



(12) 发明专利

(10) 授权公告号 CN 101238222 B

(45) 授权公告日 2013. 04. 10

(21) 申请号 200680022129. 6

(22) 申请日 2006. 06. 16

(30) 优先权数据

60/692, 174 2005. 06. 20 US

60/757, 155 2006. 01. 06 US

(85) PCT申请进入国家阶段日

2007. 12. 20

(86) PCT申请的申请数据

PCT/IS2006/000014 2006. 06. 16

(87) PCT申请的公布数据

W02006/137085 EN 2006. 12. 28

(73) 专利权人 解码遗传学私营有限责任公司

地址 冰岛雷克雅末

(72) 发明人 斯特鲁安·F·A·格兰特

(74) 专利代理机构 中原信达知识产权代理有限
责任公司 11219

代理人 杨青 樊卫民

(51) Int. Cl.

C12Q 1/68 (2006. 01)

A61P 3/10 (2006. 01)

(56) 对比文件

CN 1548553 A, 2004. 11. 24, 摘要及权利要求.

Struan F A Grant et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. 《Nature Genetics》. 2006, 第 38 卷 (第 3 期), 320-323.

审查员 张艳霞

权利要求书 1 页 说明书 58 页

序列表 47 页 附图 1 页

(54) 发明名称

作为 2 型糖尿病风险诊断标记物的 TCF7L2 基因的遗传变异体

(57) 摘要

关联分析显示基因 TCF7L2 多态性是 II 型糖尿病的易感基因。本文描述了糖尿病易感性的诊断方法, 糖尿病易感性降低的诊断方法和预防糖尿病的方法, 以及 II 型糖尿病的治疗方法。

1. 一种试剂盒,用于分析来自个体的样品以检测 II 型糖尿病易感性,其特征在于所述试剂盒包括一种或者多种试剂,用于检测 TCF7L2 外显子 4LD 块内 rs4074720 和 rs7087006 之间选自如下的一种或者多种标记物:DG10S478, rs12255372, rs7895340, rs11196205, rs7901695, rs7903146, rs12243326, 和 rs4506565, 其中所述试剂与所述外显子 4LD 块内的区域特异性杂交。

2. 如权利要求 1 所述的试剂盒,其特征在于一种或者多种标记物是标记物 rs7903146。

3. 如权利要求 2 所述的试剂盒,其特征在于标记物是 rs7903146 的 C 等位基因。

作为 2 型糖尿病风险诊断标记物的 TCF7L2 基因的遗传变异体

[0001] 发明背景

[0002] 糖尿病,一种糖利用减少、脂和蛋白利用增加的代谢疾病,由胰岛素的绝对或者相对缺乏引起。在更严重的情况下,糖尿病的特点在于长期高血糖,糖尿,水和电解质损失,酮酸中毒和昏迷。长期并发症包括神经病变、视网膜病、肾病、大小血管全身退行性病变的发展和感染易感性的增加。糖尿病最常见的类型是 II 型,非胰岛素依赖型糖尿病,其特征在于高血糖由受损的胰岛素分泌和靶组织中胰岛素抵抗造成。遗传和环境因素都对这种疾病有作用。例如,肥胖在疾病发展过程中起着重要作用。II 型糖尿病经常是渐进发病糖尿病的温和形式。

[0003] II 型糖尿病对健康的影响是巨大的。在 1995 年,全球范围内存在 1 亿 3500 万成年糖尿病患者。预计在 2025 年接近 3 亿人将患有糖尿病 (King H., 等人, Diabetes Care, 21 (9) :1414-1431 (1998))。在冰岛,成年群体中 II 型糖尿病的患病率是 2.5 % (Vilbergsson, S., 等人, Diabet. Med., 14 (6) :491-498 (1997)), 这包括大约 5,000 个年龄在 34 岁以上患有这种疾病的人。这种疾病的高患病率和增加的患病人口数目显示,更精确界定涉及 II 型糖尿病的遗传因素以定义相关风险系数的医疗需要还没有得到满足。还需要用于预防 II 型糖尿病的治疗剂。

[0004] 发明概述

[0005] 本发明涉及通过评估某些与 TCF7L2 基因 (转录因子 7-样 2 (T-细胞特定的, HMG-盒), 过去称为 TCF4 基因 (T-细胞转录因子 4)) 有关的标记物或者单倍型, 诊断 II 型糖尿病易感性增加的方法, 以及诊断 II 型糖尿病易感性降低或者诊断对 II 型糖尿病的防护的方法。该方法包括检测与 TCF7L2 基因外显子 4LD 块相关的遗传标记物。

[0006] 第一个方面, 本发明涉及在个体中诊断 II 型糖尿病易感性的方法, 包括对从个体获得的核酸样品进行与 TCF7L2 外显子 4LD 块相关的标记物或者单倍型的分析, 其中标记物或者单倍型的存在指示 II 型糖尿病易感性。在一个实施方式中, 标记物或者单倍型包括至少一种选自表 6 所列的标记物。在另一种实施方式中, 标记物或者单倍型是标记物。

[0007] 在一个优选的实施方式中, 标记物或者单倍型指示 II 型糖尿病易感性的增加。易感性增加存在于一个实施方式中, 表征为相对危险度至少为 1.2, 包括相对危险度至少为 1.3 和相对危险度至少为 1.4。在一个实施方式中, 标记物选自 DG10S478, rs12255372, rs7895340, rs11196205, rs7901695, rs7903146, rs12243326, 和 rs4506565, 其中 DG10S478 的非 -0 等位基因 (例如, -4, 4, 8, 12, 16, 20, 或者其他非 -0 等位基因), rs12255372 的 T 等位基因; rs7895340 的 A 等位基因; rs11196205 的 C 等位基因; rs7901695 的 C 等位基因; rs7903146 的 T 等位基因; rs12243326 的 C 等位基因; 或者 rs4506565 的 T 等位基因的存在, 指示 II 型糖尿病易感性的增加。在一个优选的实施方式中, 标记物选自 DG10S478 和 rs7903146, 其中 DG10S478 的非 -0 等位基因或者 rs7903146 的 T 等位基因的存在指示 II 型糖尿病易感性的增加。在另一个优选的实施方式中, 标记物是 rs7903146, 其中 rs7903146 的 T 等位基因的存在指示 II 型糖尿病易感性的增加。

[0008] 在另一个优选的实施方式中,标记物或者单倍型指示 II 型糖尿病易感性降低。易感性的降低存在于一个实施方式中,表征为相对危险度小于 0.8,包括相对危险度小于 0.7。在一个实施例中,标记物选自 DG10S478, rs12255372, rs7895340, rs11196205, rs7901695, rs7903146, rs12243326, 和 rs4506565, 其中 DG10S478 的 0 等位基因, SNPrs12255372 的 G 等位基因;rs7895340 的 G 等位基因;rs11196205 的 G 等位基因 rs7901695 的 T 等位基因;rs7903146 的 C 等位基因;rs12243326 的 T 等位基因;或者 rs4506565 的 A 等位基因的存在,指示 II 型糖尿病易感性降低。在一个优选的实施方式中,标记物是 DG10S478,其中 DG10S478 的 0 等位基因的存在指示 II 型糖尿病易感性降低。在另一个优选的实施方式中,标记物是 rs7903146,其中 rs7903146 的 C 等位基因的存在指示 II 型糖尿病易感性降低。

[0009] 第二个方面,本发明涉及一种分析来自个体的样品以检测 II 型糖尿病易感性的试剂盒,其中所述试剂盒包括用于检测与 TCF7L2 外显子 4LD 块相关的一种或者多种标记物一或者多种的试剂。在一个实施方式中,一种或者多种试剂包括至少一种连续核苷酸序列,所述核苷酸序列与包括至少一种标记物的区域完全互补,所述标记物与 TCF7L2 外显子 4LD 块相关。在一个实施方式中,一或者多种标记物选自 DG10S478, rs 12255372, rs7895340, rs11196205, rs7901695, rs7903146, rs12243326, 和 rs4506565。在一个优选的实施方式中,一或者多种标记物是 DG10S478 或者 rs7903146。在另一个优选的实施方式中,标记物是 rs7903146 的 C 等位基因。

[0010] 另一方面,本发明涉及一种评定个体对 TCF7L2 治疗剂反应可能性的方法,包括:检测与 TCF7L2 外显子 4LD 块相关的标记物,其中所述标记物存在指示对 TCF7L2 治疗剂阳性反应的可能性。在一个实施方式中,标记物选自 DG10S478, rs12255372, rs7895340, rs11196205, rs7901695, rs7903146, rs12243326, 和 rs4506565。在另一个实施方式中,标记是标记物 DG10S478 或者标记物 rs7903146,其中 DG10S478 的非 -0 等位基因或者 rs7903146 的 T 等位基因的存在指示对 TCF7L2 治疗剂阳性反应的可能性。

[0011] 本发明的另一个方面涉及利用 TCF7L2 治疗剂制备治疗 II 型糖尿病的药物。在一个实施方式中,TCF7L2 治疗剂是一种改变 Wnt 信号途径或者钙粘蛋白途径中活性的药剂。在另一个实施例中,TCF7L2 治疗剂是一种选自药剂表中所列的组中的药剂。

[0012] 附图简述

[0013] 根据下列对本发明优选实施例更详细的说明,本发明的上述和其他目标、特征和优点将更容易理解。

[0014] 图 1 描述了在 HapMap project Build 16 中感兴趣的 TCF7L2 区域相对 SNPs 的连锁不平衡 (LD)。215.9kb 基因跨越 7 个 LD 块,如黑色箭头示意图所示(根据 NCBI RefSeq),箭头显示转录的方向;外显子被标识,外显子 4 被加亮。DG10S478 位于染色体 10 上 TCF7L2 基因内含子 3 内的 114.46Mb (NCBI Build 34),在包括内含子 3 的一部分、外显子 4 的全部和内含子 4 的一部分的 74.9kb 块内(此处称为"TCF7L2 外显子 4LD 块")。等距而不是根据其物理位置绘制 SNP 标记物的曲线。图显示 LD 的两个检测值 - 即 D' (图左上部分) 和 r^2 (右下部分)。

[0015] 发明详述

[0016] 下面是本发明优选实施方式的说明。

[0017] 与 II 型糖尿病相关的基因座

[0018] II 型糖尿病的特点在于高血糖,这可能通过例如受损的胰岛素分泌、周边组织中胰岛素抵抗和肝脏中增加的葡萄糖排出等机制发生。大多数 II 型糖尿病患者患有长期高血糖的严重并发症,包括肾病、神经病变、视网膜病和心血管疾病的加速发展。当前世界范围 II 型糖尿病的患病率是 6%,但预计在未来十年将会上升 (1)。II 型糖尿病患病率的这种增加归因于人口年龄的增加和肥胖的上升。

[0019] 有证据说明遗传组分与患 II 型糖尿病风险有关,包括不同种族的组间患病率的差异 (2,3),单卵双生比异卵双生双胞胎更高的同病率 (4,5) 和欧洲人群中 II 型糖尿病的同胞 (sibling) 相对危险度 (λ_s) 大约为 3.5 (6)。

[0020] 迄今为止存在两种用于寻找 II 型糖尿病相关基因的方法。候选基因内的单核苷酸多态性 (SNPs) 已经被用于检测相关性,并且通常,不被复制或者只造成 II 型糖尿病轻型风险 – 最广泛报道的是过氧化物酶体增殖物激活的受体 gamma 基因 (PPARG2) 中保护性的 Pro 12Ala 多态性 (7) 和内向矫正性钾通道,亚族 J,成员 11 基因 (KIR6.2) 的风险多态性 (8)。

[0021] 在患有常见型 II 型糖尿病家族中进行基因组范围的连锁扫描已经得到几个基因座,国际科研社团的主要注意力集中于许多人群中观察到的染色体 1、12 和 20 上的基因座 (6)。这些基因座内的基因还有待发掘。但是,在墨西哥裔美国人中,从染色体 2q 的基因座分离出钙蛋白酶 10 (CAPN10) 基因;这代表迄今为止通过位置克隆鉴定的与常见型 II 型糖尿病有关的唯一基因 (9)。罕见的孟德尔型 II 型糖尿病,即青年发病的成人型糖尿病 (MODY),通过位置克隆获得 6 个基因 (6)。

[0022] 我们过去报道了在冰岛人群中染色体 5q 与 II 型糖尿病基因组范围的显著连锁 (10);在相同的研究中,我们还报道与 10q 和 12q 连锁的指示性证据。在墨西哥裔美国人中也观察到与 10q 区域的连锁 (11)。

[0023] 与 II 型糖尿病相关的转录因子 7-样 2 基因 (TCF7L2)

[0024] 本发明涉及编码 T-细胞转录因子 4 (TCF4- 官方基因命名 TCF7L2) 基因内 II 型糖尿病相关 LD 块 (“TCF7L2 外显子 4LD 块”) 的鉴定。TCF7L2 外显子 4LD 块内的几个标记物,包括微卫星 DG10S478 和 SNP 标记物 rs7903146 和 rs12255372,已经被发现与 II 型糖尿病有关。最先在冰岛群体中发现 DG10S478 相关性 ($P = 1.3 \times 10^{-9}$; 相对危险度 = 1.45; 人群归因危险度 = 22.7%) 的原始观察,随后在丹麦 II 型糖尿病群体和美国白种人群体中得到重复。DG10S478 位于 10q25.2 上 TCF7L2 基因的内含子 3 中,在明确限定的 74.9kb LD 块内,所述 LD 块包含内含子 3 的一部分,外显子 4 的全部和内含子 4 的一部分。TCF7L2 基因产物是一种包含高迁移率族 (HMG) 盒的转录因子,它在 Wnt 信号途径亦称为 APC3/ β -连环蛋白/TCF 途径中起作用。通过 Wnt 途径成员 β -连环蛋白和糖原合酶激酶-3 β , TCF7L2 介导胰高血糖素原基因表达 (血液葡萄糖稳态中的关键成员) 细胞类型-特定的调节 (12)。此外,通过抑制生脂转录因子 CCAAT/增强子结合蛋白 alpha (C/EBPalpha) 和过氧化物酶体增殖物激活的受体 gamma (PPARGgamma), Wnt 信号维持前成脂肪细胞在未分化状态 (13)。当前成脂肪细胞中的 Wnt 信号被显性失活 TCF7L2 的过表达阻止时,这些细胞分化成脂肪细胞 (13)。此外,有报道称,通过与结肠癌细胞中的 β -连环蛋白和 TCF7L2 的物理相互作用, Wnt/ β -连环蛋白信号途径以 PPARGgamma 活性作为靶标 (14)。多功能 β -连环蛋白对通过结合钙粘蛋白介导细胞粘着也很重要 (15)。

[0025] 作为这些发现的结果,现在已有 II 型糖尿病易感性的诊断方法,以及 II 型糖尿病易感性降低的诊断方法和 / 或对 II 型糖尿病的防护的方法。在本发明优选的实施方式中,诊断方法用于鉴定特定等位基因的存在,包括标记物 DG10S478 的 0 等位基因(与 II 型糖尿病易感性降低相关,是保护性抵抗 II 型糖尿病的等位基因);标记物 DG10S478 的非-0 等位基因(例如,-4,4,8,12,16 或者 20,或者其他等位基因)(与 II 型糖尿病易感性相关);SNP rs12255372 的 G 等位基因(与 II 型糖尿病易感性降低相关,是保护性抵抗 II 型糖尿病的等位基因);SNP rs12255372 的 T 等位基因(与 II 型糖尿病易感性相关);SNPrs7895340 的 G 等位基因(与 II 型糖尿病易感性降低相关,是保护性抵抗 II 型糖尿病的等位基因);SNP rs7895340 的 A 等位基因(与 II 型糖尿病易感性相关);SNP rs11196205 的 G 等位基因(与 II 型糖尿病易感性降低相关,是保护性抵抗 II 型糖尿病的等位基因);SNPrs11196205 的 C 等位基因(与 II 型糖尿病易感性相关);SNP rs7901695 的 T 等位基因(与 II 型糖尿病易感性降低相关,是保护性抵抗 II 型糖尿病的等位基因);SNP rs7901695 的 C 等位基因(与 II 型糖尿病易感性相关);SNPrs7903146 的 C 等位基因(与 II 型糖尿病易感性降低相关,是保护性抵抗 II 型糖尿病的等位基因);SNP rs7903146 的 T 等位基因(与 II 型糖尿病易感性相关);SNP rs12243326 的 C 等位基因(与 II 型糖尿病易感性相关);和 SNP rs4506565 的 T 等位基因(与 II 型糖尿病易感性相关)。在本发明的其他实施方式中,利用本文所述的方法鉴定的其他标记物或者 SNPs,可用于 II 型糖尿病易感性的诊断,还可用于 II 型糖尿病易感性降低的诊断或者用于鉴定保护性抵抗 II 型糖尿病的等位基因。下面介绍的诊断分析方法可用于鉴定这些特定等位基因的存在或者缺失。

[0026] 诊断分析

[0027] 诸如本文所述的那些核酸、探针、引物和抗体可用于 II 型糖尿病易感性的各种诊断方法,以及试剂盒(例如,用于 II 型糖尿病易感性的诊断)。类似地,本文所述的核酸、探针、引物和抗体可用于 II 型糖尿病易感性降低的各种诊断方法,以及保护性抵抗 II 型糖尿病的诊断方法和试剂盒。一方面,试剂盒包括可用于扩增感兴趣的标记物的引物。

[0028] 在本发明的一个方面,通过检测本文所述的 TCF7L2 核酸的多态性(例如,标记物 DG10S478 或者 SNP rs12255372, rs7895340, rs11196205, rs7901695, rs7903146, rs12243326, rs4506565 内的等位基因),进行 II 型糖尿病易感性的诊断。多态性可以是 TCF7L2 核酸的改变,例如单个核苷酸或者不止一个核苷酸的插入或者缺失,导致移码;至少一个核苷酸的改变,导致编码的氨基酸的改变;至少一个核苷酸的改变,导致终止密码子的提前产生;几个核苷酸的缺失,导致核苷酸编码的一个或者多个氨基酸的缺失;一个或者几个核苷酸的插入,例如通过不等重组或者基因转换,导致基因编码序列的中断;全部或者部分基因的重复;全部或者部分基因的转座;或者全部或者部分基因的重排。单个基因可以存在不止一种的这类改变。这类序列改变引起 TCF7L2 核酸编码多肽的差异。例如,如果差异是移码改变,移码可能导致编码氨基酸的改变,和 / 或可能导致终止密码子的提前产生,产生截短多肽。此外,与疾病或者病症相关的多态性,或者对与 TCF7L2 核酸相关的疾病或者病症相关的易感性,可能是一个或者多个核苷酸的同义突变(即,不导致 TCF7L2 核酸编码多肽改变的突变)。这类多态性可以改变剪接位点,影响 mRNA 的稳定性或者转运,或者影响基因的转录或者翻译。具有上述任意改变或者突变的 TCF7L2 核酸在此处被称为“突变的核酸”。

[0029] 在诊断 II 型糖尿病易感性的第一种方法中,可以使用杂交方法,例如 Southern 分析, Northern 分析或者原位杂交(参见 Current Protocols in Molecular Biology, Ausubel, F. 等人, eds, John Wiley & Sons, 包括 1999 年所有增引)。例如,来自待测患者(“待测个体”)基因组 DNA、RNA 或者 cDNA 的生物样品(“待测样品”),取自个体(RNA 和 cDNA 只能用于外显子标记物),例如怀疑患有 II 型糖尿病、对 II 型糖尿病敏感或者易感,或者携带 II 型糖尿病缺陷的个体。个体可能是成年人,儿童,或者胎儿。待测样品可以来自任何来源,包括基因组 DNA,例如血液样品,羊水样品,脑脊液样品,或者皮、肌肉、口腔或者结膜粘膜、胎盘、胃肠道或者其他器官的组织样品。通过适当方法,例如通过羊膜穿刺术或者绒毛膜取样,可以获得来自胎儿细胞或者组织的 DNA 待测样品。然后检测 DNA、RNA 或者 cDNA 样品,确定 TCF7L2 核酸是否存在多态性,和 / 或确定存在何种 TCF7L2 编码的剪接变体。通过基因组 DNA、RNA 或者 cDNA 内基因与核酸探针的杂交,可以显示多态性或者剪接变体的存在。此处使用的“核酸探针”,可以是 DNA 探针或者 RNA 探针;核酸探针可以包含,例如至少一个 TCF7L2 核酸多态性和 / 或包含编码 TCF7L2 核酸特定剪接变体的核酸。探针可以是上述任意的核酸分子(例如,基因或者核酸,片段,包括基因或者核酸的载体,探针或者引物,等等)。

[0030] 为诊断 II 型糖尿病型的易感性,将包含 TCF7L2 核酸的待测样品与至少一种核酸探针接触,形成杂交样品。优选的检测 mRNA 或者基因组 DNA 的探针是能够与本文所述的 mRNA 或者基因组 DNA 序列杂交的标记核酸探针。核酸探针可以是,例如,全长核酸分子,或者其一部分,例如长度至少为 15、30、50、100、250 或者 500 个核苷酸并且在严谨条件下能够与合适的 mRNA 或者基因组 DNA 特定杂交的寡聚核苷酸。用于本发明诊断方法的合适探针如上所述(参见例如,在标题“本发明的核酸”下描述的探针和引物)。

[0031] 杂交样品保存在允许核酸探针与 TCF7L2 核酸特定杂交的条件下。此处使用的“特定杂交”表示准确杂交(例如,没有错配)。特定杂交可以在高度严谨条件或者中度严谨条件下进行,例如上面所述。在特别优选的方面,用于特定杂交的杂交条件是高严谨性。

[0032] 如果存在特定杂交,则利用标准方法检测。如果核酸探针和待测样品的 TCF7L2 核酸之间发生特定杂交,则 TCF7L2 具有多态性,或者是存在于核酸探针中的剪接变体。在该方法中可以同时使用不止一个核酸探针。任意一个核酸探针的特定杂交指示 TCF7L2 核酸的多态性,或者编码特定剪接变体的 TCF7L2 核酸的存在,可用于诊断 II 型糖尿病易感性,或者用于诊断 II 型糖尿病易感性的降低(或者指示抵抗 II 型糖尿病的保护性等位基因)。

[0033] 在 Northern 分析(参见 Current Protocols in Molecular Biology, Ausubel, F. 等人, eds., John Wiley & Sons, supra)中,上述杂交方法被用于鉴定多态性或者特定剪接变体的存在,所述多态性或者特定剪接变体与 II 型糖尿病易感性相关或者与 II 型糖尿病易感性降低相关。对 Northern 分析, RNA 待测样品通过适当方法取自个体。如上所述的核酸探针与个体 RNA 的特定杂交提示 TCF7L2 核酸的多态性,或者 TCF7L2 核酸编码的特定剪接变体的存在,因此可用于诊断 II 型糖尿病易感性,或者 II 型糖尿病易感性的降低(或者指示抵抗 II 型糖尿病的保护性等位基因)。

[0034] 使用核酸探针的典型实例,参见,例如,美国专利号 5,288,611 和 4,851,330。

[0035] 此外,在上述杂交方法中可使用肽核酸(PNA)探针替代核酸探针。PNA 是一种 DNA 模拟物,具有肽-样无机骨架,例如 N-(2-氨基乙基)甘氨酸单位,一种有机碱(A, G,

C, T 或者 U) 通过甲叉羰基接头与甘氨酸上的氮连接 (参见, 例如, Nielsen, P.E. 等人, *Bioconjugate Chemistry* 5, American Chemical Society, p. 1 (1994))。可以设计 PNA 探针与 TCF7L2 核酸特定杂交。PNA 探针与 TCF7L2 核酸的杂交可以用于诊断 II 型糖尿病的易感性或者 II 型糖尿病易感性的降低 (或者指示抵抗 II 型糖尿病的保护性等位基因)。

[0036] 在本发明的另一种方法中, 如果基因的改变 (突变) 或者多态性导致限制性内切位点的产生或者消失, 利用限制性消化的突变分析可用于检测基因突变。包括基因组 DNA 的待测样品从个体获得。聚合酶链式反应 (PCR) 可用于扩增来自待测个体测试样品基因组 DNA 中的 TCF7L2 核酸 (和, 必要时, 侧翼序列)。按所述 (参见 *Current Protocols in Molecular Biology*, supra) 进行 RFLP 分析。相关 DNA 片段的消化模式提示 TCF7L2 核酸突变或者多态性的存在或者缺失, 因此指示 II 型糖尿病易感性或者 II 型糖尿病易感性降低 (或者指示抵抗 II 型糖尿病的保护性等位基因) 的存在或者缺失。

[0037] 也可以用序列分析检测 TCF7L2 核酸的特定多态性。DNA 或者 RNA 待测样品取自待测个体。如果需要, 也可以用 PCR 或者其他合适的方法扩增基因或者核酸, 和 / 或其侧翼序列。利用标准方法检测 TCF7L2 核酸序列, 或者核酸的片段, 或者 cDNA, 或者 cDNA 的片段, 或者 mRNA, 或者 mRNA 的片段。根据情况, 将核酸, 核酸片段, cDNA, cDNA 片段, mRNA 或者 mRNA 片段的序列与基因或者 cDNA 或者 mRNA 的已知核酸序列比较。TCF7L2 多态性的存在指示个体具有 II 型糖尿病易感性或者 II 型糖尿病易感性的降低 (或者指示抵抗 II 型糖尿病的保护性等位基因)。

[0038] 通过利用扩增寡聚核苷酸与等位基因 - 特定寡聚核苷酸 (ASO) 探针的斑点杂交, 也可以使用等位基因 - 特定寡聚核苷酸检测 TCF7L2 核酸多态性的存在 (参见, 例如, Saiki, R. 等人, *Nature* 324:163-166 (1986))。“等位基因 - 特定寡聚核苷酸” (此处也称为“等位基因 - 特定寡聚核苷酸探针”) 是大约 10-50 碱基对的寡聚核苷酸, 优选的大约 15-30 碱基对, 所述寡聚核苷酸与 TCF7L2 核酸特定杂交, 并且包括与 II 型糖尿病易感性相关的多态性或者与 II 型糖尿病易感性降低 (或者指示抵抗 II 型糖尿病保护性等位基因) 相关的多态性。利用标准方法 (参见 *Current Protocols in Molecular Biology*, supra) 可以制备对特定 TCF7L2 核酸多态性特定的等位基因 - 特定寡聚核苷酸探针。为鉴定与 II 型糖尿病相关基因的多态性, 从个体获得 DNA 待测样品。PCR 可用于扩增 TCF7L2 核酸及其侧翼序列的全部或者片段。利用标准方法 (参见 *Current Protocols in Molecular Biology*, supra), 包括扩增 TCF7L2 核酸 (或者基因或者核酸的片段) 的 DNA 被斑点印迹, 斑点与寡聚核苷酸探针接触。然后检测探针与扩增的 TCF7L2 核酸特定杂交的存在。等位基因 - 特定寡聚核苷酸探针与个体 DNA 的杂交提示 TCF7L2 核酸多态性, 因此指示 II 型糖尿病易感性或者指示 II 型糖尿病易感性的降低 (或者指示抵抗 II 型糖尿病的保护性等位基因)。

[0039] 本发明还提供与基因或者核酸的对照或者变异等位基因杂交的等位基因 - 特定寡聚核苷酸, 所述基因或者核酸包括单核苷酸多态性或者与其互补。这些寡聚核苷酸可以是探针或者引物。

[0040] 等位基因 - 特定引物与靶 DNA 重叠多态性的位点杂交, 只起始引物显示完全互补的等位基因形式的扩增。参见 Gibbs, *Nucleic Acid Res.* 17, 2427-2448 (1989)。这种引物与第二引物联用, 所述第二引物在远端位点杂交。扩增起始于两个引物, 产生可检测的产物, 这提示存在特定等位基因形式。通常用第二对引物作为对照, 其中一个在多态性位点显

示单个碱基错配,另一个显示与远端位点的完全互补。单碱基错配阻止扩增,没有可检测的产物形成。当错配包括在与多态性比对的寡聚核苷酸的 3' - 最远端位置内,这种方法最适用,因为这个位置对从引物延伸是最具不稳定作用的(参见,例如, W093/22456)。

[0041] 通过添加这种类似物作为锁核酸(LNAs),引物和探针的大小可以减少到只有 8 个碱基。LNAs 是双环 DNA 类似物一个新的类型,其中呋喃糖环的 2' 和 4' 位置通过 O- 甲叉(氧-LNA), S- 甲叉(硫-LNA) 或者氨基甲叉(氨基-LNA) 基团连接。所有这些 LNA 变异体的共同点是对互补核酸的亲合力,据报道迄今为止这是 DNA 类似物中最高的。例如,在分别与互补 DNA 或者 RNA 的复合物中,特定的所有氧-LNA 九聚物分别具有 64EC 和 74EC 的熔解温度,与此相对,相应 DNA 九聚物的 DNA 和 RNA 的熔解温度均是 28EC。当 LNA 单体与标准 DNA 或者 RNA 单体联用时, T_m 也有显著增加。对于引物和探针,根据在何处包括 LNA 单体(例如, 3' 端, 5' 端, 或者在中间), T_m 可以显著增加。

[0042] 另一方面,与个体靶核酸序列片段互补的寡聚核苷酸探针的阵列可用于鉴定 TCF7L2 核酸多态性。例如,一方面,可以使用寡聚核苷酸阵列。寡聚核苷酸阵列通常包括大量不同的寡聚核苷酸探针,所述探针偶联在底物表面不同的已知位置。这些寡核苷酸阵列,也称为“Genechips™”,在本领域中被广泛描述,例如美国专利号 5, 143, 854 和 PCT 专利公开号 WO 90/15070 和 92/10092。通常利用机械合成法或者光导向的合成法,包括照相平版印刷方法和固相寡聚核苷酸合成法的组合,可以产生这些阵列。参见 Fodor 等人, Science 251:767-777(1991), Pirrung 等人, 美国专利号 5, 143, 854(还可参见 PCT 申请号 WO 90/15070) 和 Fodor 等人, PCT 公开号 WO 92/10092 和美国专利号 5, 424, 186, 所有教导在此引入作为参考。利用机械合成法合成这些阵列的技术描述于,例如美国专利号 5, 384, 261; 所有教导在此引入作为参考。在另一个实施例中,可以使用线性阵列。

[0043] 一旦制备出寡聚核苷酸阵列,将感兴趣的核酸与阵列杂交,扫描多态性。杂交和扫描通常的按照此处的方法以及描述于,例如,公开的 PCT 申请号 W092/10092 和 W095/11995, 和美国专利号 5, 424, 186 的方法进行,所有教导在此引入作为参考,总之,包括一种或者多种先前鉴定的多态性标记物的靶核酸序列通过公知的扩增技术进行扩增,例如,PCR。通常,这涉及引物序列的使用,所述引物序列从多态性上下游与靶序列的两条链互补。也可以使用不对称 PCR 技术。扩增的靶,通常包括标记,然后在合适条件下与阵列杂交。当阵列杂交完成和洗涤后,扫描阵列以检测与靶序列杂交的阵列上的位置。扫描获得的杂交数据通常是作为阵列上位置的函数的荧光强度形式。

[0044] 尽管主要以单一检测块方式描述,例如,用于检测单个多态性,阵列可以包括多重检测块,因此能够分析多个、特定特定多态性。另一方面,通常应理解检测块可以组合到单一阵列内或者多个、独立的阵列,这样在靶与阵列杂交期间,可以使用变化的、最优条件。例如,通常需要提供检测与那些位于基因组序列 A-T 富集片段区的多态性分开的那些位于基因组序列 G-C 延伸区内的多态性。这允许对各种情况杂交条件的单独进行最优化。

[0045] 寡聚核苷酸阵列用于多态性检测的其他应用可见于,例如,在美国专利号 5, 858, 659 和 5, 837, 832, 全部教导在此引入作为参考。可以使用其他的核酸分析方法检测 II 型糖尿病基因或者 II 型糖尿病基因编码变异体中的多态性。典型的方法包括直接人工测序 (Church 和 Gilbert, Proc. Natl. Acad. Sci USA 81:1991-1995(1988); Sanger, F. 等人, Proc. Natl. Acad. Sci USA 74:5463-5467(1977); Beavis 等人, 美

国专利号 5,288,644) ;自动荧光测序 ;单链构象多态性检测 (SSCP) ;夹板变性凝胶电泳 (CDGE) ;变性梯度凝胶电泳 (DGGE) (Sheffield, V. C. 等人, Proc. Natl Acad. Sci USA 86 :232-236(1989)), 迁移率分析 (Orita, M. 等人, Proc. Natl Acad. Sci USA 86 :2766-2770(1989)), 限制性酶切分析 (Flavell 等人, Cell 15 :25(1978) ;Geever, 等人, Proc. Natl. Acad. Sci USA 78 :5081(1981)) ;异质双链分析 ;化学错配切割 (CMC) (Cotton 等人, Proc. Natl. Acad. Sci USA 85 :4397-4401(1988)) ;RNase 保护试验 (Myers, R. M. 等人, Science 230 :1242(1985)) ;识别核苷酸错配的多肽的使用, 例如 E. coli mutS 蛋白 ;例如等位基因 - 特定 PCR。

[0046] 在本发明的一个方面, II 型糖尿病易感性或者 II 型糖尿病易感性降低 (或者指示抵抗 II 型糖尿病的保护性等位基因) 的诊断, 也可以通过定量 PCR (动态热循环) 的表达分析实现。利用 TaqMan[®] 检测, 这项技术可以鉴定 TCF7L2 核酸编码的多肽或者 TCF7L2 核酸编码的剪接变体的表达或者组成中改变的存在。TaqMan[®] 探针还可用于鉴定多态性以及患者是纯合子还是杂合体。此外, 变异体表达可以量化为物理不同或者功能不同的。

[0047] 在本发明的另一个方面, II 型糖尿病易感性或者 II 型糖尿病易感性降低 (或者指示抵抗 II 型糖尿病保护性等位基因) 的诊断, 可以通过各种方法, 包括酶联免疫吸附 (ELISAs), Western 印迹, 免测沉淀和免疫荧光法, 检测 TCF7L2 多肽的表达和 / 或组成来实现。对个体待测样品分析 TCF7L2 核酸编码的多肽的表达中改变和 / 或组成中改变的存在, 或者 TCF7L2 核酸编码的特定特定变异体的存在。在 TCF7L2 核酸编码多肽表达中的改变可能是, 例如, 定量的多肽表达 (即, 产生多肽的数量) 的改变 ;TCF7L2 核酸编码多肽的组成的改变是定性的多肽表达 (例如, 改变的 TCF7L2 多肽或者不同剪接变体的表达) 的改变。在一个优选的方面, II 型糖尿病易感性或者 II 型糖尿病易感性降低的诊断, 可以通过检测 TCF7L2 核酸编码的特定特定剪接变体或者剪接变体的特定模式来实现。

[0048] 这两种改变 (定量的和定性的) 均可以存在。此处使用的术语多肽表达或者组成中的“改变”, 是指待测样品中表达或者组成的改变, 与对照样品中 TCF7L2 核酸编码多肽的表达或者组成比较。对照样品是与待测样品对应的样品 (例如, 来自相同种类的细胞), 并且来自不受 II 型糖尿病易感性影响的个体。与对照样品相比, 待测样品中多肽的表达或者组成的改变指示 II 型糖尿病的易感性。类似地, 与对照样品相比, 待测样品中一种或多种不同剪接变体的存在, 或者待测样品中不同剪接变体显著地不同数量的存在指示 II 型糖尿病的易感性。可以使用检测 TCF7L2 核酸编码多肽的表达或者组成的不同的方法, 包括 : 光谱法, 比色法, 电泳, 等电点聚焦, 和免疫测定 (例如, David 等人, 美国专利号 4,376,110), 例如免疫印迹 (还可参见 Current Protocols in Molecular Biology, 尤其是第 10 章)。例如, 一方面, 可以使用能够结合多肽的抗体 (例如, 如上所述), 优选具有可检测标记的抗体。抗体可以是多克隆的, 或者更优选的, 单克隆的。可以使用完整的抗体, 或者其片段 (例如, Fab 或者 F(ab')₂)。术语“标记的”, 就探针或者抗体而言, 意图是包括探针或者抗体的直接标记, 通过将可检测物质偶联 (即, 物理连接) 到探针或者抗体, 以及探针或者抗体的间接标记, 通过与直接标记的另一种试剂的反应。间接标记的实例包括利用荧光标记的第二抗体和用生物素末端标记 DNA 探针检测第一抗体, 因此它可以利用荧光标记的链亲和素进行检测。

[0049] 利用如上所述的与突变 TCF7L2 核酸编码的多肽特定结合的抗体,或者与非突变核酸编码多肽特定结合的抗体,或者与核酸编码的特定特定剪接变体特定结合的抗体的 Western 印迹分析,可用于鉴定待测样品中特定剪接变体或者多态性的或者突变的 TCF7L2 核酸编码的多肽的存在,或者待测样品中特定剪接变体或者非多态性的或者非突变的核酸编码多肽的缺失。多态性的或者突变的核酸编码多肽的存在,或者非多态性的或者非突变的核酸编码多肽的缺失,以及 TCF7L2 核酸编码的特定剪接变体的存在(或者缺失),可用于诊断 II 型糖尿病的易感性。

[0050] 在这种方法的一方面,待测样品中 TCF7L2 核酸编码多肽的水平或者数量与对照样品中 TCF7L2 编码多肽的水平或者数量进行比较。待测样品中多肽的水平或者数量比对照样品中多肽的水平或者数量更高或者更低,以致差异是统计上有显著意义的,提示 TCF7L2 核酸编码多肽表达的改变,可用于诊断 II 型糖尿病的易感性。此外,待测样品中 TCF7L2 核酸编码的多肽的组成与对照样品中 TCF7L2 核酸编码多肽的组成进行比较(例如,不同剪接变体的存在)。与对照样品中多肽组成比较,待测样品中多肽组成的差异可用于诊断 II 型糖尿病的易感性。在另一方面,在待测样品和对照样品中均可检定多肽的水平或者数量和组成。待测样品中多肽的数量或者水平相对对照样品的差异;待测样品中组成相对对照样品的差异;或者数量或者水平的差异,和组成的差异,指示 II 型糖尿病的易感性。

[0051] 当与对照(疾病)样品比较时,相同方法可以相反地用于鉴定差异的存在。与对照的差异指示糖尿病易感性的降低,和/或指示抵抗 II 型糖尿病的保护性等位基因。

[0052] 标记物和单倍型的评估

[0053] 显示遗传多样性个体的种群具有不相同的基因组。此外,基因组在基因组许多位置显示个体之间的序列变化;换言之,在种群中存在许多多态性位点。在有些情况下,没有选择参考等位基因,对多态性位点的不同等位基因做出参考。此外,参考序列可指特定多态性位点。有时参考等位基因是指“野生型”等位基因,它通常被选作第一个测序的等位基因或者作为来自“未患病”个体的等位基因(例如,不显示疾病或者异常表型的个体)。不同于参考的等位基因称为“变异”等位基因。

[0054] 本文所述的“标记物”是指表征为特定的变异等位基因(即多态性位点)的基因组序列。标记物可以包括基因组内发现的任意变异类型的任意等位基因,包括 SNPs,微卫星,插入,缺失,重复和易位。

[0055] 此处报道的 SNP 命名是指官方参考 SNP(rs) ID 识别标签,由国家生物技术信息中心(NCBI)分配给每个独特的 SNP。

[0056] 本文所述的“单倍型”是指基因组 DNA 链的片段,其特征在于沿片段排列的遗传标记物(“等位基因”)的特定组合。在某一实施方式中,单倍型可以包括一种或者更多等位基因,两种或者更多等位基因,三种或者更多等位基因,四种或者更多等位基因,或者五种或者更多等位基因。遗传标记物是在与 TCF7L2 外显子 4LD 块相关“多态性位点”的特定特定“等位基因”。本文使用的“TCF7L2 外显子 4LD 块”是指 Chr10q 上的 LD 块,其中观察到变异与 II 型糖尿病的相关性。该 LD 块的 NCBI Build 34 位置是 114,413,084-114,488,013bp。本文所述的术语“易感性”包括增加的易感性和降低的易感性。因此,本发明的特定特定标记物和/或单倍型可以表征为 II 型糖尿病易感性的增加,其特征是相对危险度大于 1。造成 II 型糖尿病易感性增加的标记物和/或单倍型还可以被认为是“具有风险”,因为它们显

示疾病风险的增加。此外,本发明的标记物和 / 或单倍型可以表征为 II 型糖尿病易感性的降低,其特征是相对危险度小于 1。

[0057] 种群 (或者自然群体或者合成群体,例如,合成分子文库) 中其位置上可能具有超过一个序列的核苷酸位置在本文中被称为 “多态性位点”。当多态性位点长度是单核苷酸时,该位点称为单核苷酸多态性 (SNP)。例如,如果在特定特定染色体位置,群体的一个成员具有腺嘌呤而该群体的另一个成员在相同位置具有胸腺嘧啶,则这个位置是多态性位点,和更具体地,该多态性位点是 SNP。当它们位于 SNP 检测中的多态性位点时,本文所述的 SNP 标记物的等位基因是指碱基 A、C、G 或者 T。本领域技术人员明白通过检测或者阅读相反链,各种情况下都可以检测互补等位基因。因此,对包括 A/G 多态性的多态性位点,采用的检测可以测定两个可能碱基的百分比或者比例,即 A 和 G。此外,通过设计测定 DNA 模板上相反链的检测法,可以测定互补碱基 T/C 的百分比或者比例。定量的 (例如,按照相对危险度) 相同结果将从任意 DNA 链 (+ 链或者 - 链) 的检测获得。多态性位点允许基于取代、插入或者缺失的序列差异。例如,多态性微卫星在特定定位点具有多个小碱基重复 (例如 CA 重复), 其中普通群体中重复长度的数目不同。本文中多态性位点有关的各种形式的序列被称为多态性位点的 “等位基因”。因此,在早先的实例中,SNP 包括了腺嘌呤等位基因和胸腺嘧啶等位基因。发现的位于 TCF7L2 外显子 4LD 块内,与 II 型糖尿病相关的 SNPs 和微卫星标记物描述于表 2-7。

[0058] 通常,参考序列是指特定序列。不同于参考的等位基因称为 “变异” 等位基因。例如,NCBI Build 34 位置 114413084 和 114488013 之间 (等于 74929bp, 或者 74.9kb) 的参考基因组 DNA 序列,这是指染色体 10 内的位置,本文中描述为 SEQ ID NO : 1。本文使用的变异序列,是指不同于 SEQ ID NO : 1 的序列,但基本上是相似的。组成与 TCF7L2 外显子 4LD 块相关单倍型的遗传标记物是变异体。其他变异体可以包括影响多肽例如 TCF7L2 基因编码多肽的变化。当与参考核苷酸序列比较时,这些序列差异可以包括单个核苷酸或者不止一个核苷酸的插入或者缺失。这种序列差异可以导致移码 ; 至少一个核苷酸的改变,导致编码氨基酸的改变 ; 至少一个核苷酸的改变,导致终止密码子的提前产生 ; 几个核苷酸的缺失,可以导致一个或者多个核苷酸编码氨基酸的缺失 ; 一个或者几个核苷酸的插入,例如通过不等重组或者基因转换,可以导致读码框编码序列的中断 ; 全部或者部分序列的复制 ; 转座 ; 或者核苷酸序列的重排,如本文详细所述。这种序列变化改变核酸编码的多肽。例如,如果核酸序列的变化造成移码,移码可能导致编码氨基酸的改变,和 / 或可能导致提前终止密码子的产生,产生截短多肽。此外,II 型糖尿病相关多态性或者 II 型糖尿病易感性可能是一种或者多种核苷酸的同义变化 (即,不导致氨基酸序列改变的变化)。这类多态性可能,例如,改变剪接位点,影响 mRNA 的稳定性或者转运,或者影响编码多肽的转录或者翻译。它还可能改变 DNA 来增加结构变化的可能性,例如在肿瘤的体细胞水平发生的扩增或者缺失。参考核苷酸序列编码的多肽是具有特定特定参考氨基酸序列的 “参考” 多肽,变异等位基因编码的多肽是指具有变异氨基酸序列的 “变异” 多肽。

[0059] 多态性微卫星在特定特定定位点具有长度在 2-8 个核苷酸 (例如 CA 重复) 的多个小的碱基重复,,其中普通群体中重复长度的数目不同。Indel 是一种常见的多态性形式,包括小的插入或者缺失,通常只有几个核苷酸长度。

[0060] 本文描述的单倍型是在多态性位点具有特定等位基因的不同遗传标记物的组合,

例如 SNPs 和微卫星。单倍型可能包括不同遗传标记物的组合,因此,可以通过本领域已知的检测多态性位点序列的方法检测单倍型。例如,可以使用检测 SNPs 和 / 或微卫星标记物存在的基因分型标准技术,例如基于荧光的技术 (Chen, X. 等人, *Genome Res.* 9(5) : 492-98(1999)), PCR, LCR, 巢式 PCR 和用于核酸扩增的其他技术。可以在风险单倍型中鉴定到这些标记物和 SNPs。鉴定相关标记物和 SNPs 的一些方法包括使用连锁不平衡 (LD) 和 / 或 LOD 值。

[0061] 在本文描述的一些方法中,具有患 II 型糖尿病风险的个体是鉴定出风险标记物或者单倍型的个体。在一方面,风险标记物或者单倍型是指造成 II 型糖尿病风险 (或者易感性) 显著增加的标记物或者单倍型。在一个实施方式中,通过相对危险度测定标记物或者单倍型相关的显著度。在另一个实施方式中,显著度利用百分比测定。在一个实施方式中,显著增加的风险测定为相对危险度至少是约 1.2,包括但不限于: 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8 和 1.9。在另一个实施方式中,至少为 1.2 的相对危险度是显著的。在另一个实施方式中,至少为大约 1.5 的相对危险度是显著的。在另一个实施方式中,风险显著增加到至少大约 1.7 是显著的。在另一个实施方式中,风险显著增加至少大约 20%,包括但不限于大约 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% 和 98%。在另一个实施方式中,风险显著增加至少大约 50%。

[0062] 在本发明的其他实施方式中,标记物或者单倍型造成 II 型糖尿病风险的降低 (易感性降低)。在一个实施方式中,显著降低的风险测定为相对危险度小于 0.9,包括但不限于 0.9, 0.8, 0.7, 0.6, 0.5 和 0.4。在另一个实施方式中,显著的相对危险度小于 0.7。在另一个实施方式中,风险 (或者易感性) 降低至少大约 20%,包括但不限于大约 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% 和 98%。在另一个实施方式中,风险显著下降至少大约 30%。

[0063] 因此,术语“II 型糖尿病易感性”表示 II 型糖尿病增加的风险或者易感性或者降低的风险或者易感性,当某一等位基因、标记物、SNP 或者单倍型存在时,数量是显著的;显著度按上述方法测定。本文使用的术语“降低的风险”、“降低的易感性”和“保护”表示当某一其他等位基因、标记物、SNP 和 / 或某一其他单倍型存在时,相对危险度相应降低。但是应了解,鉴定增减的风险是否是医学显著的还依赖于各种因素,包括特定疾病,标记物或者单倍型,并且通常还包括环境因素。

[0064] TCF7L2 基因内或者包括部分 TCF7L2 基因的风险标记物或者单倍型,是相对于健康个体 (对照) 中存在的频率,在具有患 II 型糖尿病风险个体 (患病的) 中存在频率更高的标记物或者单倍型,其中标记物或者单倍型的存在指示 II 型糖尿病易感性。简单相关检验的一个实例是对 2×2 表进行的 Fisher 精确检验。从大量染色体中构建染色体群组的 2×2 表,所述染色体可以包括标记物或者单倍型两者,只包括标记物或者单倍型中的一个而不包括另一个,和标记物或者单倍型都不包括。

[0065] 在本发明的某些方面,风险标记物或者单倍型是与 II 型糖尿病显著相关的 TCF7L2 内或者邻近 TCF7L2 的风险标记物或者单倍型。在其他方面,风险标记物或者单倍型包括 TCF7L2 内或者邻近 TCF7L2 的风险标记物或者单倍型,TCF7L2 与 II 型糖尿病易感性显著相关。在本发明的特定实施方式中,如本文描述,标记物或者单倍型与 TCF7L2 外显子 4LD 块相关。

[0066] 可以使用检测 SNPs 和 / 或微卫星标记物存在的基因分型标准技术,例如基于荧光的技术 (Chen, 等人, *Genome Res.* 9, 492 (1999)), PCR, LCR, 巢式 PCR 和用于核酸扩增的其他技术。在优选的方面,该方法包括检定个体中 SNPs 和 / 或微卫星的存在或者频率,所述 SNPs 和 / 或微卫星在 TCF7L2 基因内或者包括部分 TCF7L2 基因,其中 SNPs 和 / 或微卫星相对健康对照个体的过度或者更高频率指示个体是 II 型糖尿病易感的。这种 SNPs 和标记物可以用作筛选工具的单倍型。可以在风险单倍型中鉴定这些标记物和 SNPs。例如,风险单倍型可以包括微卫星标记物和 / 或 SNPs,例如标记物 DG10S478 和 / 或 SNP rs12255372, rs7895340, rs11196205, rs7901695, rs7903146, rs12243326 或者 rs4506565。风险单倍型的存在指示 II 型糖尿病易感性的增加,从而提示个体属于本文所述治疗方法的目标人群。

[0067] 易感性变异体的鉴定

[0068] 利用期望最大化算法可以估计患者组和对照组中单倍型的频率 (Dempster A. 等人, *J. R. Stat. Soc. B*, 39 :1-38 (1977))。可以使用能够处理缺失基因型和相不确定性的这种实现算法。在零假设下,患者和对照被认为具有相同频率。利用相似方法,检验另一种假设,其中可能包括本文所述标记物的候选风险单倍型在患者中具有比对照更高的频率,而这两组中其他单倍型频率的比例被认为是相同的。在这两个假设下相似性分别被最大化,对应 1-df 似然比统计值被用于评定统计显著性。

[0069] 为了寻找连锁区域内的风险和保护性标记物和单倍型,例如,研究基因分型的标记物所有可能组合的相关性,假设这些标记物跨越实际区域。组合的患者和对照组可随机分成两组,与原始患者和对照组规模相同。然后重复标记物和单倍型分析,确定记录的最显著的 p- 值。可以重复这种随机化方案,例如,超过 100 次以构建 p- 值的经验分布。在一个优选的实施方式中, < 0.05 的 p- 值提示显著的标记物和 / 或单倍型关联。

[0070] 下面详细讨论单倍型分析。

[0071] 单倍型分析

[0072] 单倍型分析的一种常用方法涉及利用 NEsted MModels 的似然推断 (Gretarsdottir S., 等人, *Nat. Genet.* 35 :131-38 (2003))。在程序 NEMO 中进行该方法,适合多种多态性标记物、SNPs 和微卫星。方法和软件是特意设计用于病例对照研究,其目的是鉴定造成不同风险的单倍型组。它还是研究 LD 结构的工具。在 NEMO 中,借助 EM 算法,对将其作为缺失 - 数据问题的观测数据,直接计算最大似然估计值、似然比和 p- 值。

[0073] 检测信息

[0074] 即使对观测数据直接计算基于相似性的似然比检验可以被信赖得出正确的 p- 值,所述观测数据已捕获因为相不确定性和缺失基因型导致的信息丢失,但仍有兴趣了解因为信息不完整导致多少信息丢失。单倍型分析的信息测量法描述于 Nicolae 和 Kong (Technical Report 537, Department of Statistics, University of Statistics, University of Chicago ; *Biometrics*, 60 (2) :368-75 (2004)), 作为设定用于连锁分析的信息测量法的自然延伸,在 NEMO 中应用。

[0075] 统计分析

[0076] 对单一标记物与疾病的关联, Fisher 精确检验可用于计算各个个体等位基因的双边 p- 值。全部 p- 值表现为未校正的,用于多重比较,除非特别标出。显示的频率 (对微卫星, SNPs 和单倍型) 是与载体频率相对的等位基因的频率。为了最小化患者亲缘性造成的

任何偏差,所述患者作为家族被征集用于连锁分析,一级和二级亲属可以从患者列表中除去。此外,可以重复检验以通过延长差异调整过程而进行患者中任意剩余亲缘性的关联修正,该过程描述于 Risch, N. & Teng, J. (Genome Res., 8:1273-1288(1998))),和对血缘关系进行 DNA 收集合并 (pooling) (ibid),这样它可以被用于普通家族关系,提供校正和未校正的 p - 值用于比较。与预期一致,差异通常非常小。为了评定多重检验修正的单一标记物关联的显著度,我们可以利用相同基因型数据进行随机性检验。患者和对照的群组可以随机分配,多次重复关联分析(例如,直到 500,000 次), p - 值是重复检验的分数,所述重复检验得到一些标记等位基因的 p - 值小于或者等于利用原始患者和对照群组观察到的 p - 值。

[0077] 对单一标记物和单倍型分析,相对危险度 (RR) 和群体归因危险度 (PAR) 可以利用乘法模型(单倍型相对危险度模型)计算 (Terwilliger, J. D. & Ott, J, Hum. Hered. 42:337-46(1992) 和 FaIk, CT. & Rubinstein, P, Ann. Hum. Genet. 51(Pt 5j:227-33(1987))),即,人携带两个等位基因/单倍型的风险倍增。例如,如果 RR 是 A 相对于 a 的风险,则纯合子 AA 的风险是杂合子 Aa 风险的 RR 倍,是纯合子 aa 风险的 RR^2 倍。乘法模型具有良好特性,可以简化分析和计算-单倍型是独立的,即在患病群体以及对照群体内 Hardy-Weinberg 平衡。因此,患病的和对照的单倍型计数均具有多项分布,但在另外假设下具有不同的单倍型频率。具体地,对两个单倍型 h_i 和 h_j , 风险 (h_i)/ 风险 (h_j) = 风险 (f_i/p_i)/ 风险 (f_j/p_j), 其中 f 和 p 分别表示在患病群体和对照群体中的频率。如果真实模型不是倍增的,会存在一些功能损失,除极端情况外损失倾向于适度的。最重要地, p 值总是有效的,因为它们根据零假设计算。

[0078] 利用 NEMO 的连锁不平衡

[0079] 利用 D' 和 R^2 的标准定义可以计算标记物对之间的 LD (Lewontin, R., Genetics 49:49-67(1964); Hill, W. G. & Robertson, A. Theor. Appl. Genet. 22:226-231(1968))。利用 NEMO,通过最大似然法估算两个标记等位基因组合的频率,通过似然比检验计算连锁平衡的偏差。采用边缘等位基因可能性加权,通过计算两个标记物所有可能等位基因组合的平均值,将 D' 和 R^2 的定义延伸至包括微卫星。当描绘出全部标记物组合以说明特定区域中的 LD 结构,我们将 D' 绘制在左上角, p - 值绘制在右下角。如果需要,在 LD 图中标记物可以绘制成等距离的而不是根据它们的物理位置。

[0080] 连锁分析的统计方法

[0081] 多点、只在患者中有的等位基因-共用方法 (Multipoint, affected-only allele-sharing methods) 可用于分析来评定连锁证据。结果, LOD 值和非参数的连锁 (NPL) 值,可以利用程序 Allegro 获得 (Gudbjartsson 等人, Nat. Genet. 25:12-3(2000))。我们的基线连锁分析利用 Spairs 评分函数 (Whittemore, A. S., Halpern, J. Biometrics 50:118-27(1994); Kruglyak L. 等人, Am. J. Hum. Genet. 55:1347-63(1996)),指数等位基因-共用模式 (Kong, A. 和 Cox, N. J., Am. J. Hum. Genet. 61:1179-88(1997)) 和家族权重计算方法,所述权重计算方法是在 log 级别,处于相同地加权各患病对和相同地加权各家族之间中途。我们使用的信息测量是 Allegro 程序输出值的一部分,如果标记物基因型是完全不提供信息的,则信息值等于零,如果在患病的亲属间基因型测定了精确数量的共享等位基因,则信息值等于一 (Gretarsdottir 等人, Am. J. Hum. Genet., 70:593-603(2002))。用两种不同方法计算 p - 值,这里报道的是较不显著的结果。第一个 p - 值可以根据大样本理论

计算;在无连锁的零假设下 $Z_{1r} = \square(2[\log_e(10)LOD])$ 的分布大约是标准正态变量 (Kong, A. 和 Cox, N. J., Am. J. Hum. Genet. 61:1179-88(1997))。通过观察到的 LOD 值与其在零假设下完整数据取样分布的比较,可以计算第二个 p 值 (例如, Gudbjartsson 等人, Nat. Genet. 25:12-3(2000))。当数据由好几个家族组成时,这两个 p 值倾向于非常相似。

[0082] 易感性基因座的单倍型和“单倍型块”定义

[0083] 在一些实施方式中,标记物和单倍型分析涉及定义基于“单倍型块”的候选易感性基因座 (也称为“LD 块”)。已经有报道,部分人基因组可以分解成包含一些常见单倍型的系列离散的单倍型块;对于这些块,连锁不平衡数据几乎没有提供数据表明重组 (参见,例如, Wall, J. D. 和 Pritchard, J. K., Nature Reviews Genetics 4:587-597(2003); Daly, M. 等人, Nature Genet. 29:229-232(2001); Gabriel, S. B. 等人, Science 296:2225-2229(2002); Patil, N. 等人., Science 294:1719-1723(2001); Dawson, E. 等人, Nature 418:544-548(2002); Phillips, M. S. 等人, Nature Genet. 35:382-387(2003))。

[0084] 主要有两种定义这些单倍型块的方法:块可以定义为具有有限单倍型多样性的 DNA 区域 (参见,例如, Daly, M. 等人, Nature Genet. 29:229-232(2001); Patil, N. 等人, Science 294:1719-1723(2001); Dawson, E. 等人, Nature 418:544-54% (2002); Zhang, K. 等人, Proc. Natl. Acad. Sci. USA 99:7335-7339(2002)),或者是利用连锁不平衡鉴定的在过渡带之间具有广泛历史重组的区域 (参见,例如, Gabriel, S. B. 等人, Science 296:2225-2229(2002); Phillips, M. S. 等人, Nature Genet. 33:382-387(2003); Wang, N. 等人, Am. J. Hum. Genet. 77:1227-1234(2002); Stumpf, M. P., 和 Goldstein, D. B., Curr. Biol. 73:1-8(2003))。本文使用的术语“单倍型块”或者“LD 块”包括任意一种特征定义的块。

[0085] 单倍型块鉴定的典型方法公开于,例如,美国公开的专利申请号 20030099964, 20030170665, 20040023237 和 20040146870。单倍型块可以方便的用于对表型和单倍型情况之间的关联性作图。在各单倍型块中可以鉴定主要单倍型,然后可以鉴定出一组“标签”SNPs 或者标记物 (需要最小组的 SNPs 或者标记物以区别单倍型)。然后这些标签 SNPs 或者标记物可用于鉴定来自各组个体的样品,以便鉴定表型和单倍型之间的关联性。如果需要,相邻单倍型块可以同时评定,因为单倍型块间可能存在连锁不平衡。

[0086] 单倍型和诊断

[0087] 如本文所述,包括这种标记物的一些标记物和单倍型被发现可用于 II 型糖尿病易感性的测定 - 即发现它们可用于 II 型糖尿病易感性的诊断。发现特定标记物和单倍型在患有 II 型糖尿病的个体中的频率比没有 II 型糖尿病的个体中的频率更高。因此,这些标记物和单倍型具有在个体中检测 II 型糖尿病或者 II 型糖尿病易感性的预测价值。发现包括一些标签标记物的单倍型块 (即 TCF7L2 外显子 4LD 块) 在患有 II 型糖尿病的个体中的频率比在没有 II 型糖尿病的个体中的频率更高。因此,这些单倍型块内的“风险”标签标记物也具有在个体中检测 II 型糖尿病或者 II 型糖尿病易感性的预测价值。单倍型或者 LD 块内的“风险”标签标记物还可以包括区分单倍型的其他标记物,因为这些标记物类似的具有检测 II 型糖尿病或者 II 型糖尿病易感性的预测价值。作为人类基因组单倍型块结构的结果,与单倍型块 (LD 块) 相关的大量标记物或者其他变异体和 / 或包括这种标记物或者变异体的单倍型可以被认为与某一特征和 / 或表型相关。因此,位于本文定义的 TCF7L2 外显子 4LD 块内或者与 TCF7L2 外显子 4LD 块具有强 LD (特征为 r^2 大于 0.2) 的标记物和 / 或

单倍型可能与 II 型糖尿病相关（即它们造成增加或减少的 II 型糖尿病易感性）。这包括本文描述的标记物（表 6），但还可能包括与表 6 所列的一个或者多个标记物具有强 LD（特征为 r^2 大于 0.2）的其他标记物。可以通过本领域技术人员公知的方法鉴定这类其他变异体，例如通过对特定个体组 LD 块 A 基因组区域进行 DNA 测序，本发明也包括这类其他变异体。

[0088] 如本文所述，发现 TCF7L2 外显子 4LD 块内的一些标记物的频率在患有 II 型糖尿病的个体中降低，包括两种或更多表 13、20 和 21 所列的那些标记物的单倍型也被发现在患有 II 型糖尿病的个体中频率降低。因此这些标记物和单倍型对 II 型糖尿病是保护性的，即它们造成携带这些标记物和 / 或单倍型的个体患 II 型糖尿病的风险降低。

[0089] 本文描述的单倍型和标记物，在一些情况下，是不同遗传标记物的组合，例如，SNPs 和微卫星。因此，可以通过本领域已知的和 / 或本文所述的检测多态性位点序列的方法检测单倍型。此外，利用标准技术可以验证某些单倍型或者标记物组和疾病表型之间的相关性。简单的相关性检验的一个典型实例是对 2×2 表进行 Fisher 精确检验。

[0090] 在特定实施方式中，与 TCF7L2 外显子 4LD 块相关的标记物或者单倍型，是相对于其在健康个体（对照）中存在的频率，在具有患 II 型糖尿病风险个体（患病的）中存在频率更高的标记物或者单倍型，其中标记物或者单倍型的存在指示 II 型糖尿病或者 II 型糖尿病易感性。在其他实施方式中，与 TCF7L2 外显子 4LD 块相关的一个或者多个标记物具有连锁不平衡的风险标签标记物，是相对于健康个体（对照）中存在的频率，在具有患 II 型糖尿病风险个体（患病的）中存在频率更高的标签标记物，其中标签标记物的存在指示 II 型糖尿病增加的易感性。在其他实施方式中，与 TCF7L2 外显子 4LD 块相关的一个或者多个标记物具有连锁不平衡的风险标签标记物，是相对于健康个体（对照）中存在的频率，在具有患 II 型糖尿病风险个体中存在频率更高的标记物，其中标记物的存在指示 II 型糖尿病易感性。

[0091] 在本文描述的一些方法中，具有患 II 型糖尿病风险的个体是鉴定出风险标记物或者单倍型的个体。在一个实施方式中，标记物或者单倍型关联性的强度采用相对危险度 (RR) 来测定。RR 是携带标记物或者单倍型的一个拷贝个体的发病率相对不携带标记物或者单倍型个体的发病率的比率。这个比例相当于携带两个拷贝标记物或者单倍型个体的发病率相对携带一个拷贝标记物或者单倍型个体的发病率的比率。在一个实施方式中，标记物或者单倍型具有至少为 1.2 相对危险度。在其他实施方式中，标记物或者单倍型具有至少 1.3，至少 1.4，至少 1.5，至少 2.0，至少 2.5，至少 3.0，至少 3.5，至少 4.0，或者至少 5.0 的相对危险度。

[0092] 在本发明的其他方法中，患 II 型糖尿病风险降低（或者易感性降低）的个体是鉴定出保护性标记物或者单倍型的个体。在这种情况下，相对危险度 (RR) 小于 1。在一个实施方式中，标记物或者单倍型具有小于 0.9 的相对危险度。在另一个实施方式中，标记物或者单倍型具有的相对危险度小于 0.8，小于 0.7，小于 0.6，小于 0.5 或者小于 0.4。

[0093] 遗传检验的实用性

[0094] 与造成发展 II 型糖尿病风险的遗传变异有关的知识，为能够区分患病风险增加个体（即风险变异体携带者）和患病风险降低个体（即保护性变异体的携带者）的遗传 - 检验提供了可能。对属于上述两组的个体进行遗传检验的核心价值，是在早期诊断疾

病的可能性,以及给临床医生提供关于疾病预后 / 侵袭性信息,以便能够采用最合适的治疗方法。例如,II 型糖尿病遗传检验的应用使在早期检测疾病成为可能,这导致在早期就可以进行治疗检查,从而可以最小化疾病的有害作用和 II 型糖尿病造成的严重健康后果。

[0095] 治疗方法

[0096] 在本发明的另一个实施方式中,该方法可用于治疗 II 型糖尿病。本文使用的术语“治疗”不仅指改善 II 型糖尿病相关症状,而且还指预防或者延缓 II 型糖尿病的发病;减轻 II 型糖尿病症状的严重程度或者频率;和 / 或减轻与其他药物联合治疗的必要性,所述其他药物用来改善 II 型糖尿病的相关症状。一方面,待接受治疗的个体是 II 型糖尿病易感的(增加风险)的个体(例如,具有以下等位基因存在的个体:标记物 DG10S478 除了 0 等位基因以外其他等位基因的存在;SNPrs12255372 的 T 等位基因的存在;SNP rs7895340 的 A 等位基因的存在;SNP rs11196205 的 C 等位基因的存在;SNP rs7901695 的 C 等位基因的存在;SNP rs7903146 的 T 等位基因的存在;SNP rs12243326 的 C 等位基因的存在;或者 SNP rs4506565 的 T 等位基因的存在)。

[0097] 在本发明的其他实施方式中,该方法可用于治疗其他 TCF7L2 相关的疾病或者病症。TCF7L2 治疗剂可以用于 II 型糖尿病的治疗方法,以及其他 TCF7L2 相关疾病或者病症的治疗方法。

[0098] 治疗方法(预防和 / 或治疗的)使用 TCF7L2 治疗剂。“TCF7L2 治疗剂”是一种直接地或者间接改变(例如,增强或者抑制)TCF7L2 的多肽活性和 / 或核酸表达(例如,通过改变与 TCF7L2 相互作用蛋白的活性或者核酸表达,例如 Wnt 信号途径或者钙粘蛋白途径中的蛋白(例如,beta- 连环蛋白))的药剂。在某些实施方式中,TCF7L2 治疗剂改变 TCF7L2 的活性和 / 或核酸表达。

[0099] TCF7L2 治疗剂可以通过各种方法改变 TCF7L2 多肽活性或者核酸表达,例如,通过提供额外 TCF7L2 多肽或者通过上调 TCF7L2 核酸的转录或者翻译;通过改变 TCF7L2 多肽的翻译后加工;通过改变 TCF7L2 剪接变异体的转录;或者通过干扰 TCF7L2 多肽活性(例如,通过结合 TCF7L2 多肽),或者通过结合与 TCF7L2 相互作用的另一个多肽,通过改变 TCF7L2 核酸的(例如,下调)表达、转录或者翻译,或者通过改变(例如,激动或者拮抗)活性。

[0100] 典型的 TCF7L2 治疗剂包括下列物质:本文描述的核酸或者其片段或者其衍生物,特别是编码本文描述多肽的核苷酸和包括这类核酸的载体(例如,基因,cDNA,和 / 或 mRNA,例如编码 TCF7L2 多肽或者活性片段或者其衍生物的核酸,或者寡聚核苷酸;或者其互补物,或者片段或者其衍生物,和 / 或 II 型糖尿病核酸或者片段或者其衍生物编码的其他剪接变异体);本文描述的多肽和 / 或 TCF7L2 核酸或者片段或者其衍生物编码的剪接变异体;其他多肽(例如,TCF7L2 受体);TCF7L2 结合剂;或者影响(例如,增加或者降低)活性的药剂,抗体,例如抗突变 TCF7L2 多肽的抗体,或者抗非突变 TCF7L2 多肽的抗体,或者抗上述 TCF7L2 核酸编码的特定剪接变异体的抗体;模拟肽;融合蛋白或者其前体药物;核酶;其他小分子;和其他改变(例如,增强或者抑制)TCF7L2 核酸表达的药剂,或者调节 TCF7L2 剪接变异体转录的药剂(例如,影响那个剪接变异体表达,或者影响各剪接变异体表达数量的药剂)。其他典型的 TCF7L2 治疗剂包括影响胰岛素信号和 / 或胰高血糖素, GLP-1 或者 GIP 信号的化合物。如果需要,可以同时使用不止一种 TCF7L2 治疗剂。

[0101] 在优选的实施方式中,TCF7L2 治疗剂是干扰 TCF7L2 活性的药剂,例如,干扰

TCF7L2 结合或者 TCF7L2 与 beta- 连环蛋白或其他蛋白相互作用的药剂 (参见, 例如, Fasolini, 等人, J. Biol. Chem 278 (23) :21092-06 (2003))。其他 TCF7L2 治疗剂包括影响 Wnt 信号途径的药剂或者影响钙粘蛋白途径的药剂。典型药剂包括例如那些用于癌症治疗的药剂, 包括, 例如, 蛋白例如 DKK 蛋白; APC 的 beta- 连环蛋白结合结构域或者 Axin; 因子例如 IDAX, AXAM 和 ICAT; 反义寡聚核苷酸或者干扰 RNA (RNAi), 例如伴随 Vitravene 的使用; 溶瘤细胞病毒载体; 和其他化合物 (参见, 例如, Luu 等人, Current Cancer Drug Targets 4 :6530671 (2004)); 小分子拮抗剂, 包括, 例如, ZTM00990, PKF118-310, PKF 118-744, PKF115-584, PKF222-815, CGP049090, NPDDG39.024 和 NPDDG1.024, 如 Lepourcelet 等人所述 (参见, 例如, Lepourcelet 等人, Cancer Cell 5 :91-102 (2004)); 美国专利 6,762,185 中描述的化合物; 美国专利申请 20040005313, 20040072831, 20040247593, 或者 20050059628 中描述的化合物。其他典型的 TCF7L2 治疗剂包括 gsk3 抑制剂, 包括, 例如, 在美国专利 6,057,117 ;6,153,618 ;6,417,185 ;6,465,231 ;6,489,344 ;6,512,102 ;6,608,063 ;6,716,624 ;6,800,632 ;和公开的美国专利申请 20030008866 ;20030077798 ;20030130289 ;20030207883 ;2000092535 ;和 200500851 中描述的那些 gsk3 抑制剂。说明书中引用的所有参考文献、专利和专利申请的全部教导在此完整引入。

[0102] 其他典型的 TCF7L2 治疗剂如下面药剂表所示。

[0103] 药剂表

[0104]	化合物名称	化合物名称 (利用来自 MDL 信息系统的 Autonom, ISIS Draw version 2.5 产生)	公司	化合物参考文献	适应症
--------	-------	--	----	---------	-----

AR-0133418(S N-4521)	1-(4-甲氧苯甲基)-3-(5-硝基噻唑 -2-基)-尿素	AstraZeneca		AD
AR-025028	NSD	AstraZeneca		
CT-98023	N-[4-(2,4-二氯苯基)-5-(1H-咪唑 -2-基)-嘧啶-2-基]-N'-(5-硝基-吡 啶-2-基)-乙烷-1,2-二胺	Chrion Corp		非胰岛素依赖 性糖尿病
CT-20026	NSD	Chrion Corp	Wagman 等人, Curr Pharm. Des 2004:10(10)1105-37	非胰岛素依赖 性糖尿病
CT-21022	NSD	Chrion Corp		非胰岛素依赖 性糖尿病
CT-20014	NSD	Chrion Corp		非胰岛素依赖 性糖尿病
CT-21018	NSD	Chrion Corp		非胰岛素依赖 性糖尿病
CHIR-98025	NSD	Chrion Corp		非胰岛素依赖 性糖尿病
CHIR-99021	NSD	Chrion Corp	Wagman 等人, Curr Pharm. Des 2004:10(10)1105-37	非胰岛素依赖 性糖尿病
CG-100179	NSD	CrystalGenomi cs 和 Yuyu	WO-2004065370	糖尿病 (韩国)
	4-[2-(4-二甲氨基-3-硝基-苯基氨 基)-嘧啶-4-基]-3,5-二甲基-1H- 吡咯-2-腈	Cyclacel Ltd.		非胰岛素依赖 性糖尿病,等等
NP-01139, NP-031112, NP03112, NP-00361	4- 苄 基 -2- 甲 基 -[1,2,4]thiadiazolidine-3,5-二酮	Neuropharma SA		中枢神经系统 障碍, AD
	3-[9-氟-2-(哌啶-1-羰基)-1,2,3,4- 四氢-[1,4]diazepino[6,7,1-hi] 咪 唑-7-基]-4-咪唑[1,2-a]吡啶-3-基- 吡咯-2,5-二酮	Eli Lilly & Co		非胰岛素依赖 性糖尿病
GW-784725x, GW-784775, SB-216763, SB-415286	环戊烷羧酸(6-吡啶-3-基 -furo[2,3-d]嘧啶-4-基)-酰胺	GSK	WO-03034447(参考 的化合物: 4-[2-(2- 溴苯基)-4(4-氟苯 基)-1H-咪唑-5-基] 吡啶	非胰岛素依赖 性糖尿病, 神 经退行性疾病

[0105]

[0106]

NNC-57-0511, NNC-57-0545, NNC-57-0588	1-(4-氨基-咪唑-3-基)-5-吡啶-1-基甲基-1H-[1,2,3]三唑-4-羧酸 [1-吡啶-4-基-甲基-(E)-ylidene]- 酰肼	Novo Nordisk		非胰岛素依赖性糖尿病
CP-70949	NSD	Pfizer		降血糖药
VX-608	NSD			脑血管局部缺血, 非胰岛素依赖性糖尿病
KP-403 类型	NSD	Kinetek		核因子 kappaB 调节剂, 消炎, 细胞循环抑制剂, 糖原合酶激酶-3beta 抑制剂
BYETTA®	Exenatide: C ₁₈₄ H ₂₈₂ N ₃₀ O ₆₀ S- 氨基酸序列: H-His-Gly-Glu-Gly-Thr-Phe-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 ₂	Amylin/Eli Lilly & Co		非胰岛素依赖性糖尿病
Vildagliptin (LAF237)	NSD	Novartis		非胰岛素依赖性糖尿病 - DPP-4 抑制剂

[0107] NSD = 结构未公开 (在 Iddb3 中)

[0108] 以治疗有效量 (即, 如上述足以“治疗”的数量) 的 TCF7L2 治疗剂给药。对特定个体病症或者疾病治疗有效的剂量依赖于疾病的症状和严重程度, 这可以通过标准临床技术来确定。此外, 可以选择体外或者体内检测帮助确定最适剂量范围。剂型中使用的精确剂量还依赖于给药途径和疾病或者病症的严重性, 这应该根据医师的判断和各个患者的情况来决定。可以来自体外或者动物模型检测系统的剂量反应曲线推断出有效剂量。

[0109] 在一个实施方式中, 可以使用核酸 (例如, 编码 TCF7L2 多肽的核酸); 或者编码 TCF7L2 多肽或者剪接变体、衍生物或者其片段的其他核酸, 单独的或者在如上所述的药物组合物中。例如, 编码 TCF7L2 多肽的 TCF7L2 基因或者核酸或者 cDNA, 独立的或包括在载体内, 可以被导入细胞 (体外或体内), 这样细胞产生天然的 TCF7L2 多肽。必要时, 已经转化了基因或者 cDNA 或者包括基因、核酸或者 cDNA 的载体的细胞可以被导入 (或者再导入) 患有疾病的个体。因此, 本身缺乏天然 TCF7L2 表达和活性, 或者具有突变的 TCF7L2 表达和活性, 或者具有疾病相关 TCF7L2 剪接变体表达的细胞, 可以被工程化表达 TCF7L2 多

肽或者 TCF7L2 多肽的活性片段（或者 TCF7L2 多肽的不同变异体）。在某些实施方式中，编码 TCF7L2 多肽或者活性片段或者其衍生物的核酸，可以被导入表达载体，例如病毒载体，然后载体可以导入动物的合适细胞中。可以使用其他基因转移系统，包括病毒和非病毒转移系统。此外，还可以使用非病毒基因转移方法，例如磷酸钙共沉淀，机械技术（例如，微量注射）；经由脂质体膜融合介导的转移；或者直接 DNA 吸收。

[0110] 此外，在本发明的另一个实施方式中，本发明的核酸；与本发明核酸互补的核酸；或者这类核酸的一部分（例如，如下所述的寡聚核苷酸），可用于“反义”治疗，其中在原位给药或者产生与 II 型糖尿病基因的 mRNA 和 / 或基因组 DNA 特异杂交的核酸（例如，寡聚核苷酸）。与 mRNA 和 / 或 DNA 特异杂交的反义核酸抑制 TCF7L2 多肽的表达，例如，通过抑制翻译和 / 或转录。反义核酸的结合可以通过常规碱基对互补，或者，例如，就结合 DNA 双链而言，通过在双螺旋大沟内的特异相互作用。

[0111] 本发明的反义构建体可以例如作为上述表达质粒被输送。当质粒在细胞中转录时，产生与编码 TCF7L2 多肽的 mRNA 和 / 或 DNA 的部分片段互补的 RNA。此外，反义构建体可以是离体产生并被导入细胞的寡聚核苷酸探针；然后它通过与多肽的 mRNA 和 / 或基因组 DNA 的杂交抑制表达。在一个实施方式中，寡聚核苷酸探针是修饰的寡聚核苷酸，能够抵抗内源核酸酶，例如核酸外切酶和 / 或核酸内切酶，因此它们在体内是稳定的。用作反义寡聚核苷酸的示例性核酸分子是 DNA 的氨基磷酸酯，硫代磷酸酯和甲基膦酸酯类似物（还可参见美国专利号 5, 176, 996 ; 5, 264, 564 ; 和 5, 256, 775）。此外，用于反义治疗的寡聚物的常见构建方法还描述于，例如，Van der Krol 等人，(BioTechniques 6 : 958-976 (1988)) ; 和 Stein 等人，(Cancer Res. 48 : 2659-2668 (1988))。就反义 DNA 而言，来自翻译起始位点的寡聚脱氧核糖核苷酸是优选的。

[0112] 为进行反义治疗，设计与编码 TCF7L2 基因的 mRNA 互补的寡聚核苷酸 (mRNA, cDNA 或者 DNA)。反义寡聚核苷酸与 TCF7L2 mRNA 转录产物结合，并阻止翻译。完全互补，虽然是优选的，但不是必须的。本文所指的与 RNA 部分“互补”的序列，表示序列具有足够的互补性能够与 RNA 杂交形成稳定的双链；就双链反义核酸来说，因此可以检测双链 DNA 的单链，或者也可以检测三链形成。杂交能力取决于互补程度和反义核酸长度，如上面详细所述。通常，杂交核酸越长，就可以包含与 RNA 更多的碱基错配，仍然能够形成稳定的双链（或者三链，根据情况而定）。本领域技术人员能够利用标准方法确定可耐受的错配度。

[0113] 用于反义治疗的寡聚核苷酸可以是 DNA, RNA 或者其嵌合混合物或者衍生物或者修饰形式，单链或者双链的。可以在例如碱基基团，糖基团或者磷酸盐骨架上对寡聚核苷酸进行修饰，以提高分子、杂交的稳定性。寡聚核苷酸可以包括其他附加基团例如肽（例如体内用于定位到宿主细胞受体），或者促进穿过细胞膜转运（参见，例如，Letsinger 等人，Proc. Natl. Acad. Sci. USA 86 : 6553-6556 (1989) ; Lemaitre 等人，Proc. Natl. Acad. Sci. USA 84 : 648-652 (1987) ; PCT 国际公开号 : W088/09810) 或者（参见，例如，PCT 国际公开号 : W0 89/10134) 血脑屏障转运的药剂，或者杂交触发断裂剂（参见，例如，Krol 等人，BioTechniques 6 : 958-976 (1988)) 或者嵌入剂（参见，例如，Zon, Pharm. Res. 5 : 539-549 (1988))。所以，寡聚核苷酸可以结合到另一个分子上（例如，肽，杂交触发交联剂，转运试剂，杂交触发断裂剂）。

[0114] 反义分子被递送到体内表达 TCF7L2 的细胞。可以使用多种方法将反义 DNA 或者

RNA 递送到细胞；例如，反义分子可以直接注射到组织部位，或者设计定位到目标细胞的修饰反义分子（例如，与肽或者抗体连接的反义分子，所述肽或者抗体特异结合目标细胞表面表达的受体或者抗原）可以全身给药。此外，在一个优选的实施方式中，使用重组 DNA 构建体，其中反义寡聚核苷酸置于强启动子控制下（例如，*pol III* 或者 *pol II*）。利用这类构建体转染患者的靶细胞，可以引起足量单链 RNAs 的转录，它们将与内源 TCF7L2 转录产物形成互补碱基对，从而阻止 TCF7L2 mRNA 的翻译。例如，载体可以导入体内，因此它被细胞摄取并指导反义 RNA 的转录。这类载体可以保留为附加体或者与染色体整合，只要它能够转录产生所需的反义 RNA。这类载体可以通过本领域和上述的标准的 DNA 重组技术方法进行构建。例如，质粒，粘粒，YAC 或者病毒载体可用于制备能够直接导入组织部位的重组 DNA 构建体。此外，可以使用选择性感染目标组织的病毒载体，这种情况下可以通过其他途径实现给药（例如，全身性的）。

[0115] 利用靶向同源重组，灭活或者“敲除”基因、核酸或者其启动子，也可以减少内源 TCF7L2 多肽的表达（例如，参见 Smithies 等人，*Nature* 317 :230-234(1985) ;Thomas & Capecchi, *Cell* 51 :503-512(1987) ;Thompson 等人，*Cell* 5 :313-321(1989)）。例如，可以使用与内源基因或者核酸（或者编码核酸区域或者调节区域）同源的 DNA 侧翼的突变、无功能的基因或者核酸（或者完全无关的 DNA 序列），包括或者不包括选择标记物和 / 或负选择标记物，转染体内表达基因或者核酸的细胞。通过插入靶向同源重组的 DNA 构建体，引起基因或者核酸的灭活。重组 DNA 构建体可以直接给药，或者利用如上所述的合适载体在体内定位到所需部位。此外，利用相似方法可以增加非突变基因或者核酸的表达：靶向同源重组可用于插入 DNA 构建体，所述 DNA 构建体在细胞内包括非突变的功能基因或者核酸，代替上述突变的 TCF7L2。在另一个实施方式中，靶向同源重组可用于插入包括一种核酸的 DNA 构建体，所述核酸编码与细胞中存在的不同的 II 型糖尿病多肽变异体。

[0116] 此外，通过将互补脱氧核糖核苷酸序列定位到 TCF7L2 核酸的调节区域（即，TCF7L2 启动子和 / 或增强子）形成三螺旋结构，阻止体内靶细胞中 TCF7L2 核酸的转录，内源 TCF7L2 核酸的表达可以被降低（一般参见，Helene, C, *Anticancer Drug Des.*, 6(6) : 569-84(1991) ;Helene, C. 等人., *Ann. N. Y. Acad. Sci* 660 :27-36(1992) ;和 Maher, L. J., *Bioassays* 14(12) :807-15(1992)）。同样地，本文描述的反义构建体，通过拮抗 TCF7L2 蛋白的正常生物活性，可用于组织处理，例如，组织分化，体内和体外组织培养。此外，反义技术（例如，反义分子的显微注射，或者和质粒转染，所述质粒的转录产物是 II 型糖尿病基因 mRNA 或者基因序列的反义序列），可用于研究疾病进展中的 TCF7L2 功能或者 TCF7L2 与其结合剂的相互作用，以及成人组织中 TCF7L2 的正常细胞功能或者 TCF7L2 与其结合剂的相互作用。这类技术可用于细胞培养中，但也可用于转基因动物的制备。

[0117] 在本发明的另一个实施方式中，本文描述的其他 TCF7L2 治疗剂还可用于 II 型糖尿病基因的治疗。治疗剂可以在上述组合物中，或者单独给药。它们可以全身给药，或者定位到特定组织。治疗剂可以通过各种方法制备，包括化学合成；重组产生；体内产生（例如，转基因动物，例如美国专利号：4,873,316, Meade 等人），例如，并可以用本文描述的那些标准方法分离。

[0118] 还可以使用任意上述治疗方法的组合（例如，非突变多肽给药和针对 TCF7L2 突变 mRNA 的反义治疗联用；TCF7L2 核酸编码第一种剪接变异体给药和针对 TCF7L2 核酸编码第

二种剪接的反义治疗联用)。

[0119] 评定对 TCF7L2 治疗剂反应可能性的方法

[0120] 本发明还涉及评定个体对 TCF7L2 治疗剂反应可能性的方法。在该方法中, 评定和 TCF7L2 基因相关的标记物或者单倍型, 如上所述涉及评定个体对 II 型糖尿病的易感性。与 II 型糖尿病易感性(增加的风险)相关的等位基因, 标记物, SNP 或者单倍型(例如, 标记物 DG10S478 除了 0 等位基因以外其他等位基因; SNP rs12255372 的 T 等位基因; SNP rs7895340 的 A 等位基因; SNP rs11196205 的 C 等位基因; SNP rs7901695 的 C 等位基因; SNP rs7903146 的 T 等位基因; SNP rs12243326 的 C 等位基因; 或者 SNP rs4506565 的 T 等位基因; TCF7L2 外显子 4LD 块相关标记物, 例如 TCF7L2 外显子 4LD 块相关的风险单倍型); 指示对 TCF7L2 治疗剂阳性反应的可能性。“阳性反应的可能性”表示个体比不具有本文所述 II 型糖尿病易感性(增加的风险)相关等位基因、标记物、SNP 或者单倍型的个体更可能对 TCF7L2 治疗剂显示阳性反应。对 TCF7L2 治疗剂的“阳性反应”是显示 II 型糖尿病治疗的生理反应。如上所述, “治疗”不仅指改善 II 型糖尿病相关症状, 而且还指预防或者延缓 II 型糖尿病的发病; 减轻 II 型糖尿病症状的严重程度或者频率; 和 / 或减轻与其他药物联合治疗的必要性, 所述其他药物用来改善 II 型糖尿病的相关症状。

[0121] 药物组合物

[0122] 本发明还涉及药物组合物, 包括改变 TCF7L2 活性或者另外影响 Wnt 信号途径或者钙粘蛋白途径, 或者可用作 TCF7L2 治疗剂的药物。药物组合物还可以与生理上可接受的载体或者赋形剂一起制备成药物组合物。载体和组合物是无菌的。剂型应当适合给药方式。

[0123] 合适的药学可接受的载体包括但不限于, 水, 盐溶液(例如, NaCl), 盐水, 缓冲盐水, 醇, 甘油, 乙醇, 阿拉伯树胶, 植物油, 苯甲醇, 聚乙二醇, 明胶, 糖例如乳糖, 直链淀粉或者淀粉, 葡萄糖, 硬脂酸镁, 滑石, 硅酸, 粘性石蜡, 芳香油, 脂肪酸酯, 羟甲基纤维素, 聚乙烯吡咯烷酮, 等等, 以及其组合。如果需要, 药物制剂可以与辅料混合, 例如, 润滑剂, 防腐剂, 稳定剂, 湿润剂, 乳化剂, 用于影响渗透压的盐, 缓冲液, 着色剂, 调味剂和 / 或芳香物质和类似物, 它们均不与活性药剂产生有害反应。

[0124] 如果需要, 组合物还可以包括少量湿润或者乳化剂, 或者 pH 缓冲剂。组合物可以是液体溶液, 悬浮液, 乳剂, 片剂, 药丸, 胶囊, 缓释剂型, 或者粉末。组合物可以与常规粘合剂和载体例如三酸甘油酯制备为栓剂。口服剂型可以包括标准载体例如药物等级的甘露醇, 乳糖, 淀粉, 硬脂酸镁, 聚乙烯吡咯烷酮, 糖精钠, 纤维素, 碳酸镁, 等等。

[0125] 导入这些组合物的方法包括, 但不限于, 真皮内的, 肌肉的, 腹内的, 眼内的, 静脉内的, 皮下的, 局部的, 口服的和鼻内的。其他合适的导入方法还包括基因治疗(如下所述), 可充电或者可生物降解装置, 粒子加速装置(“基因枪”)和缓释聚合装置。本发明的药物组合物还可以作为与其他药剂联合治疗的一部分给药。

[0126] 组合物可以按照常规步骤制备适于给药人类的药物组合物。例如, 用于静脉内给药的组合物通常是无菌等渗水相缓冲液中的溶液。当必要时, 组合物还可以包括增溶剂和局部麻醉剂, 以便在注射部位镇痛。通常, 成分是被分别供给或者混合为单位剂型形式, 例如, 作为密闭容器例如安瓿或者标记活性药剂数量的小袋(sachette)中的冻干粉末或者无水浓缩物。当组合物通过输液给药时, 它可以在包含无菌药物等级水、盐水或者葡萄糖 / 水的输液瓶中配制。当组合物通过注射给药时, 可以提供包含灭菌注射水或者盐水的安瓿,

这样在给药前就可以将成分混合。

[0127] 对于局部施用,可以使用非喷雾形式,包括一种载体的粘性的至半固体或者固体形式,所述载体适于局部施用并且其动态粘滞度优选的大于水。合适制剂包括但不限于溶液,悬浮液,乳剂,乳膏剂,软膏,粉末,灌肠剂,洗液,溶胶,搽剂,油膏剂,气雾剂,等等,如果需要,它们被灭菌或者与辅料混合,例如,防腐剂,稳定剂,湿润剂,缓冲液或者用于影响渗透压的盐,等等。药剂可以整合到化妆品制剂中。对于局部施用,可喷雾的气雾剂也是合适的,其中活性成分,优选的与固体或者液体惰性载体材料组合,包装在挤瓶中或者与增压的挥发性、通常是气体的喷射剂混合,例如压缩空气。

[0128] 本文描述的药剂可以配制为中性或者盐形式。药学可接受的盐包括那些与自由氨基形成的盐,例如源自盐酸,磷酸,醋酸,草酸,酒石酸,等等的盐,和那些与自由羧基形成的盐,例如来自钠,钾,铵,钙,氢氧化铁,异丙胺,三乙胺,2-乙基氨基乙醇,组氨酸,普鲁卡因,等等的盐。

[0129] 给药治疗有效量的药剂。药剂的治疗有效量部分取决于疾病的性质和/或症状的程度,可以通过标准临床技术确定,此外,可以任选用体外或者体内检测帮助确定最适剂量范围。制剂中使用的精确剂量还取决于给药途径和症状的严重性,这应该根据医师的判断和各个患者的情况来决定。可以从来自体外或者动物模型检测系统的剂量反应曲线外推出有效剂量。

[0130] 本发明还提供药物包装或者试剂盒,包括一个或者多个装满本发明药物组合物中一种或者多种成分的容器。可选的这类容器中还可以包括管理药物或者生物制品生产、使用或者销售的政府部门出具的通告,该通告表明政府批准用于人类给药的制备、使用或者销售。包装或者试剂盒可以标注关于给药方式,药物给药顺序(例如,分别、连续或者同时)或者类似的相关信息。包装或者试剂盒还可以包括一些方法提醒患者采取治疗。包装或者试剂盒可以是组合治疗的单一单位剂量,或者它可以是多个单位剂量。具体的,药剂可以是单独的,以任意组合混合的,存在于单个小瓶或者片剂中的。药剂组装成泡状包装或者其他配药装置中是优选的。对本发明的目的来说,单位剂量是为了表示一种剂量,所述剂量取决于各种药剂的个体药效学,在标准时程按 FDA 批准的剂量给药。

[0131] 筛选试验和由此鉴定的药剂

[0132] 本发明还提供鉴定药剂(例如,融合蛋白,多肽,模拟肽,前体药物,受体,结合剂,抗体,小分子或者其他药物,或者核酶)的方法,所述药剂改变(例如,增加或者降低)TCF7L2 的活性,或者与 TCF7L2 或者 Wnt 信号途径或者钙粘蛋白途径的其他成员(例如, beta- 连环蛋白)相互作用。例如,在某些实施方式中,这类药剂可以与 TCF7L2 结合;对例如 TCF7L2 的活性具有刺激或者抑制效果;或者改变(例如,增强或者抑制)TCF7L2 与 Wnt 信号途径其他成员或者与钙粘蛋白途径成员相互作用的能力,或者改变 TCF7L2 的翻译后加工。在其他实施方式中,这类药剂改变 Wnt 信号途径或者钙粘蛋白途径的活性或者功能。

[0133] 在一个实施方式中,本发明提供筛选与 TCF7L2 蛋白(或者其生物活性部分)结合或者改变其活性的候选或者待测药剂的方法,以及通过该方法鉴定的药剂。待测药剂可以利用本领域已知的组合文库方法中的任意一种方法来获得,包括:生物文库;空间寻址平行固相或者液相文库;需要去卷积(deconvolution)的合成文库方法;‘一-珠-化合物’(‘one-bead one-compound’)文库方法;和利用亲和层析选择法的合成文库。生物

文库方法限于多肽文库,而其他 4 种方法适于多肽、非肽寡聚物或者化合物的小分子文库 (Lam, K. S., *Anticancer Drug Des.* 12 :145 (1997))。

[0134] 在一个实施方式中,为鉴定改变 TCF7L2 活性的药剂,包含或者表达 TCF7L2 或者其片段或者衍生物的细胞、细胞裂解产物或者溶液,可以与待测药剂接触;或者,蛋白可以直接与待测药剂接触。TCF7L2 活性的水平(数量)被测定(例如,直接或者间接地检测 TCF7L2 活性的水平(数量)),并与对照(即,在待测药剂不存在的条件下 TCF7L2 蛋白或者其活性片段或者衍生物活性的水平)活性的水平比较。如果药剂存在时活性水平不同于药剂不存在时的活性水平,并且差值是统计上有显著意义的,则该药剂是改变 TCF7L2 活性的药剂。活性水平相对于对照的增加,表示该药剂是增强活性的药剂(是激动剂)。类似地,活性水平相对于对照的降低,表示该药剂是抑制活性的药剂(是拮抗剂)。在另一个实施方式中,待测药剂存在条件下 TCF7L2 或者其衍生物或者片段活性的水平与先前建立的对照水平进行比较。如果药剂存在条件下活性水平与对照水平的差值统计上有显著意义,则表示药剂改变 TCF7L2 的活性。

[0135] 本发明还涉及鉴定药剂(例如,反义核酸,融合蛋白,多肽,模拟肽,前体药物,受体,结合剂,抗体,小分子或者其他药物,或者核酶)的检测法,所述药剂改变(例如,增加或者降低)TCF7L2 基因的表达(例如,转录或者翻译),或者与 TCF7L2 的相互作用,以及涉及通过该检测法鉴定的药剂。例如,包含编码 TCF7L2 的核酸的溶液可以与待测药剂接触。溶液可以包括,例如,包含核酸的细胞或者包含核酸的细胞裂解产物;此外,溶液可以是包括核酸转录/翻译必需成分的其他溶液。如果需要,还可以使用未悬浮在溶液中的细胞。TCF7L2 表达的水平 and / 或模式(例如, mRNA 或者蛋白表达的水平 and / 或模式,例如不同剪接变异体的水平和 / 或模式)被测定,与对照表达的水平 and / 或模式(即,待测药剂不存在时 TCF7L2 表达的水平 and / 或模式)进行比较。如果药剂存在时的水平和 / 或模式与药剂不存在时的水平和 / 或模式有差异,并且差值在统计上有显著意义或者以统计上有显著意义的方式不同,则该药剂是改变 II 型糖尿病基因表达的药剂。TCF7L2 表达的增强表示该药剂是 TCF7L2 活性的激动剂。类似地,TCF7L2 表达的抑制表示该药剂是 TCF7L2 活性的拮抗剂。在另一个实施方式中,待测药剂存在条件下 TCF7L2 多肽(例如,不同的剪接变异体)的水平和 / 或模式与先前建立的对照水平和 / 或模式进行比较。如果药剂存在条件下水平和 / 或模式与对照水平和 / 或模式有差异,其差值在统计上有显著意义或者以统计上有显著意义的方式不同,则表示药剂改变 TCF7L2 表达。

[0136] 在本发明的另一个实施方式中,利用包含编码 TCF7L2 基因启动子区域的核酸或者与报告基因可操作连接的核酸的细胞、细胞裂解产物或者溶液,可以鉴定出改变 TCF7L2 表达或者与 TCF7L2 或者与 Wnt 信号途径或者钙粘蛋白途径其他成员相互作用的药剂。与待测药剂接触后,测定报告基因的表达水平(例如, mRNA 或者蛋白表达的水平),与对照的表达水平(即,待测药剂不存在条件下报告基因的表达水平)进行比较。如果药剂存在时水平与药剂不存在时的水平有差异,其差值是统计上有显著意义的或者以统计上有显著意义的方式不同,则该药剂是改变 TCF7L2 表达的药剂,因为显示它具有改变基因表达的能力,所述基因与 TCF7L2 基因启动子可操作的连接。报告基因表达的增强表示该药剂是 TCF7L2 活性的激动剂。类似地,报告基因表达的抑制表示该药剂是 TCF7L2 活性的拮抗剂。在另一个实施方式中,待测药剂存在条件下报告基因表达的水平与先前建立的对照水平进行比

较。如果药剂存在条件下水平与对照水平有差异,其差值在统计上有显著意义或者以统计上有显著意义的方式不同,表示该药剂改变表达。

[0137] 利用上述这些方法可以轻易的鉴定改变 TCF7L2 编码的不同剪接变异体量(例如,增强第一种剪接变异体活性的药剂,和抑制第二种剪接变异体活性的药剂)的药剂,以及是第一种剪接变异体活性激动剂和第二种剪接变异体活性拮抗剂的药剂。

[0138] 在本发明的其他实施方式中,该检测可用于评定待测药剂对 TCF7L2 结合剂相关多肽活性的影响。例如,表达与 TCF7L2 多肽相互作用的化合物(本文中称为“TCF7L2 结合剂”,它可以是与 TCF7L2 多肽直接或者间接作用的多肽或者其他分子,例如 Wnt 信号途径的成员或者钙粘蛋白途径的成员)的细胞,在待测药剂存在条件下与 TCF7L2 接触,检测待测药剂改变 TCF7L2 和 TCF7L2 结合剂相互作用的能力。此外,可以使用包含 TCF7L2 结合剂的细胞裂解产物或者溶液。通过干扰或者增强 TCF7L2 与 TCF7L2 结合剂结合、交联或者另外相互作用的能力,与 TCF7L2 或者 TCF7L2 结合剂结合的药剂可以改变相互作用。检测待测药剂结合 TCF7L2 或者 TCF7L2 结合剂的能力可以通过下述方法实现,例如,将待测药剂与放射性同位素或者酶标记物偶联,这样的话待测药剂与多肽的结合可以通过直接或者间接检测 ^{125}I , ^{35}S , ^{14}C 或者 ^3H 标记,和放射性直接计数或者闪烁计数的放射性同位素检测。此外,待测药剂可以酶联标记有,例如,辣根过氧化物酶,碱性磷酸酶,或者荧光素酶,和通过测定适合底物向产物的转化检测的酶标记物。在本发明范围内的还包括检测待测药剂与多肽相互作用的能力,所述相互作用物都没有进行标记。例如,微生理计可用于检测待测药剂与 TCF7L2 或者 TCF7L2 结合剂的相互作用,所述待测药剂,TCF7L2 或者 TCF7L2 结合剂都没有进行标记。McConnell, H. M. 等人, Science 257:1906-1912(1992)。本文使用的“微生理计”(例如, Cytosensor™) 是一种分析仪器,利用光寻址电位传感器(LAPS) 测量细胞酸化其环境的速率。酸化速率的变化可用作配体和多肽之间相互作用的指示剂。

[0139] 因此,这些受体可用于筛选作为激动剂或者拮抗剂的化合物,以用于 II 型糖尿病易感性的治疗或者研究。可以设计药物调控 TCF7L2 活化,从而可用于调控信号途径和下游基因的转录事件。

[0140] 在本发明的另一个实施方式中,检测法可用于鉴定与 TCF7L2 相互作用的多肽。例如,例如 Fields 和 Song (Fields, S. 和 Song, O., Nature 340:245-246(1989)) 描述的酵母双杂系统可用于鉴定与 TCF7L2 相互作用的多肽。在这类酵母双杂系统中,载体的构建是基于转录因子的灵活性,所述转录因子具有两个功能结构域(DNA 结合结构域和转录活化结构域)。如果两个结构域是分离的但融合到两个不同但彼此相互作用的蛋白,可以实现转录激活,特定标记物(例如,营养标记物的转录例如 His 和 Ade,或者显色标记物例如 lacZ)的转录可用于鉴定相互作用和翻译激活的存在。例如,在本发明的方法中,使用的第一种载体包括编码 DNA 结合结构域以及 TCF7L2,剪接变异体,或者其片段或者衍生物的核酸,使用的第二种载体包括编码转录激活结构域的核酸以及编码一种多肽的核酸,所述多肽可能与 TCF7L2 或者剪接变异体或者其片段或者衍生物相互作用。在合适条件(例如,例如 Clontech (Palo Alto, California, USA) 的 Matchmaker™ 系统中使用的交配条件)下孵育包含第一种和第二种载体的酵母,就可以鉴定表达感兴趣的标记物的菌落。检查这些菌落可以鉴定与 TCF7L2 或者其片段或者衍生物相互作用的多肽。这类多肽可以用作改变 TCF7L2 表达活性的药剂,如与治疗方法相关方面所述。

[0141] 在本发明上述检测法的不止一个实施方式中,可能需要将 TCF7L2 基因,TCF7L2 蛋白,TCF7L2 结合剂(例如,Wnt 信号途径或者钙粘蛋白途径的其他成员),或者检测法中的其他组分固定到固相支持物上,以便促进一种或者两种蛋白复合与非复合形式的分离,以及适应检测自动化。在待测药剂存在或者不存在的条件下,待测药剂与蛋白的结合,或者蛋白与结合剂的相互作用,可以在任何适于包含反应物的容器中进行。这类容器的实例包括微量滴定板,试管和微量离心管。在一个实施方式中,可以提供融合蛋白(例如,谷胱甘肽-S-转移酶融合蛋白),其添加一个 TCF7L2,TCF7L2 蛋白或者 TCF7L2 结合剂结合到基质或者其他固相支持物上的结构域。

[0142] 在另一个实施方式中,采用一种方法鉴定本发明核酸分子的表达调节物,其中包含 TCF7L2 的细胞,细胞裂解产物或者溶液与待测药剂接触,然后检测细胞,细胞裂解产物或者溶液中相应 mRNA 或者多肽(例如剪接变体)的表达。待测药剂存在条件下相应 mRNA 或者多肽的表达水平与待测药剂不存在条件下 mRNA 或者多肽的表达水平进行比较。基于这种比较,待测药剂可以被鉴定为表达调节物。例如,待测药剂存在条件下比其缺失条件下 mRNA 或者多肽的表达更高(统计上显著的更高),则待测药剂被鉴定为 mRNA 或者多肽表达的激发剂或者增强剂。此外,待测药剂存在条件下比其缺失条件下 mRNA 或者多肽的表达更低(统计上显著的更低),则待测药剂被鉴定为 mRNA 或者多肽表达的抑制剂。可以采用本文描述的检测 mRNA 或者多肽的方法检测细胞中 mRNA 或者多肽的表达水平。

[0143] 本发明还涉及利用上述筛选检测方法鉴定的新的药剂。因此,在本文所述治疗方法中进一步使用本文所述鉴定的药剂也在本发明的范围内。例如,通过将蛋白或者核酸(或者包括多肽或者核酸的细胞)与本文所述鉴定的药剂接触,本文所述鉴定的药剂可用于改变 TCF7L2 基因编码蛋白的活性,或者改变 TCF7L2 的表达。

[0144] 本发明的核酸

[0145] TCF7L2 核酸,部分和变体

[0146] 本发明还涉及包括人 TCF7L2 的分离的核酸分子。本发明的 TCF7L2 核酸分子可以是 RNA,例如, mRNA,或者 DNA,例如 cDNA 和基因组 DNA。DNA 分子可以是双链或者单链;单链 RNA 或者 DNA 可以是编码或者正义链,或者非编码或者反义链。核酸分子可以包括全部或者部分基因编码序列,还可以包括其他非编码序列,例如内含子和非编码 3' 和 5' 序列(例如,包括调控序列)。

[0147] 此外,本发明的核酸分子可以融合到标记物序列,例如,编码一种多肽的序列,其帮助多肽的分离或者纯化。这类序列包括,但不限于,编码谷胱甘肽-S-转移酶(GST)融合蛋白和编码流感血凝素 A(HA)多肽标记物的序列。

[0148] 本文使用的“分离的”核酸分子,是指那些与,通常位于基因或者核苷酸序列(如在基因组序列内)侧翼的核酸分离和/或已经完全或者部分从其他转录序列纯化(例如,在 RNA 文库内)。例如,本发明的分离核酸,相对于其天然产生的复杂细胞环境,或者当通过重组技术产生时的培养基,或者当化学合成时化学前体或者其他化学制剂,是基本上分离的。在有些情况下,分离的材料将成为组合物(例如,包含其他物质的粗提物),缓冲系统或者试剂混合物的一部分。在其他情况下,材料可以纯化到例如通过 PAGE 或者柱层析法例如 HPLC 检测基本均一的。优选的,分离的核酸分子包括至少大约 50,80 或者 90%(摩尔比)的所有存在的高分子种类。就基因组 DNA 而言,术语“分离的”还指与染色体分离的核酸分

子,所述染色体与基因组 DNA 是天然关联的。例如,分离核酸分子可以包含小于大约 5kb 但不限于 4kb,3kb,2kb,1kb,0.5kb 或者 0.1kb 的核苷酸,其在细胞基因组 DNA 内核酸分子的侧翼,所述细胞是核酸分子的来源。

[0149] 核酸分子可以融合到其他编码或者调控序列,仍然被看作是分离的。因此,包括在载体内的重组 DNA 属于本文使用的“分离的”定义。此外,分离核酸分子包括异源宿主细胞内的重组 DNA 分子,以及溶液中部分或者基本上纯化的 DNA 分子。“分离的”核酸分子还包括本发明 DNA 分子的体内和体外 RNA 转录产物。分离的核酸分子可以包括化学或者重组方法合成的核酸分子或者核酸序列。因此,包括在载体内的重组 DNA 属于本文使用的“分离的”定义。此外,分离的核酸分子包括异源生物内的重组 DNA 分子,以及溶液中部分或者基本上纯化的 DNA 分子。本发明 DNA 分子的体内和体外 RNA 转录产物也包括在“分离的”核酸序列中。这类分离核酸分子可用于编码多肽的制备,作为分离同源序列(例如,来自其他哺乳动物种类)的探针,用于基因作图(例如,通过与染色体的原位杂交),或者用于检测组织中基因的表达(例如,人组织),例如通过 Northern 或者 Southern 印迹分析。

[0150] 本发明还涉及核酸分子,所述核酸分子不必在自然界被发现,但其编码 TCF7L2 多肽,或者 TCF7L2 多肽的其他剪接变异体或者其多态性变异体。因此,例如,本发明涉及包括一种不同于天然存在的核苷酸序列的 DNA 分子,所述序列由于遗传密码的简并性,编码本发明的 TCF7L2 多肽。本发明还包括编码部分(片段)或者编码变异体多肽的核酸分子,所述变异体多肽例如 TCF7L2 多肽的类似物或者衍生物。这类变异体可以是天然存在的,例如等位基因变异或者单核苷酸多态性,或者非天然存在的,例如通过不同诱变剂和诱变处理进行诱导。预计的变异包括,但不限于,一个或多个核苷酸的添加,缺失和取代,这引起保守或者非保守氨基酸的改变,包括添加和缺失。优选的核苷酸(和/或生成的氨基酸)改变是沉默或者保守的;即,它们不改变 TCF7L2 多肽的特征或者活性。一方面,核酸序列是包括一种或者多种多态性微卫星标记物的片段。另一方面,核苷酸序列是包括一种或者多种 TCF7L2 基因单核苷酸多态性的片段。

[0151] 本发明核酸分子的其他改变可以包括,例如,标记,甲基化,核苷酸间修饰例如无电荷键(例如,磷酸甲酯,磷酸三酯,磷酸酰胺,氨基甲酸酯),带电荷键(例如,硫代磷酸酯,二硫代磷酸酯),侧基(例如,多肽),嵌入剂(例如,吡啶,补骨脂素),螯合剂,烷化剂(alkylators),和修饰键(例如,alpha 异头核酸)。合成分子也被包括在内,所述合成分子模拟核酸分子通过氢键作用和其他化学相互作用与指定序列结合的能力。这类分子包括,例如,那些分子骨架上肽键代替磷酸酯键的分子。

[0152] 本发明还涉及在高严谨性杂交条件下与本文所述核苷酸序列杂交的核酸分子(例如,与编码多肽的核苷酸序列特定杂交的核酸分子,该多肽是本文描述的或者具有上述多肽的活性),例如选择性杂交。一方面,本发明包括本文所述的变异体,该变异体在高严谨性杂交条件(例如,选择性杂交)下与编码氨基酸序列或者其多态性变异体的核苷酸序列杂交。另一方面,在高严谨性杂交条件下杂交的变异体具有 TCF7L2 多肽的活性。

[0153] 这类核酸分子可以通过特异杂交(例如,在高严谨性条件下)进行检测和/或分离。本文使用的“特异杂交”是指第一种核酸与第二种核酸以某种方式杂交的能力,所述方式使第一种核酸不与第二种核酸以外的所有核酸杂交(例如,当样品中第一种核酸与第二种核酸的相似性比任意其他核酸都要高时,才进行杂交)。杂交的“严谨性条件”是本领

域的术语,是指孵育和洗涤条件,例如温度和缓冲液浓度条件,允许特定核酸与第二种核酸的杂交;第一种核酸可以与第二种核酸完全(即,100%)互补,或者第一种和第二种核酸的互补程度比完全互补低(例如,70%,75%,85%,90%,95%)。例如,某些高严谨性条件可以区分完全互补的核酸和互补性稍低的核酸。“高严谨性条件”,“中等严谨性条件”和“低严谨性条件”,以及核酸杂交方法阐述于 Current Protocols in Molecular Biology 的 2.10.1-2.10.16 页和 6.3.1-6.3.6 页(Ausubel, F. 等人, " Current Protocols in MolecularBiology", John Wiley & Sons, (1998)),和 Kraus, M. 和 Aaronson, S., Methods Enzymol, 200 :546-556 (1991)。

[0154] 通过比对序列用于最优的比较目的(例如,缺口可以导入至第一个序列的序列用于最优比对),可以确定两种核苷酸或者氨基酸序列同源性或者同一性的百分比。然后比较对应位置的核苷酸或者氨基酸,两个序列之间的百分相同比是序列共有的相同位置的数目的函数(即, %同一性=相同位置的数目/位置的总数×100)。当一个序列中的一个位置被另一个序列对应位置上相同的核苷酸或者氨基酸残基占据时,则上述分子在该位置是同源的。本文使用的核酸或者氨基酸“同源性”等同于核酸或者氨基酸“同一性”。某些情况下,用于比较目的而比对的序列的长度至少是参照序列长度的 30%,例如,至少 40%,某些情况下至少 60%,和其他情况下至少 70%,80%,90%或者 95%。两个序列的有效比较可以利用公知方法进行,例如,利用一种数学算法。这类数学算法一个优选的、非限制性实例描述于 Karlin 等人, Proc. Natl. Acad. Sci. USA 90 :5873-5877 (1993)。这种算法被整合到 NBLAST 和 XBLAST 程序 (2.0 版本),描述于 Altschul 等人, Nucleic Acids Res. 25 :389-3402 (1997)。当使用 BLAST 和 Gapped BLAST 程序时,可以使用相应程序(例如, NBLAST)的缺省参数。一种情况下,序列比较的参数设为值=100,字长(W)=12,或者可以改变(例如, W=5 或 W=20)。

[0155] 用于序列比较的数学算法的另一个优选、非限制性实例是 Myers 和 Miller 的算法, CABIOS 4(1) :11-17 (1988)。这种算法被整合到 ALIGN 程序 (2.0 版本),属于 GCG 序列比对程序包的一部分 (Accelrys, Cambridge, UK)。当利用 ALIGN 程序比较氨基酸序列时,可以使用 PAM 120 权重残差表 (weight residue table), 间隙长度罚分 (gap length penalty) 12, 和缺口罚分 (gap penalty) 4。其他用于序列分析的算法是本领域公知的,包括 Torellis 和 Robotti 描述的 ADVANCE 和 ADAM, Comput. Appl. Biosci. 10 :3-5 (1994); 以及 Pearson 和 Lipman 描述的 FASTA, Proc. Natl. Acad. Sci. USA 85 :2444-8 (1988)。

[0156] 另一方面,两个氨基酸序列间的同一性百分比可以利用 GCG 程序包中的 GAP 程序计算,使用 BLOSUM63 矩阵或者 PAM250 矩阵,缺口权重 (gap weight) 为 12、10、8、6 或者 4, 长度权重 (length weight) 为 2、3 或者 4。另一方面,两个核酸序列间的同一性百分比可以利用 GCG 程序包中的 GAP 程序计算,使用缺口权重 (gap weight) 50 和长度权重 (length weight) 3。

[0157] 本发明还提供分离的核酸分子,所述核酸分子包含在高严谨性条件下与 TCF7L2 核苷酸序列或者这种序列的互补序列杂交的片段或者部分,还提供分离的核酸分子,所述核酸分子包含在高严谨性条件下与编码氨基酸序列或者其多态性变异体的核苷酸序列杂交的片段或者部分。本发明的核酸片段长度是至少大约 15, 优选的至少大约 18, 20, 23 或者

25 个核苷酸,并且长度可以是 30,40,50,100,200 或者更多个核苷酸。越长的片段,例如,长度是 30 或者更多个核苷酸,其编码本文所述抗原多肽,对例如下文所述的抗体生产是非常有用的。

[0158] 探针和引物

[0159] 在相关方面,本发明的核酸片段用作例如本文所述检测法中的探针或者引物。“探针”或者“引物”是以碱基特异方式与核酸分子互补链杂交的寡聚核苷酸。这类探针和引物包括多肽核酸,如 Nielsen 等人描述, *Science* 254 :1497-1500 (1991)。

[0160] 探针或者引物包括与核酸分子的至少大约 15 个,例如大约 20-25 个,和在一些情况下大约 40、50 或者 75 个连续核苷酸杂交的核苷酸序列区域,所述核酸分子包括 TCF7L2 或者其多态性变异体的连续核苷酸序列。在其他情况下,探针或者引物包括 100 个或者更少的核苷酸,在一些情况下从 6 到 50 个核苷酸,例如从 12 到 30 个核苷酸。在其他情况下,探针或者引物与连续核苷酸序列或者连续核苷酸序列的互补序列至少有 70% 的相似性,例如至少 80% 同一性,在一些情况下至少 90% 同一性,和在其他情况下至少 95% 同一性,或者甚至能够与连续核苷酸序列或者连续核苷酸序列的互补序列选择性杂交。通常,探针或者引物还包括标记物,例如,放射性同位素,荧光化合物,酶,或者酶辅因子。

[0161] 利用标准分子生物学技术和本文提供的序列信息,可以鉴定和分离如上所述的本发明核酸分子。例如,根据 TCF7L2 或者该序列的互补序列设计的,或者根据编码本文提供的一种或者多种氨基酸序列的核苷酸序列设计的合成寡聚核苷酸引物,利用聚合酶链式反应扩增和分离核酸分子。通产参见 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 等人, Academic Press, San Diego, CA, 1990) ;Mattila 等人, *Nucl. Acids Res.* 19 :4967 (1991) ;Eckert 等人, *PCR Methods and Applications* 1 :17 (1991) ;PCR (eds. McPherson 等人, IRL Press, Oxford) ;和美国专利 4,683,202。可以利用 cDNA, mRNA 或者基因组 DNA 作为模板扩增核酸分子,克隆到合适载体,通过 DNA 序列分析进行表征。

[0162] 其他合适的扩增方法包括连接酶链式反应 (LCR) (参见 Wu 和 Wallace, *Genomics* 4 :560 (1989), Landegren 等人, *Science* 241 :1077 (1988), 转录扩增 (Kwoh 等人, *Proc. Natl. Acad. Sci. USA* 86 :1173 (1989)), 和自主序列复制 (Guatelli 等人, *Proc. Natl. Acad. Sci. USA* 87 :1874 (1990)) 和依赖核酸的序列扩增 (NASBA)。后两个扩增方法涉及基于等温转录的等温反应,产生的扩增产物为单链 RNA (ssRNA) 和双链 DNA (dsDNA),比例分别为大约 30 或者 100 比 1。

[0163] 扩增的 DNA 可以被标记,例如,放射性标记,并用作筛选 cDNA 文库的探针,所述 cDNA 文库来自人细胞,在 zap 表达, ZIPLOX 或者其他合适载体中的 mRNA。可以分离相应克隆,体内切割后获得 DNA,利用本领域公知的方法可以对克隆的插入物进行单向或者双向测序,鉴定编码合适分子量多肽的正确读码框。例如,利用商品化供应的公知方法可以对本发明核酸分子的核苷酸序列进行直接分析。参见,例如, Sambrook 等人, *Molecular Cloning, A Laboratory Manual* (2nd Ed., CSHP, New York 1989) ;Zyskind 等人., *Recombinant DNA Laboratory Manual*, (Acad. Press, 1988))。此外,荧光方法也可用于分析核酸 (Chen 等人, *Genome Res.* 9, 492 (1999)) 和多肽。利用这些或者相似方法,多肽和编码多肽的 DNA 可以被

分离、测序和表征。

[0164] 本发明的反义核酸分子可以根据 TCF7L2 核苷酸序列和 / 或互补序列或者部分核苷酸序列设计,按照本领域公知的步骤,利用化学合成和酶连接反应构建。例如,反义核酸分子(例如,反义寡聚核苷酸)可以利用天然存在的核苷酸或者不同修饰的核苷酸化学合成,所述修饰是为了提高分子的生物稳定性或者提高反义和正义核酸之间形成的双链的物理稳定性,例如,可以使用硫代磷酸酯衍生物和吡啶取代的核苷酸。此外,反义核酸分子可以利用表达载体生物合成,在所述载体中沿反义方向已经亚克隆了核酸分子(即,从插入的核酸分子转录的 RNA 对于感兴趣的靶核酸是反义方向)。

[0165] 核酸序列还可以用于与患者的内源 DNA 序列进行比较,以鉴定如上所述的一种或者多种病症,并且可以用作探针,这样就可以从样品中杂交和发现相关 DNA 序列或者排除已知序列。核酸序列还可以用于产生遗传指纹法的引物,利用 DNA 免疫技术生产抗-多肽抗体,以及作为抗原产生抗-DNA 抗体或者激发免疫应答。本文鉴定的核苷酸序列的部分或者片段(和对应的完整基因序列)可在许多方面应用,例如多核苷酸试剂。例如,这些序列可用于:(i) 在染色体上对它们各自的基因作图;和,因此,定位与遗传疾病相关的基因区域;(ii) 从少量生物样品中鉴定个体(组织分型);和(iii) 帮助生物样品的法医鉴定。此外,本发明的核苷酸序列可用于鉴定和表达重组多肽,所述重组多肽用于分析、表征或者治疗用途,或者作为组织标记物,其中对应多肽组成性表达,在组织分化期间表达或者在患病状态下表达。该核酸序列还可用作本文所述筛选和 / 或诊断方法中的试剂,还可以作为本文所述筛选和 / 或诊断方法所用试剂盒(例如,试剂盒)的组分。

[0166] 用于诊断方法的试剂盒(例如, reagent kits)包括在本文所述任意方法中适用的组分,包括例如,本文描述的杂交探针或者引物(例如,标记的探针或者引物),检测标记分子的试剂,限制性酶(例如,用于 RFLP 分析),等位基因特异的寡聚核苷酸,与突变或者非突变(天然)TCF7L2 多肽结合的抗体,包括 TCF7L2 核酸或者 TCF7L2 的一部分的核酸扩增工具,或者如本文描述的分析 TCF7L2 核酸或者分析 TCF7L2 多肽氨基酸序列的核酸序列的工具,等等。一方面,诊断 II 型糖尿病易感性的试剂盒可以包括用于扩增 TCF7L2 核酸内区域的引物,所述区域包含在患有 II 型糖尿病或者 II 型糖尿病易感的个体频率更高的标记物 DG10S478, SNP rs12255372, rs895340, rs11196205, rs7901695, rs7903146, rs12243326 和 / 或 rs4506565,或者风险单倍型。可以根据 SNPs 侧翼的核酸部分设计引物,所述 SNPs 指示 II 型糖尿病。

[0167] 载体和宿主细胞

[0168] 本发明的另一方面涉及包含本文所述核酸分子和其互补序列(或者其一部分)的核酸构建体。该构建体包括载体(例如,表达载体),其中本发明的序列已经按正义或者反义方向插入载体中。本文使用的术语“载体”是指能够转运与其连接的另一种核酸分子的核酸分子。一类载体是“质粒”,指可以连接其他 DNA 片段的环状双链 DNA 环。另一类载体是病毒载体,其中其他 DNA 片段可以连接到病毒基因组中。某些载体能够在它们被导入的宿主细胞中自主复制(例如,具有细菌复制起点的细菌载体和附加型哺乳动物载体)。其他载体(例如,非附加型哺乳动物载体)在导入宿主细胞时整合到宿主细胞的基因组中,从而随宿主基因组进行复制。表达载体能够指导与它们可操作连接的基因的表达。重组 DNA 技术中表达载体的应用通常是采取质粒形式。但是,本发明预期包括这类其他形式的表达载

体,例如具有等同功能的病毒载体(例如,复制缺陷型逆转录病毒,腺病毒和腺相关病毒)。

[0169] 在一些情况下,本发明的重组表达载体包括本发明的核酸分子,所述核酸分子的形式适于宿主细胞中的核酸分子表达。这表明重组表达载体包括一种或者多种调控序列,根据待用于表达的宿主细胞进行选择,所述载体与待表达的核酸序列可操作的连接。在重组表达载体内,“可操作的连接”或者“操作的连接”是指感兴趣的核苷酸序列以某种方式连接到调控序列,该方式允许该核苷酸序列的表达(例如,在体外转录/翻译系统或者当载体导入宿主细胞时在宿主细胞中)。术语“调控序列”包括启动子,增强子和其他表达控制元件,(例如,多腺苷酸化信号)。这类调控序列描述于,例如,Goeddel, "GeneExpression Technology", Methods in Enzymology 185, Academic Press, San Diego, CA(1990)。调控序列包括那些在许多类型的宿主细胞中指导核苷酸序列组成性表达和只在某些宿主细胞中指导核苷酸序列表达(例如,组织-特异调控序列)的序列。本领域技术人员应了解表达载体的设计取决于这类因素,如待转化宿主细胞的选择和所需的多肽表达水平。本发明的表达载体可以导入宿主细胞从而产生本文所述核酸分子编码的多肽,包括融合多肽。

[0170] 本发明的重组表达载体可以设计成在原核或者真核细胞中表达本发明的多肽,例如,细菌细胞例如 E. coli,昆虫细胞(利用杆状病毒表达载体),酵母细胞或者哺乳动物细胞。合适的宿主细胞还描述于 Goeddel, supra。此外,重组表达载体可以在体外转录和翻译,例如利用 T7 启动子调控序列和 T7 聚合酶。

[0171] 本发明的另一方面涉及被导入本发明重组表达载体的宿主细胞。术语“宿主细胞”和“重组宿主细胞”是在本文中是可以互换使用的。应了解这类术语不仅指特定个体细胞而且还指这类细胞的后代或者潜在后代。由于突变或者环境影响后继世代可能存在某些修饰,这类后代实际上可能不同于母细胞,但仍然包括在本文使用的该术语范围内。

[0172] 宿主细胞可以是任意的原核或者真核细胞。例如,本发明的核酸分子可在细菌细胞(例如,E. coli),昆虫细胞,酵母或者哺乳动物细胞(例如中国仓鼠卵巢细胞(CHO)或者 COS 细胞)中表达。其他合适的宿主细胞是本领域技术人员公知的。

[0173] 载体 DNA 可以通过常规转化或者转染技术导入原核或者真核细胞。本文使用的术语“转化”和“转染”是指将外源核酸分子(例如,DNA)导入宿主细胞的本领域已知的各种技术,包括磷酸钙或者氯化钙共沉淀,DEAE-葡聚糖介导的转染,脂质体转染或者电穿孔。转化或者转染宿主细胞的合适方法可参见 Sambrook 等人(supra)和其他实验手册。

[0174] 对于哺乳动物细胞的稳定转染,众所周知,依赖于使用的表达载体和转染技术,只有小部分细胞可以将外源 DNA 整合到它们的基因组中。为了鉴定和选择这些整合体,编码选择性标记物(例如,抗生素抗性)的基因通常和感兴趣的基因一起导入宿主细胞。优选的选择标记物包括那些产生抗药性的标记物,例如 G418,潮霉素和氨甲蝶呤。编码选择性标记物的核酸分子可以与本发明的核酸分子在相同载体上导入宿主细胞,或者在单独载体上导入。稳定转染有导入的核酸分子的细胞可通过药物选择鉴定(例如,整合选择性标志物基因的细胞存活,而其他细胞死亡)。

[0175] 本发明的宿主细胞,例如培养的原核或者真核宿主细胞可用于产生(即,表达)本发明的多肽。因此,本发明还提供利用本发明的宿主细胞产生多肽的方法。一方面,该方法包括在合适培养基中培养本发明的宿主细胞(其中已经导入编码本发明多肽的重组表达载体)以产生多肽。另一方面,该方法还包括从培养基或者宿主细胞分离多肽。

[0176] 本发明的抗体

[0177] 本发明还提供与一种形式的基因产物特异结合而不与另一种形式的基因产物结合的多克隆抗体和/或单克隆抗体。本发明还提供与变异体或者参照基因产物的一部分结合的抗体,所述变异体或者参照基因产物包含多态性位点。本文使用的术语“抗体”是指免疫球蛋白分子和免疫球蛋白分子的免疫活性部分,即,包含特异结合抗原的抗原结合位点的分子。特异结合本发明多肽的分子是与该多肽或者其片段结合,但基本上不与样品中其他分子结合的分子,所述样品,例如,天然包含多肽的生物样品。免疫球蛋白分子免疫活性部分的实例包括 F(ab) 和 F(ab')₂ 片段,这可以通过用酶例如胃蛋白酶处理抗体产生。本发明提供结合本发明多肽的多克隆和单克隆抗体。本文使用的术语“单克隆抗体”或者“单克隆抗体组合物”,是指只包含一种抗原结合位点的抗体分子集合,所述抗原结合位点能够与本发明多肽特定抗原表位发生免疫反应。因此单克隆抗体组合物通常对与其发生免疫反应的本发明特定多肽显示单一结合亲和力。

[0178] 多克隆抗体可以如上所述制备,利用所需免疫原,例如,本发明的多肽或者其片段,免疫合适个体。通过标准技术可以随时间监控免疫个体中的抗体滴度,例如利用固定多肽的酶联免疫吸附(ELISA)。如果需要,可以从哺乳动物(例如,从血液)分离针对多肽的抗体分子,采用公知技术进一步纯化,例如蛋白 A 层析法以获得 IgG 级分。在免疫后的合适时间,例如,当抗体滴度最高时,抗体产生细胞可以从个体获得,并用于通过标准技术制备单克隆抗体,例如杂交瘤技术,最初描述于 Kohler 和 Milstein, *Nature* 256:495-497 (1975), 人 B 细胞杂交瘤技术 (Kozbor 等人, *Immunol. Today* 4:72 (1983)), EBV- 杂交瘤技术 (Cole 等人, *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, 1985, Inc., pp. 77-96) 或者三源杂交瘤技术。产生杂交瘤的技术是已知的(通常参见 *Current Protocols in Immunology* (1994) Coligan 等人., (eds.) John Wiley & Sons, Inc., New York, NY)。概括的说,永生化细胞系(通常是骨髓瘤)与淋巴细胞(通常是脾细胞)融合,所述淋巴细胞来自上述免疫原免疫的哺乳动物,筛选所得杂交瘤细胞的培养上清液,鉴定产生与本发明多肽结合的单克隆抗体的杂交瘤。

[0179] 大量公知的用于融合淋巴细胞和永生化细胞系的实验规程均可用于产生本发明多肽的单克隆抗体的目的(参见,例如, *Current Protocols in Immunology*, supra; Galfre 等人, *Nature* 266:55052 (1977); R. H. Kenneth, in *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); 和 Lerner, *Yale J. Biol. Med.* 54:387-402 (1981))。此外,普通技术人员应了解还存在这类方法的许多变形,它们也是有用的。

[0180] 制备分泌单克隆抗体的杂交瘤的另外一种方法,利用多肽筛选重组组合免疫球蛋白文库(例如,抗体噬菌体展示文库)从而分离结合多肽的免疫球蛋白库成员,可以鉴定本发明多肽的单克隆抗体。用于产生和筛选噬菌体展示文库的试剂盒是商品化供应的(例如, the Pharmacia Recombinant Phage Antibody System, Catalog No 27-9400-01; 和 the Stratagene SurfZAP™ Phage Display Kit, Catalog No 240612)。此外,特别适用于产生和筛选抗体展示文库的方法和试剂的实例可见于,例如,美国专利号 5,223,409; PCT 公开号 WO 92/18619; PCT 公开号 WO 91/17271; PCT 公开号 WO 92/20791; PCT 公开号 WO 92/15679; PCT 公开号 WO 93/01288; PCT 公开号 WO 92/01047; PCT 公开号 WO 92/09690; PCT 公开号

WO 90/02809 ;Fuchs 等人, Bio/Technology 9 :1370-1372(1991) ;Hay 等人, Hum. Antibod. Hybridomas 3 :81-85(1992) ;Huse 等人, Science 1A β :1275-1281(1989) ;和 Griffiths 等人, EMBO J. 12 :725-734(1993)。

[0181] 此外,利用标准重组 DNA 技术产生的重组抗体,例如嵌合和人源化单克隆抗体,包括人和非人部分,也包括在本发明的范围内。这类嵌合和人源化单克隆抗体可以通过本领域已知的重组 DNA 技术产生。

[0182] 通常,本发明的抗体(例如,单克隆抗体)可用于通过标准技术分离本发明的多肽,例如亲和层析法或者免疫沉淀法。多肽特定抗体可以帮助从细胞中纯化天然多肽,以及对宿主细胞中表达的重组多肽进行纯化。此外,对本发明多肽特异的抗体可用于检测多肽(例如,在细胞裂解产物,细胞上清液或者组织样品中),以便计算多肽的表达丰度和模式。抗体可以用于诊断用途,作为临床检测方法的一部分来监控组织中的蛋白水平,例如,确定给定治疗方法的效果。抗体可以偶联到可检测的物质以帮助其检测。可检测物质的实例包括各种酶,辅基,荧光材料,发光材料,生物发光材料,和放射性物质。合适酶的实例包括辣根过氧化物酶,碱性磷酸酶, beta- 半乳糖苷酶,或者乙酰胆碱酯酶;合适辅基复合物的实例包括链亲和素 / 生物素和亲和素 / 生物素;合适荧光材料的实例包括伞形酮,荧光素,异硫氰酸荧光素,若丹明,二氯三嗪胺荧光素,丹磺酰氯或者藻红蛋白;发光材料的实例包括鲁米诺;生物发光材料的实例包括荧光素酶,荧光素,和水母发光蛋白,和合适放射性物质的实例包括 ^{125}I , ^{131}I , ^{35}S 或者 ^3H 。

[0183] 现在通过下列例证解释本发明,其不应当在任一方面被认为是限制性的。

[0184] 实施例

[0185] 此处描述的是利用 10q 基因座内密集的微卫星标记物,通过单点关联分析,鉴定转录因子 7- 样 2(TCF7L2- 原先的 TCF4) 为导致患 II 型糖尿病风险的基因。

[0186] 方法

[0187] 冰岛群组

[0188] 冰岛数据保护机构(Data Protection Authority of Iceland)和冰岛国家生物伦理委员会(National Bioethics Committee of Iceland)批准了此项研究。所有研究参与者都签署了知情同意书。与血液样品、医学信息和家系相关的所有个人身份资料首先由数据保护机构利用第三方加密系统(18)(Data Protection Authority)加密。

[0189] 为了这项研究,2400 名 II 型糖尿病患者被鉴别,他们或者在冰岛心脏协会(Icelandic Heart Association)过去 30 年的长期流行病学研究中被诊断,或者在过去 12 年中在雷克雅未克(Reykjavik)的两家主要医院中被诊断。这些患者中的三分之二存活,代表现在已知的大约一半的冰岛 II 型糖尿病患者人群。为了这项研究,我们和这些患者中的大多数进行接触,合作比例超过 80%。这项研究的所有参与者都访问了冰岛心脏协会,在那里他们回答了一份问卷,进行了抽血和空腹血糖检测。问卷包括关于药物治疗和诊断时年龄的问题。这项研究中的 II 型糖尿病患者按照我们先前发表的连锁研究(10)所述的进行诊断。简单地说,II 型糖尿病的诊断由医生根据先前的医疗记录,药物治疗历史和 / 或新的实验室测量方法来确认。对于先前确诊的 II 型糖尿病患者,使用口服降糖药的报告证实 II 型糖尿病。当前接受胰岛素治疗的个体,如果他们也使用或者以前使用口服降糖药,则被划分为患有 II 型糖尿病。在这个群组里,大多数接受药物治疗的患者服用口服降糖药,

只有少部分(9%)需要胰岛素。对于至今未能诊断的个体,II型糖尿病和空腹血糖受损(IFG)的诊断基于美国糖尿病协会(American Diabetes Association)制定的标准(Expert Committee on the Diagnosis and Classification of Diabetes Mellitus 1997)。这项研究中II型糖尿病患者的平均年龄是69.7岁。

[0190] 重复群组

[0191] 丹麦研究组选自丹麦的PERF(Prospective Epidemiological Risk Factors, 预期流行病学危险因素)研究(19)。228位女性已经先前确诊患有II型糖尿病和/或检测出 $\geq 7\text{mM}$ 血糖。作为对照,539位未患病(相对II型糖尿病)女性从相同研究群组中随机抽出。

[0192] 美国的PENN CATH研究是1998年7月至2003年3月期间在宾夕法尼亚大学医疗中心(University of Pennsylvania Medical Center)接受心导管插入术的患者连续群组中进行的生化和遗传因素与冠状动脉粥样硬化关联性的抽样研究(cross sectional study)。II型糖尿病定义为具有以下病史,空腹血糖 $> 126\text{mg/dl}$,餐后2小时血糖 $> 200\text{mg/dl}$,40岁以上个体中使用口服降血糖药或者胰岛素。宾夕法尼亚大学机构审查委员会(University of Pennsylvania Institutional Review Board)批准了研究计划,所有参与者都签署了知情同意书。通过自述进行种族划分。361个白种人II型糖尿病病例来自这个群组。530位未患病(相对II型糖尿病和心肌梗死)的白种人对照从相同研究群组中随机抽出。

[0193] 用于基因分型的DNA是利用GenomiPhi Amplification试剂盒(Amersham),从丹麦和美国II型糖尿病患者和对照的外周血分离的DNA的全基因组扩增产物。

[0194] 基因分型

[0195] 在94个对照中利用串联重复发现软件(20)鉴定新的序列重复(即二核苷酸、三核苷酸和四核苷酸重复),检测多态性。CEPH样品1347-02(CEPH基因组库)低频等位基因碱基对的大小被从微卫星扩增子的大小中减去,用作参照。利用直接DNA测序(Applied BioSystems)或者Centaurus平台(Nanogen)进行SNP基因分型。

[0196] 关联分析的统计方法

[0197] 对于单一标记物与II型糖尿病的关联,我们使用似然比检验计算各等位基因的两侧p-值。对于使用的微卫星,我们提供等位基因频率而不是携带者频率。

[0198] 我们采用乘法模型(16,17)计算相对危险度(RR)和人群归因危险度(PAR)。对于CEPH白种人HapMap数据,我们利用 D' (21)和 R^2 (22)的标准定义计算SNPs配对之间的LD。当对全部SNP组合作图以说明特定区域中的LD结构,我们将 D' 绘制在左上角,p-值绘制在右下角。在我们展示的LD图中,标记物绘制成等距的,而不是根据它们的物理位置。

[0199] 结果

[0200] 基因座范围的关联研究

[0201] 我们过去报道了在冰岛人群中染色体5q与II型糖尿病基因组范围的显著连锁(10);在相同的研究中,我们还报道与10q和12q连锁的指示性证据。为了对10q基因座进行研究,我们采用关联方法,利用对应这个基因座跨越10.5Mb区域(NCBI Build 34: Chr10:114.2-124.7Mb)的高密度的基因分型的微卫星标记物。我们鉴定和分型了228个微卫星标记物--即一个标记物的平均密度为46kb(表1)。全部标记物都分型到1185位冰

岛 II 型糖尿病患者和 931 位无关的人群对照。

[0202] 表 1: 人类基因组集合 NCBI Build 34 中染色体 10 上 228 个基因分型的微卫星的位置。

[0203]

名称	起始: Build 34 Chr10 位置	结束: Build 34 Chr10 位置
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

[0204]

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

[0205]

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

[0206]

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

[0207]

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

[0208]

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

[0209]

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

[0210] 鉴定与 DG10S478 相关的微卫星标记物的单一标记物关联分析（表 2 和图 1）。

[0211] 表 2：在冰岛 DG10S478 与 II 型糖尿病的关联

[0212]

等位基因	患病组频率 (n=1185)	对照组频率 (n=931)	RR [95% CI]	两侧 P
0	0.636	0.724	0	2.1×10^{-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	4.6×10^{-7}
16	0.022	0.015	1.53	0.076
20	0.001	0.003	0.39	0.17
X	0.364	0.276	1.50[1.31,1.71]	2.1×10^{-9}

[0213] 6个等位基因中观察到这种四核苷酸重复,等位基因0,8和12占人群对照中染色体的98%。相对于其他组合的等位基因,等位基因0显示保护性关联(相对危险度(RR) = 0.67; $P = 2.1 \times 10^{-9}$)。这个P-值是两侧的,考虑了一些患者彼此具有亲缘关系。DG10S478位于10q25.2上转录因子7-样2(TCF7L2-原先的TCF4)基因的内含子3中。这个标记物在明确定义的74.9kb LD块内(基于CEPH白种人HapMap PhaseII),包含内含子3的一部分、外显子4的全部和内含子4的一部分(图1)。

[0214] 当在CEPH白种人HapMap家族DG10S478被基因分型时,清楚的观察到SNP rs12255372的等位基因G与DG10S478的等位基因0几乎完全相关($r^2 = 0.95$, $P = 5.53 \times 10^{-38}$),rs12255372的等位基因T与DG10S478的其他等位基因相关。此外,DG10S478的等位基因8和12造成的风险一致($P = 0.3$)。因此很自然将DG10S478的所有非0等位基因合并成一个复合等位基因,称为等位基因X。等位基因X在对照和患者中的频率分别是27.6%和36.4%。采用乘法模型(16,17),相对非携带者的风险,带有的每拷贝的等位基因X预测RR为1.50。

[0215] 与II型糖尿病关联的DG10S478的重复

[0216] 为了验证DG10S478与II型糖尿病的关联,在228个病例的丹麦II型糖尿病群组 and 539个对照中对微卫星进行基因分型。丹麦群组选自丹麦的PERF(Prospective Epidemiological Risk Factors,预期流行病学危险因素)研究(19)。这个女性II型糖尿病群组先前已经确诊患有II型糖尿病。冰岛观察到的关联被重复(表3)。

[0217] 表3:在丹麦DG10S478与II型糖尿病的关联

[0218]

等位基因	患病组频率 (n=228)	对照组频率 (n=539)	RR [95% CI]	两侧 P
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
X	0.331	0.260	1.41[1.11,1.79]	0.0048

[0219] 复合风险等位基因 X 在对照中频率为 26.0%，在 II 型糖尿病中为 33.1%，得到的估计 RR 为 1.41 ($P = 0.0048$)。

[0220] 随后，在 PENN CATH 研究的包括美国白种人 361 个 II 型糖尿病群组病例和 530 个对照的中对微卫星进行基因分型。这项研究是在宾夕法尼亚大学医疗中心 (University of Pennsylvania Medical Center) 接受心导管插入术的患者连续群组中进行的生化和遗传因素与冠状动脉粥样硬化关联性的横断研究。II 型糖尿病定义为具有以下病史，空腹血糖 $> 126\text{mg/dl}$ ，餐后 2 小时血糖 $> 200\text{mg/dl}$ ，40 岁以上个体中使用口服降血糖药或者胰岛素。在冰岛人中观察到的关联在这个种群中也被重复 (表 4)。

[0221] 表 4：在美国 DG10S478 与 II 型糖尿病的关联

[0222]

等位基因	患病组频率 (n=361)	对照组频率 (n=530)	RR [95% CI]	两侧 P
-4	0.001	0.000	-	-
0	0.615	0.747	0.54	3.3×10^{-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	1.2×10^{-4}
16	0.040	0.020	2.07	0.012
X	0.385	0.253	1.85[1.51,2.27]	3.3×10^{-9}

[0223] 复合风险等位基因 X 在对照中频率为 25.3%，在 II 型糖尿病中为 38.5%，得到的估计 RR 为 1.85 ($P = 3.3 \times 10^{-9}$)。利用 Mantel-Haneszel 模型 (NOTE 3)，从全部 3 个群组得到的组合得到总的两侧 P 是 4.7×10^{-18} 。

[0224] 3 个种群中复合风险等位基因与 II 型糖尿病的关联构成强有力证据，表明 TCF7L2 基因变异体增加患 II 型糖尿病风险。

[0225] 在毫无疑问的建立等位基因 X 与 II 型糖尿病的关联之后，我们更细致的研究了遗

传方式。显性模型和隐性模型可以被拒绝,因为杂合体携带者相对非携带者风险显著升高 ($P < 1 \times 10^{-6}$),相对纯合子携带者风险显著降低 ($P < 0.0001$)。乘法模型提供更好拟合,但有证据说明纯合子携带者相对杂合体携带者的风险大于杂合体携带者相对于非携带者的风险。表 5 提供杂合体携带者和纯合子携带者相对非携带者的相对危险度的无模型估算。

[0226] 表 5:相对危险度的无模型估算

群组	基因型相对危险度			PAR
	00	0X [95% CI]	XX [95% CI]	
冰岛	1	1.41[1.17,1.70]	2.27[1.70,3.04]	0.21
丹麦	1	1.37[0.98,1.90]	1.92[1.13,3.26]	0.17
美国	1	1.64[1.23,2.19]	3.29[2.13,5.07]	0.28
组合	1	1.45[1.26,1.67]	2.41[1.94,3.00]	0.21

[0228] 3 个群组对风险等位基因具有相似的种群频率,但 RR 估值不同;美国群组中效应最强,丹麦群组中效应最弱。尽管在群组中 RR 相同是没有理由的,应当注意估算相对危险度的差异不需要没达到统计显著性 ($P > 0.05$)。对采用常用相对危险度从群组得到的结果进行组合,,杂合体携带者和纯合子携带者相对非携带者的相对危险度据估算分别为 1.45 和 2.41(表 5)。假设风险等位基因的种群频率是 26%,则杂合体和纯合子的携带者分别占种群的 38%和 7%。因此,这种变异体对于临床使用具有良好的预测价值。对应种群归属风险是 21%,这从公共卫生观点来看是实质性的。

[0229] 还注意到等位基因 X 在空腹血糖受损 (IFG) 个体 (空腹血糖在 6.1 和 6.9mM 之间) 中过量。复合风险等位基因 X 在 1393 个对照中频率为 27.7%,在 278 个 IFG 病例中为 37.1%,得到估算的 RR 为 1.54 ($P = 1.36 \times 10^{-5}$)。

[0230] TCF7L2 外显子 4LD 块内 SNP 标记物与 II 型糖尿病的关联

[0231] 我们在表 6 中列出位于 TCF7L2 外显子 4LD 块内的微卫星和 SNP 标记物。该表包括公布的 SNPs,以及通过对整个 LD 块区域测序发现的 SNPs。该表还提供位于所述块内的多态性微卫星标记物。

[0232] 表 6. 位于 TCF7L2 外显子 4LD 块内的多态性标记物 (在标记物 rs4074720 和 rs7087006 之间,定位在 Build 34 坐标:rs4074720(B34:114413084)-rs7087006(B34:114488013) = 74929bp。对序列识别索引进行合适的标记,涉及各种情况下包括多态性的扩增引物和正反向引物的 SEQ ID 号,如序列表所示。

[0233] A. 公共 SNPs (包括全部 HapMap 种族)

[0234]

公共名称	染色体 10 B34 定位	碱基变化	序列 ID NO:
rs4074720	114413084	A/G	
rs4074719	114413145	C/T	
rs4074718	114413204	C/T	
rs11196181	114413605	A/G	
rs11196182	114414744	C/T	
rs4603236	114414765	G/T	
rs7922298	114414856	C/T	
rs17747324	114417090	C/T	
rs7901695	114418675	C/T	17-19
rs11196185	114420079	C/T	
rs4132115	114420083	A/C	
rs4506565	114420628	A/T	14-16
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	
s12266632	114429546	C/G	

[0235]

rs6585199	114429758	A/G	
rs7896811	114431304	C/T	
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	
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	8-10
rs12265110	114453606	C/T	
rs7077039	114453664	C/T	
rs11196198	114456472	A/G	
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	

[0236]

rs7895340	114466112	A/G	23-25
rs11196200	114466525	C/G	
rs11196201	114467894	A/T	
rs11196202	114470254	A/G	
rs11196203	114470447	A/C	
rs11196204	114470518	A/G	
rs11196205	114471634	C/G	20-22
rs10885409	114472659	C/T	
rs12255372	114473489	G/T	5-7
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	

[0237] B. 发现的新 SNPs, 随后在 TCF7L2 外显子 4LD 块内验证 (扩增引物如下):

deCODE 名称	染色体 10 B34 定位	碱基变化	序列 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

[0239] C. TCF7L2 外显子 4LD 块内的多态性微卫星（扩增引物如下）：

微卫星	C10 B34 起始	C10 B34 终止	序列 ID NO:
DG10S2164	114460344	114460627	48-50
DG10S478	114460845	114461228	2-4
DG10S479	114475487	114475632	51-53

[0240] 表 7. TCF7L2 外显子 4LD 块内选择性标记物的扩增引物和引物

[0241] > DG10S478

[0242] TTCAGGCCATTGGTGTGTATATATTCAAGATTGCTCACAGGTCCAAAGCT

[0243] TAACTTAAGCTCCCTGAGACATATCATAAATATGATTGGGGAAAAACCT

[0244] AATGGGCCATGATCAGAACATTATTATTCAACAAAGGATGAAATGCTTAAGC

[0245] CAAGATGGCCTTCTTCTTTCTTTCTTTCTTTCTTTTTTTAATGAAAGTTGA

[0246] GCAGACTCCCGTCCAACAGTTTTCAATGTAGGAATTCCCACAGCCCCATTG

[0247] ATTGCAGTTTGTTGAAAAGTTTAATGTTTTGTAGGCAATTCATAATTTCCAC

[0248] ATTGAACAGCCTGAGAGGAAGAGAGCTGGAGCCCCTGTTGTTTTTGTAGTG

[0249] GGATGGTGGGAAC TTT (SEQ ID NO :2)

[0250] 弓|物：

[0251] F :TTCAGGCCATTGGTGTGTA (SEQ ID NO :3)

[0252] R:AAAGTTCCCACCATCCCCT (SEQ ID NO:4)

```
[0253]    > rs12255372
```

[0254] TTGTCCCTTGAGGTGTACTGGAAACTAAGGCGTGAGGGACTCATAGGGGT

[0255] CTGGCTTGGAAAGTGTATTGCTATGTCCAGTTTACACATAAGGATGTGCA

[0256] AATCCAGCAGGTTAGCTGAGCTGCCCAGGAATATCCAGGCAAGAAT

[0257] K

[0258] ACCATATTCTGATAATTACTCAGGCCTCTGCCTCATCTCCGCTGCCCCC

[0259] CGCCCCCTGACTCTCTTCTGAGTGCCAGATTCAGCCTCCATTTGAATGCCAA
[0260] ATAGACAGGAAATTAGCATGCCAGAATCCACGTCTTTAGTGCACTCT
[0261] TCCCCAGCTCCAAACCTGTTACTGCTTGTGTTCAACATCTCAGTAAAGCTCA
[0262] ACAACATCGACCCATT (SEQ ID NO :5)
[0263] 引物 :
[0264] F :TTGTCCCTTGAGGTGTACTGG (SEQ ID NO :6)
[0265] R :AATGGGTCGATGTTGTTGAG (SEQ ID NO :7)
[0266] > rs12243326
[0267] GCTGTGAAATCCCCTGTGTAGTGGAAGAAGAAATAGCAAATCTTAGCTGC
[0268] CTTGGACCTGATATAATTATTTGTCTTCATTTACATGGTT
[0269] Y
[0270] ATCCTTCAAGGTGAATAAATGATGTGGGAGCTAGTCAAGGGGCTTTAGGTA
[0271] TGTGATTTTCATGCCTACTTTTTTTTAGGTAGAGAACTGAGGTCACAGGGTA
[0272] CTAGAGAATGGACTCTAAGATTCAGGTTTCTGAATTGCCTGTGGTTTTGTTG
[0273] ACTCAACTGCTCTTCTGTTGTTTTTTAGCCACATGCCTTGAAACAGTCCTCTT
[0274] TCCCATGTTTCTTCATCAGCACCATTAAACCAAGGTATACTGTCCTCTCTTAT
[0275] CTTTCACAAGGTCTTGGAGTTCCCATGCCTTTGTAAGCATCCCTCCCCGAGAT
[0276] TCAGCACCAACCAAAATCACATTTGGAAAAATTGCTTGTTTCCCAAGAAGCT
[0277] TTGGAGGATATGATTTTGTATAGAACGGGTTCACAGGTTTTCTGTTTCATTCTT
[0278] CTATGGTGGAGTGTGTGTGTATGTGACTCTGTCTTCTCTCCATTCC (SEQ IDNO :8)
[0279] 引物 :
[0280] F :GCTGTGAAATCCCCTGTGTAG (SEQ ID NO :9)
[0281] R :GGAATGGAGAGAAGACAGAGTCA (SEQ ID NO :10)
[0282] > rs7903146
[0283] AAGGGAGAAAGCAGGATTGAGCAGGGGGAGCCGTCAGATGGTAATGCAGA
[0284] TGTGATGAGATCTCTGCCGGACCAAAGAGAAGATTCCTTTTTAAATGGTGAC
[0285] AAATTCATGGGCTTTCTCTGCCTCAAAACCTAGCACAGCTGTTATTTACTGA
[0286] ACAATTAGAGAGCTAAGCACTTTTTAGATA
[0287] Y
[0288] TATATAATTTAATTGCCGTATGAGGCACCCTTAGTTTTTCAGACGAGAAAACCA
[0289] CAGTTACAGGGAAGGCAAGTAACTTAGTCAATGTCAGATAACTAGGAAAAAG
[0290] GTTAGAGGGGCCCTGGACACAGGCCTGTGTGACTGAGAAGCTTGGGCACTTC
[0291] ACTGCTACATTTTCATCTCTTCGCT (SEQ ID NO :11)
[0292] 引物 :
[0293] F :AAGGGAGAAAGCAGGATTGA (SEQ ID NO :12)
[0294] R :AGCGAAGAGATGAAATGTAGCA (SEQ ID NO :13)
[0295] > rs4506565
[0296] CTGATGAGGGTAGGGAGCATCTGTCTGCAGCTTCATCTTCATTGTCTAGGGG
[0297] CTCCAGAAATATCTGTGAGTAAATAAGTTATTTAATCTTTGCCTCAAATTTCC

[0298] AGTGACTGTAGGGATATAGCTGTGAGCCTCTAGGAGCTGAGATTTTTTAAAT
[0299] TTCCCACTTAAACATTTATTTAAAAATTTTGTGCTCAGCATGG
[0300] ACTAAGGACTTTACATTCATTAACCTCATTTACAGCTTGATCCTATGCGGT
[0301] GGGCATTCAATTTACAGAGGATCCCATTTTACAGGTGAGGAAGAGGCCAGC
[0302] TAGGGGTGCAGCCTAGGTTAGTATTCTAGAGCTCATCAGGCTGTGTTGTC
[0303] XCAGTGAAAGAATAAGCAAAGAAGTGAATGTTGTGCATTGAGAAAAATG
[0304] ACTCTCGGAGGAGGATGAGCCTCTCGGATATGGCGACCGAAGTGAT
[0305] W
[0306] TGGGGCCCTTGTCAAGGGTCTCTATTATGGCATCAAGAAAAGATGCTGCT
[0307] TTCGGTGATGCCCCAGGAGAGCCTCAATATTTTACATGGGAAACCTAAAA
[0308] AAGGGGCCATGTTGTGGTCTCTGCACCTAAGA (SEQ ID NO :14)
[0309] 引物 :
[0310] F :CTGATGAGGGTAGGGAGCA (SEQ ID NO :15)
[0311] R :TCTTAGGTGCAGAGACCACAAC (SEQ ID NO :16)
[0312] > rs7901695
[0313] TATTTAGAAACCATAAAAATCCACCTATTTGAGGTGTACAATTGAGTGATTTT
[0314] CTGTATAGTCACAGATCTGTGCAGTCATCCACACCCTCTAACTCCAGGACAT
[0315] TTTCTCACCCCCGAGGAGAAACCTCCCTTACCCATTAGCAGTCACTCCTCAT
[0316] TTCCTCTCCCCCAGCCCCCTGGCAATCACTGTGGATTTGCCTGTTCTTGACAT
[0317] TTCATATAAATGGTATCATAAAATCTA
[0318] Y
[0319] GGGCTTTTGTGTCTGTCTGCTTTCACTTAGCATACGGTTCTCAAGGTTTCATCC
[0320] AGTATTGTAGCATCTATCAGTATGTCATTCCTTTTTATGGCCAAATAATATTT
[0321] TATTGTATGGATAGACATTTTGTTTATTCAATTTATCTGTTTTTGGTTATTATGA
[0322] GTAACACTACTATGAACATTTTGCACAAATTTTGTATTGACATGTTTTTCATT
[0323] TCTCCTGGGTATAGTCCTATGAGTGGAATTGCTGG (SEQ ID NO :17)
[0324] 引物 :
[0325] F :TATTTAGAAACCATAAAAATCCACCTAT (SEQ ID NO :18)
[0326] R :CCAGCAATTCCACTCATAGGAC (SEQ ID NO :19)
[0327] > rs11196205
[0328] TTGTCTCCTTTTGTCTGCTACTGTGAATGATCCTGTGATGATCATCTTTGTG
[0329] TGTAATCTTTGTCCCTCGCCCCCTCCCTTTTATTATTTTCTTGGGATAGA
[0330] CCCCAGGACAAAAGGTAGAAAAGAACAAGTGTTAAAAAATTTCTTGATAC
[0331] ATAGCCACAGATTATTTTCTGAAAGTTCTCAACATTTATAACTAC
[0332] S
[0333] AGCAGTATGTAAGAGAGTTATGGTTGGAATGATTTTAATGTCTCTGGGGAAT
[0334] TTAACAACAAAAAACTTTAGGCTTCTTTGGAGAGAGACATGCCCTTA
[0335] ACTCCACCCCGCCCTAGAACAGAGACCCAGCCCATCCAAGTCAGCCTCCC
[0336] CAGGTCTCCACCTTCAAAACAGGCAAACGAAATCATTTCTGAATAATT

- [0337] GGTAGGCTTCAAGGTCAGATGTT (SEQ ID NO :20)
- [0338] 引物 :
- [0339] F :TTGTCTCCTTTTGTCTGCTAC (SEQ ID NO :21)
- [0340] R :AACATCTGACCTTGAAGCCTAC (SEQ ID NO :22)
- [0341] > rs7895340
- [0342] TCAGGGACAGTGCATAGGTGTAAAGAAGTTGCTGGTTGGGGTTCTAATGC
- [0343] AGGTTTCTCCAAAAGTGAATGCCCTGTAAAAAATTCTTAACAAATATA
- [0344] CAGAGATTTTTTTTTTAAAAAAGTGTGACAGTTCTAGACACCTAGAGAGTAA
- [0345] A
- [0346] R
- [0347] TGAAGAAGCCTGTTTTACAGGTTTCCCGCTCCCTGAATTTCCCAGCATGGTCC
- [0348] AGGCTTTGAAATTTATTTATCTGCTTTTGGCAATGGTTGATGGGAATTTCCCA
- [0349] CATTTATTTTTTAGCTACAGAGAAAGGACATTATCTTTAAATCTCTTCGTTG
- [0350] TTCTCTCTTTGA (SEQ ID NO :23)
- [0351] 引物 :
- [0352] F :TCAGGGACAGTGCATAGGTG (SEQ ID NO :24)
- [0353] R :TCAAGAGAGAGAACAACGAAGA (SEQ ID NO :25)
- [0354] > SG10S405
- [0355] TATTTAGAAACCATAAAATCCACCTATTTGAGGTGTACAATTGAGTGATTTT
- [0356] CTGTATAGTCACAGATCTGTGCAGTCATCCACACCTCTAACTCCAGGACAT
- [0357] TTTCTCACCCCCGAGGAGAAACCTCCCTTACCCATTAGCAGTCACTCCTCAT
- [0358] TTCCTCTCCCCCAGCCCCCTGGCAATCACTGTGGATTTGCCTGTTCTTGACAT
- [0359] TTCATATAAA
- [0360] Y
- [0361] GGTATCATAAAATCTATGGGCTTTTGTGTCTGTCTGCTTTCACTTAGCATACG
- [0362] GTTCTCAAGGTTTCATCCAGTATTGTAGCATCTATCAGTATGTCATTCCTTTTT
- [0363] ATGGCCAAATAATATTTTATTGTATGGATAGACATTTTGTATTATTCATTTATC
- [0364] TGTTTTTGGTTATTATGAGTAACACTACTATGAACATTTTGCACAAATTTTG
- [0365] TATTGACATGTTTTCATTTCTCCTGGGTATAGTCCATGAGTGGAATTGCTGG
- [0366] GTCATATAATAAATAACTGTTTAACATTTTGGGGAGCTGCCAACTTTTAAA
- [0367] ACCTTGGGTTCTGTGATGTACCAGTTGTGTTAGGCA (SEQ ID NO :26)
- [0368] 引物 :
- [0369] F :TATTTAGAAACCATAAAATCCACCTAT (SEQ ID NO :27)
- [0370] R :TGCCTAACACAACCTGGTACATC (SEQ ID NO :28)
- [0371] > SG10S428
- [0372] TGCCAGGGGTTTTATGGTTAATTTTCTCCATTATGAGGGTTGACTCAGCCTT
- [0373] GGGTATTAGATGTCTTTGAGAATCCAGGGTTCAAATACCACAGCTGGTAGAA
- [0374] TGTTTCTCAACTTGGAGCCAATCTCCATCTACTGAAGGTACGCTGGTTTAGA
- [0375] CAGACAACAGGGACATCAGCATTTTTAAAAAGCGGTGGAAAAAGTTTGCTTG

[0376] TCTTGATTGGAGCCATGACATTTTATTTTGAATTTCAAATAACATGAAGGG
[0377] AGGTTTGGAGCGGTTTTTGGTTTATCCAAAGGGCAGTGGATTGAAGGCTGAG
[0378] AAACACCAGGCTGAATGGGAGAGGGGTTGGGGTCCCCCTGTGAGATAGTGA
[0379] AACAATGGTAGTGCCATCCAATGATAGGCACTTTTCTGTCATTGAGAAGCAG
[0380] AAAGGGGGCCAGAGGCCCATTTGGCCTTACTGGG
[0381] M
[0382] AGTAAGCTGTAGAGCTGCTGCCTTTTCGTGAAAGGGTTGACACCAACCTTCT
[0383] CCCCCAGGAAGAGTGACCAGGGACCTGAGGGGCATGGTCGAGCAGATGACA
[0384] GCCTTTGTAAAACATCTCC (SEQ ID NO :29)
[0385] 引物 :
[0386] F :TGCCAGGGGTTTTATGGTTA (SEQ ID NO :30)
[0387] R :GGAGATGTTTTACAAAGGCTGTC (SEQ ID NO :31)
[0388] > SG10S422
[0389] TTGGTAGAGATGGGGTCTCCTAGGCTGGTCTTGAACCTCCTGG
[0390] R
[0391] CTCAAGCAATCTTCTGCCTCAGCCTTCCAAAGTACTGGGATTACTGGCGTG
[0392] GGCCACCATGCCTGGCTTGAAATTTTTCTATGGCTTTATTCTTTCTCCAAGTA
[0393] CAGAGTCTACCCAACCTTCTGAGATCTTTGGTTTTCTTTTCTAGGTAACCTAT
[0394] AGTACATACTTATTTATGTTAAACAACAGCAATCACACATTTCTTTTCTATA
[0395] CAGTCATGCTTTATAGGCAATAAAGCCTCCGTCTTAGGCTTTCTGGATTTTT
[0396] TCAAAAGATGCAATTCCTGGAGTATGTTTTTACTTAGAGCAAAGCAGCCTAG
[0397] TCTCCTATACCTTCTGCATCTGCAGAAAAGTTGGTTAAACAGACTTTGTAAT
[0398] GATGCCCCCTTACAATTCTGAAGGGACTTGTGAAATAGTTTCACAGAGTTTCA
[0399] GTGTTAGGTATATTTGATCAATGCTAACTTTTGAAAACTTTGGTGCCTGTAT
[0400] GATTGAGAGGGTAGGGCAGAATATTAA
[0401] ATTAATCACAACTTCTTGTATTTTAACCATCTGGGTAAATTGGGATTCC
[0402] GTGACGCCCAGGCAAAATTAT (SEQ ID NO :32)
[0403] 引物 :
[0404] F :TTGGTAGAGATGGGGTCTCC (SEQ ID NO :33)
[0405] R :ATAATTTTGCCTGGGCGTCA (SEQ ID NO :34)
[0406] > SG10S427
[0407] TATCTTATATCCCCTCCAAGCATTCACTAACTGATGGATTAGTGAGTTGG
[0408] XTTGAGAAGCATAAAGGCTCGTCTCCATGTGCTTCTAAGCATTGTGTCT
[0409] AAGTTCTGTTTGGTTTCTGAGTGAACTGTCTTAATGTTACCAACAGAAGTT
[0410] AAATGCCTAAGAG
[0411] W
[0412] TTCTTATACATGGGCTGAGTACCTCTGTGACTGGGCAAGCCACCTCACCTCA
[0413] TTTTACCTTGTCTGCAAAATGAGGAACTGGGTCAACTCATCGTTCAAATCTC
[0414] ACTGAAAGCTAATTGATCGCTTTTGACAGAAGTAGCTCCCTTGGGCCGTATA

- [0415] TTTATTTCTAGCTTGGAGGAAGGTGGGGACAGACAGAATTGATGTACACCT
- [0416] TTATTTTATCTCTATGGTAAACCTGTGCATACTAAAGCATTCTCTGGTCTT
- [0417] TTGAGATGAGTGTATACATTGTGTCTGGCCCTGTGCATTTTTTACCAAGAAAGT
- [0418] AAGTTTTGTTGAGTAAACTTGGGTTGTATGAAGAACTGCATGCTCACCCTAC
- [0419] TCAAGTAGCTTTTGTACCTAAAGGACAGCTGCTCATATGTACTTGACTTCCT
- [0420] TTAAAGTGAAGGATGATGACATTTGAAAAACGGAGGTTGAAAAGGAG (SEQ ID NO :35)
- [0421] 引物:
- [0422] F:TATCTTATATCCCCTCCAAGCATTC (SEQ ID NO :36)
- [0423] R:CTCCTTTTCAACCTCCGTTTT (SEQ ID NO :37)
- [0424] > SG10S408
- [0425] TTGAGCATGTGTTATTTAATGAGTTATACCTCTGTCATATGTGTGTGTTTATA
- [0426] TCACAAAATAAATTATTTTTATAAAACCATATTTTGAGTCATCATTTGTGACA
- [0427] ATGTCTTCTTTTCTCTGGTATAAATGAGGCATGTAGAAAAGAAGATTGACATT
- [0428] TGCTAGAAGCTTCCCCTTTCTCTAACTCCACAATAAAATGGATGCTCATAA
- [0429] TTACATCTGCTCCTATAAGGTCAAGATTTTCAAGGCTGGAAGTGACCTTAGAT
- [0430] CATTTAGGCCCAACTTGCCCTCAGGAAAGGAAACTGAGGCCAGAGATGCC
- [0431] TTAAGTGAATTGCCCAATGTCACACGCTGAGTCAGTGGCCAGAGCAAGGCTT
- [0432] GGATCCAGTTCTCTGCTCCCTTTCCAGAGCCTTGTGATGTCTTCTCTCCTACA
- [0433] GGAGGTGAAAATAACTGCTGTGGCTGGTTCTGTTTTGCTGACTGTAAATTGG
- [0434] GTCATGGTCAGGGACAGTGCATAGGTGTAAAGAAGTTGCTGGTTGGGGGTTT
- [0435] TAATGCAGGTTTCTCCAAAAGTGAATGCCCTGTAAAAAATTTCTTAACA
- [0436] AATATACAGAGATTTTTTTTTT
- [0437] W
- [0438] AAAAAAGTGTGACAGTTCTAGACACCTAGAGAGTAAAGTGAAGAAGCCTGT
- [0439] TTTTCAAGTTTCCCGCTCCCTGAATTTCCAGCATGGTCCAGGCTTTGAAATT
- [0440] TATTTATCTGCTTTTGGCAATGGTTGATGGGAATTTCCACATTTATTTTTTA
- [0441] GCTACAGAGAAAGGACATTATCTTTAAATCTCTTCGTTGTTCTCTCTTTTG
- [0442] AGTGAGGAGAGAAGATGTGAATCCTGGCAGTGGTTCAGAGTGGACACAGCC
- [0443] CCTGTGTTTGTGGCATAGGCTCTGTGGGCCCCATGCCAGGGAGCAGTACCC
- [0444] CGTGTAAGGAGTGGGGTTTGTCCATTTGGATAGAGCAAAGATCCTCCACC
- [0445] TCAAAATCCCACAAGAACAGTTGCCACAACCTGGGCCCTAAGCATCTCATTTT
- [0446] CCTATGTAGAAATTAATGATCTGGAGGAGATGGCAAAACATTCCTTCCAGAG
- [0447] CCTGTGTGGATTTTGG (SEQ ID NO :38)
- [0448] 引物:
- [0449] F:TTGAGCATGTGTTATTTAATGAGTTA (SEQ ID NO :39)
- [0450] R:CCAAAATCCACACAGGCTCT (SEQ ID NO :40)
- [0451] > SG10S409
- [0452] TAGTGCTCAGTATTTCCAACGTTCTGTTTATTTAAGATGAAAATTGCTGT
- [0453] AGTTAATAAGCACTTCCCCATGTCATTAATAATGCTTAAGGATTTTAAATG

- [0454] ACCACATAACAGTCCATAATATGATTAAACCCCAATTTACTGAATCAATG
[0455] ICATATTGTTGGGTCTTTAGATTGTCTCCTTTTGTCTGCTACTGTGAA
[0456] TGATCCTGTGATGATCATCTTTGTGTGTAATCTTTGTCCCCTCGCCCCC
[0457] TCCCCTTTTATTATTTTCTTGGGATAGACCCAGGACAAAAGGTAGAAAAGA
[0458] ACAAAGTGTAAA
[0459] M
[0460] AATTTCTTGATACATAGCCACAGATTATTTTCCTGAAAGTTCTCAACATTTAT
[0461] AACTACGAGCAGTATGTAAGAGAGTTATGGTTGGAATGATTTTAATGTCTCT
[0462] GGGGAATTTAACAACAAAAAACTTTAGGCTTCTTTGGAGAGAGA
[0463] CATGCCCTTAACCTCCACCCCGCCCTAGAACAGAGACCCAGCCCATCCAAG
[0464] TCAGCCTCCCCAGGTCTCCACCTTCAAAACAGGCAAAACGAAATCATTTCT
[0465] TTGAATAATTGGTAGGCTTCAAGGTCAGATGTT (SEQ ID NO :41)
[0466] 引物 :
[0467] F :TAGTGCTCAGTATTTCCAACGTTCT (SEQ ID NO :42)
[0468] R :AACATCTGACCTTGAAGCCTACC (SEQ ID NO :43)
[0469] > SG10S406
[0470] TAGTGCTCAGTATTTCCAACGTTCTGTTTATTTAAGATGAAAATTGCTGT
[0471] AGTTAATAAGCACTTCCCCATGTCATTAATAATGCTTAAGGATTTTAAATG
[0472] ACCACATAACAGTCCATAATATGATTAAACCCCAATTTACTGAATCAATGCC
[0473] ATATTGTTGGGTCTTTAGATTGTCTCCTTTTGTCTGCTACTGTGAATGATCC
[0474] TGTGATGATCATCTTTGTGTGTAATCTTTGTCCCCTCGCCCCCTCCCCTTTT
[0475] ATTATTTTCTTGGGATAGACCCAGGACAAAAGGTAGAAAAGAACAAGTG
[0476] TTAATAAATTTCTTGATACATAGCCACAGATTATTTTCCTGAAAGTTCT
[0477] S
[0478] AACATTTATAACTACGAGCAGTATGTAAGAGAGTTATGGTTGGAATGATTTT
[0479] AATGTCTCTGGGAATTTAACAACAAAAAACTTTAGGCTTCTTTGGAGAGA
[0480] GACATGCCCTTAACCTCCACCCCGCCCTAGAACAGAGACCCAGCCCATCCAAG
[0481] TCAGCCTCCCCAGGTCTCCACCTTCAAAACAGGCAAAACGAAATCATTTCTT
[0482] GAATAATTGGTAGGCTTCAAGGTCAGATGTT (SEQ ID NO :44)
[0483] 引物 :
[0484] F :TAGTGCTCAGTATTTCCAACGTTCT (SEQ ID NO :42)
[0485] R :AACATCTGACCTTGAAGCCTACC (SEQ ID NO :43)
[0486] > SG10S407
[0487] TGCTATGTCCAGTTTACACATAAGGATGTGCAATCCAGCAGGTTAGCTGAG
[0488] CTGCCCAGGAATATCCAGGCAAGAATGACCATATTCTGATAATTACTCAGGC
[0489] CTCTGCCTCATCTCCGCTG
[0490] S
[0491] CCCCCCGCCCCCTGACTCTCTTCTGAGTGCCAGATTGAGCCTCCATTTGAATG
[0492] CCAAAATAGACAGGAAATTAGCATGCCAGAATCCACGTCTTTAGTGCACTCT

- [0493] CTCCCCAGCTCCAAACCTGTTACTGCTTGTGTTCAACATCTCAGTAAAGCTCA
- [0494] ACAACATCGACCCATTACTTAGGCCTCAAACCTTGGGTGGCATCGTCGATTG
- [0495] CTCTTTTCTTTTCATACCCACATTCAACCCATCAGCCCATCCCACAGGCCCAA
- [0496] GTGTGTCCTCTCTACCTTCAAAGCGTGTGTGGCATCCACCGCTTATCACCACC
- [0497] TCTGCCATTACCACTGGAGTCCAGTGCCATCATC
- [0498] TCTCACTTGATGTGGCCAGAGTGTCTTTGCTGGTCTCCTTCTTGCTTCC
- [0499] TACCTTTGTAACAGCCTATCATCTATCTCTGGTCTCCATAGCTCACTCCC
- [0500] ATACTTTGAGAGGGCCTTTGAAAGCCTTAGACAGATCATATCACAGACCT
- [0501] CTATACTGAAAGTCGGG (SEQ ID NO :45)
- [0502] 引物 :
- [0503] F :TGCTATGTCCAGTTTACACATAAGG (SEQ ID NO :46)
- [0504] R :CCGACTTTTCAGTATAGAGGTCTG (SEQ ID NO :47)
- [0505] > DG10S2164
- [0506] CCATCTGTGGAGCAGAGTCACTGAAAGGAAATACTGGAAATACTGGAAGC
- [0507] CACTTGGTGTTTTATCAAGGATGTGAGGTTTCCTGGCAACTTTGTCGCCA
- [0508] TATCATCATCATCATCACCATCATCATCATCATCATCATCATCATCA
- [0509] TCATCATCATCATCATCTGCCCTTTAAGTTTTCTGCTTGTTTAGAAAAGA
- [0510] AATTTATACAGAGCCCCCAGTAGCAGCTGTAAGGGGGCAGGTTCTTGGAG
- [0511] CAGCCCATCCTCAACATTCTTGCTGCTGATGGAA (SEQ ID NO :48)
- [0512] 引物 :
- [0513] F :CCATCTGTGGAGCAGAGTCA (SEQ ID NO :49)
- [0514] R :TTCCATCAGCAGCAAGAATG (SEQ ID NO :50)
- [0515] > DG10S479
- [0516] TCCACGCAGAGAGGATCTAAATCTGGCTCTTTGCAATTGCCTTCATACAT
- [0517] GTGCATACACACCACACACACACACACACACACACACACACACACA
- [0518] CAGACACATACATATGCACACACCCCGACTCAATGGAGGACCCTC (SEQ ID NO :51)
- [0519] 引物 :
- [0520] F :TCCACGCAGAGAGGATCTAAA (SEQ ID NO :52)
- [0521] R :GAGGGTCCTCCATTGAGTCG (SEQ ID NO :53)

[0522] 为了进一步研究 TCF7L2 外显子 4LD 块内其他标记等位基因比等位基因 X 显示更高的与 II 型糖尿病相关性的可能性,我们使用 HapMapCEU 样品中产生的 DG10S478 基因型数据。HapMap Phase I 中与 DG10S478 具有最强相关性的 5 个 SNPs,按照递减顺序,是 rs12255372 ($r^2 = 0.95$), rs7903146 ($r^2 = 0.78$), rs7901695 ($r^2 = 0.61$), rs11196205 ($r^2 = 0.43$), 和 rs7895340 ($r^2 = 0.42$)。我们将这 5 个 SNPs 在 3 个群组中进行基因分型,5 个 SNPs 和 DG10S478 的相关性与在 CEU 样品中观察到的非常类似, DG10S478 用作双等位基因标记物。全部 5 个 SNPs 显示与 II 型糖尿病的关联。尽管一些 SNPs 在一个或者两个群组中显示稍微更高的估算相对危险度和较低的 p- 值,当来自 3 个群组的结果利用 Mantel-Haenszel 模型组合时,没有 SNP 比 DG10S478 显示更强的与 II 型糖尿病的关联。但是,尽管相对等位基因 X ($RR = 1.56$, $P = 4.7 \times 10^{-18}$), rs11196205 和 rs7895340 与 II 型糖尿病的关联明显

较弱,但 rs12255372 的等位基因 T ($RR = 1.52$, $p = 2.5 \times 10^{-16}$) 和 rs7903146 的等位基因 T ($RR = 1.54$, $p = 2.1 \times 10^{-17}$) 与 II 型糖尿病关联的强度是相似的。

[0523] 在 2005 年 10 月公开了 HapMap Phase II 之后,鉴定了两个其他的 SNPs,显示与微卫星 DG10S478 的强相关 -rs12243326 ($r^2 = 0.961$) 和 rs4506565 ($r^2 = 0.716$)。与 II 型糖尿病易感性相关的等位基因对 rs12243326 来说是 C(C/T SNP),和对 rs4506565 来说是 T(A/T SNP)。

[0524] 应当注意在那些载有 rs7903146C 等位基因的单倍型中间,那些载有 rs10885406A 等位基因的单倍型相对那些载有 rs10885406G 等位基因的单倍型的估算相对危险度是 1.06,但该差异在统计上没有显著意义 ($P = 0.22$)。

[0525] 为了重复和改进 II 型糖尿病的这种关联,我们在一个大的追加丹麦群组和一个遗传多样性更高的西非群组中对 DG10S478, rs12255372 和 rs7903146 进行基因分型,所述丹麦群组由 1111 个病例和 2315 个对照组成,所述西非群组由来自非裔美国人糖尿病研究的 618 个病例和 434 个对照组成 (23)。在丹麦人中,全部 3 个变异体都是与疾病风险强相关的,和先前在冰岛观察到的一致。但是,rs7903146 等位基因 T 的关联 (相对危险度 = 1.53, $P = 4.06 \times 10^{-14}$, $PAR = 24.4\%$) 显著的高于其他两种变异体提供的关联。在西非研究组中,在对亲缘关系和种族起源进行校正后,我们重复了 rs7903146 的等位基因 T 与 II 型糖尿病的关联 (相对危险度 = 1.45, 95% C. I. = 1.20-1.76, $P = 0.000146$, $PAR = 22.2\%$),但另两种变异体不适用该情况。这表明 rs7903146 的等位基因 T 或者其自身是风险变异体或者,或者是未知风险变异体的最接近的已知相关物。西非组中标记物 DG10S478 和 rs12255372 被排除为风险标记物是可能的,因为与欧洲祖先的种群不同,欧洲人中 rs7903146 的 T 等位基因几乎只存在于同时载有 DG10S478 的等位基因 X 和 rs12255372 的等位基因 T 的染色体上,在西非人中 rs7903146 的 T 等位基因与 DG10S478 和 rs12255372 的两个等位基因同时存在。这与观察到的一致,即 T 是 rs7903146 的祖先等位基因,而 DG10S478 的等位基因 X 和 rs12255372 的等位基因 T 都不同于黑猩猩参照序列。一般来说,这种发现也与预期相对多样性的种群,例如西非的种群,提供了与在同质性更高的种群中强连锁不平衡区域发现的关联信号的改进 (refine) 的关联性一致。

[0526] 讨论

[0527] 在这项研究中我们描述了一个新的 II 型糖尿病候选基因的鉴定,所述基因位于先前报道的 10q 连锁区域内 (10),编码 10q25.2 上的转录因子 7-样 2 (TCF7L2- 原先的 TCF4)。我们表明它在冰岛,丹麦和美国以相似的频率和相对危险度造成 II 型糖尿病风险。当变异体不能解释 II 型糖尿病家庭聚集性的实质部分时,从公共卫生观点来看,至少为 20% 的种群归因风险是显著的。相对于非携带者,风险复合等位基因的杂合体携带者 (大约占种群的 38%) 和纯合子携带者 (大约占种群的 7%) 的相对危险度分别是 1.45 和 2.41。因此,这种变异体对于临床使用具有良好的预测价值。

[0528] 我们报告了作为 II 型糖尿病相关微卫星的变异体, DG10S478,位于 TCF7L2 基因的第三个内含子内。TCF7L2 基因产物是一种包含高迁移率族 (HMG) 盒的转录因子,在 Wnt 信号途径中起作用。这个途径被认为是一个关键的细胞发育和生长的调节机制;它被称为 Wnts 的分泌糖蛋白调节,在结合关连受体复合物时 Wnts 启动靶细胞内许多信号级联反应, Wnts 由 Frizzled 家族的成员和 LDL 受体家族的成员 Lrp5/6 组成 (24)。Wnt 信号从降解复

合物中分离在这个途径中主要成员 β -连环蛋白,将其转移到细胞核,在细胞核中它短暂的将 TCF 因子从阻遏剂转化成翻译激活剂 (25)。通过结合钙粘蛋白, β -连环蛋白对调节细胞粘着也是非常重要的。

[0529] TCF7L2 的 NCBI RefSeq 包括 14 个外显子。但是,Duval 等人 (26) 显示 TCF7L2 具有 17 个外显子,其中 5 个是选择性的;此外,还有报导称 3 个选择性剪接受体位点被使用。这项研究还证明了位于 TCF7L2 基因 3' 端的 3 个连续外显子的选择性使用,这改变了最后一个外显子中使用的读码框,导致大量短的、中等长度的或者长的 COOH- 末端的合成。

[0530] 与 TCF7L2 类似,罕见的 Mendelian 型 II 型糖尿病即青年发病的成人型糖尿病 (MODY) 的 6 个定位克隆基因中的 5 个,是转录因子 (27)。其他的转录因子也与 II 型糖尿病的发病有关,包括过氧化物酶体增殖物激活受体 γ (PPAR γ) (7) 和叉头基因家族 (28, 29)。Noble 等人描述了 I 型糖尿病相关 TCF7 基因中的一个错义突变 (C883A) (30)。但是,还不清楚 TCF7 和 TCF7L2 是否采用相同途径对糖尿病的发病起作用。

[0531] 已经有 TCF7L2 基因突变的报道,包括在 (A)9 编码重复 (外显子 17) 的 A 缺失 (26, 31-33) 和结肠直肠细胞系中的大量突变 (26)。DG10S478 位于明确定义的 74.9kb LD 块 (CEPH Caucasian HapMapPhase II),包含外显子 4 和该外显子 5' 和 3' 端的侧翼内含子序列。有可能 DG10S478 自身就是成因变异体;还有可能 DG10S478 是基础变异体的替代物,所述基础变异体影响转录、剪接或者信息的稳定性。这类变异体可能与 DG10S478 存在强连锁不平衡 (LD),即变异体位于 TCF7L2 外显子 4LD 块内。

[0532] 多个证据表明这种基因在 II 型糖尿病发病中的肠内分泌功能。首先,已知 TCF7L2 和结肠直肠癌的进展有关 (54),致癌 TCF/ β -连环蛋白复合物的小分子拮抗剂也已经被报道 (35)。此外,出生后 24 小时内死亡的 TCF7L2^{-/-} 小鼠,缺乏肠上皮干细胞小室 (36)。TCF7L2 基因的变异体可以通过改变促胰岛素激素类胰高血糖素肽 1 (GLP-1) 的水平影响 II 型糖尿病易感性,所述类胰高血糖素肽 1 是胰高血糖素原基因编码的肽,所述高血糖素原基因在肠内分泌细胞的表达受 TCF7L2 的转录调控。与胰岛素一样,GLP-1 对血糖稳态产生重要影响 (12)。GLP-1 类似物和二肽基肽酶 IV 的抑制剂当前用于临床开发。

[0533] 尽管本发明进行了详细展示并参照其优选的实施方式进行了描述,本领域技术人员应当了解其形式和细节可以做出不脱离附加权利要求包括的本发明范围的各种改变。

[0534] 参考文献

[0535] 1. A. F. Amos, D. J. McCarty, P. Zimmet, Diabet Med 14 Suppl 5, S1 (1997).

[0536] 2. P. Zimmet 等人, Am J Epidemiol 118, 673 (Nov, 1983).

[0537] 3. W. C. Knowler, D. J. Pettitt, M. F. Saad, P. H. Bennett, DiabetesMetab Rev 6, 1 (Feb, 1990).

[0538] 4. B. Newman 等人, Diabetologia 30, 763 (Oct, 1987).

[0539] 5. A. H. Barnett, C. Eff, R. D. Leslie, D. A. Pyke, Diabetologia 20, 87 (Feb, 1981).

[0540] 6. A. L. Gloyn, Ageing Res Rev 2, 111 (Apr, 2003).

[0541] 7. D. Altshuler 等人, Nat Genet 26, 76 (Sep, 2000).

[0542] 8. A. L. Gloyn 等人, Diabetes 52, 568 (Feb, 2003).

[0543] 9. Y. Horikawa 等人, Nat Genet 26, 163 (Oct, 2000).

[0544] 10. I. Reynisdottir 等人, Am J Hum Genet 73, 323 (Aug, 2003).

- [0545] 11. R. Duggirala 等人, Am J Hum Genet 64,1127 (Apr,1999).
- [0546] 12. F. Yi, P. L. Brubaker, T. Jin, J Biol Chem 280,1457 (Jan 14,2005).
- [0547] 13. S. E. Ross 等人, Science 289,950 (Aug 11,2000).
- [0548] 14. E. A. Jansson 等人, Proc Natl Acad Sci USA 102,1460 (Feb 1,2005).
- [0549] 15. W. J. Nelson, R. Nusse, Science 303,1483 (Mar 5,2004).
- [0550] 16. C. T. Falk, P. Rubinstein, Ann Hum Genet 51 (Pt 3), 227 (Jul,1987).
- [0551] 17. J. D. Terwilliger, J. Ott, Hum Hered 42,337 (1992).
- [0552] 18. J. R. Gulcher, K. Kristjansson, H. Gudbjartsson, K. Stefansson, Eur J Hum Genet 8,739 (Oct,2000).
- [0553] 19. Y. Z. R. Bagger, B. J. ;Alexandersen, P. ;Tanko, L. B. ;Christiansen, C, J Bone Miner Res Suppl 1,1 (2001).
- [0554] 20. G. Benson, Nucleic Acids Res 27,573 (Jan 15,1999).
- [0555] 21. R. C. Lewontin, Genetics 50,757 (Oct,1964).
- [0556] 22. W. G. Hill, A. Robertson, Genetics 60,615 (Nov,1968).
- [0557] 23. C. N. Rotimi 等人, Ann Epidemiol 11,51 (Jan,2001).
- [0558] 24. C. Prunier, B. A. Heccevar, P. H. Howe, Growth Factors 22,141 (Sep,2004).
- [0559] 25. J. Huelsken, W. Birchmeier, Curr Opin Genet Dev 11,547 (Oct,2001).
- [0560] 26. A. Duval 等人, Cancer Res 60,3872 (Jul 15,2000).
- [0561] 27. S. S. Fajans, G. I. Bell, K. S. Polonsky, N Engl J Med 345,971 (Sep 27,2001).
- [0562] 28. C. Wolfrum, E. Asilmaz, E. Luca, J. M. Friedman, M. Stoffel, Nature 432, 1027 (Dec 23,2004).
- [0563] 29. J. Nakae 等人, Nat Genet 32,245 (Oct,2002).
- [0564] 30. J. A. Noble 等人, Diabetes 52,1579 (Jun,2003).
- [0565] 31. A. Duval 等人, Cancer Res 59,4213 (Sep 1,1999).
- [0566] 32. A. Duval 等人, Oncogene 18,6806 (Nov 18,1999).
- [0567] 33. H. R. Chang 等人, Cancer Lett (May 16,2005).
- [0568] 34. N. A. Wong, M. Pignatelli, Am J Pathol 160,389 (Feb,2002).
- [0569] 35. M. Lepourcelet 等人, Cancer Cell 5,91 (Jan,2004).
- [0570] 36. V. Korinek 等人, Nat Genet 19,379 (Aug,1998).

序列表

<110> 解码遗传学私营有限责任公司 (DECODE GENETICS EHF.)

<120> 作为 2 型糖尿病风险诊断标记物的 TCF7L2 基因的遗传变异体 (GENETIC VARIANTS IN THE TCF7L2GENE

AS DIAGNOSTIC MARKERS FOR

RISK OF TYPE 2 DIABETES MELLITUS)

<130>SCT075777-47

<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

cttgtgtagg aactcacgct ttgtttattc agcaatcatt cctccagaaa taaccttaat	60
agcaacaaga aaaaagaata ggtgtttttt gagctctatc tgccagtttc tctatatatg	120
gacattatat attgcaacat aacactcaca atgcctttta acatcatccc cgttatacag	180
ataagaaaac agaatttcaa agaaggtagg ggacttgccc agggatacat agctagcaag	240
tggcagcgct ggattgagtc tgggccttgt ctgaggctcg ggtcctgtca tgctctgcgg	300
ttgctatgtt gacatgcaaa gggagaggca gctgctggga gtctaggtgg gtttctcttt	360
gagaatgcta acgtgaaccc tcaagggtgaa tcagaatcct tttgcaagtg aataatcaga	420
tgtaggttcc tgtgtctccc tgtaaaatga aagcctcttt tttccaaggt ccagtataga	480
cctgaagctg ggttactctg gaatttccct ctctggctgg agtgactgag gccttgcacg	540
tgacattggt gaggactcgc agcctcaggt ctggcttccc ttagcaaccc ccctttcctg	600
tctctgcctc tggagttcac cattaaaaaa aaaaaagaa aaaaagccaa aacactttat	660
aaagttacat gctgggtttc ttctatgtcc tagaaactgt cttaattcat cttccccctt	720
actcttatat gagcaggaag aaaaaaaaaat tgctagtcaa tgctaataat tatggcatgt	780
aatgtaattg gaagtgtttc actgacatgc tcatgagagt ttgcggcttc atcttcaggc	840

tgggatgtag	cactagactt	gccttgagtg	tctgcacaag	cctttgatgc	aggtagacca	900
tattataaat	aggcgcttg	ctatggtgag	gatggcagtc	cttgcttgct	gtgggtaacc	960
ttttctacct	tctcggacac	tgttttaaaa	cacagcagcg	tgatagcatt	tcattttaatt	1020
tggaccaagg	tggggtagat	gaaatgttga	gatttagatc	taaaatgttg	ttgtggtgtt	1080
tcaggggggtt	ctggctcacc	tagtactatg	gaagattttg	cagattgggc	ttcctcatga	1140
tttattttaga	aatagatttt	ctaatagatg	gggtgagggg	agggtggtgg	gcagaaggct	1200
gggctttctt	ctcttcccc	tectccttcc	attgagcgct	tctgcgaatg	tgttggcttt	1260
gatgccccag	gagctcatac	agtgaatagg	aagttcaggt	tggcacgttg	cagaaatgat	1320
tattcctgggt	agtacgtttc	ccattactgt	taataatata	aagacaattg	cctgcctctc	1380
aggactcctg	cacgtggcta	cagtcatttc	ttcatggaat	tagacacata	gcagtgggga	1440
ccaggagtgt	tttatttagt	attgtcctcc	tgcaagtttc	cagggtatct	cagcttagac	1500
acatgaatta	ttttttcctg	ttgcttggag	ggtatacttt	taattatatt	cattcaataa	1560
cagagcagtt	caggttttga	aaatatTTTT	tctcccccaa	ccttttcccc	agcatacatc	1620
cccgctccgt	aagtttcttg	gcagagacaa	tctcaggaa	ctaaagggtg	ctaaaaaatt	1680
agctagttag	ccaggcgcat	gactcatgcc	agtaatccca	gcactttggg	aggctgaggt	1740
gggtggatcg	cttgagccca	gaaattcgag	accagcctag	acaacatggc	aaaaccctgt	1800
ctctacaaac	aaaacaaaat	ctagctgggc	atggtgggtg	atgcctgtag	tcccagctac	1860
tggggaggct	gaggtgggcg	ggcgattgag	ctcaggaggt	ccaggctgca	gtgagccgtg	1920
attgtgccac	tgcactgcag	cctggatgac	tgagtgggac	cctgtctcaa	taataaataa	1980
ataaataaat	aaaaaataaa	aaaaattagc	tagccaagct	gcttataggt	cttttacatg	2040
gccaaagccac	tttctcacct	ttaaaatggg	aataacgttt	ccgtactcat	ctcaatgggt	2100
tttgagtgcc	aagacagacc	gtttgatgga	agccctcttg	ggagaaaaat	gtacccaag	2160
acaggctttt	caattggaga	ctgatccatt	ggtgttttgg	tcagttggtg	ttgaaatccc	2220
tatttttcca	gtcaggact	gcctctctcc	ctggaactct	tcccagaggtg	agttctgcag	2280
ccttccttgg	gaactctcag	cctctggatc	ccttcttgcc	agggtggagt	gacatgccaa	2340
agttgtgggc	cagactcgga	ctgcctggct	tgtctcagca	cctttgggga	cccacttccc	2400
ctctctggga	actggggaag	ctaacagaga	tcttgctagg	ggggtggaat	cctgtatcca	2460
tgtgaggttg	tacccccagg	ctcctgagt	gtttgaaagt	ggggaaccct	ggccgggcgc	2520
ggtggctcat	gcctataatc	ccagcacctt	gggaggctga	ggcgggcgga	tcacaaggtc	2580
aggagatcga	aaccatcctg	gctaacacga	tgaaaccccg	tctctatgtg	cgtgggtggt	2640
ggcacctgta	gtcccagctg	ctcgggagtc	tgaggcagga	gaatggcgtg	aacccgggag	2700
gcggagcttg	cagttagccg	agatcgcccc	actgcactcc	agcctgggcg	acagagcgag	2760
actccatctc	aaaaaaaaaa	agaaagaaaa	aaaaagaaag	tggggaaccc	ctcccccagg	2820
atgagaagag	ccatgggggtg	agtctctgcc	accgccaagg	ggagtcaggc	tcagaggtg	2880
ctacagggac	agccagctct	ctttagatgg	tccccacat	ctagtcaggg	cttgttacat	2940
atggagcaga	gacagcgcag	gctgctgctg	ttttcctgga	gaaggccctt	gtcgggtctgt	3000
tcagctgtag	ctgacctttc	ctccttgtgc	tttttgggga	gggagccttg	gaaggagtag	3060
ggcacgtggg	gcactctgct	tcccggcccc	acactggcga	acctatggat	tctgcctctg	3120
attcctgagg	aaacatcact	gtgaagggtg	aatgagccac	atacagaggt	ggctgttggg	3180

gccggggagg	ggtgaaacgc	ccccaggggtg	tacattgcac	caaaagccag	gctgcatata	3240
gacctcagga	tgggctggct	tttctattta	tttagaagta	tttccagagg	gtaacctcat	3300
tggctacaaa	gcatgtctga	acaagagctc	cgttggttcat	tcccagccct	gttaccctgg	3360
caggatgcag	actccaggcg	gcctgttgggt	caggccttgg	actcagagag	cagtgaagcc	3420
tgaggagggg	tggggggcag	aggcgtgagt	ggtctagggc	ctcagtcctt	ccaggacacc	3480
ccttgccaag	cgcagagaaa	gctctgcccc	tccgtccctt	caggcagtggt	gattggggcaa	3540
cctgggaagc	agtgaatgtg	cgtcggtagc	atagattcca	ttccgcacgc	caccctcgcc	3600
tccgcccccc	agccctggga	gggatgcatg	ccctccggga	gacacccaga	cccgacagag	3660
aggcctttgt	tggagctgga	ggtgagaatc	tgtgggcggt	gggattcctg	ggttcgagtt	3720
ccagctcact	gccaattgcc	cgagtgtctg	gcgaacatit	ctggaatcaa	aaggagtgc	3780
gcctgcccag	cagggcctac	gggagccgga	ggctgcaggg	tgctaagatt	gcgttatctt	3840
taccaagtgc	ccggagctcc	tgggagggaa	gagagagtcc	taggactcag	gataggaggt	3900
ggttggagtt	tctcagaggaa	gactccatgc	tttgggttctg	gccccctggaa	acccctcctg	3960
aggactggac	ctccaagcag	acccccctctg	tgactccgga	atgcagtggt	actctcttat	4020
atTTTTcttt	ctTTTTTTTT	TTTTgagacg	gagtctcact	ctgtcaccca	ggctggagtg	4080
cagtggcacg	atctcggtc	actgcaacct	ccgccccctcg	agttcaagcg	attctcctgc	4140
ctcagcctcc	caagtagctg	ggattacagg	tgccctgacac	cgcgccctggc	taattttttg	4200
tatttttagt	agagatgggg	ttttaccatc	ttggccaggc	tggctcttgaa	ctcctgacct	4260
cataatccac	ccgcctcggc	ctccgaaagt	gctgggatca	caggcgtgag	ccaccgcacc	4320
cggccactgt	cttgtatttc	taacgtcccc	ctgacttttc	tgatcatgta	attcttaact	4380
ttctcaaaac	tgagatttgt	cacgtgtcct	ctccccactc	catttttgtga	atcagagtct	4440
tccaggggca	ggacctggag	aatgggtctt	tattaacaca	catgtgaaaa	tgcttttgcc	4500
agcaaggcgc	ggtggctcat	gcatgtaatc	ccggcacttt	gggaggccga	ggcaggcgga	4560
tcacttgagg	tcagcctggc	caacatggta	aaacctgtc	tctactaaaa	atacaaaaat	4620
tagctgggtg	tggctgtggg	cacctgtagt	cccacctact	cgagaggctg	aggcatgaga	4680
atcactggaa	cccaggaggt	agaagttgca	gtgagccgag	atcacaccac	tggactccag	4740
cttgggtgat	agagttagac	tctgtctcaa	aaaaagaaaa	aaaaaagaaa	atgcttttgc	4800
catgggctgt	ctcctgtctc	tgttttgcat	tgggcctctg	tacctaggtt	gcaagattcc	4860
tcaggggtgca	cctgggctta	tcgttatctg	taagttatcc	cagcaagcac	ttaaaaacaca	4920
gtgttggacg	atgaatcccc	tctacaagag	agggacaggg	caaaaacgac	acctcttgcc	4980
tcgcaagctg	tcttgggcca	aacctcaggt	ctattctttc	ttttttttga	aagtagtggc	5040
tgggcacggt	ggcttacgcc	tgtaatccta	gcactttggg	aggccaaggc	gggcggatct	5100
tgaggtcagg	agttcgagac	cagcttggcc	aacatggtaa	aactccatct	ctactaaaaa	5160
tacaaaaatt	agctgggcgt	ggtggcgcat	gcctgttagac	ccagctactc	aggaggtgga	5220
ggcaggagaa	tcacttgaac	ctgagaggca	gaggttgcag	ttagctgaga	ccatgccatt	5280
gcactccage	ctgggcggca	gagcgagact	ctgtctcaaa	aaaaaaaaaa	aagaaagtag	5340
cagctctact	gagatattta	gaaaccataa	aatccaccta	tttgaggtgt	acaattgagt	5400
gattttctgt	atagtcacag	atctgtgcag	tcattccacac	cctctaactc	caggacattt	5460
tcctcacccc	cgaggagaaa	cctcccttac	ccattagcag	tcactcctca	tttctctctc	5520

ccccagcccc	tggcaatcac	tgtggatttg	cctgttcttg	acatttcata	taaattggtat	5580
cataaaatct	atgggctttt	gtgtctgtct	gctttcactt	agcatacggg	tctcaagggt	5640
catccagtat	tgtagcatct	atcagtatgt	cattcccttt	tatggccaaa	taatatattta	5700
ttgtatggat	agacattttg	tttattcatt	tatctgtttt	tggttattat	gagtaacact	5760
actatgaaca	ttttgcacaa	atttttgtat	tgacatgttt	tcattttctcc	tgggtatagt	5820
cctatgagtg	gaattgctgg	gtcatataat	aaataactgt	ttaacatttt	ggggagctgc	5880
caaactttta	aaaccttggg	ttctgtgatg	taccagttgt	gttaggcagc	acagcaaaat	5940
gtgacttttg	attgccagaa	acaatattta	aaaagtgggt	ataaaaagtg	gtttgggagg	6000
ctgaggcagg	aggatcactt	gagcccagga	gtttgagacc	agcctgggca	acatagtgag	6060
accctgttaa	aaaaaaagaa	ggccaggcac	agtggctcat	gcctgtaatc	ccagcacttt	6120
gggagactga	ggcgagcaga	tcacctaagg	tcaggagttc	cagaccagcc	tggccaacat	6180
ggcgaaaccc	catctctact	aaaaatacaa	aaattagcca	ggcctgggtg	tgggcgcctg	6240
taatcccagc	tactcaggag	gcttgaggca	ggagaatcgc	ttgaacctgg	gagactgagg	6300
ttgcagtgag	cggagatcat	gccattgcac	tccagcctgg	gcaacaagag	cgaactgtg	6360
tctcaaaaca	aatgaaaaga	aaaggctgtc	atgttagatc	caccctcctc	ctcaggggaa	6420
cccctgggct	gctctctggg	tagagatggg	aaccaggcc	tcgggccagt	gagtgggaagg	6480
aaactttggg	atgattgact	tgggactggg	ctagagggtga	agaatctccc	agtaggcaaa	6540
gttcggcctt	acgttttttt	gtttcaagca	aaccacatca	ttaccacag	aggccattgg	6600
tgagatatatt	gtaagtctcc	tgacagtggc	tggagtctgt	tgcttggttg	ttgtttctct	6660
gtctcagccc	tggagatggg	agtgaccacc	tgctctctct	ggacagaggc	tgtccacgtt	6720
catgcaattc	cttggacacc	ggtgggtgcag	cgggaggcgt	aactgggagt	gggagaccct	6780
gaactgtgcc	ggttcttgca	gagtatcact	gtgacttcag	gcgagtcacc	ccacatcagg	6840
cagctcagaa	caagggattg	atctagaagg	acctttcacc	tgggctattc	tgtgactcaa	6900
attatcttct	cctaagccca	ctactgcctg	gtgtgtttgt	taaattagcc	taaagggtcat	6960
tccctcggag	aggccctctg	ggaaacctcc	ctttcctgag	agtcactgct	tgctggcgcc	7020
tgccccctggg	gttccttcag	agtcgtgate	atgccctggc	ctcttccttt	atttggcagt	7080
cccttccttt	ccccatccct	gatgagggtg	gggagcatct	gtctgcagct	tcactttcat	7140
tgtctagggg	ctccagaaat	atctgtgagt	aaataagtta	tttaatcttt	gcctcaaatt	7200
tccagtgact	gtagggatat	agctgtgagc	ctctaggagc	tgagattttt	taaatttccc	7260
acttaaacat	ttattttaaaa	attttgtgtc	cagcatggac	taaggacttt	acattcatta	7320
actcatttac	agcttgatcc	tatgcggtgg	gcattcattt	acagaggatc	ccattttaca	7380
ggtgaggaag	aggccagcta	gggggtgcagc	ctaggttagt	attctagagc	tcacacaggt	7440
gtgttgtccc	cagtgaaga	ataagcaaag	aagtgaatgt	tgtgcattga	gaaaaatgac	7500
tctcggagga	ggatgagcct	ctcgatgatg	gcgaccgaag	tgatatgggg	cccttgtaaa	7560
gggtctctat	tatggcatca	agaaaagatg	ctgctttcgg	tgatgcccga	ggagagcctc	7620
aatatttttac	atgggaaacc	taaaaaaggg	gccatgttgt	ggtctctgca	cctaagatac	7680
taaaggaaat	attttatgga	gagatgcaac	atgtcaggcc	ttggagggaa	accccaggat	7740
ccagatgggt	gcactctcaa	accaggggccc	ccctcacctt	ggccttcagc	atttagtggt	7800
ggaaccaata	gcataagctt	tggtcaggac	ctttgatgga	agccacagtg	ctcatttagt	7860

accacggttg	actaccttct	ctctcctaag	ctgacttctg	gagggcacct	gggatttccg	7920
gccagtgate	agtgtctggtg	aagcctgaag	gccaatgtgt	aggttttagct	gttcagtcag	7980
aacccaaaag	gggccaaga	gatggtttcc	ttcaacctcc	actgagggaa	gtgaaagtca	8040
tggttcgtta	aaaggtgag	ctgggaccag	agtctagggt	tctagagggtg	ggaatttcta	8100
cagctttggg	ggaccttgca	agggcatttg	ctcttctggg	actgcaggga	gactgtgctt	8160
ctcagagatg	ttagcatttg	gcttggggag	agagaggaaa	ggagagggttc	atgctccgcc	8220
atgatggtgg	aaagtgatgt	tggtgtggtg	aggagctgag	ctgaattcta	agtggttcca	8280
gggaattaac	aatgttctctg	cccaagtgtc	ctgttccccc	acaaactaat	gaggcagcag	8340
gtgtctgaag	agaaacattg	cagaatgtct	gccaggggtt	ttatggttaa	ttttcctcca	8400
ttatgagggt	tgactcagcc	ttgggtatta	gatgtctttg	agaatccagg	gttcaaatac	8460
cacagctggt	agaatgtttc	tcaacttgga	gccaatctcc	atctactgaa	ggtacgctgg	8520
tttagacaga	caacagggac	atcagcattt	taaaaagcgg	tggaaaaagt	ttgcttgtct	8580
tgattggagc	catgacattt	tattttgaaa	tttcaaataa	catgaaggga	ggtttggagc	8640
ggtttttggt	ttatccaaag	ggcagtggat	tgaaggctga	gaaacaccag	gctgaatggg	8700
agaggggttg	gggtccccct	gtgagatagt	gaaacaatgg	tagtgccatc	caatgatagg	8760
cacttttctg	tcattcagaa	gcagaaaggg	ggccagaggc	ccattggcct	tactgggcag	8820
taagctgtag	agctgttgcc	ttttcgtgaa	agggttgaca	ccaaccttct	cccccaggaa	8880
gagtgaccag	ggacctgagg	ggcatggtcg	agcagatgac	agcctttgta	aaacatctcc	8940
ctggctctcat	cagcgatatt	cgctctgcct	tccttctgag	taatttccat	cttaggactg	9000
gagtcagggtg	gagcaagatt	ccatgttggg	ttctgttggg	cctagagtgt	cacactgaga	9060
cctaatttca	tactttatga	attctagtac	tgctctcgaa	ggtaagagcc	gtcctctttg	9120
gctgaagggt	tttgccctgca	accttgcat	gtaatccagt	gacacctgac	gtatctgtaa	9180
atttcttcaa	atttctaagt	gtattacaac	cccgtgtgca	aaagatgatt	aattaattgc	9240
cttgacagta	aaacaaaaaa	caaaaaaaag	gtgtgggggt	atatggtatc	cctgatttac	9300
tatagaagat	gcagagagtg	aaggagatg	aggtggggag	gaggggcccc	ggttctggtc	9360
ctactttttt	tttttttttt	ctaaagagat	ggagtcttac	catgttggcc	agtctaggct	9420
tgaactcctg	gcctcaagag	gtgctctcac	ctcagcctcc	caaagtgtctg	ggattatagg	9480
cgtgagccac	cgagtttagc	ccaggttctg	tttcttgctt	agtcactttc	tgtttgaaca	9540
aaattggaat	ttcctttttg	gatctgtttc	tttaattgta	aattgaaatcg	gactaaaaacc	9600
tttccaattt	tttcacatgt	gaagacatac	acaaaagt	tattggaggg	ttgcacatgt	9660
gaaagaaaaa	gggagaaaagc	aggattgagc	agggggagcc	gtcagatggt	aatgcagatg	9720
tgatgagatc	tctgccggac	caaagagaag	attccttttt	aaatgggtgac	aaattcatgg	9780
gctttctctg	cctcaaaacc	tagcacagct	gttatttact	gaacaattag	agagctaagc	9840
actttttaga	tactatataa	tttaattgcc	gtatgaggca	cccttagttt	tcagacgaga	9900
aaccacagtt	acagggaagg	caagtaactt	agtcaatgtc	agataactag	gaaaaggtta	9960
gaggggcccct	ggacacaggc	ctgtgtgact	gagaagcttg	ggcacttcac	tgttacattt	10020
catctcttcg	ctataaacat	tttagctttt	tgtgtttgct	gactggcaac	aatacatagt	10080
gaaagtctta	ataatttgta	atgcttttgc	atgtctttgt	atttttcttg	gttatcacat	10140
cacatcaaat	taagatactg	atcagcagtg	tgagagggtta	tttttccatg	tcctcttcat	10200

tagtgttage	ttgtggatgg	atttgaggct	ctctgtgctt	tccccccagc	aaagtgaata	10260
ccagactttc	ctattaaaaa	aagtatttta	tttttcagag	acagggtctc	attctgtctc	10320
ccaggctgga	gtgcagtggc	acaatcatag	cccactgcag	cctccaactc	ttgggttcaa	10380
atgatcctcc	tgcctcagcc	tctcttaagc	agtgcctttc	cccattctca	tgggactttc	10440
caatccatga	gatactttgc	tgcaggggaag	ccctgtctgt	ccaggcctgt	gtaatagacg	10500
acttcacatg	gtcctgtgtt	gttgtttgcc	ttctgtgtgg	ctaagtttcc	atgacctggg	10560
ggcttgggaag	ccccatccct	gatttgtggg	agaggcaggg	aggcaccttg	tagcgcacta	10620
ggcgttgggc	ctgaacaagt	ctgtgtgctt	ccaatgtctt	tgtggggagg	tttacgagtc	10680
cttcttatta	tataatagta	tcttgtctta	gcttggtgcc	tttcttctca	gaagcttgag	10740
gcactctgca	gataccatct	caatttgctt	tctggggagga	ggagaggaag	ctacccaaaa	10800
gatgaagttc	tctgtgaggg	gcttgaacac	aggttgatag	cgttgctggg	tagttattct	10860
catggtgtgg	atgaaaaatg	gaatacgtcg	aaatttcagt	tactcgtcac	aaaaataagg	10920
cgtatgtaga	aaacatcctg	ggctaagggt	ttgcatgctt	ctagaacttc	ctgttactta	10980
atggctgttg	agtataaacc	tcgggaacag	tggggatcct	tggagacccc	aaataacttg	11040
tatttgtggg	tactcctgtc	ttgtctatca	ataccctgtg	ctatatcgtg	ttagaactag	11100
gacacacaga	ctggattcag	aagctggcct	ggggtttagg	agaacatggg	acctaactct	11160
ggccatctcg	atttacctcc	tggatcttgt	tttctcatct	gtaaaatgaa	ttgggggtgtg	11220
gactgtttat	ggcctgtagg	atgctagccc	tgagaatttt	ctccagatat	tctacggtta	11280
agtaatttta	ggggacactg	tctaagcagt	tgcctcttgg	agaatgaaga	tgttcattag	11340
gatattgaag	gctctgagaa	gtcctaaagt	taaagaaaat	ctgcaatgtt	ctttgtggga	11400
ccgaataatg	caacctggga	aatgagggat	tagatgacac	ttgagtagcc	ttccagatct	11460
gagacgagtc	tcactctgtt	tgtttactcc	atctgtgatg	ggtgtaggca	ccatcttggg	11520
gagcaagctg	tgatagagag	ggaacaatac	cttggttaatg	tttgtctaat	tcactaccca	11580
ggtgcatggt	agtgaattag	acactacttt	gtaggttctg	gagggaagaa	gaaaagacga	11640
gacctgcctg	gactggggct	tgagaccact	gtcaaataca	agtacagttg	tacaactggg	11700
agggagtggg	tcatagtatg	gccggtcttt	ttaaagggtga	ggaattctta	ggcccagaaa	11760
ggcaaagtga	cagatcctgg	atttaaccag	cagcccagat	ttgaggccta	gcacatagca	11820
aagcaccata	gctattcaat	agctgccaa	tgggagtgtt	gatgatggct	ttcctggaca	11880
gcgaaagcag	tgatgtttgc	ttaggatggc	ctttggcagt	gctgctgtta	tccttaccac	11940
tggcaagcca	tctcacgggc	ccggagggga	gggcaaggaa	tcctaattct	gtgagaaggc	12000
tctgggtaca	tgagtgtgag	atatggatac	cctaggctct	gccccgaag	acagtggcat	12060
cggatttact	gcactattcc	agtcggacag	gcaccttaat	ttttctcttt	ctgggtgttt	12120
gatatggttg	ggctctat	cttctctcc	aaaccccgct	agggcatttc	ccccaccctt	12180
cacttcccgg	ccttccactg	cagtctctaa	ggattctgct	tcattctttat	gtgtgaacag	12240
ggttttgaca	aacatgatta	actgggtatt	tttgggaaggc	tcaggaggaa	cgcagagtgc	12300
tccggagggc	aggcctggag	tcaggaatgc	ttcctgcaac	ctgttcgtgc	agtgagectg	12360
tcttctctgc	cctgcccttg	gctggggaa	gtgctggctt	ggagggcagg	agagtgcag	12420
gcggtttgag	aactccgggc	tctcccgtct	tcggatggct	cctgtgaaag	cagggcctga	12480
aacttttate	gtcactgctg	cagggtgaaag	actttcattt	ggctgtagt	gtccaacaaa	12540

gagtatttta	tttatgtgtt	tccaagccct	taaaaattct	tttagggcac	atcagtgggg	12600
agttaataga	aactttgaaa	taagaaaaat	gcctgcaggg	taagtagaac	cccagccagc	12660
cagctccgag	ttctgtgctg	ttagctggta	ggttggttct	cagagaagtg	gctggctggc	12720
tggtttacgg	agcccacatc	tctaattgcct	tagtgttcaa	tcattaagtg	gatttttttt	12780
tttcccttct	cttcttttgg	tttggaggga	ggactactct	aaactttact	cagggcaggg	12840
tagctcctga	aagggtcccc	taacctttct	ggtttatgac	acaagaaaag	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	taatttccta	agaattaaa	agcccccttg	13140
aaaccagca	ctaccttgtc	tctgattatc	agcataggca	ggaagggtct	ttaaggtctg	13200
agcccagctg	tttagaggct	acgagacgtg	aggcaaatcc	tggtatctct	ctttgggcct	13260
cagtttcttc	atctgtgaaa	tggcacagta	ctaccctcca	ccaaggatga	tgatgagaat	13320
taaatgggat	gacaggtttc	atccccagct	cctgttctta	ggaaggaaaa	actgtgactt	13380
atgaagcctg	taggttgtgt	tcaggtttgt	atgaggcctc	ggacttcata	caaaggtatc	13440
aaagtggcaa	accctgatcc	agatgttttc	agttcagtc	gctggctcct	gagcctgttg	13500
tgtgccagat	atcctgacca	aagaagctag	atgggagctg	ctgtgttggt	ccttggggct	13560
gctggatgca	agttgttttag	gtcggcggtt	ttcaaattgct	ggtgattttg	ctcaccagag	13620
gacatttggc	aatgtctaga	gacatttagc	atggccagtc	attgggaggt	actcctggca	13680
tctcgtgggc	agaggctaag	gatgctattg	aacatcctgc	aatgcccagg	acagccccct	13740
gtgacaggag	tcatccagcc	caacatgtca	ctagtgtctg	agtggagaag	ccctggctgt	13800
gtgtgggggt	gtgtgtgtgt	cctcttctac	atttga taag	gtaactcaca	cttgctgccc	13860
ccatgatcgc	tgtgggggat	gcttatctat	gccccagtcc	tggtgttggt	tgatgggaac	13920
atcaagattc	aggcaagatg	gaaaatagcc	cttagaacta	gcaggaaaag	aatctccttt	13980
catttgtcta	gaggttctgt	taaagtgcct	ttgcttctat	tttgagactt	gttcttaaaa	14040
aaaatgcgga	tatgaaagaa	aataaaaacc	acattatccc	tccacttttt	cttgaggagg	14100
gatgtgttga	agaagtcaaa	gttcaccatc	ccttttagata	gaatcatttt	gaacaatttc	14160
atatgtcaat	acattttgct	catctctaaa	tttcatttta	gagcctgtgg	tgttctgtgc	14220
atggatatgt	gtgcgtgtat	gcacacaaaa	ataaaaggaa	atattttattc	ttatgaataa	14280
gtatagaaat	aaattaattt	ttggaatctc	aaactatcag	agacttatgt	aataaccaga	14340
ggcaggcctg	attatgtatg	ggcaaagcat	ttgtgaacaa	tgtctccatt	gtataacata	14400
caaaacaagc	ttttcttcca	cattggatat	gcaagtcggc	cttctccaat	aagggcctgt	14460
ctctttccaa	ctccccccac	ctc ccacctt	tgagcaaaca	ttattttattg	tggtgatgt	14520
gtgatcaggt	cttgatttgg	ggcctctttt	tgatgccttc	tctttgtggg	atctcaccca	14580
cgtgccccctg	gagacccttt	ggctgccagg	gcctttgttt	cccagccacc	catgtggtgc	14640
cagtagtgtc	tgctttgtag	cgagctgtcc	ccagagcctc	agcatggctt	ggggatggtc	14700
tctgaggttg	ggcttggatc	cctcccactt	ttgggctcag	aaagaatgac	tgccctctat	14760
ttccctgtcc	ctgccctctc	ttatcctgtt	tcccageccg	catcatgtta	tctttgcttc	14820
ttgtaactta	ccaaacgatt	tatgggcaag	taggggaggt	gaagagggaa	ctcatctatc	14880

aagataacct	actttgtgcc	aaccactgag	catagcaatt	gtcccttctc	cagccctctg	14940
aggaccgtgg	atgggattct	catgaaatga	aacagggtgag	gaacttttct	tttaggggaac	15000
ttgcttgagg	tcccacaggc	agcgagtatc	aatcaacgtc	aggatctgag	ccctgttctg	15060
ttcggctgaa	aatatactcc	ctgagatggt	gtaggccacc	atggctttca	gcaggctctg	15120
tgcttggtgg	aaggaagctg	gaagctgtgt	acacaccac	ggggaacagg	gaccatagag	15180
gagcaccttt	tgagtgcaga	acctggcgaa	acatacacct	ttagagggat	tttaggtacc	15240
cttgaggctg	ggagaatcaa	gcagagctaa	gtttccatt	ggggtgtcac	agactgaaga	15300
aacagagccc	taggtagcac	agggaagtgt	attgccagt	atcagttagt	ttggctttaa	15360
tgactgagaa	gagattccac	cagttcattg	aagagagggc	ggacttttta	ttggaggaaa	15420
gaagagtgcc	tgtaagtaga	gaagtctccg	gggtgtagt	ctgtttgggg	caggaagaac	15480
agtgtgagcc	actgtggaga	gaaagcccaa	agagtcttgg	cagggcagg	g agtaggatgg	15540
atttgaagcc	agaggaagta	tgggtctct	gtagactcca	ggcaagccat	gttaatat	15600
taggaagccg	tgatggagct	gcagatgggt	gtggaagtta	aagtttaact	gttcattcac	15660
cagtccttcc	cctggagaat	gtgcagcacg	tggacagtgg	aactttaagg	tccttggctt	15720
gtatttcaca	cccaagagat	gaataggtcc	aggtatgtca	tagaccagac	taatgaaata	15780
acaaatttct	tttcaaaaat	tttacttttt	gtaggaaagc	ttctctgtct	ggcatttttc	15840
ttctcccagt	tgtgactcaa	tcttaaacgt	cttcagacaa	ttagcataaa	atttcccaca	15900
gtgaattgac	gtatactttt	gagggttcca	tttctttttt	atTTTTTTTT	tcttttgaga	15960
tggagtttct	cgtcaccag	gttgagtg	aatgggtcca	tcttggctcg	ctgcaacctc	16020
cgcctcccgg	gttcaagcga	ttctcctgcc	tcagcctcct	gagtagctgg	gatgtcaggc	16080
acccgccacc	atgcccggt	aattttcgtg	tgttttagta	gagatgggg	tccaccgtgt	16140
tggccaggct	ggtcacaaac	tcccagctc	aggcaatccg	cccgcctcgg	cctcccaaag	16200
gctgtgatta	caggtgtgag	ccactgtgcc	cagcctaggg	ttccatttct	taaccctcc	16260
ttctgatgcc	tcagaaagtc	ttgctctgta	agcctcttgt	agctgcctcg	gttcagggga	16320
agggggaggc	ttttgtttta	ggaccgtcca	gaccatagac	acatttcctg	gcacctagca	16380
cgtgttgggt	caaacaggaa	tgatgaatgc	atgcatgaat	gaggttctta	gcgctgaaga	16440
cgggtgtcata	ggtggtctac	cacgccgcct	gatcattcca	atggccatt	atgaatgtgt	16500
gtgctgcagg	gccctccac	gatcccgta	gcactgtgca	tgttgtgggg	aggtgctggg	16560
agaaagactg	ggtctcagaa	gatgggttag	aggtgggtcc	ttctctgctg	ctggctagca	16620
gggtagctgt	ggaggggtgc	ccatcttgc	tggctcttaa	tttctcact	gtaggcagg	16680
agcatgacct	ggctgaattc	taagtccctt	tctactctga	ggttcattgt	gggtgtgacc	16740
tgctgggctc	agctctggct	ttgggagaca	ccctctcccc	ttgatctcga	caaccctta	16800
gcagagccca	gtggctccta	cagtgcctg	agctgcttgc	ccgaaggatg	cggttgtggt	16860
tatctcacc	cctgccacc	tgtttgcgca	agggtttgag	attgtgtggc	ccctccttgt	16920
acttcgggg	gaggcttgct	ccagaaaggt	ggtctgcaaa	gggttggct	ggggggagg	16980
aggaagtcat	tctccaagt	ttgtcctca	tgttatccc	aaattgcttg	cctggaataa	17040
ggaaggaaag	aaaaaaaaat	actcttgagt	ggtttgggcc	aggattttag	ctgatggatc	17100
tggtagtcc	ctctgtcaga	tttgttttct	ttgaactgtc	tgggccggtc	acagtgtcat	17160
tgtttaaatg	tggaatgtag	gtgttctgtg	ttctgggaaa	taaaaaccaa	aactggtcca	17220

ggggatccac	agaggtaaga	aaagaacatt	ccaataggaa	tgtttcagaa	ccaggagggg	17280
aggagagaaa	aacggctctg	ttggctctct	agaggaagaa	cttgtagat	ttggggagag	17340
tcaggataaa	tttgacccta	agagtctctg	attcctttta	gagacttttc	ttataagaaa	17400
taaaatggaa	cttgggagag	gcggcaactt	gggaaacagc	acattctgcc	gtaatgaaag	17460
tcgtcccata	agaattttct	tatcccttta	gccaaatttc	tgtttctaaa	aggggaaaag	17520
gggctagaga	taggcttggt	tgttttctta	gttgaatctt	actttttgta	tttccagccc	17580
attctgcagg	gtaagaacaa	gcacagcccg	agggctcact	cagtgtgatg	ttctagagcc	17640
tggctctgcc	tcaatccctc	acgctggagg	atcaggcagc	aggggccagt	gatggatttt	17700
tttttcttcc	tttccctccc	tattaatatt	tactgaggta	taaattacag	caaagtgcgc	17760
agacctaggt	atctaggact	gtgaggtttt	cctgtgttac	ctgtgtaacc	acgaccaga	17820
tcaagataag	gaacttttct	ggcatctcag	aggctcttcc	tgctcccttt	cagactccgt	17880
ctcccagaag	gaactacttc	tgattcctat	agccatagac	tgaattttct	tttccaactt	17940
catatgcata	ggatcatcat	gggtgtttta	attttaattt	catgtctgct	tgccactccc	18000
aaatggaaat	gtgttggcat	ctctggatgt	ttcttcataa	gaaacatgcc	ctgtggggca	18060
aagcccagga	cagggctgtg	ctgctgctgg	aagtcctgtg	cagctggcca	gcctctgctc	18120
accctccgg	ccacgtggc	actttcagct	tctccagcct	cctgcccttc	ccacttccag	18180
tcctgcacc	t gctgtcctca	ctgatgcacc	tgcccttttc	cttccgtcct	ttatgtggca	18240
cacccttaag	ggagacatct	tcctgtctgt	gttttgcacc	ctcttaaaac	tacattcctt	18300
tcctttcagc	attggcatct	ctgtccttgt	gtattacctg	ggatgactat	tcagttaaca	18360
aatgctttct	tcctaggctg	tgagcccaag	tttgttggat	gattggatgg	gggcacgttg	18420
tgtgagagaa	ggatcatggg	gtagcatctg	gctctcttag	aggtgtgtgg	gggcgtgtga	18480
tgccctgcaa	ggcgctttcg	ttctgggggg	ttctgtgtgt	ttgaagcact	tgggttgtgt	18540
gtccctgagg	cctccgtcac	gggcaacctc	attccttctc	tagcctccat	cccctgcccc	18600
ctgcccaccc	caggcctctg	gagctggctc	cctttcctgc	tcactctctt	ttggccagga	18660
ttttaacata	tatcacaggc	tggtaggcta	agagcttggg	acttcccctc	accacactca	18720
aagcctttga	tcttttgctt	tggaggtaac	atcaaaagga	aggctgagga	agacagccag	18780
gctgtgaagt	tcaacgttca	agttaatagc	ttgactgaag	gttgtgtgc	gttgtggcag	18840
catcaccgag	gctggagtaa	acagagtgat	tctgccacat	tttcttgaa	atgcaccca	18900
atattggaag	agggtcttct	ttacattcgg	aatgaattca	ggctgtagtc	agagctgctt	18960
ttccctttcc	ccattttcct	tggaagtgtg	aaaacttggg	ggagaagatg	tttgtaggag	19020
ggcatgatga	ggggtagagg	aagcccaaag	agaggatctg	gggaggggaa	gccccatggg	19080
atgagactct	gaagttatcc	ttgccccgat	tccgggactt	gctatctgcc	tgcccttttg	19140
cgtgggtgtct	ctgtgccctt	gac tgttctt	gatttagcga	ggtgtttctg	aattctgatg	19200
gaattcaaag	aagcctgggc	aggcaggcag	cttgacttgg	ggcttgggga	agcgtgcagc	19260
ccagacatag	cagcgatgag	agggcctcag	ggctgagggc	tgagatgaga	atttcacac	19320
atgcaaaagt	gaaagcgacc	catcgtcttc	tccacttgat	ctcttgctga	gctttgcaga	19380
cactttgggt	gttgtttaat	ttaacatttt	ctgcaatgct	ccttttttca	gattttcatc	19440
caaagctctg	tatgagaggt	tttcaaacc	attttggccc	tgattctatt	tggcatacga	19500
ttcaactctg	gggatgggtca	tcttccccac	acctgcgttg	ggtacctttt	tgggtgtatgc	19560

tcagagcatc	cttggacatc	ttcctggtea	gtgtccagca	tcgtgaagct	gccctttagc	19620
ctctcagtgc	ccccagatac	acctgtctct	ctgcgtagcg	gcactcagcg	tcacctttct	19680
gtggggctct	gagaccctga	tgatatcagc	actatgctgc	cagaattccc	cttggattct	19740
ttagtgtggc	ttctcaagca	tccttatcgc	ctataacgcc	ttcatggttt	ttggcataac	19800
tgtatactac	ctgtgctatt	atttatttga	tgcatc aaa	catttgattc	atttatttaa	19860
actcagtctc	actgtaatcc	ttaattaaca	cctgtgaaat	tataggtttg	atgtgctact	19920
tatttattta	ttttttaata	cacattagta	taatcccgta	acggctaaag	taacactttg	19980
tactgcctaa	aacctatgct	gggagcgcca	cagtttgaga	aagtgccttag	ccttcctttc	20040
cctccttttag	tgacttgtgg	tttggggcat	ctgttgactc	ctagggctcc	cttgttcate	20100
tttctgttcc	taagctcagg	gattagttgc	tcaaccagg	tgtggcctca	aaattctgct	20160
catggaatag	cctcaggctt	ctataaatct	catctttttt	gttttgtttt	gttttgtttt	20220
ttgagactga	gtcttgctct	gttgcccagg	ctggagcaca	gtggcgcaat	ccactgtaac	20280
cattgcgttc	tgggttcaag	cgatcctccc	atctcagact	ccaagtagc	tgggactgta	20340
ggctgggtacc	accaggcccc	actaatTTTT	aaattttttg	aagagatggg	gtctcactat	20400
attgcccatg	ccggaagtct	agttttatag	tgatgagaat	tcatctgggg	tccaaggggc	20460
cctcctgtgt	tgcttcctgt	gtccccctct	aaataaagat	actccttcca	agttgtcctg	20520
attttcaggt	catcaccatt	ttttgagctg	gatggggaag	ttggcctgga	gcagccttcc	20580
ctgtctccga	gttgcatcac	ctcctgagag	gtctcagcaa	atcactgcca	tctcttgatc	20640
agagttgctg	gcaagagtcc	tctgtggttc	taggttttca	gccctggaga	ctctcgcctg	20700
cattcattat	acatgtcctt	ttgggtgcctt	gttgaaaggc	atctcctgcc	accgaagggt	20760
gtgggcttct	ggaaattctc	agaaaacaca	atatgccagc	ctccagggat	gggtctccaa	20820
agcttcagga	acatacctg	gggtgttgag	gaaacacca	ccttaaaatg	ttcctcaagg	20880
gggaatgtta	ctgcttgccc	taaccctctt	gagctgatgc	tcacatgacg	tccctgagat	20940
gggcttcttt	tttgcccgta	cttaaagctg	ttaaaggcca	ttgtcaaatt	tgttttagctt	21000
ctcaattcat	gttccttaga	ggatggtaaa	ttaaagttag	cattcctgga	cagagccttt	21060
catacattga	agacaacccg	gtgagtctca	aggggagagg	taaggagagg	atgaaagggt	21120
ttctccaggc	ctgttcggca	gcatggactg	ttcttttagg	taattaaggg	agaccataaa	21180
agacaattgt	gtgagtccat	ttacctttca	cttgggggtc	ttaagtcttt	ggttgggctt	21240
ctttaaccct	gtgtgtcacc	cacgggctcc	tatgggtgct	gttttcattg	ttccgttate	21300
tagttggctg	gaacacacct	ttggggattg	gagaatggag	ttctgggggc	tttgggaact	21360
ttgagttttc	ctgcaatgtc	ctatagaagc	ttgagtctgt	gattcctggg	cagggccttc	21420
tcctagttga	gtgagattgg	tggggcaggg	cagccagtta	gggggtcatg	ggagcaggtg	21480
tggaaaagg	tatatgtctt	agtaattctt	tgtgacaatc	accctcattc	attgatatct	21540
tcttcctate	atgtatttagg	gcagtgggtc	ccccaatgtg	ctgcacatta	ggttcacctg	21600
gagagctttt	ataaaaaatgc	caatgcccgg	ggccactttt	gggaggagcc	aggcatcagt	21660
aatttcaaag	gtctctaaat	gatttacagt	ttgggaatca	ccgtatgagg	atagtaagct	21720
ctgagtcccta	tgcgttctgt	gccgaacacc	catgaagcag	tcttccaagc	attttacctg	21780
catcatctca	attctcacac	tgtttaaggag	atagacagta	tcattctccat	ttttagagaca	21840
agacaactga	atctcagaga	ggttttaagtc	tcaggacacc	aaggctcatta	ttaatcaggg	21900

ggactgtgat	tgctcccttt	ataaaatgta	ggagatattg	tggagtacgg	ttgagaaacc	21960
attgcaatag	ttttcttact	ttgttaagaa	attaggctgg	gcgtggtggc	tcaggcctat	22020
aatcccagca	cattgggaat	ccgaggtgga	cagatctctt	gagctcggga	gttccagacc	22080
agcttgggca	acagggtgaa	accccatctc	gactaaaaat	acaaaaatat	tagccggggc	22140
tggtggtgtg	cacctgtagt	ctcagctact	tgagaggctg	aggtgggagg	atcacctgag	22200
tccggctgca	gtgagctggc	attgtgccac	tgtactccag	cctgggcaat	gagagtgaga	22260
tcctgtctca	aaaaaaagaa	aaaaaaggaa	attagtgggtg	gaaggtgact	ttgcatctgg	22320
gcgtatctgc	ctgcagagtt	ggtgtcctta	ccttgaagaa	accctgcttt	agttggagta	22380
tccttaatgg	ttagtggcag	gaggggagga	gtggttcctg	ggagactgga	acaaaatatg	22440
gtacctgaat	gcttaaggct	tggcagatga	gcagtcattt	tcttacacag	agcttaggaa	22500
agggcaccca	ggtagaggaa	tcagcatgaa	caaaagcaca	gggccataga	gttctcagaa	22560
ggaaagatgg	ggttaaccgg	agccaagcca	gagatctggt	ggtagtgggg	ggtttccaag	22620
ctagaatggt	tgtgtggtat	tctgtcctca	ggggctttga	actctgtgtg	ctaattaggc	22680
ctcaaattct	ctggggctct	ggttaaaatg	tagattctga	tatcagttgg	cttgggtggg	22740
gccttgcatt	tctgtaagcc	cttagcagtt	gcactgctgc	tactaccgtg	agtattgctg	22800
ttgagcatta	ctaccttgag	tattgctgtc	aagtgttact	accttgagta	ttgctgttga	22860
gtattactgt	cgaattttac	taccttgagt	gttgctgttg	agtattacta	ccttgagtgt	22920
tgctgttgaa	tattactact	ttgagtatta	ctgttgagca	taaccacttt	gagtattgct	22980
cttgagtatt	accaccttga	gtattgcttt	tgagtgtctac	tgacctgagt	atcgctgttg	23040
agtattgcta	ccttgaatat	tactgttgag	tattaccacc	ttgagtattg	ctcttgagta	23100
ttaccacctt	gagttttgtt	cttgagtatt	gctaccttga	gtattgctgt	tgagcattac	23160
taccttgagt	attgctgttg	agcattacta	ccttgagtat	tgctgttgag	cattactacc	23220
tcaaggattg	ctcttgagct	ttaccgcctc	aagtattgct	cttgagcggt	actgcctcga	23280
gtattgccgt	tgagtattac	tcccttgagt	attgccattg	agtttagtcc	tgtgagtatt	23340
gctgctactg	cgccttggca	atggttttca	aactttgcaa	cacatcagaa	tcacttggga	23400
aacctttaaa	attctaacgc	ccaggtcaca	tcccattcca	actagatcag	aacatctggg	23460
gaatgcgagc	catgcaccag	tagttataaa	acctgcccag	gtgattccaa	agtgtgggaa	23520
cctttgagaa	gcactgcttt	aggggttgga	atagtcctgg	ctgaatttta	atcaggggaag	23580
actgactgct	ccgtttatga	aacgtaggag	agtggagcag	ggttgagaaa	ccatcgggat	23640
agtgttctta	ctttgttacg	tgagcaatat	ttgttgagtc	tctgtggtgg	gttctagggg	23700
ttcagaggac	agcagtgtgc	tgctaggatg	gtggtctgaa	ctagtggaaa	ggcactcaaa	23760
ggaagaaaga	cagaattcta	agaggagagg	aatttttagga	aggagatacc	caggactttt	23820
gaattacagg	taatttgatc	agaacccaaa	actgaaatgt	ctctgctctg	tgatgaaagg	23880
gtttgctggc	attgagtaag	gagctgcagg	aaggccttta	acttgtctcc	aggtctctta	23940
acagctttgt	catttacata	caagcacctg	cctggctaaa	ccattcattt	ctgtagcttc	24000
cttctggatc	tgtctaggga	atatttgctt	tgcatatttt	ggggttatct	taagtgtttg	24060
aaggaaccaa	aatatTTTTc	ttaaaaataa	cactcaaatg	tagttcacat	gattaatttt	24120
gactgatttg	tgagaatcag	taagtgtctga	ctgactgagg	cgccccacac	atccggcttc	24180
cttctgttac	tctacgcgtg	ttgctgaaac	ttaacgaacc	catgtggggg	cttctcgcct	24240

ggtgcagtc	ggccccagtat	tcatactgag	gtttgcagtg	ggagaaagga	aggtatttat	24300
ttgtaggtca	ccaagcaggg	caaattccagc	agctcacgct	taagacctga	cctctcccat	24360
ggtttataag	caagtggttt	tttttttttt	tttttttttc	agactgagtc	ttgctctgtc	24420
accagggctg	gagtgcagtg	gcgtgatctc	agctcactgc	aacctccgcc	tcccaggttc	24480
aagcgattct	cctgcctcag	cctcctgaat	agctgggact	acaggcgtgc	gccccacac	24540
ctggctaagt	tttgtctttt	tagtagagat	ggggtttcac	catgttgccc	aggctagttt	24600
ccagctcctg	acctcaagtg	atcctcctgc	cttgacctcc	cagagtgtctg	ggattacggg	24660
catgagccac	agtgcctggc	ctgtaagcaa	gtgtttttta	agaaaggggt	aaattttagg	24720
gaaacagaag	ttctaggcaa	aatggtaaat	taatacaggg	aggtaagaca	ttggtttggc	24780
ctaaaaagat	gggatatatt	gaagtggggg	ctcataggtc	ataagtggat	ttaaagattt	24840
ttttggtttg	taattggtta	aggaagataa	gctttgatta	aagatttggt	gtcagcagaa	24900
agaaatgtta	ggtctggctc	gtgggcatgt	ctttttctag	gcccctcctt	ggaaagaact	24960
ttagagcaaa	gaaaggcagt	tggagcttag	tccccacttt	ctcctgatct	gaggtctacg	25020
gaccactgga	tccatttggg	ggggctccatc	tttctgaaaa	acaagtcagg	gacatgtatt	25080
gagatgatat	tatttggatt	tatagggaac	caaacaacgc	cccatgactc	ttttttggct	25140
attgttttaa	gccactgttt	ttttttgttt	attgagttgt	taacttattt	tttaaagcta	25200
gctagctgcc	tggaatttct	ttagaaggaa	ctgaagtttt	taaaaatttt	tatgttgggg	25260
ggtattgccc	tgcaggcccc	taaaaggggt	ccctgcgctg	tctcaaaact	tggatgcaaa	25320
aagaagttga	gttaacacag	gaggacaggg	gtagacgcac	caagggcattg	tgcctcgagt	25380
gcgtggctct	tattaagaag	ggtgggttaga	cagggaatgg	gttagttccc	aggctggcat	25440
tcagctgaaa	cagtgatggg	taaaattctg	aaaaatgtcc	acgctctgca	ttctcttctt	25500
aacaccagg	accagtaac	tataaagccc	cctac cctgg	ggcatagcag	ggggcttcag	25560
ggacccatga	gaaggtcatc	tgctgctagt	tacactcctt	ctgggacctg	atttagacag	25620
tttgggtgga	gttttgcgag	ggttaatttc	agggccaagg	atgcttctag	aatggaaata	25680
ccttcttgac	attgggagct	ttattggttg	attatgtcaa	tgtgagaatt	caggaagccc	25740
agtgttaatc	ctccatccta	aaaggagtag	attggctggg	cgtgggtggcg	catgcctgta	25800
atcccagcac	tttgggaggc	cgagggggcg	cggatcacct	gaggtcagga	gttcaagacc	25860
aacatggcga	aaccccgctt	ctactaaaaa	tacataaatt	agccagggtgt	ggtgggtgggc	25920
gcctgtaatg	ccacctactc	gggaggctga	ggcagggaga	attgcttgat	cccaggaggc	25980
ggaggctgca	gtgagccaag	atttgtccac	tgccctccag	cctgggagac	agagcgagac	26040
ttcatctcac	aaaaacaaac	aaacaaacaa	acaaaaacta	aaaggagatt	tcctccttct	26100
gtcctttatg	ggagacttca	accttgggaa	agtctggaat	ccttggacat	tagaaattct	26160
gaagttttgg	ctggctgtag	tggctcatgc	ctataatccc	agcacgctgg	gaggccgagg	26220
caggttggtca	cttaggccag	gagtttgaga	ccagcctggc	caacatgggtg	aaaccccatc	26280
tctactaaaa	atacaaaaat	tagctggggcg	tggtagcgga	cgcctgtaag	cccagctact	26340
tgggaggctg	aggcaggaga	atctccagaa	cctatgaggt	ggaggttgca	gtgagctgag	26400
atcacaccat	tgcactccag	cctgggcaac	agaacaagat	tccgtttcaa	gaaagcagaa	26460
actctgaaat	ttttgcctgt	ccaggccaca	tcaatcccat	tcctctgctg	tctctgcagg	26520
attctgtgag	gaataattag	ttaatgtttg	cagagcactt	tgaaatcctc	agatgaaagg	26580

caccggagaa	gcacaaagta	ttattattta	ttattagctt	gccccagaat	ggaggcgcat	26640
gaggccctgg	cagctccctg	cctcgtgcc	ggtgtgatcc	tcctgctggg	cttttcctgc	26700
ctgatgagct	tttttttttt	tttttttttt	gagatcaggt	tcagctctgt	cgcccaggct	26760
ggagtgcagt	ggcatgaaaa	cagttcactg	cacacagctc	actgcactgc	agcctcaaac	26820
acctgggctc	aagcaatccc	cctgcctcag	cctcccaggt	aactgggact	atatactaca	26880
ggcatgcgcc	accactcctg	gctaattaaa	aaaaattttt	ttttgtagag	atgggggtct	26940
cactatgttg	cccaggctgg	tctcaaactc	ctgggcctca	aagatgccaa	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	aggttgccgc	tgccattcct	tggggtctgg	ggttaagacc	27240
atctgcctga	gggtaacgca	gtaataaatc	agtactaaag	ggcgtactaa	agtactgtat	27300
tgctaggcta	ggccatgctt	ggtgtatttt	tttttttttt	taattgagac	ggagtcttgc	27360
tttgttgccc	aggctggagt	gcagtgggtg	gatctcggct	cactacaacc	tctgctgccc	27420
agtttcaagt	gattctcctg	ccttagcctc	ctgagtagct	gggattacag	gcacgtgcta	27480
ccatgcttgg	ctagttttta	aatattttta	gtagagattg	ggttttgccg	tgttgtccaa	27540
gctggtctca	aactcctgac	ctcaagggat	cagcccacct	cggcctcccg	aagtgtctggg	27600
attacaggca	tgagcctggc	tggtgtattt	gttttaaat	ttaaagtttac	taaatttaat	27660
gatatctggg	gaatcagctt	gcttcctggg	gatctggatg	tacttgaggt	gagagggttg	27720
ggattcagaa	ttatcctttc	tatcgagca	tgttctggat	tgattcatgt	aggtctcaag	27780
tgtgtgtaat	atttcatttc	tttgtgcaat	tttggcatgc	cgaggcgggc	accctgaagc	27840
tccggcagag	cctggagaca	gagtggggag	ctctcgcctc	tttcccttcc	ttcatcccag	27900
ctgacttcga	ctggaattga	attcatcagc	tgctggagag	ttgttttatt	tgccctgctg	27960
gtggagaggg	aggaaaggaa	catcatgggg	ccaggctttt	tttttttaaa	ggaaagattt	28020
gatttacttt	cccccttagt	agcatgatgg	gcacctgcac	ccgccagcta	atcagaagcc	28080
actgtccctt	gaatgcctcc	gctgcccacc	agatcctgac	agcatcccac	gcgggagcac	28140
tctcgtgtgc	ccctggcagc	ttctgctgcc	tggcagttct	ctaaacttgc	tgggtgtctct	28200
ctgcccggag	gctcagaaac	ccagaggact	gaccacttct	tgaggctcat	gtccagtttg	28260
caaagagccc	ccagcaagca	gagaagggga	tttttgtacc	agcgatatct	cttctccact	28320
cctcaacaca	ctcctttcca	ctctgtctcc	tataaacatg	gaacagccag	gaataactcaa	28380
atcctagcct	gtcatgaagc	caaaaattga	tagagatcta	ctgtccagaa	tgattttctta	28440
tagtgaccct	gtgttttagt	ggtaagactt	tcttaaacca	tgagggatcc	tgggtcccaca	28500
gggcagtaat	atctggggca	gagcctgaga	cttttctcat	tgatttctct	tgtgagccag	28560
gagtgactgc	tctgatgcag	ggtgctgtgt	ggttggtaga	agctggcggt	atcccatttt	28620
accacagagg	aaacaaatga	ccagtgggtg	agcgggagct	cagcatccca	tgtgcccact	28680
tctctctcgg	gtggactttt	cacctgcccc	tgcctctctc	tttgcaaact	ttactgcagt	28740
gacggagaca	tcttttaata	caaattcttg	ggggaaccct	gtgttccctg	gctggagcct	28800
ggctgggaag	gaggaggag	cagagggtct	tcttgggtgt	ggcctattgc	agttgagcca	28860
gggaaaggct	ggtccactgg	agacaccctc	tctggtcacc	gcagacttcc	tgccttccat	28920

ccagtgtcct	tctacttgca	ggatgtgtgc	ccagcagaga	gaatctctga	agccatgtca	28980
ttattgggat	aacattcctg	tcccagtcac	cttattttctc	agaaaaagga	caatgggaaa	29040
caagttttta	ttgaatccta	tgctgggcct	attaatgggg	tctcttactt	ttcatagcag	29100
cactgcaaac	agagttacgt	ttctattcat	tttatggatt	agaaagctga	gatccagagc	29160
gggcagatgt	aaacctgggg	tctttaaaat	gcaccccttt	tgcaaacaaa	taaacttagt	29220
gtattaaaag	ggctggagag	agcagagtaa	ggtaacattt	gggtgggtcag	catgtagttc	29280
tgggtcccca	cagtggagat	ggcacagtgc	tgggtgctgg	gggaactatg	gtcactaaga	29340
gacactgaat	aatttaaatgc	atgcccctga	ttccatcact	gactgttgag	gtaacacata	29400
catttatatt	gtcagtgggtg	gtgatgatta	catgagctgc	gtaaagcggt	tgaccagtgc	29460
ctgcacataa	catagtaggt	gctcaataaa	gatcaccac	tcttaagagg	tgggaggagg	29520
tgaagtcac	tttctgggga	gtgttgccct	gttgttctct	gctgcattct	ttctgtcctt	29580
tgggctccga	gaatgctggg	ttgggcagtgc	tgagtgggtct	tctcaggcct	ctgtgacatg	29640
ttgctttcat	gaaagggttc	cctctagcca	aagactgagt	ggtccttgca	ggctttctcc	29700
tgagtccttt	tttttttttt	tttttttttt	ttaaagacag	agactctgtt	gccagattgg	29760
agtgcagtga	cgcggctctg	gctcactgca	accgctgcct	cccaggttca	agcaatctac	29820
aaaatgcac	tataaaatga	tgcatcagcc	tcctgagtat	ttgggatcac	aggtgcccac	29880
taccatgcct	gggtattttt	ttgtattttt	agtagagaca	gggtttcacc	ctgttgacca	29940
gtttgggtctc	aaactcctga	cctcgagtga	tccgcctgtc	ttgacctccc	aaagtgcctgg	30000
gattacaggc	gtgagccact	gcacctggcc	tctctcctga	gtccttttgt	ttgtgcctgc	30060
tttggggatt	ccctctggct	gggggtggact	gccgggatct	gtttgtccag	tgtacatttc	30120
ctggtcacct	agcacgggcc	agctgcgggtg	ctgggaggaa	cagggcctgg	ctctgggagg	30180
cagctgggag	agtcaggaag	tgaagaaagt	tcttggtgggt	gtgatgggtg	aaacccaagc	30240
agcgtccaga	gggagcacaa	gagggaggga	caaactcttg	gaggggtccc	ggccaatggg	30300
accagtgta	agaaattgca	cctgtccttg	cagatagaga	aggtggaagc	agtgaatggt	30360
agagcatcct	cactcttctc	tctgccagca	agcacctttg	gggaagtcct	cacggacagg	30420
aatgtcgtgt	gtcttggtt	gagatgtcaa	agaaacatgt	tggacacacc	atggtgacag	30480
agcaggagtc	tcttaacccc	ggcgtgggtg	aggctgccgt	tctgggtggga	tctgggggtca	30540
gtcaggggtt	aacagtcgct	cctgcttgcc	tgattgacac	agtaataaag	gcagtgacac	30600
caaactaggt	ctcaggaatg	tgtcctcggt	agaaagactc	actaatggtt	gtgggggggt	30660
ggcccatgag	tccttctggg	tgggtggcgag	aagtagggga	ccctttgggc	tttgcccttt	30720
ttgggtcatag	gacttcactc	cacagacata	attgaaccgt	tgggtttctg	cagccaaatt	30780
caaatgtcac	caatcttggt	cacccctttc	atctcttggg	tcctctgtaa	gttatagcta	30840
tctgatagtt	tactgaaaaa	taaactgaaa	atatgtttta	aattgtactt	tcgattttaaa	30900
ataatgttta	gagacaaaaa	aaaagggtcc	aatccacttg	gagaaaagca	ttgtcaaagg	30960
tgggttgattt	ttttcttttg	ctgtttttaa	gtggtaagtgc	gatgagtgtt	ttggatatat	31020
tgattttttca	ggtgtgcagg	cggtcacatg	aacagctgac	attttttttt	ttttcatgtg	31080
gacttcagcc	agtcttgaca	cctgccccct	aacgaaaagt	aaaccatcgc	cttgtttgac	31140
agtttaagtgc	cagtgatacg	gatggaggca	ggtttacgtt	atgttaaagg	cttgacaacc	31200
cagaaccccc	ctgttggttt	ctttgttgta	acctttgagc	cgggtggcctg	ctgaaatgtc	31260

acctttgccc	ttctttaaaa	gcaggaataa	taggtggtga	gtgggtggat	gcctcttaaa	31320
atactgga	gtgctgtggc	ccgagggtaa	gctttttaga	agtgagtgtg	tgtgtttgtg	31380
tgttttaatt	aatgaatctt	ctgggcctga	agataatgag	gtcagtgagg	gcagccatgc	31440
tgcctcacag	ctcaccttag	ggtccttggt	gtccagaacg	tgcctgacct	actggagggg	31500
cctgggaatg	cttctttgat	tgacgtgggt	aggaagacag	atgtggcggc	ctccatgctg	31560
atgggaggca	gctgggaaga	aggtcatggg	caccatctca	ggagtggcag	agccacctcc	31620
ccctctcctc	accccggtg	tctggattct	tccagctgtg	tggtccttct	tcctgcctgg	31680
aaatgagcat	cctgcagagc	tcggctcctg	ttcacaccct	cctcctaacc	ccctactctc	31740
cctctccctt	tcatccagg	ctggaggacc	agatgggctt	tacctgatgg	agtgtgcttt	31800
gctgacatgg	tgcaaagagc	caattcctgg	ttgcaaagag	gcagctgggt	gcagaggcgg	31860
ggtgcattcc	tgtaataata	ataacttggt	tttttataat	actttacagt	ctaagtactt	31920
ttcaaatact	tgacctcatt	tagttctcac	cacagccctc	tgaagggata	ttactattac	31980
cttcatttta	tagatgcgtt	aaccagggtc	tgttttggga	ggtagagggg	gtgtgagggg	32040
gacagagggg	agggaaccag	tgttgaatga	attctgaggc	cctgccaaag	cagccagcta	32100
gctaggtgtt	gtttaatgga	ctctttgcat	ctacagaatg	agggaggtgg	gatgagggga	32160
aattattttc	caataactga	aggtcggaga	gactaactct	ctgctcattg	tcacacagca	32220
gttgagtgcc	agctgggatt	tgtcaccag	gtcacctgac	tcctaagccc	tgtgatcggt	32280
ctcttctgtc	tctagtatac	ccagcataat	gcccggcaaa	gtgctggcat	caataaatat	32340
ttgtcgaatg	ttaaatgagg	cttaaagaga	accattcatg	cttggcacag	gggcacagtg	32400
agacaaacat	gtttcctgcc	ctcgtaacct	tcgcttccaa	attctgtgac	cttgggcggg	32460
ttgcctgagc	tcttttccag	ctcagtttcc	tgaaaactca	tccggaaaat	gggtaaaata	32520
tcagagtgea	ttctgtatgg	taagactgea	aatgt tagat	gattctgcta	cttattattg	32580
ttttcttttt	ctcaccacac	accttccctt	tttatgtaca	cctcctagga	agtaagtttc	32640
tgataacata	ctgcattgtt	ggataacagc	aacaaaaagc	acttcctgac	attattgccc	32700
aaatcaccaa	atgaggcaat	taccaacttt	ggaataagaa	tagcaacggt	ggtaagagct	32760
gacatttctt	gagcgcttgc	catatgctgg	gtgtactact	ataagctgct	gcctgcaatt	32820
atgatttggg	aacaactctg	aaagtagtta	cctcccattt	tatagaagag	taaactgagg	32880
ttcagagagg	ttaagtaacc	ccccagggt	ctctcaggaa	gtagttgggt	gcctgggatt	32940
caaacaccag	aattggtctg	acttcccact	ctttaaacca	cactcaaaac	tgaactctcc	33000
acgtgtgtgt	gttctgggca	ttttcgcatc	tcccttggct	tgttgacagc	gtggactttt	33060
gctttcccat	tctcatagaa	catggccagt	gcaggaggag	gaaatccaca	ctggtctttg	33120
gactgaacca	gaggctggcg	atgggtccga	aacagggtgc	caagtggctg	acccttgttt	33180
ttatgccttg	cgctggtaag	ct tggggcac	caggagtctt	ttgaatttct	ctttcttgac	33240
tgtccacgcc	ttcttttaggc	aatcttttaa	caggctatgt	tttaaactct	attgacatct	33300
ctgaagaaag	aggaggaaaa	aaaaatcaag	acatggcctt	agagaagtga	caggttttct	33360
ttgagatttt	gttttctgtt	ttccttttta	ttttgtgcac	attgcaaaac	ctctttggga	33420
tgatgatttc	gtgtggtttc	ttggtagccc	ttgggcagct	gctgccaggt	ttcacccaaa	33480
tgcattgtga	ccccctgttt	cgtgggacgg	ctttgcctcc	acatggctga	ttgtgctctg	33540
tgtgtccgct	gtgggcagag	tgtgattgta	agaatcagaa	ttctgctggg	cttgcaagca	33600

tttaaaaaat	ctctataagt	ttgagaactg	gttggaaggg	agagatgcag	cgacttagaa	33660
cacccggcct	gcagctgagc	ttccgtgtgc	ctgggaggag	cccatatgga	gaaacaggaa	33720
aattccactt	caccagaaag	ctggggaaat	gagtgggagt	aggggccagg	ctggttaact	33780
aggaagactt	ttggtcactc	tgctttactt	agcctaaagt	gttcatttcc	cccttaagcg	33840
gtggttaata	cgcgtatctg	cagatttact	ttttggatga	tttaaaatct	tgcaacatct	33900
caagggattg	tatcctgatg	atactgatta	tattaataac	aagataatag	cttataatat	33960
aatagctagc	aattaccaag	cacttacctt	acaccaggta	caatgccagg	catttgatcc	34020
ttacacgaac	tctgagagat	ggctgtttgt	attcccattt	tacagatgag	ggaagctgtg	34080
ctgagagagg	tgagttcatt	tgcccaagct	caccacattt	gagatggctt	gcttttatat	34140
cctcttagcc	caattctttt	ggatgtcagc	acttggcatg	tattaggcac	tggataagtg	34200
ttttgttgaa	tgaacaaaat	gatggaactt	ggatttgaac	ccaggtctga	gctggctttg	34260
agtcttcaga	atagtaggtc	caattagtgg	agtgggggct	cagtagtcca	aagggaagg	34320
agcaaggaga	cattgtgggg	gccgagaaga	gggcttctgg	ggtgtttcct	gggcacattg	34380
gcattaaagg	catagtgtga	agtgccattc	aagaccatgt	gctggattag	tgtttttcct	34440
ctccacttga	ggtgcttgcg	attggctttg	tgccccggtg	tctgcaaagt	gagttgggct	34500
gaactcagga	agaccttttg	gttgggatga	ctgtgtattc	acctgcacct	gagtagggac	34560
tgagtttcac	ttgccagttt	taccgcagca	agacctcggt	aagttggctt	cctctcattt	34620
aggcatttgg	gaaacttttag	gcggctggag	tttaattctc	aaggcaaagc	cccttttcaa	34680
gggacatgaa	gaaaggcaga	gggatatatt	taaaatacct	gaatgaactg	tttctttttc	34740
tttttatttt	ttgagatgga	gtctccctct	gtcacccaag	ctggagtgca	gtggcacgat	34800
ctcagctcac	tgcaaccttc	acctcccagg	ttcaagcgat	tctcctgcct	cagcctcccc	34860
agtagctggg	actgcagggt	tgcaccacca	caccagctta	atttttctat	tgtttttattt	34920
tattttattt	attttttaat	tttttttttg	agacggagtc	tcgctctgtt	gcccaggctg	34980
gagtgcaatg	gcgtgatctc	ggctcactgc	aagctccacc	tcccgggttc	atgccattct	35040
cctgaatcag	cctcccaggt	agctgggact	acaggcacct	gccaccacac	ccggctcatt	35100
ttttgtattt	ttagtagaga	tggggtttca	ccatgttggc	caggtctctt	gaccttgtga	35160
tccgcccgcc	tcggcctccc	aaagtgtggg	gattacaggc	gtgagccact	gcgacttgca	35220
tgtagacagt	aatggcaggt	cactatcagt	gggtctgtta	atcaggtgtc	caacctgggtg	35280
ctgggcttgg	tggtcatgc	ctgtaatcct	agcactctgg	gaggccaagg	cgagtggatc	35340
atctgaggtc	aggagtacaa	gaccagcctg	gccaacatag	taaaacccca	tctctactaa	35400
aaatacaaaa	attagctagg	catggtggga	tgcatctgta	gtcccagcta	ctcaagaaga	35460
tgaggcagga	gaatggcctt	aacctgggag	gcggagattg	cagttagcca	agatcatgcc	35520
actgcactcc	atccagcctg	gacaacaaag	cgagactctc	acaacaaaac	aaaacaaaca	35580
aacaaacaaa	caaaaagtca	cttgccttct	tttttgcttg	cttatggaca	taaaccctgt	35640
aaactatctc	atacatcatg	ggagttaggt	tgcagtgggt	agactgctat	tcacaaactc	35700
atatacatcc	tatgaaggag	tacaggttaa	ataaccatta	tccaaaatgc	ttggggctga	35760
aagtgttttg	gatttttaat	ttttttcaga	ttttggaata	ttcgcatata	cataatgaga	35820
tatcctgggg	atgggacca	aatctgaaca	ggaaattcat	ttatgtttta	tataaacctt	35880
tttttttttt	tttttttttg	agacagagtt	tcacgcttgt	tccccaggct	ggaatgcagt	35940

ggtgtgatct	cggtcactg	caacctctac	ctcccaggtc	gaaacgattc	tcctgcctca	36000
gcctcctgag	tagctgagat	tacaggggct	tcccaccacg	cacagctaata	ttttgtatatt	36060
ttagtagaga	tgaggtttca	ccctgttagc	caggctggtc	tcgaactcct	gacctcaagt	36120
gatccacccg	ctttggcctc	ccaaagtgct	gggattacaa	tgtgagccac	tgtgcatggc	36180
cttatataaa	ccttaggtaa	ttttatacaa	tattttaaat	aatttttgtg	catgaaacag	36240
agttttgact	gcattttgac	tgtgactcct	cacttgagggt	cagggtgtaga	cttttccact	36300
tgtggtgtca	aatttcagat	tttgaagctt	tatataatga	gatagggtct	tgctctgttg	36360
cccaggctga	agtacgggtg	cacaatcaca	gctcactgca	accatgacct	cctgggctca	36420
agtgatectc	ccatctcagc	cacctgagta	gctgggacta	caggcatgca	ctgtgactgg	36480
atTTTTTTTT	TTTTTTTTTT	ttttgagacg	gagctctggaa	tctcaagtct	cgctctgggtg	36540
cccaggctgg	agtgaagtg	gcgcgttctt	ggctcactgc	aatctccgcc	tcctgggttc	36600
aagtgattct	cctgtctcag	cctcctgagt	agctgggatt	ataggcgtgt	gccaccactt	36660
ctggctaatt	ttttagtatt	tagtaggggtc	ggagtttcac	tgtgttggcc	aggttgggtct	36720
tgaactcctg	aactgaagtg	atctgcccac	cttggcctcc	cagagtgatg	ggattatagg	36780
catgagccac	cgtgcccagc	cttggctaatt	tttttatatt	ttttgtagag	acagggtttc	36840
gctatgtttg	ccaggttgggt	cttgaattcc	tggactcaag	caatctgccc	accttggcct	36900
cgcaaagtgc	tgggattaca	ggtgtgagtc	accgctcctg	gcctgaagca	ttttggattt	36960
ttgggttaga	gttgccacagc	ctttactgtt	attatcctga	tgttattatc	cacattttac	37020
aggcaaggat	ctggaggcgt	agagaggtaa	aatcattttt	tcaaagcccc	agaagtacta	37080
agccgcagat	tctgaatttg	aactcaggca	ttctgggtca	gaattagtga	ggttttaagt	37140
taattttttt	TTTTTTTTTT	gatagagtct	tgctctgtta	cccagggtgg	agtgacagtg	37200
gtgctatctc	ggctcactgc	aacctctgcc	tcccgggttc	aagtgattct	cctgcctcag	37260
cttctagagt	agctgggact	acagacatgt	gccaccacgc	ctggctaatt	tttgtatttt	37320
tattagagat	ggtgtttcgc	cacgttggcc	aggctgggtct	tgaactcctg	acctcagggtg	37380
atctaaccac	ctcggcctcc	agaagtgctg	ggattacagg	cgtgagccac	tgcgccctggc	37440
ctacccccctg	tcgcccaggc	tggagtgcaa	gtggcacagt	ctcggtcac	tgcaacctct	37500
gcctcccagg	ttcaagcgat	tctcctgcct	cagcctcctg	agtagctggg	attacagata	37560
cccaccacca	tgcctggcta	atTTTTTTTT	tttaagtatt	tttagtagag	acagagtttc	37620
aacaagttag	tcaggctcct	cttgaacttc	tgacctcatg	atctgcctgc	ctcggcctcc	37680
caaagtgctg	ggattacagg	catgagccac	catgcctggc	ctaagtttgg	tttttaacca	37740
tgtctgtttt	tctagaccct	tctgtcagcc	agctccacaa	tggatgaatca	gggagttagg	37800
tccgtctgta	agagagggcc	caggagctgg	gtcagatagg	taaagagaca	ttccttagtt	37860
cttatectct	gctaccaagc	catttcttgg	gatcaagccc	tctttggcct	gtgtcattcc	37920
acctccatta	agttcagcct	tctttccttc	tttccacatc	tttccactgc	tgttgaaaac	37980
ttgagacctg	aaatcccatc	tcttgaattc	ctgggagctc	tagaagtgga	gatggccagg	38040
ttctgtggtc	agagctgggt	gggattacaa	ataaaccaaa	gcctgggaaa	ctttcttget	38100
attaatagcg	cagacctttt	ggggagggaa	taccctaaact	cgatgctgtt	ggaattgatt	38160
ttgcctgtct	agatgacata	ctaatgagct	aagtgggttag	cttcggatca	ttattgctct	38220
ttcccaagcc	aagttctttt	aaagactaaa	accacaaaag	cagagaacga	gttgggttag	38280

agaggcatag	tggctgggtc	cagagaaggg	agagtgggtca	gccctggcct	taaacatgag	38340
aaaataaagg	tggctccttg	tttggaatga	gttagtgggtg	ctgactaatt	caactggttt	38400
ttctttttct	ttttggaagt	ggaatttgat	ttgggtgtctg	gattttgata	gggccattta	38460
tattttcttca	gcactttttg	gttcttgcag	aaagttacat	tcctagtctc	tcaactgctt	38520
atcttctttt	ggtttttgaa	gcaggaatgt	gattttgggtg	ctgggttttg	ataggggtgt	38580
ttatgtttca	tcaatgtctt	ttggttcttt	cagcgtttct	ctccttgtct	gtcatgtgtc	38640
agagaggggtg	cctgtcaacg	attctctctc	tccagggaga	aagtctttct	taaaacagcc	38700
ctaagatcct	tatctcctca	aatcacccac	tttgatgata	atatccattg	ttctcttctc	38760
tgttctttgg	catattctag	tcagtctatg	tatacaatta	aaaaaacaaa	acagccctct	38820
gtgtccaaag	tgcttggaat	atcccagtgt	ttcacaggag	acttttggag	tggacaaaac	38880
tgtatttctt	tccccaaatg	aggttattgt	gtctgaaata	tctcctggta	gtttattaaa	38940
ggaaaccgca	ggcaggggta	agagaggcag	tttctacacg	ctgcaaccct	attatcttgc	39000
ctcttctttt	cgaccctcct	tcctccctct	cctcttctct	ccccccctca	ccctattcaa	39060
cccagcccca	catgtcatgc	cgtccccagg	aggtagccct	gcagccctgc	ttctctggga	39120
tggctctgtt	ttgccacccg	tcccatggaa	cgtgaagaag	gaatttgggg	tgtggacttc	39180
cttgagtgc	taggattaga	cccgctgggt	ctgcagtcag	acgagaagcg	tgtgggcaaa	39240
gggaactatt	gtgtgaggct	tctctggaca	gaaagcctgc	cttcactctt	tactgtgcct	39300
aatggacaat	tgagacattc	agcttatgtc	tgaaggaaa	gtgggccggg	atggtctagc	39360
agacctcca	gatgaaggct	tgtaggagga	gcaaataagag	acaaggatta	ccaacagggg	39420
gaacaactgg	ggcagagtcc	tgggagagaa	tgtttatttc	cttgctctct	aggagggatt	39480
tggaaagagc	cataatcctg	ggttaggagt	aatttgttac	agcaagatga	tacttgagtg	39540
acaggctgct	tctggctgag	gcagcaagac	ttgcatgcag	gggggtcgtg	gggcctccag	39600
aaggctcagc	tcccgtaaat	cttcaccctg	gctttggggg	ttgttccctc	ccaagcaaaa	39660
ttaaccagag	gcactgctga	cctttgggct	tcctgggtgt	agcgttacga	agcatctcca	39720
catgtttgtc	acagctagaa	tttgacaata	aaaatttggg	caggagagacc	ctgccagagc	39780
cactgacctc	tttccaatgt	gacaagggga	aaaaaaaaca	aaggaaaacg	cagcacgggg	39840
tgcggtttca	gttgaagttg	gaggacacgg	agcccagcct	gtctcgcatt	tgtgtcttat	39900
gtagactcac	taaagcaagt	taattcattg	ctctttgacc	gccaaagtct	tcgttgtctt	39960
ttgtttgtgt	ggaatggggg	aaagaaatac	agaatgggga	ggagaacctc	attagaaaaa	40020
tcaagccttg	agagctccca	gccatggaga	aagaaaggga	tttttttagaa	gttgtgattt	40080
taatatctgc	tgcaatctga	tgattcatgg	atttaaaata	acccttccag	gtcaccagga	40140
ccctgttact	tgtctggctt	gtacctctca	aaggctcatt	gttggcttcg	tctcttaaca	40200
atttccatgg	tagacctaaa	atttctggct	gtgaaatccc	ctgtgtagtg	ggaagaagaa	40260
atagcaaate	ttagctgcct	tggacctgat	ataattattt	gtcttcattt	acatggttta	40320
tccttcaagg	ttgaataaat	gatgtgggag	ctagtcaagg	ggcttttaggt	atgtgatttc	40380
atgcctactt	tttttttaggt	agagaaactg	aggctcacagg	gtactagaga	atggactcta	40440
agattcaggt	ttctgaattg	cctgtgggtt	tgttgactca	actgctcttc	tgttgttttt	40500
tagccacatg	ccttgaaaca	gtcctctttc	ccatgtttct	tcatcagcac	cattaacceca	40560
aggtatactg	tcctctctta	tctttcacaa	ggctttggag	ttcccatgcc	tttgtaagca	40620

tcctccccg	agattcagca	ccaacaaaa	tcacatttgg	aaaaattgct	tgtttcccaa	40680
gaagcttttg	aggatatgat	tttgtataga	acgggttcac	aggttttctg	ttcattcttc	40740
tatggtggag	tgtgtgtgta	tgtgactctg	tcttctctcc	attcctcttt	tttttttttt	40800
ttttttgaga	tggaatttcg	cttttgtggc	ccaggcttga	gtgcaatggc	gtgatctcgg	40860
ctcactgcaa	cctccacctc	ctgggttcaa	gcgattctcc	tgtctcagcc	tcgcaagtag	40920
ctaggattac	aggcatgcgc	caccacgtcc	agctaatttt	tgtattttta	gtagagatgg	40980
agtttcatca	ctttggtcag	tctggtcaca	caaactcctg	acctcagggtg	atccaaccgc	41040
ctcggcctcc	caaagtgcgtg	ggattacagg	tgtgagccac	cgcgccaagc	ctccccatcc	41100
ccttttatct	cttaaatgaa	tgtggtcacc	atcaaagatg	gtgcctgact	cttttttgtt	41160
ttcagttcat	cttaaatcca	catataattc	acacgtcata	aaatgtaccc	atttaagggtg	41220
tacagttcag	tggtttttta	gtctatttag	tatatattaca	agattgtaca	aacataaccag	41280
tatcttaata	tttttatcat	ccccaaaaga	aacactgtaa	ccctagcagc	cagtctctac	41340
ccgccttccc	catagctcct	ggcaatcact	aatttacttc	ctgtctctat	gaatttgcct	41400
attttggtta	tttcatataa	aaagaatcat	acaccatgaa	actttcttca	tctgccttga	41460
agtttagcata	ttttcaaggg	ctacatgtt	gtggcatgtg	tcagtactcc	atttgttttt	41520
attactgaat	agtattccat	tttatggctg	taccgcattt	tagttatcca	gctatcggtt	41580
gagacttgg	gcattcttat	cccagaacat	accatattca	gtccccagt	acaccacat	41640
tcattcctgg	gctgtctcct	gtcttccagc	tattttcctg	gtctcctgtt	gcctctgcct	41700
acttcagcat	gctgtagaga	catgggtagt	aactaaaaca	ttccaattaa	ctgcattgta	41760
cttggccttt	ttataagaag	cagtaattag	aaaatatgg	ggccacaaga	ttgatattaa	41820
agtgaagat	tgtaaatact	tttctgcctg	aaggtagatg	gcctctggcc	tgcctcttag	41880
tgggagggtc	ttccaggagc	ttgcaagcat	ccattatttg	ttagtcatca	gcttagcggc	41940
caaggagcat	tagcctgtct	tgtctgtct	gctgaagact	ctgagagaca	tgggagggca	42000
agggtctgct	cttttgaatt	cttccaatgt	cttcatgtcc	tttaacctcc	tggcttaggg	42060
acttgtgtgc	tgggtgtgga	gctgacattt	gtttggaatc	cacagccctt	tgggtgggac	42120
tcaatcttgg	ggttgcctga	agactttgag	atggctaggt	ctgggcctct	tttggtcact	42180
atggaacaag	actgtctcag	aggccagagt	ctgtctcacc	agctccctgt	cttgggactg	42240
caccattgca	gggtctttgc	cctccccctg	agatttctct	tcctgcctgg	gcacccattg	42300
gccattctgc	ccgtaagctc	agtaggggtg	aggcaaaaaga	gttctggcct	ggaagtacca	42360
aagtccctgc	ttcttggttc	agtcctcat	aactgtgtat	aactaagtca	cttagttttc	42420
tgtgcctcac	tttcttctgt	tttaagatgg	atttggagat	tattggcttt	gaccacctaa	42480
aaaggatgta	gtgacaatca	atttagaggt	ctaaaagagc	ctttgaggaa	gtaaaatgga	42540
atcttcaaat	ggactacatg	ctgattattg	acactgccct	agcactgata	gttgatgttg	42600
actgatggtc	agaattgctt	ggcaagttag	aaaaaagtac	gtacagatcc	tgggccacta	42660
ccaagtttca	tttaacagat	ctggagtgc	tcaggaaaaa	agtcctctca	aacaagccag	42720
caaggttttg	atactgtgca	accttttttt	tttttttttt	ttccttttga	gatggagtct	42780
ggctctgttg	cccaagctgg	agtgcagttg	cacaatcttg	gtctactata	acctctgcct	42840
cccaggttca	agcaattctc	ctgcttttagc	ctcccagagta	gctggcataa	caggcgctg	42900
ccaccacacc	cagctaattt	ttatatTTTT	tggagagatg	gggtctcacc	atgttggcca	42960

ggctgggtctc	gaactactga	cctcaaagtg	atccgcccac	ctctgcctct	taaagtgtctg	43020
ggattacagg	catgagccac	tgtgtctggc	cctacttacc	ttcttttgtgt	taattcctgc	43080
accattgatt	agcttattgt	ccattgact	gtgtcttttag	atgacttctc	tgggcctcag	43140
aatatctagt	ccatagctga	cacagagcat	ctgttttaatg	gtaaatgctg	caggaatcca	43200
tgcattggag	tagaaagagt	tttagatcat	gttcctcatt	tcttgctaca	gacttaggca	43260
aagcgtggag	aagaggttgt	ccaatgaaga	aatgaagtga	catgccaggt	cagtggcaga	43320
gctaggcctg	gaaaataggt	ttccagactc	ttccctttct	accatacttt	tcctgggagt	43380
acgcactcgt	aatttgaaga	gcgacttttg	ggagagggtg	gaaggaaggc	ctgggcctca	43440
gcctaagggg	ccattgggtt	gtgagaggag	ggtctgggtga	aattccatac	cgattgtccg	43500
tgtgtgagct	gctgtacat	agcctccctg	cagaaccact	aacctgtcaa	atgcagaaat	43560
agttcaggga	cagagctgtt	aaaggattgg	cgggttaaag	aaaacagtga	atcccaagtt	43620
ttgttaattg	gatttttttg	ttgttttagt	atttgttttg	cttcattgtc	ttcatcacac	43680
caggggcctc	cttaaatctg	gtggaaaaat	ttccttggaa	aacaattcag	tgtttgtcca	43740
tagacttggg	agggagagat	gctagatgct	ggaaagtctt	gcttattact	ttggggacac	43800
tgagatgttc	ccttcacat	gtactttgag	acacacatcc	tggttgagtt	caggcaagga	43860
tgccataacag	ttgataagaa	aactgggaaa	gatagaaggg	atttgtaagg	taagtcaggg	43920
tgagtgaana	cacatccggt	atgctggaga	cctagatgct	tgactgccac	tcgctcctgt	43980
cacctcagtc	aatctgggtc	ttgctctgtt	ggccttagtt	tcctccttgc	taacagggtta	44040
gttccacctt	tctgcccatt	tattttgtag	ggttattgtg	gatgtcattc	tgaactctaa	44100
aataccctct	aaatatgaag	tgatattagt	gctctttaca	ttgttatgat	taaaaaatatt	44160
tatgagaaaa	aggttaactg	taaggatttc	attgaaaatc	ttataacaac	caactgatag	44220
agatagaaga	taaggctatt	aaattgttca	cacagatgcc	ttgatatcct	acctttttcc	44280
ccctatatte	cttttatgtg	agaaatgaga	tagtgattta	agggaanaac	ttaaaagagt	44340
tccgactatg	ttggtttttt	ttcccccaag	tcaaccttaa	tatcttactt	aaatcttttt	44400
ctttttttate	ttttcttttc	tttttttctt	ttccctccct	ccctcctttc	ctctccttcc	44460
ttcccttctt	cctcctcctt	ctgctgcttc	tctctctctc	tctcgtttcc	ttttcttttc	44520
tattcttctt	ttttcttttg	agaccaggtc	ttgctctgtt	gctcaggctg	gagtgcagtg	44580
gcaccttctt	ggcttattgc	aacctctgcc	tcctgggctc	aagtgatcct	cccacctcag	44640
cctcccaagt	agctgggacc	acaggcacgc	gccaccacac	tcagctaatt	tttttttttt	44700
ggtagagatg	gggtctccta	ggctgggtctt	gaactcctgg	actcaagcaa	tcttcctgcc	44760
tcagccttcc	aaagtactgg	gattactggc	gtgggccacc	atgcctggct	tgaatttttt	44820
ctatggcttt	attctttctc	caagtacaga	gtctacccaa	ccttctgaga	tctttggttt	44880
tcttttctta	ggtaactata	gtacatactt	atttatgtta	aacaacagca	atcacacatt	44940
tctttttcta	tacagtcatt	ctttatagtc	aaataaagcc	tccgtcttag	gctttctgga	45000
ttttttcaaa	agatgcaatt	cctggagtat	gttttttactt	agagcaaagc	agcctagtct	45060
cctatacctt	ctgcatctgc	agaaaagttg	gttaaacaga	ctttgtaatt	atgcccctta	45120
caattctgaa	gggactttgt	aaatagtttc	acagagtttc	agtgttaggt	atattttgatc	45180
aatgctaact	tttggaaaac	tttgggtgct	gtatgattca	gagggtaggg	cagaatatta	45240
aattaatcac	aacttcttgt	attttaacca	ttctgggtaa	attgggattc	cgtgacgccc	45300

aggcaaaatt	atttgtttat	agaagatggg	ctgaattttc	catcgtccat	ttctgagaaa	45360
tgaggtaggt	ttagaaagag	acaatcaggc	ctcttcttta	acagaaatgt	ttgtgtctac	45420
taggtgtgtg	tcacaatatg	agttcctgaa	gaaataagtg	tccgctattg	ggttgataac	45480
ttgtacttcc	tattttctta	ttttgcacat	ttttctggta	tttccctttc	tatggtgagt	45540
ggcttctgat	cgtctttcct	tttgtaaagt	gtaatgatat	gagaatcata	atcgtgggtgc	45600
ggtcttttgt	gttgcatatt	tgtagggggg	cagtatgaat	ggcccgtggg	gaggctgcac	45660
tgaagatta	ggagcagcca	ccttgatgcg	gaggaggcct	agtgactttg	gacatgatgg	45720
gctatggctg	gctatactct	cagctttggg	cgcataagca	gagtattgat	tttgtatttg	45780
gttaaaacca	gaagtacaac	tttctggcac	cagaggatta	ggaaaattta	acagcggaaa	45840
gccatcatga	ggatagtaac	caattaattc	gatttttttg	gtcagacatg	gtccccacct	45900
gtaatccag	cactttggga	ggctgagggtg	ggagggtcat	ctgagggtcag	gagtttgaga	45960
ccagcctgac	caacatggta	aaacccgac	tctactaaaa	atacaaaaat	tagcctggcg	46020
tggtgatacg	cgcctgtaat	cccagctact	cgggaggctg	aggcaggaga	atcacttgaa	46080
gctggaaggt	agaggttgca	gtgagtcgag	cttgcgtcac	tgcactccag	cctaggcaac	46140
agagtaagac	tgtatctcaa	aaataccata	attcgttttg	tcttttcttt	acttttttct	46200
ttccttttcc	ttccctctc	ccctccctc	cttccctttc	ctcccttttc	cttccctttc	46260
catctctttc	cttcccttct	tttctctctt	tctctctttc	tttcaacagg	gtctcgctct	46320
gaaccttttc	cagtcagaat	tgctcagggg	tttttagact	tccattcttg	aaaagagggg	46380
gtagttat	tggtgagatt	gtggctctgt	ggttagacct	tgtgatgggg	gcctcagcca	46440
aagggttcag	gatttttttc	caagcttttc	cctcacaact	tgagttaatc	cgaacggtg	46500
ctattaggcc	accggacatg	cttttctgca	tgccctgtgt	gggctgtttg	gattgaaggc	46560
ccagcaaggg	aaggcacct	cgcccatctg	acacaggcag	gcctctacaa	ttttattccc	46620
taaccagggc	atgacaaact	atggcccata	gacaaaaatt	ggcttgccac	gtgctttttt	46680
ctggccagtg	agttaagaat	gactttttat	tatcattatt	attaatat	tttgagccag	46740
gttctcattt	tgtcaccag	gctggagtgc	agtgggtgca	tcacggctcc	tgcagcgtga	46800
aactcctggg	ctctagcaat	cctcctgcta	acttttttta	tttttgtaca	gtcttgctgt	46860
tgttgcccag	gctggctctg	aactcctggc	cttaagcaat	cttccggcct	tggccttcca	46920
aaatgttggg	actacaggcc	tgagccgctg	catccagcac	ttttattatt	tttaaatggt	46980
tgaacacat	caagagagga	ataatat	ctgacacagg	aaaatgat	gaaattcaca	47040
tttcagtatc	tgtaaataag	cttttatttg	agcacagcca	tgatacaaga	catatactga	47100
ctgcctgtgg	ctgctttcga	gttacaatgg	ctgagtcgag	tagttatgac	agagattgtg	47160
tgggcccga	agcctaagat	atttgctgtc	tggcactttg	cagaaaaagt	ttgccaaccc	47220
tgccctgaac	aaataaagg	acaaattcca	cttgccccgt	ccatctgtgg	agcagagtca	47280
ctgaaaggaa	atactggaaa	tactggaagc	cacttggtgt	tttatcaagg	atgtgaggtt	47340
tcttggaac	tttgtcgcca	tatcatc	atcatcacca	tcatcatcat	catcatc	47400
atcatcatca	tcatcatcat	catcatctgc	ccittaaagt	ttctgcttgt	ttagaaaaga	47460
aatttataca	gagccccag	tagcagctgt	aagggggcag	gttcttggag	cagcccatcc	47520
tcaacattct	tgtctgtgat	ggaagattct	caaggatgaa	ggccccctcta	tgggagcagg	47580
atcagtctgg	ctttagtaga	tgccaat	tgctaagact	atttcctaaa	ggagcctctc	47640

ctcatttgcc	ttttctccct	gttttcattg	ggggagggtg	aagaggagaa	aaataattag	47700
agatgctcac	ctttttcttt	ttgctggcaa	tttaacagtc	ttttcagctg	ctttgattcc	47760
tttcaggcca	ttggtgttgt	atatatttca	agatttgctc	acaggtecaa	agcttaactt	47820
aagctccctg	agacatatca	taaaatatga	tttggggaaa	aaccctaata	ggccatgatc	47880
agaacattat	tattcaacaa	aggatgaaat	gcttaagcca	agatggcctt	ctttctttct	47940
ttctttcttt	cttttttttt	aatgaaagtt	gagcagactc	ccgtccaaca	gttttcaatg	48000
taggaattcc	cacageccca	tttgattgca	gtttgttgaa	aagttaaata	ttttttagg	48060
caattcataa	tttccacatt	gaacagcctg	agaggaagag	agctggagcc	cactgttgtt	48120
tttgtagtgg	gatggtggga	actttttttt	tccctccccc	aaaaggatat	aaaactaagt	48180
cagatggttg	ggaaaacgtg	gcacagggtt	ccagcccttt	tgtaaactct	agatgcccc	48240
tccttttaggt	cttccttttag	gacccaacag	aatagaaatt	cctgctgctt	aatgtctcca	48300
ggaaggaaaa	aaatttttct	ctaggctgta	atagtagcta	atttcctttt	tcttctcttt	48360
atttatttat	tttccctatt	aataagcacc	aattgtagaa	gatgaaggaa	gctgggaaac	48420
ccatcacttt	tggagaaggt	taatagcttc	ctttagaaaa	tcctgacata	atacttattt	48480
ccccaaaagg	cacttcatca	gcctgaatgc	cagttaagat	tcaaggaatg	ggcttggatt	48540
tgtgtgtacc	cagcggttct	gtggcatcaa	gttgcaactg	gaaggagagt	ttggggctgt	48600
cactgtggag	tccctgcaag	tcagcaggac	cagggtgtgc	ttcctgcacc	atctggattt	48660
ggttagctct	ctctgggcag	tggggccgag	tctcatttcc	tccaacaata	atgttatata	48720
ggcaatgate	ctgggctgcc	ctaacataat	tgaaaattat	gtgtattgta	ggcttggagt	48780
gctgaaatgt	gggctcataa	aaatatgtgg	tgcaggtagc	ctatggagat	tggatgtggc	48840
acacaatgaa	gcttttatgt	aaagtaagaa	ttataagtct	ccatgttaat	attgtattat	48900
gagtatgaca	gttcttgggt	gggtcctcag	ggcaggctct	tcaccttcaa	caaagcccga	48960
gtttcctaata	tctacagagc	tggatatttg	atgtaataca	atcggttttg	cagggtggcca	49020
aagatgaaaa	cttgtccacc	aatccagctc	tccccactga	gggatagcat	gggatgtaga	49080
tgggtttgac	tccatttggc	atttttgttc	acgggttttt	atgagatgga	gaggtgagtg	49140
ttggtgggtg	tccatttttg	ttggcctcaa	ggaaatgact	ctattgagtg	gttttgacca	49200
atgcagctca	tatagttaat	tggtaagtga	gaatgggaag	aagttaggat	gagatggggc	49260
agtttagatt	cccagagccc	tctggccttg	gttacagatg	gagactggaa	atatttactt	49320
tagtggttct	caacttgaga	tgatactgct	cccagagaag	gtatttggaa	gtgatgagat	49380
ggtaaggata	accaaggggg	ttcctgtttg	tatttactgt	ctgggggctt	ggagtcctac	49440
aagtccttca	gtgtttgggg	cagactcccc	acctaatacc	ctgtcgcaga	taggacaact	49500
cattcagtac	acagatgaaa	aaaacagaga	tcactgaagc	aaggggagtc	gatgcagggt	49560
cttgtggcaa	gatgcagaca	caaccggact	aataactagg	ttgctcacca	cgggaggcct	49620
ctaggtgaaa	gctctgaatt	tgtagcagac	acaccacact	cgtatagatc	ctagacgtca	49680
tgggaaaate	gactgtgtac	tttggcaagt	agtctttggg	caatgatctt	ccagcttttag	49740
gtataaccaa	atttggtttg	aatttgccaa	gcagtcgtat	cttcgaggaa	ctccgtcggc	49800
tggcttgttg	atggcttttg	cacttctgtc	tctcgtggga	tttgtcaaaa	cccttctttc	49860
tgtattatec	tttctgtctt	ttttcttttc	tattgaaatt	gttctgacca	tcaagacctc	49920
actctgtgca	gccttcccca	gtctattgtc	ccagaaattc	tgtcatcttt	cttggcattt	49980

cctgagtc	ccc	tgagtctctg	tcacagtgtc	accatgttct	gtcttgattt	acctgtgtct	50040		
gtaaggct	ccc	tcattgctggc	aaaactcccc	gagagcggac	atctttgtct	ctcctagtgc	50100		
ttgtcacagc		ctgtacacaa	agcaagtagt	actcagtgtt	cattgagtaa	agttttctat	50160		
agaattaata		ttaaaaccag	ccatttat	ttt	tgcttgagga	ggctctccgaa	atgaccaagg	50220	
tgtctcctta		tatcttata	t	ccccccaag	cattcattaa	ctgatggatt	agtgagttgg	50280	
ccttgagaag		cataaaggct	cgtctccatg	tgcttctaag	cattgtgtct	aagtctgtt		50340	
tggtttcctg		agtgaactg	tcttaatgtt	accaacagaa	gttaaagcc	taagagtttc		50400	
ttatacatgg		gctgagtacc	tctgtgactg	ggcaagccac	ctcacctcat	tttaccttgt		50460	
ctgcaaaatg		aggaactggg	tcaactc	atc	gttcaa	atct	cactgaaagc	taattgatcg	50520
cttttgacag		aagtagctcc	cttggg	ccgt	atatttat	ttt	cctagcttgg	aggaaggtgg	50580
ggacagacag		aattgatgta	cacctttatt	tttatctcta	tggtaaacct	gtgcatacta			50640
aagcattcct		ctggctttt	gagatgagt	gt	tatacattgt	gtctggccct	gtgcattttt		50700
taccaagaag		taagttttgt	tgagtaaact	tgggttgtat	gaagaactgc	atgctcaccg			50760
tactcaagta		gcttttgcta	cctaaaggac	agctgctcat	atgtacttga	cttcctttaa			50820
agtgaaggat		gatgacattt	gaaaaacgga	ggttgaaaag	gagcagattt	ggaattgatg			50880
gtttcctagg		acattcttgg	cttgagattt	gtgttttact	ttcttccctt	ggaatagctc			50940
tatattcttt		cctctccctc	cccacctctc	ccactccctc	ccagccccc	ccaagttaag			51000
gtagtagtaa		tgaaatcatt	tttctgaag	ctaccctgta	ctttgaaatgc	aaagacaaaa			51060
aatacagttg		ctagtaacat	taatcttcta	tatgtgtact	tactgaactt	gagctctgag			51120
gaagacccta		ttggaattgc	atgctttttt	at	tttttttaa	tgattatttg	catgcttgta		51180
tgtttttcag		tttctgacct	atgtcacagt	tatttcttgg	gctagttgtt	ctgcatttac			51240
tttctgaatt		cattgttttt	catttcactt	ttgtttccctc	tcgccagtat	ctccagatga			51300
aatggccact		gcttgatgtc	caggcaggga	gcctccagag	tagacaagcc	ctcaaggatg			51360
cccggtcccc		atcacccgga	cacattgtcg	taagtaacct	cccagagatg	atggcttccct			51420
ttattgaggg		ggtgaaaaag	aaaatgcctt	tttgatgata	acaggcccta	tttgtcattt			51480
ttttctttct		ttaa	acacat	tttcttttga	aata	ttgttg	ggtatatagttt	atatctataa	51540
ggtattcatt		ttctgtctatt	ggacctta	at	tt	gtgtaacc	tacctggaaa	ttttacaaac	51600
ctttccctcca		ctcttttcca	tgtatttgg	t	aaaaatctag	ccttgtgggc	tctagtttat		51660
aggacacaat		caccatggta	tggaggagac	tagagg	ttgg	atcaaagcag	ttataaaaaat		51720
acattcaggg		caggtgaagt	gaagaagagg	gaattagaaa	actcaaaagg	gggtcctgga			51780
tttgaaactt		gcctattatc	ctctccccc	atttatctta	atattttgtg	gcaacattct			51840
acactaacat		tagaaaaatt	tcatctgggc	tggctgactt	gtaaacctag	agtagaaatg			51900
aactttgaaa		ggctaaaatg	gaatttaatc	tatacatcca	tggctttgaa	agtatgtagg			51960
tttgatagag		aaagcatttg	tttttagtac	taagagacta	caagtgtgtg	tctacatata			52020
tttttaatgt		at	tttctttag	ggttttgtag	gctctaagag	tggaatttat	aaattaacct		52080
cttgagaaga		tagctcagcc	ttatttgaag	attcccttct	atgtatttat	atcatgagct			52140
ggacttcata		cttttgaaat	aattaatgga	aggcatattt	ttataatgaa	tccatccatg			52200
acaggtagaa		ttatgcaaag	catgaatcaa	tcatgggttt	ttcattttgag	tatcacaaaa			52260
tgttaatcat		aaatacattt	tgcctctata	ttgtaatttc	taaaaattgc	aaaataagtt			52320

tcttaagtag	aaaaatctta	agatgcattc	tgccattttg	ggctaactgc	ctccttattt	52380
tggagcttgc	tgtaatlgag	catgtgttat	ttaatgagtt	atacctctgt	catatgtgtg	52440
tgtttatata	acaaaataac	ttatTTTTat	aaaaccatat	tttgagtcac	catttgtgac	52500
aatgtcttct	tttctctggg	ataaatgagg	catgtagaaa	gaagattgac	atttgcctaga	52560
agcttccctt	ttcctctaac	tccacaataa	aatggatgct	cataattaca	tctgctccta	52620
taaggtcaag	atttcagggc	tggaagtgc	cttagatcat	ttaggcccaa	cttgccctca	52680
ggaaaggaaa	ctgaggccca	gagatgcctt	aagtgaattg	cccaatgtca	cacgctgagt	52740
cagtggccag	agcaaggctt	ggatccagtt	ctctgctccc	tttccagagc	cttgtgatgt	52800
cttctctcct	acaggagggtg	aaaataactg	ctgtggctgg	ttctgttttg	ctgactgtaa	52860
attgggtcat	ggtcaggggc	agtgcatagg	tgtaaagaag	ttgctgggtg	ggggttctaa	52920
tgcaggtttc	tccaaaagtgc	aatgccctgt	taaaaaaaaa	ttcttaacaa	atatacagag	52980
atTTTTTTTT	taaaaaagtgc	tgacagttct	agacacctag	agagtaaagt	gaagaagcct	53040
gttttcagggt	ttcccgccctc	cctgaatttc	ccagcatggg	ccaggctttg	aaatttattt	53100
atctgctttt	ggcaatgggt	gatgggaatt	tcccacattt	atTTTTtagc	tacagagaaa	53160
ggacattata	tttaaaatct	cttcgttggt	ctctctcttt	gagtgaggag	agaagatgtg	53220
aatcctggca	gtgggttcaga	gtggacacag	ccccgtgtgt	tgtggcatag	gctctgtggg	53280
cccatgccca	gggagcagta	ccccgtgtga	aaggagtggg	ggtttgtcca	tttgataga	53340
gcaaagatcc	tccacctcaa	atcccacaag	aacagttgcc	acaacctggg	ccctaagcat	53400
ctcattttcc	tatgtagaaa	ttaatgatct	ggaggagatg	gcaaaacatt	ccttccagag	53460
cctgtgtgga	ttttggccag	gggtgcagca	agggggctta	ggcacctttt	tcctctgctg	53520
tgtcttagca	ggcgtgttga	ccatagcaac	tcccctgggg	catacacacc	ctcttgtaga	53580
tggagacctt	tgtccaaagc	agccacagct	ggcaactgtc	tacaatcttt	tgggctttct	53640
gctgtgtcct	aggggatctg	ggaatggcca	ttgcctagag	gggatgggct	ggtggaggaa	53700
ggtgggctct	gggagccggg	gagaagggaa	aagccatgaa	tttggaacaa	aggacaaatg	53760
tggtttacat	ttgtgaaata	cttgaatgct	tgtcatgaat	ggtgactttg	gttctatgag	53820
tcagccctgt	gatgggggtat	ttctgcagtc	ttcacctgac	accaggggtg	agaaggagga	53880
tttctgggga	ggaggaaaga	gttgaggggag	ataggaaaagt	agagtggaaag	aaaggccttg	53940
cgttgttgac	ctctatccac	ctggtcacct	atagtttttg	ggattgagga	tgcatacacc	54000
ttgagactac	aaatttatga	ttatatTTTT	gctgaacata	aggcaatgtg	ccaacacaaa	54060
ccagctgttc	tttggctggg	acagtgtgtc	tttgtttgta	aagggtgcat	tctgaatggg	54120
ggctgataca	tcatttgggt	ctttgtacag	ttaaaccattg	gccagagggt	ctggttcgtg	54180
tttagagtgc	ccgatgaagg	gctaactttt	ctccagacac	ttggggctct	tgttcacact	54240
ttgcttttca	ctctttttaag	taagacatag	tcacatcaca	gtgttttcac	agacatgttt	54300
caaaataatt	gtctaaggat	tgtttcttaa	tttccccgaa	atttgggaatt	gttgtaaactt	54360
ttgggcccaag	ctatttcata	attattttcta	atgtctcgct	tgaagaatag	ggatgtatct	54420
agtgttgatt	attaatcatt	cgaaactaca	actttacaga	ttgctaagaa	gaataaacttc	54480
ttccagtacc	catatggggc	agaatcttca	cgtgggaatt	cagagcattt	tgttggacta	54540
ttttaatctg	atttgattat	tttcatgtgg	tatgtgggtt	accacattag	aaacgattga	54600
tgtgtagaat	aaatgttctt	aacaagtggg	ggtcaactta	tcaaatgata	tttacattaa	54660

gaatagactc	cacaaatfff	agttcctgta	gctgatatag	catctcattt	gttatataat	54720
ccagtgattc	ctaattctgtg	ttcagaggag	agaggaaatc	gattgcaaca	gggacgatgc	54780
cttcattggc	tggcccaaaa	ctgggagttt	atacaaggcg	tcagtctttg	ccttcctcct	54840
ccctgccttc	cctcttccct	cttccctccc	catactcccc	aacaaattca	tggacttctt	54900
aacaactcag	agacattagc	cacaagttcc	aagacacccc	cacccccccag	cctccccagt	54960
cctatffffc	cattcatata	actaaactct	ttttctttct	tgggtggagt	ttgaaattta	55020
tatffffaat	tctttgctcc	cttttttccct	cttacaaaat	gagtgcccaag	cagctaagtt	55080
gtgctgagt	gtagagtttg	agtcagtcct	ggctggtaag	ctgtgggggt	aggagccgct	55140
ccctggatac	cacctctggg	gtctttgcta	tacaaagact	ttcatttagc	ctcctttgta	55200
tccagcaaaa	aaagattcag	tacccaaaat	gggtggtatff	tgggtatagta	tgtatcttac	55260
aaaacggcaa	aagacttcaa	aagttccctac	aatfffftatct	tgggggtttc	cttttgaagt	55320
cgatgtagaa	ttttaccttg	gggtggatff	tttgtacttc	ttgggtctgg	gtgttttggt	55380
gtgtaatgag	catggagggt	tgggataaga	aagcagactg	aatccccagg	aacaaagcct	55440
gccagactgt	gggtggtgtac	ttttcttggt	gttattgctt	aaatgctgca	agagagtggg	55500
aaactcttac	gaaataatgc	acgatgggta	gaacttcaga	gaaaatctct	gccgtctacc	55560
ctgtgcattt	tcgaggaagc	tcagagggca	tgctgaacct	ttgctttttg	tttctgaaga	55620
gttcagggga	acctacccat	aattaatfff	ttaaaacact	acctagagag	caccctcttg	55680
gttattaaac	acatgcgctg	tttcgatggg	atgtttgacc	tggattgtgg	atgcttgctg	55740
ggacgtggca	tgtgttggga	ggctctgtgc	tgctgtctga	gcaccagcaa	agccacagt	55800
ggccctacct	ctgtgggagg	ccctgtgcc	gggtgccctca	aagagtaggg	ggcccatgag	55860
ggtatgacca	gggggacctg	atttcggctg	agaagttggc	ggggattaca	ggcctgggcg	55920
gctccctgag	gaaattgcat	taaaaatgag	atctgaaggc	ttgattgggg	ttggcccaat	55980
gaagggatag	gagaagggat	ggggagtggg	cagaaggaaa	cacatgtgtg	aaggtcctca	56040
agggaaaagt	gcttggcctt	gacagaggca	ggaaatcagg	taggaggcta	gaggtcgggc	56100
agggtccgg	gagagtga	tgggggtgcag	catatggtga	ggatctgaca	ctggggagtc	56160
atttgagcag	gttggctgtt	tctgtaggag	cgtgtgttaa	gctgctggca	gtggggatgg	56220
tgaaaataga	gatgtggagg	aaacagcagc	ggaacttgct	gacaggttag	atattggcat	56280
tgaggggagaa	aggagagtca	aaggtaggta	gatggagatg	cttcaactgag	tggggagtat	56340
tggaggagga	gcaggtttgg	gggtggaagcg	ttgtcctttt	agagagattg	tatttgccat	56400
tgattgattc	attcattgtt	tctgcaaata	tttagtgtgg	gaaaaagcat	gctagacacc	56460
aagagagagt	ggagtcaatg	aagaacgata	acagcaacaa	agactgtagc	gcttcctatg	56520
cgaggcttgt	tccagttgct	tcagaggctg	tgttaccctt	gttctagaga	ggaggaaacta	56580
ggcccaggga	gggtggggatt	tgcccagtcg	tgggagtcag	gatgtgaaac	aaggcaccct	56640
ggctccagag	cacaccgtcc	tctcaaccac	tgcagagaag	ctgggaaaga	gacaaataag	56700
tgggtgctta	gagcacaatg	tgtgtgggtg	gccaagagca	gctgggagcc	ctgggacccc	56760
cagggaaccc	cagccccacc	tgggcatggg	gggcatggct	ggaggaggcc	tgtgtgcttt	56820
gctggagagt	gggacatgca	tcaaggtggc	cagagactgg	gcttctgggt	gtcgtgctgt	56880
gactgtctga	aagggtcat	tgacatatgg	tggggagggg	cagcgtatff	tctgcgggca	56940
ggacatttgg	gggatattgg	gtgtgaccct	gtactatcta	aaatctttta	cttctggatt	57000

atctccactt	tctctactgc	atatatactt	tgttttttatt	tatttttattc	atttatctat	57060
gactcagcca	gactctctaa	aagagttgac	tigtgtttcc	tagcagccac	tgagtcagaa	57120
ctttccatt	tcgcagtcag	ggctgtggtc	aggggtgtctg	tgttgtctaa	ggatataaag	57180
caagccttcg	ggcactacca	aaacattatt	ttataaggag	aactatgagt	acctaatagg	57240
aagaaccagg	caatcagggt	atcttttggg	gaggaagaag	tggtagatgg	gatcattggg	57300
gctttgaagg	gagtgggtgg	tgtagactcc	aaagtgtaca	tggggccatg	atagagtcta	57360
tgtcagatgt	ccaaagcttc	cttctctcct	cccagaaact	ctgtcctctg	gtgaagagtt	57420
ttgaagtttc	ctgaggtttg	ggttcatggg	gtggcagggtg	ataccatggc	aatagaaaat	57480
atcccatcaa	gaaggattgt	gtgacctcag	ttgtagcccc	tgcattgttg	aatcacaaca	57540
atttgcaggg	ccttaaaaac	aaatgccatt	tcaccaactg	ccctcccccg	tttttttcag	57600
cactgtttgg	tagctatctg	tttccccga	tattcttggg	cacttccaga	gatgggggct	57660
ctatctcctg	gtggtagact	gtttcttttt	ggtacaatat	gaactcttaa	gagagttcta	57720
cctttagggg	gctgcagtct	ctctcctgga	aatgctcaac	tccttaattc	atgttttgct	57780
gttaaattct	gctaattgct	caccttacat	gtcttgacaa	tttgaaggta	gctattgtat	57840
tccccgcaac	cccaagtctt	ctcttcaaaa	tgattattaa	ttgtaattca	aatcatcagt	57900
gactgggtatc	ttagactact	taaggatggg	aattgctaata	tttgtattta	aaagttgtac	57960
ctctaaagta	agtgaatttt	attttttaac	gtagctttct	tcattcataa	agtttatgtt	58020
cattgttaggc	agtttggaaa	acagcccata	atctcaccac	tcggagatta	cattgtgaat	58080
aatttgggtat	atttcttttt	agaaatatac	caaattatcc	ttttttcctc	tgagtgtatg	58140
aatattttata	tttgtttttta	acatacttga	gctcatagtg	ctcagtattt	ccaacgttct	58200
gtttattttta	gatgaaaatt	gctgtagtta	ataagcactt	ccccatgtca	ttaaaatgct	58260
taaggattttt	taatgaccac	ataacagtcc	ataatatgat	taaaccccaa	tttactgaat	58320
caatgccata	ttgttgggtc	tttagattgt	ctccttttgt	ttctgctact	gtgaatgac	58380
ctgtgatgat	catcttttgt	tgtaaaatct	tgtccccctc	ccccctcccc	ttttattatt	58440
ttcttgggat	agacccaggg	acaaaaggta	gaaaagaaca	aagtgtttaa	aaatttcttg	58500
atacatagcc	acagattatt	ttcctgaaag	ttctcaacat	ttataactac	gagcagtatg	58560
taagagagtt	atggttggaa	tgattttta	gtctctgggg	aatttaacaa	caaaaaaact	58620
ttaggcttct	ttggagagag	acatgccctt	aactccaccc	cgccctagaa	cagagaccca	58680
gcccatecaa	gtcagcctcc	ccaggctctc	caccttcaaa	acaggcaaac	gaaatcattt	58740
cttgaataat	tggtaggctt	caaggtcaga	tgtttatttt	agataattca	cagcataaat	58800
ttatatgttt	taggtacctt	agcccctgaa	tatactcagt	tcatttagga	ctattttaga	58860
ggtcttgagt	ttactcttat	aacctcacat	ttttttgtga	atttttagtt	ctattatctt	58920
tgttttcatg	gcatattatt	gggcaaagat	actattttatt	cgatgctatg	tgtgagctgg	58980
gtcaggatta	tgacctgag	ttatgtttct	gggaaaatgt	accacttgt	caaagatgcc	59040
gttggctcct	gtgattaagg	tcagcccaca	atgaatgtgg	ggagggtctg	cagcctctca	59100
aatcagctct	tgaccatttc	tcaagctggg	gcctgtttgt	cttgggggaa	gagtcttttg	59160
cagctcagct	cggggctagc	gtttcctgac	atttgtttcg	ctgaatgtta	acaaggttac	59220
tggaaaaaag	ggttctctcc	taaaataggt	ttagggaagc	actgggatat	gcgaagtga	59280
tgagtttctt	tagggcagga	tcttgactct	gcagggggct	tggaggcctt	ccctagagtg	59340

```

gggcttccta acactgcaga gctcttccca ggacgagggg caagattggg acctactttg 59400
gaaggttggt tttgtttcgg cacctgctct gtttacgaag cgtgggagcc tgttttaaata 59460
taatgtgcgc ctacttagag ctacactcat ggttttgact atgtttatct ttccagtaaa 59520
taaaacaaaa ttgttcattt ggcacccagc ctgtcctgct tgtcatttct tgtcttgctg 59580
attaactcta tggatggggc atgtttctcc aaccagattg taagtttctt gaagccaagg 59640
agccctgtgg ttgatttctt cacatgtggc tctctctcct cccacaatgg tgcttcgtta 59700
attaagcaga aaacccatct ctggttaggg actggagttg atttcgtttg gaatgagtgt 59760
gacttcatca tgacctgaaa gtgttcagaa ccatcttggt tagcacaagg gcgtggacgt 59820
gtgtctactt tctacctgat gggatagcat gtttaatttg gggttatgac actgaatgg 59880
ttgccagtaa ctgtctaate caaccttata cattccagct cacagtggag cgtgtctaata 59940
tgccacagca gcatttatgt ggaacgtggg tgcacaaaag ctccagaaaag tcaggctgag 60000
ggctcctatc tctcctcaat ctgggtttac gatgtctgtt tctgaggaat cctgggatgg 60060
ggccactggc tctttaagag agagcccgat ttggaaatct aggacttgat tgttgattat 60120
gggcaataga tacattttta gaatgatgtt gtaggctgta tgaagtcatt tgatgattgt 60180
tttgttaatg gcttgcaggt cagattttca tctttttaaa ttaattatca tagaaggaga 60240
aaacaactgg atttcagaat tgtcccttga ggtgtactgg aaactaaggc gtgagggact 60300
cataggggtc tggcttggaa agtgtattgc tatgtccagt ttacacataa ggatgtgcaa 60360
atccagcagg ttagctgagc tgcccaggaa tatccaggca agaataacca tattctgata 60420
attactcagg cctctgcctc atctccgctg ccccccgcc ccttgactct cttctgagt 60480
ccagattcag cctccatttg aatgccaaat agacaggaaa ttagcatgcc cagaatccac 60540
gtcttttagtg cactctctcc ccagctccaa acctgttact gcttgtgttc aacatctcag 60600
taaagctcaa caacategac ccattactta ggccctcaaac cttgggtggc atcgctgatt 60660
gctcttttct ttcatacccc acattcaacc catcagccca tcccacaggc ccaagtgtgt 60720
cctctctacc ttcaaagcgt gtgtggcata caccgcttat caccacctct gccattacca 60780
ctggagtcca gtgcatcat ctctcacttg gatgtggcca gagtgtcttt gctggctctcc 60840
ttcttgcttc ctacctttgt aacagcctat catctatctc tggctctccat agctcactcc 60900
catactttga gagggccttt gaaagcctta gacagatcat atcacagacc tctatactga 60960
aagtcgggat aaattttatc tctggaaaga gtcccaaaagc agcgatgaac agatattttg 61020
tcctgtcact tgatgaagag gtggggcttt gagaccaag agcttagaat ggagagccta 61080
gatgccacta agcccaggca ctggccatgc ttcgagtgga gcttttgtgc tgggtggagga 61140
gagatggctg ggggacacct gtaggctgag caagtccccg ttcacagac cctggctcat 61200
ccagcagggc gtggctgatg ttttcaatgt tgtatcctga gtgggaccca gatgcttccc 61260
aactgtgcca catctgagcc ctgcatgcca tctgtccagt tgcagcctga ctgcaatgtg 61320
aggctgctga agagctctgg atgggttgaa gcaatctgtt ttctagcccc agcctgcata 61380
gctggtggat cctggaccgt gattaagtgc atcacctagg cttcaatgag atggagtcac 61440
tgtgtgtcca aacagtggga taaaggcttt actctttgtc ttctgtctct gagggcacaa 61500
gctgcttggt tctctcacia ggacaccgtc tgtgttgctc aggtgctggg gtgaaaaaaa 61560
cagcaagcat ttgaaaaggc tgaagaagga aagaaagctg agagcggtac agccttgggg 61620
actgagccat cccattgtcc cagagggtgg ggtgttatca agacctgttt ttgagccata 61680

```

cctctgactc	ttcctggaaa	gtagacca	actcaagaac	acactaagag	aagtgtttcc	61740
ccctagccct	ttcagattga	aaggagacgc	caaccttgat	gggtggaggt	agaaaataaa	61800
gtcccaaaac	agtgtcttgt	aagcgaaggg	gaacatggct	gggcagaggg	cttctgggtga	61860
aacttttggg	agtattcagt	tggaaactcag	gaaaaaaaaa	ttgttttttt	ggaaagaggt	61920
agcagccccc	ttcagccaaa	gctcataaat	gaaggaatgt	ctgagactca	gaattacagt	61980
gaccaaggca	agacattgtc	aaaggctgaa	taagtgaagt	tgactgacag	aggccatctc	62040
catttttagt	atatggccaa	gcacttttcc	cacagtcttc	cttgagcccc	ttcccatccc	62100
acttctgaaa	agcactgagt	tggccattat	tatgcttttt	tcttaaatta	tgaagttggt	62160
ttcaggtatt	gagaataaca	cccaggtgct	gaactcccag	cataagaaat	caaacattca	62220
aaatggagta	aggttctgaa	gctgacatct	gtctctacac	atTTTTTTTT	ttctgataat	62280
ggcatttcct	atctccaccc	tcactctttt	tgttgtgggtg	aactacactt	cccttgttcc	62340
actcggttct	gttgccatgt	tgattaggca	aggggcagat	atgtgatatt	tattatgagt	62400
cttttccacg	cagagaggat	ctaaatctgg	ctctttgcaa	ttgccctcat	acatgtgcat	62460
acacaccaca	cacacacaca	cacacacaca	cacacacaca	cacacagaca	catacatatg	62520
cacacacccc	gactcaatgg	aggaccctca	tttgtagaag	ggtaaaatgg	gtgaggcgga	62580
aatgcctgta	tggcaccatg	gagttctgtg	tagccagttc	taatcctggg	ctatttggtga	62640
aggaatgaag	ttggagatag	tcttctgtcc	cttacaacca	aaggaattct	aactaatagt	62700
ttgccaaagt	ttatgtttat	aataaaaaat	gacatgcttt	ttcttttggg	tttttaatgc	62760
ttttgaatta	aaaatgctag	aacatgaact	gattcttcta	tcgctattta	gatagagcct	62820
tgcaagagca	gagcacgcat	gctttcttta	agaacagggt	ggtttgtggt	cgtctgagga	62880
ctgttttaag	gagacttatt	atacacaatc	atccccaca	aatgatttct	aaagagaggc	62940
tggtatgaaa	gaaggagttt	ccatgattct	gtcctgtggt	tctggggaat	tctgaaaatg	63000
aacttttagat	atTTTTTgtga	aattcttatt	ttcatatttt	tggtatctca	gagttttctt	63060
ttctggcttc	tgtttaacat	actcttcttt	gccctaaatc	tctcttattt	ttgtcccttg	63120
ggacaactga	agaatcctta	gataattaat	agtatgaaat	actgcccttt	tagttgaaaa	63180
atgtcacaat	aatgtaataa	gataaataag	gagggtgtgc	tttaacctgt	atcgtgtagt	63240
ctcctctact	tactaacact	tacttgtatt	actagaagca	ttatttttta	aatcatggaa	63300
aattgggtggc	aagctgagca	tacagttgtt	tatttctgtt	tgactgatta	ttacaacttc	63360
attatttgat	gaaggttctg	tacgttttcc	tttaagacac	atagaaattg	tgagaagatc	63420
ctgcagcccc	gaaaggctac	agtgttgatc	caaggactct	gagccgagtg	cagggtttgt	63480
acttggacct	gcaggctggg	tggcgtctgt	gggagcagtg	tgttgagaga	gattctgagg	63540
ctgtatgtgt	cagggcctcc	aggggaagga	tgcattgatg	gattaatttc	tgccaaggct	63600
gaaagaggag	agagtaagag	gctgtagagg	gtcacagct	gtcattgctg	ttttaggcag	63660
tcaagctttt	gggaaagtgt	cagaaattga	gccccctact	ggatctatcg	gagccctgtc	63720
aaatgtccat	ttagatgtcc	tgggtgaacaa	aagtctctgt	actcaccatt	taaaaaacttg	63780
ttccaaatga	aattatggga	gaaaggaaca	tttttcatcc	gaaccagaa	tgaggatgta	63840
cccaaggaaa	aggacgtagg	ctcaggagct	ggactgtggc	tcagctggcc	tgatgtatcc	63900
cactttgttc	ctcccatggc	tgggatgtct	ctttgtcttc	catgacccat	gtatcttgag	63960
gacatgacac	atggaccaag	cttgaactgc	ggattcattt	ttatgcattc	tacctgtgaa	64020

tgattgcagc	ggatctagtc	gtatttctga	gagttactca	aactggactt	cagcagtga	64080
ctctacagtt	ctcttttcc	cccaccttc	tattagacat	tgcatgatac	aaaaatcaag	64140
atatttctaa	gagggtgata	acttcaatgt	tatctaaact	tttaatttgg	aagaagaggg	64200
gttctttgtt	ctttttaaaa	agatacaaac	gaacttcttt	atctgattct	ttttttggtg	64260
caaaccatg	atgccttctt	cctgattcat	ctgctacact	gtgagttcaa	gcctggcgtg	64320
ggacacaggc	acagctctca	tgccaacgat	ctcatgggta	agtttttgaa	cataatttga	64380
aaaatgtaac	ccattgagag	gcagtaagga	catacgggta	gctagtgcgt	gtttggacgt	64440
ctgtgtggaa	taagtgagtg	ggtagagagg	acatttgtca	aggagcggga	gggcgggcca	64500
ttggcttggg	ggaaatgggc	tgagactcta	gggggtggcca	gcaccgcata	cggaggccag	64560
cagggttggg	cttggctaag	tgctgtgggtg	tctggatgcc	tatgtgagtt	tcctccagaa	64620
gttttcagtt	ggcaaagtag	aacctgctgg	atatgtagca	agggtgtgga	ttgtcgggat	64680
cctgctgggc	gcaggcgtgt	gataccagag	gtcagaacag	aagctgaggg	atgaggcttt	64740
gggagctttt	tgtcatgcac	tgtcctggag	cctcagttac	tacaaagtct	gcaaataata	64800
gaccggagct	ttggttctgc	ctgatgctag	ctccctgttt	cctgattttt	cttttcaata	64860
ttagacttaa	tcccagaatt	cacatgttga	aagaaaactt	agaggtctag	tgacataaaa	64920
gcctcatttt	gatcgttaca	gaactgatgc	cttgagaaat	ggagagagaa	gtacacgac	64980
atggtaatac	tggatgttca	ctgagcactc	actagctcca	ggccttttct	aagtaattta	65040
tgaagtgtgc	aggtttaatc	ctcacaacgc	ccttatgaat	gagctattgt	tattatccc	65100
atttggcaga	tgaggaaact	gaggcttgag	gggaggatga	cgtactcaag	gtcacacagc	65160
tgggagcg	caagctggaa	gttgaaccca	aggagtctca	catcgagacc	aggactctca	65220
cccttcagtg	ttatgctgcc	ttaatcaggc	acacatacag	gcggggagag	gcaggtttcc	65280
ggacaccaga	ctaggctggg	gccggtcagg	ctacaccagg	gaacctggag	gcctgtcatt	65340
cttttgtgat	gctgttagtt	cctgttgagg	aagtgaggct	ttgtgggttc	ccaggaggaa	65400
aaggtatgaa	ctcatggcaa	aagaaaggaa	ccaaaaagg	gagatttgca	tcacaatgag	65460
ccttctattc	atcctaaatt	atacctcctt	ttataccatg	tgtgtctgca	aacttgtggg	65520
taaatcacia	atctttctgg	taagttacaa	tggatggaag	gtttttgcat	ttctctcaaa	65580
tcaccaacca	tttaatgcta	tgtgtagtca	ctccctaate	tatcttttgt	ataaatttgg	65640
atctttgagt	attgggggtt	tccatgatgt	ttggcagttc	cccttagggg	gtctatctca	65700
aagtttgtca	cactgacaag	ctttggggag	agaagttaga	ggtgggcttc	cctgttttta	65760
gtggctgtgt	ctgattgttc	tgtctgttct	ccaggacagg	agagattgat	tgctttctag	65820
ctttttttta	aattaaaaca	acaacaacaa	aaaaatacag	aaaggtacaa	aggataacaa	65880
acacattcat	gtacctgcca	cctaaaataa	caattactaa	tcttttcacc	ctcctagccc	65940
atgatcttcc	ctccaggct	gttattaata	tgaaaaccga	gttcagggtt	ttatactttt	66000
cgacatctat	ttatattaac	gtatgtatta	taaataatct	tagtagtttt	taactttgac	66060
ataagtggct	tcacattcca	cataacattc	tcagcatgt	tttcttttat	ttttattttt	66120
ttctttat	ttaaattttt	attttgcagc	atgcttttct	tattcaacat	tacatttgaa	66180
ttttttcaac	attgtacatt	gaaatttagc	tcattctttt	taactgctct	gtagtattta	66240
ttgtatgcat	atactacagc	tttctatttc	tgtattgatg	gttaattagg	ttgcttacag	66300
ttttttaaga	ttacagattc	tgctgtaata	accatccttt	gggcaagtgt	atgtaggtac	66360

ctatatatga	gtttctctag	gattcatacc	aaagtagagg	aattggtagg	gcattggttt	66420	
gctggtttta	attttaattc	acatgctatt	gtcaagctct	ccagaacaac	tgatgagtt	66480	
gattggatca	atgagtat	ccatcaccag	catataaact	ctttcctcat	aatcacacca	66540	
atgcttgatc	ctgttggact	taaaat	gtccaatttgc	tgggtatgca	acggcatctt	66600	
acctaatttg	cctttatttg	atgactcctg	aggttgaaca	tctggtcata	tgtttat	66660	
ctcctctgtg	gcttgcctgg	tttaatgcct	tcttcatttt	aaagaatcag	atagt	66720	
gttattgatt	tataggaact	ctttatataa	gttgaaaact	tgattatatg	tgttggaaat	66780	
actttttcta	ggctgtgatg	ttttaaata	ttgctttaga	tgggttttca	tttttacctt	66840	
ttat	ttttaga	gatggagtct	cactgcattg	cccaggctgg	attgcagtgg	ctattcacag	66900
gaaagagcat	agtatgttac	agcctccacc	tcctgggtacc	aagaggctcct	cctgccccag	66960	
cctcctgaat	aggtgggacc	acagggtgcac	atcactgtgc	ctagcttttg	atgggttttg	67020	
aaagaaagaa	gttttaaat	ttaatgcct	caaattcatt	tgtattttcc	tctgtgcttt	67080	
tattttgtac	ccactctaag	tagctccgaa	ttctgcagat	agttgggtgca	ggaattctga	67140	
ttttgagtg	acatctgctc	tctaacagtc	acattgaagg	aaattaggtt	tttttggtag	67200	
gaatctaagc	aaggggttga	tttgtaaact	aggctttaaa	tatgatttta	agcaactcac	67260	
ttagaacaag	atacaaaaat	tgtggactgg	acctatatct	ggaaaacttg	aaagtgcag	67320	
ggcaataaat	aattcttgg	cacatacagc	cgagatcctg	ggctcctgac	tctgggacag	67380	
aagctttcta	tattttatct	catcagctct	tgcaacaggc	tccttgaagc	aattttatcc	67440	
ccatttttaga	gataagaaaa	ccagagctta	aagcagttag	ataatttatg	aagtaagtgg	67500	
cagagccaag	attcaaatcc	agacctttct	gaccacaaag	ctcgttgctg	aataccgcgc	67560	
ctcattgcct	tcttgcgaat	tacttgggat	ttgtttgaat	cccaaatct	ttatatgtta	67620	
ttttaaat	gaatctaatt	ggaagtgggg	cagtggagggt	agaggacaga	aagaagggga	67680	
agagcttgag	actcaataat	agaaacaaaa	aaccgcgtct	caggaggcg	gttcaaaagg	67740	
aagaattcca	tatttcatgt	aactgaaacg	ttaaaagccc	aaataattgc	atcatgcaag	67800	
tctgatgctg	agtaatcacc	ctccccata	ttattgggga	gagggggcaa	gaagtctggg	67860	
aagctgtttt	tgcctaagga	attacattcc	aggggactct	gaggatttag	gtaaccacaa	67920	
aagccattta	tttcgagtac	actgagattt	ctaccacttt	gatcccta	ccatagcata	67980	
attaataaat	gaaatgtgct	gtagcatggg	ttttttacaa	agtgtacttt	taaaatggct	68040	
tttggcttga	catgattcat	ttgccacttg	gaaaagcgtc	atcgccctcag	atgggcaggc	68100	
tgggagaggc	tgcctgggtg	gtagctgagg	gcggtttcct	ggggcacagt	tcctgccttg	68160	
ggcctctaca	gagcgggtct	atccaaacat	ctcccagact	ctgcgttttc	caggaagcgt	68220	
gcagaaatag	gaggccagta	ctgaaatgct	atctgctctg	tgtatgtcag	aagaccacaa	68280	
accacttata	acaaatgaag	atctttttat	ttgttcttat	ccctttatgt	cacttgagga	68340	
aagtgtctgt	gagtaggtga	tgatcattac	agtgatcact	ggttgcccaa	actgagaagc	68400	
cagacatttg	gcttgggttc	tctcccttcc	tcttgtctct	cctaccctgt	aaacacatac	68460	
ttgggtgatta	cccatgggga	gacaagacag	gctgggaata	tatacttctg	caacttcagc	68520	
ctcctgggtt	ccagegattc	tcctgcctca	gtctccagaa	gagctcggtat	tacaggtgtg	68580	
caccaccagg	cccagctaac	tttttgtatt	tttagtaaag	atgggggttc	atcatgttgg	68640	
ccaggctgg	ctcgaactcc	tgacctcagg	tgatctgccc	ttctcggcct	cccgaagtgg	68700	

tgggattata	ggcgtgagtc	accgagcctg	gccccaggca	ataatatacc	agtgggcaag	68760
aaaatattct	tgctctcatg	ggacttctgt	tgggggtcag	ggtataggga	ggaaggcata	68820
gagatgaaaa	ccagtaaata	agtaacaggg	gaaaacattt	taaatacatt	aataactaat	68880
aaaatagaaa	taaatctgtt	ggctacttaa	caggatgtgc	cacattccag	atacattacg	68940
ttaatcctta	tgatcttttg	gggctaagta	ttagtattcc	attttacgga	tgaagagact	69000
gaggctcaga	gggaagggag	gtggcttttc	tcagggtgaa	agccagacct	tttcagtggg	69060
cattcagttc	atagctaagg	tcttattttc	tgtgctctct	gtcggctgaa	aatgggcaag	69120
gtaatttcac	atagtacag	gagccatgtc	agagaaagag	caggacagtg	ggacagagag	69180
ggaccaggct	gggggctgtt	tgagatggag	ggtcaggaag	aaccaaacta	agatgtgaac	69240
agtgggaggt	gttggagctg	tggtgcttgc	ctagaaggac	cctcatcgag	caaatagaag	69300
cttctggcag	gaagaagtta	atgtcttgcg	tgtgccctat	gtaggttcat	tagggccttt	69360
aaagggggaa	gaaggtgggtg	gctataaatg	ttacaatctt	acctttggcc	cctaggggatt	69420
ctgtctttca	accttgggtc	agtaacaact	tgtgactgcc	caacagggct	tcctttcggg	69480
agagaatggc	ttgttacatt	caaatatgcc	atgaaagtat	caccatttat	ttcagtgtct	69540
gatgccccag	cttgggcagc	ctgagcaggc	tctgaatggg	tctgaagagg	cccttttagag	69600
tagagatgaa	gaggggggtg	ggaatcctca	attctaaaca	aagagtctgc	aatgggaaga	69660
tggccaaatg	ctgttttttg	agtgggtgag	agggaaaaga	aaggtataga	tggttcgttg	69720
gaaaatgtgg	ttttataacc	ggttttgggtg	tcagggtccc	gagggcaaca	tggactccac	69780
actgtgatcc	tccgggcagc	tcatagcccc	agcccccttc	ttttgcttcc	tggtcagttt	69840
gtgagaagga	gggggttgtgt	ctccaatctg	agcaataagg	ggtctgaggg	gggttggatc	69900
catgtggctt	tcctgtgtct	tgttccttgt	aaaagttcca	ggttttgggt	cgtgagctgt	69960
gtgtgtgtgt	gtgcgtgtgt	gtgcgctgta	cgttaatatg	gagagatggg	cttggggccag	70020
tgggaaatag	agagaccgc	aagcacagag	tgacagggtt	tgatagtaag	cagcaggcca	70080
gcgttgctgc	ttttattcct	cggtaaatec	ttgcacaatg	ccatatgctc	ttgcattccg	70140
tagctgctgc	atagggtgtg	atttagttaa	tgcccgctct	gcaaacagga	aacggtgctc	70200
actgctgtgt	atgcttttca	tggagataaa	gtgtcaggag	caagacccca	aacctgcgaa	70260
atcactaatg	caaccgcccc	ccatgccccca	aaaggtggga	gtgggggata	aaaagagtag	70320
gaaagtgggtg	tggggagggg	aagcttttag	gccataactc	agacaatttg	tcaggcagtg	70380
gcategggtg	ggaggaaaat	attgatgtac	actttttgtt	tttgaacctg	aagtttgggt	70440
tttttcggat	gcatttgagg	acttttaaat	gttttcggag	tgccagagtt	tggactgtta	70500
ggtcaccgta	ggtaccggct	tgcatatcat	ttcagaggaa	tattttcaaa	actccataaa	70560
aacatgcggc	tttcaaggct	ggaccacttg	ttcaggctcct	cctccccacc	cccacccttt	70620
ttggcaaaac	catgcaaaac	ttgggtattca	aaaatatatt	gttacttttc	ttggcaaaagt	70680
gttccaagaa	ggaattgcaa	cacagtctca	gagttaggag	gcaactttct	ggggaaaagg	70740
cgggggttgg	ggagggttgg	agtttgaatc	aaaaacagac	accgaagctt	taataaaaata	70800
aatgaagegg	agccctttca	gtcacagggtg	gactgtgttg	gtgcgcggtg	caggttttaa	70860
cgtgcctagt	ggaaattgac	agtctgagaa	ctgggacata	aacaaaaaatg	tcagtcctctg	70920
ggagtcttgt	tacttggaca	atgtctcaat	tgttcctttg	gttttcaagg	cagcaggggag	70980
agtggaatat	taactgttta	ctgccc aaag	ctggctcgga	aattgcttgg	agaagggggag	71040

aaaaaagaca	gaaaatcaca	ttttttatatt	agaaactatt	aaacatgtca	gtaagagata	71100
ggaaaagagc	agattgtttt	ctccttaatt	atctgccatt	cacttccata	tttctgcata	71160
ccatttttgg	ggtgtgtgtg	tgtgaaggaa	cagcaggggtg	tttcttttta	aatttgaatg	71220
ttagccttgc	atattgtcag	tttttaaagc	ttgctggcat	gtagattatc	cgcccccggt	71280
ggatatgaca	gtgggcttta	ggaaaggaag	tgtgatttct	gataacattt	acatcttagc	71340
tgttcagcgg	ataccctgtt	agtgtttgtt	cttcagaatg	ctcagataga	acaaaaatca	71400
agtggttgga	atttttaaaaa	acaaaatgta	tttggctctc	cataaaaaatg	catttagtga	71460
taaagggggg	cagcaagtaa	ctatgtctga	gagaaggaat	tgcaggcaca	gaggagatcc	71520
agaattctgt	tcacacttga	atttacttga	ttcgagaaac	aaacagcaaa	gcctgggtga	71580
ttggccttta	tctgggcaaa	gttcaaaact	caactggtaa	ttatgtcctt	agaagcctta	71640
aaaggactgt	gttggttaca	aagcagtgac	tgagcttact	tcttcaggac	cgaatgcact	71700
cgagttgttt	gttagataaa	cttgttttaa	taaatggggg	ggtcagggga	gaggtttctg	71760
ttcttgggaag	attccctgat	aagtagcttt	cttctcttgg	agaacttcag	gctttctctc	71820
caagcgaggg	gtttgcaggc	agctaaagtc	agcttcggct	tctgcttcct	gtcagtcagg	71880
aagtcacttc	cttaacccaa	attacaagct	agagcacaac	tccccagcca	taccgaaaag	71940
agcaggtttt	tcccagaaga	ctgtgtttct	agatgcggaa	gtgtaaattg	gtacgctgtg	72000
tgatcatgga	atgccccaaa	tacataggga	acagtgttgt	tggaaagagg	cgctgtgtcc	72060
ccaaggagaa	gacgcgcgcc	agaatggctg	gatcgccctgt	tgtggctgag	tgcgaggcag	72120
ctgtggctgg	ctgctgtgtg	acgatgacct	agtagccacc	catgtggagt	cctggctgcc	72180
tcagaacctt	atcacatcta	ggcaaaatct	tgcatttttt	atctgggagg	cctgaggact	72240
tcagggtgg	tggatagtaa	gtctcttggg	tatctcacag	atacaagagg	tcttgggaat	72300
ccacgatcaa	acttgatgtg	tgcgtttacc	ctctctccct	tgaatctgtt	attcaaatat	72360
ttaagcctcc	aaccttgtgg	cccctacctg	caccacccct	cacccccccg	acaaaaatca	72420
agctcttgac	ctcatggctt	ctttcagtga	cccttggggg	acagggtttc	ccaaggctgg	72480
ttgccagctg	gcatgggtccc	ccgttgggtga	agtgagagacc	tgtgtttttt	tgggtcatttt	72540
gcaaagagct	tatggatgac	agcagttctc	tgtgcctcgc	tgggacagag	tgtattctga	72600
ggtccagcgt	ctgcatggag	atctgcctat	ccttcacttg	gggtgctcag	tagataacgc	72660
ggccactttc	ctatacattt	ccttaattta	agggaacagc	gtaaactcag	cccagggtgga	72720
ttaatctctc	cagtgcattt	tgaaaactca	atttccaatt	tccctcttat	gtctaggtgt	72780
gagtgaggat	acgtgtagta	attgtcgcag	gtatttagtga	gaaaggggtgc	agatcacaca	72840
aatatttcac	acgttatttag	ttggaccaga	ctttggaggc	aaggaggggc	cgtgtcacct	72900
aggaaatttg	ctcttccgtg	gagatgaaag	ggcagtgaaat	taagtgcctg	ctttttctcc	72960
ctttttccct	ctgacggtta	ttgatectcc	ccttggaactg	tacagttcac	gttctgatct	73020
ttttcttgac	aaagggaatt	cccagtttgt	tgcgtggcga	acgcactagc	aggtgaggag	73080
ttaaaaagtig	gcaacgcctg	ccctctcgag	agtgtcagga	tttttagtct	cttccttgag	73140
agctagaaga	tgtttctaaa	agaatctctt	tgggtgactta	gaagtggaga	gagctttaga	73200
agcatggcac	aaataaaaagg	aaagaggcaa	acaccgtcat	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 cttaaagtttg gagccacttg cagaaggctc	73500
tcctcaaacc cctgtgttct ttgcctaccc ctgctgttgc cacatcatct tggagagctg	73560
gctgcttccc tcctcaacta gaagtcccta gtgcctgctt agttcttgtc tcttgcttcc	73620
caagtgtca caaaatacat ccatgttcgc tacgaggaaa tggaccacat aaggtttccg	73680
tgaaaacctt agcccttagg tctaacacag taggaacaga agttaatgtt ttcctgacgt	73740
agaagtttct cttgctgctt ctggtcacat ttctttcttg tgtggttctt ctatggctac	73800
tgcacttttt ttttttctt actgtctccc ccttccccca cacaccacct tttggggata	73860
gggtggcagg tgagaatata aacagataat ggttaagaga tagtttagtc tttctaggcc	73920
agattattta gtttttgcca tctaggtaaa attcggcca attaacgctc cattaagtgt	73980
tttaataata 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 cattcgagga agcaatgcac	74280
ggaatgaatt ctggacctgc gaactctgaa ttccaattct ctgtctccta cttttactgg	74340
agtgtttgca aacagtacag tgtttttgtt gtgaagttat accgtgcctg taatctctct	74400
gcgggtggcc ctctaagcc ctacttcaag aaatagctct aagctcatga cccccgccc	74460
acccgatgcc tacatatgtc ttatatcctt ggagtagtgt ttgggggttg aaatttgact	74520
ttaggagac atactctctg atgataggct aatgcttata ttactgata aacttccttt	74580
ttgacggtca tgggcttcgg gggccacca accaaactgt gtggctgctt ttatgttggg	74640
ccaaaagaca ggctccttgt gtctctccag ttctttaaac aatgaagtca tggcatttta	74700
cagtgttgtt gaattgattg agattgttgt ggccctggaa tgtggcactg ctctggctgg	74760
agggaagatg agagtgaggg atggagagga gaggagagcg ggagatggga acctggtgga	74820
cacaggaggg agtgtgagtt ctgagggcc aaggaaactt gacaccggat gggacattaa	74880
tctgattctg ttatctgagg ctgtcaccag tcttccctgt cctcctggca	74930

<210>2

<211>384

<212>DNA

<213>Homo sapiens

<400>2

ttcaggccat tgggtgttgta tatatttcaa gatttgctca cagggtccaaa gcttaactta	60
agctccctga gacatatcat aaaatatgat ttggggaaaa accctaattgg gccatgatca	120
gaacattatt attcaacaaa ggatgaaatg cttaaagccaa gatggccttc tttctttctt	180
tctttcttct ttttttttta atgaaagttg agcagactcc cgtccaacag ttttcaatgt	240
aggaattccc acagcccat ttgattgcag ttgttgaaa agtttaaatgt tttttaggc	300
aattcataat ttccacattg aacagcctga gaggaagaga gctggagccc actgttgttt	360

ttgtagtggg atggtgggaa cttt	384
<210>3	
<211>20	
<212>DNA	
<213>Homo sapiens	
<400>3	
ttcaggccat tggtgttgta	20
<210>4	
<211>20	
<212>DNA	
<213>Homo sapiens	
<400>4	
aaagttccca ccateccact	20
<210>5	
<211>366	
<212>DNA	
<213>Homo sapiens	
<400>5	
ttgtcccttg aggtgtactg gaaactaagg cgtgaggac tcataggggt ctggcttgga	60
aagtgtattg ctatgtccag tttacacata aggatgtgca aatccagcag gttagctgag	120
ctgcccagga atatccaggc aagaatkacc atattctgat aattactcag gcctctgcct	180
catctecgct gccccccgc cccctgactc tcttctgagt gccagattca gcctccattt	240
gaatgccaaa tagacaggaa attagcatgc ccagaatcca cgtctttagt gcactctctc	300
cccagctcca aacctgttac tgcttgtgtt caacatctca gtaaagctca acaacatcga	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 tgttggtgag

20

<210>8

<211>558

<212>DNA

<213>Homo sapiens

<400>8

gctgtgaaat cccctgtgta gtgggaagaa gaaatagcaa atcttagctg ccttggacct	60
gatataatta ttigtcttca ttacatggt tyatccttca aggttgaata aatgatgtgg	120
gagctagtca aggggcttta ggtatgtgat ttcattgccta ctttttttta ggtagagaaa	180
ctgaggtcac agggtagtag agaattggact ctaagattca gggtttctgaa ttgcctgtgg	240
ttttgttgac tcaactgctc ttctgttggt ttttagccac atgccttgaa acagtcctct	300
ttcccatggt tcttcatcag caccattaac ccaaggtata ctgtcctctc ttatctttca	360
caaggtcttg gagttcccat gcctttgtaa gcatccctcc ccgagattca gcaccaacca	420
aaatcacatt tggaaaaatt gcttgtttcc caagaagctt tggaggatat gattttgtat	480
agaacgggtt cacagggttt ctgttcattc ttctatgggt 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

aaggagaaaa gcaggattga gcagggggag ccgtcagatg gtaatgcaga tgtgatgaga 60
tctctgccgg accaaagaga agattccttt ttaaattgggtg acaaattcat gggctttctc 120
tgcctcaaaa cctagcacag ctgttattta ctgaacaatt agagagctaa gcacttttta 180
gataytatat aatttaattg ccgtatgagg cacccttagt tttcagacga gaaaccacag 240
ttacagggaa ggcaagtaac ttagtcaatg tcagataact aggaaaaggt tagaggggcc 300
ctggacacag gcctgtgtga ctgagaagct tgggcacttc actgctacat ttcattcttt 360
cgct 364

<210>12

<211>20

<212>DNA

<213>Homo sapiens

<400>12

aaggagaaaa 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

ctgatgaggg tagggagcat ctgtctgcag cttcatcttc attgtctagg ggctccagaa	60
atatctgtga gtaaataagt tatttaatct ttgcctcaaa tttccagtga ctgtagggat	120
atagctgtga gcctctagga gctgagatct tttaaatttc ccacttaaac atttatttaa	180
aaattttgtg ctgagcatgg actaaggact ttacattcat taactcattt acagcttgat	240
cctatgcggt gggcattcat ttacagagga tcccatttta caggtgagga agaggccagc	300
taggggtgca gcctagggtta gtattctaga gctcatcagg ctgtgttgtc cccagtga	360
gaataagcaa agaagtgaat gttgtgcatt gagaaaaatg actctcggag gaggatgagc	420
ctctcggata tggcgaccga agtgatwtgg ggcccttgtc aagggtctct attatggcat	480
caagaaaaga tgctgcttct ggtgatgccc gaggagagcc tcaatatttt 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 cacctatttg aggtgtacaa ttgagtgatt ttctgtatag	60
tcacagatct gtgcagtcac ccacaccctc taactccagg acattttcct caccctcgag	120
gagaaacctc ccttaccat tagcagtcac tctcatcttc ctctcccccc agccctggc	180
aatcactgtg gatttgctg ttcttgacat ttcataataa tggatcata aaatctaygg	240
gcttttgtgt ctgtctgctt tcacttagca tacggttctc aagggttcac cagtattgta	300
gcatttatca gtatgtcatt cctttttatg gccaaataat attttattgt atggatagac	360

atTTtGttta tTcattttatc tGttttttggT tattatgagt aacactacta tgaacatttt	420
gcacaaattt ttgtattgac atgttttcat ttctcctggg tatagtccta tgagtggaat	480
tgctgg	486

<210>18

<211>27

<212>DNA

<213>Homo sapiens

<400>18

tatttagaaa ccataaaatc cacctat	27
-------------------------------	----

<210>19

<211>22

<212>DNA

<213>Homo sapiens

<400>19

ccagcaattc cactcatagg ac	22
--------------------------	----

<210>20

<211>428

<212>DNA

<213>Homo sapiens

<400>20

ttgtctcctt ttgtttctgc tactgtgaat gatcctgtga tgatcatctt tgtgtgtaaa	60
tctttgtccc ctgcgccct ccccttttat tattttcttg ggatagaccc caggacaaaa	120
ggtagaaaag aacaaagtgt taaaaaattt cttgatacat agccacagat tattttcctg	180
aaagttctca acatttataa ctacsagcag tatgtaagag agttatgggt ggaatgattt	240
taatgtctct ggggaattta acaacaaaaa aactttaggc ttctttggag agagacatgc	300
ccttaactcc accccgccct agaacagaga cccagcccat ccaagtcagc ctccccaggt	360
cctccacctt caaaacaggc aaacgaaatc atttcttgaa taattggtag gcttcaaggt	420
cagatggt	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 tgcataggtg taaagaagtt gctggttggg ggttctaatg caggtttctc	60
caaaagtga tgccctgtta aaaaaaatt cttacaaat atacagagat tttttttta	120
aaaaagtgtg acagttctag acacctagag agtaaatga agaagcctgt tttcaggttt	180
cccgcctccc tgaatttccc agcatgggcc aggccttgaa atttatttat ctgcctttgg	240
caatggttga tgggaatttc ccacatttat ttttagcta cagagaaagg acattatctt	300
taaaatctct tcgtttgtct ctctctttga	330

<210>24

<211>20

<212>DNA

<213>Homo sapiens

<400>24

tcagggacag tgcataggtg 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 acattttcct ccccccgag	120
gagaaacctc ccttaccat tagcagtcac tcctcatttc ctctcccccc agcccctggc	180
aatcactgtg gatttgcctg ttcttgacat ttcataataa yggatcata aaatctatgg	240
gcttttgtgt ctgtctgctt tcacttagca tacggttctc aaggttcatc cagtattgta	300
gcactatca gtatgtcatt cctttttatg gccaaataat attttattgt atggatagac	360
attttgttta ttcatttate tgtttttggg tattatgagt aacactacta tgaacatttt	420
gcacaaatth ttgtattgac atgttttcat ttctcctggg tatagtccta tgagtggaat	480
tgctgggtca tataataaat aactgtttta cattttgggg agctgccaaa cttttaaaac	540
cttgggttct gtgatgtacc agttgtgtta 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

tgcctaacac aactggtaca tc	22
--------------------------	----

<210>29

<211>571

<212>DNA

<213>Homo sapiens

<400>29

tgccaggggt tttatggtta attttctcc attatgaggg ttgactcagc cttgggtatt	60
agatgtcttt gagaatccag ggttcaaata ccacagctgg tagaatgttt ctcaacttgg	120
agccaatctc catctactga aggtacgctg gtttagacag acaacaggga catcagcatt	180
ttaaaaagcg gtggaaaaag tttgcttgtc ttgattggag ccatgacatt ttattttgaa	240
atttcaaata acatgaaggg aggtttggag cggtttttgg tttatccaaa gggcagtgga	300
ttgaaggctg agaaacacca ggctgaatgg gagaggggtt ggggtcccc tgtgagatag	360
tgaacaatg gtagtgccat ccaatgatag gcacttttct gtcattcaga agcagaaagg	420
gggccagagg cccattggcc ttactgggma gtaagctgta gagctgctgc cttttcgtga	480
aagggttgac accaaccttc tccccagga agagtgacca 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 tacaaggct gtc	23
--------------------------	----

<210>32

<211>614

<212>DNA

<213>Homo sapiens

<400>32

ttggtagaga tggggtctcc taggctggtc ttgaactcct ggrrtcaage aatcttctg	60
cctcagcett ccaaagtact gggattactg gcgtgggcca ccatgcctgg cttgaaattt	120
ttctatggct ttattctttc tccaagtaca gagtctaccc aaccttctga gatctttggt	180
tttcttttcc taggtaacta tagtacatac ttatttatgt taaacaacag caatcacaca	240
tttcttttcc tatacagtea tgctttatag gcaaataaag cctccgtctt aggtttctg	300

gattttttca aaagatgcaa ttcctggagt atgtttttac ttagagcaaa gcagcctagt	360
ctcctataacc ttctgcatct gcagaaaagt tggttaaaca gactttgtaa tgatgcccct	420
tacaattctg aagggacttg tgaaatagtt tcacagagtt tcagtgttag gtatatattga	480
tcaatgctaa cttttggaaa actttgggtgc 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 tggggtctcc	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 aagtctctgt tggtttcctg	120
agtgaaactg tcttaatgtt accaacagaa gttaaatagcc taagagwtc ttatacatgg	180
gctgagtacc tctgtgactg ggcaagccac ctcacctcat tttaccttgt ctgcaaaatg	240
aggaactggg tcaactcatc gtccaatct cactgaaagc taattgatcg cttttgacag	300
aagtagctcc cttgggccgt atatttatatt cctagcttgg aggaagggtg ggacagacag	360
aattgatgta cacctttatt tttatctcta tggtaaacct gtgcatacta aagcattcct	420
ctggctcttt gagatgagtg tatacattgt gtctggccct gtgcattttt taccaagaag	480
taagttttgt tgagtaaact tgggttgtat gaagaactgc atgctcaccg tactcaagta	540
gcttttgcta cctaaaggac agctgctcat atgtacttga cttcctttta agtgaaggat	600

gatgacattt gaaaaacgga ggttgaaaag gag 633

<210>36

<211>25

<212>DNA

<213>Homo sapiens

<400>36

tatcttatat cccctccaag catte 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 gtcatcattt gtgacaatgt cttcttttct	120
ctggtataaa tgaggcatgt agaaagaaga ttgacatttg ctagaagctt cccctttcct	180
ctaactccac aataaaatgg atgctcataa ttacatctgc tcctataagg tcaagatttc	240
agggctggaa gtgaccttag atcathtag cccaacttgc cctcaggaaa ggaaactgag	300
gcccagagat gccttaagtg aattgcccac tgtcacacgc tgagtcagt gccagagcaa	360
ggcttggatc cagttctctg ctccctttcc agagccttgt gatgtcttct ctctacagg	420
aggtgaaaat aactgctgtg gctggttctg ttttctgac tgtaaattgg gtcatggtca	480
gggacagtgc ataggtgtaa agaagttgct ggttgggggt tctaatagcag gtttctccaa	540
aagtgaatgc cctgttaaaa aaaaattctt aacaaatata cagagatttt tttttwaaaa	600
aagtgtgaca gttctagaca cctagagagt aaagtgaaga agcctgtttt caggtttccc	660
gcctccctga atttccagc atgggtccagg ctttgaaatt tatattctctg cttttggcaa	720
tggttgatgg gaatttccca catttatatt tttagctacag agaaaggaca ttatctttaa	780
aatctcttcg ttgttctctc tctttgagtg aggagagaag atgtgaatcc tggcagtggt	840
tcagagtgga cacagcccct gtgtttgtgg cataggctct gtgggcccga tgccagggag	900
cagtaccccc gtgtaaagga gtgggggttt gtccatttgg atagagcaaa gatcctccac	960

ctcaaattccc acaagaacag ttgccacaac ctgggcccta agcatctcat tttcctatgt	1020
agaaattaat gatctggagg agatggcaaa acatttcctc 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 agtccataat	120
atgattaaac cccaattttac tgaatcaatg ccatattgtt gggctcttag attgtctcct	180
tttgtttctg ctactgtgaa tgatcctgtg atgatcatct ttgtgtgtaa atctttgtcc	240
cctcgcccc tcccctttta ttattttctt gggatagacc ccaggacaaa aggtagaaaa	300
gaacaaagtg ttaaamaatt tcttgataca tagccacaga ttattttcct gaaagttctc	360
aacatttata actacgagca gtatgtaaga gagttatggt tggaatgatt ttaatgtctc	420
tggggaattt aacaacaaaa aaactttagg ctcttttgga gagagacatg cccttaactc	480
cacccgccc tagaacagag acccagccca tccaagtcag cctccccagg tcctccacct	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 ttttaagatga aaattgctgt agttaataag	60
cacttcccca tgtcattaaa atgcttaagg atttttaatg accacataac agtccataat	120
atgattaaac cccaatttac tgaatcaatg ccatattgtt gggctcttag attgtctcct	180
tttgtttctg ctactgtgaa tgatcctgtg atgatcatct ttgtgtgtaa atctttgtcc	240
cctcgcccc tcccccttta ttatcttctt gggatagacc ccaggacaaa aggtagaaaa	300
gaacaaagtg ttaaaaaatt tcttgataca tagccacaga ttatcttcct gaaagttcts	360
aacatttata actacgagca gtatgtaaga gagttatggt tggaatgatt ttaatgtctc	420
tggggaattt aacaacaaaa aaactttagg cttctttgga gagagacatg cccttaacte	480
cacccgccc tagaacagag acccagecca tccaagtcag cctccccagg tcctccacct	540
tcaaacagg caaacgaaat catttcttga ataattggta ggcttcaagg tcagatgtt	599

<210>45

<211>641

<212>DNA

<213>Homo sapiens

<400>45

tgctatgtcc agtttacaca taaggatgtg caaatccagc aggttagctg agctgcccag	60
gaatatccag gcaagaatga ccatattctg ataattactc aggcctctgc ctcatctccg	120
ctgscceccc gcccctgac tctcttctga gtgccagatt cagcctccat ttgaatgcc	180

aatagacagg aaattagcat gccagaatc cacgtcttta gtgcactctc tccccagctc	240
caaacctgtt actgcttggt ttcaacatct cagtaaagct caacaacatc gaccattac	300
ttaggcctca aaccttggtt ggcatcgctg attgctcttt tctttcatac cccacattca	360
acccatcagc ccatccaca ggcccaagtg tgtcctctct accttcaaag cgtgtgtggc	420
atccaccgct taccaccacc tctgccatta ccactggagt ccagtgccat catctctcac	480
ttggatgtgg ccagagtgtc ttgctgggtc tccttcttgc ttcctacctt tgtaacagcc	540
tatcatctat ctctggtctc catagctcac tcccatactt tgagagggcc ttgaaagcc	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

cccgactttc 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 cttttaagtt	180
ttctgcttgt ttagaaaaga aatttataca gagccccag tagcagctgt aagggggcag	240
gttcttggag cageccatcc tcaacattct tgctgctgat 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

caccccgact caatggagga ccctc 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

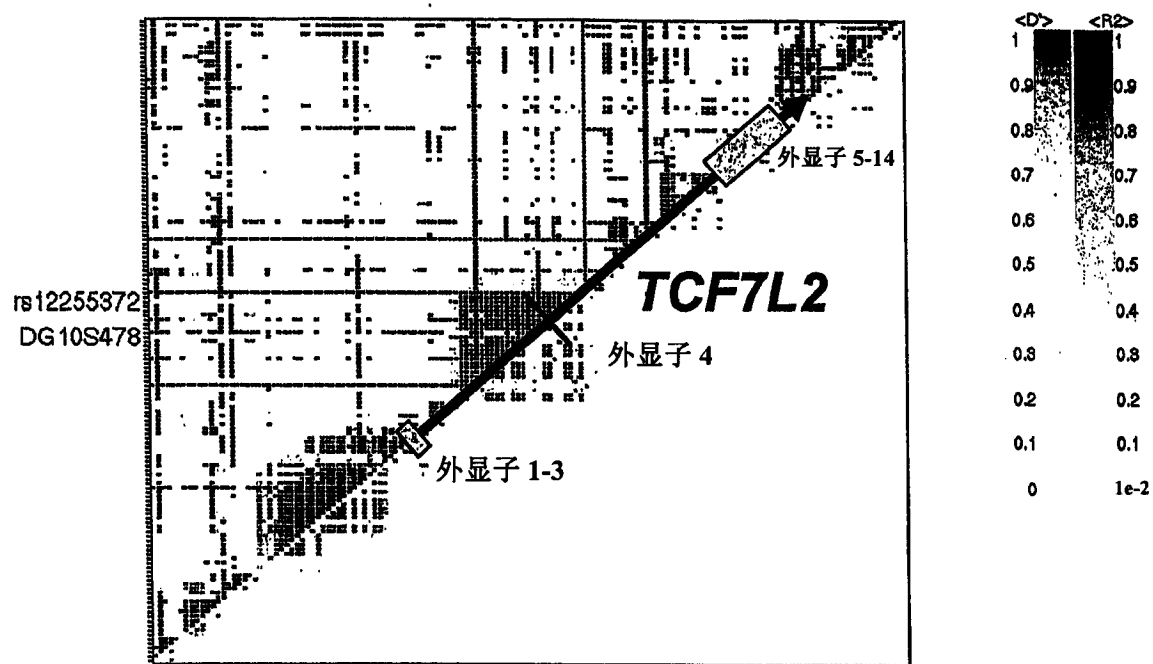


图 1