

(19)



**Евразийское  
патентное  
ведомство**

(11) **016397**(13) **B1**

## (12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ

(45) Дата публикации и выдачи патента  
**2012.04.30**

(21) Номер заявки  
**200800087**

(22) Дата подачи заявки  
**2006.06.16**

(51) Int. Cl. **C12Q 1/68** (2006.01)  
**A61P 3/10** (2006.01)

### (54) ГЕНЕТИЧЕСКИЕ ВАРИАНТЫ ГЕНА TCF7L2 КАК ДИАГНОСТИЧЕСКИЕ МАРКЕРЫ РИСКА ВОЗНИКНОВЕНИЯ САХАРНОГО ДИАБЕТА II ТИПА

(31) **60/692,174; 60/757,155**

(32) **2005.06.20; 2006.01.06**

(33) **US**

(43) **2008.06.30**

(86) **PCT/IS2006/000014**

(87) **WO 2006/137085 2006.12.28**

(71)(73) Заявитель и патентовладелец:  
**ДЕКОУД ДЖЕНЕТИКС ЕХФ. (IS)**

(72) Изобретатель:  
**Грант Струан Ф. А. (US)**

(74) Представитель:  
**Соболев А.Ю. (RU)**

(56) DATABASE SNPDB NCBI,  
"rs7903146", 04.07.2003, Найдено в  
Интернет:<URL: [http://www.ncbi.nlm.nih.gov/SNP/snp\\_ref.cgi?rs=7903146](http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=7903146), 02.06.2008  
DATABASE SNPDB NCBI, "rs4506565",  
16.03.2003, Найдено в Интернет:<URL::<http://www.ncbi.nlm.nih.gov>, 02.06.2008  
WO-A-2004042358

Twyman R.M. Snp discovery and typing  
technologies for pharmacogenomics. Curr Top Med  
Chem, 2004; 4(13):1423-31, реферат [Найдено в  
Интернет 04.06.2008], Medline [он-лайн] PMID:  
15379655

HALPERIN E. et al. "Tag SNP selection  
in genotype data for maximizing SNP prediction  
accuracy". Bioinformatics, 2005 Jun., 21 Suppl. 1;  
195-203, реферат [Найдено в Интернет 04.06.2008],  
Medline [он-лайн], PMID: 15961458

Yi F. "TCF-4 Mediates Cell Type-specific  
Regulation of Proglucagon Gene Expression by B-  
Catenin and Glycogen Synthase Kinase-3B", Journal  
of Biological Chemistry, 2005, January 14; 280(2):  
1457-64

(57) С помощью ассоциативного анализа продемонстрировано, что полиморфизм гена TCF7L2 определяет восприимчивость к диабету II типа. Приведены способы диагностики восприимчивости к диабету, пониженной восприимчивости к диабету и защиты против диабета, а также способы лечения диабета II типа.

**B1****016397****016397****B1**

### Уровень техники изобретения

Сахарный диабет - заболевание, связанное с нарушением метаболизма, при котором снижается утилизация углеводов и усиливается утилизация липидов и белков, возникает вследствие абсолютного или относительного дефицита инсулина. В более тяжелых случаях диабет характеризуется хронической гипергликемией, глюкозурией, потерей воды и электролитов, кетоацидозом и комой. Долговременные осложнения включают развитие нейропатии, ретинопатии, нефропатии, генерализованных дегенеративных изменений в крупных и малых кровеносных сосудах и повышенную чувствительность к инфекциям. Наиболее общей формой диабета является диабет II типа, независимый от инсулина диабет, который характеризуется гипергликемией, обусловленной нарушенной секрецией инсулина и резистентностью к инсулину тканей-мишеней. В развитие заболевания вносят вклад генетические факторы и влияние окружающей среды. Например, в развитии заболевания важную роль играет тучность. Диабет II типа часто является мягкой формой постепенного начала сахарного диабета.

Последствия диабета II типа для здоровья являются огромными. В 1995 г. в мире было 135 миллионов людей с диабетом. Подсчитано, что в 2025 г. около 300 миллионов людей будут больны диабетом (King H., et al., *Diabetes care*, 21(9): 1414-1431 (1998)). Распространенность диабета II типа в популяции взрослых жителей Исландии составляет 2,5% (Vilbergsson, S., et al., *Diabet. Med.*, 14(6): 491-498 (1997)), что включает приблизительно 5000 людей в возрасте выше 34 лет, которые больны этим заболеванием. Высокая распространенность заболевания и увеличивающаяся популяция больных свидетельствуют о неудовлетворенной потребности в определении генетических факторов, участвующих в развитии диабета II типа, для более точного определения ассоциированных факторов риска. Нужны также терапевтические агенты для предотвращения развития диабета II типа.

### Раскрытие изобретения

Настоящее изобретение относится к способам диагностики повышенной восприимчивости к диабету II типа, а также к способам диагностики пониженной восприимчивости к диабету II типа или диагностики действия защиты против диабета II типа путем выявления определенных маркеров или гаплотипов, относящихся к гену TCF7L2 (фактор транскрипции 2, подобный 7 (специфический по отношению к HMG-боксу Т-клеток), предварительно названный как TCF4 ген (транскрипционный фактор 4 Т-клеток). Способы включают определение генетического маркера, ассоциированного с блоком LD (нарушение равновесия сцепления) экзона 4 гена TCF7L2.

В первом аспекте изобретение относится к способу диагностики восприимчивости индивидуума к диабету II типа, включающему анализ образца нуклеиновой кислоты, полученного из индивидуума для определения признака или гаплотипа, ассоциированного с блоком LD экзона 4 гена TCF7L2, где присутствие признака или гаплотипа является индикатором восприимчивости к диабету II типа. В одном воплощении признак или гаплотип включает по меньшей мере один признак, выбранный из группы признаков, перечисленных в табл. 6. В другом воплощении признак или гаплотип является маркером.

В одном предпочтительном воплощении маркер или гаплотип является индикатором повышенной восприимчивости к диабету II типа. В одном воплощении повышенная восприимчивость характеризуется относительным риском по меньшей мере 1,2, включая относительный риск по меньшей мере 1,3 и относительный риск по меньшей мере 1,4. В одном воплощении маркер выбирается из группы, состоящей из DG10S478, rs12255372, rs7895340, rs11196205, rs7901695, rs7903146, rs12243326 и rs4506565, где присутствие не ноль-0 аллеля (т.е. -4, 4, 8, 12, 16, 20 или другого не-0 аллеля) в DG10S478, Т аллеля в rs12255372; А аллеля в rs7895340; С аллеля в rs11196205; С аллеля в rs7901695; Т аллеля в rs7903146; С аллеля в rs4506565 или Т аллеля в rs4506565 является индикатором повышенной восприимчивости к диабету II типа. В предпочтительном воплощении маркер выбирается из группы, состоящей из DG10S478 и rs7903146, и где присутствие не ноль-0 аллеля является маркером повышенной восприимчивости к диабету II типа. Еще в одном предпочтительном воплощении маркером является rs7903146, где присутствие Т аллеля в rs7903146 является индикатором восприимчивости к диабету II типа.

В другом предпочтительном воплощении маркер или гаплотип является индикатором пониженной восприимчивости к диабету II типа. Пониженная восприимчивость в одном воплощении характеризует относительный риск менее чем 0,7. В одном воплощении маркер выбирается из группы, состоящей из DG10S478, rs12255372, rs7895340, rs11196205, rs7901695, rs7903146, rs12243326 и rs4506565, где присутствие 0 аллеля в DG10S478, G аллеля в SNP rs12255372, G аллеля в rs7895340; Т аллеля в rs7901695; С аллеля в rs7903146; Т аллеля в rs12243326 или А аллеля в rs4506565 является индикатором пониженной восприимчивости к диабету II типа. В предпочтительном воплощении маркером является DG10S478, где присутствие 0 аллеля в DG10S478 является индикатором пониженной восприимчивости к диабету II типа. В другом предпочтительном воплощении маркером является rs7903146, где присутствие С аллеля в rs7903146 является индикатором пониженной восприимчивости к диабету II типа.

Во втором аспекте настоящее изобретение относится к набору реактивов для анализа образца из индивидуума с целью определения восприимчивости к диабету II типа, где набор реактивов включает один или более реагентов для обнаружения одного или более маркеров, ассоциированных с LD экзона 4, блок TCF7L2. В одном воплощении один или более реагентов включает по меньшей мере одну прилегающую нуклеотидную последовательность, которая полностью комплементарна области, включающей

по меньшей мере один маркер, ассоциированный с LD экзона 4, блок TCF7L2. В одном воплощении один маркер выбран из группы, состоящей из DG10S478, rs12255372, rs7895340, rs11196205, rs7901695, rs7903146, rs12243326 и rs4506565. В предпочтительном воплощении маркер является С аллелем в rs7903146.

В другом аспекте настоящее изобретение относится к способу анализа индивидуума для определения вероятности ответа на терапевтический агент TCF7L2, включающему определение маркера, ассоциированного с LD экзона 4, блок TCF7L2, где присутствие маркера является индикатором вероятности позитивного ответа на терапевтический агент TCF7L2. В одном воплощении маркер выбирается из группы, состоящей из DG10S478, rs12255372, rs7895340, rs11196205, rs7901695, rs7903146, rs12243326 и rs4506565. В другом воплощении маркер является маркером DG10S478 или маркером rs7903146, где присутствие не-0 аллеля в DG10S478 или Т аллеля в rs7903146 является индикатором вероятности позитивного ответа на терапевтический агент TCF7L2.

Другой аспект изобретения относится к применению терапевтического агента TCF7L2 для производства медикамента для лечения диабета II типа. В одном воплощении терапевтический агент TCF7L2 является агентом, который изменяет активность сигнального пути Wnt или кадгеринового сигнального пути. В другом воплощении терапевтический агент TCF7L2 является агентом, выбранным из группы, указанной выше в таблице агентов.

### **Краткое описание чертежей**

Вышеупомянутые и другие объекты, особенности и преимущества изобретения будут очевидны из следующего более детального описания предпочтительных воплощений изобретения.

Фигура описывает интересующую область TCF7L2 по отношению к нарушению равновесия сцепления (LD) SNP в проекте НарМар, версия 16. Участки гена в 215,9 т.п.о. семи LD блоков, которые указаны схематично черными стрелками (на основе данных NCBI RefSeq), показывают направление транскрипции; указаны экзоны с выделенным экзоном 4. DG10S478 расположен в 114,46 м.п.о. на хромосоме 10 (NCBI конструкция 34) в интроне 3 гена TCF7L2, с 74,9 т.п.о. блоком, который включает часть интрона 3, весь экзон 4 и часть интрона 4 (здесь называется "блок экзона 4 TCF7L2"). Маркеры SNP размещены равноудалено, а не в соответствии с их физическим расположением. На фигуре показаны два параметра LD: D' (верхняя левая часть фигуры) и  $r^2$  (нижняя правая часть).

### **Осуществление изобретения**

Описание предпочтительных воплощений изобретения даны ниже.

Локусы, ассоциированные с диабетом II типа.

Диабет II типа характеризуется гипергликемией, которая может возникать благодаря таким механизмам, как пониженная секреция инсулина, резистентность к инсулину периферических тканей и повышенная продукция глюкозы печенью. Большинство пациентов с диабетом II типа страдают от серьезных осложнений хронической гипергликемии, включая нефропатию, нейропатию, ретинопатию и ускоренное развитие сердечно-сосудистых заболеваний. Распространенность диабета II типа в мире в настоящее время составляет 6%, но предсказывается увеличение этого параметра в течение следующего десятилетия (1). Это увеличение распространенности диабета II типа является следствием увеличения возраста популяции и роста тучности.

Имеются свидетельства о вкладе генетического компонента в риск возникновения диабета II типа, включая влияние различий между группами различных рас (2, 3), более высокий уровень согласованности между монозиготными, чем между дизиготными близнецами (4, 5), и относительный риск появления диабета II типа у братьев и сестер ( $\lambda_s$ ) в европейской популяции, составляющий приблизительно 3,5 (6).

Таким образом, для поиска генов, ассоциированных с диабетом II типа, было применено два подхода. Для выявления ассоциации с заболеванием в пределах генов-кандидатов тестировали единичный нуклеотидный полиморфизм (SNP), и в основном он не дал или дал лишь умеренный риск диабета II типа, о котором наиболее широко сообщалось, что он характеризуется защитным Pro-12-Ala полиморфизмом в активированного пролифератором пероксисом рецептором гамма гена (PPARG2) (7), и риск полиморфизма в калиевом канале, регулируемом изнутри, подсемейство J, член 11 гена (KIR6,2) (8).

Сканирование связей во множестве геномов семейств с общей формой диабета II типа выявило несколько локусов, и первично международный исследовательский консорциум сфокусировался на хромосомах 1, 12 и 20, наблюдаемых во многих популяциях (6). Гены этих локусов до этого не рассматривались. Однако у мексиканских американцев ген кальпаина 10 (CAPN10) был изолирован из локуса хромосомы 2q; это отражало единственный ген для общей формы диабета II, идентифицированный до настоящего времени благодаря позиционному клонированию (9). Редкая Менделева форма диабета II, а именно диабет зрелого возраста у молодых (MODY), дала при позиционном клонировании семь генов (6).

Предварительно авторы изобретения сообщали о существенной связи диабета II типа с хромосомой 5q для исландской популяции для множества геномов (10); в том же исследовании также сообщали о предположительном доказательстве связи с 10q и 12q. Связь с 10q областью наблюдалась также для мексиканских американцев (11).

Ассоциация диабета II типа с геном белка 2, подобного транскрипционному фактору 7 (TCF7L2).

Настоящее изобретение относится к идентификации ассоциированного с диабетом II типа блока ("LD блок экзона 4 TCF7L2) в пределах гена, кодирующего транскрипционный фактор 4 (TCF4 - официальный символ гена TCF7L2). Было обнаружено, что с диабетом II типа ассоциировано несколько маркеров в пределах LD экзона 4 блока TCF7L2, включая микросателлитные маркеры DG10S478 и SNP rs7903146 и rs12255372. Оригинальные наблюдения, впервые сделанные в исландской популяции ( $P=1,3 \times 10^{-9}$ ; относительный риск=1,45; риск, отнесенный к популяции=22,7%), были впоследствии подтверждены на датской популяции больных диабетом II типа и на популяции белых, больных диабетом, в Соединенных Штатах Америки. DG10S478 находится в интроне 3 гена TCF7L2 в 10q25,2 и в пределах хорошо определенного LD блока размером 74,9 т.п.о., который включает часть интрона 3, весь экзон 4 и часть интрона 4. Продукт гена TCF7L2 представляет собой высокомолекулярную группу (HMG) включающего бокс транскрипционного фактора, который играет важную роль в Wnt сигнальном пути, известном также как сигнальный путь APC3/ $\beta$ -катенин/TCF путь. TCF7L2 опосредует специфическую для клеток регуляцию экспрессии гена проглюкагона (ключевого игрока в гомеостазе глюкозы в крови) через участника Wnt сигнального каскада  $\beta$ -катенина и киназу-3-бета гликогенсинтазы (12). Кроме того, сигнальный путь Wnt поддерживает преадипоциты в недифференцированном состоянии за счет ингибирования белка альфа (C/EBP-альфа), связывающегося с энхансером CCAAT адипогенных факторов транскрипции и активированного пролифератором пероксисом рецептора гамма (PPAR-гамма) (13). Когда Wnt сигнальный путь в преадипоцитах предотвращается за счет суперэкспрессии доминантно-негативного TCF7L2, эти клетки дифференцируют в адипоциты (13). Кроме того, сообщалось, что сигнальный каскад Wnt/ $\beta$ -катенин направлен на PPAR-гамма активность путем физического взаимодействия с  $\beta$ -катенином и TCF7L2 в клетках рака кишечника (14). Мультифункциональный белок  $\beta$ -катенин важен также для обеспечения клеточной адгезии через его связывание с кадгеринами (15).

В результате этого открытия в настоящее время доступны способы диагностики восприимчивости к диабету II типа, а также диагностики пониженной восприимчивости к диабету II типа и/или защиты против диабета II типа. В предпочтительных воплощениях изобретения диагностический анализ был применен для идентификации присутствия специфических аллелей, включая 0 аллель в маркере DG10S478 (ассоциированный с пониженной восприимчивостью к диабету II типа, который является защитным против диабета II типа); не-0 аллеля (т.е. -4, 4, 8, 12, 16 или 20 или другой аллель) в маркере DG10S478 (ассоциированный с восприимчивостью к диабету II типа); G аллель в SNP rs12255372 (ассоциированный с пониженной восприимчивостью к диабету II типа и являющийся аллелем, защищающим против диабета II типа); T аллель в SNP rs12255372 (ассоциированный с восприимчивостью к диабету II типа); G аллель в rs7895340 (ассоциированный с пониженной восприимчивостью к диабету II типа и являющийся аллелем, защищающим против диабета II типа); A аллель в SNP rs7895340 (ассоциированный с восприимчивостью к диабету II типа); G аллель в SNP rs11196205 (ассоциированный с восприимчивостью к диабету II типа); G аллель в rs7895340 (ассоциированный с пониженной восприимчивостью к диабету II типа и являющийся аллелем, защищающим против диабета II типа); C аллель в SNP rs1196205 (ассоциированный с восприимчивостью к диабету II типа); T аллель в rs7901695 (ассоциированный с пониженной восприимчивостью к диабету II типа и являющийся аллелем, защищающим против диабета II типа); C аллель в SNP rs7901695 (ассоциированный с восприимчивостью к диабету II типа); C аллель в SNP rs7903146 (ассоциированный с пониженной восприимчивостью к диабету II типа и являющийся аллелем, защищающим против диабета II типа); T аллель в SNP rs7903146 (ассоциированный с восприимчивостью к диабету II типа); C аллель в SNP rs12243326 (ассоциированный с восприимчивостью к диабету II типа) и T аллель в SNP rs4506565 (ассоциированный с восприимчивостью к диабету II типа). В дополнительных воплощениях изобретения для диагностики восприимчивости к диабету II типа или для идентификации аллеля, который защищает от диабета II типа, могут применяться другие маркеры или SNP, идентифицированные с применением способов, описанных здесь. Диагностические анализы, представленные ниже, могут применяться для идентификации присутствия или отсутствия этих специфических аллелей.

Диагностические анализы.

Нуклеиновые кислоты, зонды, праймеры и антитела, такие как описанные здесь, могут применяться в различных способах диагностики восприимчивости к диабету II типа, а также в наборах реактивов (например, полезных при диагностике восприимчивости к диабету II типа). Подобным образом, нуклеиновые кислоты, зонды, праймеры и антитела, описанные здесь, могут применяться в способах диагностики пониженной восприимчивости к диабету II типа, а также в способах диагностики защиты против диабета II типа, а также в наборах реактивов. В одном аспекте набор реактивов включает праймеры, которые могут применяться для амплификации интересующих маркеров.

В одном аспекте изобретения диагноз восприимчивости к диабету II типа ставится путем детектирования полиморфизма нуклеиновой кислоты в TCF7L2, как описано здесь (например, аллелей в маркере DG10S478 или в SNP rs12255372, rs7895340, rs11196205, rs7901695, rs7903146, rs12243326, rs4506565). Полиморфизм может быть изменением в нуклеиновой кислоте TCF7L2, таким как вставка или деления единственного нуклеотида или более чем одного нуклеотида, приводящая к сдвигу рамки считывания;

изменением по меньшей мере одного нуклеотида, приводящим к изменению кодируемой аминокислоты; изменением по меньшей мере одного нуклеотида, приводящим к генерации преждевременного стоп-кодона; делецией нескольких нуклеотидов, приводящей к делеции одной или более аминокислот, кодируемых нуклеотидами; вставкой одного или нескольких нуклеотидов, такой как неадекватная рекомбинация или конверсия гена, приводящая к прерыванию кодирующей последовательности гена; дупликацией всего или части гена; транспозицией всего или части гена или реаранжировкой всего или части гена. В единичном гене может быть представлено более чем одно такое изменение. Такие изменения последовательности вызывают различия в полипептиде, кодируемом нуклеиновой кислотой TCF7L2. Например, если различие представляет собой сдвиг рамки считывания, то сдвиг рамки приводит к изменению кодируемых аминокислот и/или может приводить к генерации преждевременного стоп-кодона, вызывая генерацию труктурированного (укороченного) полипептида. Альтернативно, полиморфизм, ассоциированный с заболеванием или состоянием или восприимчивостью к заболеванию или состоянию, ассоциированный с нуклеиновой кислотой TCF7L2, может быть синонимичным изменением в одном или более нуклеотидах (т.е. изменением, которое не приводит к изменению в полипептиде, кодируемом нуклеиновой кислотой TCF7L2). Такой полиморфизм может изменять сайты сплайсинга, влиять на стабильность или транспорт мРНК или другим образом влиять на транскрипцию или трансляцию гена. Нуклеиновая кислота TCF7L2, которая имеет любое изменение или перестройки, описанные выше, называется здесь "измененная нуклеиновая кислота".

В первом способе диагностики восприимчивости к диабету II типа могут применяться методы гибридизации, такие как саузерн анализ, нозерн анализ или гибридизации *in situ* (см. Current Protocols in Molecular Biology, Ausubel, F. et al., eds, John Wiley & Sons, включая все дополнения с 1999 года). Например, биологический образец ("тестируемый образец") из тестируемого субъекта ("тестируемый индивидуум") геномной ДНК, РНК или кДНК получают из индивидуума (для экзонных маркеров может применяться только РНК и кДНК), такого как индивидуум, подозреваемый в том, что он является восприимчивым к или предрасположенным к или несущим дефект диабета II типа. Индивидуум может быть взрослым, ребенком или плодом. Тестируемый образец может быть из любого источника, который включает геномную ДНК, таким как образец крови, образец амниотической жидкости, образец цереброспинальной жидкости или образец ткани из кожи, мышцы, буккальной слизистой или слизистой оболочки глаза, плацента, желудочно-кишечный тракт или другие органы. Тестируемый образец ДНК из фетальных клеток или ткани может быть получен подходящими способами, такими как амниоцентез или образцы ворсинок хориона. Затем образец ДНК, РНК или кДНК исследуется для определения того, имеется ли полиморфизм в нуклеиновой кислоте TCF7L2 и/или для определения, какой сплайсированный вариант(ы), кодируемый TCF7L2, представлен в нем. Присутствие полиморфизма или сплайсированного варианта(ов) может быть показано гибридизацией гена геномной ДНК, РНК или кДНК с нуклеотидным зондом. "Нуклеотидный зонд", как это применяется здесь, может быть зонд ДНК или зонд РНК; нуклеотидный зонд может включать, к примеру, по меньшей мере один полиморфизм нуклеиновой кислоты TCF7L2 и/или включать нуклеиновую кислоту, кодирующую специфический сплайсированный вариант нуклеиновой кислоты TCF7L2. Зонд может быть любой молекулой нуклеиновой кислоты, описанной выше (например, ген или нуклеиновая кислота, фрагмент, вектор, включающий ген или нуклеиновую кислоту, зонд или праймер, и т.д.).

Для диагностики восприимчивости к диабету II типа образец для гибридизации может быть сформирован путем контакта нуклеиновой кислоты TCF7L2 по меньшей мере с одним нуклеотидным зондом. Предпочтительным зондом для детектирования мРНК или геномной ДНК является меченый нуклеотидный зонд, способный к гибридизации с последовательностями мРНК или геномной ДНК, описанными выше. Нуклеотидный зонд может быть, к примеру, молекулой нуклеиновой кислоты полной длины или ее частью, такой как олигонуклеотид, имеющий длину по меньшей мере в 15, 30, 50, 100, 250 или 500 нуклеотидов, и достаточный для специфической гибридизации в жестких условиях с подходящей мРНК или геномной ДНК. Зонды, подходящие для применения в диагностических анализах изобретения, описаны выше (см., к примеру, зонды и праймеры, обсуждавшиеся под названием "Нуклеиновые кислоты изобретения").

Образец для гибридизации поддерживается в условиях, которые достаточны для того, что позволить специфическую гибридизацию нуклеотидного зонда с нуклеиновой кислотой TCF7L2. "Специфическая гибридизация", как это применяется здесь, означает правильную гибридизацию (например, без несоответствий). Специфическая гибридизация может быть выполнена в очень жестких условиях или в умеренных условиях, например, как описано выше. В особенно предпочтительном аспекте условия гибридизации для специфической гибридизации являются очень жесткими.

Специфическая гибридизация, если она присутствует, обнаруживается затем с применением стандартных способов. Если в тестируемом образце имеет место специфическая гибридизация между нуклеотидным зондом и нуклеиновой кислотой TCF7L2, то TCF7L2 имеет полиморфизм или является сплайсированным вариантом, который представлен в нуклеотидном зонде. Одновременно в этом способе может применяться более чем один нуклеотидный зонд. Специфическая гибридизация любого одного из нуклеотидных зондов является индикатором полиморфизма в нуклеиновой кислоте TCF7L2 или присут-

ствия определенного сплайсированного варианта, кодируемого нуклеиновой кислотой TCF7L2, и может быть диагностирована, как восприимчивость к диабету II типа или как пониженная восприимчивость к диабету II типа (или быть указанием на наличие защитного аллеля против диабета II типа).

В нозерн анализе (см. *Current Protocols in Molecular Biology*, Ausubel, F. et al., eds, John Wiley & Sons, supra) применяются способы гибридизации, описанные выше, для идентификации присутствия полиморфизма или специфического сплайсированного варианта, ассоциированного с восприимчивостью к диабету II типа или ассоциированного с пониженной восприимчивостью к диабету II типа. Для нозерн анализа тестируемый образец ДНК получают из индивидуума подходящим способом. Специфическая гибридизация нуклеотидного зонда, как описано выше, с РНК из индивидуума является индикатором полиморфизма в нуклеиновой кислоте TCF7L2, присутствия специфического сплайсированного варианта, кодируемого нуклеиновой кислотой TCF7L2, и является, таким образом, диагностикой восприимчивости к диабету II типа или с пониженной восприимчивостью к диабету II типа (или индикатором защитного аллеля против диабета II типа).

Для репрезентативных примеров применения нуклеотидных зондов см., к примеру, патенты US 5288611 и 4851330.

Альтернативно, вместо нуклеотидного зонда в способах гибридизации, описанных выше, может применяться зонд из пептидонуклеиновой кислоты (PNA). PNA является имитатором ДНК, имеющим пептидоподобный неорганический скелет, такой как N-(2-аминоэтил)глициновые единицы с органическими основаниями (A, G, C, T или U), прикрепленными к азоту глицина через метиленкарбонильный линкер (см., например, Nielsen, P.E. et al., *Bioconjugate Chemistry* 5, American Chemical Society, p.1 (1994)). Зонд PNA может быть сконструирован для специфической гибридизации с нуклеиновой кислотой TCF7L2. Гибридизация зонда PNA с нуклеиновой кислотой TCF7L2 может быть диагностикой восприимчивости к диабету II типа или пониженной восприимчивости к диабету II типа (или индикатором защитного аллеля против диабета II типа).

В другом способе изобретения может быть применен анализ изменения путем рестрикционного переваривания для обнаружения изменения в гене, если изменение (мутация) или полиморфизм в гене является результатом создания или элиминации рестрикционного сайта. Тестируемый образец, включающий геномную ДНК, получают из индивидуума. Для амплификации нуклеиновой кислоты TCF7L2 в тестируемом образце геномной ДНК из тестируемого индивидуума может быть применена полимеразная цепная реакция (ПЦР) (и, если необходимо, фланкирующие последовательности). RFLP анализ проводится, как описано (см. *Current Protocols in Molecular Biology*, Ausubel, F. et al., eds, John Wiley & Sons, supra). Профиль переваривания релевантного фрагмента ДНК указывает на присутствие или отсутствие изменения или полиморфизма нуклеиновой кислоты TCF7L2 и, таким образом, указывает на присутствие или отсутствие восприимчивости к диабету II типа или пониженную восприимчивость к диабету II типа (или указывает на наличие защитного аллеля против диабета II типа).

Для обнаружения специфического полиморфизма в нуклеиновой кислоте TCF7L2 может быть также применен анализ последовательности. Тестируемый образец ДНК или РНК получают из тестируемого индивидуума. ПЦР или другие подходящие методы могут быть применены для амплификации гена или нуклеиновой кислоты и/или его фланкирующих последовательностей, если это желательно. Последовательность нуклеиновой кислоты TCF7L2 или фрагмента нуклеиновой кислоты, или фрагмента кДНК, или мРНК или фрагмента мРНК определяется с применением стандартных методов. Последовательность нуклеиновой кислоты TCF7L2 или фрагмента нуклеиновой кислоты, или фрагмента кДНК, или мРНК или фрагмента мРНК сравнивается с известной последовательностью нуклеиновой кислоты или гена, или кДНК, или мРНК, в зависимости от того, что является более подходящим. Присутствие полиморфизма в TCF7L2 показывает, что индивидуум обладает восприимчивостью к диабету II типа или пониженной восприимчивостью к диабету II типа (или указывает на наличие защитного аллеля против диабета II типа).

Специфичные по отношению к аллелю олигонуклеотиды также могут применяться для выявления присутствия полиморфизма в нуклеиновой кислоте TCF7L2 путем применения дот-блот гибридизации амплифицированных нуклеотидов со специфичными по отношению к аллелю олигонуклеотидными зондами (ASO) (см., например, Saiki, R. et al., *Nature*. 324:163-166 (1986)). "Специфичные по отношению к аллелю олигонуклеотиды" (ASO) (называемые здесь также "специфичными по отношению к аллелю олигонуклеотидными зондами") представляют собой олигонуклеотиды с длиной приблизительно 10-15 пар оснований, предпочтительно приблизительно 15-30 пар оснований, которые специфически гибридизуются с нуклеиновой кислотой TCF7L2 и которые включают полиморфизм, ассоциированный с восприимчивостью к диабету II типа, или полиморфизм, ассоциированный с пониженной восприимчивостью к диабету II типа (или указывает на наличие защитного аллеля против диабета II типа). С применением стандартных способов может быть получен ASO зонд, специфичный для определенного полиморфизма в нуклеиновой кислоте TCF7L2 (см. *Current Protocols in Molecular Biology*, supra). Для идентификации полиморфизмов в гене, который ассоциирован с диабетом II типа, из индивидуума получают тестируемый образец ДНК. Для амплификации всей нуклеиновой кислоты TCF7L2 или ее фрагмента и ее фланкирующих последовательностей может быть применена ПЦР. ДНК, включающая амплифицированную

нуклеиновую кислоту TCF7L2 или фрагмента гена или нуклеиновой кислоты, подвергается дот-блоттингу с применением стандартных способов (см. *Current Protocols in Molecular Biology*, supra), и блот контактирует с олигонуклеотидным зондом. Затем детектируется наличие специфической гибридизации зонда с амплифицированной нуклеиновой кислотой TCF7L2. Гибридизация AOS с ДНК из индивидуума является индикатором полиморфизма в нуклеиновой кислоте TCF7L2 и, таким образом, индикатором восприимчивости к диабету II типа или индикатором пониженной восприимчивости к диабету II типа (или указывает на наличие защитного аллеля против диабета II типа).

Дополнительно изобретение обеспечивает AOS, которые гибридизуются с контрольным или различающимися аллелями гена или нуклеиновой кислоты, включающей единственный нуклеотидный полиморфизм, или к комплементарной ей молекуле. Эти нуклеотиды могут быть зондами или праймерами.

Специфичный к аллелю праймер гибридизуется с участком на ДНК-мишени, перекрывающим полиморфизм, и только начинает амплификацию аллельной формы, к которой праймер демонстрирует совершенную комплементарность. См. Gibbs, *Nucleic Acid Res.* 17, 2427-2448 (1989). Праймер применяется в соединении со вторым праймером, который гибридизуется в дистальном участке. Амплификация происходит с двух праймеров, приводя к получению детектируемого продукта, который указывает на присутствие определенной аллельной формы. Контроль обычно получают со второй парой праймеров, один из которых показывает единственное несоответствующее основание в полиморфном участке, а другой из них демонстрирует совершенную комплементарность с дистальным участком. Единственное несоответствующее основание предотвращает амплификацию, поэтому детектируемый продукт не образуется. Способ работает наилучшим образом, когда несоответствие включено в большинстве случаев в 3'-позицию олигонуклеотида, выровненного с полиморфизмом, поскольку эта позиция является наиболее дестабилизирующей для элонгации с праймера (см., например, WO 93/22456).

С добавлением таких аналогов, как заблокированные нуклеиновые кислоты (LNA), размер праймеров и зондов может быть редуцирован лишь до 8 оснований. LNA являются новым классом бициклических аналогов ДНК, в которых позиции 2' и 4' в фуранозном кольце соединены через О-метиленовую (окси-LNA) или аминометиленовую (амино-LNA) группировку. Общим для всех этих вариантов LNA является сродство к комплементарной нуклеиновой кислоте, которая общепризнанно является наиболее высоким из известных для аналогов ДНК. Например, было показано, что, в частности, все окси-LNA нуклеотиды имеют температуру плавления 64°C и 74°C, когда находятся в комплексе с комплементарной ДНК или РНК соответственно в отличие от 28°C для ДНК и РНК соответствующих ДНК нуклеотидов. Существенное увеличение температуры плавления получают также, когда LNA мономеры применяются в комбинации с мономерами стандартной ДНК или РНК. Для праймеров и зондов, в зависимости от того, где включены LNA мономеры (например, на 3'-конце, 5'-конце или посередине), температура плавления может быть значительно увеличена.

В другом аспекте матрица олигонуклеотидной пробы, которая является комплементарной к последовательности сегментов нуклеиновой кислоты-мишени из индивидуума, может быть применена для идентификации полиморфизма в нуклеиновой кислоте TCF7L2. Например, в одном аспекте может применяться олигонуклеотидная матрица. Олигонуклеотидные матрицы обычно включают множество различных олигонуклеотидных зондов, которые сопряжены с поверхностью субстрата в различных известных местах. Эти олигонуклеотидные матрицы, известные также как "Genechips<sup>TM</sup>", были описаны, к примеру, в патенте US 5143854 и в патентных публикациях PCT WO 90/15070 и 92/10092. Эти матрицы в основном могут быть получены с применением методов механического синтеза или направленного светом синтеза, который включает комбинацию фотолитографических методов и методов твердофазного синтеза олигонуклеотидов. См. Fodor et al., *Science*. 251:767-777 (1991), Pirrung et al., патент US 5143854 (см. также заявку PCT WO 90/15070 и Fodor et al., публикация PCT WO 92/10092 и патент US 5424186, включенные сюда в их целостности путем ссылки). Методы синтеза этих матриц с применением методов механического синтеза описаны, к примеру, в патенте US 5384261, включенном сюда в его целостности путем ссылки. В другом примере могут быть применены линейные матрицы.

Как только олигонуклеотидная матрица приготовлена, интересующая нуклеиновая кислота гибридизуется с матрицей и сканируется на полиморфизм. Гибридизация и сканирование проводятся в основном способами, описанными здесь, а также в опубликованной публикации PCT WO 92/10092 и WO 95/11995 и патенте US 5424186, которые включены сюда в их целостности путем ссылки. Кратко, последовательность нуклеиновой кислоты-мишени, которая включает один или более предварительно идентифицированных полиморфных маркеров, амплифицируется с применением хорошо известных методов амплификации, например ПЦР. Обычно для этого применяются последовательности праймеров, которые комплементарны двум нитям последовательности-мишени, расположенным ниже и выше полиморфизма. Может применяться техника асимметричной ПЦР. Затем амплифицированная мишень, в основном включающая метку, гибридизуется с матрицей в подходящих условиях. После выполнения гибридизации и промывания матрицы матрица сканируется для определения позиции, с которой гибридизуется последовательность-мишень. Данные гибридизации, полученные из скана, представлены обычно в виде интенсивности флуоресценции как функции расположения матрицы.

Хотя первично это было описано в терминах детекции единственного блока, например, для детектирования единичного полиморфизма, матрицы могут включать множество блоков детекции и, таким образом, способны анализировать множественные специфические полиморфизмы. В альтернативных аспектах в основном будет понятно, что блоки детекции могут быть сгруппированы в пределах единственной матрицы или множества отдельных матриц, так что различные оптимальные условия могут быть применены во время гибридизации мишени с матрицей. Например, часто может быть желательным обеспечить детекцию таких полиморфизмов, которые попадают в пределы отрезка геномной последовательности, обогащенного G-C, отдельно от тех, которые попадают в сегменты, обогащенные A-T. Это позволяет провести отдельную оптимизацию условий для каждой ситуации.

Дополнительное применение олигонуклеотидных матриц для детекции полиморфизма может быть обнаружено, к примеру, в патентах US 5858659 и 5837832, которые включены сюда в их целостности путем ссылки. Другие способы анализа нуклеиновой кислоты могут применяться для детекции полиморфизма в гене диабета II типа или вариантов, кодируемых геном диабета II типа. Репрезентативные способы включают, к примеру, прямое ручное секвенирование (Church and Gilbert, *Proc. Natl. Acad. Sci. USA.* 81:1991-1995 (1988); Sanger, F. et al., *Proc. Natl. Acad. Sci. USA.* 74:5463-5467 (1977), Beavis et al., патент US 5288644; автоматизированное флуоресцентное секвенирование; анализ полиморфизма одноцепочечной конформации (SSCP); зажатый электрофорез в геле в денатурирующих условиях (CDGE); градиентный электрофорез в геле в денатурирующих условиях (DGGE) (Sheffield, V.C. et al., *Proc. Natl. Acad. Sci. USA.* 86:232-236 (1989); анализ сдвига мобильности (Orita, M. et al., *Proc. Natl. Acad. Sci. USA.* 86:2766-2770 (1989); рестрикционный анализ (Flavell et al., *Cell.* 15:25 (1978); Greever, et al., *Proc. Natl. Acad. Sci. USA.* 78:5081 (1981); анализ гетеродуплекса; химическое расщепление в точке несоответствия (CMC) (Cotton et al., *Proc. Natl. Acad. Sci. USA.* 85:4397-4401 (1985); анализ защиты от РНКаз (Myers, R.M. et al., *Science.* 230:1242 (1985)); применение полипептидов, которые распознают несоответствие нуклеотидов, таких как mutS белок из *E.coli*; специфическая по отношению к аллелю ПЦР.

В одном аспекте изобретения диагноз восприимчивости к диабету II типа или пониженной восприимчивости к диабету II типа (или индикатор защитного аллеля против диабета II типа) может быть сделан также путем экспрессионного анализа путем количественной ПЦР (кинетические температурные циклы). Этот метод с применением TaqMan® анализа может определить присутствие изменения в экспрессии или состав полипептида, кодируемого нуклеиновой кислотой TCF7L2 или сплайсированными вариантами, кодируемыми нуклеиновой кислотой TCF7L2. Для идентификации полиморфизма могут применяться также TaqMan® зонды, независимо от того, является пациент гомо- или гетерозиготным. Дополнительно экспрессия вариантов может быть оценена количественно как физически или функционально различных.

В другом аспекте изобретения диагноз восприимчивости к диабету II типа или пониженной восприимчивости к диабету II типа (или индикатор защитного аллеля против диабета II типа) может быть сделан путем исследования экспрессии и/или состава полипептида TCF7L2 с применением различных способов, включая иммуноферментный анализ с использованием сорбента (ELISA), Вестерн-блоты, иммунопреципитацию и иммунофлуоресценцию. Тестируемый образец из индивидуума оценивается на присутствие изменения в экспрессии и/или изменения в составе полипептида, кодируемого нуклеиновой кислотой TCF7L2, или на присутствие определенного варианта, кодируемого нуклеиновой кислотой TCF7L2. Изменение в экспрессии полипептида, кодируемого нуклеиновой кислотой TCF7L2, может быть, например, изменением в количественной экспрессии полипептида (т.е. в количестве производимого полипептида); изменением в составе полипептида, кодируемого нуклеиновой кислотой TCF7L2, представляет собой изменение в качественной экспрессии полипептида (например, экспрессии измененного TCF7L2 полипептида или другого сплайсированного варианта). В предпочтительном аспекте диагноз восприимчивости к диабету II типа или пониженной восприимчивости к диабету II типа может быть сделан путем детекции определенного сплайсированного варианта, кодируемого этой нуклеиновой кислотой TCF7L2, или определенного спектра сплайсированных вариантов.

Могут быть представлены оба таких изменения (количественные и качественные). Термин "изменение" в экспрессии или составе полипептида, как он применяется здесь, относится к изменению в экспрессии или составе тестируемого образца по сравнению с экспрессией или составом полипептида нуклеиновой кислоты TCF7L2 в контрольном образце. Контрольный образец является образцом, который соответствует тестируемому образцу (например, происходит из того же типа клеток и из индивидуума, у которого не затронута восприимчивость к диабету II типа). Изменение в экспрессии или составе полипептида в тестируемом образце по сравнению с контрольным образцом является индикатором восприимчивости к диабету II типа. Подобным образом присутствие одного или более сплайсированных вариантов в тестируемом образце или присутствие значительно различающихся количеств различных сплайсированных вариантов в тестируемом образце по сравнению с контрольным образцом является индикатором чувствительности к диабету II типа. Могут применяться различные способы анализа экспрессии или состава полипептида, кодируемого нуклеиновой кислотой TCF7L2, включая спектроскопию, колориметрию, электрофорез, изоэлектрическое фокусирование и иммуноанализ (например, David et al., патент

US 4376110), такой как иммуноблоттинг (см. также Current Protocols in Molecular Biology, особенно главу 10). Например, в одном аспекте антитело способно связываться с полипептидом (например, как описано выше), предпочтительно может применяться антитело с детектируемой меткой. Антитела могут быть поликлональными или более предпочтительно, моноклональными. Может применяться интактное антитело или его фрагмент (например, Fab или F(ab')<sub>2</sub>). Термин "меченый" по отношению к зонду или антителу включает непосредственно меченый зонд или антитело, сопряженное (т.е. физически соединенное) с детектируемым веществом зонда, а также непрямо меченый зонд или антитело за счет его взаимодействия с другим реагентом, который помечен непосредственно. Примеры непрямого мечения включают детекцию первичного антитела с применением флуоресцентно меченого вторичного антитела и зонд ДНК, помеченный на конце биотином, так что это может быть обнаружено с флуоресцентно меченым стрептавидином.

Вестерн-блот анализ с применением антитела, как описано выше, которое специфически связывается с полипептидом, кодируемым измененной нуклеиновой кислотой TCF7L2, или антитела, которое специфически связывается с полипептидом, кодируемым неизменной нуклеиновой кислотой, или антитела, которое специфически связывается с определенным сплайсированным вариантом, кодируемым нуклеиновой кислотой, может применяться для идентификации присутствия в тестируемом образце определенного сплайсированного варианта или полипептида, кодируемого полиморфной или измененной нуклеиновой кислотой TCF7L2, или отсутствия в тестируемом образце определенного сплайсированного варианта или полипептида, кодируемого неполморфной или неизменной нуклеиновой кислотой. Присутствие полипептида, кодируемого полиморфной или измененной нуклеиновой кислотой, или отсутствие полипептида, кодируемого неполморфной или неизменной нуклеиновой кислотой, являются диагнозом на восприимчивость диабета II типа, как и присутствие (или отсутствие) определенных сплайсированных вариантов, кодируемых нуклеиновой кислотой TCF7L2.

В одном аспекте этого способа уровень или количество полипептида, кодируемого нуклеиновой кислотой TCF7L2, в тестируемом образце сравнивается с уровнем или количеством полипептида, кодируемого TCF7L2, в контрольном образце. Уровень или количество полипептида в тестируемом образце, которое является более высоким или низким, чем уровень или количество полипептида в контрольном образце, такие, что различие является статистически значимым, является индикатором изменения в экспрессии полипептида, кодируемого нуклеиновой кислотой TCF7L2, и диагностикой восприимчивости к диабету II типа. Альтернативно, состав полипептида, кодируемого нуклеиновой кислотой TCF7L2, в тестируемом образце сравнивается с составом полипептида, кодируемого нуклеиновой кислотой TCF7L2, в контрольном образце (например, присутствие различных сплайсированных вариантов). Различие в составе полипептида в тестируемом образце по сравнению с составом полипептида в контрольном образце является диагностикой восприимчивости к диабету II типа. В другом аспекте может быть определен и уровень, и состав полипептида в тестируемом и в контрольном образцах. Различие в количестве или уровне полипептида в тестируемом образце по сравнению с контрольным образцом; различие состава в тестируемом образце по сравнению с контрольным образцом или оба различия в количестве или уровне и различие в составе являются индикатором восприимчивости к диабету II типа.

Те же способы могут, наоборот, применяться для идентификации присутствия различия, когда сравниваются с контрольным (при заболевании) образцом. Отличие от контроля является индикатором пониженной восприимчивости к диабету II типа и/или индикатором присутствия защитного аллеля против диабета II типа.

Оценка маркеров и гаплотипов.

Популяции индивидуумов, демонстрирующих генетическое многообразие, не имеют идентичных геномов. Точнее, геном демонстрирует вариабельность последовательности между индивидуумами во многих участках генома; другими словами, в популяции имеется много полиморфных участков. В некоторых случаях делается ссылка на различные аллели в полиморфном участке без выбора аллеля сравнения. Альтернативно, контрольная последовательность может сравниваться с определенным полиморфным участком. Контрольный аллель иногда рассматривают как аллель "дикого типа", и он обычно выбирается либо как первый секвенированный аллель, либо как аллель из "не подверженного" болезни индивидуума (например, индивидуума, который не демонстрирует признаков заболевания или аномального фенотипа). Аллели, которые отличаются от контрольного, называют "вариантными" аллелями.

"Маркер", как описано здесь, относится к последовательности генома, характерной для определенного вариантного аллеля (т.е. полиморфного участка). Маркер может включать любой аллель любого вариантного типа, обнаруженного в геноме, включая SNP, микросателлиты, вставки, делеции, дупликации и транслокации.

Номенклатура SNP, как сообщается здесь, относится к официальному ID SNP (rs) идентификационному ярлыку, назначенному каждому уникальному SNP Национальным Центром Биотехнологической Информации (NCBI).

"Гаплотип", как это описано здесь, относится к сегменту цепи геномной ДНК, которая характеризуется специфической комбинацией генетических маркеров ("аллелей"), расположенных внутри сегмента. В определенном воплощении гаплотип может включать один или более аллелей, два или более аллелей,

три или более аллелей, четыре или более аллелей или пять или более аллелей. Генетический маркер является определенными "аллелями" в "полиморфных участках", ассоциированными с LD блоком 4 экзона TCF7L2. Как это применяется здесь, "LD блок экзона 4 TCF7L2" относится к LD блоку на Chr10q, в пределах которого наблюдается ассоциация вариантов с диабетом II типа. В соответствии с NCBI существует 34 позиции этого блока в пределах от 114413084 до 114488013 пар оснований. Термин "восприимчивость", как описывается здесь, включает как повышенную, так и пониженную восприимчивость. Таким образом, определенные маркеры и/или гаплотипы изобретения могут характеризовать повышенную восприимчивость к диабету II типа, что характеризуется относительным риском заболевания более чем единица. Более того, маркеры и/или гаплотипы, которые придают повышенную восприимчивость к диабету II типа, рассматриваются как наличие "риска", поскольку они обеспечивают повышенный риск заболевания. Альтернативно, маркеры и/или гаплотипы изобретения, характеризующие пониженную восприимчивость к диабету II типа, характеризуются относительным риском менее чем единица.

Позиция нуклеотида, в которой в популяции возможна более чем одна последовательность (либо в природной популяции, либо в синтетической популяции, например в библиотеке синтетических молекул), называется здесь "полиморфным участком". Там, где полиморфный участок имеет длину в один нуклеотид, участок называется SNP. Например, если в определенном участке хромосомы один член популяции имеет аденин, а другой член популяции имеет тимин в одной и той же позиции, то эта позиция является полиморфным участком, более точно SNP. Аллели для маркеров SNP, как это называется здесь, относятся к основаниям А, С, G или Т, поскольку они находятся в полиморфном участке в примененном анализе SNP. Специалист поймет, что путем определения или чтения противоположной нити в каждом случае может быть определен комплементарный аллель. Таким образом, для полиморфного участка, включающего полиморфизм А/С, примененный анализ может измерить либо процент, либо соотношение двух возможных оснований, т.е. А и С. Альтернативно, путем создания анализа, который определяет противоположную нить на матрице ДНК, может быть измерен процент или соотношение комплементарных оснований Т/С. Количественно (например, в терминах относительного риска) должны быть получены идентичные результаты из измерения любой нити ДНК (+нить или -нить). Участки полиморфизма могут позволить различать последовательности на основании замен, вставок или делеций. Например, полиморфные микросателлиты имеют множественные малые повторы оснований (такие как СА повторы) в определенном участке, в котором ряд дин повторов в основной популяции варьирует. Каждая версия последовательности в отношении участка полиморфизма называется здесь как "аллель" полиморфного участка. Таким образом, в предыдущем примере SNP позволяет присутствие обоих аденинового и тиминового аллелей. Обнаружено, что SNP и микросателлитные маркеры, локализованные в пределах LD блока экзона 4 TCF7L2, ассоциированные с диабетом II типа, описаны в табл. 2-7.

Типично контрольная последовательность относится к определенной последовательности. Аллели, которые отличаются от контрольной, называются "вариантные" аллели. Например, контрольная геномная последовательность ДНК между позициями 114413084 и 114488013 NCBI Конструкция 34 (равная 74929 парам оснований или 74,9 т.п.о.), которая относится к локализации в пределах хромосомы 10, описана здесь как SEQ ID NO: 1. Вариантная последовательность, как это применяется здесь, относится к последовательности, которая отличается от SEQ ID NO:1, но во всем остальном, по существу, подобна. Генетические маркеры, которые составляют гаплотипы, ассоциированные с LD блоком экзона 4 TCF7L2, являются вариантными. Дополнительно варианты могут включать изменения, которые влияют на полипептид, например полипептид, кодируемый геном TCF7L2. Эти различия в последовательности, когда они сравниваются с контрольной нуклеотидной последовательностью, могут включать вставки или делеции единственного нуклеотида или более чем одного нуклеотида. Такие различия в последовательности могут приводить к сдвигу рамки считывания; изменение по меньшей мере одного нуклеотида может приводить к изменению кодируемой аминокислоты; изменение по меньшей мере одного нуклеотида может приводить к генерации преждевременного стоп-кодона; делеция нескольких нуклеотидов может приводить к делеции одной или более аминокислот, кодируемых нуклеотидами; вставка одного или нескольких нуклеотидов, такая как неравная рекомбинация или конверсия гена, может приводить к прерыванию кодирующей последовательности рамки считывания; дубликации всей или части последовательности; транспозиции или реаранжировке нуклеотидной последовательности, как это описано здесь детально. Такие изменения последовательности изменяют полипептид, кодируемый нуклеиновой кислотой. Например, если изменение в последовательности нуклеиновой кислоты вызывает сдвиг рамки, то сдвиг рамки может привести к изменению в кодируемых аминокислотах и/или может привести к генерации преждевременного стоп-кодона, вызывая получение укороченного полипептида. Альтернативно, полиморфизм, ассоциированный с диабетом II типа или с восприимчивостью к диабету II типа, может быть синонимичным изменением в одном или более нуклеотидах (т.е. изменением, которое не приводит к изменению в аминокислотной последовательности). Такой полиморфизм может, например, изменять участки сплайсинга, влиять на стабильность или транспорт мРНК или по-другому влиять на транскрипцию или трансляцию кодируемого полипептида. Он может изменять также ДНК, увеличивая возможность структурных изменений, таких как амплификация или делеция, имеющих место на соматическом уровне в опухлях. Полипептид, кодируемый контрольной нуклеотидной последовательностью, является "кон-

трольным" полипептидом с определенной контрольной аминокислотной последовательностью, и полипептиды, кодируемые вариантными аллелями, называются "вариантными" полипептидами с вариантными аминокислотными последовательностями.

Полиморфные микросателлиты имеют множественные малые повторы оснований, которые имеют длину в 2-8 нуклеотидов (такие как CA повторы) в определенном участке, в котором ряд длин повторов варьирует в общей популяции. Вставка является общей формой полиморфизма, включающего малые вставки или делеции, которые типично имеют длину лишь несколько нуклеотидов.

Гаплотипы, описанные здесь, представляют собой комбинацию различных генетических маркеров, например SNP и микросателлитов, имеющих определенные аллели в полиморфных участках. Гаплотипы могут включать комбинацию различных генетических маркеров, детектирование гаплотипов, таким образом, может быть выполнено способами, известными специалисту, для детектирования последовательностей в полиморфных участках. Например, могут применяться стандартные методы для генотипирования присутствия SNP и/или микросателлитных маркеров, такие как основанные на флуоресцентных методах (Chen, X. et al., *Genome Res.* 9(5):492-98 (1999)), ПЦР, LCR, вложенная ПЦР и другие методы амплификации нуклеиновой кислоты. Эти маркеры и SNP могут быть идентифицированы в гаплотипах, придающих риск возникновения заболевания. Определенные способы идентификации важных маркеров и SNP включают LD и/или оценки LOD.

В определенных способах, описанных здесь, индивидуум, который имеет риск возникновения диабета II типа, является индивидуумом, у которого идентифицированы маркеры риска или гаплотип риска. В одном аспекте маркер или гаплотип риска является тем, что придает значительно увеличенный риск (или восприимчивость) диабета II типа. В одном воплощении значимость, ассоциированная с маркером или гаплотипом, измеряется относительным риском. В дополнительном воплощении значимость определяется процентом. В одном воплощении значительно увеличенный риск измеряется как относительный риск по меньшей мере примерно 1,2, включая, но не ограничиваясь, 1,2; 1,3; 1,4; 1,5; 1,6; 1,7; 1,8 и 1,9. В дополнительном воплощении относительный риск 1,2 является значимым. В дополнительном воплощении относительный риск по меньшей мере 1,5 является значимым. В дополнительном воплощении значимое увеличение риска, являющееся равным по меньшей мере 1,7, является значительным. В дополнительном воплощении значимое увеличение риска составляет по меньшей мере примерно 20%, включая, но не ограничиваясь этим, примерно 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 и 98%. В дополнительном воплощении значимое увеличение риска составляет по меньшей мере примерно 50%.

В других воплощениях изобретения маркер или гаплотип придает пониженный риск (пониженную восприимчивость) диабета II типа. В одном воплощении значимое уменьшение риска измеряется как относительный риск менее чем 0,9, включая, но не ограничиваясь, 0,9, 0,8, 0,7, 0,6, 0,5 и 0,4. В дополнительных воплощениях значимый относительный риск составляет менее чем 0,7. В другом воплощении снижение риска (или восприимчивости) составляет по меньшей мере примерно 20%, включая, но не ограничиваясь, примерно 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 и 98%. В дополнительном воплощении значимое уменьшение риска составляет по меньшей мере примерно 30%.

Таким образом, термин "восприимчивость к диабету II типа" указывает либо на повышенный риск либо на восприимчивость или пониженный риск или восприимчивость диабета II типа в количестве, которое является значимым, когда представлен определенный аллель, маркер, SNP или гаплотип; значимость измеряется, как указано выше. Термины "пониженный риск", "пониженная восприимчивость" и "защита против", как они применяются здесь, показывают, что относительный риск является пониженным соответственно, когда представлен определенный другой аллель, маркер, SNP и/или определенный гаплотип. Понятно, однако, что идентификация того, является ли пониженный или повышенный риск значимым в медицине, может также зависеть от множества факторов, включая специфическое заболевание, маркер или гаплотип и часто факторы окружающей среды.

Маркер или гаплотип риска, включая часть, в гене TCF7L2 является таковым, где маркер или гаплотип наиболее часто представлены у индивидуумов с риском диабета II типа (подверженным) по сравнению с частотой его присутствия у здорового индивидуума (контрольного) и где присутствие маркера или гаплотипа является индикатором восприимчивости к диабету II типа. Примером простого теста на корреляцию будет строгий тест Фишера попарно в таблице. Данная группа хромосом попарно в таблице сконструирована из ряда хромосом, которые включают оба маркера или гаплотипа, один из маркеров или гаплотипов, но не другой, и ни один из маркеров или гаплотипов.

В определенных аспектах изобретения маркер или гаплотип риска является маркером или гаплотипом риска в пределах или вблизи TCF7L2, что значимо коррелирует с диабетом II типа. В других аспектах маркер или гаплотип риска включает маркер или гаплотип риска в пределах или вблизи TCF7L2, что значимо коррелирует с восприимчивостью к диабету II типа. В определенном воплощении маркер или гаплотип ассоциирован с LD блоком экзона 4 TCF7L2, как это описано здесь.

Могут применяться стандартные способы генотипирования присутствия SNP и/или маркера микросателлита, такие как методы, основанные на флуоресценции (Chen, et al., *Genome Res.* 9, 492 (1999)), ПЦР, LCR, вложенная ПЦР и другие методы амплификации нуклеиновой кислоты. В предпочтительном аспекте способ включает оценку у индивидуума присутствия или частоты SNP и/или, включая часть,

гена TCF7L2, где избыток или более высокая частота SNP и/или микросателлитов по сравнению со здоровыми контрольными индивидуумами является индикатором того, что индивидуум восприимчив к диабету II типа. Такие SNP и маркеры могут формировать гаплотипы, которые могут применяться как инструмент для скрининга. Эти маркеры и SNP могут быть идентифицированы как гаплотипы риска. Например, гаплотип риска может включать маркеры микросателлиты и/или SNP, такие как маркер DG10S478 и/или SNP rs12255372, rs7895340, rs11196205, rs7901695, rs7903146, rs12243326 и rs4506565. Присутствие гаплотипа риска является индикатором повышенной восприимчивости к диабету II типа и, таким образом, индикатором индивидуума, который попадает в популяцию для лечения способами, описанными здесь.

Идентификация вариантов восприимчивости.

Частоты гаплотипов у пациентов и контрольных групп могут быть оценены с применением алгоритма максимизации ожидания (Dempster A. et al., J. R. Stat. Soc. D, 39:1-38 (1977)). Может быть применено осуществление этого алгоритма, которое может контролировать отсутствующие генотипы и неопределенность с фазой. В нулевой гипотезе предполагается, что пациенты и контроли имеют идентичные частоты. С применением вероятностного подхода тестируется альтернативная гипотеза, где кандидату на гаплотип риска, который может включать маркеры, описанные здесь, позволено иметь более высокую частоту у пациентов, чем в контроле, в то время как соотношение частот других гаплотипов предполагается одинаковым в обеих группах. Вероятность максимизируется отдельно для обеих гипотез, и применяется соответствующий 1-df статистический коэффициент вероятности для оценки статистической значимости.

Для поиска маркеров риска и защитных маркеров в пределах сцепленной области, к примеру, исследуется ассоциация всех возможных комбинаций генотипированных маркеров, в том случае, если эти маркеры простираются в настоящую область. Комбинированные группы пациентов и контрольные группы могут быть беспорядочно разделены на два набора, равные по размеру оригинальной группе пациентов и контрольных индивидуумов. Затем анализ маркера и гаплотипа повторяется, и определяются наиболее значимые зарегистрированные значения  $p$ . Эта рандомизированная схема затем может быть повторена, например, более 100 раз для конструирования эмпирического распределения значений  $p$ . В предпочтительном воплощении значение  $p < 0,05$  является индикатором значимой ассоциации маркера и/или гаплотипа.

Далее следует подробное обсуждение анализа гаплотипа.

Анализ гаплотипа.

Один общепринятый подход к анализу гаплотипа вовлекает применение основанного на вероятности заключения, применимого к Nested Models (Gretarsdottir S., et al., Nat. Genet. 35:131-38 (2003)). Способ выполняется с помощью программы NEMO, которая учитывает многие полиморфные маркеры, SNP и микросателлиты. Метод и программное обеспечение специально созданы для изучения пары случай-контроль, где цель заключается в идентификации группы гаплотипов, которая придает различные риски заболевания. Это также инструмент для исследования LD структур. В NEMO максимальная вероятность оценивается, коэффициенты вероятности и значения  $p$  непосредственно рассчитываются с поддержкой EM алгоритма для наблюдаемых данных, обрабатывая их как в случае проблемы отсутствия данных.

Информация об измерениях.

Даже хотя тесты коэффициента вероятности, основанные на вероятности, вычисленной непосредственно для наблюдаемых данных, которые фиксировали информацию о потере, обусловленную неопределенностью в фазе, и отсутствующих генотипах, позволяют надеяться на получение ценных значений  $p$ , было бы интересно узнать, как много информации было потеряно, вследствие того, что информация была неполной. Измерение информации для анализа гаплотипа описано у Nicolae and Kong (Technical Report 537, Department of Statistics, University of Chicago; Biometrics, 60(2):368-75 (2004)) как натуральное распространение информационных пределов для анализа связи и выполняется с помощью NEMO.

Статистический анализ.

Для ассоциации единственного маркера с заболеванием может быть применен строгий тест Фишера для расчета двусторонних значений  $p$  для каждого индивидуального аллеля. Все значения  $p$  представлены не подведенными до множественных сравнений, пока это не указано специально. Представленные частоты (для микросателлитов, SNP и гаплотипов) являются частотами аллелей в противоположность несущим частотам. Для минимизации любого смещения, обусловленного связанностью пациентов, которые были набраны как семейства для анализа взаимосвязи, родственники первой и второй степени могли быть удалены из списка пациентов. Более того, тест для ассоциации можно повторить, чтобы ввести поправки на любую сохраняющуюся связанность среди пациентов путем процедуры распространения подведения расхождений, описанной у Rich, N. & Teng, J. (Genome Res., 8:1273-1288 (1998)), создания общего фонда ДНК (ibid) для кровных родственников так, чтобы он мог применяться для основных семейных взаимоотношений и для сравнения представляться как доведенными, так и не доведенными значениями  $p$ . Как ожидалось, различия в основном были незначительными. Для оценки значимости ассоциации с единственным маркером, скорректированной для множественного тестирования, авторы изобретения могли проводить тест рандомизации с применением данных того же генотипа. Группа пациентов и кон-

трольная группа могли быть рандомизированы, и анализ ассоциации сделан повторно много раз (например, до 500000 раз), и значение  $p$  представляет собой фракцию повторных анализов, которые дают значение  $p$  для того же маркерного аллеля, которое ниже или равно значению  $p$ , наблюдаемому с применением оригинальной группы пациентов или контрольной группы. Для обоих типов анализа (единичного маркера и гаплотипа) относительный риск (RR) и приписанный к популяции риск могут быть рассчитаны на основе предположения множественной модели (гаплотип относительно модели риска) (Terwilliger, J.D. & Ott, J. Hum. Hered. 42:337-46 (1992) and Falk, C.T. & Rubinstein, H., Ann. Hum. Genet. 51 (Pt. 3):227-33 (1987)), т.е. что риски двух аллелей/гаплотипов индивидум несет, умножая. К примеру, если RR представляет собой риск А относительно а, то риск гомозиготной персоны АА будет  $RR^2$  раз от такового для гетерозиготы Аа и  $RR^2$  раз от риска гомозиготной персоны аа. Мультипликативная модель имеет замечательное свойство, которое упрощает анализ и расчеты - гаплотипы являются независимыми, т.е. находясь в равновесии Харди-Вайнберга в пределах подверженной заболеванию популяции, а также в пределах контрольной популяции. Вследствие этого каждый гаплотип, рассчитанный для подверженных заболеванию и контролей, имеет полиномиальное распределение, но с различными частотами гаплотипа при альтернативных гипотезах. В частности, для двух гаплотипов,  $h_i$  и  $h_j$ ,  $\text{risk}(h_i/h_j) = (f_i/p_i)(f_j/p_j)$ , где  $f$  и  $p$  означают соответственно частоты в подверженной заболеванию популяции и в контрольной популяции. В то время как имеется некоторая потеря силы, если истинная модель не является мультипликативной, потеря имеет тенденцию быть умеренной, исключая экстремальные случаи. Наиболее важным является то, что значения  $p$  всегда являются правдоподобными, поскольку они рассчитаны по отношению к нулевой гипотезе.

Нарушение равновесия сцепления с применением NEMO.

LD между парами маркеров можно рассчитать с применением стандартного определения  $D'$  и  $R^2$  (Lewontin, R., Genetics. 49:49-67 (1964); Hill, W.G. & Robertson, F. Theor. Appl. Genet. 22:226-231 (1968)). С применением программы NEMO частоты комбинаций двух маркерных аллелей оцениваются с максимальной вероятностью, а отклонение от равновесия взаимосвязи оценивается с помощью теста коэффициента вероятности. Определения  $D'$  и  $R^2$  расширены с включением микросателлитов путем усреднения значений для всех возможных аллельных комбинаций двух маркеров, взвешенных посредством крайних вероятностей аллелей. Размещая все комбинации маркеров для выявления структуры LD в определенной области, помещали  $D'$  в верхний левый угол, а значение  $p$  - в нижний правый угол. В LD графике, если это желательно, маркеры могут быть размещены равноудаленно, а не в соответствии с их физической локализацией.

Статистические способы анализа сцепления.

В анализах оценки свидетельств сцепления может быть применен многоточечный, приписанный только к доле одного аллеля, метод. Результаты, такие как LOD баллы и непараметрическая оценка сцепления (NPL), могут быть получены с применением программы Allegro (Gudbjartsson et al., Nat. Genet. 25:12-3 (2000)). Наш базовый анализ сцепления применяет бальную функцию  $S_{\text{pairs}}$  (Wittemore, A.S., Halpern, J. Biometrics. 50:118-27 (1994); Kruglyak L., et al., Amer. J. Hum. Genet. 58:1347-63 (1996)), экспоненциальную модель распределения аллеля (Kong, A. and Cox, N.J. Am. J. Hum. Genet. 61:1179-88 (1997)) и семейно взвешенную схему, которая на логарифмической шкале представляет собой половину расстояния между равным взвешиванием каждой подверженной заболеванию пары и равным взвешиванием каждой семейной пары. Информационный показатель, который применяется, является частью данных на выходе программы Allegro, и информационный показатель равен нулю, если маркер генотипов является полностью информативным, и равен единице, если генотипы определяют точное количество распределения аллеля среди подверженных заболеванию родственников нестрогим образом (Gretarsdottir et al., Am. J. Hum. Genet., 70:593-603 (2002)). Значения  $P$  были рассчитаны двумя различными путями, и здесь приводится менее значимый результат. Первое значение  $P$  может быть рассчитано на основе теории больших выборок; распределение  $Z_{lr} = \sqrt{2[\log_e(10)LOD]}$  аппроксимирует стандартную нормальную переменную в нулевой гипотезе сцепления (Kong, A. and Cox, N.J. Am. J. Hum. Genet. 61:1179-88 (1997)). Второе значение  $P$  может быть рассчитано путем сравнения наблюдаемой LOD-оценки с его полными данными распределения образца в нулевой гипотезе (например, Gudbjartsson et al., Nat. Genet. 25:12-3 (2000)).

Когда данные состоят из более чем нескольких семейств, эти два значения  $P$  проявляют тенденцию быть очень похожими.

Определение гаплотипов и "гаплотипного блока" локуса восприимчивости.

В определенных воплощениях анализ маркера и гаплотипа вовлекает определение локуса-кандидата на восприимчивость на основе "блоков гаплотипа" (называемых также "LD блоками"). Сообщалось о том, что части генома человека могут быть разделены на серии дискретных гаплотипных блоков, включающих несколько общих гаплотипов; для этих блоков данные о нарушении сцепления обеспечивали мало доказательств, указывающих на рекомбинацию (см., например, Wall J.D. and Pritchard J.K. Nature Reviews Genetics. 6:587-597 (2003); Daly, M. et al., Nature Genet. 29:229-232 (2002); Patil, N. et al., Science. 294:1719-1723 (2001); Dawson, E. et al., Nature. 418:544-548 (2002); Phillips, M.S. et al., Nature. Genet. 33:382-387 (2003)).

Существует два основных способа определения этих гаплотипных блоков: блоки могут быть определены как области ДНК, которые имеют ограниченное разнообразие гаплотипов (см., например, Daly M. et al., *Nature Genet.* 29:229-232 (2002); Patil, N. et al., *Science.* 294:1719-1723 (2001); Dawson, E. et al., *Nature.* 418:544-548 (2002); Zang, K. et al., *Proc. Natl. Acad. Sci. USA.* 99:7335-7339 (2002)), или как области между переходными зонами, имеющие большую историческую рекомбинацию, идентифицированную с применением выявлений нарушения сцепления (см., например, Gabriel, S.B. et al., *Science.* 296:2225-2229 (2002); Phillips, M.S. et al., *Nature Genet.* 33:382-387 (2003); Wang et al., *Am. J. Hum. Genet.* 71:1227-1234 (2002); Stumpf, M.P., and Goldstein, D.B. *Curr. Biol.* 13:1-8 (2003)). Как это применяется здесь, термины "гаплотипный блок" или "LD блок" включают блоки, определенные с помощью любой из двух характеристик.

Репрезентативные способы идентификации блоков гаплотипов приведены выше, например, в опубликованных патентных заявках US 20030099964, 200330170665, 20040023237 и 20040146870. Блоки гаплотипов могут легко применяться для ассоциации на карте между фенотипов и статусом гаплотипа. Основные гаплотипы могут быть идентифицированы в каждом блоке гаплотипов, а затем может быть идентифицирован набор "меченых" SNP или маркеров (самый малый набор SNP или маркеров, необходимых для различения гаплотипов). Эти меченые SNP или маркеры могут затем применяться в оценке образцов из группы индивидуумов для идентификации ассоциации между фенотипом и гаплотипом. Если желательно, одновременно могут быть оценены соседние блоки гаплотипов, поскольку может также существовать нарушение сцепления между блоками гаплотипов.

#### Гаплотипы и диагностика.

Как описано здесь, было обнаружено, что определенные маркеры и гаплотипы, включающие такие маркеры, полезны при определении восприимчивости к диабету II типа, т.е. обнаружено, что они полезны при диагностике восприимчивости к диабету II типа. У индивидуумов с диабетом II типа они обнаруживаются чаще, чем у индивидуумов без диабета II типа. Таким образом, эти маркеры и гаплотипы имеют предсказательное значение для обнаружения диабета II типа или восприимчивости к диабету II типа у индивидуума. Блок гаплотипов (т.е. LD блок экзона 4 TCF7L2), включающий определенные меченые маркеры, может быть обнаружен более часто у индивидуумов с диабетом II типа, чем у индивидуумов без диабета II типа. Таким образом, эти несущие "риск" меченые маркеры в пределах блока гаплотипов также обладают предсказательной силой для выявления диабета II типа или восприимчивости к диабету II типа у индивидуума. Несущие "риск" меченые маркеры в пределах гаплотипа или LD блоков могут также включать другие маркеры, которые проводят различие среди гаплотипов, поскольку они подобным образом имеют предсказательную силу для выявления диабета II типа или восприимчивости к диабету II типа. Вследствие гаплотипной блоковой структуры человеческого генома может быть обнаружено, что большое число маркеров или других вариантов и/или гаплотипов, включающих такие маркеры или варианты, в ассоциации с блоком гаплотипов (LD блок) ассоциированы с определенной особенностью и/или фенотипом. Таким образом, возможно, что маркеры и/или гаплотипы, находящиеся в пределах LD блока экзона 4 TCF7L2, как это определено здесь, или в строгом LD (характеризуясь  $r^2$  более чем 0,2) с LD блоком экзона 4 TCF7L2, ассоциированы с диабетом II типа (т.е. они придают увеличенную или пониженную восприимчивость к диабету II типа). Это включает маркеры, которые описаны здесь (табл. 6), но может также включать другие маркеры, которые находятся в строгом LD (характеризуясь  $r^2$  более чем 0,2) с одним или более маркерами, перечисленными в табл. 6. Идентификация таких дополнительных вариантов может быть достигнута способами, хорошо известными специалисту, например путем секвенирования ДНК LD блока геномной области А у определенной группы индивидуумов, и настоящее изобретение также включает такие дополнительные варианты.

Как это описано здесь, определенные маркеры в пределах LD блока экзона 4 TCF7L2 обнаруживаются с более низкой частотой у индивидуумов с диабетом II типа, и гаплотипы, включающие два или более таких маркеров, перечисленных в табл. 13, 20 и 21, также обнаруживаются с более низкой частотой у индивидуумов с диабетом II типа. Следовательно, эти маркеры и гаплотипы являются защитными для диабета II типа, т.е. они придают пониженный риск развития диабета II типа для индивидуумов, несущих эти маркеры и/или гаплотипы.

Гаплотипы и маркеры, описанные здесь, являются в некоторых случаях комбинацией различных генетических маркеров, например SNP и микросателлитов. Таким образом, детектирование гаплотипов может быть проведено способами, известными специалисту и/или описанными здесь для выявления последовательностей в полиморфных участках. Более того, корреляция между определенными гаплотипами или набором маркеров и фенотипом заболевания может быть верифицирована с применением стандартных методов. Репрезентативный пример простого теста для корреляции будет строгий тест Фишера попарно в таблице.

В специфических воплощениях маркер или гаплотип, ассоциированный с LD блоком экзона 4 TCF7L2, является таким, в котором маркер или гаплотип наиболее часто представлен у индивидуума с риском для диабета II типа (подверженные заболеванию) по сравнению с частотой его присутствия у здорового индивидуума (контроль), где присутствие маркера или гаплотипа является индикатором диабета II типа или восприимчивости к диабету II типа. В других воплощениях меченые маркеры, несущие

риск заболевания, в LD с одним или более маркерами, ассоциированными с LD блоком экзона 4 TCF7L2, являются мечеными маркерами, которые наиболее часто представлены у индивидуумов с риском диабета II типа (подверженные заболеванию) по сравнению с частотой его присутствия у здорового индивидуума (контроль), где присутствие меченого маркера или гаплотипа является индикатором повышенной восприимчивости к диабету II типа. В дополнительном воплощении маркеры риска в LD с одним или более маркерами, ассоциированными с LD блоком экзона 4 TCF7L2, являются маркерами, которые наиболее часто представлены у индивидуума с риском диабета II типа по сравнению с частотой их присутствия у здорового индивидуума (контроль), где присутствие маркеров является индикатором восприимчивости к диабету II типа.

В определенных способах, описанных здесь, индивидуум, который имеет риск развития диабета II типа, представляет собой индивидуума, у которого идентифицирован маркер риска. В одном воплощении сила ассоциации маркера или гаплотипа измеряется относительным риском (RR). RR является соотношением долей состояния у субъектов, которые несут одну копию маркера или гаплотипа, и долей состояния у субъектов, которые не имеют такой маркер или гаплотип. Это соотношение эквивалентно соотношению доли состояния среди субъектов, которые несут две копии маркера или гаплотипа к доле состояния у субъектов, которые несут одну копию маркера или гаплотипа. В одном воплощении маркер или гаплотип имеет RR по меньшей мере 1,2. В другом воплощении маркер или гаплотип имеет RR по меньшей мере 1,3, по меньшей мере 1,4, по меньшей мере 1,5, по меньшей мере 2,0, по меньшей мере 2,5, по меньшей мере 3,0, по меньшей мере 3,5, по меньшей мере 4,0 или по меньшей мере 5,0.

В других способах изобретения индивидуум, который имеет пониженный риск (или пониженную восприимчивость) диабета II типа, является индивидуумом, у которого идентифицирован защитный маркер или гаплотип. В таких случаях RR составляет менее чем 1,0. В одном воплощении маркер или гаплотип имеет RR менее чем 0,9. В других воплощениях маркер или гаплотип имеет RR менее чем 0,8, менее чем 0,7, менее чем 0,6, менее чем 0,5, менее чем 0,4.

Полезность генетического тестирования.

Знание о генетическом варианте, который несет риск развития диабета II, открывает возможность генетического тестирования для разграничения между индивидуумами с повышенным риском развития заболевания (т.е. несущих вариант риска) и теми, кто имеет пониженный риск развития заболевания (т.е. несет защитный вариант). Основным достоинством генетического тестирования для индивидуумов, принадлежащих к обеим упомянутым выше группам, являются возможности диагностики заболевания на ранней стадии и обеспечение информации для клиницистов о прогнозе/агрессивности заболевания, для того чтобы быть способными применить наиболее подходящее лечение. Например, применение генетического теста для диабета II типа может обеспечить возможность детекции заболевания на ранней стадии, которое может привести к использованию терапевтических мер на ранней стадии и, таким образом, может минимизировать неблагоприятные эффекты симптомов и серьезные последствия для здоровья, характерные для диабета II типа.

Способы терапии.

В другом воплощении изобретения могут быть применены способы лечения диабета II типа. Термин "лечение", как он применяется здесь, относится не только к облегчению симптомов, ассоциированных с диабетом II типа, но также к предотвращению или задержке начала диабета II типа; к снижению тяжести или частоты симптомов диабета II типа и/или к снижению также потребности в сопутствующей терапии другими лекарствами, которые облегчают симптомы, ассоциированные с диабетом II типа. В одном аспекте индивидуум, который лечится, является индивидуумом, который восприимчив (имеет повышенный риск), диабета II типа (например, индивидуум, у которого присутствуют аллели, иные, чем 0 аллель в маркере DG10S478, присутствует Т аллель в SNP rs12255372, присутствует А аллель в SNP rs7895340, присутствует С аллель в SNP rs11196205, присутствует С аллель в SNP rs7901695, присутствует Т аллель в SNP rs7903146, присутствует С аллель в SNP rs12243326 или присутствует Е аллель в SNP rs4506565).

В дополнительном воплощении изобретения могут быть применены способы лечения других заболеваний или состояний, ассоциированных с TCF7L2. Терапевтический агент TCF7L2 может применяться как в способах лечения диабета II типа, так и в способах лечения других заболеваний или состояний, ассоциированных с TCF7L2.

В способах лечения (профилактических и/или терапевтических) применяется терапевтический агент TCF7L2. "Терапевтический агент TCF7L2" представляет собой агент, который изменяет (например, усиливает или ингибирует) активность полипептида и/или экспрессию нуклеиновой кислоты TCF7L2, непосредственно или непрямым способом (например, путем изменения активности экспрессии нуклеиновой кислоты, кодирующей белок, взаимодействующий с TCF7L2, такой как белок в Wnt сигнальном пути или в кадгериновом сигнальном каскаде (например,  $\beta$ -катенина)). В определенных воплощениях терапевтический агент TCF7L2 изменяет активность и/или экспрессию нуклеиновой кислоты TCF7L2.

Терапевтический агент TCF7L2 может изменять активность полипептида TCF7L2 или экспрессию нуклеиновой кислоты различными способами, такими как, к примеру, обеспечение дополнительного полипептида TCF7L2 или путем активации регуляции транскрипции или трансляции нуклеиновой кислоты TCF7L2; путем изменения посттрансляционного процессинга полипептида TCF7L2; путем изменения транскрипции сплайсированных вариантов TCF7L2 или путем вмешательства в активность полипептида TCF7L2 (например, путем связывания с полипептидом TCF7L2) или путем связывания с другим полипептидом, который взаимодействует с TCF7L2, путем изменения (например, путем обратной регуляции) экспрессии, транскрипции или трансляции нуклеиновой кислоты TCF7L2 или путем изменения (например, агонистами или антагонистами) активности.

Репрезентативные терапевтические агенты TCF7L2 включают следующее: нуклеиновые кислоты или фрагменты или производные этого, описанные здесь, определенные нуклеотиды, кодирующие полипептиды, описанные здесь, и векторы, включающие такие нуклеиновые кислоты (например, ген, кДНК и/или мРНК, кодирующие полипептид TCF7L2 или активный фрагмент, или производное этого; или олигонуклеотид, или дополнительный по отношению к нему, или фрагмент, или производное этого, и/или сплайсированные варианты, кодирующие нуклеиновую кислоту диабета II типа; или фрагменты, или производные этого); полипептиды, описанные здесь, и/или сплайсированные варианты, кодируемые нуклеиновой кислотой TCF7L2, или фрагменты, или производные этого; другие полипептиды (например, рецепторы TCF7L2); агенты, связывающие TCF7L2, или агенты, которые влияют (например, увеличивают или уменьшают) на активность, антитела, такие как антитело на неизмененный полипептид TCF7L2 или антитело против определенного сплайсированного варианта, кодируемого нуклеиновой кислотой TCF7L2, как описано выше; имитаторы пептида, слитые белки или пролекарства, рибозимы, другие малые молекулы и другие агенты, которые изменяют (например, усиливают или ингибируют) экспрессию нуклеиновой кислоты TCF7L2, или которые регулируют транскрипцию сплайсированных вариантов TCF7L2 (например, агенты, которые влияют на то, какой из сплайсированных вариантов экспрессируется), или которые влияют на количество каждого сплайсированного варианта, который экспрессируется. Дополнительные репрезентативные терапевтические агенты TCF7L2 включают соединения, которые влияют на сигнальный каскад инсулина и/или глюкагона, GLP-1 или GIP сигнальные пути. Если желательно, одновременно может применяться более чем один терапевтический агент TCF7L2.

В предпочтительных воплощениях терапевтический агент TCF7L2 представляет собой агент, который мешает проявлению активности TCF7L2, такой, к примеру, как агент, который мешает связыванию или взаимодействию TCF7L2 с бета-катенином (см., например, Fasolini, et al., J. Biol. Chem. 278(23):21092-21096 (2003)) или с другими белками. Другие терапевтические агенты TCF7L2 включают агенты, которые влияют на сигнальный путь Wnt, или агенты, которые влияют на кадгериновый сигнальный каскад. Репрезентативные агенты включают агенты, такие как те, что применяются в терапии рака, белки, такие как DKK белки; бета-катенин-связывающий домен APC или аксин; факторы, такие как IDAX, AXAM и ICAT; антисенс олигонуклеотиды или интерферирующая РНК (РНК-и), такие как с применением Vitravet, онколитические вирусные векторы и другие соединения (см., к примеру, Liu, et al., Current Cancer Drug Target. 4:6530671 (2004)); малые молекулы антагонистов, включающие, например, ZTM00990, PKF118-310, PKF118-584, PKF22-815, CGPO49090, NPDDG39.024 и NPDDG1.024, как описано у Lepourcelet et al. (см., к примеру, Lepourcelet et al., Cancer Call. 5:91-102 (2004)); соединения, описанные в патенте US 6762185, соединения, описанные в заявках на патент US 200400053136, 20040072831, 20040247593 или 20050059628. Другие репрезентативные терапевтические агенты TCF7L2 включают ингибиторы gsk3, включая, к примеру, таковые, описанные в патентах US 6057117, 6153618, 6417185, 6465231, 6489344, 6512102, 6608063, 6716624, 6800632 и опубликованных заявках на патенты US 200300088866, 20030077798, 20030130289, 20030207883, 1000092535 и 200500851. Все описанное в этих ссылках, патентах и заявках на патенты повторно цитируется в Спецификации и включено сюда в их целостности.

Дополнительные репрезентативные терапевтические агенты TCF7L2 показаны в таблице агентов.

Таблица агентов

Название соединения	Название соединения (созданное с помощью Autonom, ISIS Draw версии 2,5 от MDL Information Systems)	Компания	Ссылка на соединение	Показания
AR-0133418 (SN-4521)	1-(4-Метокси-бензил-3-(5-нитрогиазо-2-ил)-мочевина	AstraZeneca		AD
AR-025028	NSD	AstraZeneca		
CT-98023	N-[4(2,4-дихлоро-фенил)-5-(1H-имидазол-2-ил)-пиримидин-2-ил]-N'-(5-нитро-пиридин-2-ил)-этан-1,2-диамин	Chiron Corp		Инсулин-независимый диабет
CT-20026	NSD	Chiron Corp	Wagman et al., Curr. Pharm. Des 2004;10(10)1105-37	Инсулин-независимый диабет
CT-21022	NSD	Chiron Corp		Инсулин-независимый диабет
CT-20014	NSD	Chiron Corp		Инсулин-независимый диабет
CT-21018	NSD	Chiron Corp		Инсулин-независимый диабет
CHIR-98025	NSD	Chiron Corp		Инсулин-независимый диабет

CHIR-99021	NSD	Chiron Corp	Wagman et al., Curr. Pharm. Des 2004;10(10)1105-37	Инсулин-независимый диабет
CG-100179	NSD	Crystal Genomics and Yuyu	WO-2004065370	Сахарный диабет (Корея)
	4-[2-(4-диметидамино-3-нитрофениламино)-пиримидин-4-ил]-3,5-диметил-1Н-пиррол-2-карбонитрил	Cyclacel Ltd		Инсулин-независимый диабет среди Других
NP-01139, NP-031112, NP-03112, NP-036,	4-бензил-2-метил-[1,2,4]гиадиазолидин-3,5-дион	Neuropharma SA		Расстройства ЦНС, AD
	3-[9-фтор-2-(пиперидин-1-карбонил)-1,2,3,4-тетрагидро[1,4]дiazепино [6,71-хи]индол-7-ил]-4-имидазо[1,2-а]пиридин-3-ил-пиррол-2,5-дион	Eli Lilly&Co		Инсулин-независимый диабет
GW-784752x, GW-784775, SB-216763, SB-415286,	Циклопентанкарбоновой кислоты (6-пиридин-3-ил-фуро[2,3-d]пиримидин-4-ил)амид	GSK	WO-0302447 (ссылка на соединение 4-[2-(бромфенил)-4-(4-фторфенил)-1Н-имидазол-5-ил]пиридин	Инсулин-независимый диабет, нейродегенеративные заболевания
NNC-57-0511, NNC-57-0545, NNC-57-0588	1-(4-амино-фуразан-3-ил)-5-пиперидин-1-илметил-1Н-[1,2,3]триазол-4-карбоновой кислоты [1-пиридин-4-ил-мет-(Е)-илиден]-гидразид	Novo Nordisk		Инсулин-независимый диабет
CP-70949	NSD	Pfizer		Гипогликемический агент
VX-608	NSD			Цереброваскулярная ишемия, Инсулин-независимый диабет
Класс KP-403	NSD	Kinetek		Модулятор ядерного фактора каппа В, противовоспалительный, ингибитор клеточного цикла, ингибитор киназы-3 гликогенсинтазы
БЫЕТТА (экзенамид)	Эксенамид: C <sub>184</sub> H <sub>282</sub> N <sub>50</sub> O <sub>60</sub> S Аминокислотная последовательность: H-His-Gly-Glu-Gly-Thr-Phe-Thr-Ser-Asp-Leu-Ser-Lys-Gln-Met-Glu-Glu-Glu-Ala-Val-Arg-Leu-Phe-Ile-Glu-Trp-Leu-Lys-Asn-Gly-Gly-Pro-Ser-Ser-Gly-Ala-Pro-Pro-Pro-Ser-NH <sub>2</sub>	Amylin/Eli Lilly&Co		Инсулин-независимый диабет
Vildagliptin (LAF237)	NSD	Novartis		Инсулин-независимый диабет, ингибитор DPP-4

NSD = структура не раскрыта (в Iddb3).

Терапевтический(е) агент(ы) TCF7L2 вводится в терапевтически эффективном количестве (т.е. в количестве, которое "достаточно" для лечения, как описано выше). Количество, которое будет терапевтически эффективным при лечении определенного индивидуального расстройства или состояния, будет зависеть от симптомов и тяжести заболевания и может быть определено стандартными клиническими методами. Кроме того, необязательно могут применяться анализы *in vitro* или *in vivo*, чтобы помочь идентифицировать оптимальные диапазоны дозировки. Точная доза, применяемая в формулировке, будет также зависеть от пути введения и серьезности заболевания или расстройства и должна быть выбрана в соответствии с решением практикующего врача и состоянием каждого пациента. Эффективные дозировки могут быть экстраполированы из кривых доза-ответ, полученных на тест-системах *in vitro* или на животных моделях.

В одном воплощении могут применяться нуклеиновая кислота, например нуклеиновая кислота, кодирующая TCF7L2 полипептид, или другая нуклеиновая кислота, которая кодирует полипептид TCF7L2 или сплайсированный вариант, производное или фрагмент этого, одна или в фармацевтической композиции, как описано выше. Например, ген TCF7L2, или нуклеиновая кислота, или кДНК, кодирующая TCF7L2 полипептид, сама по себе или включенная в вектор, могут быть введены в клетки (*in vitro* или *in vivo*), так что клетки производят нативный полипептид TCF7L2. Если необходимо, клетки, которые были трансформированы геном или кДНК или вектором, включающим ген, нуклеиновую кислоту или кДНК, могут быть введены (или введены повторно) индивидууму, подверженному заболеванию. Таким образом, клетки, у которых в природе отсутствует нативная экспрессия и активность TCF7L2 или которые имеют измененную экспрессию или активность TCF7L2 либо имеют экспрессию сплайсированного варианта TCF7L2, ассоциированного с заболеванием, могут быть сконструированы для экспрессии полипептида TCF7L2 или активного фрагмента полипептида TCF7L2 (или отличающегося варианта TCF7L2). В определенных воплощениях нуклеиновые кислоты, кодирующие полипептид TCF7L2, или активный фрагмент, или производное этого могут быть введены в экспрессионный вектор, такой как вирусный вектор, а вектор может быть введен в подходящие клетки животного. Могут быть применены другие переносящие ген системы, включая вирусные и невирусные системы транспорта. Альтернативно, могут применяться также невирусные способы транспорта генов, такие как совместная преципитация с фосфатом кальция, механические методы (например, микроинъекции); опосредованный слиянием с мембранами транспорт с помощью липосом; прямое поглощение ДНК.

Альтернативно, в другом воплощении изобретения нуклеиновая кислота изобретения; нуклеиновая кислота, комплементарная нуклеиновой кислоте изобретения, или часть такой нуклеиновой кислоты (например, олигонуклеотид, как описано ниже) могут быть применены в "антисенс"-терапии, в которой нуклеиновая кислота (например, олигонуклеотид), которая специфически гибридизуется с мРНК и/или геномной ДНК гена диабета II типа вводится или производится *in situ*. Антисенс-нуклеиновая кислота, которая специфически гибридизуется с мРНК и/или геномной ДНК, ингибирует экспрессию полипептида TCF7L2, например, путем ингибирования трансляции или транскрипции. Связывание антисенс-нуклеиновой кислоты может осуществляться за счет комплементарности традиционных пар оснований или, к примеру, в случае связывания дуплекса ДНК может осуществляться через специфическое взаимодействие в большой борозде двойной спирали.

Антисенс-конструкт настоящего изобретения может доставляться, например, как экспрессионная плаزمид, как описано выше. Когда плазмид транскрибируется в клетке, она производит РНК, которая комплементарна части мРНК и/или ДНК, которая кодирует полипептид TCF7L2. Альтернативно, антисенс-конструкт может быть олигонуклеотидным зондом, который производится *ex vivo* и вводится в клетки; затем он ингибирует экспрессию путем гибридизации с мРНК и/или геномной ДНК полипептида. В одном воплощении олигонуклеотидный зонд модифицируется олигонуклеотидами, которые резистентны к эндогенным нуклеазам, например экзонуклеазам и/или эндонуклеазам, таким образом сохраняя их стабильными *in vivo*. Типичными молекулами нуклеиновой кислоты для применения в качестве антисенс-олигонуклеотидов являются фосфорамидатные, фосфотиоатные и метилфосфонатные аналоги ДНК (см. также патенты US 5176996, 5264564 и 5256775). Дополнительно общепринятым подходом к конструированию олигомеров, полезных при антисенс-терапии, являются также описанные, например, у Van der Krol et al. (Biotechniques. 6:958-976 (1988)) и у Stein et al. (Cancer Res. 48:2659-2668 (1988)). По отношению к антисенс-ДНК предпочтительными являются олигодезоксирибонуклеотиды, происходящие из сайта инициации трансляции.

Для осуществления антисенс-терапии конструируются олигонуклеотиды (мРНК, кДНК или ДНК), которые являются комплементарными к мРНК, кодируемой геном TCF7L2. Антисенс-олигонуклеотиды связываются с транскриптами мРНК TCF7L2 и предотвращают трансляцию. Абсолютная комплементарность, хотя и является предпочтительной, не требуется. Последовательность, "комплементарная" части РНК, как это называется здесь, показывает, что последовательность имеет достаточную комплементарность, чтобы быть способной гибридизоваться с РНК, образуя стабильный дуплекс; в случае двутяжной антисенс-нуклеиновой кислоты единственная нить дуплекса ДНК может быть теснирована таким образом, или может быть проанализировано образование триплекса. Способность гибридизоваться будет зависеть как от степени комплементарности, так и от длины молекулы антисенс-нуклеиновой кислоты, как

подробно описано выше. В основном, чем длиннее гибридизуемая нуклеиновая кислота, тем больше оснований, являющихся некомплементарными с РНК, она может включать и все еще образовывать стабильный дуплекс (или триплекс, в случае, если он существует). Специалист может выявить степень толерантности неспаривания путем применения стандартных процедур.

Олигонуклеотиды, применяемые в антисенс-терапии, могут быть ДНК, РНК или химерными смесями или производными либо модифицированными версиями этого, одноцепочечными или двуцепочечными. Олигонуклеотиды могут быть модифицированы по основаниям, сахарам или фосфатному скелету, например, для увеличения стабильности молекулы, гибридизации и т.д. Олигонуклеотиды могут включать другие прикрепленные группы, такие как пептиды (например, для направления их к рецепторам клетки-хозяина *in vivo*) или агенты, облегчающие транспорт через клеточную мембрану (см., например, Letsinger et al., Proc. Natl. Acad. Sci. USA. 86:6553-6556 (1989); Lemaître et al., Proc. Natl. Acad. Sci. USA. 84:648-652 (1987); Международная публикация PCT: WO 88/09810) или через гематоэнцефалический барьер (см., например, Международную публикацию PCT WO 89/10134), или запускаемые гибридизацией расщепляющие агенты (см., например, Krol et al., Biotechniques. 6:958-976 (1988)), или встраиваемые в ДНК агенты. (См., например, Zon, Pharm. Res. 5:539-549 (1988)). Наконец, олигонуклеотиды могут быть конъюгированы с другими молекулами (например, пептидом, запускаемым гибридизацией кросс-сшивающим агентом, транспортирующим агентом, запускаемым гибридизацией расщепляющим агентом).

Антисенс-молекулы доставляются к клеткам, которые экспрессируют TCF7L2 *in vivo*. Для доставки антисенс-ДНК или РНК в клетки может быть использован ряд способов, например антисенс-молекулы могут быть инъецированы непосредственно в участок ткани, или модифицированные антисенс-молекулы, сконструированные так, чтобы направляться в желаемые клетки (например, антисенс-молекулы связывают с пептидами или антителами, которые специфически связываются с рецепторами или антигенами, экспрессируемыми на поверхности клетки-мишени), могут быть введены системно. Альтернативно, в предпочтительном воплощении применяется рекомбинантный ДНК-конструкт, в котором антисенс-олигонуклеотид помещен под контролем сильного промотора (например, *pol* III или *pol* II). Применение такого конструкта для транспорта к клеткам-мишеням у пациента приводит к транскрипции достаточных количеств одноцепочечной РНК, которая будет образовывать комплементарные пары оснований с эндогенным транскриптом TCF7L2 и таким образом предотвращать трансляцию мРНК TCF7L2. Например, *in vivo* может быть введен такой вектор, который аккумулируется клеткой и направляет транскрипцию антисенс-РНК. Такой вектор может оставаться эпизомальным или интегрироваться в хромосомы до тех пор, пока он может быть транскрибирован для получения желаемой антисенс-РНК. Такие векторы могут быть сконструированы с применением технологий рекомбинантной ДНК, стандартных для отрасли и описанных выше. Например, плазида, космида, YAC или вирусный вектор могут применяться для приготовления рекомбинантного конструкта ДНК, который может быть введен непосредственно в участок ткани. Альтернативно, могут применяться вирусные векторы, которые избирательно инфицируют желаемую ткань, в этом случае введение может быть выполнено другим путем (например, системно).

Эндогенная экспрессия полипептида TCF7L2 может быть также снижена путем инактивации или "нокаута" гена, нуклеиновой кислоты или его промотора с применением направленных гомологичных рекомбинаций (например, см. Smithies et al., Nature. 317:230-234 (1985); Thomas & Capecchi, Cell. 51:503-512 (1987); Thompson et al., Cell. 5:313-321 (1989)). Например, могут быть применены измененный нефункциональный ген или нуклеиновая кислота (или полностью неродственная последовательность ДНК), фланкированная ДНК, гомологичная эндогенному гену или нуклеиновой кислоте (кодирующими или регуляторными областями нуклеиновой кислоты) с или без маркера для селекции и/или негативного маркера для селекции, для трансфицирования клеток, которые экспрессируют ген или нуклеиновую кислоту *in vivo*. Встраивание ДНК-конструкта путем направленной гомологичной рекомбинации приводит к инактивации гена или нуклеиновой кислоты. Рекомбинантные ДНК-конструкты могут быть введены непосредственно или направлены к нужному участку *in vivo* с применением подходящих векторов, как описано выше. Альтернативно, экспрессия неизмененных генов или нуклеиновых кислот может быть увеличена с применением подобного способа: для встраивания ДНК-конструкта, включающего неизмененный функциональный ген или нуклеиновую кислоту в месте измененного TCF7L2 в клетке, может быть применена направленная гомологичная рекомбинация, как это описано выше. В другом воплощении направленная гомологичная рекомбинация может применяться для встраивания ДНК-конструкта, включающего нуклеиновую кислоту, которая кодирует вариантный полипептид диабета II типа, который отличается от представленного в клетке.

Альтернативно, экспрессия эндогенной нуклеиновой кислоты TCF7L2 может быть снижена путем направления дезоксирибонуклеотидных последовательностей, комплементарных к регуляторной области нуклеиновой кислоты TCF7L2 (т.е. к промотору или энхансерам TCF7L2), для формирования тройных спиральных структур, которые предотвращают транскрипцию нуклеиновой кислоты TCF7L2 в клетках-мишенях организма (см. в основном Helene, C., Anticancer Drug Des., 6(6):569-84 (1991); Helene, C., et al., Ann. N.Y. Acad. Sci. 660:27-36 (1992) и Maher, L.J. Bioassay. 14(12):807-15 (1992)). Подобным образом

антисенс-конструкты, описанные здесь, за счет антагонизма с нормальной биологической активностью одного из белков TCF7L2 могут применяться в манипуляциях с тканью, например в дифференциации ткани, как *in vivo*, так и *ex vivo* в культурах ткани.

Дополнительно антисенс-методы (например, микроинъекция антисенс-молекул или трансфекция плазидами, транскрипты которых являются антисенс-молекулами по отношению к мРНК гена диабета II типа или к последовательности гена) могут применяться для исследования роли TCF7L2 или взаимодействия TCF7L2 и связывающих его агентов в процессе развития, а также в нормальном функционировании TCF7L2 в клетке, или взаимодействия TCF7L2 и связывающих его агентов с тканью взрослых. Такие методы могут применяться на культуре клеток, но могут также применяться в создании трансгенных животных.

Еще в одном воплощении изобретения в обработке гена диабета II типа могут также применяться другие TCF7L2 терапевтические агенты, как те, что описаны здесь.

Терапевтические агенты могут доставляться в композиции, как описано выше, или сами по себе. Они могут вводиться системно или могут быть направлены в определенную ткань. Терапевтические агенты могут быть получены различными способами, включая химический синтез; рекомбинантную продукцию; продукцию *in vivo* (например, в трансгенных животных, таких как в патенте US 4873316 Meade et al.), и могут быть изолированы с применением стандартных способов, таких как описанные выше.

Могут также применяться комбинации любых из описанных выше способов (например, введение неизмененных полипептидов в сочетании с антисенс-терапией, направляющей измененную мРНК TCF7L2; введение первого сплайсированного варианта, кодируемого нуклеиновой кислотой TCF7L2 в сочетании с антисенс-терапией, направленной доставкой второго сплайсированного варианта, кодируемого нуклеиновой кислотой TCF7L2).

Способы оценки вероятности ответа на терапевтические агенты TCF7L2.

Настоящее изобретение дополнительно относится к способам оценки вероятности ответа индивидуума на терапевтический агент TCF7L2. В способах оцениваются маркеры или гаплотипы, относящиеся к TCF7L2 гену, как это описано выше в отношении к оценке индивидуума на восприимчивость к диабету II типа. Присутствие аллеля, маркера, SNP или гаплотипа, ассоциированного с восприимчивостью (повышенным риском) к диабету II типа (например, аллеля, иного, чем 0 аллель в маркере DG10S478, T аллель в SNP rs12255372, A аллель в SNP rs7895340, C аллель в SNP rs11196205, C аллель в SNP rs7901695, T аллель в SNP rs7903146, C аллель в SNP rs12243326 и T аллель в SNP rs4506565; маркера, ассоциированного с LD блоком экзона 4 TCF7L2, такого как гаплотип риска, ассоциированный с LD блоком экзона 4 TCF7L2), является индикатором вероятности положительного ответа на терапевтический агент TCF7L2. "Вероятность положительного ответа" показывает, что индивидуум наиболее вероятно имеет положительный ответ на терапевтический агент TCF7L2, чем индивидуум, не имеющий такого аллеля, маркера, SNP или гаплотипа, ассоциированных с восприимчивостью (повышенным риском) к диабету II типа, как описано здесь. "Положительный ответ" на терапевтический агент TCF7L2 является физиологическим ответом, который служит признаком лечения диабета II типа. Как описано выше, "лечение" относится не только к облегчению симптомов, ассоциированных с диабетом II типа, но также к предотвращению или задержке начала диабета II типа; уменьшение тяжести или частоты симптомов диабета II типа и/или также снижение потребности в сопутствующей терапии другими лекарствами, которые облегчают симптомы, ассоциированные с диабетом II типа.

Фармацевтические композиции.

Настоящее изобретение относится также к фармацевтическим композициям, включающим агенты, которые изменяют активность TCF7L2, или которые другим способом влияют на Wnt сигнальный каскад или на кадгеринный сигнальный путь, или которые могут применяться как TCF7L2 терапевтические агенты. Фармацевтические композиции могут быть составлены с физиологически приемлемым носителем или эксципиентом для приготовления фармацевтической композиции. Носитель и композиция могут быть стерильными. Композиции должны соответствовать способу введения.

Подходящие фармацевтически приемлемые носители включают, но не ограничиваются этим, воду, солевые растворы (например, NaCl), буферно-солевые растворы, спирты, глицерин, этанол, гуммиарабик, растительные масла, бензиловые спирты, полиэтиленгликоли, углеводы, такие как лактоза, амилоза или крахмал, декстроза, стеарат магния, тальк, кремниевая кислота, вязкий парафин, парфюмерное масло, сложные эфиры жирных кислот, гидроксиметилцеллюлоза, поливинилпирролидон и т.д., а также комбинации этого. Если желательно, то фармацевтические препараты могут быть смешаны со вспомогательными агентами, например смазочными агентами, консервантами, стабилизаторами, увлажняющими агентами, эмульгаторами, солями для воздействия на осмотическое давление, буферами, красителями, вкусовыми и/или ароматическими веществами и подобным, которые не реагируют вредным образом с активными агентами.

Если желательно, композиции могут также включать минорные количества увлажняющих или эмульсифицирующих агентов или агентов, забуферивающих pH. Композиция может быть жидким раствором, суспензией, эмульсией, таблеткой, пилюлей, капсулой, формулировкой с пролонгированным освобождением или порошком. Композиция может быть составлена как суппозиторий с традиционными связывающими агентами и носителями, такими как триглицериды. Пероральные композиции могут включать стандартные носители, такие как фармацевтической чистоты маннитол, лактоза, крахмал, стеарат магния, поливинилпирролидоны, сахарин натрия, целлюлоза, карбонат магния и т.д.

Способы введения этих композиций включают, но не ограничиваются этим, внутрикожный, внутримышечный, внутрибрюшинный, внутриглазной, внутривенный, подкожный, локальный, пероральный или интраназальный. Другие подходящие способы введения могут включать также генную терапию (как описано ниже), перезаряжаемые или биodeградируемые устройства, устройства, обеспечивающие ускорение частиц ("генный пистолет"), и медленно освобождающие вещество полимерные устройства.

Фармацевтическая композиция этого изобретения может быть также введена как часть комбинационной терапии с другими агентами.

Композиции могут быть составлены в соответствии с обычными процедурами, как фармацевтические композиции, адаптированные для введения людям. Например, композиции для внутривенного введения типично являются растворами в стерильных изотонических буферах на водной основе. Где это необходимо, композиция может также включать солибутилизирующий агент и локальный анестетик для снижения боли в месте инъекции. В основном ингредиенты поставляются либо отдельно, либо в смешанном виде в единичной форме дозировки, например, как сухой лиофилизированный порошок или свободный водный концентрат в герметически запечатанном контейнере, таком как ампула или саше с указанием количества активного агента. Там, где композиция вводится путем инфузии, она может быть суспендирована в инфузионной бутылочке, включающей стерильную воду фармацевтической чистоты, солевой раствор или смесь декстроза/вода. Там, где композиция вводится путем инъекции, может быть обеспечена ампула со стерильной водой для инъекций или солевым раствором, так что ингредиенты могут быть смешаны до введения.

Для локального введения могут применяться нераспыляемые формы, от вязких до полутвердых или твердых форм, включающих носитель, совместимый с локальным применением и имеющим динамическую вязкость, предпочтительно большую, чем вода. Подходящие композиции включают, но не ограничиваются этим, растворы, эмульсии, кремы, мази, порошки, клизмы, примочки, золи, линименты, бальзамы, аэрозоли и т.д., которые по желанию стерилизуются или смешиваются со вспомогательными агентами, например консервантами, стабилизаторами, увлажняющими агентами, буферами или солями для воздействия на осмотическое давление и т.д. Агент может быть включен в косметическую композицию. Для локального введения подходящими являются также распыляемые аэрозольные препараты, где активный ингредиент, предпочтительно в комбинации с твердым или жидким инертным носителем упаковывается в запечатанную бутылочку или смешивается с прессуемым летучим веществом, обычно газообразной сжатой в баллончике жидкостью, например со сжатым воздухом.

Агенты, описанные здесь, могут быть сформулированы как нейтральные солевые формы. Фармацевтически приемлемые соли включают те, которые образуются со свободными аминогруппами, такие как происходящие из соляной, фосфорной, уксусной, щавелевой, винной кислот и т.д., и тех, что образуются со свободными карбоксильными группами, такие как происходящие из гидроксидов натрия, калия, аммония, кальция, железа, изопропиламина, триэтиламина, 2-этиламиноэтанола, гистидина, прокаина и т.д.

Агенты вводятся в терапевтически эффективном количестве. Количество агентов, которое будет терапевтически эффективным, частично зависит от природы расстройства и/или степени симптомов и может быть определено стандартными химическими способами. Кроме того, необязательно могут быть сделаны анализы *in vitro* и *in vivo* с тем, чтобы помочь идентифицировать оптимальные интервалы дозирования. Точная доза для применения в композиции будет также зависеть от способа введения и серьезности симптомов и должна быть определена в соответствии с решением практикующего врача и обстоятельствами болезни каждого пациента. Эффективные дозы могут быть экстраполированы из кривых доза-ответ, полученных на тест-системах *in vitro* или на животных моделях.

Изобретение обеспечивает также фармацевтическую упаковку или набор, включающие один (или более) контейнер, заполненный одним (или более) ингредиентом фармацевтических композиций изобретения. Необязательно в наборе с таким контейнером(и) может быть уведомление в форме предписанных государственным агентством правил, регулирующих производство, применение или продажу фармацевтических или биологических продуктов, это уведомление отражает разрешение, данное агентством на применение или продажу для введения людям. Упаковка или набор могут быть снабжены информацией относительно способа введения, последовательности введения лекарств (например, отдельно, последовательно или одновременно) или подобной информацией. Упаковка или набор могут также включать средство напоминания пациенту о принятии лекарства. Упаковка или набор могут быть стандартной дозированной комбинационной терапии или они могут содержать множество стандартных дозирования. В частности, агенты могут быть разделены, смешаны вместе в любой комбинации, представлены в одном пузырьке

или таблетке. Агенты, смонтированные в блистерной упаковке, или другие средства дозирования являются предпочтительными. Для целей этого изобретения стандартная дозировка означает дозировку, которая зависит от индивидуальной фармакодинамики каждого агента и вводится в одобренных FDA дозировках в виде стандартных временных режимов.

Скрининг-анализы и идентифицированные таким образом агенты.

Изобретение обеспечивает также способы идентификации агентов (например, слитых белков, полипептидов, имитаторов пептидов, пролекарств, рецепторов, связывающих агентов, антител, малых молекул или других лекарственных средств, или рибозимов), которые изменяют (например, увеличивают или уменьшают) активность TCF7L2, которые другим образом взаимодействуют с TCF7L2 или с другим членом Wnt сигнального каскада или кадгеринового сигнального каскада (например, бета-катенин). К примеру, в определенных воплощениях такие агенты могут быть агентами, которые связываются с TCF7L2, которые производят стимулирующее или ингибирующее воздействие, например на активность TCF7L2; или которые изменяют (например, усиливают или ингибируют) способность TCF7L2 взаимодействовать с другими членами Wnt сигнального пути или с членами кадгеринового сигнального пути, или которые изменяют посттрансляционный процессинг TCF7L2. В других воплощениях такие агенты могут быть агентами, которые изменяют активность или функцию Wnt сигнального каскада или кадгеринового сигнального пути.

В одном воплощении изобретение обеспечивает анализы для скрининга кандидатов или тестовых агентов, которые связываются с или модулируют активность белка TCF7L2 (или биологически активных частей этой молекулы), а также агенты, идентифицируемые этим анализом. Тестируемые агенты могут быть получены с применением любого из многочисленных подходов методами комбинаторной библиотеки, известными специалисту, включая биологические библиотеки; библиотеки, пространственно расположенные параллельно твердой фазе или жидкой фазе; способы создания синтетических библиотек, требующие деконволюции; метод библиотек "одна бусина-одно соединение" и методы синтетической библиотеки с применением селекции путем аффинной хроматографии. Подход с применением биологической библиотеки ограничен полипептидными библиотеками, в то время как другие четыре подхода применимы к полипептидам, непептидным олигомерам или библиотекам малых молекул (Lam, K.S., *Anticancer Drug Des.* 12:145 (1997)).

В одном воплощении для идентификации агентов, которые изменяют активность TCF7L2, клетка, клеточный лизат или раствор, включающий или экспрессирующий TCF7L2, или фрагмент, или производное этого, может контактировать с тестируемым агентом; альтернативно, белок может контактировать непосредственно с тестируемым агентом. Оценивается уровень (количество) активности TCF7L2 (например, измеряется уровень (количество) активности TCF7L2 либо непосредственно, либо непрямым способом) и сравнивается с уровнем активности в контроле (т.е. с уровнем активности белка TCF7L2 или активного фрагмента, или производного этого в отсутствие тестируемого агента). Если уровень активности в присутствии агента отличается на количество, которое является статистически значимым, от уровня активности в отсутствие агента, то агент является агентом, который изменяет активность TCF7L2. Увеличение в уровне активности относительно контроля указывает, что агент является агентом, который усиливает (является агонистом) активность. Подобным образом уменьшение уровня активности относительно контроля указывает, что агент является агентом, который ингибирует (является антагонистом) активности. В другом воплощении уровень активности TCF7L2 или производного, или фрагмента этого в присутствии тестируемого агента сравнивается с контрольным уровнем, который был установлен предварительно. Уровень активности в присутствии агента, который отличается от контрольного уровня по количеству, которое является статистически значимым, указывает, что агент изменяет активность TCF7L2.

Настоящее изобретение относится также к любому способу анализа по идентификации агентов, которые изменяют экспрессию гена TCF7L2 (например, антисенс-нуклеиновые кислоты, слитые белки, полипептиды, имитаторы пептидов, рецепторы, связывающие агенты, антитела, малые молекулы или другие лекарственные средства, или рибозимы), которые изменяют (например, увеличивают или уменьшают) экспрессию (например, транскрипцию или трансляцию) гена или которые другим образом взаимодействуют с TCF7L2, а также агентов, идентифицируемых анализом. Например, раствор, включающий нуклеиновую кислоту, кодирующую TCF7L2, может контактировать с тестируемым агентом. Раствор может включать, к примеру, нуклеиновую кислоту или клеточный лизат, включающий нуклеиновую кислоту; альтернативно, раствор может быть иным раствором, который включает элементы, необходимые для транскрипции/трансляции нуклеиновой кислоты. По желанию могут быть также применены клетки, которые не суспендируются в растворе. Оценивается уровень и/или характер экспрессии TCF7L2 (например, уровень и/или характер экспрессированных мРНК или белка, такой как уровень и/или характер различных сплайсированных вариантов) и сравнивается с уровнем и/или характером экспрессии в контроле (т.е. уровнем и/или характером экспрессии TCF7L2 в отсутствие тестируемого агента). Если уровень и/или характер в присутствии агента отличается по количеству или по способу, которые являются статистически значимыми, от уровня и/или характера в отсутствие агента, то агент является агентом, который изменяет экспрессию гена диабета II типа. Усиление экспрессии показывает, что агент является

агонистом активности TCF7L2. Подобным образом ингибирование экспрессии показывает, что агент является антагонистом активности TCF7L2. В другом воплощении уровень и/или характер полипептида(ов) TCF7L2 (т.е. различных сплайсированных вариантов) в присутствии тестируемого агента сравнивается с контрольным уровнем и/или характером, который был установлен предварительно. Уровень и/или характер в присутствии агента, который отличается от контрольного уровня и/или характера по количеству или по способу, которые являются статистически значимыми, указывает, что агент изменяет экспрессию TCF7L2.

В другом воплощении изобретения агенты, которые изменяют экспрессию TCF7L2 или которые другим образом взаимодействуют с TCF7L2 или с другими членами Wnt сигнального пути или кадгеринового сигнального пути, могут быть идентифицированы с применением клетки, клеточного лизата или раствора, включающего нуклеиновую кислоту, кодирующую область промотора гена TCF7L2 или нуклеиновой кислоты, операбельно связанной с репортерным геном. После контакта с тестируемым агентом определяется уровень экспрессии репортерного гена (например, уровень экспрессируемых мРНК или белка) и сравнивается с уровнем экспрессии в контроле (т.е. уровнем экспрессии репортерного гена в отсутствие тестируемого агента). Если уровень в присутствии агента отличается по количеству или по способу, которые являются статистически значимыми, от уровня в отсутствие агента, то агент является агентом, который изменяет экспрессию TCF7L2, как показано по его способности изменять экспрессию гена, который операбельно связан с промотором гена TCF7L2. Усиление экспрессии репортера показывает, что агент является агонистом TCF7L2 активности. Подобным образом ингибирование экспрессии репортера показывает, что агент является антагонистом активности TCF7L2. В другом воплощении уровень экспрессии репортера в присутствии тестируемого агента сравнивается с контрольным уровнем, который установлен предварительно. Уровень в присутствии агента, который отличается от контрольного уровня по количеству или по способу, которые являются статистически значимыми, показывает, что агент изменяет экспрессию.

Агенты, которые изменяют количество различных сплайсированных вариантов, кодируемых TCF7L2 (например, агент, который усиливает активность первого сплайсированного варианта и который ингибирует активность второго сплайсированного варианта), а также агенты, которые являются агонистами активности первого сплайсированного варианта и антагонистами второго сплайсированного варианта, могут быть легко идентифицированы с применением способов, описанных выше.

В другом воплощении изобретения для оценки вклада тестируемого агента в активность полипептида в отношении связывающего TCF7L2 агента могут быть применены анализы. Например, клетка, которая экспрессирует соединение, взаимодействующее с полипептидом TCF7L2 (называемое здесь "агент, связывающий TCF7L2", который может быть полипептидом или другой молекулой, взаимодействующей прямо или непрямо с полипептидом TCF7L2, такой как член сигнального пути Wnt или член кадгеринового сигнального каскада), контактирует с TCF7L2 в присутствии тестируемого агента, и определяется способность тестируемого агента изменять взаимодействие между TCF7L2 и агентом, связывающим TCF7L2. Альтернативно, может быть использован клеточный лизат или раствор, включающий агент, связывающий TCF7L2. Агент, который связывается с TCF7L2 или TCF7L2 связывающий агент, изменяет взаимодействие, мешая ему, или усиливает способность TCF7L2 связываться или ассоциировать либо другим образом взаимодействовать с агентом, связывающим TCF7L2. Определение способности тестируемого агента связываться с TCF7L2 или агент, связывающий TCF7L2, может быть выполнено, к примеру, путем сопряжения тестируемого агента с радиоизотопом или ферментативной такой меткой, что связывание тестируемого агента с полипептидом может быть прямо или непрямо определено путем детекции меченого  $^{125}\text{I}$ ,  $^{35}\text{S}$ ,  $^{14}\text{C}$  или  $^3\text{H}$ , а радиоизотоп определяется путем прямого счета радиоэмиссии или путем сцинтилляционного счета. Альтернативно, тестируемые агенты могут быть помечены ферментом, например пероксидазой из хрена, щелочной фосфатазой или люциферазой, и ферментативная метка детектируется путем определения превращения подходящего субстрата в продукт. В рамках этого изобретения находится определение способности тестируемого агента взаимодействовать с полипептидом без мечения любых взаимодействующих компонентов. К примеру, для детектирования взаимодействия тестируемого агента с TCF7L2 или TCF7L2 связывающего агента без мечения тестируемого агента, TCF7L2 или TCF7L2 связывающего агента может применяться микрофизиометр. McConnell, H.M. et al., Science. 257:1906-1912 (1992). Как это применяется здесь, "микрофизиометр" (например, Cytosensor<sup>TM</sup>) представляет собой аналитический инструмент, который измеряет скорость, с которой клетка закисляет свое окружение с применением чувствительного к свету потенциометрического сенсора (LAPS). Изменение в скорости закисления может применяться в качестве индикатора взаимодействия между лигандом и полипептидом.

Таким образом, эти рецепторы могут быть полезны для скрининга соединений, которые являются агонистами или антагонистами, для их применения в лечении или исследовании чувствительности к диабету II типа. Могут быть созданы лекарства для регуляции активации TCF7L2, которые, в свою очередь, могут применяться для регуляции сигнальных путей и транскрипционных, происходящих ниже гена.

В другом воплощении изобретения может быть применен анализ для идентификации полипептидов, которые взаимодействуют с TCF7L2. Например, для идентификации полипептидов, взаимодействующих с TCF7L2, может быть применена дрожжевая двухгибридная система (Fields, S. and Song, O., Nature. 340:245-246 (1989)). В такой дрожжевой двухгибридной системе конструируются векторы, основанные на гибкости транскрипционного фактора, который имеет два функциональных домена (ДНК-связывающий домен и домен активации транскрипции). Если два домена разделены, но слиты в два различных белка, которые взаимодействуют друг с другом, может быть достигнута активация транскрипции и транскрипция специфического маркера (например, для идентификации присутствия взаимодействия и активации транскрипции могут применяться маркеры питательной среды, такие как His и Ade, или окрашенные маркеры, такие как lacZ). Например, в способе изобретения применяется первый вектор, который включает нуклеиновую кислоту, кодирующую ДНК-связывающий домен, а также TCF7L2, сплайсированный вариант, или фрагмент, или производное этого, и применяется второй вектор, который включает нуклеиновую кислоту, кодирующую домен активации транскрипции, а также нуклеиновую кислоту, кодирующую полипептид, который потенциально может взаимодействовать с TCF7L2 или сплайсированным вариантом, фрагментом или производным этого. Инкубация дрожжей, включающих первый вектор со вторым вектором в подходящих условиях (например, в условиях спаривания, таких как применяются в системе Matchmaker™ от Clontech (Palo Alto, California, USA), позволяет идентифицировать колонии, которые экспрессируют интересующий маркер. Эти колонии могут быть исследованы для идентификации полипептида(ов), которые взаимодействуют с TCF7L2 или его фрагментом или производным. Такие полипептиды могут применяться как агенты, которые изменяют активность экспрессии TCF7L2, как описано в отношении способов лечения.

В более чем одном воплощении приведенного выше способа анализа настоящего изобретения может быть желательным иммобилизовать на твердой подложке либо TCF7L2 ген, либо белок TCF7L2, либо агент, который связывает TCF7L2 (например, другой член Wnt сигнального каскада или член кадгеринового сигнального пути) или другие компоненты анализа, для того чтобы облегчить разделение формы в виде комплекса от невзаимодействующих белков, а также для приспособления к автоматизации анализа. Связывание тестируемого агента с белком или взаимодействие белка со связывающим агентом в присутствии или в отсутствие тестируемого агента может быть выполнено в любом сосуде, подходящем для включения реагентов. Примеры таких сосудов включают пластины для микротитрования, пробирки или микроцентрифужные пробирки. В одном воплощении может быть обеспечен слитый белок (например, белок, слитый с глутатион-S-трансферазой), к которому добавлен домен, который позволяет TCF7L2, белку TCF7L2 или агенту, связывающему TCF7L2, быть связанным с матриксом или другой твердой подложкой.

В другом воплощении модуляторы экспрессии молекул нуклеиновой кислоты изобретения идентифицируются способом, в котором клетка, клеточный лизат или раствор, включающий TCF7L2, контактирует с тестируемым агентом, и определяется экспрессия подходящей мРНК или полипептида (например, сплайсированного варианта(ов)) в клетке, лизате или растворе. Уровень экспрессии подходящей мРНК или полипептида(ов) сравнивается с уровнем экспрессии мРНК или полипептида(ов) в отсутствие тестируемого агента. Тестируемый агент может быть идентифицирован как модулятор экспрессии на основе этого сравнения. Например, когда экспрессия мРНК или полипептида больше (статистически значимо больше) в присутствии тестируемого агента, чем в его отсутствие, тестируемый агент идентифицируется как стимулятор или энхансер экспрессии мРНК или полипептида. Альтернативно, когда экспрессия мРНК или полипептида меньше (статистически значимо меньше) в присутствии тестируемого агента, чем в его отсутствие, тестируемый агент идентифицируется как ингибитор экспрессии мРНК или полипептида. Уровень экспрессии мРНК или полипептида в клетках может быть определен способами, описанными здесь для детектирования мРНК или полипептида.

Дополнительно, изобретение относится к новым агентам, идентифицированным с помощью описанного выше скринирующего анализа. Соответственно в рамках настоящего изобретения далее применяется агент, идентифицированный, как описано здесь, в способах лечения, описанных здесь. Например, агент, идентифицированный, как описано здесь, может быть применен для изменения активности белка, кодируемого геном TCF7L2, или для изменения экспрессии TCF7L2 путем контактирования белка или нуклеиновой кислоты (или контактирования с клеткой, включающей полипептид или нуклеиновую кислоту) с агентом, идентифицированным, как описано здесь.

Нуклеиновые кислоты изобретения.

Нуклеиновые кислоты TCF7L2, их части и варианты.

Настоящее изобретение относится также к молекулам изолированных нуклеиновых кислот, включающих человеческий TCF7L2. Молекулы нуклеиновой кислоты TCF7L2 настоящего изобретения могут быть РНК, например мРНК, или ДНК, такой как кДНК и геномная ДНК. Молекулы ДНК могут быть дву-тяжными или однотяжными, одноцепочечная РНК или ДНК может быть кодирующей или смысловой нитью либо не кодирующей или антисмысловой нитью. Молекула нуклеиновой кислоты может включать всю или часть кодирующей последовательности гена и может дополнительно включать дополнительные не-кодирующие последовательности, такие как интроны и 3' и 5' не кодирующие последовательности (вклю-

чая, к примеру, регуляторные последовательности).

Дополнительно, молекулы нуклеиновой кислоты изобретения могут быть слиты с маркерной последовательностью, например последовательностью, которая кодирует полипептид, способствующий изоляции или очистке белка. Такие последовательности включают, но не ограничиваются, те, которые кодируют белок, слитый с глутатион-S-трансферазой (GST), и те, которые кодируют маркерный полипептид гемагглютинин А (НА) из вируса гриппа.

"Изолированная" молекула нуклеиновой кислоты, как это применяется здесь, является такой, которая отделена от нуклеиновых кислот, которые в норме фланкируют ген или нуклеотидную последовательность (как в геномных последовательностях) и/или были полностью или частично очищены от других транскрибируемых последовательностей (например, как в библиотеке РНК). К примеру, изолированная нуклеиновая кислота изобретения может, по существу, быть изолированной по отношению к комплексу клеточной среды, в которой она присутствует, или культуральной среды, когда она производится с помощью рекомбинантных методов, или химических предшественников, или других химических веществ, когда они синтезируются химическим путем. В некоторых случаях изолированный материал будет формировать часть композиции (например, грубый экстракт, включающий другие вещества) буферной системы или смешанных реагентов. В других обстоятельствах материал может быть очищен, по существу, до гомогенности, например, как определяется методом ЭФ в ПААГ колоночной хроматографии, такой как ВЭЖХ. Предпочтительно изолированная молекула нуклеиновой кислоты включает по меньшей мере 50, 80 или 90% (на молярной основе) всех представленных видов макромолекул. По отношению к геномной ДНК термин "изолированная" также может относиться к молекулам нуклеиновых кислот, которые выделены из хромосомы, с которой геномная ДНК ассоциирована в природе. Например, молекула изолированной нуклеиновой кислоты может включать менее чем примерно 5 т.п.о., но не ограничиваться 4 т.п.о., 3 т.п.о., 2 т.п.о., 1 т.п.о., 0,5 т.п.о. или 0,1 т.п.о. нуклеотидов, которые фланкируют молекулу нуклеиновой кислоты в геномной ДНК клетки, из которой молекула нуклеиновой кислоты происходит.

Молекула нуклеиновой кислоты может быть слита с другими кодирующими или регуляторными последовательностями и все еще рассматриваться как изолированная. Таким образом, рекомбинантная ДНК, содержащаяся в векторе, также включена в определение "изолированная", как это применяется здесь. Также молекулы изолированной нуклеиновой кислоты включают рекомбинантные молекулы ДНК в гетерологичных клетках-хозяевах, а также частично или по существу очищенные молекулы ДНК в растворе. Молекулы "изолированной" нуклеиновой кислоты включают также транскрипты РНК молекул ДНК настоящего изобретения *in vivo* и *in vitro*.

Молекула изолированной нуклеиновой кислоты может включать молекулу нуклеиновой кислоты или последовательность нуклеиновой кислоты, которая синтезирована химическим или рекомбинантным способом. Следовательно, рекомбинантная ДНК, содержащаяся в векторе, включена в определение "изолированная", как оно применяется здесь. Также молекулы нуклеиновой кислоты включают рекомбинантные молекулы ДНК в гетерологичных организмах, а также частично или по существу очищенные молекулы ДНК в растворе. РНК транскрипты молекул ДНК настоящего изобретения *in vivo* и *in vitro* также включены в "изолированные" последовательности нуклеиновой кислоты. Такие изолированные молекулы нуклеиновой кислоты полезны при получении кодируемого полипептида как зондов для выделения гомологичных последовательностей (например, из других видов млекопитающих) для картирования гена (например, путем гибридизации *in situ* с хромосомами) или для детектирования экспрессии гена в ткани (например, ткани человека), такого как анализ с помощью нозерн-блота или саузерн-блота.

Настоящее изобретение относится также к молекулам нуклеиновой кислоты, которые не обнаружены в природе, но которые кодируют полипептид TCF7L2 или другие сплайсированные варианты полипептида TCF7L2 или полиморфных вариантов этого. Таким образом, к примеру, изобретение относится к молекулам ДНК, включающим последовательность, которая отличается от существующей в природе нуклеотидной последовательности, но которая благодаря вырожденности генетического кода кодирует полипептид TCF7L2 настоящего изобретения. Изобретение также включает молекулы нуклеиновой кислоты, кодирующие части (фрагменты) или варианты полипептидов, такие как аналоги или производные полипептида TCF7L2. Такие варианты могут существовать в природе, например, в случае аллельных вариаций или SNP или неприродных, таких как те, что индуцированы воздействием различных мутагенов и мутагенных процессов. Обозначенные варианты включают, но не ограничиваются этим, вставки, делеции или замены одного или более нуклеотидов, которые могут приводить к консервативным или неконсервативным заменам аминокислот, включая вставки и делеции. Предпочтительно изменение нуклеотида (и/или полученной аминокислоты) является не проявляющимся или консервативным; это означает, что они не изменяют характеристик или активности полипептида TCF7L2. В одном аспекте последовательности нуклеиновой кислоты являются фрагментами, которые включают один или более полиморфных микросателлитных маркеров. В другом аспекте нуклеотидные последовательности являются фрагментами, которые включают один или более SNP в гене TCF7L2.

Другие изменения молекул нуклеиновых кислот изобретения могут включать, например, мечение, метилирование, межнуклеотидные модификации, такие как незаряженные соединения (например, метилфосфонаты, фосфотриэфиры, фосфоамидаты, карбаматы), заряженные соединения (например, фосфоротиоаты, фосфородитиоаты), прикрепленные компоненты (например, полипептиды), интеркаляторы (например, акридин, псорален), хелаторы, алкиляторы и модифицированные соединения (например, альфа аномерные нуклеиновые кислоты). Включены также синтетические молекулы, которые имитируют молекулы нуклеиновой кислоты в способности связываться с указанной последовательностью через водородные связи и другие химические интеркаляторы. Такие молекулы включают, к примеру, те, в которых в скелете молекулы пептидные связи заменены на фосфатные связи.

Изобретение относится также к молекулам нуклеиновой кислоты, которые гибридизуются в жестких условиях гибридизации, таких как селективная гибридизация, с нуклеотидной последовательностью, описанной здесь (например, молекулами нуклеиновой кислоты, которая специфически гибридизуется с последовательностью, кодирующей полипептиды, описанные здесь, и необязательно имеет активность полипептида). В одном аспекте изобретение включает варианты, описанные здесь, которые гибридизуются в жестких условиях гибридизации (например, при селективной гибридизации) с нуклеотидной последовательностью, кодирующей аминокислотную последовательность или ее полиморфный вариант. В другом аспекте вариант, который гибридизуется в жестких условиях гибридизации, имеет активность полипептида TCF7L2.

Такие молекулы нуклеиновых кислот могут быть обнаружены и/или изолированы путем специфической гибридизации (например, в жестких условиях гибридизации).

"Специфическая гибридизация", как это применяется здесь, относится к способности первой нуклеиновой кислоты гибридизоваться со второй нуклеиновой кислотой таким способом, что первая нуклеиновая кислота не гибридизуется с любой нуклеиновой кислотой, иной, чем вторая нуклеиновая кислота (например, когда первая нуклеиновая кислота имеет более высокое подобие со второй нуклеиновой кислотой, чем с любой другой нуклеиновой кислотой в образце, где выполняется гибридизация). "Жесткие условия" гибридизации - это термин, применяемый специалистами, который относится к условиям инкубации и промывания, например, такие условия, как температура и концентрация буфера, которые позволяют осуществить гибридизацию определенной нуклеиновой кислоты с второй нуклеиновой кислотой: первая нуклеиновая кислота может быть полностью (т.е. на 100%) комплементарна второй или первая и вторая могут делить некоторую степень комплементарности, которая меньше, чем полная (например, 70, 75, 80, 85, 90, 95%). Например, могут быть применены определенные жесткие условия, которые позволяют безукоризненно разделить комплементарные нуклеиновые кислоты от таковых с меньшей комплементарностью. "Условия высокой жесткости", "условия средней жесткости" и "условия низкой жесткости", так же как и способы гибридизации нуклеиновых кислот, объяснены на страницах 2.10.1-2.10.16 и страницах 6.3.1-6.3.6 в *Current Protocols in Molecular Biology* (Ausubel, F., "Current Protocols in Molecular Biology", John Wiley & Sons (1998) и Kraus, M. and Aaronson, S. *Methods Enzymol.*, 200:546-556 (1991)).

Процент гомологии или идентичности двух нуклеотидных или аминокислотных последовательностей может быть определен путем выравнивания последовательностей для оптимальных целей сравнения (например, для оптимального выравнивания в первую последовательность могут быть введены пропуски). Затем нуклеотиды и аминокислоты в соответствующих позициях сравниваются, и процент идентичности между двумя последовательностями представляет собой функцию числа идентичных позиций, общих для двух последовательностей (т.е. % идентичности = # идентичных позиций/общее количество позиций  $\times$  100). Когда позиция в одной последовательности занимает тот же нуклеотидом или аминокислотным остатком, что и в соответствующей позиции другой последовательности, то молекулы являются гомологичными в этой позиции. Как это применяется здесь, "гомология" нуклеиновой кислоты или аминокислоты представляет собой эквивалент "идентичности" нуклеиновой кислоты или аминокислоты. В определенных аспектах длина последовательности, которая выровнена для целей сравнения, составляет по меньшей мере 30%, например по меньшей мере 40%, в определенных аспектах по меньшей мере 60%, и в других аспектах по меньшей мере 70, 80, 90 или 95% от длины последовательности сравнения. Действительное сравнение двух последовательностей может быть выполнено хорошо известными способами, например с применением математических алгоритмов. Предпочтительный нелимитирующий пример такого математического алгоритма описан у Karlin et al., *Proc. Natl. Acad. Sci. USA.* 90:5873-5877 (1993). Такой алгоритм включен в программы NBLAST и XBLAST (версия 2.0), которые описаны у Altschul et al., *Nucleic Acid Res.* 25:389-3402 (1997). При применении программ BLAST и Gapped BLAST могут применяться параметры дефолта соответствующих программ (например, NBLAST). В одном аспекте параметры для сравнения последовательностей могут быть установлены с баллами = 100, длиной слова = 12 или могут варьировать (например, W=5 или W=20).

Другим предпочтительным нелимитирующим примером математического алгоритма, применяемого для сравнения последовательностей, является алгоритм Мейерса и Миллера (Myers and Miller, *CABIOS* 4(1):11-17 (1988)). Такой алгоритм включен в программу ALIGN (версия 2.0), которая является частью пакета программ для выравнивания последовательностей GCG (Accelrys, Cambridge, UK). Когда

для сравнения последовательностей применяется программа ALIGN, могут быть применены параметры PAM120 взвешенной таблицы остатков, штраф за длину делеции 12, штраф за деление 4. Дополнительные алгоритмы для анализа последовательностей известны специалисту и включают ADVANCE и ADAM, как описано у Torellis and Robotti, *Comput. Appl. Biosci.* 10:3-5 (1994), и FASTA, описанная у Pearson and Lipman, *Proc. Natl. Acad. Sci. USA.* 85:2444-8 (1988).

В другом аспекте процент идентичности между двумя аминокислотными последовательностями может быть определен с применением пакета программ GCG с использованием либо матрикса BLOSUM, либо матрикса PAM250 и весом делеции 12, 10, 8, 6 или 4 и весом длины 2, 3 или 4. Еще в одном аспекте процент идентичности между двумя последовательностями нуклеиновых кислот может быть определен с применением программы GAP в пакете программ GCG с применением веса делеции 50 и веса длины 3.

Настоящее изобретение обеспечивает также молекулы изолированных нуклеиновых кислот, которые включают фрагмент или часть, которая гибридизуется в очень жестких условиях с нуклеотидной последовательностью TCF7L2 или комплементом такой последовательности, а также обеспечивает молекулы, которые включают фрагмент или часть, которая гибридизуется в очень жестких условиях с нуклеотидной последовательностью, кодирующей аминокислотную последовательность или полиморфный вариант этого. Фрагменты нуклеиновой кислоты изобретения имеют длину по меньшей мере примерно 15, предпочтительно по меньшей мере примерно 18, 20, 23 или 25 нуклеотидов и могут быть длиной 30, 40, 50 100, 200 или более нуклеотидов. Более длинные фрагменты, например длиной 30 или более нуклеотидов, которые кодируют антигенные полипептиды, описанные здесь, являются особенно полезными, например, для генерации антител, как описано ниже.

Зонды и праймеры.

В родственном аспекте фрагменты нуклеиновой кислоты изобретения применяются как зонды и праймеры в анализах, таких как те, что описаны здесь. "Зонды" или "праймеры" - это нуклеотиды, которые гибридизуются специфическим по отношению к основаниям способом с комплементарной нитью молекул нуклеиновых кислот. Такие зонды и праймеры включают полипептид нуклеиновых кислот, как описано у Nielsen et al., *Science.* 254:1497-1500 (1991).

Зонды и праймеры включают область нуклеотидной последовательности, которая гибридизуется по меньшей мере примерно с 15, например приблизительно 20-25 и в определенных аспектах примерно с 40, 50 или 75 последовательными нуклеотидами молекулы нуклеиновой кислоты, включающей смежную нуклеотидную последовательность TCF7L2, или полиморфных вариантов этого. В других аспектах зонд или праймер включает 100 или несколько нуклеотидов, в определенных аспектах от 6 до 50 нуклеотидов, например от 12 до 30 нуклеотидов. В других аспектах зонд или праймер по меньшей мере на 70% идентичен смежной нуклеотидной последовательности или комплементарной нити смежной нуклеотидной последовательности, например по меньшей мере на 80% идентичен, в определенном аспекте по меньшей мере на 90% идентичен и в других аспектах по меньшей мере на 95% идентичен или даже способен селективно гибридизоваться с прилегающей нуклеотидной последовательностью или с нитью, комплементарной прилегающей нуклеотидной последовательности. Часто зонд или праймер дополнительно включает метку, например радиоизотоп, флуоресцентное соединение, фермент или кофактор фермента.

Молекулы нуклеиновой кислоты изобретения, такие как те, что описаны выше, могут быть идентифицированы и изолированы с применением стандартных методов молекулярной биологии, и при этом обеспечивается информация о последовательности. Например, молекулы нуклеиновой кислоты могут быть амплифицированы и изолированы путем ПЦР с применением синтетических олигонуклеотидных праймеров, созданных на основе последовательности TCF7L2 или последовательности, комплементарной ей, или созданных на основе нуклеотидных оснований последовательности, кодирующей одну или более аминокислотных последовательностей, обеспеченной при этом. См. в основном *PCR Technology: Principles and Applications for DNA Amplification* (ed. H.A. Erlich, Freeman Press, NY, NY, (1992)); *PCR Protocols: A Guide to Methods and Applications* (Eds Innis et al., Academic Press, San Diego, CA, (1990); Mattila et al., *Nucl. Acids Res.* 19:4967 (1991); Eckert et al., *PCR Methods and Applications.* 1:17 (1991)); *PCR* (eds. McPherson et al., IRL Press, Oxford и патент US 4683202). Молекулы нуклеиновой кислоты могут быть амплифицированы с применением в качестве матрицы кДНК, мРНК или геномной ДНК, клонированы в подходящий вектор и охарактеризованы путем анализа последовательности ДНК.

Другие подходящие способы амплификации включают лигазную цепную реакцию (LCR) (см. Wu and Wallace, *Genomics.* 4:560 (1989), Landegren et al., *Science.* 241:1077 (1988)), амплификацию транскрипции (Kwoh et al., *Proc. Natl. Acad. Sci. USA.* 86:1173 (1989)) и самоподдерживаемую репликацию последовательности (Guatelli et al., *Proc. Natl. Acad. Sci. USA.* 87:1874 (1990)) и амплификацию последовательности на основе нуклеиновой кислоты (NASBA). Последние два способа амплификации вовлекают изотермальную реакцию на основе изотермальной транскрипции, которая производит как одноцепочечную РНК (ssРНК), так и двухцепочечную РНК (ds-РНК) в качестве продуктов амплификации в соотношении примерно 30 или 100 к 1 соответственно.

Аmplифицированная ДНК может быть помечена, например, радиоактивными изотопами и применена как зонд для скрининга библиотеки кДНК, происходящей из клеток человека, мРНК в *zap express*, ZIPLOX или другом подходящем векторе. Соответствующие клоны могут быть изолированы, с последующим удалением *in vivo* может быть получена ДНК, и клонированная вставка может быть секвенирована в одной или обеих ориентациях способом, известным специалисту, для идентификации правильной рамки считывания полипептида подходящего молекулярного веса. Например, прямой анализ нуклеотидной последовательности молекул нуклеиновой кислоты настоящего изобретения может быть выполнен с применением хорошо известных способов, которые являются коммерчески доступными. См., например, Sambrook et al., *Molecular Cloning, A Laboratory Manual* (2<sup>nd</sup> Ed., CSHP, New York, 1989); Zyskind et al., *Recombinant DNA Laboratory Manual* (Acad Press, 1988). Дополнительно, для анализа нуклеиновых кислот и полипептидов полезны также флуоресцентные методы (Chen et al., *Genome Res*, 9, 492 (1999)). С применением этих или похожих методов могут быть изолированы, секвенированы и дополнительно охарактеризованы полипептид и ДНК, кодирующая полипептид.

Антисенс-молекулы нуклеиновой кислоты изобретения могут быть придуманы с применением нуклеотидной последовательности TCF7L2 и/или комплементарной ей или ее части, сконструированы с применением химического синтеза и реакций ферментативного лигирования с применением процедур, известных в данной области. Например, антисенс-молекула нуклеиновой кислоты (например, антисенс-нуклеотид) может быть синтезирована химическим путем с применением природных нуклеотидов или различных модифицированных нуклеотидов, задуманных для увеличения биологической стабильности молекул или для увеличения физической стабильности дуплекса, сформированного между антисенс и смысловой нуклеиновыми кислотами, например могут быть применены фосфоротиоатные производные и замещенные акридином нуклеотиды. Альтернативно, антисенс-молекула нуклеиновой кислоты может быть получена биологически с применением экспрессионного вектора, в котором молекула нуклеиновой кислоты была субклонирована в антисенс-ориентации (т.е. РНК, транскрибируемая с встроенной молекулой нуклеиновой кислоты, будет в антисенс-ориентации по отношению к интересующей нуклеиновой кислоте-мишени).

Последовательности нуклеиновых кислот могут также применяться для сравнения с последовательностями эндогенной ДНК у пациентов для идентификации одного или более расстройств, описанных выше, и как зонды, например, для гибридизации и открытия последовательностей родственных ДНК или для вычитания известных последовательностей из образца. Последовательности нуклеиновых кислот могут дополнительно применяться для извлечения праймеров при генетическом фингерпринтинге, для выработки антиполипептидных антител с применением технологии иммунизации ДНК и как антиген для выработки анти-ДНК антител или вызывания иммунного ответа. Части или фрагменты нуклеотидных последовательностей, идентифицированные здесь (и соответствующие полные последовательности гена) могут применяться многими путями, например как полинуклеотидный реагент. К примеру, эти последовательности могут использоваться для (i) картирования соответствующих им генов в хромосоме, и таким образом, для выявления локализации областей гена, ассоциированных с генетическим заболеванием; (ii) идентификации индивидуума исходя из мельчайшего образца (типирование ткани) и (iii) помощи при судебной идентификации биологического образца. Дополнительно, нуклеотидные последовательности могут применяться для идентификации и экспрессии рекомбинантных полипептидов для анализа, характеристики или терапевтического применения или как маркеры для тканей, в которых экспрессируются соответствующие полипептиды либо конститутивно во время дифференциации ткани, либо при заболевании. Последовательности нуклеиновых кислот могут дополнительно применяться как реагенты в скрининге и/или диагностических анализах, описанных здесь, и могут также включаться как компоненты в наборы (например, наборы реактивов) для применения в скрининге и/или в диагностических анализах, описанных здесь.

Наборы (например, наборы реактивов), полезные в способах диагностики, включают компоненты, полезные в любых методах, описанных здесь, включая, к примеру, зонды или праймеры для гибридизации, как описано здесь (например, меченые зонды или праймеры), реагенты для детекции меченых молекул, рестрикционные ферменты (например, для RFLP анализа), аллель-специфичные олигонуклеотиды, антитела, которые связываются с измененным или с неизменным (нативным) полипептидом TCF7L2, средства для амплификации нуклеиновых кислот, включая нуклеиновую кислоту TCF7L2, или для части TCF7L2 или средства для анализа последовательности нуклеиновой кислоты или для анализа аминокислотной последовательности полипептида TCF7L2, как описано здесь, и т.д. В одном аспекте, набор реагентов для диагностики восприимчивости к диабету II типа может включать праймеры для амплификации области нуклеиновой кислоты TCF7L2, включающей маркер DG10S478, SNP rs12255372, rs7895340, rs11196205, rs7901695, rs7903146, rs12243326 и/или rs4506565, или гаплотип риска, который более часто представлен у индивидуумов, имеющих диабет II типа, или которые восприимчивы к диабету II типа. Праймеры могут быть сконструированы с применением частей нуклеиновых кислот, фланкирующих SNP, которые являются индикаторами диабета II типа.

Векторы и клетки-хозяева.

Другой аспект настоящего изобретения относится к конструктам нуклеиновой кислоты, включающим молекулы нуклеиновой кислоты, описанные здесь, и комплементарные им молекулы (или часть этого). Конструкты включают вектор (например, экспрессионный вектор, в который встроена последовательность изобретения в смысловой или антисенс-ориентации). Как применяется здесь, термин "вектор" относится к молекуле нуклеиновой кислоты, способной транспортировать другую нуклеиновую кислоту, с которой она была связана. Один тип векторов является "плазмидой", термин, который относится к циклической двутяжной петле ДНК, с которой может быть лигирован дополнительный сегмент ДНК. Другим типом векторов является вирусный вектор, где дополнительные сегменты ДНК могут быть лигированы с вирусным геномом. Определенные векторы способны к автономной репликации в клетках-хозяевах, в которые они интродуцированы (например, бактериальные векторы, имеющие бактериальное происхождение репликации, и эписомальные векторы млекопитающих). Другие векторы (например, не-эписомальные векторы млекопитающих) интегрируются в геном клетки-хозяина при интродукции в клетку-хозяин и, таким образом, реплицируются вместе с геномом хозяина. Экспрессионные векторы способны к направленной экспрессии генов, с которыми они операбельно связаны. В основном экспрессионные векторы, полезные в рекомбинантных технологиях ДНК, часто являются плазмидами. Однако изобретение намерено включить такие другие формы экспрессионных векторов, как вирусные векторы (например, дефективные по репликации ретровирусы, аденовирусы и вирусы, ассоциированные с аденовирусами), которые имеют эквивалентные функции.

В определенных аспектах векторы рекомбинантной экспрессии изобретения включают молекулу нуклеиновой кислоты изобретения в форме, подходящей для экспрессии молекулы нуклеиновой кислоты в клетке-хозяине. Это означает, что векторы рекомбинантной экспрессии включают один или более регуляторных последовательностей, выбранных на основе клеток хозяина для применения при экспрессии, которые операбельно связаны с последовательностью нуклеиновой кислоты, которая экспрессируется. В пределах рекомбинантного вектора экспрессии "операбельно связанный" или "оперативно связанный" предназначается для обозначения того, что интересующая нуклеотидная последовательность связана с регуляторной последовательностью(ями) способом, который позволяет обеспечить экспрессию нуклеотидной последовательности (например, система транскрипции/трансляции *in vitro* или в клетке-хозяине, когда вектор интродуцирован в клетку-хозяин). Термин "регуляторная последовательность" предназначен для включения промоторов, энхансеров и других регуляторных элементов экспрессии (например, сигналов полиаденилирования). Такие регуляторные последовательности описаны, к примеру, у Goeddel, "Gene Expression Technology", Methods in Enzymology. 185, Academic Press, San Diego, CA (1990). Регуляторные последовательности включают те, которые направляют конститутивную экспрессию нуклеотидной последовательности во многих типах клеток-хозяев, и те, которые направляют экспрессию нуклеотидной последовательности только в определенных клетках-хозяевах (например, тканеспецифичные регуляторные последовательности). Специалист оценит, что конструирование экспрессионных векторов может зависеть от таких факторов, как выбор клетки-хозяина, которая должна быть трансформирована, и уровня экспрессии желаемого полипептида. Экспрессионные векторы изобретения могут быть введены в клетки-хозяева для производства таким образом полипептидов, включая слитые полипептиды, кодируемые молекулами нуклеиновых кислот, как описано здесь.

Векторы рекомбинантной экспрессии изобретения могут быть сконструированы для экспрессии полипептида изобретения в клетках прокариот и эукариот, например в бактериальных клетках, таких как клетки *E.coli*, клетках насекомых (с применением бакуловирусного вектора экспрессии), дрожжевых клетках или клетках млекопитающих. Подходящие клетки-хозяева обсуждаются дополнительно у Goeddel, *supra*. Альтернативно, векторы рекомбинантной экспрессии могут быть транскрибированы и транслированы *in vitro*, например, с применением регуляторных последовательностей промотора T7 и T7 полимеразы.

Другой аспект изобретения относится к клеткам-хозяевам, в которые был интродуцирован вектор рекомбинантной экспрессии. Термины "клетка-хозяин" и "рекомбинантная клетка-хозяин" применяются здесь как взаимозаменяемые. Понятно, что такие термины относятся не только к определенным клеткам субъектам, но также к потомкам или потенциальным потомкам такой клетки. Поскольку в последующих поколениях могут иметь место определенные модификации, обусловленные либо мутациями, либо воздействием окружающей среды, такие потомки фактически могут не быть идентичны родительской клетке, но они еще включены в рамки термина, как он применяется здесь.

Клетка-хозяин может быть любой прокариотической или эукариотической клеткой. Например, молекула нуклеиновой кислоты изобретения может быть экспрессирована в бактериальных клетках (например, *E.coli*), клетках насекомых, клетках дрожжей или млекопитающих (таких как клетки яичника китайского хомячка (CHO) или клетках COS). Другие подходящие клетки-хозяева известны специалистам.

Вектор ДНК может быть интродуцирован в прокариотическую или эукариотическую клетку с применением традиционных методов трансформации или трансфекции. Как это применяется здесь, термины "трансформация" и "трансфекция" предназначены для отсылки ко многим известным специалистам ме-

тодам интродукции чужой молекулы нуклеиновой кислоты (например, ДНК) в клетку-хозяин, включая совместную преципитацию с фосфатом или хлоридом кальция, опосредованную DEAE-декстраном трансфекцию, липофекцию или электропорацию. Подходящие способы для трансформации или трансфекции клеток-хозяев можно найти у Sambrook et al. (supra) и других лабораторных пособиях.

Известно, что для стабильной трансфекции клеток млекопитающих в зависимости от примененного вектора и техники трансфекции, лишь малая часть клеток может интегрировать чужую ДНК в их геном. Для того чтобы идентифицировать и отобрать эти интегранты в клетку хозяин вместе с интересующим геном вводится обычно маркер для отбора (например, определяющий резистентность к антибиотикам). Предпочтительные маркеры для отбора включают такие, которые придают резистентность к лекарствам, таким как G418, гигромицин и метотрексат. Молекулы нуклеиновой кислоты, кодирующие маркер для отбора, могут вводиться в клетку-хозяин с тем же вектором, что и молекула нуклеиновой кислоты изобретения, или могут вводиться с помощью отдельного вектора. Клетка, стабильно трансфицированная введенной молекулой нуклеиновой кислоты, может быть идентифицирована путем селекции с применением лекарств (например, клетки, которые имеют включенный маркерный ген для селекции, будут выживать, в то время как другие клетки умрут).

Клетка-хозяин изобретения, такая как прокариотическая или эукариотическая клетка, может применяться в культуре для производства (т.е. экспрессии) полипептида изобретения. Соответственно изобретение дополнительно обеспечивает способы для производства полипептида с применением клеток-хозяев изобретения. В одном аспекте способ включает культивирование клетки-хозяина изобретения (в которую был интродуцирован вектор рекомбинантной экспрессии, кодирующий полипептид изобретения) в подходящей среде, так что производится полипептид. В другом аспекте способ дополнительно включает выделение полипептида из среды или клетки-хозяина.

Антитела изобретения.

Обеспечиваются также поликлональные антитела и/или моноклональные антитела, которые специфически связывают одну, но не другую форму продукта гена. Обеспечиваются также антитела, которые связывают часть или вариант, или продукт контрольного гена, который включает полиморфный участок или участки. Термин "антитело", как он применяется здесь, относится к молекулам иммуноглобулинов и иммунологически активным частям молекул иммуноглобулинов, т.е. к молекулам, которые включают антигенсвязывающие участки, которые специфически связывают антиген. Молекула, которая специфически связывает полипептид изобретения, представляет собой молекулу, которая связывается с полипептидом или его фрагментом, но, по существу, не связывается с другими молекулами в образце, например в биологическом образце, который в природе включает полипептид. Примеры активных частей иммуноглобулинов включают F(ab) и F(ab')<sub>2</sub> фрагменты, которые можно получить путем обработки антитела ферментом, таким как папаин. Изобретение обеспечивает поликлональные и моноклональные антитела, которые связываются с полипептидом изобретения. Термин "моноклональное антитело" или "композиция моноклонального антитела", как он применяется здесь, относится к популяции молекул антитела, которая включает только один вид антигенсвязывающего участка, способного к иммунореакции со специфическим эпитопом полипептида изобретения. Таким образом, композиция моноклонального антитела демонстрирует обычно единственную связывающую аффинность к специфическому полипептиду изобретения, с которым оно вступает в иммунную реакцию.

Поликлональные антитела могут быть получены, как описано выше, путем иммунизации подходящего субъекта желаемым иммуногеном, например полипептидом изобретения или его фрагментом. За титром антитела у иммунизированного субъекта можно следить во времени с применением стандартных методов, таких как иммуноферментный анализ на сорбенте (ELISA) с применением иммобилизованного полипептида. Если желательно, молекула полипептида, направленная против полипептида, может быть изолирована из млекопитающего (например, из крови) и дополнительно очищена с помощью хорошо известных методов, таких как хроматография на протеин А, для получения фракции IgG. В подходящее после иммунизации время, например, когда титры антитела являются наиболее высокими, клетки, производящие антитело, могут быть получены из субъекта и применены для приготовления моноклональных антител с использованием стандартных методов, таких как гибридомная техника, оригинально описанная у Kohler and Milstein, *Nature*. 256:495-497 (1975), техника с использованием человеческих гибридом B-клеток (Kozbor et al., *Immunol. Today*. 4:72 (1983)), EBV-гибридомная техника (Cole et al., *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, 1985, Inc. p. 77-96) или триомный метод. Технология для производства гибридом хорошо известна (см. главным образом *Current Protocols in Immunology* (1994) Coligan et al., (eds) John Wiley & Sons, Inc., New York, NY). Коротко, иммортализованная клеточная линия (обычно миелома) сливается с лимфоцитами (обычно спленоцитами) из млекопитающего, иммунизированного иммуногеном, как описано выше, и супернатанты культуры полученных гибридных клеток подвергаются скринингу для идентификации гибридомы, производящей моноклональное антитело, которое связывается с полипептидом изобретения.

Для получения моноклонального антитела к полипептиду изобретения может быть применен любой из многих хорошо известных протоколов, применяемых для слияния лимфоцитов и иммортализованных клеток (см., к примеру, *Current Protocols in Immunology*, supra; Galfe et al., *Nature*. 266:55052 (1977);

R.H. Kenneth, in *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980) и Lerner, *Yale J. Biol. Med.* 54:387-402 (1981)). Более того, обычный специалист оценит, что имеется много вариаций таких методов, которые также будут полезны.

Альтернативно, для получения гибридом, секретирующих моноклональное антитело, моноклональное антитело к полипептиду изобретения может быть модифицировано и изолировано путем скрининга рекомбинантной комбинаторной библиотеки иммуноглобулинов (например, библиотеки антител фагового дисплея) к полипептиду, к изолированному, таким образом, члену иммуноглобулиновой библиотеки, который связывает полипептид. Наборы реактивов для скрининга библиотек фагового дисплея доступны коммерчески (например, у Pharmacia Recombinant Phage Antibody System, каталожный номер 27-9400-01; и Stratagene SurfZAP™ Phage Display Kit, каталожный номер 240612). Дополнительно, примеры способов и реагентов, особенно подходящих для применения в получении и скрининге библиотеки антител фагового дисплея, могут быть обнаружены, например, в патенте US 5223409, патентной публикации PCT WO 92/18619; патентной публикации PCT WO 91/17271; патентной публикации PCT WO 92/20791; патентной публикации PCT WO 92/15679; патентной публикации PCT WO 93/01288; патентной публикации PCT WO 92/01047; патентной публикации PCT WO 92/09690; патентной публикации PCT WO 90/02809; Fuchs et al., *Bio/Technology*. 9:1370-1372 (1991); Hay et al., *Hum. Antibod. Hybridomas*. 3:81-85 (1992); Huse et al., *Science*. 246:1275-1281 (1989) и Griffiths et al., *EMBO J.* 12:725-734 (1993).

Дополнительно, в рамках изобретения находятся рекомбинантные антитела, такие как химерные и гуманизированные моноклональные антитела, включающие части, происходящие из человека и не из человека, которые могут быть получены с применением стандартных методов рекомбинантной ДНК. Такие химерные и гуманизированные моноклональные антитела могут быть получены с помощью методов рекомбинантной ДНК, известных в отрасли.

В основном антитело изобретения (например, моноклональное антитело) может применяться для выделения полипептида изобретения с применением стандартных методов, таких как аффинная хроматография или иммунопреципитация. Специфичное по отношению к полипептиду антитело может облегчать очистку природного полипептида из клеток и рекомбинантно получаемых полипептидов, экспрессируемых в клетках-хозяевах. Более того, антитело, специфичное по отношению к полипептиду изобретения, может быть применено для детектирования полипептида (например, в клеточном лизате, клеточном супернатанте или образце ткани), для того чтобы оценить концентрацию и характер экспрессии полипептида. Антитела могут применяться диагностически для регистрации уровней белка в тканях как часть клинической процедуры тестирования, например для определения эффективной данной схемы лечения. Антитело может быть сопряжено с детектируемым веществом для облегчения его детектирования. Примеры детектируемых веществ включают различные ферменты, простетические группы, флуоресцентный материал, люминесцентный материал, биолюминесцентный материал и радиоактивный материал. Примеры подходящих ферментов включают пероксидазу из хрена, щелочную фосфатазу, бета-галактозидазу или ацетилхолинэстеразу; примеры подходящих комплексов с простетическими группами включают стрептавидин/биотин и авидин/биотин; примеры подходящего флуоресцентного материала включают умбеллиферон, флуоресцеин, флуоресцеинизотиоцианат, родамин, дихлоротриазиниламин-флуоресцеин, дансилхлорид или фикоэритрин; пример люминесцентного материала включает люминол; примеры биолюминесцентных материалов включают люциферазу, люциферин и акворин, и примеры подходящего радиоактивного материала включают  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{35}\text{S}$  или  $^3\text{H}$ .

Настоящее изобретение теперь иллюстрируется следующими примерами, которые не ограничивают его любым путем.

### Примеры

Описанное здесь является идентификацией транскрипционного фактора 7, подобного фактору 2 (TCF7L2 - ранее TCF4), как гена, обеспечивающего риск развития диабета II типа путем точечного анализа ассоциаций с применением плотного набора микросателлитных маркеров в пределах локуса 10q.

#### Методы.

#### Исландская группа.

Агентство по Защите Информации Исландии и Национальный Биоэтический Комитет Исландии одобрили это исследование. Все участники исследования дали на него информированное согласие. Все персональные идентификаторы, ассоциированные с образцами крови, медицинской информацией и генеалогией, были сначала зашифрованы Агентством по Защите Информации с применением системы кодирования, полученной от третьей стороны (18).

Для первого исследования было идентифицировано 2400 пациентов с диабетом II типа, которые были диагностированы либо при долговременном эпидемиологическом исследовании, проведенном Исландской Ассоциацией Сердца в последние 30 лет, либо в одном из двух госпиталей в Рейкьявике в течение последних 12 лет. Две трети этих пациентов были живы, представляя примерно половину популяции известных пациентов с диабетом II типа в Исландии в настоящее время. Большая часть пациентов была в контакте с этим исследованием, доля кооперации превышала 80%. Все участники исследования посетили Исландскую Сердечную Ассоциацию, где они ответили на вопросы и где у них были взяты образцы кро-

ви и определен уровень глюкозы натощак. Были включены вопросы о лечении и возрасте, в котором был поставлен диагноз. Пациенты с диабетом II типа этого исследования были диагностированы, как описано в предварительном исследовании, связанном с проектом (10). Коротко, диагноз диабет II типа был подтвержден путем исследования врачами путем ранее сделанных медицинских регистраций, истории лечения и/или новых лабораторных измерений. Для пациентов с ранее диагностированным диабетом II типа, сообщивших о пероральном применении агента, снижающего уровень глюкозы, диагноз диабет II типа был подтвержден. Индивидуумы, которые к моменту исследования применяли инсулин, были классифицированы, как имеющие диабет II типа, если они также применяли или ранее перорально использовали агенты, снижающие уровень глюкозы. В этой группе большинство лечатся пациентов принимали агенты, снижающие уровень глюкозы, и лишь небольшая часть (9%) нуждалась в инсулине. Не диагностированным до исследования индивидуумам диагноз диабет II типа и нарушение уровня глюкозы натощак (IFG) был дан на основе критериев, перечисленных Американской Ассоциацией Диабета (Экспертный Комитет по Диагностике и Классификации Сахарного Диабета 1997). Средний возраст пациентов с диабетом II типа в этом исследовании составил 69,7 лет.

Дублирующие группы.

Датская группа проекта была отобрана из проекта PERF (Prospective Epidemiological Risk Factors - Предполагаемые эпидемиологические факторы риска) в Дании (19). Предварительно было диагностировано 228 женщин, как имеющие диабет II типа и/или уровень глюкозы в крови  $\geq 7$  mM. В качестве контроля случайным образом были отобраны 539 женщин той же группы, не имеющих диабета II типа.

Проект PENN CATN в США представляет собой перекрестное исследование ассоциации биохимических и генетических факторов при коронарном атеросклерозе в следующей друг за другом группе пациентов, которые подвергались коронарной катетеризации в Медицинском Центре Университет Пенсильвании между июлем 1998 и мартом 2003 г. Диабет II типа был определен как история повышения уровня глюкозы натощак  $\geq 126$  мг/дл, через 2 ч после обеда  $>200$  мг/дл, пероральное применение гипогликемических агентов или инсулина и оральная гипогликемия у субъектов в возрасте старше 40 лет. XXX Университета Пенсильвании одобрило протокол исследования, и все субъекты дали письменное информированное согласие. Этичность определяли путем самоотчетов. Из этой группы происходил 361 пациент белой расы с диабетом II типа. В том же исследовании 530 неподверженных заболеванию (в отношении диабета II типа и инфаркта миокарда) белых пациентов были отобраны случайным образом в качестве контроля.

ДНК, примененная для генотипирования, была продуктом амплификации всего генома, проведенной с применением набора реактивов GenomiPhi Amplification. ДНК для амплификации была получена из периферической крови датских пациентов и пациентов из США с диабетом II типа и контрольных субъектов.

Генотипирование.

Новые повторы, полученные при секвенировании (т.е. динуклеотиды, тринуклеотиды и тетрануклеотиды), были идентифицированы с применением программ Tandem repeats finder (20) и тестированы на полиморфизм у 94 контрольных субъектов. Размер в парах оснований меньшего аллеля CEPH образца 1347-02 (CEPH genomics repository) вычитали из размера ампликона микросателлита и применяли в качестве контроля. SNP генотипирование было проведено с применением прямого секвенирования ДНК (Applied Biosystems) или платформы Centaurus (Nanogen).

Статистические методы анализа ассоциации.

Для установления ассоциации единственного маркера с диабетом II типа применяли тест коэффициента вероятности для расчета двусторонних значений  $p$  для каждого аллеля. Для исследованных микросателлитов представим частоты аллелей, а не несущую частоту.

Авторы изобретения рассчитали RR и риск, приписанный к популяции, на основе мультипликативной модели (16, 17). Для данных по белым пациентам HarMap CEPH рассчитали LD между парами SNP с применением стандартного определения  $D'$  (27) и  $R^2$  (22). При построении графика для всех комбинаций SNP для выявления LD структуры в отдельной области помещали  $D'$  в верхний левый угол, а значения  $p$  - в нижний правый угол. В представленном графике LD маркеры размещены равноудаленно, а не в соответствии с их физическим расположением.

### Результаты

Широколокусное исследование ассоциации.

Предварительно авторы изобретения сообщали о широкогеномной существенной связи сахарного диабета II типа с хромосомой 5q в исландской популяции (10); в том же исследовании также сообщали о предположительном доказательстве связи заболевания с 10q и 12q. Для завершения исследования локуса 10q применяли ассоциативный подход, используя высокую плотность генотипированных микросателлитных маркеров в 10,5 м.п.о. области (NCBI, версия 34: Chr10:114,2 - 124,7 м.п.о.), соответствующей этому локусу. Идентифицировали и типировали 228 микросателлитных маркеров, т.е. до средней плотности 1 маркер на каждые 46 т.п.о. (табл. 1). Все маркеры были типированы у 1185 исландских пациентов с диабетом II типа и у 931 индивидуума из неродственной контрольной популяции.

Таблица 1

Локализация 228 генотипированных микросателлитов в хромосоме 10 в конструкции человеческого генома NCBI, версия 34

Название	СТАРТ: Локализация в хромосоме 10, версия 34	КОНЕЦ: Локализация в хромосоме 10, версия 34
D10S1269	114186051	114186276
DG10S475	114389853	114390116
D10S168	114410102	114410266
DG10S478	114460845	114461228
DG10S479	114475488	114475632
DG10S480	114507574	114507829
DG10S481	114542657	114542924
DG10S1624	114545990	114546237
DG10S1625	114568323	114568715
DG10S488	114713594	114714008
DG10S1630	114770344	114770609
DG10S1631	114778307	114778598
DG10S492	114811884	114812269
DG10S494	114852114	114852280
DG10S495	114879344	114879474
DG10S496	114919414	114919678
DG10S498	114964123	114964270
DG10S500	115024471	115024854
DG10S501	115045332	115045710
DG10S508	115241356	115241602
DG10S1634	115267106	115267460
DG10S512	115357290	115357439
DG10S514	115400157	115400338
DG10S17	115463773	115464048
DG10S1635	115519619	115519900
DG10S520	115536945	115537130
D10S554	115695920	115696071
D10S1237	115784580	115784977
DG10S535	115858565	115858720
D10S1158	115937134	115937433
DG10S1636	115966165	115966382
DG10S540	115983225	115983471
DG10S1637	116025219	116025491
DG10S542	116054130	116054255
DG10S1638	116062921	116063264
D10S1776	116140681	116140897
DG10S546	116141340	116141590
DG10S547	116173634	116173887
DG10S1639	116184720	116184898
DG10S548	116202775	116203174
DG10S550	116288175	116288560
D10S562	116304948	116305132
DG10S1640	116344030	116344279
DG10S1641	116638155	116638540
DG10S566	116866173	116866431
D10S468	116869582	116869674
DG10S567	116904174	116904433
D10S1731	117001692	117001870
DG10S573	117070087	117070192
DG10S576	117153566	117153823
DG10S578	117196538	117196813
DG10S1644	117206992	117207391
DG10S579	117226056	117226234
DG10S580	117240674	117240858
DG10S584	117336471	117336821
DG10S585	117364742	117364845
DG10S586	117385650	117385816
DG10S589	117481892	117482165
DG10S590	117508690	117508966

DG10S591	117520912	117521057
DG10S593	117567541	117567800
D10S1748	117589638	117589885
DG10S596	117629981	117630119
DG10S597	117654759	117654928
DG10S523	117691905	117692329
DG10S598	117691905	117692156
D10S1773	117708786	117708989
DG10S599	117713714	117714115
DG10S524	117713997	117714115
DG10S600	117742602	117743019
DG10S525	117742701	117742986
DG10S1250	117861226	117861405
DG10S604	117867801	117868010
DG10S1293	117932494	117932721
DG10S1144	117950298	117950606
DG10S609	118014503	118014752
DG10S610	118041410	118041787
DG10S1252	118085912	118086081
DG10S612	118092869	118093247
DG10S613	118126058	118126312
DG10S614	118150018	118150178
D10S544	118164684	118164979
D10S1683	118211053	118211180
D10S1657	118287426	118287695
D10S545	118299618	118299851
DG10S1649	118306954	118307121
D10S187	118317655	118317730
DG10S1295	118375973	118376205
DG10S624	118401694	118402073
DG10S1203	118440472	118440835
DG10S627	118514695	118515072
DG10S1650	118521021	118521210
DG10S1681	118522946	118523333
DG10S628	118553693	118553836
DG10S634	118566844	118567191
DG10S639	118712208	118712596
DG10S640	118743450	118743821
D10S221	118766458	118766560
DG10S1686	118766464	118766561
DG10S641	118788135	118788401
DG10S1651	118794961	118795267
DG10S1255	118834290	118834438
DG10S644	118857362	118857745
DG10S1652	118862172	118862311
DG10S1654	118954536	118954869
DG10S1688	118972583	118972717
DG10S1689	118987319	118987480
DG10S1690	119004704	119004986
D10S1425	119004742	119004920

DG10S651	119030166	119030595
DG10S1655	119044005	119044188
DG10S1691	119078576	119078943
DG10S1207	119094382	119094722
D10S1693	119109493	119109731
DG10S1258	119131611	119131788
DG10S656	119177278	119177672
DG10S1694	119177430	119177614
DG10S1695	119204432	119204655
DG10S657	119204769	119205174
DG10S658	119223917	119224102
DG10S1696	119243071	119243408
DG10S1657	119282299	119282586
DG10S1658	119290241	119290632
DG10S661	119305067	119305226
DG10S662	119317406	119317660
DG10S663	119330718	119331131
DG10S1699	119364904	119365188
DG10S665	119396863	119397144
DG10S1659	119412611	119412992
DG10S667	119448478	119448736
DG10S1701	119473676	119473914
D10S1236	119473739	119473870
DG10S669	119485378	119485552
DG10S670	119505799	119505905
D10S190	119510348	119510554
DG10S1702	119510362	119510479
DG10S1153	119526060	119526329
DG10S673	119606691	119606963
DG10S1305	119615268	119615484
DG10S675	119659153	119659532
DG10S1661	119663175	119663453
DG10S1662	119700563	119700948
DG10S1306	119703996	119704204
DG10S1663	119783538	119783739
DG10S1704	119783569	119783694
DG10S631	119788517	119788678
D10S1148	119803465	119803663
D10S1150	119803465	119803662
D10S503	119803476	119803653
DG10S632	119811193	119811621
DG10S681	119811347	119811621
DG10S633	119833701	119833987
D10S2473	119833724	119833869
DG10S682	119838539	119838806
DG10S683	119853558	119853862
DG10S684	119880412	119880572
DG10S685	119909682	119910062
DG10S686	119923527	119923790
DG10S687	119954835	119955083

DG10S1212	119972358	119972707
DG10S1261	119995566	119995727
DG10S1350	120004924	120005036
DG10S1	120030830	120031131
DG10S693	120100794	120101005
DG10S1263	120132349	120132528
D10S542	120417003	120417230
DG10S1664	120444685	120444808
DG10S1163	120506796	120507066
DG10S703	120538236	120538484
DG10S704	120570334	120570593
DG10S706	120642052	120642312
DG10S708	120699520	120699811
DG10S709	120723780	120724158
D10S1701	120849161	120849428
DG10S716	120893782	120894153
DG10S1669	120969521	120969659
DG10S720	121016792	121017048
D10S1792	121042408	121042574
DG10S722	121070320	121070693
DG10S1181	121101362	121101685
DG10S724	121117025	121117286
DG10S1670	121162511	121162898
DG10S726	121217327	121217580
DG10S1167	121247552	121247838
DG10S729	121283257	121283429
DG10S730	121318865	121319131
DG10S731	121342622	121342893
DG10S1278	121384227	121384464
DG10S734	121425229	121425633
DG10S735	121446549	121446695
DG10S1185	121466936	121467248
DG10S1129	121472295	121472600
DG10S1085	121494260	121494657
DG10S1327	121526700	121526830
DG10S1271	121559895	121560066
DG10S741	121638254	121638391
DG10S1087	121647884	121648273
DG10S1359	121713760	121713892
DG10S1120	121726128	121726519
DG10S1671	121750886	121750993
DG10S1673	121823695	121823925
DG10S749	121841816	121841997
DG10S1134	121901381	121901668
DG10S1674	121931406	121931809
DG10S755	121976143	121976435
D10S1757	121989325	121989539
D10S209	121995173	121995376
DG10S757	122029990	122030248
DG10S1283	122045222	122045429
DG10S1191	122071761	122072115
DG10S761	122141102	122141322
DG10S1678	122146312	122146535
DG10S762	122167889	122168135
DG10S763	122185793	122185925
DG10S1284	122207287	122207508
DG10S1137	122220809	122221073
DG10S766	122257534	122257929
DG10S767	122283871	122284250
DG10S1361	122318975	122319081
DG10S1680	122390160	122390294
D10S1230	122407279	122407403
DG10S772	122421708	122421845
DG10S775	122463781	122463941
DG10S777	122524358	122524547
DG10S779	122580228	122580603
DG10S784	122719087	122719236
D10S1483	122948181	122948324
D10S587	124728937	124729112

Анализ ассоциации единичного маркера с маркерами микросателлитами идентифицировал ассоциацию с DG10S478 (табл. 2 и фигура).

Таблица 2

## Ассоциация DG10S478 с диабетом II типа в Исландии

Аллель	Частота у страдающих заболеванием (n=1185)	Контрольная частота (n=931)	RR [95% Дов.инт]	Двустороннее Р
0	0,636	0,724	0,67	$2,1 \times 10^{-9}$
4	0,005	0,002	2,36	0,12
8	0,093	0,078	1,21	0,090
12	0,242	0,178	1,48	$4,6 \times 10^{-7}$
16	0,022	0,015	1,53	0,076
20	0,001	0,003	0,39	0,17
<b>X</b>	<b>0,364</b>	<b>0,276</b>	<b>1,50[1,31; 1,71]</b>	<b><math>2,1 \times 10^{-9}</math></b>

Наблюдалось шесть аллелей с этим тетрауклеотидным повтором, с аллелями 0, 8 и 12, несущие ответственность за 98% хромосом в контрольной популяции. Аллель 0 демонстрировал ассоциацию с защитой от диабета II типа (Relative Risk (RR)=0,67;  $P=2,1 \times 10^{-9}$ ) относительно других аллелей, взятых вместе. Это значение Р является двусторонним и принимает во внимание, что некоторые пациенты являются родственниками. DG10S478 локализован в интроне 3 гена транскрипционного фактора 2, подобного 7 (TCF7L2 - бывший TCF4) гена на 10q25,2. Этот маркер находится в пределах хорошо определенного LD блока в 74,9 т.п.о. (на основе CEPH HarMap, фаза II для белой расы), которая включает часть интрона 3, весь экзон 4 и часть интрона 4 (см. фигура).

Когда DG10S478 был генотипирован в семьях CEPH HarMap европейцев, стало ясно, что наблюдается почти совершенная корреляция аллеля G SNP rs12255372 с аллелем 0 DG10S478 ( $r^2=0,95$ ,  $P=5,53 \times 10^{-38}$ ), а аллель Т в rs12255372 коррелирует с другими аллелями DG10S478. Более того, риск возникновения заболевания, который придают аллелями 8 и 12 DG10S478, не отличается ( $P=0,3$ ). Поэтому естественно свести все не-0 аллели DG10S478 в составной аллель, который был назван X аллелем. Аллель X имеет частоту 27,6 и 36,4% в контроле и у пациентов соответственно. На основе предположения о мультипликативной модели (16, 17) по сравнению с риском для пациентов, не являющихся носителем, частота аллеля X была оценена значением RR 1,50 на имеющуюся в геноме копию.

Повторное исследование ассоциации DG10S478 с диабетом II типа.

Для верификации ассоциации DG10S478 с диабетом II типа микросателлиты были генотипированы в датской популяции для 228 больных пациентов и 539 контрольных. Датская группа была отобрана из проекта PERF, осуществляемого в Дании (19). Эта женская группа больных диабетом II типа была диагностирована предварительно. Ассоциация, наблюдаемая в Исландии, повторилась (табл. 3).

Таблица 3

## Ассоциация DG10S478 с диабетом II типа в Дании

Аллель	Частота у пораженных заболеванием (n=228)	Контрольная частота (n=539)	RR [95% Дов.инт.]	Двустороннее Р
0	0,669	0,740	0,71	0,0048
4	0,002	0,004	0,59	0,62
8	0,070	0,048	1,49	0,091
12	0,239	0,190	1,34	0,032
16	0,020	0,018	1,12	0,78
<b>X</b>	<b>0,331</b>	<b>0,260</b>	<b>1,41 [1,11;1,79]</b>	<b>0,0048</b>

Составной аллель риска X имеет частоту 26,0% в контроле и 33,1% в случае диабета II типа, давая рассчитанный RR 1,41 ( $P=0,0048$ ).

Позднее микросателлиты были генотипированы у представителей белой расы в США в 361 случае и у 530 контролей, участвовавших в исследовании PENN CATH. Это исследование является перекрестным исследованием ассоциации биохимических и генетических факторов при коронарном атеросклерозе в следующей друг за другом группе пациентов, которые подвергались коронарной катетеризации в Медицинском Центре Университета Пенсильвании. Диабет II типа был определен как история повышения уровня глюкозы натощак  $\geq 126$  мг/дл, через 2 ч после обеда  $\geq 200$  мг/дл, пероральное применение гипогликемических агентов или инсулина и оральная гипогликемия у субъектов в возрасте старше 40 лет. В этой популяции также повторилась ассоциация, наблюдаемая в Исландии (табл. 4).

Таблица 4

## Ассоциация DG10S478 с диабетом II типа в США

Аллель	Частота у пораженных заболеванием (n=361)	Контрольная частота (n=530)	RR [95% Дов.инт.]	Двустороннее Р
-4	0,001	0,000	-	-
0	0,615	0,747	0,54	$3,3 \times 10^{-9}$
4	0,003	0,004	0,73	0,72
8	0,085	0,049	1,79	0,0029
12	0,256	0,180	1,57	$1,2 \times 10^{-4}$
16	0,040	0,020	2,07	0,012
X	0,385	0,253	1,85 [1,51, 2,27]	$3,3 \times 10^{-9}$

Составной аллель риска X имеет частоту 25,3% в контроле и 38,5% в случае диабета II типа, давая рассчитанный RR 1,85 ( $P=3,3 \times 10^{-9}$ ). Объединение результатов из всех трех групп с применением модели Мантел-Ханесцел (Mantel-Haneszel) (NOTE 3) дает общее значение двустороннего Р  $4,7 \times 10^{-18}$ .

Ассоциация составного аллеля риска для диабета II типа в трех популяциях дает строгое свидетельство, что варианты гена TCF7L2 вносят вклад в риск развития диабета II типа.

После устранения сомнений относительно ассоциации аллеля X с диабетом II типа исследовали характер наследования более подробно. Доминантная и рецессивная модели могут быть отвергнуты, поскольку гетерозиготные носители явно имели повышенный риск относительно не носителей ( $P < 1 \times 10^{-6}$ ) и пониженный риск по сравнению с гомозиготными носителями ( $P < 0,0001$ ). Мультипликативная модель обеспечивает лучшее соответствие, но имеется свидетельство, что риск у гомозиготных носителей относительно гетерозиготных носителей больше, чем таковой у гетерозиготных носителей относительно не носителей. Табл. 5 обеспечивает свободные от модели оценки относительного риска гетерозиготных носителей по сравнению с не носителями.

Таблица 5

## Свободные от модели оценки относительных рисков

Группа	Относительный риск генотипа			
	00	OX [95% Дов.инт.]	XX [95% Дов.инт.]	
Исландия	1	1,41 [1,17;1,70]	2,27 [1,70;3,04]	0,21
Дания	1	1,37 [0,98;1,90]	1,92 [1,13;3,26]	0,17
США	1	1,64 [1,23; 2,19]	3,29 [2,13; 5,07]	0,28
Объединенная	1	1,45 [1,26, 1,67]	2,41 [1,94; 3,00]	0,21

Три группы имели похожую популяционную частоту для аллеля риска, но оценки RR варьировали с наиболее сильным эффектом, наблюдаемым в группе из США, и наиболее слабым в датской группе. Хотя нет причины для того, чтобы RR был идентичен в группах, отмечается, что различия в оцененном относительном риске не вполне достигают статистической значимости ( $P > 0,05$ ). При комбинации результатов из групп и предполагая общие относительные риски, гетерозиготные носители и гомозиготные носители оцениваются, как имеющие относительные риски 1,45 и 2,41 соответственно по сравнению с не-носителями (табл. 5). Если предполагать популяционную частоту аллеля риска 26%, то гетерозиготные и гомозиготные носители составляют 38 и 7% популяции соответственно. Следовательно, этот вариант имеет достаточно предсказуемое значение для клинического применения. Соответствующий риск, приписанный популяции, составляет 21%, что является существенным с точки зрения здоровья общества.

Необходимо также отметить, что аллель X имеется в избытке у индивидуумов с нарушением уровня глюкозы в крови натощак (IFG) (концентрация глюкозы в сыворотке крови натощак между 6,1 и 6,9 мМ). Составной аллель X имеет частоту 27,7% у 1393 контролей и 37,1% у 278 IFG, давая оцененный RR 1,54 ( $P=1,36 \times 10^{-5}$ ).

Ассоциация SNP маркеров в LP блоке экзона 4 TCF7L2 с диабетом II типа.

В табл. 6 приведен список микросателлитов и SNP маркеров, находящихся в LD блоке в пределах экзона 4 TCF7L2. Таблица включает как общедоступные SNP, так и SNP, открытые при секвенировании полной области LD блока. Кроме того, таблица обеспечивает данные о полиморфных микросателлитных маркерах, находящихся в пределах блока.

Полиморфные маркеры, локализованные в пределах LD блока экзона 4 TCF7L2 (между маркерами rs4074720 и rs7087006, позиции в соответствии с координатами версии 34: rs4074720 (B34: 114413084) - rs7087006 (B34: 114488013) = 74929 п.о. Ссылки на идентификацию последовательностей указаны как присвоенные, отсылающие в каждом случае к номеру SEQ ID для амплимера, включающего полиморфизм, и прямые и обратные праймеры, как раскрыто в Списке последовательностей.

Таблица 6

Общедоступные SNP (включая все этнические группы HarMap)

Название в общедоступном источнике	Локализация в хромосоме 10, вер.34	Изменение основания	Sequence ID NO:
rs4074720	114413084	A/G	17-19
rs4074719	114413145	C/T	
rs4074718	114413204	C/T	
rs11196181	114413605	A/G	
rs1196182	114414744	C/T	
rs4603236	114414765	G/T	
rs7922298	114414856	C/T	
rs17747324	114417090	C/T	
rs7901695	114418675	C/T	
rs11196185	114420079	C/T	
rs4132115	114420083	A/C	14-16
rs4506568	114420628	A/T	
rs7068741	114420845	C/T	
rs7069007	114420872	C/G	
rs7903146	114422936	C/T	11-13
rs11196187	114424032	A/G	
rs7092484	114425520	A/G	
rs10885402	114426284	A/C	
rs12098651	114426306	A/G	
rs6585198	114426824	A/G	
rs7910244	114427209	C/G	
rs12266632	114429546	C/G	
rs6585199	114429758	A/G	
rs7896811	114431304	C/T	8-10
rs6585200	114433196	A/G	
rs6585201	114433370	A/G	
rs4319449	114433993	G/T	
rs12220336	114434854	A/G	
rs7896091	114436550	A/G	
rs12354626	114437016	A/G	
rs7075199	114437307	C/G	
rs7904519	114438514	A/G	
rs13376896	114441336	A/C	
rs10885405	114442257	C/T	8-10
rs10885406	114442311	A/G	
rs11196192	114446874	G/T	
rs6585202	114447390	C/T	
rs7924080	114451599	C/T	
rs7907610	114451677	A/G	
rs12262948	114452313	C/G	
rs12243326	114453402	C/T	
rs12265110	114453606	C/T	
rs7077039	114453664	C/T	

rs11196198	114456472	A/G	23-25
rs12775336	114459590	G/T	
rs7904948	114459672	A/T	
rs7100927	114460635	A/G	
rs11196199	114460704	A/G	
rs17685538	114462058	C/G	
rs11592706	114463573	C/T	
rs7081912	114463678	A/G	
rs7895340	114466112	A/G	
rs11196200	114466525	C/G	
rs11196201	114467894	A/T	20-22
rs11196202	114470254	A/G	
rs11196203	114470447	A/C	
rs11196204	114470518	A/G	
rs11196205	114471634	C/G	5-7
rs10885409	114472659	C/T	
rs12255372	114473489	G/T	
rs12265291	114474827	C/T	
rs7904443	114475774	A/G	
rs11196208	114475903	C/T	
rs7077247	114476658	C/T	
rs11196209	114477314	A/G	
rs4077527	114477628	A/G	
rs12718338	114477634	C/T	
rs11196210	114478558	C/T	
rs7907632	114481823	A/G	
rs7071302	114482114	G/T	
rs12245680	114484778	C/T	
rs11196213	114486141	C/T	
rs4918789	114486394	G/T	
rs7085785	114487050	C/T	
rs7085989	114487326	A/G	
rs7087006	114488013	A/G	

В. Новые SNP, открытые и впоследствии подтвержденные в LD блоке экзона 4 TCF7L2 (амплимеры ниже).

Название согласно decode	Локализация в хромосоме 10, вер. 34	Замена основания	Sequence ID NO:
SG10S405	114418658	C/T	26-28
SG10S428	114421901	A/C	29-31
SG10S422	114457824	A/G	32-34
SG10S427	114463480	A/T	35-37
SG10S408	114466074	A/T	38-40
SG10S409	114471574	A/C	41-43
SG10S406	114471618	C/G	42-44
SG10S407	114473534	C/G	45-47

С. Полиморфные микросателлиты в пределах LD блока экзона 4 TCF7L2 (амплимеры ниже).

Микросателлиты	Начало C10, вер.34	Конец C10, вер.34	Sequence ID NO:
DG10S2164	114460344	114460627	48-50
DG10S478	114460845	114461228	2-4
DG10S479	114475487	114475632	51-53

Таблица 7

Амплимеры и праймеры для избранных маркеров в пределах LD блока экзона  
TCF7L2 4

>DG10S478

TTCAGGCCATTGGTGTGTATATATTTCAAGATTGCTCACAGGTCCAAA  
GCTTAACTTAAGCTCCCTGAGACATATCATAAAATATGATTGGGGAAAA  
ACCCTAATGGGCCATGATCAGAACATTATTATTCAACAAAGGATGAAATG  
CTTAAGCCAAGATGGCCTTCTTTCTTTCTTTCTTTCTTTCTTTTAAAA  
ATGAAAGTTGAGCAGACTCCCGTCCAACAGTTTTCAATGTAGGAATTCCC  
ACAGCCCCATTGATTGCAGTTTGTGAAAAGTTTAATGTTTTTGTAGGC  
AATTCATAATTTCCACATTGAACAGCCTGAGAGGAAGAGAGCTGGAGCCC  
ACTGTTGTTTTGTAGTGGGATGGTGGGAACCTT (SEQ ID NO:2)

**Праймеры:**

F: TTCAGGCCATTGGTGTGTGTA (SEQ ID NO:3)

R: AAAGTTCCCACCATCCCACT (SEQ ID NO:4)

>rs12255372

TTGTCCCTTGAGGTGTACTGGAAACTAAGGCGTGAGGGACTCATAGGGG  
CTGGCTTGGAAGTGATTTGCTATGTCCAGTTTACACATAAGGATGTGCA  
AATCCAGCAGGTTAGCTGAGCTGCCCAGGAATATCCAGGCAAGAAT

K

ACCATATTCTGATAATTACTCAGGCCTCTGCCTCATCTCCGCTGCCCCC  
CGCCCCCTGACTCTCTTCTGAGTGCCAGATTCAGCCTCCATTTGAATGCC  
AAATAGACAGGAAATTAGCATGCCCAGAATCCACGTCTTTAGTGCACTC  
CTCCCCAGCTCCAAACCTGTTACTGCTTGTGTTCAACATCTCAGTAAAGC  
TCAACAACATCGACCCATT (SEQ ID NO:5)

**Праймеры:**

F: TTGTCCCTTGAGGTGTACTGG (SEQ ID NO:6)

R: AATGGGTCGATGTTGTTGAG (SEQ ID NO:7)

&gt;rs12243326

GCTGTGAAATCCCCTGTGTAGTGGGAAGAAGAAATAGCAAATCTTAGCTG  
CCTTGGACCTGATATAATTATTGTCTTCATTTACATGGTT

Y

ATCCTTCAAGGTTGAATAAATGATGTGGGAGCTAGTCAAGGGGCTTTAGG  
TATGTGATTTTCATGCCTACTTTTTTTTAGGTAGAGAACTGAGGTCACAG  
GGTACTAGAGAATGGACTCTAAGATTCAGGTTTCTGAATTGCCTGTGGTT  
TTGTTGACTCAACTGCTCTTCTGTTGTTTTTTAGCCACATGCCTTGAAAC  
AGTCCTCTTTCCCATGTTTCTTCATCAGCACCATTAAACCAAGGTATACT  
GTCCTCTCTTATCTTTTACAAGGTCTTGGAGTTCCCATGCCTTTGTAAGC  
ATCCCTCCCCGAGATTCAGCACCAACCAAAATCACATTTGGAAAAATTGC  
TTGTTTCCCAAGAAGCTTTGGAGGATATGATTTTGTATAGAACGGGTTC  
CAGGTTTCTGTTCATTCTTCTATGGTGGAGTGTGTGTGTATGTGACTCT  
GTCTTCTCTCCATTCC (SEQ ID NO:8)

**Праймеры:**

F: GCTGTGAAATCCCCTGTGTAG (SEQ ID NO:9)

R: GGAATGGAGAGAAGACAGAGTCA (SEQ ID NO: 10)

&gt;rs7903146

AAGGGAGAAAGCAGGATTGAGCAGGGGAGCCGTCAGATGGTAATGCAGA  
TGTGATGAGATCTCTGCCGGACCAAAGAGAAGATTCCTTTTTAAATGGTG  
ACAAATTCATGGGCTTTCTCTGCCTCAAAACCTAGCACAGCTGTTATTTA  
CTGAACAATTAGAGAGCTAAGCACTTTTAGATA

Y

TATATAATTTAATTGCCGTATGAGGCACCCTTAGTTTTTCAGACGAGAAAC  
CACAGTTACAGGGAAGGCAAGTAACTTAGTCAATGTCAGATAACTAGGAA  
AAGGTTAGAGGGGCCCTGGACACAGGCCTGTGTGACTGAGAAGCTTGGGC  
ACTTCACTGCTACATTTTCATCTCTTCGCT (SEQ ID NO: 11)

**Праймеры:**

F: AAGGGAGAAAGCAGGATTGA (SEQ ID NO: 12)

R: AGCGAAGAGATGAAATGTAGCA (SEQ ID NO: 13)

&gt;rs4506565

CTGATGAGGGTAGGGAGCATCTGTCTGCAGCTTCATCTTCATTGTCTAGG  
GGCTCCAGAAATATCTGTGAGTAAATAAGTTATTTAATCTTTGCCTCAAA

TTTCCAGTGACTGTAGGGATATAGCTGTGAGCCTCTAGGAGCTGAGATT  
 TTTAAATTTCCCACTTAAACATTTATTTAAAAATTTTGTGCTCAGCATGG  
 АСТААГГАСТТТАСАТТКАТТААСТКАТТТАСАГСТТГАТСТТАТГССГТ  
 ГГГСАТТКАТТТАСАГАГГАТСССАТТТТАСАГСТГАГГААГАГГССАГ  
 ТАГГГГТГСАГССТАГГТТАГТАТТТАГАГСТКАТСАГСТТГТГТГТ  
 СССАГТГААГААТААГСААГААГТААТГТТГТГАТТГАААААТГ  
 АСТСТСГГАГГАГГАТГАГСТСТСГГАТАТГГСАССГААГТАТ  
 W  
 TGGGGCCCTTGTCAAGGGTCTCTATTATGGCATCAAGAAAAGATGCTGCT  
 TTCCGTGATGCCCAGGAGAGCCTCAATATTTTACATGGGAAACCTAAAA  
 AAGGGGCCATGTTGTGGTCTCTGCACCTAAGA (SEQ ID NO: 14)

**Праймеры:**

F: CTGATGAGGGTAGGGAGCA (SEQ ID NO: 15)

R: TCTTAGGTGCAGAGACCACAAC (SEQ ID NO: 16)

&gt;rs7901695

TATTTAGAAAACCATAAAAATCCACCTATTTGAGGTGTACAATTGAGTGATT  
 TTCTGTATAGTCACAGATCTGTGCAGTCATCCACACCTCTAACTCCAGG  
 ACATTTTCCTCACCCCGAGGAGAAACCTCCCTTACCCATTAGCAGTCAC  
 TCCTCATTTCTCTCCCCCAGCCCCTGGCAATCACTGTGGATTTGCCTG  
 TTCTTGACATTTTCATATAAATGGTATCATAAAATCTA  
 Y  
 GGGCTTTTGTGTCTGTCTGCTTTCACTTAGCATACGGTTCTCAAGGTTC  
 TCCAGTATTGTAGCATCTATCAGTATGTCATTCTTTTATGGCCAAATA  
 ATATTTTATTGTATGGATAGACATTTTGTATTATCATTTATCTGTTTTG  
 GTTATTATGAGTAACACTACTATGAACATTTGCACAAATTTTGTATTG  
 ACATGTTTTCATTTCTCCTGGGTATAGTCCTATGAGTGGAATTGCTGG  
 (SEQ ID NO: 17)

**Праймеры:**

F: TATTTAGAAAACCATAAAAATCCACCTAT (SEQ ID NO: 18)

R: CCAGCAATCCACTCATAGGAC (SEQ ID NO: 19)

&gt;rs11196205

TTGTCTCSTTTTGTCTCTGCTACTGTGAATGATCCTGTGATGATCATCTT  
 TGTGTGTAATCTTTGTCCCCTCGCCCCCTCCCCTTTTATTATTTCTTG

GGATAGACCCCAGGACAAAAGGTAGAAAAGAACAAAGTGTTAAAAAATTT  
CTTGATACATAGCCACAGATTATTTTCCTGAAAAGTTCTCAACATTTATAA  
CTAC

S

AGCAGTATGTAAGAGAGTTATGGTTGGAATGATTTTAATGTCTCTGGGGA  
ATTTAACAACAAAAAACTTTAGGCTTCTTTGGAGAGAGACATGCCCTTA  
ACTCCACCCCGCCCTAGAACAGAGACCCAGCCCATCCAAGTCAGCCTCCC  
CAGGTCCTCCACCTTCAAAACAGGCAAACGAAATCATTCTTGAATAATT  
GGTAGGCTTCAAGGTCAGATGTT (SEQ ID NO:20)

**Праймеры:**

F: TTGTCTCCTTTTGTCTGCTAC (SEQ ID NO:21)

R AACATCTGACCTTGAAGCCTAC (SEQ ID NO:22)

>rs7895340

TCAGGGACAGTGCATAGGTGTAAGAAGTTGCTGGTTGGGGGTTCTAATG  
CAGGTTTCTCCAAAAGTGAATGCCCTGTAAAAAAAATTCTTAACAAAT  
ATACAGAGATTTTTTTTTTAAAAAAGTGTGACAGTTCTAGACACCTAGAG  
AGTAAA

R

TGAAGAAGCCTGTTTTTCAGGTTTCCCGCCTCCCTGAATTTCCCAGCATGG  
TCCAGGCTTTGAAATTTATTTATCTGCTTTTGGCAATGGTTGATGGGAAT  
TTCCACATTTATTTTTTAGCTACAGAGAAAGGACATTATCTTTAAATC  
TCTTCGTTGTTCTCTCTTTGA (SEQ ID NO:23)

**Праймеры:**

F: TCAGGGACAGTGCATAGGTG (SEQ ID NO:24)

R: TCAAAGAGAGAGAACAACGAAGA (SEQ ID NO:25)

>SG10S405

TATTTAGAAACCATAAAAATCCACCTATTTGAGGTGTACAATTGAGTGATT  
TTCTGTATAGTCACAGATCTGTGCAGTCATCCACACCCTCTAACTCCAGG  
ACATTTTCCTCACCCCGAGGAGAAACCTCCCTTACCCATTAGCAGTCAC  
TCCTCATTTCTCTCCCCCAGCCCTGGCAATCACTGTGGATTTGCCTG  
TTCTTGACATTTTCATATAAA

Y

GGTATCATAAAATCTATGGGCTTTTGTGTCTGTCTGCTTTCACTTAGCAT

ACGGTTCTCAAGGTTTCATCCAGTATTGTAGCATCTATCAGTATGTCATT  
 CTTTTATGGCCAAATAATATTTTATTGTATGGATAGACATTTTGTAT  
 TCATTTATCTGTTTTTGGTTATTATGAGTAACACTACTATGAACATTTG  
 CACAAATTTTGTATTGACATGTTTTCATTTCTCCTGGGTATAGTCCTAT  
 GAGTGGAATTGCTGGGTCATATAATAAATAACTGTTTAACATTTTGGGGA  
 GCTGCCAAACTTTTAAACCTTGGGTTCTGTGATGTACCAGTTGTGTTAG  
 GCA(SEQIDNO:26)

**Праймеры:**

F: TATTTAGAAACCATAAAATCCACCTAT (SEQ ID NO:27)

R: TGCCTAACACAACCTGGTACATC (SEQ IDNO:28)

**>SG10S428**

TGCCAGGGGTTTTATGGTTAATTTTCCTCCATTATGAGGGTTGACTCAGC  
 CTTGGGTATTAGATGTCTTTGAGAATCCAGGGTTCAAATACCACAGCTGG  
 TAGAATGTTTCTCAACTTGGAGCCAATCTCCATCTACTGAAGGTACGCTG  
 GTTTAGACAGACAACAGGGACATCAGCATTTTAAAAAGCGGTGGAAAAAG  
 TTTGCTTGTCTTGATTGGAGCCATGACATTTTATTTGAAATTTCAAATA  
 ACATGAAGGGAGGTTTGGAGCGGTTTTTGGTTTATCCAAAGGGCAGTGGA  
 TTGAAGGCTGAGAAACACCAGGCTGAATGGGAGAGGGGTGGGGTCCCCC  
 TGTGAGATAGTGAAACAATGGTAGTGCCATCCAATGATAGGCACTTTTCT  
 GTCATTCAGAAGCAGAAAGGGGGCCAGAGGCCCATTGGCCTTACTGGG  
 М  
 AGTAAGCTGTAGAGCTGCTGCCTTTTCGTGAAAGGGTTGACACCAACCTT  
 CTCCTCCAGGAAGAGTGACCAGGGACCTGAGGGGCATGGTCGAGCAGATG  
 ACAGCCTTTGTAAACATCTCC (SEQ ID NO:29)

**Праймеры:**

F: TGCCAGGGGTTTTATGGTTA (SEQ ID NO:30)

R: GGAGATGTTTTACAAAGGCTGTC (SEQ ID NO:31)

**>SG10S422**

TTGGTAGAGATGGGGTCTCCTAGGCTGGTCTTGAACCTCCTGG  
 R  
 CTCAGCAATCTTCCTGCCTCAGCCTTCCAAAGTACTGGGATTACTGGCG

TGGGCCACCATGCCTGGCTTGAAATTTTCTATGGCTTTATTCTTTCTCC  
 AAGTACAGAGTCTACCCAACCTTCTGAGATCTTTGGTTTTCTTTCTAG  
 GTAACSTATAGTACATACTTATTTATGTTAAACAACAGCAATCACACATTT  
 CTTTTCTATACAGTCATGCTTTATAGGCAAATAAAGCCTCCGTCTTAGG  
 CTTTCTGGATTTTTTCAAAAAGATGCAATTCCTGGAGTATGTTTTACTTA  
 GAGCAAAGCAGCCTAGTCTCCTATACCTTCTGCATCTGCAGAAAAGTTGG  
 TTAACAGACTTTGTAATGATGCCCCTTACAATTCTGAAGGGACTTGTGA  
 AATAGTTTCACAGAGTTTCAGTGTTAGGTATATTTGATCAATGCTAATT  
 TTGGAACCTTTGGTGCCTGTATGATTCAGAGGGTAGGGCAGAATATTAA  
 ATTAATCACAACCTTCTTGATTTTAACCATTCTGGGTAAATTGGGATTCC  
 GTGACGCCCAGGCAAAATTAT (SEQ ID NO:32)

**Праймеры:**

F: TTGGTAGAGATGGGGTCTCC (SEQ ID NO:33)

R: ATAATTTGCCTGGGCGTCA (SEQ ID NO:34)

>SG10S427

TATCTTATATCCCCTCCAAGCATTCATTAACCTGATGGATTAGTGAGTTGG  
 CCTTGAGAAGCATAAAGGCTCGTCTCCATGTGCTTCTAAGCATTGTGTCT  
 AAGTTCTGTTTGGTTTCCTGAGTGAACTGTCTTAATGTTACCAACAGAA  
 GTTAAATGCCTAAGAG

W

TTCTTATACATGGGCTGAGTACCTCTGTGACTGGGCAAGCCACCTCACCT  
 CATTTTACCTTGTCTGCAAAATGAGGAACTGGGTCAACTCATCGTTCAAA  
 TCTCACTGAAAGCTAATTGATCGCTTTTGACAGAAGTAGCTCCCTTGGGC  
 CGTATATTTATTTCTAGCTTGGAGGAAGGTGGGACAGACAGAATTGAT  
 GTACACCTTTATTTTATCTCTATGGTAAACCTGTGCATACTAAAGCATT  
 CCTCTGGTCTTTGAGATGAGTGTATACATTGTGTCTGGCCCTGTGCATT  
 TTTTACCAAGAAGTAAGTTTGTGAGTAACTTGGGTTGTATGAAGAAC  
 TGCATGCTCACCCTACTCAAGTAGCTTTTGCTACCTAAAGGACAGCTGCT  
 CATATGTAAGTACTTCCCTTTAAAGTGAAGGATGATGACATTGAAAAAC  
 GGAGGTTGAAAAGGAG (SEQ ID NO:35)

**Праймеры:**

F: TATCTTATATCCCCTCCAAGCATTC (SEQ ID NO:36)

R: CTCCTTTTCAACCTCCGTTTT (SEQ ID NO:37)

>SG10S408

TTGAGCATGTGTTATTTAATGAGTTATACCTCTGTCATATGTGTGTGTTT  
 ATATCACAATAAATACTTATTTTATAAAACCATATTTTGAGTCATCATTT  
 GTGACAATGTCTTCTTTTCTCTGGTATAAATGAGGCATGTAGAAAGAAGA  
 TTGACATTTGCTAGAAAGCTTCCCCTTTCTCTAACTCCACAATAAAATGG  
 ATGCTCATAATTACATCTGCTCCTATAAGGTCAAGATTTTCAGGGCTGGAA  
 GTGACCTTAGATCATTAGGCCCAACTTGCCCTCAGGAAAGGAAACTGAG  
 GCCCAGAGATGCCTTAAGTGAATTGCCCAATGTCACACGCTGAGTCAGTG  
 GCCAGAGCAAGGCTTGGATCCAGTTCTCTGCTCCCCTTTCCAGAGCCTTGT  
 GATGTCTTCTCTCCTACAGGAGGTGAAAATAACTGCTGTGGCTGGTTCTG  
 TTTTGCTGACTGTAAATTGGGTCATGGTCAGGGACAGTGCATAGGTGTAA  
 AGAAGTTGCTGGTTGGGGGTTCTAATGCAGGTTTCTCCAAAAGTGAATGC  
 CCTGTAAAAAAAATTCTTAACAAATATACAGAGATTTTTTTTTT

W

AAAAAAGTGTGACAGTTCTAGACACCTAGAGAGTAAAGTGAAGAAGCCTG  
 TTTTCAGGTTTCCCGCCTCCCTGAATTTCCAGCATGGTCCAGGCTTTGA  
 AATTTATTTATCTGCTTTTGGCAATGGTTGATGGGAATTTCCACATTTA  
 TTTTTAGCTACAGAGAAAGGACATTATCTTTAAATCTCTTCGTTGTTT  
 TCTCTCTTTGAGTGAGGAGAGAAGATGTGAATCCTGGCAGTGGTTCAGAG  
 TGGACACAGCCCCTGTGTTTGTGGCATAAGGCTCTGTGGGCCCCATGCCAG  
 GGAGCAGTACCCCCGTGTAAAGGAGTGGGGGTTTGTCCATTTGGATAGAG  
 CAAAGATCCTCCACCTCAAATCCCACAAGAACAGTTGCCACAACCTGGGC  
 CCTAAGCATCTCATTTTCTATGTAGAAATTAATGATCTGGAGGAGATGG  
 CAAAACATTCCTTCCAGAGCCTGTGTGGATTTTGG (SEQ ID NO:38)

**Праймеры:**

F: TTGAGCATGTGTTATTTAATGAGTTA (SEQ ID NO:39)

R: CCAAAATCCACACAGGCTCT (SEQ ID NO:40)

>SG10S409

TAGTGCTCAGTATTTCCAACGTTCTGTTTATTTAAGATGAAAATTGCTGT  
 AGTTAATAAGCACTTCCCCATGTCATTTAAATGCTTAAGGATTTTAAATG  
 ACCACATAACAGTCCATAATATGATTAAACCCCAATTTACTGAATCAATG

CCATATTGTTGGGTCTTTAGATTGTCTCCTTTTGTCTGCTACTGTGAA  
 TGATCCTGTGATGATCATCTTTGTGTGTAAATCTTTGTCCCCTCGCCCC  
 TCCCCTTTTATTATTTTCTTGGGATAGACCCAGGACAAAAGGTAGAAAA  
 GAACAAAGTTTAAA

M

AAATTCTTGATACATAGCCACAGATTATTTTCCTGAAAGTTCTCAACATT  
 TATAACTACGAGCAGTATGTAAGAGAGTTATGGTTGGAATGATTTTAATG  
 TCTCTGGGGAATTTAACAACAAAAAACTTTAGGCTTCTTTGGAGAGAGA  
 CATGCCCTTAACCTCACCCCGCCCTAGAACAGAGACCCAGCCCATCCAAG  
 TCAGCCTCCCCAGGTCTCCACCTTCAAAACAGGCAAACGAAATCATTTT  
 TTGAATAATTGGTAGGCTTCAAGGTCAGATGTT (SEQ ID NO:41)

**Праймеры:**

F: TAGTGCTCAGTATTTCCAACGTTCT (SEQ ID NO:42)

R: AACATCTGACCTTGAAGCCTACC (SEQ ID NO:43)

>SG10S406

TAGTGCTCAGTATTTCCAACGTTCTGTTTATTTAAGATGAAAATTGCTGT  
 AGTTAATAAGCACTTCCCCATGTCATTAATAATGCTTAAGGATTTTAAATG  
 ACCACATAACAGTCCATAATATGATTAAACCCCAATTTACTGAATCAATG  
 CCATATTGTTGGGTCTTTAGATTGTCTCCTTTTGTCTGCTACTGTGAA  
 TGATCCTGTGATGATCATCTTTGTGTGTAAATCTTTGTCCCCTCGCCCC  
 TCCCCTTTTATTATTTTCTTGGGATAGACCCAGGACAAAAGGTAGAAAA  
 GAACAAAGTTTAAAAAATTTCTTGATACATAGCCACAGATTATTTTCCT  
 GAAAGTTCT

S

AACATTTATAACTACGAGCAGTATGTAAGAGAGTTATGGTTGGAATGATT  
 TTAATGTCTCTGGGGAATTTAACAACAAAAAACTTTAGGCTTCTTTGGA  
 GAGAGACATGCCCTTAACCTCACCCCGCCCTAGAACAGAGACCCAGCCCA  
 TCCAAGTCAGCCTCCCCAGGTCTCCACCTTCAAAACAGGCAAACGAAAT  
 CATTTCTGAATAATTGGTAGGCTTCAAGGTCAGATGTT (SEQ ID NO:44)

**Праймеры:**

F: TAGTGCTCAGTATTTCCAACGTTCT (SEQ ID NO:42)

R: AACATCTGACCTTGAAGCCTACC (SEQ ID NO:43)

>SG10S407

TGCTATGTCCAGTTTACACATAAGGATGTGCAAATCCAGCAGGTTAGCTG

AGCTGCCCAGGAATATCCAGGCAAGAATGACCATATTCTGATAATTACTC  
 AGGCCTCTGCCTCATCTCCGCTG  
 S  
 CCCCCCGCCCCCTGACTCTCTTCTGAGTGCCAGATTGAGCCTCCATTGGA  
 ATGCCAAATAGACAGGAAATTAGCATGCCAGAAATCCACGTCTTTAGTGC  
 ACTCTCTCCCCAGCTCCAAACCTGTTACTGCTTGTGTTCAACATCTCAGT  
 AAAGCTCAACAACATCGACCCATTACTTAGGCCTCAAACCTTGGGTGGCA  
 TCGTCGATTGCTCTTTCTTTTCATACCCACATTCAACCCATCAGCCCAT  
 CCCACAGGCCCAAGTGTGCTCTCTACCTTCAAAGCGTGTGTGGCATCC  
 ACCGCTTATCACCACCTCTGCCATTACCACTGGAGTCCAGTGCCATCATC  
 TCTACTTGGATGTGGCCAGAGTGTCTTTGCTGGTCTCCTTCTCTCTCC  
 TACCTTTGTAACAGCCTATCATCTATCTCTGGTCTCCATAGCTCACTCCC  
 ATACTTTGAGAGGGCCTTTGAAAGCCTTAGACAGATCATATCACAGACCT  
 CTATACTGAAAGTCGGG (SEQ ID NO:45)

**Примеры:**

F: TGCTATGTCCAGTTTACACATAAGG (SEQ ID NO:46)

R: CCCGACTTTCAGTATAGAGGTCTG (SEQ ID NO:47)

>DG10S2164

CCATCTGTGGAGCAGAGTCACTGAAAGGAAATACTGGAAATACTGGAAGC  
 CACTTGGTGTTTTATCAAGGATGTGAGGTTTCCTGGCAACTTTGTGCGCA  
 TATCATCATCATCATCACCATCATCATCATCATCATCATCATCATCATCA  
 TCATCATCATCATCATCTGCCCTTTAAGTTTCTGCTTGTTTAGAAAAGA  
 AATTTATACAGAGCCCCCAGTAGCAGCTGTAAGGGGGCAGGTTCTTGGAG  
 CAGCCCATCTCAACATTCTTGCTGCTGATGGAA (SEQ ID NO:48)

**Примеры:**

F: CCATCTGTGGAGCAGAGTCA (SEQ ID NO:49)

R: TTCCATCAGCAGCAAGAATG (SEQ ID NO:50)

> DG10S479

TCCACGCAGAGAGGATCTAAATCTGGCTCTTTGCAATTGCCTTCATACAT  
 GTGCATACACACCACACACACACACACACACACACACACACACACACA  
 CAGACACATACATATGCACACACCCCGACTCAATGGAGGACCCTC  
 (SEQ ID NO: 51)

**Примеры:**

F: TCCACGCAGAGAGGATCTAAA (SEQ ID NO:52)

R: GAGGGTCTCCATTGAGTCG (SEQ ID NO:53)

Для дальнейшего исследования возможности того, что другие маркерные аллели в LD блоке экзона 4 TCF7L2 проявляют более высокую корреляцию с диабетом II типа, чем аллель X, использовали данные DG10S478 генотипа, полученные в HarMap CELT образцах. Пять SNP из HarMap фаза I были с более сильной корреляцией DG10S478, в порядке уменьшения, rs12255372 ( $r^2=0,95$ ), rs7903146 ( $r^2=0,78$ ), rs7901695 ( $r^2=0,61$ ), rs 11196205 ( $r^2=0,43$ ), и rs7895340 ( $r^2=0,42$ ). Генотипировали эти SNP в трех группах и корреляции между пятью SNP и DG10S478, причем последний рассматривали как биаллельный маркер, были очень похожи на те, что наблюдались в CEU образцах. Все пять SNP продемонстрировали ассоциацию с диабетом II типа. В то время как некоторые SNP показали несколько большие рассчитанные относительные риски и более низкие значения p в одной или двух группах, ни один из них не продемонстрировал более сильной ассоциации с диабетом II типа, чем DG10S478, когда результаты для всех трех групп были объединены с применением модели Мантиеля-Хензеля (Mantel-Haenszel). Однако, хотя rs11196205 и rs7895340 отчетливо имели более слабую ассоциацию с диабетом II типа по сравнению с аллелем X (RR=1,56,  $P=4,7 \times 10^{-18}$ ), сила ассоциации с диабетом II типа аллеля Т для rs12255372 (RR=1,52,  $P=2,5 \times 10^{-16}$ ) и аллеля Т для rs7903146 (RR=1,54,  $P=2,1 \times 10^{-17}$ ) являются сравнимыми.

Согласно последнему выпуску HarMap, фаза II в октябре 2005 г., были идентифицированы два дополнительных SNP, которые демонстрируют сильную корреляцию с микросателлитом DG10S478 - rs12243326 ( $r^2=0,961$ ) и rs4506565 ( $r^2=0,716$ ). Аллели, ассоциированные с восприимчивостью к диабету II типа, будут С для rs12243326 (C/T SNP) и Т для rs4506565 (A/T SNP).

Необходимо отметить, что среди тех гаплотипов, которые несут С аллель rs7903146, те, что несут аллель А rs10885406, имеют рассчитанный относительный риск 1,06 по сравнению с индивидами, несущими G аллель rs10885406, но различия не являются статистически значимыми ( $P=0,22$ ).

В попытке повторить и детализировать эту ассоциацию с диабетом II типа, генотипировали DG10S478, rs12255372 и rs7903146 в дополнительной большой датской группе, состоящей из 1111 больных и 2315 контрольных индивидуумов, а также в более генетически разнородной группе из Западной Африки, состоящей из 618 больных и 434 контрольных пациентов, набранных из Афро-Американского исследования сахарного диабета (23). У датчан все три варианта были строго ассоциированы риском заболевания, как это предварительно наблюдалось в Исландии. Однако ассоциация аллеля T rs7903146 ( $RR=1,53$ ,  $P=4,06 \times 10^{-14}$ ,  $PAR=24,4\%$ ) была значительно сильнее, чем того, что обеспечивали два других варианта. В Западно-Африканской группе после введения поправок на родственные связи и этническое происхождение, повторно наблюдали ассоциацию аллеля T rs7903146 с диабетом II типа ( $RR=1,45$ , 95% дов.инт. = 1,20-1,76,  $P=0,000146$ ,  $PAR=22,2\%$ ), но не наблюдали ее в случае других двух вариантов. Это позволяет предполагать, что аллель T rs7903146 является либо вариантом риска сам по себе, либо наиболее близко из известных коррелирует с неидентифицированным вариантом риска. Исключение маркеров DG10S478 и rs12255372 из маркеров риска в Западно-Африканской группе было возможным, поскольку в популяции европейского происхождения, где T аллель rs7903146 имеется почти исключительно в хромосомах, несущих как аллель X DG10S478, так и аллель T rs12255372, у индивидуумов из Западной Африки T аллель rs7903146 имеется с обоими аллелями DG10S478 и rs12255372. Это согласуется с наблюдением, что T является исходным аллелем rs7903146, в то время как аллель X DG10S478 и аллель T rs12255372 отличаются от контрольной последовательности шимпанзе. Если говорить более широко, то это открытие согласуется с ожиданием, что относительно разнообразные популяции, такие как популяция Западной Африки, обеспечивают возможность для очищения сигналов ассоциации, обнаруженных в областях сильного LD в более гомогенных популяциях.

#### Обсуждение.

В этом исследовании описывается идентификация нового гена-кандидата для диабета II типа в пределах предварительно обнаруженной области взаимодействия в 10q (10), кодирующего транскрипционный фактор 2, подобный 7 (TCF7L2, бывший TCF4) на 10q25,2. Показано, что он придает риск развития диабета II типа в Исландии, Дании и США с похожей частотой и относительными рисками. В то время как этот вариант не объясняет существенной части семейного диабета II типа, приданный популяции риск по меньшей мере 20% является существенным с точки зрения здравоохранения. По сравнению с индивидуумами, не являющимися носителями, относительные риски гетерозиготных носителей составного аллеля риска (приблизительно 38% популяции) и гомозиготных носителей (около 7% популяции) составляют 1,45 и 2,41 соответственно. Следовательно, этот вариант имеет достаточное предсказательное значение, чтобы быть клинически применимым.

Авторы изобретения сообщают о варианте, как ассоциированном с диабетом II типа микросателлите, DG10S478, находящемся в третьем интроне гена TCF7L2. Продукт гена TCF7L2 представляет собой транскрипционный фактор, содержащий бокс группы высокой подвижности (HMG), который играет важную роль в Wnt сигнальном каскаде. Этот каскад рассматривается как один из ключевых в механизмах развития и регуляции роста клетки; он опосредован секретируемыми гликопротеинами, известными как Wnt, которые инициируют многие сигнальные каскады в клетках-мишенях при связывании с родственным рецепторным комплексом, состоящим из члена семейства Frizzled и члена семейства LDL рецепторов, Lrp5/6 (24). Wnt каскад обеспечивает несопряженность центрального игрока каскада,  $\beta$ -катенина за счет деградации комплекса и транслокации его в ядро, где он временно превращает TCF факторы из репрессоров в активаторы транскрипции (25). Белок  $\beta$ -катенин является также важным для клеточной адгезии, благодаря его связыванию с кадгеринами (15).

NCBI RefSeq для TCF7L2 включает 10 экзонов. Однако Дюваль и соавт. (Duval et al.) (26) показали, что TCF7L2 имеет 17 экзонов, 5 из которых являются альтернативными, кроме того, дополнительно сообщалось, что используются три акцепторных участка альтернативного сплайсинга. Это исследование также продемонстрировало альтернативное использование трех следующих друг за другом экзонов, расположенных на 3' конце гена TCF7L2, которое изменяет рамку считывания, используемую в последнем экзоне, приводя к синтезу большого количества изоформ TCF7L2 с короткими, средними или длинными C-концами.

Подобно TCF7L2, пять из шести позиционно клонированных генов для редких Менделевых форм диабета II типа, а именно диабета зрелого возраста у молодых (MODY), являются генами транскрипционных факторов (27). Дополнительные транскрипционные факторы вовлечены в патогенез диабета II типа, включая активированный пролифератором пероксисом рецептор гамма (PPAR $\gamma$ ) (7) и ген семейства forkhead (28, 29). Noble et al. описали миссенс-мутацию (C883A) в родственном TCF7 гене при диабете I типа (30). Однако не ясно, работают ли TCF7 и TCF7L2 в одном каскаде по отношению к патогенезу диабета.

Были описаны мутации в гене TCF7L2, включая делецию A в (A)<sub>9</sub> кодирующем повторе (экзон 17) (26, 31-33), и ряд мутаций в колоректальных клеточных линиях (26). DG10S478 находится в хорошо определенном LD блоке в 74,9 т.п.о. (CEPH НаpMap для белой расы, фаза II), который включает экзон 4 и фланкирующие интронные последовательности, расположенные 5' и 3' по отношению к экзону. Возмож-

но, что DG10S478 является сам по себе причиной; возможно также, что DG10S478 является заместителем для основного варианта, который влияет на транскрипцию, сплайсинг или стабильность мРНК. Вероятно, такой вариант находится в сильной LD с DG10S478, т.е. вариант сохраняется в пределах LD экзона 4 LD TCF7L2.

Несколько линий доказательств позволяют предполагать энтероэндокринную роль этого гена в патогенезе диабета II типа. Во-первых, TCF7L2 вовлечен в развитие колоректального рака (34) и уже описаны малые молекулы-антагонисты онкогенного белкового комплекса TCF/  $\beta$ -катенин (35). Кроме того, у TCF7L2-I-мышей, которые умирают в течение 24 ч после рождения, отсутствует компартмент стволовых клеток эпителия кишечника (36). Варианты TCF7L2 гена могли бы влиять на восприимчивость к диабету II типа через изменение уровней инсулиноподобного гормона - глюкагонподобного пептида 1 (GLP-1), одного из пептидов, кодируемого геном проглюкагона, экспрессия которого в энтероэндокринных клетках на уровне транскрипции регулируется TCF7L2. Согласованно с инсулином GLP-1 оказывает решающее действие на гомеостаз глюкозы в крови (12). Аналоги GLP-1 и ингибиторы дипептидилпептидазы IV в настоящее время находятся в стадии клинической разработки.

Хотя это изобретение было конкретно продемонстрировано и описано со ссылками на его предпочтительные воплощения, специалисту будет понятно, что могут быть проведены различные изменения в форме и деталях без отклонения от рамок изобретения, охватываемых приложенной формулой изобретения.

#### Ссылки

1. A.F. Amos, D.J. McCarty, P. Zimmet, *Diabet Med.* 14 Suppl. 5, SI (1997).
2. P. Zimmet et al., *Am. J. Epidemiol.* 118, 673 (Nov., 1983).
3. W.C. Knowler, D.J. Pettitt, M.F. Saad, P.H. Bennett, *Diabetes Metab. Rev.* 6, 1 (Feb., 1990).
4. B. Newman et al., *Diabetologia* 30, 763 (Oct., 1987).
5. A.H. Barnett, C. Eff, R.D. Leslie, D.A. Pyke, *Diabetologia* 20, 87 (Feb., 1981).
6. A.L. Gloyn, *Ageing Res. Rev.* 2, 111 (Apr., 2003).
7. D. Altshuler et al., *Nat. Genet.* 26, 76 (Sep., 2000).
8. A.L. Gloyn et al., *Diabetes* 52, 568 (Feb., 2003).
9. Y. Horikawa et al., *Nat. Genet.* 26, 163 (Oct., 2000).
10. I. Reynisdottir et al., *Am. J. Hum. Genet.* 73, 323 (Aug., 2003).
11. R. Duggirala et al., *Am. J. Hum. Genet.* 64, 1127 (Apr., 1999).
12. F. Yi, P.L. Brubaker, T. Jin, *J. Biol. Chem.* 280, 1457 (Jan. 14, 2005).
13. S.E. Ross et al., *Science*. 289, 950 (Aug. 11, 2000).
14. E.A. Jansson et al., *Proc. Natl. Acad. Sci. USA*. 102, 1460 (Feb. 1, 2005).
15. W.J. Nelson, R. Nusse, *Science*. 303, 1483 (Mar. 5, 2004).
16. C.T. Falk, P. Rubinstein, *Ann. Hum. Genet.* 51 (Pt. 3), 227 (Jul., 1987).
17. J.D. Terwilliger, J. Ott, *Hum. Hered.* 42, 337 (1992).
18. J.R. Gulcher, K. Kristjansson, H. Gudbjartsson, K. Stefansson, *Eur. J. Hum. Genet.* 8, 739 (Oct., 2000).
19. Y.Z.R. Bagger, B.J. Alexandersen, P. Tanko, L.B.; Christiansen, C., *J. Bone Miner. Res. Suppl.* 1, 1 (2001).
20. G. Benson, *Nucleic Acids Res.* 27, 573 (Jan. 15, 1999).
21. R.C. Lewontin, *Genetics* 50, 757 (Oct., 1964).
22. W.G. Hill, A. Robertson, *Genetics* 60, 615 (Nov., 1968).
23. C.N. Rotimi et al., *Ann. Epidemiol.* 11, 51 (Jan., 2001).
24. C. Prunier, B.A. Hecovar, P.H. Howe, *Growth Factors*. 22, 141 (Sep., 2004).
25. J. Huelsken, W. Birchmeier, *Curr. Opin. Genet. Dev.* 11, 547 (Oct., 2001).
26. A. Duval et al., *Cancer Res.* 60, 3872 (Jul. 15, 2000).
27. S.S. Fajans, G.I. Bell, K.S. Polonsky, *N Engl. J. Med.* 345, 971 (Sep. 27, 2001).
28. C. Wolfram, E. Asilmaz, E. Luca, J.M. Friedman, M. Stoffel, *Nature*. 432, 1027 (Dec. 23, 2004).
29. J. Nakae et al., *Nat. Genet.* 32, 245 (Oct., 2002).
30. J.A. Noble et al., *Diabetes* 52, 1579 (Jun., 2003).
31. A. Duval et al., *Cancer Res.* 59, 4213 (Sep. 1, 1999).
32. A. Duval et al., *Oncogene* 18, 6806 (Nov. 18, 1999).
33. H.R. Chang et al., *Cancer Lett* (May 16, 2005).
34. N.A. Wong, M. Pignatelli, *Am. J. Pathol.* 160, 389 (Feb., 2002).
35. M. Lepourcelet et al., *Cancer Cell* 5, 91 (Jan., 2004).
36. V. Korinek et al., *Nat. Genet.* 19, 379 (Aug., 1998).

## Список последовательностей

<110> DECODE GENETICS BHF.  
GRANT, Struan F. A.

<120> GENETIC VARIANTS IN THE TCF7L2 GENE AS DIAGNOSTIC MARKERS FOR  
RISK OF TYPE 2 DIABETES MELLITUS

<130> 2069-PC

<150> US 60/692,174  
<151> 2005-06-20

<150> US 60/757,155  
<151> 2006-01-06

<160> 53

<170> PatentIn version 3.3

<210> 1  
<211> 74930  
<212> DNA  
<213> Homo sapiens

<400> 1  
ctgtgttagg aactcaogct ttgtttattc agcaatcatt cctccagaaa taaccttaat 60  
agcaacaaga aaaaagaata ggtgtttttt gagctctatc tgccagtttc tctatatatg 120  
gacattatat attgcaacat aacactcaca atgcctttaa acatcatccc cgttatacag 180  
ataagaaaac agaatttcaa agaaggtagg ggacttgccc agggatacat agctagcaag 240  
tggcagcgct ggattgagtc tgggccttgt ctgaggctcg ggtcctgtca tgctctcgcg 300  
ttgctatggt gacatgcaaa gggagaggca gctgctggga gtctaggtgg gtttctcttt 360  
gagaatgcta acgtgaaccc tcaaggtgaa tcagaatcct ttgcaagtg aataatcaga 420  
tgtagggtcc tgtgtctccc tgtaaaatga aagcctcttt tttccaaggt coagtataga 480  
cctgaagctg ggttactctg gaatttcctt ctctggctgg agtgactgag gccttgcaag 540  
tgacattggt gaggactcgc agcctcaggt ctggcttccc ttagcaaccc ccttttctctg 600  
tctctgcctc tggagttcac cattaaaaa aaaaaaagaa aaaaagccaa aacactttat 660  
aaagttacat gctgggttct ttctatgtcc tagaaactgt ottaattcat cttccctttt 720  
actcttatat gagcaggaag aaaaaaaaat tgctagtcaa tgctaataat tatggcatgt 780  
aatgtaattg gaagtgttct actgacatgc tcatgagagt ttgcggcttc atcttcaggc 840  
tgggatgtag cactagactt gccttgagtg totgcacaag cctttgatgc aggttagacca 900  
tattataaat aggcgcgttg ctatggtgag gatggcagtc cttgcttgct gtgggtaacc 960  
ttttctacct tctcggaacac tgttttaaaa cacagcagcg tgatagcatt tcatttaatt 1020

tggaccaagg tggggtagat gaaatgttga gatttagatc taaaatgttg ttgtggtgtt	1080
tcaggggggtt ctggctcacc tagtactatg gaagattttg cagattgggc ttctcatga	1140
tttattttaga aatagatttt ctaatagatg gggtagggg agggtaggtg gcagaaggct	1200
gggttttctt ctcttcccc tctctcttct attgagcgt tctgcgaatg tgttggttt	1260
gatgccccag gagctcatac agtgaaatgg aagttcaggt tggcacgttg cagaaatgat	1320
tattctcgtt agtacgttct ccattactgt taataatata aagacaattg cctgcctctc	1380
aggactcctg cacgtggcta cagtcatttc ttcatggaat tagacacata gcagtgggga	1440
ccaggagtgt tttatttagt attgtctctc tgcaagtctt cagggtatct cagcttagac	1500
acatgaatta ttttttctg ttgcttgag ggtatacttt taattatatt cattcaataa	1560
cagagcagtt caggtttcta aaatattttt tctcccccac ctttttcccc agcatacatc	1620
cccgccccgt aagtttcttg gcagagacaa tctcaggaac ctaaagggtg ctaaaaaatt	1680
agctagtgtg ccaggcgcat gactcatgcc agtaatccca gcactttggg aggctgaggt	1740
gggtggatcg cttgagccca gaaattcag accagcctag acaacatggc aaaacccgt	1800
ctctacaaac aaaacaaaat ctgctgggc atggtaggtg atgcctgtag tcccagctac	1860
tggggaggct gaggtggcg ggcgattgag ctcaggaggt ccaggctgca gtgagccgtg	1920
attgtgccac tgcactgcag cctggatgac tgagtgggac cctgtctcaa taataaataa	1980
ataaataaat aaaaaataa aaaaattagc tagccaagct gcttataggt cttttacatg	2040
gccaagccac tttctcact ttaaaatggt aataacgttt ccgtactcat ctcaatgggt	2100
tttgagtgc aagacagacc gtttgatgga agccctcttg ggagaaaaat gctaccaag	2160
acaggctttt caattggaga ctgatccatt ggtgttttgg tcagttggtg ttgaaatccc	2220
tatttttcca gctcaggact gcctctctcc ctggaactct tcccgaggtg agttctgcag	2280
ccttcttgga gaactctcag cctctggatc cttctctgcc aggtggagtg gacatgccaa	2340
agttgtgggc cagactcgga ctgcctggct tgtctcagca ctttgggga cccacttccc	2400
ctctctggga actggggaag ctaacagaga tottgctagg ggggtggaat cctgtatcca	2460
tgtgaggttg taccgccagg ctctgagtg gtttgaagt ggggaacctt ggccggggcg	2520
gggtggctcat gcctataatc ccagcacttt gggaggctga ggcgggggga tcacaaggtc	2580
aggagatcga aaccatcctg gctaacacga tgaaaacctt tctctatgtg cgtggtggct	2640
ggcacctgta gtcccagctg ctggggagtc tgaggcagga gaatggcgtg aaccggggag	2700
goggagcttg cagtgagcgg agatcgcccc actgcactcc agcctggggc acagagcgag	2760

actccatctc aaaaaaaaaa agaaagaaaa aaaaagaaag tggggaaccc ctccccagg 2820  
 atgagaagag ccatggggtg agtctctgcc accgccaagg ggagtcaggc tcagaggctg 2880  
 ctacagggac agccagctct ctttagatgg tccccacat ctagtcaggc cttgttacat 2940  
 atggagcaga gacagcgag gctgctgctg ttttcctgga gaaggccct gtcggctgtg 3000  
 tcagctgtag ctgaccttc ctcttctgctg tttttggga gggagccttg gaaggagtag 3060  
 ggcaagtgg gcactctgct tcccgcccc aactggcga acctatggat tctgcctctg 3120  
 attcctgagg aaacatcact gtgaagggtg aatgagccac atacagagg ggctgttggg 3180  
 gcgggggagg ggtgaaacgc cccaggggtg tacattgcac caaagccag gctgcatata 3240  
 gacctcagga tgggctggct tttctattta tttagaagta tttccaggag gtaacctcat 3300  
 tggctacaaa gcatgtctga acaagagctc cgttgttcat tcccagcct gttaccttg 3360  
 caggatgcag actccaggcg gcctgttggg caggccttgg actcagagag cagtgaagcc 3420  
 tgaggagggg tggggggcag aggcgtgagt ggtctagggc ctcagtccct ccaggacacc 3480  
 ccttgccaag cgcagagaaa gctctgccca tccgtccct caggcagtg gattgggcaa 3540  
 cctgggaagc agtgaatgtg cgtcggtagc atagattcca ttcgcacgc caccctcgcc 3600  
 tcgcgcccc agccctggga gggatgcatg cctccggga gacaccaga cccgacagag 3660  
 aggcctttgt tggagctgga ggtgagaatc tgtggcgctt gggattcctg ggttcagatt 3720  
 ccagctcact gccaatgcc cgagtgtg gcgaaacatt ctggaatcaa aaggagtga 3780  
 gcctgcccag cagggcctac gggagcggga ggctgcaggg tgctaagatt gcgttatctt 3840  
 taccaagtgc ccggagctcc tgggagggaa gagagagtc taggactcag gataggagg 3900  
 ggttgagatt tctcaggaa gactccatgc tttggttctg gccctggaa accctctctg 3960  
 aggactggac ctccaagcag accccctctg tgactcggga atgcagtggt actctcttat 4020  
 atttttcttt cttttttttt ttttgagacg gactctcact ctgtcaccca ggtggagtg 4080  
 cagtggcacg atctcggctc actgcaacct ccgcctccg agttcaagcg attctctgc 4140  
 ctcagcctcc caagtagctg ggattacagg tgctgacac cgcgcctggc taattttttg 4200  
 tatttttagt agagatgggg ttttaccatc ttggccaggc tggctttgaa ctctgacct 4260  
 cataatccac ccgcctcggc ctccgaaagt gctgggatca caggcgtgag ccaccgcacc 4320  
 cggccactgt cttgtatttc taactgccc ctgacttttc tgatcatgta attcttaact 4380  
 ttctcaaaac tgagatttgt cacgtgtcct ctcccactc cattttgtga atcagagct 4440  
 tccaggggca ggacctggag aatgggtctt tattaacaca catgtgaaaa tgcttttgcc 4500  
 agcaaggcgc ggtggctcat gcatgtaac ccggcacttt gggaggccga ggcaggcga 4560

tcacttgagg	tcagcctggc	caacatggta	aaaccctgtc	tctactaaaa	atacaaaaat	4620
tagctgggtg	tggctgtggg	caoctgtagt	cccacctact	cgagaggctg	aggcatgaga	4680
atcactggaa	cccaggaggt	agaagttgca	gtgagccgag	atcacaccac	tggactccag	4740
cttgggtgat	agagtgagac	tctgtctcaa	aaaaagaaaa	aaaaagaaa	atgcttttgc	4800
catgggtgtg	ctcctgcttc	tgctttgcat	tgggcctctg	tacctaggtt	gcaagattcc	4860
tcagggtgca	cctgggctta	tcgttatctg	taagttatcc	cagcaagcac	ttaaaacaca	4920
gtgttggacg	atgaatcccc	tctacaagag	agggacaggg	caaaaacgac	acctcttgcc	4980
tgcgaagctg	tcttggggcca	aacctcaggt	ctattctttc	ttttttttga	aagtagtggc	5040
tgggcacggt	ggcttacgcc	tgtaatccta	gcactttggg	aggccaaggc	gggcggatct	5100
tgaggtcagg	agttcgagac	cagcttggcc	aacatggtaa	aactccatct	ctactaaaaa	5160
tacaaaaatt	agctgggcgt	ggctggcgcat	gcctgtagac	ccagctactc	aggaggctga	5220
ggcaggagaa	tcacttgaa	ctgagaggca	gaggttgcag	ttagctgaga	ccatgccatt	5280
gcactccago	ctgggcggca	gagcgagact	ctgtctcaaa	aaaaaaaaaa	aagaaagtag	5340
cagctctact	gagatattta	gaaaccataa	aatccacctc	tttgaggtgt	acaattgagt	5400
gattttctgt	atagtccacg	atctgtgcag	tcateccac	cctctaactc	caggacattt	5460
tcctcacccc	cgaggagaaa	cctcccttac	ccattagcag	tcactcctca	tttctctccc	5520
ccccagcccc	tggcaatcac	tgtggatttg	cctgttcttg	acatttcata	taaatggtat	5580
cataaaatct	atgggccttt	gtgtctgtct	gctttcaact	agcatacggg	tctcaaggtt	5640
catccagtat	tgtagcatct	atcagtatgt	cattcctttt	tatggccaaa	taatatttta	5700
ttgtatggat	agacattttg	tttattcatt	tatctgtttt	tggttattat	gagtaacact	5760
actatgaaca	ttttgocaaa	atttttgtat	tgacatgttt	tcattttctc	tgggtatagt	5820
cctatgagtg	gaattgctgg	gtcatataat	aaataactgt	ttaacatttt	ggggagctgc	5880
caaaacttta	aaacottggg	ttctgtgatg	taccagtgtg	gtaggcagc	acagcaaaat	5940
gtgacttttg	attgccagaa	acaatattta	aaaagtgggt	ataaaaagtg	gtttgggagg	6000
ctgaggcagg	aggatcactt	gagcccagga	gtttgagacc	agcctgggca	acatagttag	6060
accctgttaa	aaaaaaagaa	ggccaggcac	agtggctcat	gcctgtaato	ccagcacttt	6120
gggagactga	ggcgagcaga	tcacctagg	tcaggagtgc	cagaccagcc	tggccaacat	6180
ggcgaaaccc	catctctact	aaaaatacaa	aaattagcca	ggcctggttg	tgggcgcctg	6240
taatcccgag	tactcaggag	gcttgaggca	ggagaatcgc	ttgaacctgg	gagactgagg	6300

ttgcagtgag cggagatcat gccattgcac tccagcctgg gcaacaagag cgaaactgtg 6360  
 tctcaaaaaca aatgaaaaga aaaggctgtc atgttagatc caccctctct ctcaggggaa 6420  
 cccctgggct gctctctggg tagagatggg aaccocaggcc togggccagt gagtgggaagg 6480  
 aaaotttggg atgattgact tgggactggg ctagagggtga agaattctccc agtaggcaaa 6540  
 gttcggcctt acgttttttt gtttcaagca aaccacatca ttaccocag aggccattgg 6600  
 tgagatattt gtaagtctcc tgacagtggc tggagttcgt tgcctgggtt ttgtttctct 6660  
 gtctcagccc tggagatggg agtgaocacc tgcctctctc ggacagaggc tgtocacgtt 6720  
 catgcaatcc cttggacacc ggtgggtcag cgggaggcgt aactgggagt gggagaccct 6780  
 gaactgtgcc ggttcttgca gagtatcaat gtgaactcag gcgagtcacc ccacatcagg 6840  
 cagctcagaa caagggattg atctagaagg acctttcacc tgggctatcc tgtgactcaa 6900  
 attatcttct cctaagccca ctactgctg gtgtgttggg taaattagcc taaaggtcat 6960  
 tccctcggag aggcctctg ggaacacctc ctttctgag agtcaactgct tgcctggcgc 7020  
 tgccctggg gttccttcag agtcgtgatc atgcctcggc ctcttccttt atttggcagt 7080  
 cccttccctt ccccatccct gatgagggtg gggagcatct gtctgcagct tcatcttcat 7140  
 tgtctagggg ctccagaaat atctgtgagt aaataagtta tttaattctt gcctcaaatt 7200  
 tccagtgact gtagggatat agctgtgagc ctctaggagc tgagattttt taaatttccc 7260  
 acttaaacat ttatttaaaa attttgtgct cagcatggac taaggacttt acattcatta 7320  
 actcatttac agcttgatcc tatcggtgg gcattcattt acagaggatc ccattttaca 7380  
 ggtgaggaag aggccagcta ggggtgcagc ctaggttagt attctagagc tcatcaggct 7440  
 gtgttgtccc cagtgaaga ataagcaaag aagtgaatgt tgtgcattga gaaaaatgac 7500  
 tctcggagga ggtgagcct ctcggatatg gcgaccgaag tgatatgggg cccttgtcaa 7560  
 gggctctctat tatggcatca agaaaagatg ctgctttcgg tgatgccga ggagagcctc 7620  
 aatattttac atgggaaacc taaaaagggt gccatgttgt ggtctctgca cctaagatac 7680  
 taaaggaaat attttatgga gagatgaac atgtcaggcc ttggagggaa accocaggat 7740  
 ccagatggtt gcactctcaa accagggccc cctcacctt ggccttcagc atttagtgtt 7800  
 ggaaccaata gcataagctt tggtcaggac ctttgatgga agccacagtg ctcattagtg 7860  
 accacgggtg actacottct ctctcctaag ctgacttctg gagggcacct gggatttccg 7920  
 gccagtgatc agtgctggtg aagcctgaag gccaatgtgt aggttttagct gttcagtcag 7980  
 aacccaaaag gggccaaaga gatggtttcc ttcaacctcc actgagggaa gtgaaagtca 8040  
 tggttcgtta aaaggctgag ctgggaccag agtctagggt tctagagggt ggaatttcta 8100

cagctttggg ggaacctgca agggcatttg ctcttctggg actgcagga gactgtgctt 8160  
 ctcagagatg ttagcatttg gcttggggag agagaggaaa ggagagggtc atgctccgcc 8220  
 atgatggtgg aaagtgatgt tgggtggtg aggagctgag ctgaattcta agtgggtcca 8280  
 gggaattaac aatgttcctg cccaagtgtc ctgttcccc acaactaat gaggcagcag 8340  
 gtgtctgaag agaaacattg cagaatgtct gccagggtt ttatgggtaa ttttctcca 8400  
 ttatgaggtg tgactcagcc ttgggtatta gatgtcttg agaaccagg gttcaaatac 8460  
 cacagctggt agaatgttc tcaacttga gccaatctcc atctactgaa ggtacgctgg 8520  
 ttttagacaga caacagggac atcagcattt taaaagcgg tggaanaagt ttgctgtct 8580  
 tgattggagc catgacattt tttttgaaa tttcaaata catgaaggga ggttgggagc 8640  
 ggtttttggt ttatccaaag gccagtgat tgaaggctga gaaacaccag gctgaatggg 8700  
 agaggggttg ggtccccct gtgagatagt gaaacaatgg tagtgccatc caatgatagg 8760  
 cacttttctg tcattcagaa gcagaaagg gccagaggc ccattggcct tactgggcag 8820  
 taagctgtag agctgtgcc tttctgtgaa aggggtgaca ccaaccttct ccccaggaa 8880  
 gagtgaccag ggaacctgag gccatggtcg agcagatgac agcctttgta aaacatctcc 8940  
 ctggtctcat cagcgatatt cgtcctgcct tcctctgag taatttccat ottaggaotg 9000  
 gagtccagtg gagcaagatt ccattgtgtt ttctgttggg cctagagtgt cacactgaga 9060  
 cctaatttca tactttatga attctagtac tgctctcgaa ggtaagagcc gtcctcttg 9120  
 gctgaagggt tttgctgca accttgcat gtaatccagt gacacctgac gtatctgtaa 9180  
 atttcttcaa atttctaagt gtattacaac ccctgtgca aaagatgatt aattaattgc 9240  
 cttgacagta aaacaaaaa caaaaaaag gtgtgggggt atatggtatc cctgatttac 9300  
 tatagaagat gcagagagtg aaggagatg aggtggggag gaggggcca ggttctggtc 9360  
 ctactttttt ttttttttt ctaagagat ggagtcttac catgttggcc agtctaggct 9420  
 tgaactctg gctcaagag gtgctctcac ctcagcctcc caaagtgtg ggattatagg 9480  
 cgtgagccac cgagtttagc ccaggttctg tttctgtgt agtcactttc tgtttgaaca 9540  
 aaattggaat ttctttttt gatctgttcc tttaattgta aattgaatcg gactaaaacc 9600  
 tttccaattt ttccacatgt gaagacatac acaaaagttt tattggaggg ttgcacatgt 9660  
 gaaagaaaaa ggagaaaagc aggattgagc agggggagcc gtcagatggt aatgcagatg 9720  
 tgatgagatc tctgocggac caaagagaag attcctttt aaatggtgac aaattcatgg 9780  
 gctttctctg cctcaaaacc tagcacagct gttatttact gaacaattag agagctaagc 9840

acttttttaga tactatataa ttttaattgcc gtatgaggca cocttagttt tcagacgaga 9900  
 aaccacagtt acaggggaagg caagtaactt agtcaatgtc agataactag gaaaagggtta 9960  
 gagggggccct ggacacaggc ctgtgtgact gagaagcttg ggcacttcac tgotacattt 10020  
 catctcttcg ctataaacat tttagctttt tgtgtttgct gactggcaac aatacatagt 10080  
 gaaagttcta ataatttgta atgcttttgc atgtctttgt atttttcttg gttatcacat 10140  
 cacatcaaat taagatactg atcagcagtg tgagagggtta tttttccatg tctcttcat 10200  
 tagtgttagc ttgtggatgg atttgaggct ctctgtgctt tccccccagc aaagtgaata 10260  
 ccagactttc ctattaaaaa aagtatttta tttttcagag acagggcttc attctgtctc 10320  
 ccaggctgga gtgcagtggc acaatcatag cccactgcag cctccaaetc ttgggttcaa 10380  
 atgactctcc tgcctcagcc totcttaagc agtgctttc cccattctca tgggactttc 10440  
 caatccatga gatactttgc tgcaggaag ccctgtctgt ccaggcctgt gtaatagacg 10500  
 acttccatg gtctgtgtt gttgtttgcc ttctgtgtgg ctaagtttcc atgacctggt 10560  
 ggcttggaag ccccatccct gatttgtggg agaggcaggg aggcacctg tagcgacta 10620  
 ggctgtgggc ctgaacaagt ctgtgtgctt ccaatgtctt tgtggggagg ttacgagtc 10680  
 cttcttatta tataatagta tctgtctta gcttggtgcc tttcttctca gaagcttgag 10740  
 gcactctgca gataccatct caatttgctt tctgggagga ggagaggaag ctaccaaaa 10800  
 gatgaagttc tctgtgaggg gottgaacac aggttgatag cgttgctggt tagttattct 10860  
 catggtgttg atgaaaaatg gaatacgtg aaatttcagt tactcgtcac aaaaataagg 10920  
 cgtatgtaga aaacatctg ggctaagggt ttgcatgctt ctagaacttc ctgttactta 10980  
 atggctgttg agtataaaco tggggaacag tggggatcct tggagacccc aaataacttg 11040  
 tatttgtggg tactcctgto ttgtctatca ataccctgt ctatctcgtg ttagaactag 11100  
 gacacacaga ctggattcag aagctggcct ggggtttagg agaacatggg acctaatcct 11160  
 ggcatctcgc atttacctcc tggatcttgt tttctcatct gtaaaatgaa ttggggtgtg 11220  
 gactgtttat ggctgtagg atgctagccc tgagaatttt ctccagatat tctacgggtta 11280  
 agtaatttta ggggacactg tctaagcagt tgcctcttgg agaatagaaga tgttcattag 11340  
 gataattgaag gctctgagaa gtcctaaagt taaagaaaat ctgcaatggt ctttgtggga 11400  
 ccgaataatg caacctggga aatgagggat tagatgacac ttgagtagcc ttccagatct 11460  
 gagacgagtc tcaactctgtt tgtttactcc atctgtgatg ggtgtaggca ccatcttggg 11520  
 gagcaagctg tgatagagag ggaacaatac cttgttaatg tttgtctaat tcaactacca 11580  
 ggtgcatggt agtgaattag acactacttt gtaggttctg gagggaagaa gaaaagacga 11640

gacctgcctg gactggggct tgagaccact gtcaaatata agtacagtgt tacaactggt 11700  
agggagtggg toatagtatg gccggtcttt ttaaagggtga ggaattctta ggcccagaaa 11760  
ggcaaagtga cagatcctgg atttaaccag cagcccagat ttgaggccta gcacatagca 11820  
aagcaccata gctattcaat agctgccaaag tgggagtttg gatgatggct ttccctggaca 11880  
gcgaaagcag tgatgtttgc ttaggatggc ctttggcagt gctgctgtta tccttaccac 11940  
tggcaagcca totcacgggc ccggagggga gggcaaggaa tcctaattct gtgagaaggc 12000  
totgggtaca tgagtgtgag atatggatac cctaggtctc gccctgaag acagtggcat 12060  
cggatttact gcactattcc agtcggacag gcacottaat tttctcttt ctgggtgttt 12120  
gatatggttg ggtcctatct cttctctcc aaaccccgct agggccattc cccaccctt 12180  
cacttcccg ccttccactg cagtctctaa ggattctgct tcactttat gtgtgaacag 12240  
ggttttgaca aacatgatta actgggtatt tttggaaggc tcaggaggaa cgcagagtgc 12300  
tccggagggc aggcctggag tcaggaatgc ttctgcaac ctgttogtgc agtgagcgtg 12360  
tcttctctgc cctgcccttg gctggggaat gtgctggctt ggagggcagg agagtgcag 12420  
gcggtttgag aactccgggc tctccgtct tcggatggct cctgtgaaag cagggcctga 12480  
aacttttato gtcaatgctg cagggtgaaag aacttctatt ggcgtgtagt gtccaacaaa 12540  
gagtatttta tttatgtgtt tccaagccct taaaaattct tttagggcac atcagtgggg 12600  
agttaataga aactttgaaa taagaaaaat gcctgcaggg taagtagaac ccagccagc 12660  
cagctccgag ttctgtgctg ttagctggta ggttggttct cagagaagtg gctggctggc 12720  
tgggttacgg agcccacato tctaagcct tagtgttcaa tcattaagtg gatttttttt 12780  
tttccctct cttcttttg tttggaggga ggactactct aaactttact cagggcaggg 12840  
tagctcctga aagggtccc taacctttct ggtttatgac acaaagaaag tttggaggta 12900  
ctgggataag agatggcttg ggtgaccccc ctatcatgcc ccctaacaca tacacagcaa 12960  
accaaaccaa ctcacccttg atcatactcg ttgtttacac gaagggaatt tttattgtct 13020  
tgtgagtgtt gagtgatgat taaacagaag agatgtgact ccaagcctgg cttcactaag 13080  
atagtcttgt ttgtttcttt tcctccaaag taatttcta aagaattaaa agccctttg 13140  
aaaccagca ctacctgtc totgattatc agcataggca ggaagggtt ttaagggtotg 13200  
agcccagctg tttagaggct acgagacgtg aggc aaatcc tggatctct ctttgggcct 13260  
cagttttct atctgtgaaa tggcacagta ctaccctcca ccaaggatga tgatgagaat 13320  
taaätgggat gacagggttc atccccagct cctgttctta ggaaggaaaa actgtgactt 13380

atgaagcctg taggttgtgt tcaggtttgt atgaggccto ggacttcata caaaggtato 13440  
 aaagtggcaa accotgatcc agatgttttc agttcagtoa gctggtoctt gagcctgttg 13500  
 tgtccagat atcctgacca aagaagctag atgggagctg ctgtgttgtt ccttggggct 13560  
 gctggatgca agttgttttag gtcggcggtt ttcaaatgct ggtgattttg ctcaccagag 13620  
 gacatttggc aatgtctaga gacatttagc atggccagtc attgggaggt actcctggca 13680  
 tctcgtgggc agaggctaag gatgctattg aacatcctgc aatgccagag acagccccc 13740  
 gtgacaggag tcaccagcc caacatgtca ctagtgtgc agtggagaag cctggctgt 13800  
 gtgtgggggt gtgtgtgtgt ootottotac atttgataag gtaactcaca cttgctgccc 13860  
 ccatgatcgc tgtgggggat gcttatctat gccccagtc tggtgttgtt tgatgggaac 13920  
 atcaagatto aggcaagatg gaaaatagcc cttagaacta gcaggaaaag aatctccttt 13980  
 catttgtcta gaggttctgt taaagtgcct ttgcttctat tttgagactt gttottaaaa 14040  
 aaaatgcgga tatgaaagaa aataaaaaac acattatccc tccacttttt cttggaggag 14100  
 gatgtgttga agaagtcaaa gttcaccatc ctttagata gaatcatttt gaacaatttc 14160  
 atatgtcaat acattttgct catctctaaa ttcatcttta gagcctgttg tgttctgtgc 14220  
 atggatatgt gtgcgtgtat gcacacaaaa ataaaaggaa atattttatc ttatgaataa 14280  
 gtatagaaat aaattaattt ttggaatctc aaactatcag agacttatgt aataaccaga 14340  
 ggcaggcctg attatgtatg ggcaaagcat ttgtgaacaa tgtctccatt gtataacata 14400  
 caaaacaagc ttttcttcca cattggatat gcaagtcgga cttctccaat aagggcctgt 14460  
 ctotttccaa ctccccccac ctcccacctt tgagcaaaca ttattttattg tggtgatgt 14520  
 gtgatcaggt cttgatttgg ggcctctttt tgatgccttc tctttgtggg atctcaccac 14580  
 cgtgcccctg gagacccttt ggcctgccag gcctttgttt ccagccacc catgtggtgc 14640  
 cagtagtgtc tgctttgtag cgagctgtcc ccagagctc agcatggctt ggggatggtc 14700  
 tctgaggttg ggcttgatc cctcccactt ttgggtcag aaagaatgac tgccctctat 14760  
 ttccctgtcc ctgcctctc ttatcctgtt tccagcccg catcatgtta tctttgcttc 14820  
 ttgtaactta ccaaacgatt tatgggcaag taggggaggt gaagagggaac ctcatctatc 14880  
 aagataacct actttgtgoc aaccactgag catagcaatt gtcccttctc cagccctctg 14940  
 aggaccgtgg atgggattct catgaaatga aacagggtgag gaacttttct tttagggaac 15000  
 ttgcttgagg tcccacagga agcgagtatc aatcaacgtc aggatctgag cctgtttctg 15060  
 ttcggtgaa aatatactcc ctgagatggt gtaggccacc atggctttca gcaggctctg 15120  
 tgcttggtgg aaggaagctg gaagctgtgt acacaccac ggggaacagg gaaccatagag 15180

gagcaccttt tgagtgcaga acctggcgaa acatacacct ttagaggat tttagggtacc 15240  
cttgaggctg ggagaatcaa gcagagctaa gtttccatt ggggtgtcac agactgaaga 15300  
aacagagccc taggtagcac aggaagtgtg attgccagt atcagttagt ttggctttaa 15360  
tgactgagaa gagattccac cagttcattg aagagagggc ggacttttta ttggaggaaa 15420  
gaagagtgcc tgtaagtaga gaagtctccg ggggtgtagt ctgtttggg caggaagaac 15480  
agtgtgagcc actgtggaga gaaagcccaa agagtcttg cagggcaggg agtaggatg 15540  
at ttgaagcc agaggaagta tggggtctct gtagactcca ggcaagccat gttaatattt 15600  
taggaagccg tgatggagct gcagatgggt gtggaagtta aagtttaact gttcattcac 15660  
cagtccttcc cctggagaat gtgcagcacg tggacagtgg aactttaagg tccttggtt 15720  
gtatttcaca cccaagagat gaataggtcc aggtatgtca tagaccagac taatgaaata 15780  
acaaatttct tttcaaaaat ttacttttt gtaggaaagc ttctctgtct ggcatTTTT 15840  
ttctccagct tgtgactcaa tcttaaacgt ctccagacaa ttgcataaa atttccaca 15900  
gtgaattgac gtatactttt gagggttcca ttctttttt atttttttt tcttttgaga 15960  
tggagtttct cgtcaccag gttggagtgc aatggtgcca tcttggtctg ctgcaacctc 16020  
cgctcccggt ttcaagoga ttctctgccc toagctctt gagttagctg gatgtcaggc 16080  
accgccacc atgcccggt aattttctgt tgttttagta gagatggggt tccacctgt 16140  
tggccaggct ggtcacaaac tcccgacctc aggcaatccg ccgcctcgg cctcccaaag 16200  
gctgtgatta cagggtgtgag ccaactgtgc cagcctaggg ttccatttct taaccctctc 16260  
ttctgatgcc tcagaaagtc ttgctctgta agcctcttgt agctgcctg gttcagggga 16320  
agggggaggc ttttgtttta ggaccgtcca gaccatagac acatttctg gcacctagca 16380  
cgtgttgggt caaacaggaa tgatgaatgc atgcatgaat gaggttctta gcgctgaaga 16440  
cgggtgcata ggtgtctac cagcccgct gatcattcca atggccatt atgaatgtgt 16500  
gtgctgcagg gccctccac gatcccgta gcaactgtga tgttgtggg aggtgctggg 16560  
agaaagactg ggtctcagaa gatgggttag aggtgggtcc ttctctgctg ctggctagca 16620  
gggtagctgt ggaggggtgc cccatcttgc tgggtctaaa ttttctcact gtaggcaggg 16680  
agcatgacct ggctgaattc taagtcttt tctactctga ggttcattgt ggggtgtgacc 16740  
tgctgggctc agctctggct ttgggagaca ccctctccc ttgatctcga caaccctta 16800  
gcagagccca gtggctccta cagtgcctg agctgcttgc ccgaaggatg cggttgtggt 16860  
tatctcacc cctgcaccc tgtttgcga aggggttgag attgtgtggc ccctccttgt 16920

acttcggggt gaggtttgct ccagaaaggt ggtctgcaaa ggggttggct gggggggagg 16980  
 aggaagtcac tctccaagtg tttgtcctca tcgttatccc aaattgottg cctggaataa 17040  
 ggaagaaaaa aaaaaaaaaa actcttgagt ggtttgggcc aggatcttag ctgatggatc 17100  
 tggtagttcc ctctgtcaga tttgttttct ttgaactgtc tgggocggtc acagtgtcat 17160  
 tgtttaaatg tggaatgtag gtgttctgtg ttctgggaaa taaaaaccaa aactgggtcca 17220  
 ggggatccac agaggttaaga aaagaacatt ccaataggaa tgtttcagaa ccaggagggg 17280  
 aggagagaaa aacggctctg ttggtctcct agaggaagaa cttgttagat ttggggagag 17340  
 tcaggataaa tttgacccta agagtctctg attcctttta gagacttttc ttataagaaa 17400  
 taaaatggaa cttgggagag gcggcaactt gggaaacagc acattctgoc gtaatgaaag 17460  
 tcgtcccata agaattttct tacccttta gccaaatttc tgtttctaaa aggggaaaaa 17520  
 gggctagaga taggcttggtt tgttttctta gttgaatctt actttttgta tttccagccc 17580  
 attctgcagg gtaagaacaa gcacagcccg agggctcact cagtgtgatg ttctagagcc 17640  
 tggctctgoc tcaatccctc acgctggagg atcaggcagc aggggccagt gatggatttt 17700  
 tttttcttcc tttcctccc tattaatatt tactgaggta taaattacag caaagtgcgc 17760  
 agacctaggt atctaggact gtgaggtttt cctgtgttac ctgtgtaacc acgaccaga 17820  
 tcaagataag gaacttttct ggcactctcag aggtctctcc tgctcccttt cagactccgt 17880  
 ctccagaag gaactacttc tgattcctat agccatagac tgaattttct tttccaactt 17940  
 catatgcata ggatcatcat ggggttttta attttaattt catgtctgct tgccactccc 18000  
 aaatggaaat gtgttggcat ctctggatgt ttcttcataa gaaacatgcc ctgtggggca 18060  
 aagcccagga cagggtctgt ctgtctgtg aagtcctgtg cagctggcca gcctctgtc 18120  
 accctctcgg ccacgtctgc actttcagct tctccagcct cctgcccttc ccaactccag 18180  
 tctgtcacct gctgtcctca ctgatgcacc tgcccttttc cttccgtcct ttatgtggca 18240  
 cacccttaag ggagacatct tcctgtctgt gttttgcacc ctcttaaaac tacattcctt 18300  
 tccttcagc attggcatct ctgtccttgt gtattacctg ggatgactat tcagttaaca 18360  
 aatgctttct tcctaggtct tgagcccaag tttgttgat gattggatgg gggcacgttg 18420  
 tgtgagagaa ggatcatggg gtagcatctg gctctcttag aggtgtgttg gggcgtgtga 18480  
 tgcoctgcaa ggogctttcg ttctgggggg ttctgtgtgt ttgaagcaact tgggttgtgt 18540  
 gtcoctgag cctccgtcac gggcaacctc attccttctc tagcctccat cccctgcccc 18600  
 ctgccacccc caggcctctg gagctggctc cctttcctgc tcaactctctt ttggccagga 18660  
 ttttaacata tatcacagga tggtaggcta agagcttggg acttccctc accacactca 18720

aagcctttga tcttttgctt tggaggtaac atcaaaagga aggcctgagga agacagccag 18780  
gctgtgaagt tcaacgttca agttaatagc ttgactgaag gttgtgctgc gttgtggcag 18840  
catcaccgag gctggagtaa acagagtgat tctgccacat tttcctggaa atgcacccca 18900  
atattggaag agggcttctt ttacattcgg aatgaattca ggctgtagtc agagctgctt 18960  
ttccctttcc ccatttttct tggaggtgtg aaaacttggg ggagaagatg tttgtaggag 19020  
ggcatgatga ggggtagagg aagcccaaag agaggatctg gggaggggaa gcccatggg 19080  
atgagactct gaagttaatcc ttgcccgat tcogggatt gctatctgcc tgccttttg 19140  
cgtgggtct ctgtgcccct gactgttctt gatttagcga ggtgtttctg aattctgatg 19200  
gaattcaaag aagcctgggc aggcaggcag cttgacttgg ggcttgggga agcgtgcagc 19260  
ccagacatag cagcgatgag agggcctcag ggctgagggc tgagatgaga atttcacac 19320  
atgcaaaagt gaaagcgacc catogtcttc tccacttgat ctcttgctga gctttgcaga 19380  
cactttggtt gttgtttaat ttaacatttt ctgcaatgct ccttttttca gattttcatc 19440  
caaagctctg tatgagaggt tttcaaacc attttggccc tgattctatt tggcatacga 19500  
ttcaactctg gggatggtca tcttcccac acctgcgttg ggtacctttt tgggtgatgc 19560  
tcagagcact cttggacatc ttctgtgtca gtgtccagca togtgaagct gcccttttag 19620  
ctctcagtgc cccagatag acctgtctct ctgcgtagcg gcaactcagcg tcacctttct 19680  
gtggggtctt gagaccctga tgatcagc actatgctgc cagaattccc cttggattct 19740  
ttagtgtggc ttctcaagca tcccttatcg ctataacgcc ttcattggtt tggcataac 19800  
tgtatactac ctgtgctatt atttatttga tgcattcaaa catttgattc atttatttaa 19860  
actcagtctc actgtaatcc ttaattaaca cctgtgaaat tataggtttg atgtgctact 19920  
tatttattta ttttttaata cacattagta taatcccgta aoggtcaaag taacactttg 19980  
tactgcctaa aaccatgctt gggagcgcca cagtttgaga aagtgcctag ccttcctttc 20040  
cctcctttag tgacttgttg tttggggcat ctgttgactc ctagggtccc cttgttcate 20100  
tttctgttcc taagctcagg gattagtgc tcaaccagg tgtggcctca aaattctgct 20160  
catggaatag cctcaggctt ctataaatct catctttttt gttttgtttt gttttgttt 20220  
ttgagactga gtcttgctct gttgccagg ctggagcaca gtggcgcaat ccactgtaac 20280  
cattgcgttc tgggttcaag cgatcctccc atctcagact cccaagtagc tgggactgta 20340  
ggctgggtacc accaggcccg actaattttt aaattttttg aagagatggg gtctcactat 20400  
attgccatg ccggaagtct agttttatag tgatgagaat tcatctgggg tccaaggggc 20460

cctcctgtgt tgcttctgt gctccctct aaataaagat actccttcca agttgtcctg 20520  
 attttcaggt catcaccatt ttttgagctg gatggggaag ttggcctgga gcagccttcc 20580  
 ctgtctccga gttgcattac ctctgagag gtctcagcaa atcactgcca tctcttgatc 20640  
 agagttgctg gcaagagtcc tctgtggttc taggttttca gccctggaga ctctcgctg 20700  
 cattcattat acatgtcctt ttggtgcctt gttgaaaggc atctcctgcc accgaagggt 20760  
 gtgggcttct ggaaattctc agaaaacaca atatgccagc ctccagggat gggctctcaa 20820  
 agcttcagga acatatcctg ggggtgtgag gaaacacoca ccttaaaatg ttctcagagg 20880  
 gggaatgtta ctgcttgccc taacctctt gagctgatgc tcacatgacg tccctgagat 20940  
 gggcttcttt ttgcccga cttaaagctg taaagggcca ttgtcaaatt tgtttagctt 21000  
 ctcaattcat gttccttaga ggatggtaaa ttaaagttag cattcctgga cagagccttt 21060  
 catacattga agacaaccog gtgagtctca aggggagagg taaggagag atgaaagggt 21120  
 ttctccaggc ctgttcggca gcattgactg ttcttttagg taattaaggg agaccataaa 21180  
 agacaattgt gtgagtcctt ttacctttca ctgggggttc ttaagtcttt ggttgggctt 21240  
 cttaaccct gtgtgtcacc caccggctcc tatgggtgct gttttcattg ttccgttato 21300  
 tagttgctg gaacacacct ttggggattg gagaatggag ttctgggggc ttggggaact 21360  
 ttgagtttct ctgcaatgtc ctatagaagc ttgagtctgt gattcctggg cagggccttc 21420  
 tcctagttag gtgagattgg tggggcaggg cagccagtta gggggtcatg ggagcagggt 21480  
 tggaaaagg tatatgtctt agtaattctt tgtgacaatc accctcattc attgatatct 21540  
 tcttctatc atgtattagg gcagtggctc ccccaatgtg ctgcacatta ggttcacctg 21600  
 gagagctttt ataaaaatgc caatgcccg ggcccacttt gggaggagcc aggcacaggt 21660  
 aatttcaaag gtctctaaat gatttacagt ttgggaatca ccgtatgagg atagtaagct 21720  
 ctgagtccta tgcgttctgt gccgaacacc catgaagcag tcttccaagc attttacctg 21780  
 catcatctca attctcacac tgttaaggag atagacagta tcatctccat tttgtagaca 21840  
 agacaactga atctcagaga ggtttaagtc tcaggacacc aaggtcatta ttaatcaggg 21900  
 ggaactgtgat tgctccctt ataaaatgta ggagatattg tggagtacgg ttgagaaaac 21960  
 attgcaatag ttttcttact ttgttaagaa attaggctgg gogtgggtgc tcaggcctat 22020  
 aatcccagca cattgggaat ccgagggtga cagatctctt gagctcggga gttccagacc 22080  
 agcttgggca acagggtgaa accccatctc gactaaaaat acaaaaatat tagccggggc 22140  
 tgggtggtgt cacctgtagt ctgagctact tgagaggctg aggtgggagg atcacctgag 22200  
 tccggctgca gtgagctggc attgtgccac tgtactccag cctgggcaat gagagtgaga 22260

toctgtctca aaaaaaagaa aaaaaaggaa attagtgggtg gaaggtgaot ttgcatctgg 22320  
 gcgtatctgc ctgcagagtt ggtgtcotta cttgaagaa accotgcttt agttggagta 22380  
 tcottaatgg ttagtggcag gaggggagga gtggttcctg ggagactgga aaaaaatatg 22440  
 gtacctgaat gcttaaggct tggcagatga gcagtcattt tcttacacag agcttaggaa 22500  
 agggcatcca ggtagaggaa tcagcatgaa caaaagcaca gggccataga gttctcagaa 22560  
 ggaaagatgg ggttaacogg agccaagcca gagatctggt ggtagtgggg ggtttccaag 22620  
 ctagaatggt tgtgtggtat tctgtcctca ggggctttga actctgtgtg ctaatgaggc 22680  
 ctcaaattct ctggggctct ggttaaaatg tagattctga tatcagttgg cttgggtggg 22740  
 gccttgcat tctgtaagcc cttagoagtt goactgctgc tactaccgtg agtattgctg 22800  
 ttgagcatta ctaccttgag tattgtgtgc aagtgttact accttgagta ttgctgttga 22860  
 gtattactgt cgaattttac tacttgagt gttgtgttg agtattacta cttgagtggt 22920  
 tgctgttgaa tattactact ttgagtatta ctgttgagca taaccacttt gagtattgct 22980  
 cttgagtatt accaccttga gtattgcttt tgagtgtctac tgacctgagt atcgctgttg 23040  
 agtattgcta cttgaatat tactgttgag tattaaccacc ttgagtattg ctcttgagta 23100  
 ttaccaoctt gagttttgtt cttgagtatt gotacctga gtattgotgt tgaogattac 23160  
 tactctgagt attgtgttg agcattacta cttgagtat tgctgttgag cattactacc 23220  
 tcaaggatg ctcttgagct ttacgcctc aagtattgct cttgagcgtt actgcctcga 23280  
 gtattgccgt tgagtattac tcccttgagt attgocattg agtttagtcc tgtgagtatt 23340  
 gctgctactg cgccttgga atggttttca aactttgcaa cacatcagaa tcaactggga 23400  
 aacctttaa attctaacgc ccaggtcaca tccattcca actagatcag aacatctggg 23460  
 gaatgcgagc catgcaccag tagttataaa acctgcccag gtgattccaa agtgtgggaa 23520  
 ctttgagaa gcaactgctt aggggttgga atagtctgg ctgaatttta atcagggag 23580  
 actgactgct cgtttatga aacgtaggag agtggagcag ggttgagaaa ccatcgggat 23640  
 agtgttotta ctttgtaag tgagcaatat ttgttgagtc tctgtgttg gttctagggg 23700  
 ttcagaggac agcagtgtgc tgctaggatg gtggtctgaa ctagtggaaa ggcactcaaa 23760  
 ggaagaaaga cagaattcta agaggagagg aattttagga aggagatacc caggactttt 23820  
 gaattacag taatttgatc agaaccocaa actgaaatgt ctctgctctg tgatgaaagg 23880  
 gtttgctgac attgagtaag gagctgcag aaggccttta acttgctctc aggtctotta 23940  
 acagctttgt catttacata caagcacctg cctggctaaa ccattcattt ctgtagcttc 24000

cttctggatc tgtctagga atatttgctt tgcattttt ggggttatct taagtgttg 24060  
 aaggaaacaa aatattttt ttaaaaaaa cactcaaatg tagttcacat gattaatttt 24120  
 gactgatttg tgagaatcag taagtgtga ctgactgagg cgcaccacac atccggcttc 24180  
 cttctgttac tctacgcgtg ttgctgaaac ttaagaaco catgtgggt cttctcgcct 24240  
 ggtgcagtc gccccagtat tcatactgag gtttgcatg ggagaaagga aggtatttat 24300  
 ttgtaggta ccaagcagg ccaatccagc agctcacgct taagacotga cctctcccat 24360  
 ggtttataag caagtgttt ttttttttt ttttttttt agactgagtc ttgctctgct 24420  
 acccaggctg gagtgcagtg gcgtgatctc agctcactgc aacctccgcc tcccaggctc 24480  
 aagcgattct cctgcctcag cctcctgaat agctgggact acaggcgctg gccccacac 24540  
 ctggctaagt tttgtcttt tagtagagat ggggtttcac catgttgccc aggctagttt 24600  
 ccagctcctg acctcaagt atoctcctgc ctgacctcc cagagtgtg ggattacggg 24660  
 catgagccac agtgccctgg ctgtaagcaa gtgtttttaa agaaaggggt aaattttagg 24720  
 gaaacagaag ttctaggcaa aatggtaa atatacagg aggtaagaca ttggtttggc 24780  
 ctaaaaagat gggatatttt gaagtgggg ctcataggtc ataagtggat ttaaagattt 24840  
 ttttggttg taattggtta aggaagataa gctttgatta aagatttggg gtcagcagaa 24900  
 agaaatgta ggtctggctc gtgggcatgt ctttttctag gccctcctt ggaagaact 24960  
 ttagagcaaa gaaaggcagt tggagcttag tcccacttt ctctgatct gaggtctacg 25020  
 gacctgga tccatttggt ggggtccatc ttctgaaaa acaagtcagg gacatgtatt 25080  
 gagatgatat tattggtatt tatagggaac caaacacgc cccatgactc tttttggct 25140  
 attgtttta gccactgttt tttttgttt attgagttgt taacttattt tttaaagcta 25200  
 gctagctgcc tggaatttct ttagaaggaa ctgaagttt taaaaattt tatgttggg 25260  
 ggtattgccc tgcaggcccc taaaaggggt ccctgcgctg tctcaaaact tggatgcaaa 25320  
 aagaagttga gttaacacag gaggacagg gtagacgcac caagggcagtg tgcctcgagt 25380  
 ggtgtgctct tattaagaag ggtggttaga cagggaatgg gttagttccc aggtcggcat 25440  
 tcagctgaaa cagtgtggt taaaattctg aaaaatgtcc acgctctgca ttctcttct 25500  
 aacacccagg acccagtaac tataagccc cctaccctgg ggcatagcag ggggcttcag 25560  
 ggacccatga gaaggtcac tgctgctagt tacactcctt ctgggacctg atttagacag 25620  
 tttggtgta gttttgcgag ggttaattt agggccaagg atgcttctag aatggaaata 25680  
 cttcttgac attgggagct ttattggtt attatgtcaa tgtgagaatt caggaagccc 25740  
 agtgctaact ctccatccta aaaggagtag attggctggg cgtggtggcg catgcotgta 25800

atccacagcac ttgaggagcg cgagggggcg cggatcacct gaggtcagga gttcaagacc 25860  
 aacatggcga aacccogtct ctactaaaaa tacataaatt agccaggtgt ggtggtgggc 25920  
 gcctgtaatg ccacctactc gggaggctga ggcagggaga attgcttgat ccaggaggcg 25980  
 ggaggctgca gtgagccaag attgtgccac tggcctccag cctgggogac agagcgagac 26040  
 ttcattctcac aaaaacaaac aaacaaacaa acaaaaacta aaaggagatt tcctccttct 26100  
 gtcccttatg ggagacttca accttgggaa agtctggaat ccttggacat tagaaattct 26160  
 gaagtttttg ctggctgtag tggctcatgc ctataatccc agcacgttg gagggcgagg 26220  
 cagggtgtca cttaggccag gagtttgaga ccagcctggc caacatggtg aaaccccatc 26280  
 tctactaaaa atacaaaaat tagctgggcg tggtagcgga cgctgtgaag ccagctact 26340  
 tgggaggctg aggcaggaga atctccagaa cctatgaggt ggagggttga gtgagctgag 26400  
 atcacaccat tgcactccag cctgggcaac agaacaagat tccgtttcaa gaaagcagaa 26460  
 actctgaat ttttgctgt ccaggccaca tcaatccat tcctctgctg tctctgcagg 26520  
 attctgtgag gaataattag ttaatgttg cagagcactt tgaaatcctc agatgaaagg 26580  
 caccggagaa gcacaaagta ttattattta ttattagctt gcccagaat ggaggcgcat 26640  
 gagggccctg cagctccctg cctcgtgcca ggtgtgatcc tectgctggg cttttcctgc 26700  
 ctgatgagct tttttttttt tttttttttt gagatcaggt tcagctctgt cggccaggct 26760  
 ggagtgcagt ggcatgaaaa cagtccactg cacacagctc actgcactgc agcctcaaac 26820  
 acctgggctc aagcaatccc cctgcctcag cctcccaggt aactgggact atatactaca 26880  
 ggcattgcgc accactcctg gctaattaaa aaaaattttt tttttagag atgggggtct 26940  
 cactatgttg ccaggctgg tctcaaacct ctgggcctca aagatgcaa aggttcacac 27000  
 cttggcctct caaagtgtg agatgacagg cgtgagccac tgtgcctgtg ctcaattgat 27060  
 tttctttatt aaagaaacat ggaagaaagt gaaggatgag aatcagtaac gtaacgtgtg 27120  
 cttcagattg tggacaagtg atgtgaagga aacacattgg tcccactgtg gtgacagagc 27180  
 aggggtttcc ttacctggca aggttgcggc tgccattcct tggggctctg ggtaagacc 27240  
 atctgcctga gggtaacgca gtaataaatc agtaactaaag ggcgtactaa agtactgtat 27300  
 tgctaggcta ggocatgctt ggtgtatttt tttttttttt taattgagac ggagtcttgc 27360  
 tttgttgccc aggotggagt gcagtggtgt gatctcggt cactacaacc tctgctgccc 27420  
 agtttcaagt gattctcctg ccttagcctc ctgagtagct gggattacag gcacgtgcta 27480  
 ccattgcttg ctagttttaa aatattttta gtagagattg ggttttgccg tgttgtccaa 27540

gctgggtotca aactcctgac ctcaagggat cagcccacct cggcctcccg aagtgcctggg 27600  
 attacaggca tgagcctggc tgggtatatt gttttaaatt taaagtttac taaatttaatt 27660  
 gatatctggg gaatcagctt gcttcctggg gatctggatg taacttgaggt gagagggtgg 27720  
 ggattcagaa ttatcctttc tatcgacgca tgttctggat tgattcatgt aggtotcaag 27780  
 tgtgtgtaat atttcatttc tttgtgcaat tttggcatgc cgaggcgggc accotgaagc 27840  
 tccggcagag cctggagaca gagtggggag ctctcgcctc tttcccttcc ttcacccag 27900  
 ctgaactoga ctggaattga attcatcagc tgctggagag ttgttttatt tgccctgctg 27960  
 gtggagaggg aggaaaggaa catcatgggg ccaggctttt tttttttaa ggaaagattt 28020  
 gatttacttt ccccttagt agcatgatgg gcacctgcac ccgccagcta atcagaagcc 28080  
 actgtccctt gaatgcctcc gctgccacc agatcctgac agcatccac gcgggagcac 28140  
 tctcgtgtgc cctggcagc tctcgtgtgc tggcagttct cttaacttgc tgggtgtctct 28200  
 ctgcccggag gctcagaaac ccagaggact gaccacttct tgaggctcat gtccagtttg 28260  
 caaagagccc ccagcaagca gagaagggga tttttgtacc agcgatatct cttctccact 28320  
 cctcaacaca ctcttttcca ctctgtctcc tataaacatg gaacagccag gaatactcaa 28380  
 atcctagcct gtcataagc caaaaattga tagagatcta ctgtccagaa tgattttctta 28440  
 tagtgacctt gtgttttagt ggtaagactt tcttaaacca tgagggattc tgggtccaca 28500  
 gggcagtaat atctggggca gagcctgaga cttttctcat tgatttcctc tgtgagccag 28560  
 gagtgactgc tctgatgcag ggtgctgtgt ggttggtaga agctggcgtt atcccatttt 28620  
 acccacgagg aaacaaatga ccagtgggtg agcgggagct cagcatccca tgtgccact 28680  
 tcctcctcgg gtggactttt cacctgccca tgccgtotct tttgcaact ttactgcagt 28740  
 gacggagaca tctttaata caaattcttg ggggaaccct gtgttccttg gctggagcct 28800  
 ggctgggaag gagggaggag cagagggtc tcttgggtgt ggctatttgc agttgagcca 28860  
 gggaaaggct ggtccactgg agacacccct cctggtcacc gcagacttcc tgccctccat 28920  
 ccagtgtcct tctacttga ggatgtgtgc ccagcagaga gaatctctga agccatgtca 28980  
 ttattgggat aacattcctg tcccagtcac cttatttcto agaaaaagga caatgggaaa 29040  
 caagttttta ttgaatccta tgcctggcct attaatgggg tctcttactt ttcatagcag 29100  
 cactgcaaac agagttaagt ttctattcat tttatggatt agaaagctga gatccagagc 29160  
 gggcagatgt aaacctggg tctttaaaat gcatcctttt tgcaaacaaa taaacttagt 29220  
 gtattaaaag ggctggagag agcagagtaa ggtaacattt ggggtgtcag catgtagtct 29280  
 tgggtcccca cagtggagat ggcacagtgc tgggtgctgg gggaactatg gtcactaaga 29340

gacactgaat aatttaaatgc atgccctga ttccatcact gactgttgag gtaacacata 29400  
catttatatt gtcagtggg gtgatgatta catgagctgc gtaaagcgtt tgaccagtgc 29460  
ctgcacataa catagtaggt gotcaataaa gatcacccac tottaagagg tgggaggagg 29520  
tgaagtcatc tttotgggga gtgttgccct gttgtttctc gctgcattct tttctgcctt 29580  
tgggctccga gaatgtggg ttgggcagtg tgagtggctc totcaggcct ctgtgacatg 29640  
ttgctttcat gaaaggttcc cctctagcca aagactgagt ggtccttgca ggctttctcc 29700  
tgagtctttt tttttttttt tttttttttt ttaaagacag agactctgtt gccagattgg 29760  
agtgcagtga cgcggtctcg gtcactgca accgctgcct ccaggttca agcaatctac 29820  
aaaatgcac tataaatga tgcacagcc tctgagtat ttgggatcac aggtgccac 29880  
taccatgcct gggatatttt ttgtattttt agtagagaca gggtttcacc ctgttgacca 29940  
gtttggtctc aaactcctga cctcgagtga tccgcctgtc ttgacctccc aaagtgtgg 30000  
gattacaggg gtgagccact gcacctggc tctctcctga gtccttttgt ttgtgcctgc 30060  
tttggggatt ccctctggct ggggtggact gccgggatct gttgtccag tgtacatttc 30120  
ctggtcacct agcaccggc agctgcggtg ctgggaggaa cagggcctgg ctctgggagg 30180  
cagctgggag agtcaggaag tgaagaaagt tcttgtgggt gtgatggagg aaaccaagc 30240  
agcgtccaga gggagcacia gaggaggga caaatcttgg gagggtccc ggccaatggg 30300  
accagtgtga agaaattgca cctgtcctgg cagatagaga aggtggaagc agtgaatgg 30360  
agagcatcct cactcttctc tctgcagca agcaccttg gggaagtcc caccgacagg 30420  
aatgtcgtgt gtcttggtt gagatgtcaa agaaacatgt tggacacacc atggtgacag 30480  
agcaggagtc tcttaacccc ggcgtggtt aggctgcgt tctggtggga tctgggttca 30540  
gtcaggggtt aacagtgcct cctgcttgcc tgattgacac agtaataaag gcagtgacac 30600  
caaaactagg ctcaggaatg tgcctcgtt agaaagactc actaatggtt gtgggggggt 30660  
ggcccatgag tcttctggg tggggcgag aagtagggga ccctttggg tttgcccttt 30720  
ttggtcatag gacttcaact cacagacata attgaaccgt tgggtttctg cagccaaatt 30780  
caaatgtcac caatcttgg caccctttc atctcttggg tcctctgtaa gttatagcta 30840  
totgatagtt tactgaaaa taaactgaaa atatgtttta aattgtactt tctgatttaa 30900  
ataatgttta gagacaaaa aaaagggtcc aatccacttg gagaaaagca ttgtcaaagg 30960  
tgggtgattt tttcttttg ctgttttaa gtggttaagt gatgagtgtt ttggatatat 31020  
tgatttttca ggtgtgcagg cgtcacatg aacagctgac atttttttt tttcatgtg 31080

gacttcagcc agtcttgaca cctgcccctt aacgaaaagt aaaccatcgc ottgtttgac 31140  
agtttaagtg cagtgtacg gatggaggca ggtttaogtt atgttaaagg ottgacaacc 31200  
cagaaccccc ctgttggttt ctttgttgta acccttgagc cgggtggcctg ctgaaatgtc 31260  
acctttgccc ttctttaaaa gcaggaataa taggtggtga gtgggtggat gcctcttaaa 31320  
atactggaaa gtgctgtggc ccgagggtaa gctttttaga agtgagtgtg tgtgttgtgt 31380  
tgttttaatt aatgaatott ctgggcctga agataatgag gtcagtgagg gcagccatgc 31440  
tgccacacag ctacacttag ggtccctgtt gtccagaacg tgcccgacct actggagggg 31500  
cctgggaatg cttctttgat tgacgtgggt aggaagacag atgtggcggc ctccatgctg 31560  
atgggaggca gctgggaaga aggtcatggg caccatctca ggagtggcag agccacotcc 31620  
ccctctctc accccgtgtg tctggattct tccagctgtg tggctcttct tcctgcctgg 31680  
aatgagcat cctgcagagc tcggctcctg ttcacacct cctcctaacc ccctactctc 31740  
cctctccctt tcattccagg ctggaggacc agatgggctt tacctgatgg agtgtgcttt 31800  
gctgacatgg tgcaaagagc caattcctgg ttgcaaagag gcagctgggt gcagaggcgg 31860  
gggtcattcc tgtaataata ataacttggt tttttataat actttacagt ctaagtactt 31920  
ttcaaatact tgacctcatt tagttctcac cacagcccto tgaagggata ttactattac 31980  
cttcatttta tagatgcgtt aaccagggtt tgttttggga ggtagagggg gtgtgagggg 32040  
gacagagggg agggaaccag tgttgatga attctgaggc cctgccaaag cagccagcta 32100  
gctagggtgt gtttaagtga ctctttgcat ctacagaatg agggaggtgg gatgagggga 32160  
aattatttca caataactga aggtcggaga gaactaactct ctgctcattg tcacacagca 32220  
gttgagtgcc agctgggatt tgcacccag gtcacctgac tccaaagccc tgtgatcgtt 32280  
ctcttctgtc tctagtatac ccagcataat gcccgccaaa gtgctggcat caataaatat 32340  
ttgtogaatg ttaaatgagg ottaaagaga accattcatg cttggcacag gggcacagt 32400  
agacaaacat gtttctgtcc ctcgtaacct tcgcttccaa attctgtgac cttgggcggg 32460  
ttgcctgagc tcttttccag ctgagtttc tgaaaactca tccggaaaat gggtaaaata 32520  
tcagagtga ttctgtatgg taagactgca aatgttagat gattctgcta cttattattg 32580  
ttttcttttt ctacccacac acctccctt tttatgtaca cctcctagga agtaagtttc 32640  
tgataacata ctgcattgtt ggataacagc aacaaaaagc acttctgac attattgccc 32700  
aatcaccaa atgaggcaat taccaacttt ggaataagaa tagcaacggg ggtaagagct 32760  
gacatttctt gagcgcttgc catatgctgg gtgtactact ataagctgct gcctgcaatt 32820  
atgatttggg aacaactctg aaagtagtta cctccattt tatagaagag taaactgagg 32880

ttcagagagg ttaagtaacc cccccagggc ctctcaggaa gtagttggtg gcctgggatt 32940  
 caaacaccag aattggtctg acttcccact ctttaaacca cactcaaac tgaactctcc 33000  
 acgtgtgtgt gttctgggca ttttcgcac tccttggct tgttgacagc gtggactttt 33060  
 gctttcccat tctcatagaa catggccagt gcaggaggag gaaatccaca ctggtctttg 33120  
 gactgaacca gaggtggcg atggtccga aacaggggtgc caagtggctg acccttgttt 33180  
 ttatgccttg cgtggtaag cttggggcac caggagtctt ttgaatttct ctttcttgac 33240  
 tgtccacgcc ttcttttaggc aatcttttaa caggctatgt tttaaatctt atgacatct 33300  
 ctgaagaaaag aggaggaaaa aaaaatcaag acatggcctt agagaagtga caggttttct 33360  
 ttgagatttt gttttctgtt ttcttttta tttgtgcac attgcaaac ctcttggga 33420  
 tgatgatttc gtgtggttcc ttggtagccc ttgggcagct gctgccaggt ttcacccaaa 33480  
 tgcattgtga cccctgttt cgtgggacgg ctttgctcc acatggctga ttgtgctctg 33540  
 tgtgtccgct gtgggcagag tgtgattgta agaatoagaa ttctgctggg cttgcaagca 33600  
 tttaaaaat ctctataagt ttgagaactg gttggaagg agagatgcag cgaactagaa 33660  
 cccccggcgc gcagctgagc ttccgtgtgc ctgggaggag cccatagga gaaacaggaa 33720  
 aattccactt caccagaaag ctggggaaat gagtgggagt aggggccagg ctggttaact 33780  
 aggaagactt ttggtcactc tgccttactt agcctaaagt gttcatttcc cccttaagcg 33840  
 gtggttaata cgcgtatctg cagatttact ttttgatga tttaaaatct tgcaacatct 33900  
 caagggattg tatcctgatg atactgatta tattaataac aagataatag cttataatat 33960  
 aatagctagc aattaccaag cacttacctt acaccaggta caatgccagg catttgatcc 34020  
 ttacacgaac tctgagagat ggcgttttgt attccattt tacagatgag ggaagctgtg 34080  
 ctgagagagg tgagttcatt tgcccaagct caccacttt gagatggtct gcttttatat 34140  
 cctcttagcc caattctttt ggatgtcagc acttggcatg tattaggcac tggataagtg 34200  
 ttttgttgaa tgaacaaaat gatggaactt ggatttgaac ccaggctctga gctggctttg 34260  
 agtcttcaga atagtaggtc caattagtgg agtgggggct cagtagtoca aagggaaagg 34320  
 agcaaggaga cattgtgggg gccgagaaga gggcttctgg ggtgtttcct gggcacattg 34380  
 gcattaaagg catagtgtga agtgccattc aagaccatgt gctggattag tgttttctct 34440  
 ctccaactga ggtgcttgcg attggctttg tgcccoggtg tctgcaaagt gagttgggct 34500  
 gaactcagga agaccttttg gttgggatga ctgtgtatcc acctgcacct gagtagggac 34560  
 tgagtttcac ttgccagttt tacgcagca agacctcgtt aagttggctt cctctcattt 34620

aggcatttgg gaaactttag gcggtggag tttaattctc aaggcaaagc cctttttcaa 34680  
 gggacatgaa gaaaggcaga gggatatatt taaaatacct gaatgaactg tttctttttc 34740  
 tttttatttt ttgagatgga gtctccctct gtcacccaag ctggagtgca gtggcacgat 34800  
 ctccagctcac tgcaaccttc acctcccagg ttcaagcgat tctcctgcct cagcctcccc 34860  
 agtagctggg actgcaggtg tgcaccacca caccagcta atttttctat tgttttattt 34920  
 tattttattt attttttaat ttttttttg agacggagtc tcgctctgtt gccaggctg 34980  
 gagtgcattg gcgtgatctc ggctcactgc aagctccacc tccgggttc atgccattct 35040  
 cctgaatcag cctcccaagt agctgggact acaggcacct gccaccacac ccggctcatt 35100  
 ttttgtattt ttagtagaga tggggtttca ccatgttggc caggtctctt gacctgtgta 35160  
 tcggcccgcc tggcctccc aaagtgtgg gattacagcg gtgagccact gcgacttgca 35220  
 tgtagacagt aatggcaggc cactatcagt gggctctgta atcaggtgtc caacctgggtg 35280  
 ctgggcttgg tggctcatgc ctgtaacct agcactctgg gagccaagg cgagtggatc 35340  
 atctgaggtc aggagtacaa gaccagctg gccaacatag taaaacccca tctctactaa 35400  
 aaatacaaaa attagctagg catggtggga tgcattctga gtcccagcta ctcaagaaga 35460  
 tgaggcagga gaatggcttg aacctgggag gcggagattg cagtggacca agatcatgoc 35520  
 actgcactcc atccagcctg gacaacaaag cgagactctc acaacaaaac aaaacaaaca 35580  
 aacaaacaaa caaaaagtca cttgcttctt tttttgcttg cttatggaca taaacctgt 35640  
 aaactatctc atacatcatg ggagtgaatt tgcagtggtg agactgctat tcacaaactc 35700  
 atatacatcc tatgaaggag tacagggtta ataaccatta tccaaaatgc ttggggctga 35760  
 aagtgttttg gatttttaat ttttttcaga ttttggaata ttgcatata cataatgaga 35820  
 tatcctgggg atgggaccca aatctgaaca ggaaattcat ttatgtttta tataaacct 35880  
 tttttttttt ttttttttg agacagagtt tcacgcttgt tcccaggct ggaatgcagt 35940  
 ggtgtgatct cggtcactg caacctctac ctcccaggct gaaacgattc tcctgcctca 36000  
 gcctcctgag tagctgagat tacaggggct tcccaccag cacagcta at ttttgtattt 36060  
 ttagtagaga tgaggtttca ccctgttagc caggctggtc tcgaactcct gacctcaagt 36120  
 gatccaccgc ctttggcctc ccaaagtgtc gggattacaa tgtgagccac tgtgcatgac 36180  
 cttatataaa cottaggtta ttttatacaa tatttttaaat aatttttgtg catgaaacag 36240  
 agttttgact gcattttgac tgtgactcct cacttgaggt cagggtgtaga cttttccact 36300  
 tgtgggtgca aatttcagat tttgaagctt tatataatga gataggtgtc tgcctctgtg 36360  
 ccaggctga agtacggtg cacaatcaca gctcactgca accatgacct cctgggctca 36420

agtgatcctc ccatctcagc cacctgagta gctgggacta caggcatgca ctgtgactgg 36480  
 attttttttt tttttttttt ttttgagaag gactctggaa totcaagtct cgtctctggg 36540  
 cccaggctgg agtgcaagtg gcgcgttctt ggctcactgc aatctccgcc tcttggttgc 36600  
 aagtgattct cctgtctcag cctcctgagt agctgggatt ataggcgtgt gccaccactt 36660  
 ctggctaatt tttgtagttt tagtagggtc ggagtttcac tgtgttggtc aggttggtct 36720  
 tgaactcctg aactgaagtg atctgccac cttggcctcc cagagtgatg ggattatagg 36780  
 catgagccac cgtgccacag cttggctaatt tttttatatt tttgtagag acagggttcc 36840  
 gctatgttgc ccagggttgg cttgaattcc tggactcaag caatctgcc accttggcct 36900  
 cgcgaagtgc tgggattaca ggtgtgagtc acgcctcctg gcctgaagca ttttggtatt 36960  
 ttgggttaga gttgcacagc ctttactgtt attatcctga tgttattatc cacattttac 37020  
 aggcaaggat ctggaggcgt agagaggtaa aatcattttt tcaaagcccc agaagtacta 37080  
 agccgcagat totgaatttg aactcaggca ttctgggtca gaattagtga ggttttaagt 37140  
 taattttttt ttttttttta gatagagtct tgcctctgta ccagggttg agtgacagtg 37200  
 gtctatctc ggctcactgc aacctctgcc tccgggttcc aagtgttct cctgcctcag 37260  
 cttctagagt agctgggact acagacatgt gcaaccaagc ctggctaatt tttgtatttt 37320  
 tattagagat ggtgtttcgc cactgtggcc aggtgtgtct tgaactcctg acctcaggtg 37380  
 atctaaccac ctggcctcc agaagtgtg ggattacagg cgtgagccac tgcgcctggc 37440  
 ctacccctg tgcgccaggc tggagtgcaa gtggcacagt ctggctcac tgcacctct 37500  
 gcctccagg ttcaagcgt totcctgcct cagcctcctg agtagctggg attacagata 37560  
 cccaccacca tgcctggcta attttttttt tttaagtatt tttagtagag acagagtttc 37620  
 aacaagttag tcaggctcct cttgaacttc tgacctcatg atctgcctgc ctgggcctcc 37680  
 caaagtgtg ggattacagg catgagccac catgcctggc ctaagtttgg tttttaacca 37740  
 tgcctgtttt tctagacct totgtcagcc agctccaca tgggtgaatca gggagttagg 37800  
 tccgtctgta agagagggcc caggagctgg gtcagatagg taaagagaca ttcttagtt 37860  
 cttatcctct gctaccaagc cattctgtg gatcaagccc totttggcct gtgtcattcc 37920  
 acctcatta agttcagct tctttccttc tttcacatc tttccactgc tgttgaaaac 37980  
 ttgagacctg aaatcccatc tctgaattc ctgggagctc tagaagtgga gatggccagg 38040  
 ttctgtgtgc agagctggtt gggattacaa ataaaccaa gcctgggaaa ctttctgtct 38100  
 attaatagcg cagacctttt ggggagggaa taccacaaat cgatgctgtt ggaattgatt 38160

ttgcctgtct agatgacata ctaatgagct aagtgggttag cttcggatca ttattgctct 38220  
 ttccaagcc aagttctttt aaagactaaa accacaaaag cagagaaoga gttgggttag 38280  
 agaggcatag tggctgggtc cagagaaggg agagtggtoa gccctggcct taaacatgag 38340  
 aaaataaagg tggtccttgc tttggaatga gttagtgtg ctgactaatt caactggtt 38400  
 ttctttttct ttttggaagt ggaatttgat ttggtgtctg gattttgata gggccattta 38460  
 tattttctca gcaotttttg gttcttgcag aaagttacat tctagtctc tcaactgctt 38520  
 attttttttt ggtttttgaa gcagggaattt gatttggtgt ctgggttttg ataggggtgt 38580  
 ttatgttca tcaatgtctt ttggttctt cagcgtttct ctcctgtct gtcagtgtc 38640  
 agagagggtg cctgtcaacg attctctctc tcaggggaga aagtctttct taaaacagcc 38700  
 ctaagatcct tatctcctca aatcacccac tttgatgata atatccattg ttctcttctc 38760  
 tgcttcttgg cataattctag tcagtctatg tatacaatta aaaaaacaaa acagccctct 38820  
 gtgtccaaag tgcttggaat atcccagtgt ttcacaggag actttggaag tggacaaaac 38880  
 tgtatttctc tcccaaatg aggttattgt gctcgaaata tctcctggta gtttattaaa 38940  
 ggaaaccgca ggcagggtga agagaggcag tttctacagc ctgcaacct attatcttgc 39000  
 ctctctttt cgacctctc tctccctct cctcttctc ccccccctca ccctattcaa 39060  
 cccagcccca catgtcatgc cgtcccagc aggtagccct gcagccctgc ttctctggga 39120  
 tggctgttgc ttgccaccg tcccatggaa cgtgaagaag gaatttgggg tgtggacttc 39180  
 cttgagtgc taggattaga cccgtcgggt ctgcagtcag acgagaagcg tgtgggcaaa 39240  
 gggaactatt gtgtgaggt tctctggaca gaaagcctgc cttcatctt tactgtgct 39300  
 aatggacaat tgagacattc agcttatgtc tgaaaggaaa gtgggcccgg atggtctagc 39360  
 agacctcca gatgaaggct tgtaggagga gcaaatagag acaaggatta ccaacagggg 39420  
 gaacaactgg ggcagagtcc tgggagagaa tgtttatttc cttgctctc aggagggtt 39480  
 tggaaagagc cataatcctg ggttaggagt aatttgttac agcaagatga tacttgagt 39540  
 acaggctgct tctggctgag gcagcaagac ttgcatgcag gggggtcgtg gggcctocag 39600  
 aaggctagcc tcccgtaaat cttcacctg gctttggggt ttgttctcc ccaagcaaaa 39660  
 ttaaccagag gcactgtga cctttgggt tctgggtgt agcgttacga agcatctoca 39720  
 catgtttgtc acagctagaa tttgacaata aaaatttgga caggagagcc ctgccagagc 39780  
 cactgacctc tttccaatgt gacaagggga aaaaaacaa agggaaaacg cagcacgggg 39840  
 tgcggttca gttgaagtg gaggacaagg agcccagcct gtctcgatt tgetgtctat 39900  
 gtagactcac taaagcaagt taattcattg ctcttgacc gccagtcct tcgttgtctt 39960

ttgttgttgt ggaatggggg aaagaaatac agaatgggga ggagaacota attagaaaaa 40020  
 tcaagccttg agagctccca gccatggaga aagaaaggga ttttttagaa gttgtgattt 40080  
 taatatctgc tgcaatctga tgattcatgg atttaaaata acccttccag gtcaccagga 40140  
 ccctgttact tgetggcttt gtacctctca aaggtcattt gttggcttcg tctcttaaca 40200  
 atttccatgg tagacctaaa atttctggct gtgaaatccc ctgtgtagtg ggaagaagaa 40260  
 atagcaaato ttagctgcct tggacctgat ataattattt gtcttcattt acatggttta 40320  
 tocttcaagg ttgaataaat gatgtgggag ctagtoaagg ggcttttagt atgtgatttc 40380  
 atgcctactt ttttttaggt agagaaactg aggtcacagg gtactagaga atggactcta 40440  
 agattcaggt ttctgaattg cctgtgggtt tgttgactca actgctcttc tgttgttttt 40500  
 tagccacatg ccttgaaaca gtccctcttc ccatgtttct tcatcagcac cattaaccca 40560  
 aggtatactg toctctctta tctttcacia ggtcttggag ttcccatgcc tttgtaagca 40620  
 toccctcccg agattcagca ccaacaaaaa tcacatttgg aaaaattgct tgtttcccaa 40680  
 gaagccttgg aggatatgat tttgtataga acgggttcac aggttttctg ttcatcttc 40740  
 tatgggtggag tgtgtgtgta tgtgactctg tcttctctcc attcctcttt tttttttttt 40800  
 ttttttgaga tggaaattcg othttgtggo ccaggottga gtgoaatggo gtgatctogg 40860  
 ctcaactgca cctccacctc ctgggttcaa gcgattctcc tgtctcagcc tcgcaagtag 40920  
 ctaggattac aggcattgag caccacgtcc agctaatttt tgtattttta gtagagatgg 40980  
 agtttcatca ctttggtcag tctggtcaca caaactcctg acctcaggtg atccaaccgc 41040  
 ctgcgcctcc caaagtgtg ggattacagg tgtgagccac cgcgcgaagc ctcccatcc 41100  
 ccttttatct cttaaataaa tgtggtcacc atcaaagatg gtgcctgact cttttttgtt 41160  
 ttcagttcat cttaaattca catataattc acacgtcata aaatgtaccc atttaaggtg 41220  
 tacagttcag tgggttttta gtctatttag tatatttaca agattgtaca aacataccag 41280  
 tatcttaata tttttatcat ccccaaaaga aacactgtaa ccctagcagc cagtctctac 41340  
 ccgccttccc catagctcct ggcaatcact aatttacttc ctgtctctat gaatttgcct 41400  
 attttggtta tttcatataa aaagaatcat acacatgaa actttcttca tctgccttga 41460  
 agttagcata ttttcaaggg ctaccatgtt gtggcatgtg tcagtaactcc atttggtttt 41520  
 attactgaat agtattccat tttatggctg taccgcattt tagttatoca gctatcggtt 41580  
 gagaattggt gcattcttat ccagaaacat accatattca gctcccagtg acacccacat 41640  
 tcattcctgg gctgctcctt gtcttccagc tattttcctg gtctcctggt gcctctgcct 41700

acttcagcat gctgtagaga catgggtagt aactaaaaca ttccaattaa ctgcattgta 41760  
 cttggccttt ttataagaag cagtaattag aaaatatggt ggccacaaga ttgatattaa 41820  
 agtgaaagat tgtaaatact tttctgcctg aaggtagatg gcctctggcc tgcctcttag 41880  
 tgggagggtt ttccaggagc ttgcaagcat ccattatttg ttagtcatca gcttagcggc 41940  
 caaggagcat tagcctgtct tgcctgtctt gctgaagact ctgagagaca tgggagggca 42000  
 agggctgtct cttttgaatt cttccaatgt cttcatgtcc tttaacctcc tggcttaggg 42060  
 acttgtgtgc tggtggtgga gctgacattt gtttggaaac cacagccctt tgggtgggac 42120  
 tcaatcttgg ggttgctga agactttgag atggctaggt ctgggcctct tttggctact 42180  
 atggaacaag actgtctcag aggccagagt ctgtctcacc agctccctgt cttgggactg 42240  
 caccattgca ggttctttgc cctccctgag agatttctct tctgtcctgg gcaccattg 42300  
 gccattctgc ccgtaagctc agtaggggtg agggaaaaga gttctggcct ggaagtacca 42360  
 aagtcctgag ttttggttgc agtccctcat aactgtgtat aactaagtca cttagttttc 42420  
 tgtgcctcac tttctctgtt ttaagatgg atttggagat tattggcttt gaccacctaa 42480  
 aaaggatgta gtgacaatca atttagaggt ctaaaagagc ctttgaggaa gtaaaatgga 42540  
 atcttcaaat ggactacatg ctgattattg acactgcctt agcactgata gttgatgttg 42600  
 actgatgtgc agaattgctt ggcaagttgg aaaaaagtac gtacagatcc tgggccacta 42660  
 ccaagtttca ttaacagat ctggagtgca tcaggaaaaa agtccctcta aacaagccag 42720  
 caaggtttgg atactgtgca accctttttt tttttttttt ttccctttga gatggagtct 42780  
 ggctctgttg cccaagctgg agtgcagttg cacaatcttg gctcactata acctctgctt 42840  
 cccaggttca agcaattctc ctgcttttag cctccgagta gctggcataa caggcgcctg 42900  
 ccaccacacc cagctaattt ttatattttt tggagagatg ggggtctcacc atgttgacca 42960  
 ggctgggtct gaactactga cctcaaagt atccgcccac ctctgcctct taaagtgtg 43020  
 ggattacagg catgagccac tgtgtctggc cctacttaac tttttgtgt taattcctgc 43080  
 accattgatt agcttattgt cccattgact gtgtctttag atgacttctc tgggcctcag 43140  
 aatatctagt ccatagtga cacagagcat ctgtttaatg gtaaatgctg caggaaatcca 43200  
 tgcattggag tagaaagagt tttagatcat gttcctcatt tcttgctaca gacttaggca 43260  
 aagcgtggag aagaggttgt ccaatgaaga aatgaagtga catgccaggt cagtggcaga 43320  
 gctaggcctg gaaaataggt ttccagactc ttccctttct accatacttt tccctggagt 43380  
 acgcaactgt aatttgaaga gcgacttttg ggagaggggt gaaggaaagg cttgggcctca 43440  
 gctaagggg cccattgggt gtgagaggag ggtctggtga aattccatac cgattgtccg 43500

tgtgtgagct gctgtacat agcctcctg cagaaccaact aacotgtcaa atgcagaaat 43560  
 agttcaggga cagagctgtt aaaggattgg cgggttaaag aaaacagtga atcccaagtt 43620  
 ttgttaattg gatttttttg tttgttagtt atttgttttg ottcattgtc ttcacacac 43680  
 caggggcctc cttaaatctg gtggaaaaat ttcocttgaa aacaattcag tgtttgtcca 43740  
 tagacttggg agggagagat gctagatgct ggaaagtctt gcttattact ttggggacac 43800  
 tgagatgttc ccttcacat gtactttgag acacacatcc tggttgagtt caggcaagga 43860  
 tgcctaacag ttgataagaa aactgggaaa gatagaaggg atttgtaagg taagtcaggg 43920  
 tgagtgaana cacatccggt atgctggaga cctagatgct tgactgcac tcgctcctgt 43980  
 cacctcagtc aatctgggtc ttgctctgtt ggccttagtt tccctcctgc taacaggtta 44040  
 gttccacott totgccatt tattttgtag ggttattgtg gatgtcattc tgaactotaa 44100  
 aataccctct aaatatgaag tgatattagt gctctttaca ttgttatgat taaaaatatt 44160  
 tatgagaaaa aggttaactg taaggatttc attgaaaato ttataacaac caactgatag 44220  
 agatagaaga taaggctatt aaattgttca cacagatgcc ttgatatcct acctttttcc 44280  
 cccatatctc cttttatgtg agaaatgaga tagtgattta agggaaaaac ttaaaagagt 44340  
 toogactatg ttggtttttt ttcccccagg tcaacotta tatottaott aaatcttttt 44400  
 cttttttatc ttttcttttc tttttttott ttccctccct cccctccttc ctcctcctcc 44460  
 ttcccttctc cctcctcctt ctgctgcttc tctctctctc tctcgtttcc ttttcttttc 44520  
 tattottcct ttttcttttg agaccaggtc ttgctctgtt gctcaggctg gagtgcagtg 44580  
 gcaccttctt ggcttattgc aacotctgcc tcctgggctc aagtgatcct cccacctcag 44640  
 octcccaagt agctgggacc acaggcacgc gccaccacac tcagctaatt tttttttttt 44700  
 ggtagagatg gggctctcta ggcgtgtctt gaaactcctgg actcaagcaa tcttcctgcc 44760  
 tcagccttcc aaagtactgg gattactggc gtggggccac atgcctggct tgaatttttt 44820  
 ctatggcttt attctttctc caagtacaga gtctacccaa cctctctgaga tctttggttt 44880  
 tcttttctta ggtaactata gtacatactt atttatgtta aacaacagca atcacacatt 44940  
 tctttttcta tacagtcatg ctttatagga aaataaagcc tccgtcttag gctttctgga 45000  
 ttttttcaaa agatgcaatt octggagtat gtttttactt agagcaaagc agcctagtct 45060  
 cctatacctt ctgcatctgc agaaaagttg gttaaacaga ctttgtaatg atgccoctta 45120  
 caattctgaa gggacttgtg aaatagtttc acagagtttc agtggttaggt atatttgatc 45180  
 aatgctaact ttgggaaaac ttggtgcct gtatgattca gagggtaggg cagaatatta 45240

aattaatcac aacttcttgt attttaacca ttctgggtaa attgggattc cgtgacgccc 45300  
aggcaaaatt atttgtttat agaagatggg ctgaattttc catcgtccat ttctgagaaa 45360  
ttaggtaggt ttagaaagag acaatcaggc ctcttcttta acagaaatgt ttgtgtctac 45420  
taggtgtgtg tcacaatatg agttcctgaa gaaataagtg tcggtattg ggttgatata 45480  
ttgtacttcc tttttcttta ttttgacat tttcttgga tttccctttc tatggtgagt 45540  
ggcttctgat cgtctttcct tttgtaaagt gtaatgatat gagaatcata atcgtggtgc 45600  
ggtcttttgt gttgcataatt ttaggggggt cagtatgaat ggcccggtgt gaggtgcac 45660  
tgaaagatta ggagcagcca ccttgatgog gaggaggtt agtgactttg gacatgatgg 45720  
gotatggctg gctatactct cagctttggg cgcataagca gagtattgat tttgtatttg 45780  
gttaaaacca gaagtacaac tttctggcac cagaggatta ggaaaattta acagcgga 45840  
gccatcatga ggatagtaac caattaattc gatttttttg gtcagacatg gctccacct 45900  
gtaatcccag cactttggga ggtgaggtg ggagggtcat ctgaggtcag gagtttgaga 45960  
ccagcctgac caacatggta aaaccgatc tctactaaaa atacaaaaat tagcctggcg 46020  
tgggtatacg cgctgtaat ccagctact cgggaggtg aggcaggaga atcacttgaa 46080  
gctggaaggt agaggttgca gtgagtcgag cttgcgtcac tgcactccag cctaggcaac 46140  
agagtaagac tgtatctcaa aaataccata attcgttttg tctttctttt actttttct 46200  
ttcttttcc ttccctctc cctccctc cttcccttc ctccctttc cttcccttc 46260  
catctcttcc cttctttct tttctctct tctctcttcc tttcaacagg gtctcgtct 46320  
gaaccttttc cagtcagaat tgcacaggga tttttagact tccattctgg aaaagagggg 46380  
gtagttattt tggtagagatt gtggtcttgt ggttagacct tgtgatgggg gcctcagcca 46440  
aagggttcag gatttttttc caagcttttc cctcacaact tgagttaatc cgaaacgttg 46500  
ctattagacc accggacatg cttttctgca tgcctgtgtt gggctgtttg gattgaaggo 46560  
ccagcaaggg aaggcaccct cgcccatctg acacaggcag gcctctacaa ttttattccc 46620  
taaccagggc atgacaaact atggcccata gacaaaaatt ggcttgccac gtgctttttt 46680  
ctggcagtg agttaagaat gactttttat tatcattatt attaatattt tttgagccag 46740  
gttctcattt tgtcaccag gctggagtgc agtgggtgaa tcacggctcc tgcagcgtga 46800  
aactcctggg ctctagcaat cctcctgcta acttttttta tttttgtaca gtcttctgt 46860  
tgttgccag gctgggtctg aactcctggc cttaagcaat cttccggcct tggccttcca 46920  
aaatgttggg actacaggcc tgagccgctg catccagcac ttttattatt tttaaatggt 46980  
tgaaacacat caagagagga ataattttt ctgacacagg aaaatgatat gaaattcaca 47040

ttccagtatc tgtaaataag cttttattgg agcacagcca tgatacaaga catatactga 47100  
 ctgcotgtgg ctgctttcga gttacaatgg ctgagtcgag tagttatgac agagattgtg 47160  
 tggggccgcaa agcctaagat atttgctgtc tggcactttg cagaaaaagt ttgccaaccc 47220  
 tgccotgaac aaataaaggg acaaatcca cttgccccgt ccatctgtgg agcagagtca 47280  
 ctgaaaggaa atactggaaa tactggaagc cacttggtgt tttatcaagg atgtgaggtt 47340  
 tcctggcaac tttgtcgcca tatcatcaco atcatoacca tcatcatcat catcatcatc 47400  
 atcatcatca tcatcatcat catcatctgc cctttaagtt ttctgcttgt ttagaaaaga 47460  
 aattttataca gagccccag tagcagctgt aagggggcag gttcttgagg cagcccatcc 47520  
 tcaacattct tgctgtgat ggaagattct caaggatgaa ggccctcta tgggagcagg 47580  
 atcagtctgg ctttagtaga tgccaatttc tgctaagact atttctaaa ggagcctctc 47640  
 ctcatctgcc tttctccct gtttctattg ggggaggtgg aagaggagaa aaataattag 47700  
 agatgtctac cttttctttt ttgtggcaa tttaacagtc ttttcagctg ctttgattcc 47760  
 tttcaggcca ttgtgttgt atatatctca agatttgctc acaggtccaa agcttaactt 47820  
 aagctccctg agacatatca taaaatatga ttggggaaa aaccctaag ggccatgatc 47880  
 agaacattat tattcaacaa aggatgaaat gottaaagca agatggcctt ctttctttct 47940  
 ttctttcttt cttttttttt aatgaaagt gagcagactc ccgtccaaca gttttcaatg 48000  
 taggaattcc cacagcccca tttgattgca gtttggtgaa aagtttaatg ttttttagg 48060  
 caattcataa tttccacatt gaacagcctg agaggaagag agctggagcc cactgttgtt 48120  
 tttgtagtgg gatgggtgga actttttttt tccctcccc aaaaggatat aaaactaagt 48180  
 cagatggttg ggaacagtg gcacagggtt ccagcccttt tgtaaatctg agatgcccc 48240  
 tccttttagt ctcccttag gacccaacag aatagaaatt cctgctgctt aatgtctcca 48300  
 ggaaggaaaa aaattttct ctaggtgtga atagtacctt atttctttt tcttctcttt 48360  
 atttatttat tttccctatt aataagcacc aattgtagaa gatgaaggaa gctgggaaac 48420  
 ccatcacttt tggagaaggt taatagcttc ctttagaaaa tctgacata atacttattt 48480  
 ccccaaaagg cacttcatca gctgaatgc cagttaagat tcaaggaaat ggcttgatt 48540  
 tgtgtgtacc cagcggttct gtggcatcaa gttgcactgg gaaggagagt ttggggctgt 48600  
 cactgtggag tccctgcaag tcagcaggac cagggtgtc ttctgcaac atctggattt 48660  
 ggtagctct ctctgggcag tggggccgag totcatttcc tocaacaata atgttatata 48720  
 ggcaatgac ctgggctgco ctaacataat tgaaaattat gtgtattgta ggcttgagg 48780

gotgaaatgt gggctoataa aaatatgttg tgcaggtagc ctatggagat tggatgtggc 48840  
 acacaatgaa gcttttatgt aaagtaagaa ttataagtct ccatgttaat attgtattat 48900  
 gagtatgaca gttcttgggt gggctcctcag ggcaggctctg tcaccttcaa caaagcccg 48960  
 gtttcctaataa totacagagc tgggtatttg atgtaataca atcggttttg cagggtggcca 49020  
 aagatgaaaa cttgtocacc aatccagctc tcccactga gggatagcat gggatgtaga 49080  
 tgggtttgac tccatttggc atttttgttc acgggttttt atgagatgga gaggtgagtg 49140  
 ttggtgggtg tccatttggg ttggcctcaa ggaaatgact ctattgagtg gttttgacca 49200  
 atgcagctca tatagttagt tggtaagtga gaatgggaag aagttgggat gagatggggc 49260  
 agtttagatt cccagagccc tctggcctgg gttacagatg gagactggaa atatttactt 49320  
 tagtggttct caacttgaga tgatactgct cccagagaag gtatttggaa gtgatgagat 49380  
 ggtaaggata accaaggggg ttctgttggt tatttactgt ctgggggctt ggagtcctac 49440  
 aagtccttca gtgtttggg cagactcccc acctaatacc ctgtcgaga taggacaact 49500  
 cattcagtag acagatgaaa aaaacagaga tcaactgaag aaggggagtc gatgcagggt 49560  
 cttgtggcaa gatgcagaca caacgggact aataactagg ttgctacca cgggaggcct 49620  
 ctaggtgaaa gctctgaatt ttagcagac acaccacct cgtatagatc ctacagctca 49680  
 tgggaaaatc gactgtgtac tttggcaagt agttcttggg caatgatctt ccagctttag 49740  
 gtataaccaa atttggttg aatttgcaa gcagtcgtat ctctaggaa ctccgtoggc 49800  
 tggctgttg atggcttgg cactctgtgc tctcgtgga tttgtgcaa cccttcttct 49860  
 tgtattatcc ttctctgtct ttttcttctc tattgaaatt gttctgacca tcaagacct 49920  
 actctgtgca gccttcccca gtctattgtc ccagaaatto tgtcatcttt ctgggcattt 49980  
 cctgagtccc tgagtctctg tcacagtgtc accatgttct gtcttgattt acctgtgtct 50040  
 gtaaggctcc tcatgtgtgc aaaactcccc gagagcggac atctttgtct ctctagtgc 50100  
 ttgtcacagc ctgtacacaa agcaagtagt actcagtgtt cattgagtaa agtttctat 50160  
 agaattaata ttaaaaccag ccattttatt tgettggaga ggtctccgaa atgaccaagg 50220  
 tgtctcctta tatcttatat cccctccaag cattcattaa ctgatggatt agtgagttgg 50280  
 ccttgagaag cataaaggct cgtctccatg tgcttctaag cattgtgtct aagttctgtt 50340  
 tggtttctct agtgaaactg tcttaatgtt accaacagaa gttaaatgcc taagagtttc 50400  
 ttatacatgg gctgagtacc tctgtgactg ggcaagccac ctccactcat ttaccttgt 50460  
 ctgcaaaatg aggaactggg tcaactcact gttcaaactc cactgaaagc taattgatcg 50520  
 cttttgacag aagtagctcc cttgggocgt atatttattt cctagcttgg aggaaggtgg 50580

ggacagacag aattgatgta cacotttatt tttatctcta tggtaaacct gtgcatacta 50640  
 aagcattcct ctggtctttt gagatgagtg tatacattgt gtctggccct gtgcattttt 50700  
 taccaagaag taagttttgt tgagtaaacct tgggttgat gaagaactgc atgctcaccg 50760  
 tactcaagta gcttttgcta cctaaaggac agctgctcat atgtacttga cttccttta 50820  
 agtgaaggat gatgacattt gaaaaacgga ggttgaaaag gagcagattt ggaattgatg 50880  
 gtttcttagg acacttctgg cttgagattt gtgttttact ttcttctttt ggaatagctc 50940  
 tataattctt cctctccctc cccacctctc ccactccctc ccagccccc ccaagttaag 51000  
 gtagtagtaa tgaatcatt tttctgaag ctaccctgta ctttgaatgc aaagacaaaa 51060  
 aatacagttg ctagtaacat taatcttcta tatgtgtact tactgaactt gagctctgag 51120  
 gaagacccta ttggaattgc atgctttttt atttttttta tgattatttg catgcttgta 51180  
 tgtttttcag tttctgacc atgtccagct tatttcttgg gctagttgtt ctgcatttac 51240  
 tttctgaatt cattgttttt catttcaatt ttgtttctct tcgccagtat ctccagatga 51300  
 aatggccact gcttgatgct caggcaggga gcctccagag tagacaagcc ctcaaggatg 51360  
 cccggtcccc atcaccggca cacattgtog taagtaacct ccagagatg atggcttctc 51420  
 ttattgaggg ggtgaaaaag aaaatgcttt ttgatgata acaggcctta ttgtctattt 51480  
 tttctttct ttaaacacat tttcttggga aatattgttg ggtatagttt atatctataa 51540  
 ggtattcatt ttctgctatt ggacctaat gattgtaacc tacctggaaa tttacaaac 51600  
 ctttctcca ctcttttcca tgtatttgg taaaatctag ccttgtgggc tctagtttat 51660  
 aggacacaat caccatggta tggaggagac tagagggtgt atcaaagcag ttataaaaaat 51720  
 acattcaggg cagggtgaagt gaagaagagg gaattagaaa actcaaaagg gggtcctgga 51780  
 tttgaaactt gcttattatc ctctcccca atttatctta atatttgttg gcaacattct 51840  
 acactaacat tagaaaaatt tcatctgggc tggctgactt gtaaacctag agtagaaatg 51900  
 aactttgaaa ggctaaaatg gaatttaac tatacatcca tggctttgaa agtatgtagg 51960  
 tttgatagag aaagcatttg tttttagta taagagacta caagtgtgtg tctacatata 52020  
 tttttaatgt attttcttag ggtttttag gctctaagag tgggaatttat aaattaacct 52080  
 cttgagaaga tagctcagcc ttatttgaag attcccttct atgtatttat atcatgagct 52140  
 ggacttcata cttttgaaat aattaatgga aggcattatt ttataatgaa tccatccatg 52200  
 acaggtagaa ttatgcaaag catgaatcaa tcatgggttt ttcatttgag tatcacaaaa 52260  
 tgtaaatcat aaatacattt tgctctata ttgtaatttc taaaaattgc aaaataagtt 52320

tcttaagtag aaaaatotta agatgcatto tgccattttg ggctaactgc ctccattatt 52380  
 tggagccttg tgaattgag catgtgttat ttaatgagtt atacctctgt catatgtgtg 52440  
 tgtttatato aaaaaataac ttatttttat aaaaccatat tttagtcat catttgtgac 52500  
 aatgtcttct tttctctggt ataatgagg catgtagaaa gaagattgac atttgctaga 52560  
 agcttccctt ttctctaac tcacacataa aatggatgct cataattaca tctgctccta 52620  
 taaggccaag atttcagggc tggaagtgc cttagatcat ttaggcccaa cttgccctca 52680  
 ggaaaggaaa ctgaggccca gagatgcctt aagtgaattg cccaatgtca cagctgagt 52740  
 cagtggccag agcaaggctt ggatccagtt ctctgctcc tttccagagc cttgtgatgt 52800  
 ottctctct acaggagggt aaaaataactg ctgtggctgg ttctgttttg ctgactgtaa 52860  
 attgggtcat ggtcaggagc agtgcatagg tgtaagaag ttgctggttg ggggttctaa 52920  
 tgcaggtttc tccaaaagt aatgcctgt taaaaaaaa ttcttaacaa atatacagag 52980  
 atttttttt taaaaaagt tgacagttct agacacctag agagtaaagt gaagaagcct 53040  
 gttttcaggt ttcccgctc cctgaatttc ccagcatggt ccaggctttg aaatttattt 53100  
 atctgctttt ggcaatggtt gatgggaatt tcccacattt atttttttagc tacagagaaa 53160  
 ggacattatc tttaaaatct ctctgtgtt ctctctctt gagtgaggag agaagatgtg 53220  
 aatcotggca gtggttcaga gtggacacag ccctgtgtt tgtggcatag gctctgtggg 53280  
 ccccatgcca gggagcagta ccccggtga aaggagtggg ggtttgtcca tttggataga 53340  
 gcaagatcc tccacctcaa atcccacaag aacagttgcc acaacctggg ccctaagcat 53400  
 ctcattttcc tatgtagaaa ttaatgatct ggaggagatg gcaaacatt ccttccagag 53460  
 cctgtgtgga ttttggccag ggtgcagca agggggctta ggcacctttt tctctgctg 53520  
 tgtcttagca ggcgtgttga ccatagcaac tcccctggg catacacacc ctctttaga 53580  
 tggagacctt tgtccaaagc agccacagct ggcaactgtc tacaatcttt tgggctttct 53640  
 gctgtgctca aggggatctg ggaatggcca ttgcctagag gggatgggct ggtggaggaa 53700  
 ggtgggctct gggagccggg gagaaggaa aagccatgaa tttggacaaa aggacaaatg 53760  
 tggtttacat ttgtgaaata cttgaatgct tgcataaat ggtgactttg gttctatgag 53820  
 tcagocctgt gatgggttat ttctgcagtc ttcacotgac accaggggtg agaaggagga 53880  
 tttctgggga ggaggaaaga gttgaggag ataggaaagt agagtggag aaaggccttg 53940  
 cgttgtgac ctctatccac ctggtcacct atagtttttg ggattgagga tgcatacac 54000  
 ttgagactac aaatttatga ttatattttt gctgaacata aggcaatgtg ccaacaaaa 54060  
 ccagctgttc tttggctggt acagtgtgtc tttgtttgta aagggtgcat tctgaatggt 54120

ggctgataca tcatttgggt cttgtacag ttaaacattg gccagagggg ctggttogtg 54180  
 ttttagagtcg ccgatgaagg gctaaccttt ctcagacac ttggggctct tgttcacact 54240  
 ttgcttttca ctcttttaag taagacatag tcacatcaca gtgtttcatc agacatgttt 54300  
 caaataaatt gtctaaggat tgcctcttaa ttccccgaa atttggaatt gttgtaactt 54360  
 ttggggcaag ctatttcata attatttcta atgtctcgtc tgaagaatag ggatgtatc 54420  
 agtgttgatt attaatcatt cgaaactaca actttacaga ttgctaagaa gaataaactc 54480  
 ttccagtacc catatggggc agaactctca cgtgggaatt cagagcattt tgttgacta 54540  
 ttttaactcg attggattat ttcatgtgg tatgtgggtt accacattag aaacgattga 54600  
 tgtgtagaat aaatgttctt aacaagtga ggtcaactta tcaaatgata ttacattaa 54660  
 gaatagactc cacaattttt agttcctgta gctgatatag catctcattt gttatataat 54720  
 ccagtgttc ctaatctgtg ttccaggag agaggaaatc gattgcaaca gggacgatgc 54780  
 cttcattggc tggcccaaaa ctgggagttt atacaaggcg tcagtctttg ccttctcct 54840  
 ccctgccttc cctcttctt cttccttccc catactccc aacaaattca tggacttctt 54900  
 aacaactcag agacattagc cacaagttcc aagacacccc cccccccag cctccccagt 54960  
 cctattttcg cattoatata actaaactct ttttcttctt tgggtggagt ttgaaattta 55020  
 tatttttaat tctttgtccc ctttttctt cttacaaaat gagtccaag cagctaagtt 55080  
 gtgctgagtg gtagagtttg agtcagttt ggtgggtaag ctgtgggtt aggagccgct 55140  
 ccctggatcc cactctgtgt gtctttgcta tacaagact ttcatttagc ctcctttgta 55200  
 tccagcaaaa aaagattcag taccacaaat ggtggtattt tggatatagta tgtatcttac 55260  
 aaaacggcaa aagacttcaa agttctac aattttatct tgggggtttc cttttgaagt 55320  
 cgatgtagaa ttttaccttg ggggtgattt tttgtacttc ttggtctggt gtgttttgtt 55380  
 gtgtaatgag catggaggtg tgggataaga aagcagactg aatcccgagg aacaaagcct 55440  
 gccagactgt ggtggtgtac tttcttgtt gttattgctt aaatgctgca agagagtga 55500  
 aaactottac gaaataatgc acgatgggta gaacttcaga gaaaatctct gcctctacc 55560  
 ctgtgcattt tcgaggaagc tcagagggca tgotgaacot ttgctttttg tttctgaaga 55620  
 gttcagggga acctacccat aattaatttt ttaaaacact acctagagag caccctcttg 55680  
 gttattaaac acatgcgctg ttccgatggg atgtttgacc tggattgtgg atgcttgctg 55740  
 ggacgtggca tgtgttggga ggctctgtgc tgctgtgta gcaccagcaa agccacagt 55800  
 gccctacct ctgtgggagg ccctgtgcca ggtgccctca aagagtaggg ggcccatgag 55860

ggtatgacca gggggacctg atttcggctg agaagttggc ggggattaca ggcctgggog 55920  
 gctccctgag gaaattgcat taaaaatgag atctgaaggc ttgattgggg ttggcccaat 55980  
 gaagggatag gagaagggat ggggagtggg cagaaggaaa cacatgtgtg aaggtcctca 56040  
 agggaaaagt gcttggtctg gacagaggca ggaatcagg taggaggcta gaggtcgggc 56100  
 agggctccgg gagagtgaat tggggtgcag catatggtga ggatctgaca ctggggagtc 56160  
 atttgagcag gttggctgtt tctgtaggag cgtgtgttaa gctgctggca gtggggatgg 56220  
 tgaaaataga gatgtggagg aaacagcagc ggaacttgot gacaggttag atattggcat 56280  
 tgaggagaa aggagagtca aaggtaggta gatggagatg cttcactgag tggggagtat 56340  
 tggaggagga gcaggtttg ggtggaagcg ttgtcctttt agagagattg tatttgccat 56400  
 tgattgattc attcattgtt tctgcaaata tttagtgtgg gaaaaagcat gctagacacc 56460  
 aagagagagt ggagtcaatg aagaacgata acagcaacaa agactgtagc gcttcctatg 56520  
 cgaggcttgt tccagttgct tcagaggctg tgttaccctt gttctagaga ggaggaacta 56580  
 ggcccaggga ggtggggatt tgcccagtcg tgggagtcag gatgtgaaac aaggcaccct 56640  
 ggctccagag cacaccgtcc tctcaaccac tgcagagaag ctgggaaaga gacaaataag 56700  
 tgggtgctta gagcacaatg tgtgtggtgt gccaaagaca gctgggagcc ctgggacccc 56760  
 cagggaaccc cagccccacc tgggcatggt gggcatggct ggaggaggcc tgctggcttt 56820  
 gctggagagt gggacatgca tcaagtggtc cagagactgg gcttctgggt gtcgtgctgt 56880  
 gactgctgca aagggtcat tgacatatgg tggggagggc cagcgtattt tctgcgggca 56940  
 ggacatttgg gggatatggg gtgtgaccct gtactatcta aaatctttta cttctggatt 57000  
 atctccactt tctctactgc atatatactt tgtttttatt tttttattc atttatctat 57060  
 gactcagcca gactctctaa aagagttgac ttgtgtttcc tagcagccac tgagtcagaa 57120  
 ctttccattt tgcagtcag ggtgtggtc aggggtgtctg tgtgtctaa ggatataaag 57180  
 caagccttcg ggcactacca aaacattatt ttataaggag aactatgagt acctaatagg 57240  
 aagaaccagg caatcaggtt atcttttggg gaggaagaag tggtagatgg gatcattggt 57300  
 gctttgaagg gagtgggtgg tgtagactcc aaagtgtaca tggggccatg atagagtcta 57360  
 tgtcagatgt ccaaagcttc cttctctctc ccagaaact ctgtcctctg gtgaagagtt 57420  
 ttgaagtttc ctgaggtttg ggttcattgt gtggcagggt ataccatggc aatagaaaat 57480  
 atcccatcaa gaaggattgt gtgacctcag ttgtagcccc tgcattgttg aatcacaaca 57540  
 atttgagggt ccttaaaatc aaatgocatt tcaccaactg cctcccccg tttttttcag 57600  
 cactgttttg tagctatctg tttccctga tattotttga cacttccaga gatgggggct 57660

ctatctcctg gtggttagact gtttcttttt ggtacaatat gaactcttaa gagagttcta 57720  
 ccttttagga gctgcagctt ctctcctgga aatgetcaac tcottaatto atgttttgct 57780  
 gttaaattct gctaatagct caccctacat gtcttgacaa tttgaaggta gctattgtat 57840  
 tccccgcaac cccaagtctt ctcttcaaaa tgattattaa ttgtaattca aatcatcagt 57900  
 gactggtatc ttagactact taaggatggg aattgctaatt tttgtattta aaagttgtac 57960  
 ctotaaagta agtgaaattt atttttaaac gtagctttct tcattcataa agtttatgtt 58020  
 cattgtaggc agtttggaac acagcccata atctcaccac tcggagatta cattgtgaat 58080  
 aatttggtat atttcttttt agaaatatac caaattatcc ttttttcttc tgagtgtatg 58140  
 aatatttata tttgttttta acatacttga gctcatagtg ctgagtattt ccaacgttct 58200  
 gtttatttaa gatgaaaatt gctgtagtta ataagcactt ccccatgtca ttaaaatgct 58260  
 taaggatttt taatgaccac ataacagtcc ataatatgat taaaccccaa tttactgaat 58320  
 caatgccata ttgttgggtc tttagattgt ctctttttgt ttctgtact gtgaatgatc 58380  
 ctgtgatgat catcttttggt tgtaaatctt tgtccctcgc cccctctccc ttttattatt 58440  
 ttcttgggat agaccccagg acaaaaaggta gaaaagaaca aagtgttaaa aaatttcttg 58500  
 atacatagcc acagattatt ttctgaaag ttctcaacat ttataactac gagcagtatg 58560  
 taagagagtt atggttggaac tgattttaat gtctctgggg aatttaacaa caaaaaaact 58620  
 ttaggcttct ttggagagag acatgcctt aactccacc cgccttagaa cagagacca 58680  
 gcccatccaa gtcagcctcc ccaggctctc cacttcaaaa acaggcaaac gaaatcattt 58740  
 cttgaataat tggtaggctt caaggtcaga tgtttatttt agataattca cagcataaat 58800  
 ttatatgttt taggtacctt agccctgaa tatactcagt tcatttagga ctattttaga 58860  
 ggtcttgagt ttactcttat aacctcacat ttttttgta attttttagt ctattatctt 58920  
 tgtttctatg gcataattatt gggcaaagat actatttatt cgatgctatg tgtgagctgg 58980  
 gtcaggatta tgaccctgag ttatgtttct gggaaaatgt acccacttgt caaagatgcc 59040  
 gttggctcct gtgattaagg tcagcccaca atgaatgtgg ggagggctgg cagcctctca 59100  
 aatcagctct tgaccatttc tcaagctggg gcctgttggt cttgggggaa gagtctttgg 59160  
 cagctcagct cggggctagc gtttctgac atttgttctg ctgaatgtta acaaggttac 59220  
 tggaaaaag ggttctctcc taaaataggt ttagggaagc actgggatat gcgaagtga 59280  
 tgagtttctt tagggcagga tcttgactct gcagggggtc tggaggcctt cctagagtg 59340  
 gggcttcta aactgcaga gctcttcca ggaagagggg caagattggg acctactttg 59400

gaaggttggt tttgtttcgg cactgtctct gtttacgaag cgtgggagcc tgttttaaat 59460  
 taatgtggcg ctacttagag ctacactcat ggttttgact atgtttatct tccagtaaa 59520  
 taaaacaaaa ttgttoatth ggcaaccagc ctgtcctgct tgtcatttct tgtcttgctg 59580  
 attaaactta tggatggggc atgtttctcc aaccagattg taagtttctt gaagccaagg 59640  
 agccctgtgg ttgatttctt cacatgtggc tctctctcct cccacaatgg tgccttcgta 59700  
 attaaagcaga aaacccatct ctgggttaggg actggagtgg atttogtttg gaatgagtgt 59760  
 gacttcatca tgacctgaaa gtgttcagaa ccatcttggt tagcacaagg gcgtggacgt 59820  
 gtgtctactt tctacctgat gggatagcat gttaatttg gggttatgac actgaatggg 59880  
 ttgccagtaa ctgtctaac caacctata cattccagct cacagtggag cgtgtctaata 59940  
 tgccacagca gcatttatgt ggaacgtggg tgcacaaaag ctccagaaag tcaggctgag 60000  
 ggctcctatc tctcctcaat ctgggtttac gatgtctgtt tctgaggaat cctgggatgg 60060  
 ggccactggc tctttaagag agagcccgat ttggaaatct aggacttgat tgttgattat 60120  
 gggcaataga tacattttaa gaatgatgtt gtaggctgta tgaagtcat tcatgattgt 60180  
 tttgttaatg gcttgcaggc cagattttca tcttttttaa ttaattatca tagaaggaga 60240  
 aaacaactgg atttcagaat tgtccctga ggtgtactgg aaactaaggc gtgagggact 60300  
 cataggggtc tggcttgga agtgattgc tatgtccagt ttacacataa ggatgtgcaa 60360  
 atccagcagg ttagctgagc tgcccaggaa tatccaggca agaataacca tattctgata 60420  
 attactcagg cctctgcctc atctccgctg cccccccgcc ccctgactct cttctgagt 60480  
 ccagattcag cctccatttg aatgccaaat agacaggaaa ttagcatgcc cagaatccac 60540  
 gtcttttagt cactctctcc ccagctccaa acctgttact gcttgtgttc aacatctcag 60600  
 taaagctcaa caacatcgac ccattactta ggcctcaaac cttgggtggc atcgtcgatt 60660  
 gctcttttct ttcatacccc acattcaacc catcagccca tcccacaggc ccaagtgtgt 60720  
 cctctctacc ttcaaagcgt gtgtggcatc caccgcttat caaccctct gccattacca 60780  
 ctggagtcca gtgccatcat ctctcaattg gatgtggcca gagtgtcttt gctggtctcc 60840  
 ttcttgcttc ctacctttgt aacagocctat catctatctc tggctctcat agctcactcc 60900  
 catactttga gagggccttt gaaagcctta gacagatcat atcacagacc tctatactga 60960  
 aagtcgggat aaattttatc tctggaaaga gtcccaaagc agcgatgaac agatattttg 61020  
 tctgtcact tgatgaagag gtggggcttt gagaccaag agcttagaat ggagagccta 61080  
 gatgccacta agcccaggca ctggccatgc ttogagtggg gcttttgtgc tgggtggagga 61140  
 gagatggctg ggggacacct gtaggctgag caagtcccg ttcatcagac cctggctcat 61200

ccagcagggc gtggtgatg ttttcaatgt tgtatcctga gtgggaccca gatgcttccc 61260  
 aactgtgcc aatctgagcc ctgcatgcc tctgtccagt tgcagcctga ctgcaatgtg 61320  
 aggctgctga agagctctgg atggtgtgaa gcaatctgtt ttctagcccg agcctgcata 61380  
 gctggtggat cctggaccgt gattaagtgc atcacctagg cttcaatgag atggagtcac 61440  
 tgtgtgtcca aacagtggga taaaggcttt actctttgtc ttctgtctct gagggcacia 61500  
 gctgcttggt tctctcacia ggacacgctc tgtgttgctc aggtgctggg gtgaaaaaaa 61560  
 cagcaagcat ttgaaaaggc tgaagaagga aagaaagctg agagcgtac agccttgggg 61620  
 actgagccat cccattgtcc cagaggtggg ggtgttatca agacctgttt ttgagccata 61680  
 cctctgactc ttcttgaaa gttagacca actcaagaac aactaagag aagtgtttcc 61740  
 ccctgacctt ttcagattga aaggagacgc caaccttgat ggggtggagg agaaaaataa 61800  
 gtocaaaaac agtgtcttgt aagcgaaggg gaacatggct gggcagaggg cttctggtga 61860  
 aacttttggg agtattcagt tggaactcag gaaaaaaaaa ttgttttttt ggaaagagg 61920  
 agcagccccc ttcagccaaa gctcataaat gaaggaatgt ctgagactca gaattacagt 61980  
 gaccaaggca agacattgtc aaaggctgaa taagtgagtt tgactgacag aggccatctc 62040  
 cttttttagt atatggccaa gcatctttcc cacagtcttc cttgagcccc tccccatccc 62100  
 acttctgaaa agcactgagt tggccattat tatgtctttt tottaaattt tgaagttgtt 62160  
 ttcagggtatt gagaataaca cccagggtct gaactcccag cataagaaat caaacattca 62220  
 aaatggagta aggttctgaa gctgacatct gtctctacac attttttttt ttctgataat 62280  
 ggcatttctc atctccaccc tcaactcttt tgttgtgtg aactacaott ccttgttcc 62340  
 actcggttct gttgcacatg tgattaggca aggggcagat atgtgatatt tattatgagt 62400  
 cttttccacg cagagaggat ctaaactctg ctctttgcaa ttgcottcat acatgtgcat 62460  
 acacaccaca cacacacaca cacacacaca cacacacaca cacacagaca catacatatg 62520  
 cacacacccc gactcaatgg aggacccca tttgtagaag ggtaaaatgg gtgaggcgga 62580  
 aatgcctgta tggcaccatg gagttctgtg tagccagttc taatcctggg ctatttggt 62640  
 aggaatgaag ttggagatag tcttctgtcc cttacaacca aaggaattct aactaatagt 62700  
 ttgccaagtt ttatgtttat aataaaaaat gacatgcttt ttcttttgga tttttaatgc 62760  
 ttttgaatta aaaatgctag aacatgaact gattcttcta tcgtatttta gatagagcct 62820  
 tgcaagagca gagcacgcat gctttcttta agaacagggt ggtttgtggt cgtctgagga 62880  
 ctgttttaag gagacttatt atacacaatc atccccaca aatgatttct aaagagaggc 62940

tggatatgaaa gaaggagttt ccatgattct gtcoctgtggt totggggaat tctgaaaatg 63000  
 aacttttagat atttttgtga aattcttatt ttcatatttt tggatatca gagttttctt 63060  
 ttctggcttc tgtttaacat actcttcttt gccctaaato tctcttattt ttgctccttg 63120  
 ggacaactga agaactccta gataattaat agtatgaaat actgcccttt tagttgaaaa 63180  
 atgtcacaaat aatgtaataa gataaataag gaggtgtcgc tttaacctgt atcgtgtagt 63240  
 ctctctact tactaacact tacttgatt actagaagca ttatttttta aatcatggaa 63300  
 aattggtggc aagctgagca tacagttgtt tatttctgtt tgactgatta ttacaacttc 63360  
 attatttgat gaagggtctg taogttttcc tttaagacac atagaaattg tgagaagatc 63420  
 ctgcagcccc gaaaggctac agtggtgac caaggactct gagccgagtg cagggtttgt 63480  
 acttggacct gcaggctggg tggcgtctgt gggagcagtg tgttgagaga gattctgagg 63540  
 ctgtatgtgt cagggcctcc aggggaagga tgcattgatg gattaatttc tgccaaggct 63600  
 gaaagaggag agagtaagag gctgtagagg tgcacagct gtcattgctg ttttaggcag 63660  
 tcaagctttt gggaaagtgt cagaaattga gcccctact ggatctatcg gagcctgtc 63720  
 aaatgtccat ttagatgtcc tggngaaca aagttctctg actcaccatt taaaaacttg 63780  
 ttccaaatga aattatggga gaaaggaaca tttttcatcc gaaccagaa tgaggatgta 63840  
 cccaagaaa aggacgtagg ctgagagct ggactgtggc tcagctggcc tgatgtatcc 63900  
 cactttgttc ctcccatggc tgggatgtct ctttgctctc catgaoccat gtatcttgag 63960  
 gacatgacac atggaccaag cttgaactgc ggattcattt ttatgcatto tacctgtgaa 64020  
 tgattgcagc ggatctagtc gtatttctga gagttactca aactggactt cagcagtga 64080  
 ctctacagtt ctcttttctt cccacctttc tattagacat tgcattgata aaaaatcaag 64140  
 atatttttaa gaggtgata acttcaatgt tatctaaact tttaatttgg aagaagaggg 64200  
 gttctttgtt ctttttaaaa agatacaaac gaacttcttt atctgattct ttttttggtg 64260  
 caaaccatg atgccttctt cctgattcat ctgctacact gtgagttcaa gcctggcgtg 64320  
 ggacacaggc acagctctca tgccaacgat ctcatggtta agttttggaa cataatttga 64380  
 aaaatgtaac ccattgagag gcagtaagga catacgggtg gctagtgcgt gtttggaagt 64440  
 ctgtgtgga taagttagtg ggtagagagg acatttctca aggagcggga gggcgggcca 64500  
 ttggcttggg ggaatgggc tgagactcta ggggtggcca gcaccgata cggaggccag 64560  
 cagggttggg cttggctaag tgcgtggtg tctggatgcc tatgtgagtt tctocagaa 64620  
 gttttcagtt ggcaaagtag aacctgctg atatgtagca aggggtgtgga ttgtcgggat 64680  
 cctgctgggc gcaggcgtgt gataccagag gtcagaacag aagctgaggg atgaggtttt 64740

gggagctttt tgtcatgcac tgctctggag cctcagttac tacaaagtct gcaaatgata 64800  
 gaocggagct ttggtttctgc ctgatgctag ctccctgtt cctgattttt cttttcaata 64860  
 ttagacttaa tcccagaatt cacatgttga aagaaaactt agaggtctag tgacataaaa 64920  
 gcttcatttt gatcgttaca gaactgatgc cttgagaaat ggagagagaa gtacacgac 64980  
 atggtataac tggatgttca ctgagcactc actagctcca ggctttttct aagtaattta 65040  
 tgaagtgtc aggtttaatc ctcaacaagc cttatgaat gagctattgt tattatcccg 65100  
 atttggcaga tgaggaaact gaggtttgag gggaggatga cgtactcaag gtcacacagc 65160  
 tgggagggcg caagctggaa gttgaaccca aggagtctca catcggagcc aggactctca 65220  
 ccttcagtgt ttatgtgtcc ttaatcaggc acacatacag gcggggagag gcagggttcc 65280  
 ggacaccaga ctaggctggg gccggtcagg ctacaccagg gaacctggag gcctgtcatt 65340  
 ctttttgat gctgttagt cctgttgagg aagtgaggct ttgtgggttc ccaggaggaa 65400  
 aagggtatgaa ctcatggcaa aagaaaggaa ccaaaaaagg gagatttgca tcacaatgag 65460  
 ccttctattc atoctaaatt atacctcctt ttataccatg tgtgtctgca aacttgtggg 65520  
 taaatcacia atctttctgg taagttaaa tggatggaag gtttttgcat ttctctcaaa 65580  
 tcaccaacca tttaatgcta tgtgtagtca ctccctaact tatcttttgt ataaatttgg 65640  
 atctttgagt attgggggtt tccatgatgt ttggcagttc cccttagggg gtctatctca 65700  
 aagtttgtca cactgacaag ctttggggag agaagttaga ggtgggcttc cctgttttta 65760  
 gtggctgtgt ctgattgttc tgtctgttct ccaggacagg agagattgat tgctttctag 65820  
 ctttttttaa aattaaaaca acaacaacaa aaaaatacag aaaggtaaaa aggataaaaa 65880  
 acacattcat gtacctgcca cctaaaataa caattactaa tcttttcacc ctctagccc 65940  
 atgatcttcc ctcccaggct gttattaata tgaaaaccga gttcagggtt ttatactttt 66000  
 cgacatctat ttatattaac gtatgtatta taaataatct tagtagtttt taactttgac 66060  
 ataagtggtc tcacattcca cataacatto tgcagcatgt tttcttttat ttttattttt 66120  
 ttctttattt ttaaatTTTT attttgcagc atgtttttct tattcaacat tacatttgaa 66180  
 ttttttcaac attgtacatt gaaatttagc toattctttt taactgctct gtagtattta 66240  
 ttgtatgcat atactacagc tttctatttc tgtattgatg gtttaattagg ttgcttacag 66300  
 ttttttaaga ttacagatto tgctgtaata aocatcttt gggaagtgat atgtaggtag 66360  
 ctatatatga gtttctctag gattcatacc aaagtagagg aattggtagg gcattgggtt 66420  
 gctgggttta attttaatto acatgctatt gtcaagctct ccagaacaac tggatgagtt 66480

gattggatca atgagtatct ccatcaccag catataaact ctttcctcat aatcacacca 66540  
atgcttgatc ctgttggaat taaaattttt gccaatcttg tgggtatgca acggcatctt 66600  
acctaatttg cctttatttg atgactcctg aggttgaaca tctggtcata tgtttatttt 66660  
ctcctctgtg gcttgctgg tttaatgcct tcttcatttt aaagaatcag atagtcttct 66720  
gttattgatt tataggaact ctttatataa gttgaaaact tgattatatg tgttggaat 66780  
actttttcta ggctgtgatg ttttaaaata ttgctttaga tgggttttca tttttacct 66840  
ttattttaga gatggagtct cactgcattg ccaggtctgg attgcagtgg ctattcacag 66900  
gaaagagcat agtatgttac agcctccacc tctggtacc aagaggtcct cctgccccag 66960  
cctcctgaat aggtgggacc acaggtgcac atcactgtgc ctacttttg atgggttttg 67020  
aaagaaagaa gttttaaatt ttaatgcct caaatcact tgtattttcc tctgtgcttt 67080  
tattttgtac ccactctaag tagctcogaa ttctgcagat agttggtgca ggaattctga 67140  
ttttgagtgg acatctgtct tctaacagtc acattgaagg aaattaggtt tttttggtag 67200  
gaatctaagc aaggggttga tttgtaaaact aggtctttaa tatgatttta agcaactcac 67260  
ttagaacaag atacaaaaat tgtggactgg acctatatct ggaaaacttg aaagtgtatg 67320  
ggcaataaat aattcttggc cacatacagc cgagatcctg ggctcctgac tctgggacag 67380  
aagctttcta tattttatct catcagctct tgcaacagga tcttgaagc aattttatcc 67440  
ccattttaga gataagaaaa ccagagctta aagcagttag ataatttatg aagtaagtgg 67500  
cagagccaag attcaaatcc agacctttct gaccacaaag ctggttgctg aataccgcgc 67560  
ctcattgcct tcttgcaat tacttgggat ttgtttgaat cccaaaatct ttatatgtta 67620  
ttttaaattt gaatctaatt ggaagtggg cagtgagggt agaggacaga aagaagggga 67680  
agagcttgag actcaataat agaaacaaaa aaccgtctc caggaggggc gttcaaaagg 67740  
aagaattcca tatttcatgt aactgaaacg ttaaaagccc aaataattgc atcatgcaag 67800  
tctgatgctg agtaatcacc ctccccata ttattgggga gagggggcaa gaagtctggg 67860  
aagctgtttt tgctaagga attacattcc aggggactct gaggatttag gtaaccacaa 67920  
aagccattta tttogagtac actgagattt ctaccacttt gatccataat coatagcata 67980  
attaataaat gaaatgtgct gtagcatggg ttttttacia agtgtacttt taaaatggct 68040  
tttggctgca catgattcat ttgccacttg gaaaagcgtc atcgctcag atgggcaggc 68100  
tgggagaggc tgctgtggg gtagctgagg gcggtttcct ggggcacagt tctgccttg 68160  
ggcctctaca gagcgtctc atccaaacat ctcccagact ctgcgttttc caggagcgt 68220  
gcagaaatag gagccagta ctgaaatgct atctgctctg tgtatgtcag aagaccacaa 68280

accacttata acaaatgaag atctttttat ttgttcttat ccotttatgt cacttgagga 68340  
 aagttgctgt gagtaggtga tgatcattac agtgatcact ggttgcccaa actgagaagc 68400  
 cagacatttg gcttggttcc tctcccttcc tottgtctct cctaccctgt aaacacatac 68460  
 ttggtgatta cccatgggga gacaagacag gctgggaata tatacttctg caacttcagc 68520  
 ctcttggtt ccagcgattc toctgcctca gtctccagaa gagctcggat tacaggtgtg 68580  
 caccaccagg ccagcctaac tttttgtatt tttagtaaag atgggggttc atcatgttg 68640  
 ccaggctggt ctogaactcc tgacctcagg tgatctgccc ttctcggcct ccgaagtgg 68700  
 tgggattata ggcgtgagtc accgagcctg gcccaggca ataataacc agtgggcaag 68760  
 aaaatattct tgcctctcatg ggacttctgt tgggggtcag ggtatagga ggaaggcata 68820  
 gagatgaaaa ccagtaaata agtaacagg gaaaacattt taaatacatt aataactaat 68880  
 aaaatagaaa taaatctgtt ggctacttaa caggatgtgc cacattccag atacattacg 68940  
 ttaatcotta tgatctttgg gggctaagta ttagtattcc attttacgga tgaagagact 69000  
 gaggtcaga ggaaggag gtggcttttc tcaggtgaa agccagacct ttccagtgt 69060  
 cattcagttc atagctaagg tcttattttc tgtgctctct gtcggctgaa aatgggcaag 69120  
 gtaatttcac atagtgcag gagccatgtc agagaaagag caggacagtg ggacagagag 69180  
 ggaccaggct gggggctgtt tgagatggag ggtcagggaag aaccaaaacta agatgtgaac 69240  
 agtgggaggt gttggagctg .tggtgcttgc ctagaaggac cctcatcgag caaatagaag 69300  
 cttctggcag gaagaagta atgtcttgcg tgtgccctat gtaggttcat tagggccttt 69360  
 aaaggggaa gaaggtgtg gctataaatg ttacaatctt acotttggcc cctagggatt 69420  
 ctgtcttca acottggtc agtaacaact tgtgactgcc caacagggtc tcotttcggg 69480  
 agagaatggc ttgttacatt caaatatgcc atgaaagtat caccatttat ttcagtgtct 69540  
 gatgccccag cttgggcagc ctgagcaggc totgaatggg tctgaagagg ccctttagag 69600  
 tagagatgaa gaggggttg ggaatcctca attctaaaca aagagtctgc aatgggaaga 69660  
 tggccaaatg ctgtttttg agtgggtgag agggaaaaga aaggtataga tggttcgttg 69720  
 gaaaaatgtg ttttataccg ggttttggtg tcaggtcccc gagggcaaca tggactccac 69780  
 actgtgatcc tcgggcagc tcatagcccc agccccctcc ttttgcttcc tggtcagttt 69840  
 gtgagaagga ggggttgtgt ctccaatctg agcaataagg ggtctgaggg gggttgatc 69900  
 catgtggctt tcctgtgtct tgttcottgt aaaagtcca ggttttgggt cgtgagctgt 69960  
 gtgtgtgtgt gtgcgtgtgt gtgcgtgtga cgtaatatg gagagatggg cttgggccag 70020

tgggaaatag agagaccocg aagcacagag tgacagggtt tgatagtaag cagcaggcca 70080  
 gogttgctgc ttttattocg cggtaaatcc ttgcacaatg ccatatgctc ttgcattccg 70140  
 tagctgctgc atagggtgtg atttagttaa tgcccgctct gcaaacagga aacgggtgctc 70200  
 actgctgtgt atgcttttca tggagataaa gtgtcaggag caagacccca aacctgcgaa 70260  
 atcactaatg caaccgcccc ccatacccca aaagggtgga gtgggggata aaaagagtag 70320  
 gaaagtgggtg tggggagggg aagctttagg gccataactc agacaatttg tcaggcagtg 70380  
 gcatcggttg ggaggaaaat attgatgtac actttttgtt tttgaacctg aagtttggtt 70440  
 tttttcggat gcattggagg acttttaaat gttttcggag tgccagagt tggactgtta 70500  
 ggtcacogta ggtaccggct tgcatacat ttcagaggaa ttttttcaa actccataaa 70560  
 aacatgcggc tttcaaggct ggaccacttg ttcaggtcct cctcccacc cccacccttt 70620  
 ttggcaaaac catgcaaaac ttgtatttca aaaatatttt gttacttttc ttggcaaatg 70680  
 gttccaagaa ggaattgcaa cacagtctca gagttaggag gcaactttct ggggaaaagg 70740  
 cgggggttgg ggaggtttgg agtttgaatc aaaaacagac accgaagctt taataaaata 70800  
 aatgaagcgg agccctttca gctcacgggtg gactgtgttg gtgcgcgggt caggctttta 70860  
 cgtgcctagt ggaaattgac agtctgagaa ctgggacata aacaaaaatg tcagtccctg 70920  
 ggagtcttgt tcaactggaca atgtctcaat tgttccttg gtttcaagg cagcaggag 70980  
 agtggaatat taactgttta ctgcccagg ctggctcgga aattgcttg agaaggggag 71040  
 aaaaagaca gaaaatcaca ttttttattt agaaactatt aaacatgtca gtaagagata 71100  
 ggaaaagago agattgtttt ctccttaatt atctgccatt caattccata tttctgcata 71160  
 ccatttttgg ggtgtgtgtg tgtgaaggaa cagcagggtg tttcttttta aatttgaatg 71220  
 ttagccttgc atattgtcag tttttaagc ttgctggcat gtagattatc cgcctccggt 71280  
 ggatatgaca gtgggcttta ggaaaggaag tgtgatttct gataacattt acatcttagc 71340  
 tgttcagcgg ataccctgtt agtgtttgtt cttcagaatg ctcagataga acaaaaatca 71400  
 agtggttggg attttaaaaa acaaatgtg tttggctctc cataaaaatg catttagtga 71460  
 taaagggggg cagcaagtaa ctatgtctga gagaaggaat tgcaggcaca gaggagatcc 71520  
 agaattctgt tcacaactga atttacttga ttogagaaac aaacagcaaa gcctgggtga 71580  
 ttggccttta tctgggcaaa gttcaaaact caactggtaa ttatgtcctt agaagcotta 71640  
 aaaggactgt gttgttacaa aagcagtgac tgagcttact tcttcaggac cgaatgcact 71700  
 cgagttgttt gttagataaa cttgttttaa taaatggggg ggtcagggga gaggtttctg 71760  
 ttcttgaag attccctgat aagtagcttt cttctcttgg agaacttcag gctttctctc 71820

caagcgaggg gtttgcaggc agctaaagtc agcttoggot totgcttcct gtcagtcagg 71880  
 aagtcacttc ctttaacccaa attacaagct agagcacaac tocccagoca taocgaaaag 71940  
 agcagggtttt tcccagaaga ctgtgtttct agatgcggaa gtgtaaattg gtaogctgtg 72000  
 tgatcatgga atgoccaaaa tacataggga acagtgttgt tggaaagagg ogotgtgtcc 72060  
 ccaaggagaa gacgcgccc agaatggctg gatcgctgt tgtggctgag tgcgaggcag 72120  
 ctgtggctgg ctgctgtgtg acgatgacct agtagccacc catgtggagt cctggctgcc 72180  
 tcagaacctc atcacatcta ggcaaatct tgcatttttt atctgggagg cctgaggact 72240  
 tcagggtgtg tggatagtaa gctccttggg tatctcacag atacaagagg tcttgggaat 72300  
 ccacgatcaa acttgatgtg tgcgtttacc ctctccctt tgaatctgtt attcaaatat 72360  
 ttaagcctcc aacotttggg cccctacctg caccacctc ccccccccg acaaaaatca 72420  
 agctcttgac ctcatggctt ctttcagtga cccttggggg acagggtttc ccaaggctgg 72480  
 ttgccagctg gcattgttcc cgttgggtga agtggagacc tgtgtttttt tggctatttt 72540  
 gcaaagagct tatggatgac agcagttctc tgtgcctcgc tgggacagag tgtattctga 72600  
 ggtccagcgt ctgcatggag atctgcctat ccttcacttg ggggtctcag tagataacgc 72660  
 ggccactttc ctatacatct ccttaattta agggaacagc gtaaaactcag cccagggtga 72720  
 ttaatctctc cagtgaactt tgaacttca atttccaatt tccctcttat gtctagggtg 72780  
 gagtggagat acgtgtagta attgtcgcag gtattagtga gaaagggtgc agatcacaca 72840  
 aatatttcac acgttattag ttggaccaga ctttggaggo aaggaggggc cgtgtcacct 72900  
 aggaaatttg ctcttcctg gagatgaaag ggcagtgaat taagtgcctg ctttttctcc 72960  
 ctttttccct ctgacgggta ttgatcctcc cctggaactg tacagttcac gttctgatct 73020  
 ttttcttgac aaagggaatt ccagtttgt tgcgtggcga acgcactagc aggtgaggag 73080  
 ttaaaagttg gcaacgcctg cctctcagag agtgtcagga tttttagtot cttccttgag 73140  
 agctagaaga tgtttctaaa agaactctt tgggtgactta gaagtggaga gagctttaga 73200  
 agcatggcac aaataaaagg aaaggggcaa acacgcctat tctacatctg tttattttgt 73260  
 tattaacaaa aggcaaggcg attttcatta aagttttgct ggggttgggg ttgagggtgt 73320  
 agagagcaaa agtgtgagtt gtacaccatg actggaatcg cttggacata ctcttcagca 73380  
 gacatcgtgt gactgtggaa gaaatgagtt tcatgaagat gactgataga aggaagccac 73440  
 tgaaccagtc ctctatcacc tcttccaagg cttaaagttg gagccacttg cagaaggctc 73500  
 tcctcaaac cctgtgttct ttgctaccc ctgctgttgc cacatcatct tggagagctg 73560

```

gtcgtctccc tcctcaacta gaagttccta gtgcctgctt agttcttgtc tcttgcttcc 73620
caagtgtctca caaaatacat ccatgttcgc tacgaggaaa tggaccacat aagggtttccg 73680
tgaaaacctt agcccttagg tctaacacag taggaacaga agttaatgtt ttctgaegt 73740
agaagtttct cttgtgctt ctggtcacat ttctttcttg tgtgggttct ctatggctac 73800
tgcacttttt tttttttctt actgtctccc cctccccca cacaccacct tttggggata 73860
gggtggcagg tgagaatata aacagataat ggttaagaga tagtttagtc tttctaggcc 73920
agattattta gtttttgcca tctaggtaaa attoggtcca attaagcgtc cattaagtgt 73980
tttaataataa gctggagaag gagttgaacc tggaggtcag ggctctgtgg tctattacag 74040
tccccctggg gtctctagcc caagggagac tccagggtct taataaatga ctgggggttt 74100
cattttgagg cctttactac caaagactga ataatacatt gggcatgatg gttttgtcct 74160
aaacattaac agccacaaaa ggtagagagt gtgtctgttt atagatacac atgtatcatg 74220
aataattagt tggggactgt gcatcaggtc tctcatttta cattogagga agcaatgcac 74280
ggaatgaatt ctggacctgc gaactctgaa ttccaattct ctgtctccta cttttactgg 74340
agtgtctgca aacagtacag tgtttttgtt gtgaagttat accgtgcctg taatctctct 74400
gcggytggcc ctctaagcc ctacttcaag aaatagctct aagctcatga caccgcgcc 74460
accgatgcc tacatatgtc ttatatcctt ggagtagtgt ttggggttgc aaatttgact 74520
ttagggagac atactctctg atgataggct aatgcttata tttactgata aacttccttt 74580
ttgacggtca tgggcttcgg gggccacca accaaactgt gtgctgctt ttatgttggg 74640
ccaaaagaca ggctccttgt gtcctcccag tttcttaaac aatgaagtca tggcatttta 74700
cagtgtctgt gaatggattg agattgtggt ggccctggaa tgtggcactg ctctggctgg 74760
aggaagatg agagtgggg atggagagga gaggagagcg ggagatggga acctggtgga 74820
cacaggaggg agtgtgagtt ctgaggcca aaggaaactt gacaccgat gggacattaa 74880
tctgattctg ttatctgagg ctgtcaccag tctccctgt cctcctggca 74930

```

```

<210> 2
<211> 384
<212> DNA
<213> Homo sapiens

```

```

<400> 2
ttcaggccat tgggtgtgta tatatttcaa gatttgctca cagggtccaa gcttaactta 60
agctccctga gacatatcat aaaatatgat ttggggaaaa accctaattg gccatgatca 120
gaacattatt attcaacaaa ggatgaaatg cttaagccaa gatggccttc tttctttctt 180

```

tctttctttc ttttttttta atgaaagttg agcagactcc cgtccaacag ttttcaatgt 240  
 aggaattccc acagcccat ttgattgcag tttgttgaaa agtttaatgt tttgttaggc 300  
 aattcataat ttccacattg aacagcctga gaggaagaga gctggagccc actgttggtt 360  
 ttgtagtggg atggtgggaa cttt 384

<210> 3  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens

<400> 3  
 ttcaggccat tgggtgtgta 20

<210> 4  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens

<400> 4  
 aaagttccca ccatcccact 20

<210> 5  
 <211> 366  
 <212> DNA  
 <213> Homo sapiens

<400> 5  
 ttgtcccttg aggtgtactg gaaactaagg cgtgaggagc tcataggggt ctggcttgga 60  
 aagtgtattg ctatgtccag tttacacata aggatgtgca aatccagcag gtttagctgag 120  
 ctgcccagga atatccaggc aagaatkacc atattctgat aattactcag gcctctgcct 180  
 catctccgct gccccccgc cccctgactc tottctgagt gccagattca gcctccattt 240  
 gaatgccaaa tagacaggaa attagcatgc ccagaatcca cgtctttagt gcactctctc 300  
 cccagctcca aacctgttac tgcttgtgtt caacatctca gtaaagctca acaacatoga 360  
 cccatt 366

<210> 6  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens

<400> 6  
 ttgtcccttg aggtgtactg g 21

<210> 7  
 <211> 20

```

<212> DNA
<213> Homo sapiens

<400> 7
aatgggtcga tgttgttgag 20

<210> 8
<211> 558
<212> DNA
<213> Homo sapiens

<400> 8
gctgtgaaat cccctgtgta gtggaagaa gaaatagcaa atottagctg ccttggaact 60
gatataatta tttgtcttca tttacatggt tyatccttca aggttgaata aatgatgtgg 120
gagctagtca aggggcttta ggtatgtgat ttcatgccta ctttttttta ggtagagaaa 180
ctgaggtcac agggtagtag agaatggact ctaagattca ggtttctgaa ttgcctgtgg 240
ttttgttgac tcaactgctc ttctgttgtt ttttagccac atgccttgaa acagtcctct 300
ttcccatgtt totteatcag caocattaac ccaaggata ctgtcctctc ttatctttca 360
caaggtcttg gagttcccat gcctttgtaa gcacccctcc ccgagattca gcaccaacca 420
aaatcacatt tggaaaaatt gcttgtttcc caagaagctt tggaggatat gattttgtat 480
agaacgggtt cacagggttt ctgttcattc ttctatgggtg gagtgtgtgt gtatgtgact 540
ctgtcttctc tccattcc 558

<210> 9
<211> 21
<212> DNA
<213> Homo sapiens

<400> 9
gctgtgaaat cccctgtgta g 21

<210> 10
<211> 23
<212> DNA
<213> Homo sapiens

<400> 10
ggaatggaga gaagacagag tca 23

<210> 11
<211> 364
<212> DNA
<213> Homo sapiens

<400> 11
aaggagagaa gcaggattga gcagggggag cagtcagatg gtaatgcaga tgtgatgaga 60

```

tctctgccgg accaaagaga agattccttt ttaaattggtg acaaattcat gggctttctc 120  
 tgcctcaaaa cctagcacag ctgttattha ctgaacaatt agagagctaa gcacttttta 180  
 gataytatat aatttaattg ccgtatgagg cacccttagt tttcagacga gaaaccacag 240  
 ttacagggaa ggcaagtaac ttagtcaatg tcagataact aggaaaaggt tagaggggoc 300  
 ctggacacag gcctgtgtga ctgagaagct tgggcacttc actgctacat ttcattctctt 360  
 cgct 364

<210> 12  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens

<400> 12  
 aaggagagaa gcaggattga 20

<210> 13  
 <211> 22  
 <212> DNA  
 <213> Homo sapiens

<400> 13  
 agcgaagaga tgaaatgtag ca 22

<210> 14  
 <211> 579  
 <212> DNA  
 <213> Homo sapiens

<400> 14  
 ctgatgagg tagggagcat ctgtctgcag cttoactctc attgtctagg ggctccagaa 60  
 atatctgtga gtaaataagt tattaatct ttgcctcaaa tttccagtga ctgtagggat 120  
 atagctgtga gcctctagga gctgagattt tttaaatttc ccacttaaac atttatttaa 180  
 aaattttgtg ctcagcatgg actaaggact ttacattcat taactcattt acagcttgat 240  
 cctatgcggg gggcattcat ttacagagga tcccatthta caggtagga agaggccagc 300  
 taggggtgca gcctagggtta gtattctaga gctcatcagg ctgtgttgtc ccagtgaaa 360  
 gaataagcaa agaagtgaat gttgtgcatt gagaaaaatg actctcggag gaggatgagc 420  
 ctctcggata tggcgaccga agtgatwtgg ggcccttgtc aagggtctct attatggcat 480  
 caagaaaaga tgctgcttct ggtgatgccc gaggagagcc tcaatattht acatgggaaa 540  
 cctaaaaaag gggccatggt gtggtctctg cacctaaga 579

```

<210> 15
<211> 19
<212> DNA
<213> Homo sapiens

<400> 15
ctgatgaggg tagggagca 19

<210> 16
<211> 22
<212> DNA
<213> Homo sapiens

<400> 16
tcttaggtgc agagaccaca ac 22

<210> 17
<211> 486
<212> DNA
<213> Homo sapiens

<400> 17
tatttagaaa ccataaaatc cacotatttg aggtgtacaa ttgagtgatt ttctgtatag 60
tcacagatct gtgcagtcac ccacaccctc taactccagg acattttcct ccccccgag 120
gagaaaacct ccttaccatc tagcagtcac tctcatttc ctctccccc agccctggc 180
aatcactgtg gatttgctg ttcttgacat ttcataaaa tggatcata aaatctaygg 240
gcttttgtgt ctgtctgctt tcaacttagca tacggttctc aagggtcacc cagtattgta 300
goatctatca gtatgtcatt cctttttatg gccaaataat attttattgt atggatagac 360
attttgttta ttcattttatc tgtttttggt tattatgagt aacactacta tgaacatttt 420
gcacaaattt ttgtattgac atgttttcat ttctcctggg tatagtccca tgagtggaat 480
tgctgg 486

<210> 18
<211> 27
<212> DNA
<213> Homo sapiens

<400> 18
tatttagaaa ccataaaatc cacotat 27

<210> 19
<211> 22
<212> DNA
<213> Homo sapiens

<400> 19
ccagcaatto cactcatagg ac 22

```

```

<210> 20
<211> 428
<212> DNA
<213> Homo sapiens

<400> 20
ttgtctcctt ttgtttctgc tactgtgaat gatcctgtga tgatcatctt tgtgtgtaaa 60
tctttgtccc ctgcgccctt ccccttttat tattttcttg ggatagacc caggacaaaa 120
ggtagaaaag aacaaagtgt taaaaaattt cttgatacat agccacagat tttttctctg 180
aaagtcttca acatttataa ctacsagcag tatgtaagag agttatgggt ggaatgattt 240
taatgtctct ggggaattta acaacaaaaa aacttttagg ttctttggag agagacatgc 300
ccttaactcc accccgccct agaacagaga ccagoccat ocaagtcagc ctcccaggt 360
cctccacctt caaaacaggc aaacgaaatc atttcttgaa taattggtag gtttcaaggt 420
cagatgtt 428

<210> 21
<211> 23
<212> DNA
<213> Homo sapiens

<400> 21
ttgtctcctt ttgtttctgc tac 23

<210> 22
<211> 22
<212> DNA
<213> Homo sapiens

<400> 22
aacatctgac cttgaagcct ac 22

<210> 23
<211> 330
<212> DNA
<213> Homo sapiens

<400> 23
tcagggacag tgcatagggt taaagaagtt gctggttggg ggttotaatg caggtttctc 60
caaaagtga tgcctgtta aaaaaaatt cttacaaat atacagagat tttttttta 120
aaaaagtgtg acagttctag acacctagag agtaaaatga agaagcctgt ttccaggttt 180
ccgcctccc tgaatttccc agcatggtcc aggttttgaa atttatttat ctgcttttgg 240
caatgggtga tgggaatttc ccacatttat ttttagcta cagagaaagg acattatctt 300

```

```

taaaatctct tcgttgttct ctctcttga 330

<210> 24
<211> 20
<212> DNA
<213> Homo sapiens

<400> 24
tcagggacag tgcataagtg 20

<210> 25
<211> 23
<212> DNA
<213> Homo sapiens

<400> 25
tcaaagagag agaacaacga aga 23

<210> 26
<211> 574
<212> DNA
<213> Homo sapiens

<400> 26
tatttagaaa ccataaaatc cacctatttg aggtgtacaa ttgagtgatt ttctgtatag 60
tcacagatct gtgcagtcac ccacaccctc taactccagg acatttttct cccccccgag 120
gagaaacctc ccttaccatc tagcagtcac tctcatttc ctctccccc agccctggc 180
aatcactgtg gatttgctg ttcttgacat ttcataaaa yggatcata aaatctatgg 240
gcttttggtg ctgtotggt tcaottagoa tacggttctc aaggttcato cagtattgta 300
gcatotatca gtatgtcatt cctttttatg gccaaataat attttattgt atggatagac 360
attttgttta ttcatttatc tgtttttggt tattatgagt aacactacta tgaacatttt 420
gcacaaattt ttgtattgac atgttttcat ttctcctggg tatagtctta tgagtggaat 480
tgctgggtca tataataaat aactgtttaa cattttgggg agctgccaaa cttttaaaac 540
cttgggttct gtgatgtacc agttgtgta ggca 574

<210> 27
<211> 27
<212> DNA
<213> Homo sapiens

<400> 27
tatttagaaa ccataaaatc cacctat 27

<210> 28
<211> 22

```

```

<212> DNA
<213> Homo sapiens

<400> 28
tgcttaacac aactggtaca tc 22

<210> 29
<211> 571
<212> DNA
<213> Homo sapiens

<400> 29
tgccaggggt tttatggtta attttcootcc attatgaggg ttgactcagc cttgggtatt 60
agatgtcttt gagaatccag ggttcaaata ccacagctgg tagaatgttt ctcaacttgg 120
agccaatctc catctactga aggtacgctg gtttagacag acaacagga catcagcatt 180
ttaaaaagcg gtggaaaag tttgcttgc ttgattggag ccacgacatt ttattttgaa 240
atttcaaata acatgaaggg aggtttggag cggtttttgg tttatccaaa gggcagtgga 300
ttgaaggctg agaaacacca ggctgaatgg gagaggggtt ggggtccccc tgtgagatag 360
tgaaacaatg gtagtgccat ccaatgatag gcacttttct gtcattcaga agcagaaagg 420
gggcagaggg cccattggcc ttactgggma gtaagctgta gagctgctgc cttttcgtga 480
aagggttgac accaaccttc tccccagga agagtgaacca gggacctgag gggcatggtc 540
gagcagatga cagcctttgt aaaacatctc c 571

<210> 30
<211> 20
<212> DNA
<213> Homo sapiens

<400> 30
tgccaggggt tttatggtta 20

<210> 31
<211> 23
<212> DNA
<213> Homo sapiens

<400> 31
ggagatgttt tacaaaggct gtc 23

<210> 32
<211> 614
<212> DNA
<213> Homo sapiens

<400> 32
ttggtagaga tggggtctcc taggctgggc ttgaactcct ggrotcaagc aatcttctg 60

```

```

cctcagcctt ccaaagtact gggattactg gcgtgggcca coatgccttg cttgaaattt 120
ttctatggct ttattcttcc tccaagtaca gagtctaccc aaccttctga gatctttggt 180
tttcttttcc taggtaacta tagtacatac ttatttatgt taaacaacag caatcacaca 240
tttcttttcc tatacagtca tgctttatag gcaaataaag cctccgtctt aggctttctg 300
gattttttca aaagatgcaa ttcttgaggt atgtttttac ttagagcaaa gcagcctagt 360
ctcctatacc ttctgcatct gcagaaaagt tggtaaaca gactttgtaa tgatgccct 420
tacaattctg aagggaactg tgaaatagtt tcacagagtt tcagtgttag gtatatttga 480
tcaatgctaa cttttgaaa actttggtgc ctgtatgatt cagagggtag ggcagaatat 540
taaattaatc acaacttctt gtattttaac cattctgggt aaattgggat tccgtgacgc 600
ccaggcaaaa ttat 614

```

```

<210> 33
<211> 20
<212> DNA
<213> Homo sapiens

```

```

<400> 33
ttggtagaga tggggctctc 20

```

```

<210> 34
<211> 20
<212> DNA
<213> Homo sapiens

```

```

<400> 34
ataattttgc ctgggcgtca 20

```

```

<210> 35
<211> 633
<212> DNA
<213> Homo sapiens

```

```

<400> 35
tatcttatat cccctccaag cattcattaa ctgatggatt agtgagttgg ccttgagaag 60
cataaaggct cgtctccatg tgcttctaag catttgtctt aagttctggt tggtttccctg 120
agtgaactg tcttaatggt accaacagaa gttaaatgoc taagagwttc ttatacatgg 180
gtgagtaacc tctgtgactg ggcaagccac ctcacctcat ttaccttgt ctgcaaatg 240
aggaactggg tcaactcatc gttcaaatct cactgaaagc taattgatcg cttttgacag 300
aagtagctcc cttgggcccgt atatttatct cctagcttgg aggaaggtgg ggacagacag 360
aattgatgta cacctttatt ttatctcta tggtaaacct gtgcatacta aagcattcct 420

```

```

ctgggtctttt gagatgagtg tatacattgt gtctggccct gtgcattttt taccaagaag 480
taagttttgt tgagtaaact tgggttgat gaagaactgc atgctcaccg tactcaagta 540
gcttttgcta cctaaaggac agctgctcat atgtacttga ctctctttaa agtgaaggat 600
gatgacattt gaaaaacgga ggttgaaaag gag 633

<210> 36
<211> 25
<212> DNA
<213> Homo sapiens

<400> 36
tatottatat cccctccaag cattc 25

<210> 37
<211> 21
<212> DNA
<213> Homo sapiens

<400> 37
ctccttttca acctccgttt t 21

<210> 38
<211> 1081
<212> DNA
<213> Homo sapiens

<400> 38
ttgagcatgt gttatttaat gagttatacc tctgtcatat gtgtgtgttt atatcacaaa 60
ataacttatt ttataaaaac catattttga gtcacattt gtgacaatgt cttcttttct 120
ctgggtataaa tgaggcatgt agaaagaaga ttgacatttg ctagaagott cccctttcct 180
ctaactccac aataaaatgg atgctcataa ttacatctgc tctataagg tcaagatttc 240
agggctggaa gtgaccttag atcatttagg occaacttgc cctcaggaaa ggaaactgag 300
gccagagat gccttaagtg aattgcccaa tgtcacacgc tgagtoagtg gccagagcaa 360
ggcttggtac cagttctctg ctccctttcc agagccttgt gatgtcttct ctctacagg 420
aggtgaaaat aactgctgtg gctgggtotg ttttgctgac tgtaaattgg gtcattggtca 480
gggacagtgc ataggtgtaa agaagttgct ggttgggggt tctaatagcag gtttctccaa 540
aagtgaatgc cotgttaaaa aaaaattcctt aacaaatata cagagatttt tttttwaaaa 600
aagtgtgaca gttctagaca cctagagagt aaagtgaaga agcctgtttt caggtttccc 660
gcctccotga atttccacgc atggtccagc ctttgaaatt tatttatctg cttttggcaa 720
tggttgatgg gaatttccca catttatatt ttagctacag agaaaggaca ttatotttaa 780

```

```

aatctcttcg ttgttctctc tctttgagtg aggagagaag atgtgaatcc tggcagtggt 840
tcagagtggg cacagcccct gtgtttgtgg cataggctct gtgggcccga tggcagggag 900
cagtaccccc gtgtaaagga gtgggggttt gtccatttgg atagagcaaa gatcctccac 960
ctcaaatccc acaagaacag ttgccacaac ctggggccota agcatctcat tttcctatgt 1020
agaaattaat gatctggagg agatggcaaa acattccttc cagagcctgt gtggattttg 1080
g 1081

<210> 39
<211> 26
<212> DNA
<213> Homo sapiens

<400> 39
ttgagcatgt gttatttaaat gagtta 26

<210> 40
<211> 20
<212> DNA
<213> Homo sapiens

<400> 40
ccaaaatcca cacaggctct 20

<210> 41
<211> 599
<212> DNA
<213> Homo sapiens

<400> 41
tagtgctcag tatttccaac gttctgttta ttttaagatga aaattgctgt agttaataag 60
cacttcccca tgtcattaaa atgcttaagg atttttaatg accacataac agtcataat 120
atgattaaac cccaatttac tgaatcaatg ccatattggt gggctcttag attgtctcct 180
tttgtttctg ctactgtgaa tgatcctgtg atgatcatct ttgtgtgtaa atctttgtcc 240
cctcgccccc tcccctttta ttattttctt gggatagacc ccaggacaaa aggtagaaaa 300
gaacaaagtg ttaamaatt tcttgataca tagccacaga ttattttcct gaaagttctc 360
aacatttata actacgagca gtatgtaaga gagttatggt tggaatgatt ttaatgtctc 420
tggggaattt aacaacaaaa aaactttagg cttctttgga gagagacatg ccottaactc 480
caccocgccc tagaacagag acccagccca tccaagtcag cctccccagg tctccacct 540
tcaaaacagg caaacgaaat catttcttga ataattggta ggcttcaagg tcagatgtt 599

```

```

<210> 42
<211> 25
<212> DNA
<213> Homo sapiens

<400> 42
tagtgctcag tatttccaac gttct 25

<210> 43
<211> 23
<212> DNA
<213> Homo sapiens

<400> 43
aacatctgac cttgaagcct acc 23

<210> 44
<211> 599
<212> DNA
<213> Homo sapiens

<400> 44
tagtgctcag tatttccaac gttctgttta tttaagatga aaattgctgt agttaataag 60
cacttcccca tgccattaaa atgcttaagg atttttaatg accacataac agtccataat 120
atgattaaac cccaatttac tgaatcaatg coatatgtt gggctcttag attgtctcct 180
tttgtttctg ctactgtgaa tgatcctgtg atgatcatct ttgtgtgtaa atctttgtcc 240
cctgcggccc tccccttita ttattttctt gggatagacc ccaggacaaa aggtagaaaa 300
gaacaaagtg ttaaaaaatt tottgataca tagccacaga ttattttcct gaaagttcts 360
aacatttata actacgagca gtatgtaaga gagttatggt tggaatgatt ttaatgtctc 420
tggggaattt aacaacaaaa aaactttagg cttctttgga gagagacatg cccttaactc 480
caccccgccc tagaacagag acccagcca tccaagtcag cctccccagg tctccacct 540
tcaaaacagg caaacgaaat catttcttga ataattggtt ggcttcaagg tcagatgtt 599

<210> 45
<211> 641
<212> DNA
<213> Homo sapiens

<400> 45
tgctatgtcc agtttacaca taaggatgtg caaatccagc aggttagctg agctgcccag 60
gaatatccag gcaagaatga coatatctg ataattactc aggcctctgc ctcatctccg 120
ctgscggccc gcccctgac totcttctga gtgccagatt cagcctccat ttgaatgcc 180
aatagacagg aaattagcat gccagaatc cacgtcttta gtgcactctc tccccagctc 240

```

```

caaacctgtt actgottgtg ttcaacatct cagtaaagct caacaacatc gacccattac 300
ttaggcotca aaccttgggt ggcacgtcg attgtcttt totttcatac cccacattca 360
accocatcagc ccatcccaca ggoccaaagt tgcctctct accttcaaag cgtgtgtggc 420
atccacogct tatcaccacc totgccatta ccactggagt ccagtgcacat catctctcac 480
ttggatgtgg ccagagtgtc ttgtctgtc tccttcttgc ttctacactt tgtaacagcc 540
tatcatctat ctctgtctc catagctcac tccataact tgagagggcc tttgaaagcc 600
ttagacagat catatcacag acctctatac tgaaagtcgg g 641

<210> 46
<211> 25
<212> DNA
<213> Homo sapiens

<400> 46
tgctatgtcc agtttacaca taagg . 25

<210> 47
<211> 24
<212> DNA
<213> Homo sapiens

<400> 47
cccgaattc agtatagagg tctg 24

<210> 48
<211> 284
<212> DNA
<213> Homo sapiens

<400> 48
ccatctgtgg agcagagtca ctgaaaggaa atactggaaa tactggaagc cacttggtgt 60
tttatcaagg atgtgaggtt tcctggcaac ttgtcgcca tatcatcatc atcatcacca 120
tcatcatcat catcatcatc atcatcatca tcatcatcat catcatctgc cctttaagtt 180
ttctgcttgt ttgaaaaaga aatttataca gagccccag tagcagotgt aagggggcag 240
gttcttggag cagcccatcc tcaacattct tgotgtgat ggaa 284

<210> 49
<211> 20
<212> DNA
<213> Homo sapiens

<400> 49
ccatctgtgg agcagagtca 20

<210> 50
<211> 20
<212> DNA
<213> Homo sapiens

<400> 50
ttccatcagc agcaagaatg 20

<210> 51
<211> 145
<212> DNA
<213> Homo sapiens

<400> 51
tccacgcaga gaggatctaa atctggctct ttgcaattgc cttcatacat gtgcatacac 60
accacacaca cacacacaca cacacacaca cacacacaca cagacacata catatgcaca 120
caccocgact caatggagga ccttc 145

<210> 52
<211> 21
<212> DNA
<213> Homo sapiens

<400> 52
tccacgcaga gaggatctaa a 21

<210> 53
<211> 20
<212> DNA
<213> Homo sapiens

<400> 53
gagggtcctc cattgagtcg 20

```

## ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Способ диагностики восприимчивости к диабету II типа у индивидуума, включающий анализ образца нуклеиновой кислоты, полученного из индивидуума, на маркер или гаплотип в LD блоке экзона 4 TCF7L2, между маркерами rs4074720 и rs7087006 или с сильным нарушением равновесия сцепления с LD блоком экзона 4 TCF7L2, характеризующимся численным значением  $r^2$  более чем 0,2, где присутствие маркера или гаплотипа является индикатором восприимчивости к диабету II типа.

2. Способ по п.1, где маркер или гаплотип включает по меньшей мере один маркер, выбранный из маркеров, перечисленных в табл. 6, или по меньшей мере один маркер с сильным нарушением равновесия сцепления по меньшей мере с одним маркером, перечисленным в табл. 6, характеризующимся численным значением  $r^2$  более чем 0,2.

3. Способ по п.2, где маркер или гаплотип включает по меньшей мере один маркер, выбранный из маркеров, перечисленных в табл. 6.

4. Способ по любому из предшествующих пунктов, где маркер или гаплотип являются маркером.

5. Способ по любому из предшествующих пунктов, где маркер или гаплотип являются индикатором повышенной восприимчивости к диабету II типа.

6. Способ по п.5, где повышенная восприимчивость характеризуется относительным риском по меньшей мере 1,2, включая относительный риск по меньшей мере 1,3 и относительный риск по меньшей мере 1,4.

7. Способ по любому из пп.1-4, где маркер или гаплотип являются индикатором пониженной восприимчивости к диабету II типа.

8. Способ по п.6, где пониженная восприимчивость характеризуется относительным риском менее чем 0,8, включая относительный риск менее чем 0,7.

9. Способ по любому из пп.4-6, где маркер выбирают из группы, состоящей из DG10S478, rs12255372, rs7895340, rs11196205, rs7901695, rs7903146, rs12243326 и rs4506565, и где присутствие не-0 аллеля в DG10S478; T аллеля в rs12255372; A аллеля в rs7895340; C аллеля в rs11196205; C аллеля в rs7901695; T аллеля в rs7903146; C аллеля в rs12243326 или T аллеля в rs4506565 является индикатором повышенной восприимчивости к диабету II типа.

10. Способ по п.9, где маркером является rs7903146 и где присутствие T аллеля в rs7903146 является индикатором повышенной восприимчивости к диабету II типа.

11. Способ по п.7 или 8, где маркер выбирают из группы, состоящей из DG10S478, rs12255372, rs7895340, rs11196205, rs7901695, rs7903146, rs12243326 и rs4506565, и где присутствие 0 аллеля в DG10S478; G аллеля в SNP rs12255372; G аллеля в rs7895340; G аллеля в rs11196205; T аллеля в rs7901695; C аллеля в rs7903146; T аллеля в rs12243326 или A аллеля в rs4506565 является индикатором пониженной восприимчивости к диабету II типа.

12. Способ по п.11, где маркером является rs7903146 и где присутствие C аллеля в rs7903146 является индикатором пониженной восприимчивости к диабету II типа.

13. Набор для анализа образца из индивидуума для обнаружения восприимчивости к диабету II типа, где набор включает один или более реагентов для обнаружения одного или более маркеров в LD блоке экзона 4 TCF7L2, между маркерами rs4074720 и rs 7087006 или с сильным нарушением равновесия сцепления с LD блоком экзона 4 TCF7L2, характеризующимся численным значением  $r^2$  более чем 0,2.

14. Набор по п.13, где один или более маркеров выбраны из маркеров, перечисленных в табл. 6.

15. Набор реактивов по п.14, где один или более маркеров выбраны из группы, состоящей из DG10S478, rs12255372, rs7895340, rs11196205, rs7901695, rs7903146, rs12243326 и rs4506565.

16. Набор по любому из пп.13-15, где один или более маркеров являются маркером rs7903146.

17. Набор реактивов по п.16, где маркером является C аллель в rs7903146.

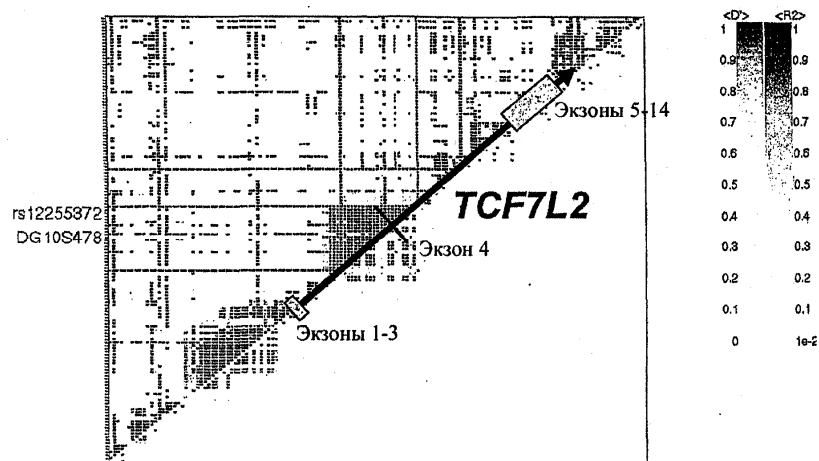
18. Набор по любому из пп.13-17, где один или более реагентов включает по меньшей мере одну прилегающую нуклеотидную последовательность, которая полностью комплементарна области, включающей один или более маркеров.

19. Способ оценки индивидуума на вероятность ответа на терапевтический агент TCF7L2, включающий детекцию маркера в LD блоке экзона 4 TCF7L2, между маркерами rs4074720 и rs7087006 или с сильным нарушением равновесия сцепления с LD блоком экзона 4 TCF7L2, характеризующимся численным значением  $r^2$  более чем 0,2.

20. Способ по п.19, где маркер выбирают из группы, состоящей из DG10S478, rs12255372, rs7895340, rs11196205, rs7901695, rs7903146, rs12243326 и rs4506565.

21. Способ по п.20, где маркер является маркером rs7903146 и где присутствие T аллеля в rs7903146 является индикатором вероятности позитивного ответа на терапевтический агент TCF7L2.

22. Способ по любому из пп.19-21, где терапевтический агент TCF7L2 является агентом, выбранным из группы, приведенной в таблице агентов.



Евразийская патентная организация, ЕАПВ  
Россия, 109012, Москва, Малый Черкасский пер., 2