGGGGTGCAGCTGGAGGACCCTGAGTTCAGG  
CCAGCAACATCATGCACAGCATCAATGGCTATGTGTTTGACAGCCTGCAGCTGTCTGTGTGCC  
TGCATGAGGTGGCCTACTGGTACATCCTGAGCATTGGGGCCAGACTGACTTCTGTCTGTGT  
TCTTCTCTGGCTACACCTTCAAGCACAAGATGGTGTATGAGGACACCCTGACCCTGTCCCCT  
TCTCTGGGGAGACTGTGTTTCATGAGCATGGAGAACCCTGGCCTGTGGATTCTGGGTGCCACA  
ACTCTGACTTCAGGAACAGGGGCATGACTGCCCTGCTGAAAGTCTCCAGCTGTGACAAGAACA  
CTGGGGACTACTATGAGGACAGCTATGAGGACATCTCTGCCTACCTGCTGAGCAAGAACAATG

CCATTGAGCCCAGGAGCTTCAGCCAGAATGCCACTAATGTGTCTAACAAACAGCAACACCAGCA  
ATGACAGCAATGTGTCTCCCCAGTGCCTGAAGAGGCACCAGAGGGAGATCACCAGGACCACCC  
TGCAGTCTGACCAGGAGGAGATTGACTATGATGACACCATCTCTGTGGAGATGAAGAAGGAGG  
ACTTTGACATCTACGACGAGGACGAGAACCAGAGCCCCAGGAGCTTCAGAAGAAGACCAGGC  
ACTACTTCATTGTCTGTGGAGAGGCTGTGGGACTATGGCATGAGCAGCAGCCCCATGTGC  
TGAGGAACAGGGCCCAGTCTGGCTCTGTGCCCCAGTTCAGAAGGTGGTGTCCAGGAGTTCA  
CTGATGGCAGCTTCACCCAGCCCCGTACAGAGGGGAGCTGAATGAGCACCTGGGCCCTGCTGG  
GCCCCTACATCAGGGCTGAGGTGGAGGACAACATCATGGTGACCTTCAGGAACCAGGCCAGCA  
GGCCCTACAGCTTCTACAGCAGCCTGATCAGCTATGAGGAGGACCAGAGGCAGGGGGCTGAGC  
CCAGGAAGAATTTGTGAAGCCCAATGAAACCAAGACCTACTTCTGGAAGGTGCAGCACCACA  
TGGCCCCCACCAGGATGAGTTTGACTGCAAGGCCCTGGGCCCTACTTCTCTGATGTGGACCTGG  
AGAAGGATGTGCACTCTGGCCTGATTGGCCCCCTGCTGGTGTGCCACCAACACCCCTGAACC  
CTGCCCATGGCAGGCAGGTGACTGTGCAGGAGTTTGCCCTGTTCTTACCATCTTTGATGAAA  
CCAAGAGCTGGTACTTCACTGAGAACATGGAGAGGAACTGCAGGGCCCCCTGCAACATCCAGA  
TGGAGGACCCACCTTCAAGGAGAATACAGGTTCCATGCCATCAATGGCTACATCATGGACA  
CCCTGCCTGGCCTGGTGTGATGGCCAGGACCAGAGGATCAGGTGGTACCTGCTGAGCATGGGCA  
GCAATGAGAACATCCACAGCATCCACTTCTCTGGCCATGTGTTCACTGTGAGGAAGAAGGAGG  
AGTACAAGATGGCCCTGTACAACCTGTACCCTGGGGTGTTTGAGACTGTGGAGATGCTGCCCA  
GCAAGGCTGGCATCTGGAGGTGGAGTGCCTGATTGGGGAGCACCTGCATGCTGGCATGAGCA  
CCCTGTTCTGTGTGTACAGCAACAAGTGCCAGACCCCCCTGGGCATGGCCTCTGGCCACATCA  
GGGACTTCCAGATCACTGCCTCTGGCCAGTATGGCCAGTGGGCCCCCAAGCTGGCCAGGCTGC  
ACTACTCTGGCAGCATCAATGCCTGGAGCACCAGGAGCCCTTCAGCTGGATCAAGGTGGACC  
TGCTGGCCCCATGATCATCCATGGCATCAAGACCCAGGGGGCCAGGCAGAAGTTCAGCAGCC  
TGTACATCAGCCAGTTTCATCATCATGTACAGCCTGGATGGCAAGAAGTGGCAGACCTACAGGG  
GCAACAGCACTGGCACCCCTGATGGTGTCTTTGGCAATGTGGACAGCTCTGGCATCAAGCACA  
ACATCTTCAACCCCCCATCATTGCCAGATACATCAGGCTGCACCCACCCACTACAGCATCA  
GGAGCACCTGAGGATGGAGCTGATGGGCTGTGACCTGAACAGCTGCAGCATGCCCCGGGCA  
TGGAGAGCAAGGCCATCTCTGATGCCAGATCACTGCCAGCAGCTACTTCACCAACATGTTTG  
CCACCTGGAGCCCCAGCAAGGCCAGGCTGCACCTGCAGGGCAGGAGCAATGCCTGGAGGCCCC  
AGGTCAACAACCCCAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATGAAGGTGACTGGGG  
TGACCACCCAGGGGGTGAAGAGCCTGCTGACCAGCATGTATGTGAAGGAGTTCCTGATCAGCA  
GCAGCCAGGATGGCCACCAGTGGACCCCTGTTCTTCCAGAATGGCAAGGTGAAGGTGTTCCAGG  
GCAACCAGGACAGCTTCAACCCCTGTGGTGAACAGCCTGGACCCCCCTGCTGACCAGATACC  
TGAGGATTCAACCCAGAGCTGGGTGCACCAGATTGCCCTGAGGATGGAGGTGCTGGGCTGTG  
AGGCCCAGGACCTGTACTGA

SEQ ID NO: 31

TACGTCTAACTCGACTCGTGGACGAAGAAGGACACGGACGACTCCAAGACGAAGAGAC  
GGTGGTCTCTATGATGGACCCCGACACCTCGACTCGACCCTGATGTACGTGAGACTGGACC  
CCCTCGACGGACACCTACGGTCCAAGGGGGGTCTCACGGGTCTCGAAGGGGAAGTTGTGGA  
GACACCACATGTTCTTCTGGGACAAACACCTCAAGTGACTGGTGGACAAGTTGTAACGGTTCG  
GGTCCGGGGGACCTACCCGGACGACCCGGGGTGGTAGGTCGGACTCCACATACTGTGACACC  
ACTAGTGGGACTTCTTGTACCGGTCCGGTGGGACACTCGGACGTACGACACCCCGACTCGATGA  
CCTTCCGGAGACTCCCCGACTCATACTACTGGTCTGGTCCGTCTCCCTCTTCTCCTACTGT  
TCCACAAGGGACCCCGTCCGTGTGGATACACACCGTCCACGACTTCCCTTACCGGGGTACC  
GGAGACTGGGGGACACGGACTGGATGTCGATGGACTCGGTACACCTGGACCCTTCTGGACT  
TGAGACCGGACTAACCCGGGACGACCACACGTCCCTCCCGTCCGACCGGTCTCTTCTGGG  
TCTGGGACGTGTTCAAGTAGGACGACAAACGACACAACTACTCCCGTCTCGACCGTGAGAC  
TTTGGTCTTGTGCGGACTACGTCTTCTTCCCTACGACGGAGACGGTCCCGGACCGGGTCTACG  
TGTGACACTTACCGATACTTGTCTCGGACGGACCGGACTAACCGACCGGTCTCTCAGAC  
ACATGACCGTACACTAACCGTACCCGTGGTGGGACTCCACGTGTCGTAGAAGGACCTCCCGG  
TGTGGAAGGACCAGTCTTGGTGTCCGTCCGGTCCGACCTCTAGTCCGGGTAGTGAAGGACT  
GACGGGTCTGGGACGACTACCTGGACCCGGTCAAGGACGACAAGACGGTGTAGTCGTGGTGG  
TCGTACTACCGTACCTCCGGATACTTCCACCTGTCGACGGGACTCCTCGGGTTCGACTCCT  
ACTTCTTGTACTCCTCCGACTCCTGATACTACTACTGGACTGACTGAGACTCTACCTACACC  
ACTCCAACTACTACTGTTGTCCGGGTGGAAGTAGGCTAGTCCAGACACCGGTCTTCTCGTGG  
GGTCTGAGCCACGTGATGTAACGACGACTCCTCCTCGACCTGATACGGGGGACACG  
ACCGGGGACTACTGTCTCGATGTTCTCGGTATGACTTGTACCAGGGGTCTCCTAACCGT  
CCTTCATGTTCTTCCAGTCCAAGTACCGGATGTGACTACTTTGGAAGTCTGGTCCCTCCGGT  
AGGTCTGACTCAGACCGTAGGACCCGGGGACGACATACCCCTCCACCCCTGTGGGACGACT  
AGTAGAAGTCTTGGTCCGGTCCGCGGATGTTGTAGATGGGGTACCGTAGTGACTACACT  
CCGGGGACATGTCGTCTCCGACGGGTCCCCACTTCGTGGACTTCCCTGAAGGGTAGGACG  
GACCCCTCTAGAAGTTCATGTTACCTGACACTGACACCTCCTACCGGGTGGTTCAGACTGG  
GGTCCACGGACTGGTCTATGATGTCGTCGAAACACTGTACCTCTCCCTGGACCGGAGACCGG  
ACTAACCGGGGACGACTAGACGATGTTCTCAGACACCTGGTCTCCCGTGGTCTAGTACA  
GACTGTTCTCCTTACACTAGGACAAGAGACACAACTACTCTTGTCTCGACCATGGACTGAC  
TCTTGTAGGTCTCCAAGGACGGGTGGGACGACCCACGTCGACCTCCTGGGACTCAAGTCC  
GGTCTGTTAGTACGTGTCGTAGTTACCGATAACAACTGTCGGACGTGACAGACACACGG  
ACGTACTCCACCGGATGACCATGTAGGACTCGTAACCCGGGTCTGACTGAAGGACAGACACA  
AGAAGAGACCGATGTGGAAGTTCGTGTTCTACCACATACTCCTGTGGGACTGGGACAAGGGGA  
AGAGACCCCTCTGACACAAGTACTCGTACCTCTTGGGACCGGACACCTAAGACCCGACGGTGT  
TGAGACTGAAGTCTTCTCCCGTACTGACGGGACGACTTTCAGAGGTCGACACTGTTCTTGT  
GACCCCTGATGATACTCCTGTCGATACTCCTGTAGAGACGGATGGACGACTCGTCTTGTTC

GGTAACTCGGGTCCCTCGAAGTCGGTCTTACGGTGATTACACAGATTGTTGTCGTTGTGGTTCGT  
TACTGTGCTTACACAGAGGGGGTACGACTTCTCCGTGGTCTCCCTCTAGTGGTCCCTGGTGGG  
ACGTCAGACTGGTCCCTCTAACTGATACTACTGTGGTAGAGACACCTCTACTTCTTCTCC  
TGAAACTGTAGATGCTGCTCCTGCTTGGTCTCGGGGTCCTCGAAGGTCTTCTTCTGGTCCG  
TGATGAAGTAACGACGACACCTCTCCGACACCTGATACCGTACTCGTGGTCCGGGGTACACG  
ACTCCTTGTCCCGGGTCAGACCGAGACACGGGGTCAAGTCTTCCACCACAAGGTCCCAAGT  
GACTACCGTTCGAAGTGGGTCGGGACATGTCTCCCTCGACTTACTCGTGGACCCGGACGACC  
CGGGGATGTAGTCCCGACTCCACCTCCTGTTGTAGTACCCTGGAAGTCCCTGGTCCGGTTCGT  
CCGGGATGTGAAAGATGTGTCGGACTAGTCGATACTCCTCCTGGTCTCCGTCCCCGACTCG  
GGTCTTCTGAAACACTTCGGGTTACTTTGGTCTGGATGAAGACCTTCCACGTGCTGGTGT  
ACCGGGGGTGGTTCCTACTCAAAGTACGTTCCGGACCCGGATGAAGAGACTACACCTGGACC  
TCTTCTTACACGTGAGACCGGACTAACCGGGGACGACCACACGGTGTGGTGTGGGACTTGG  
GACGGGTACCGTCCGCTTACTGACACGTCCTCAAACGGGACAGAAGTGGTAGAAACTACTTT  
GGTTCGACCATGAAGTACTCTTGTACCTCCTTGACGTCGCGGGGACGTTGTAGGTCT  
ACCTCCTGGGGTGAAGTTCCTCTTGATGTCCAAGGTACGGTAGTTACCGATGTAGTACCTGT  
GGGACCGACCGGACACTACCGGTCCTGGTCTCCTAGTCCACCATGGACGACTCGTACCCGT  
CGTTACTCTTGTAGGTGTCGTAGGTGAAGAGACCGGTACACAAGTACACTCCTTCTTCTCC  
TCATGTTCTACCGGACATGTTGGACATGGGACCCCAAACTCTGACACCTCTACGACGGGT  
CGTTCGACCGTAGACCTCCACCTCACGGACTAACCCCTCGTGGACGTACGACCGTACTCGT  
GGGACAAGGACACATGTCGTTGTTACGGTCTGGGGGACCCGTACCGGAGACCGGTGTAGT  
CCCTGAAGGTCTAGTGACGGAGACCGGTCATACCGGTCACCCGGGGTTCGACCGGTCCGACG  
TGATGAGACCGTCTAGTTACGGACCTCGTGGTTCCTCGGGAAGTGCACCTAGTTCACCTGG  
ACGACCGGGGTACTAGTAGGTACCGTAGTTCTGGGTCCCCGGTCCGCTTCAAGTCTGTCGG  
ACATGTAGTCCGTCGAAGTAGTAGTACATGTCGGACCTACCGTCTTCCACCGTCTGGATGTCCC  
CGTTGTGTCGACCGTGGGACTACCACAAGAAACCGTTACACCTGTCGAGACCGTAGTTCGTGT  
TGTAAGTGGGGGGTAGTAACGGTCTATGTAGTCCGACGTGGGGTGGGTGATGTCGTAGT  
CCTCGTGGGACTCCTACCTCGACTACCCGACACTGGACTTGTGACGTCGTACGGGGACCCGT  
ACCTCTCGTTCGCGGTAGAGACTACGGGTCCTAGTGACGGTTCGTGATGAAGTGGTGTGTAACA  
GGTGGACCTCGGGGTCGTTCCGGTCCGACGTGGACGTCCCGTCTCCTGTTACGGACCTCCGGG  
TCCAGTTGTTGGGGTTCCTCACCGACGTCCACCTGAAGTCTTCTGGTACTTCCACTGACCCC  
ACTGGTGGTCCCCACTTCTCGGACGACTGGTCTGATACACTTCTCAAGGACTAGTCTGT  
CGTGGTCTTACCGTGGTCACTGGGACAAGAAGTCTTACCGTTCACCTTCCACAAGGTCC  
CGTTGGTCTGTGAAAGTGGGGACACCACTTGTGCGACCTGGGGGGGACGACTGGTCTATGG  
ACTCTAAGTGGGGTCTCGACCCACGTGGTCTAACGGGACTCTACTCTCCACGACCCGACAC  
TCCGGTCTTGGACATGACT

SEQ ID NO: 32

MQIELSTCFFLCLLRFCFSATTRYYLGAVELSWDYMQSDLGELPVDARFPFRVPKSFP  
 FNTSVVYKKTLE  
 VEFTDHLFNIAKPRPPWMLLGPITQAEVYDVTVVITLKNMASHPVSLHAVGVSYWKAS  
 EGAEYDDQTSQR  
 EKEDDKVFPGGSHYVWQVLKENGPMASDPLCLTYSYLSHVLDLVKDLNSGLIGALLVC  
 REGSLAKEKTQT  
 LHKFILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNRSLPGLIGCH  
 RKSVMYWHVIGMG  
 TTPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTLLMDLGQFLLFCHISSHQHDG  
 MEAYVKVDSCE  
 EPQLRMKNNEEAEDYDDDLTDSEMDVVRFDNNSPSFIQIRSVAKKHPKTWVHYIAAE  
 EEDWDYAPLVLA  
 PDDRSYKSYQLNNGPQRIGRKYKVRFMAYTDETFKTREAIQHESGILGPLYGEVGD  
 TLLIIFKNQASR  
 PNYIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCL  
 TRYSSFVNMER  
 DLASGLIGPLLIICYKESVDQRGNQIMSDKRNVLFSVFDENRSWYLTENIQRFLPNPA  
 GVQLEDPEFQAS  
 NIMHSINGYVFDLSQLSVLHEVAYWYILSIGAQTDFLSVFFSGYTFKHKMVEDTLT  
 LFPFSGETVFMMS  
 MENPGLWILGCHNSDFRNRGMTALLKVSSCDKNTGDYEDSYEDISAYLLSKNNAIEP  
 RSFSQATNVSN  
 NSNTSNDNSNVSPPVLKRHQREITRITLQSDQEEIDYDDTISVEMKKEDFDIYDEDENQ  
 SPRSFQKTRHY  
 FIAAVERLWDYGMSSSPHVLNRNRAQSGSVPQFKKVVQEFETDGSFTQPLYRGELNEHL  
 GLLGPYIRAEVE  
 DNIMVTFRNQASRPYSFYSSLISYEEDQRQGAEPKRFVKNPNETKTYFWKVQHMAPT  
 KDEFDCKAWAYF  
 SDVDLEKDVHSGLIGPLLVCHTNTLNPAGRQVTVQEFALFFTIFDETKSWYFTENME  
 RNCRAPCNIQME  
 DPTFKENYRFHAINGYIMDTLPGLVMAQDQRIRWYLLSMGSNENIHSIHFSGHVFTVR  
 KKEYKMALYNL  
 YPGVFETVEMLPSKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPMGASGHIRDFQ  
 ITASGQYQWAP  
 KLARLHYSGSINAWSTKEPFSWIKVDLLAPMI IHGIKTQGARQKFSSLYISQFIIMYS  
 LDGKKWQTYRGN

STGTLMVFFGNVDSSGIKHNI FNPPI IARY IRLHPTHYSIRSTLRMELMGCDLNSCSM  
 PLGMESKAISDA  
 QITASSYFTNMFATWSPSKARLHLQGRSNAWRPQVNNPKEWLQVDFQKTMKVTGVTTQ  
 GVKSLLTSMYVK  
 EFLISSSQDGHQWTLFFQNGKVKVFQGNQDSFTPVVNSLDPPLLRYLRIHPQSWVHQ  
 IALRMEVLGCEA  
 QDLY  
**SEQ ID NO: 33**  
 GGCGCCTCTAGAGTTATAACCGGTAATCGGTATAATAAGTAACCAATATATCGTATTT  
 AGTTATAACCGA  
 TAACCGGTAACGTATGCAACATAGATATAGTATTATACATGTAATATAACCGAGTAC  
 AGGTTATACTGG  
 CCGGTACAACCGTAACATAACTGATCAATAATTATCATTAGTTAATGCCCCAGTAAT  
 CAAGTATCGGGT  
 ATATACCTCAAGGCGCAATGTATTGAATGCCATTTACCGGGCGGACCGACTGGCGGGT  
 TGCTGGGGGCGG  
 GTAACCTGCAGTTATTACTGCATACAAGGGTATCATTCGCGTTATCCCTGAAAGGTAAC  
 TGCAGTTACCCA  
 CCTCATAAATGCCATTTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGGTTCA  
 GCGGGGGATAA  
 CTGCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTATGTACTGGAATGCC  
 TGAAAGGATGAA  
 CCGTCATGTAGATGCATAATCAGTAGCGATAATGGTACCCTACGCCAAAACCGTCAT  
 GTGGTTACCCGC  
 ACCTATCGCCAAACTGAGTGCCCTAAAGGTTTCAGAGGTGGGGTAACTGCAGTTACCC  
 TCAAACAAAACC  
 GTGGTTTTAGTTGCCCTGAAAGGTTTTACAGCATTATTGGGGCGGGGCAACTGCGTTT  
 ACCCGCCATCCG  
 CACATGCCACCCTCCAGATATATTCGTCTCGAGCAAATCACTTGGCAGTCTAGTGATC  
 TTCGAAATAACG  
 CCATCAAATAGTGTCAATTTAACGATTGCGTCAGTCACGAAGACTGTGTTGTCAGAGC  
 TTGAATTCGACG  
 TCTTCAACCAGCACTCCGTGACCCGTCCGATCG

#### ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Лентивирусный вектор для генной терапии, псевдотипированный гемагглютинин-нейраминидазой HN и F-белками из респираторного парамиксовируса, содержащий гибридный энхансер/промотор человеческих CMV и EF1a (hCEF) и трансген, кодирующий терапевтический белок, причем между hCEF и трансгеном отсутствует какой-либо интрон.

2. Лентивирусный вектор по п.1, выбранный из вектора на основе вируса иммунодефицита человека (HIV), вируса иммунодефицита обезьян (SIV), вируса иммунодефицита кошачьих (FIV), вируса инфекционной анемии лошадей (EIAV) и вируса висна/маэди.

3. Лентивирусный вектор по п.1 или 2, представляющий собой вектор на основе SIV.

4. Лентивирусный вектор по любому из пп.1-3, где респираторный парамиксовирус представляет собой вирус Сендай.

5. Лентивирусный вектор по любому из пп.1-4, в котором терапевтический белок представляет собой

(a) секретируемый терапевтический белок, в частности  $\alpha$ -1-антитрипсин (A1AT), фактор VIII, поверхностно-активный белок В (SFTPВ), фактор VII, фактор IX, фактор X, фактор CI, фактор фон Виллебранда, гранулоцитарно-макрофагальный колониестимулирующий фактор (GM-CSF) или моноклональное антитело к инфекционному агенту; или

(b) CFTR, DNAH5, DNAH11, DNAI1 или DNAI2.

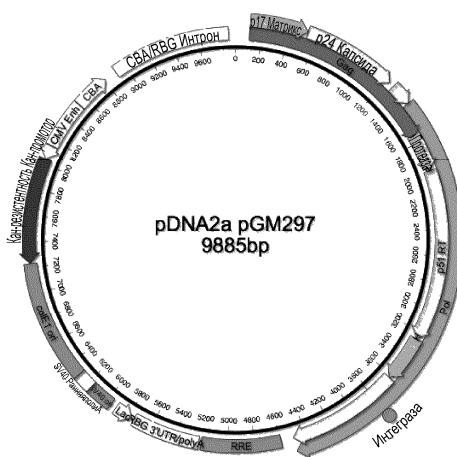
6. Лентивирусный вектор по любому из пп.1-5, в котором терапевтический белок представляет собой CFTR.

7. Лентивирусный вектор по любому из пп.1-5, в котором терапевтический белок представляет собой A1AT.

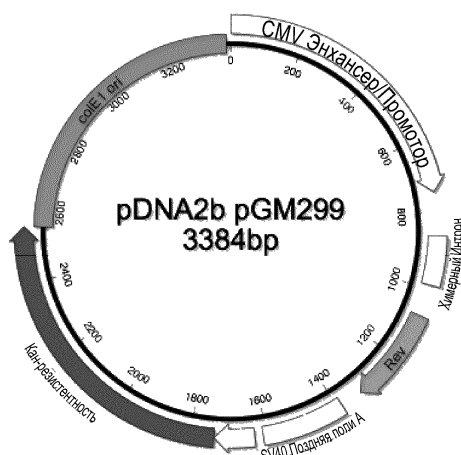
8. Лентивирусный вектор по любому из пп.1-5, в котором терапевтический белок представляет собой FVIII.

9. Способ получения лентивирусного вектора, определенного в любом из пп.1-8, включающий сле-



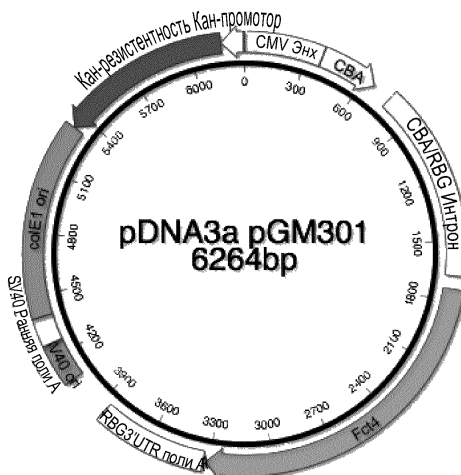


Фиг. 1В



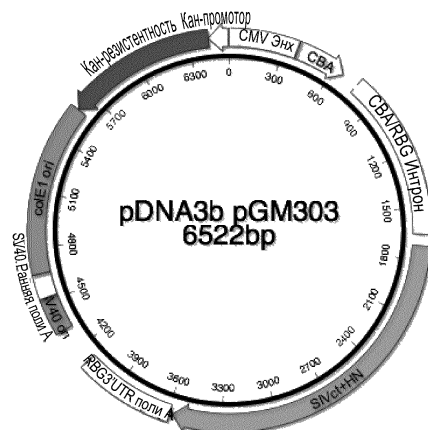
Кан-промотор

Фиг. 1С

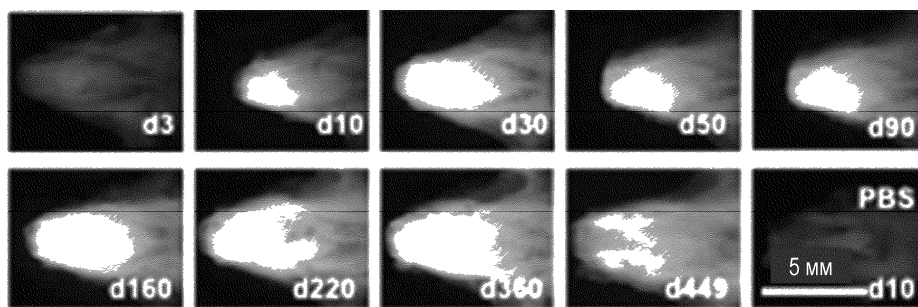


Фиг. 1D

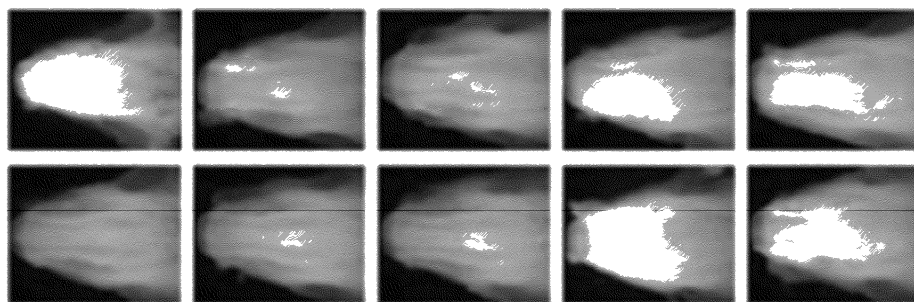




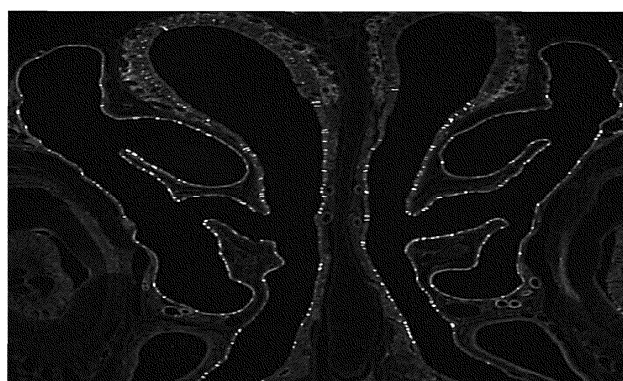
Фиг. 1Е



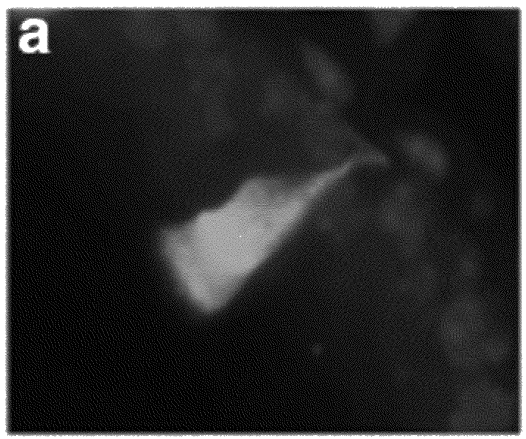
Фиг. 2



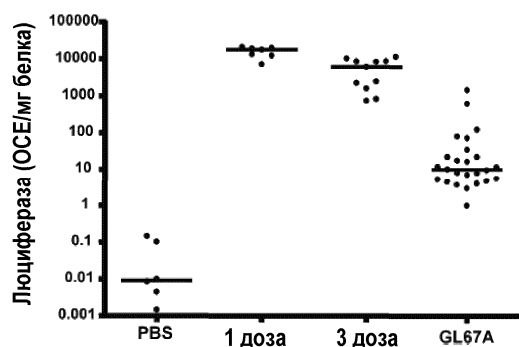
Фиг. 3



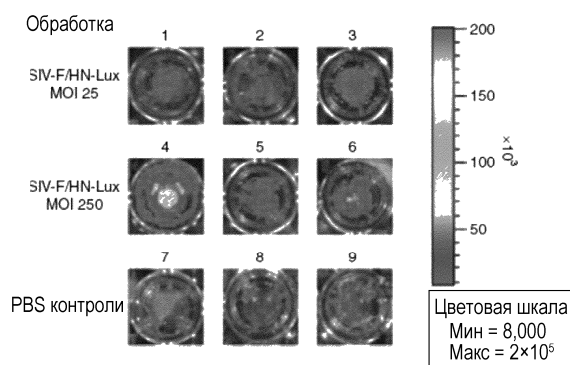
Фиг. 4



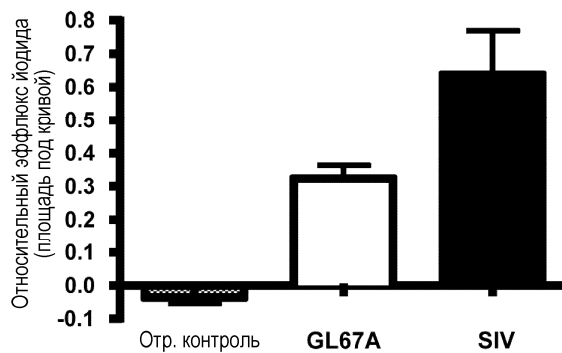
Фиг. 5



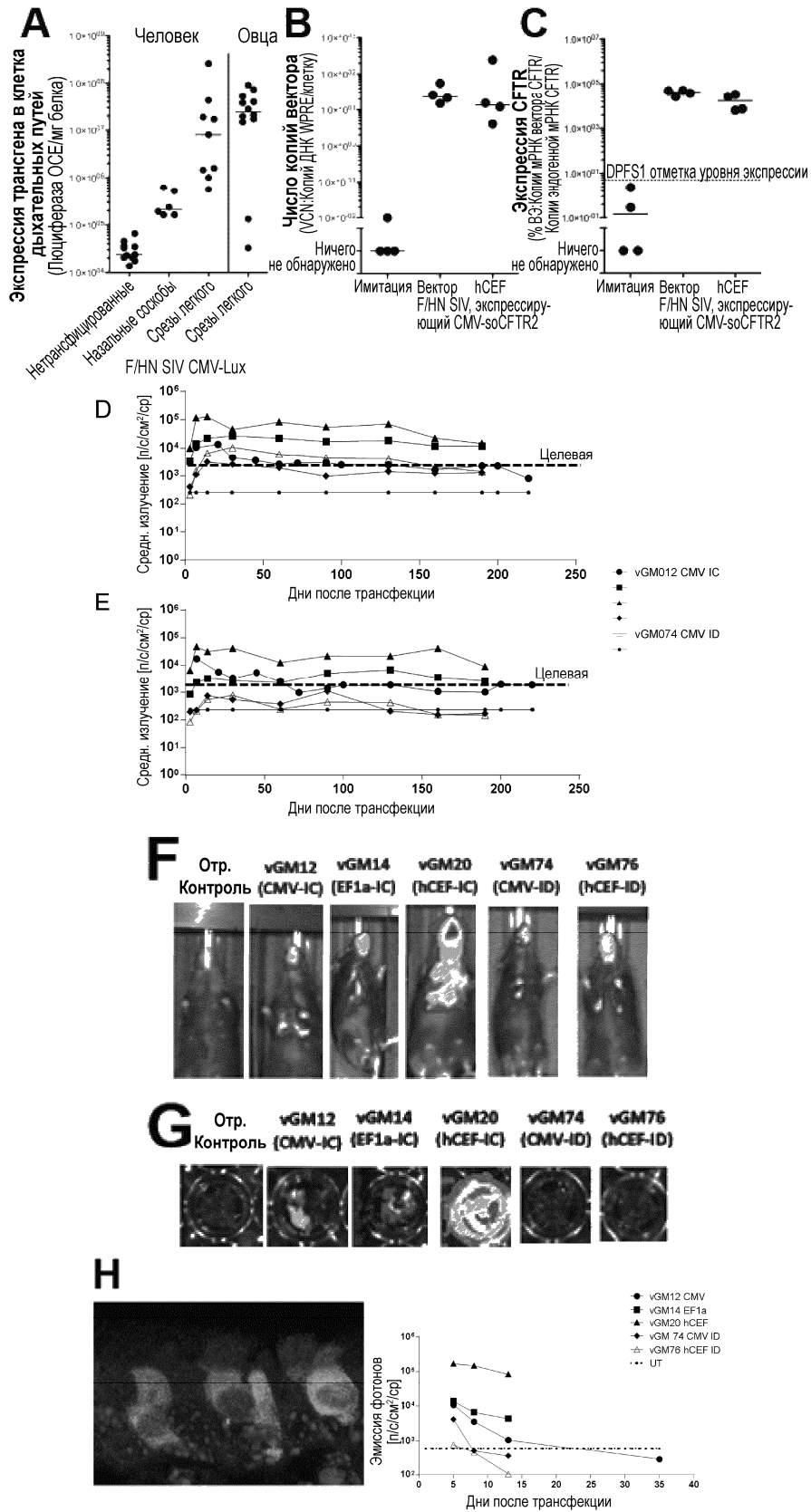
Фиг. 6



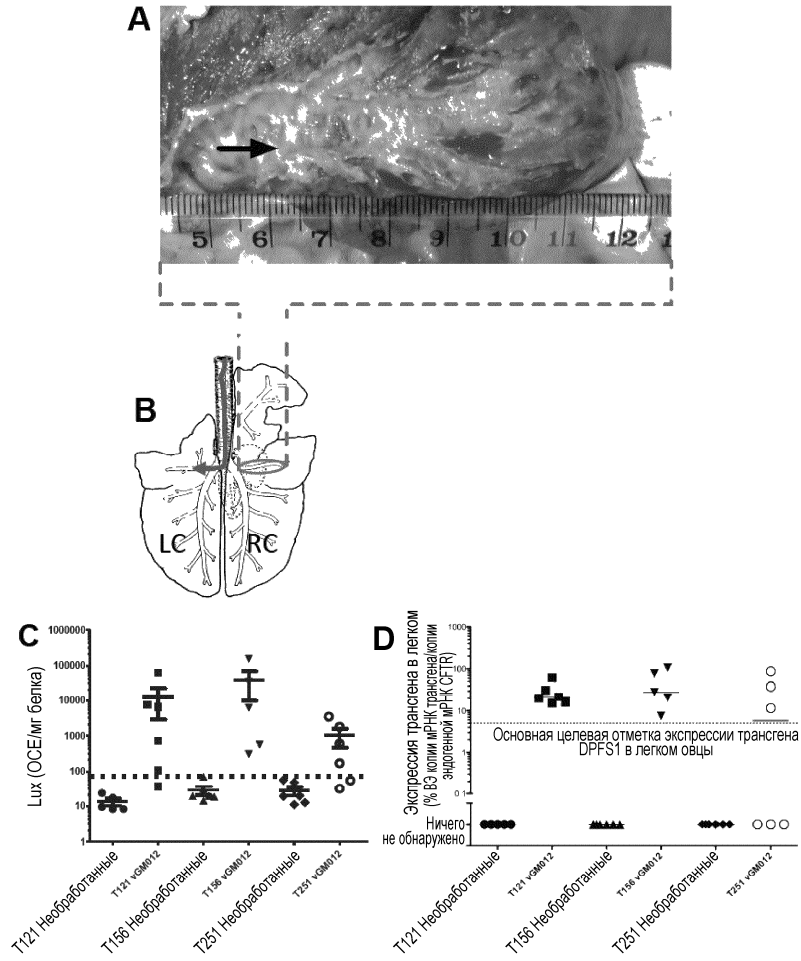
Фиг. 7



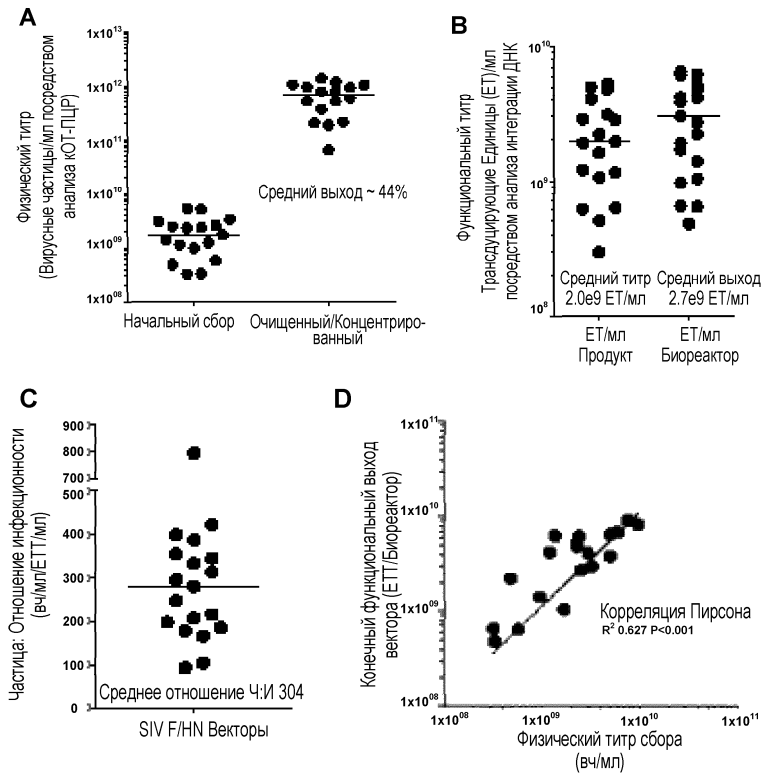
Фиг. 8



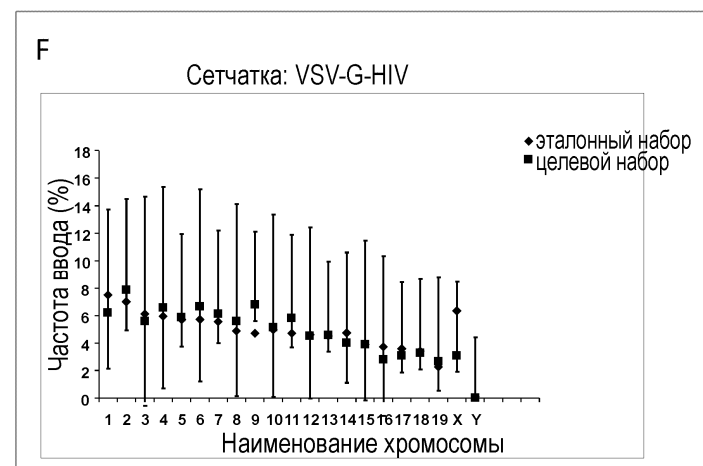
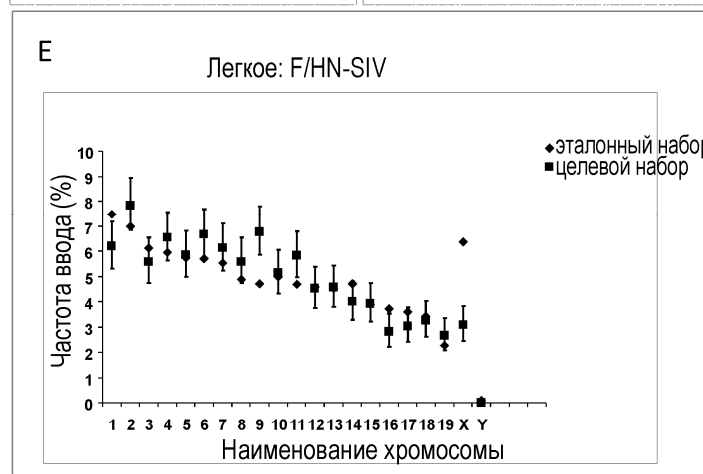
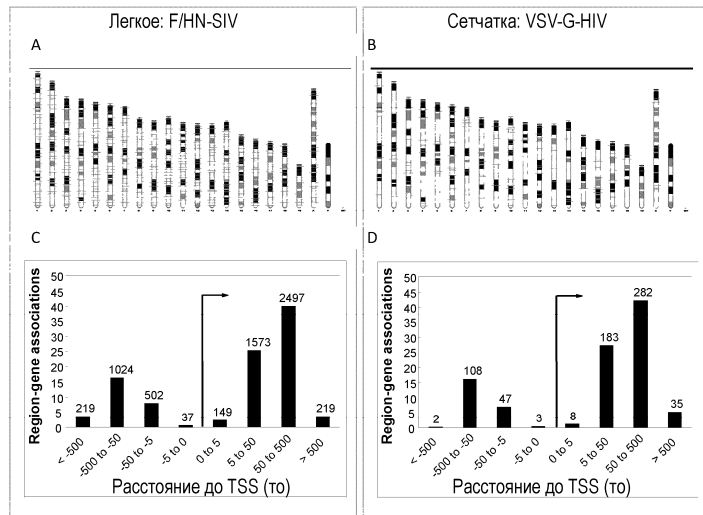
Фиг. 9

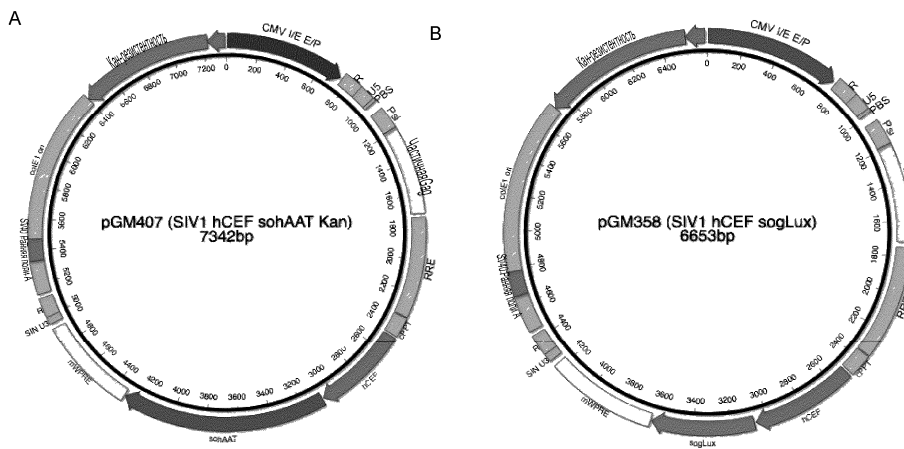
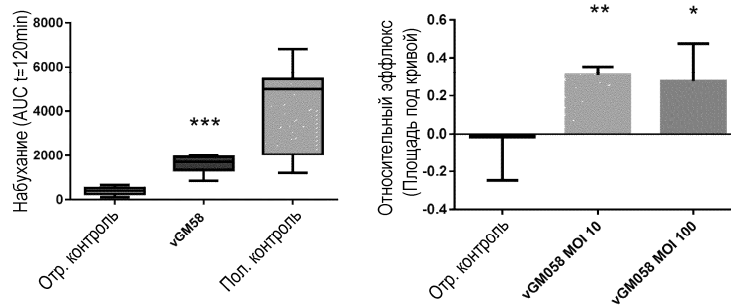
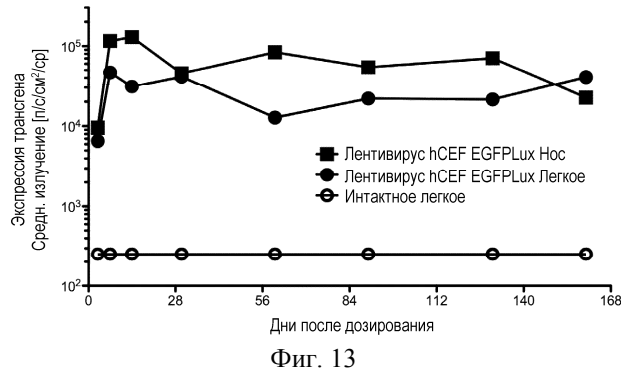
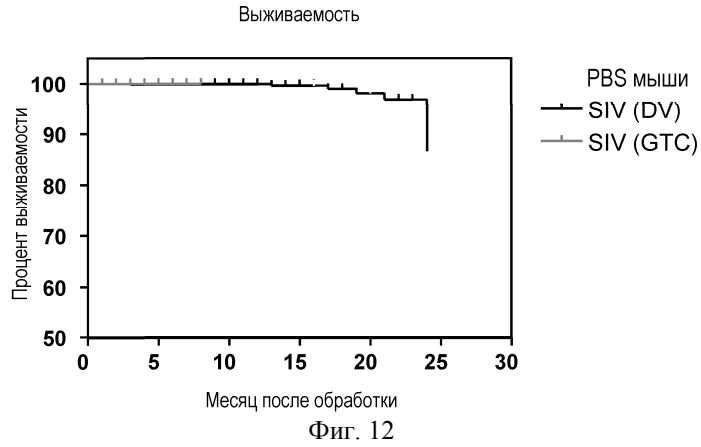


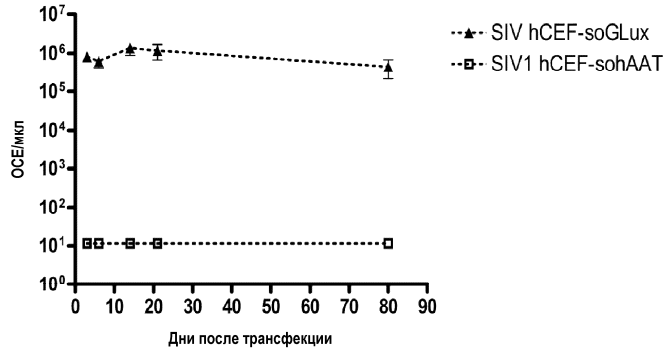
Фиг. 10



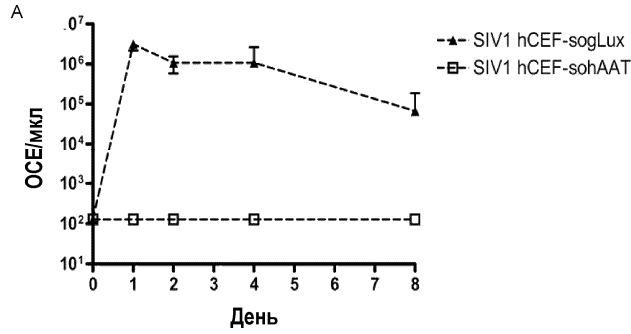
Фиг. 11



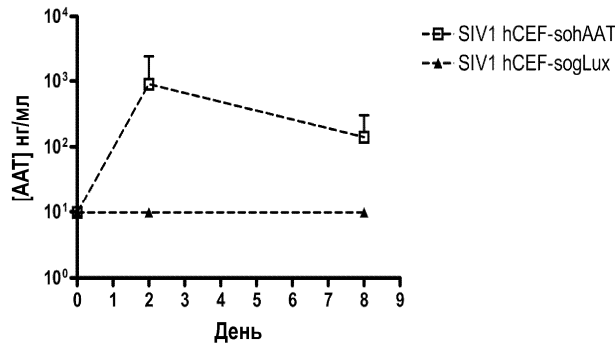




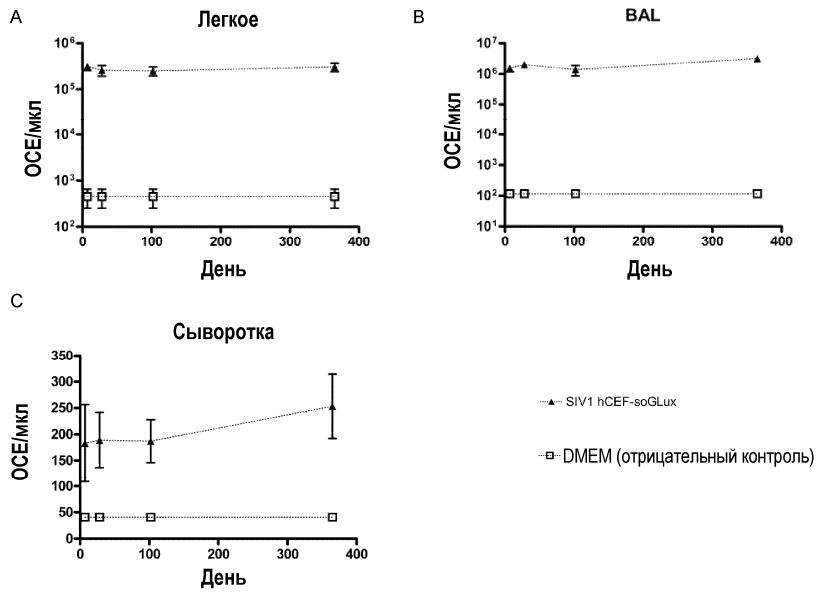
Фиг. 16



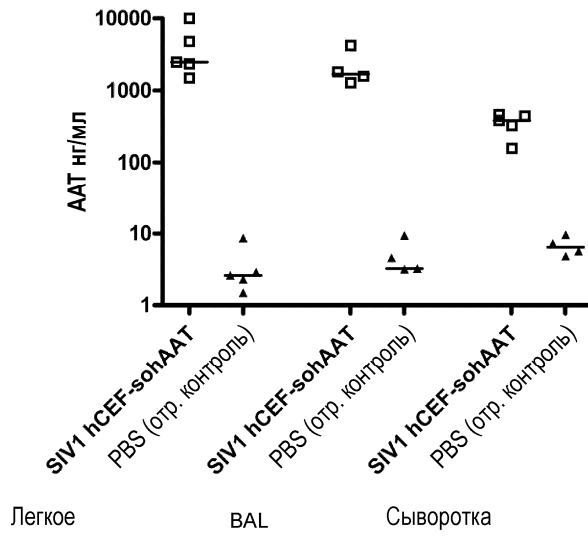
В



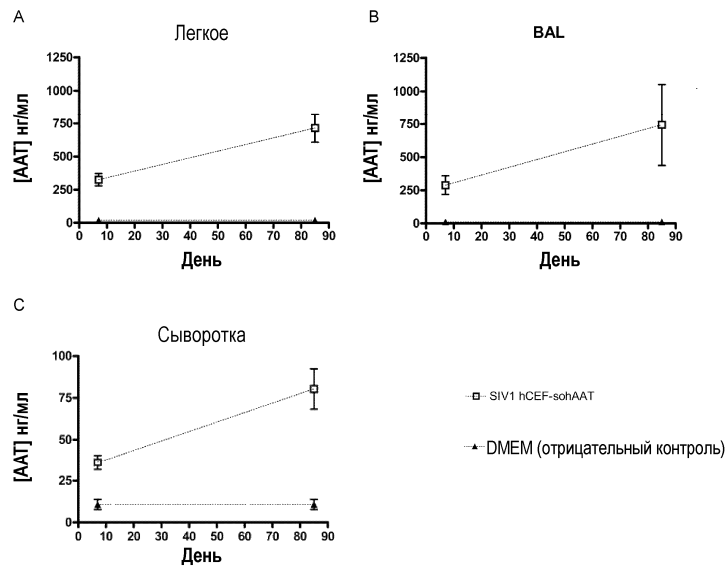
Фиг. 17



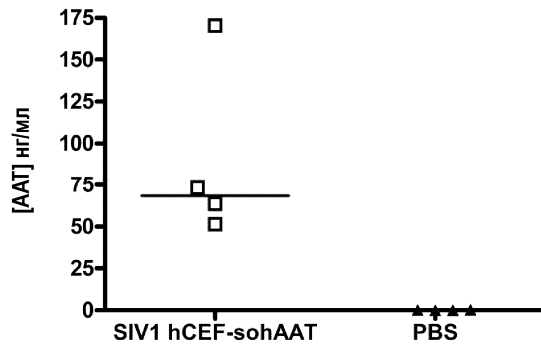
Фиг. 18



Фиг. 19

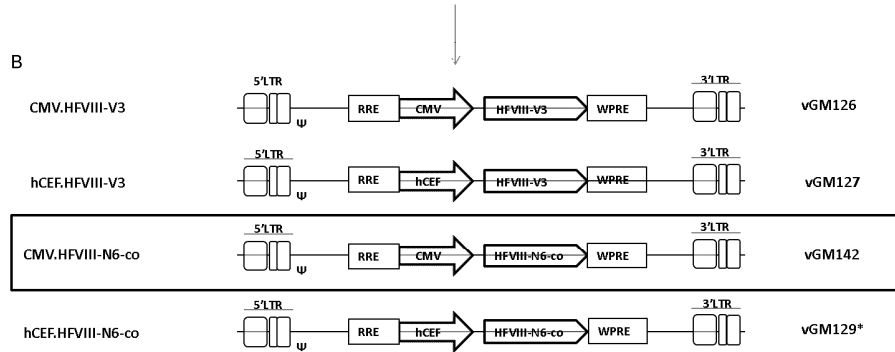
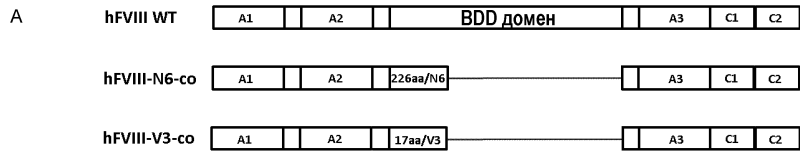


Фиг. 20

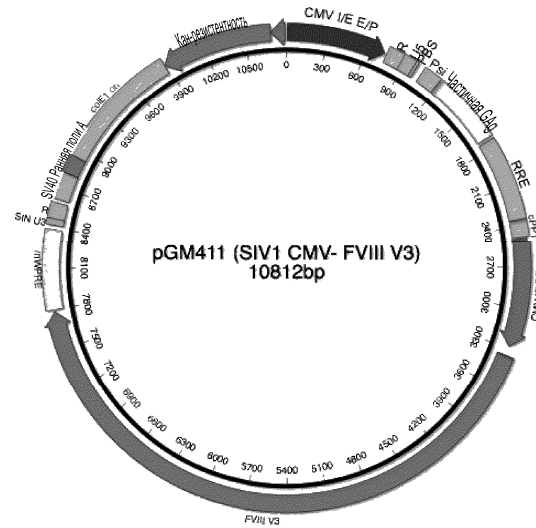


Фиг. 21

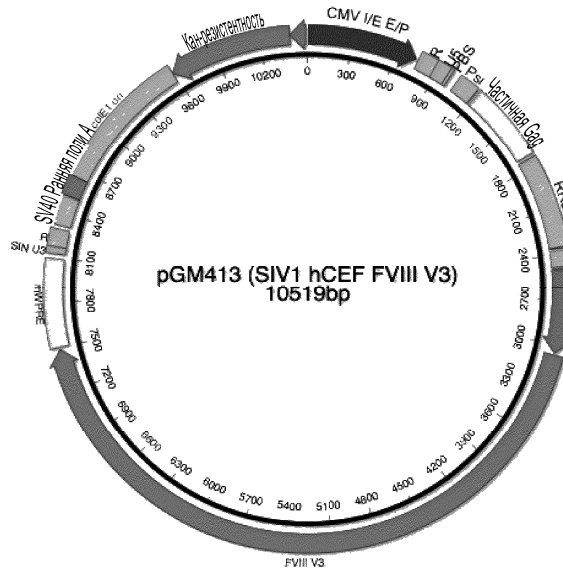


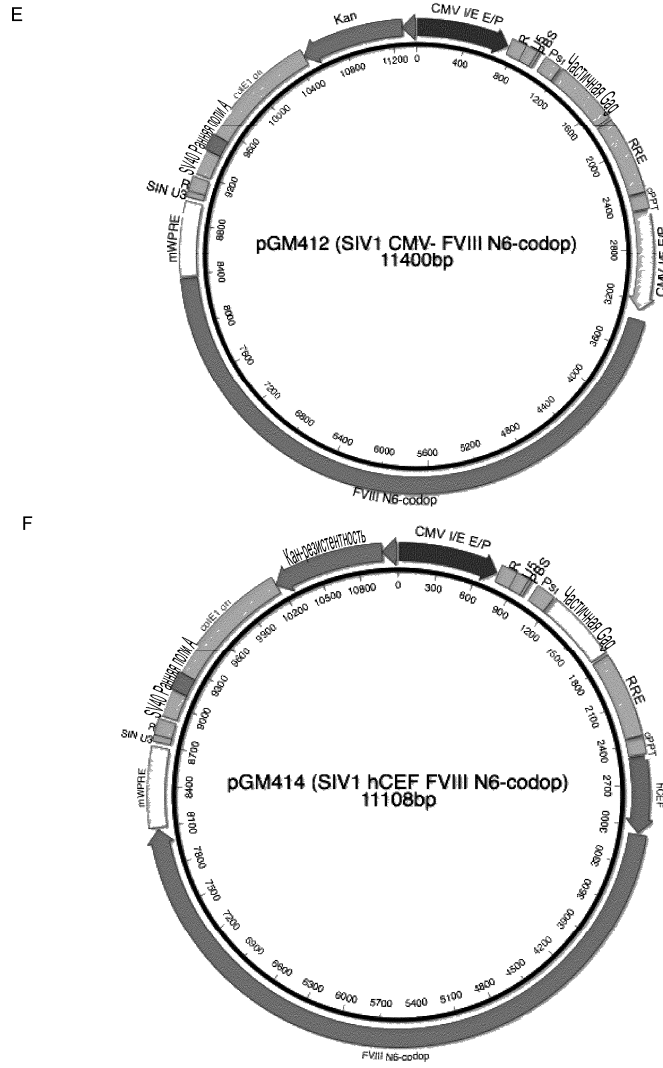


C

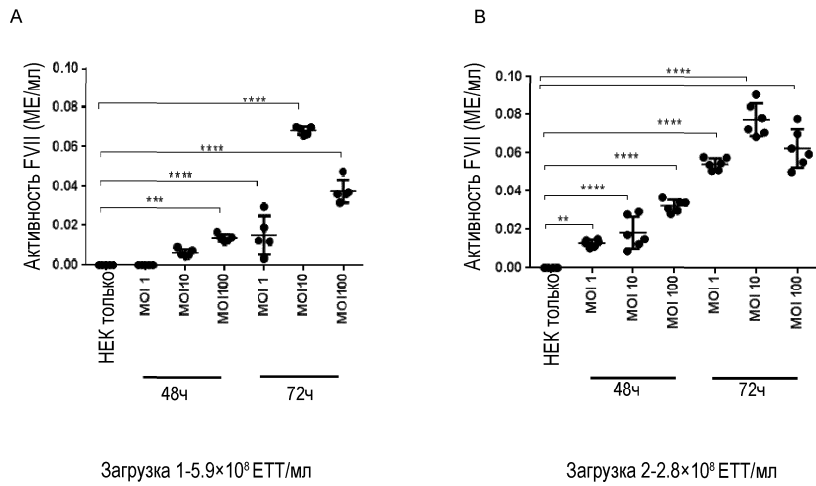


D

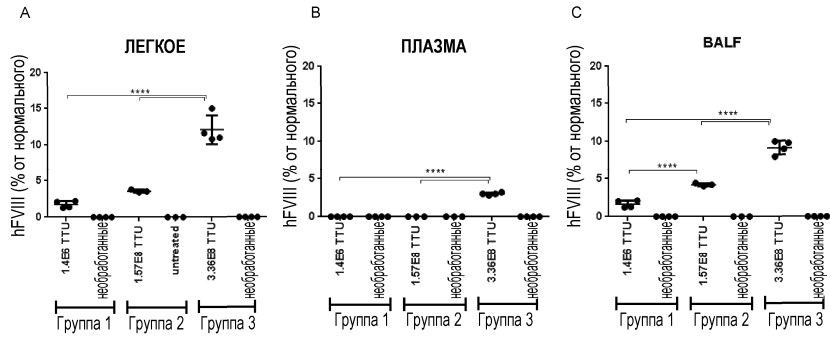




Фиг. 22



Фиг. 23



Фиг. 24



Евразийская патентная организация, ЕАПО

Россия, 109012, Москва, Малый Черкасский пер., 2