



(12) 发明专利

(10) 授权公告号 CN 1541265 B

(45) 授权公告日 2014. 09. 10

(21) 申请号 02815708. 7

约瑟夫·特维利格

(22) 申请日 2002. 08. 09

(74) 专利代理机构 北京三友知识产权代理有限公司

(85) PCT国际申请进入国家阶段日

公司 11127

2004. 02. 10

代理人 丁香兰

(86) PCT国际申请的申请数据

(51) Int. Cl.

PCT/EP2002/008963 2002. 08. 09

C12N 15/00 (2006. 01)

(87) PCT国际申请的公布数据

(56) 对比文件

W02003/014342 EN 2003. 02. 20

BG189020. GenBank. 2001,

AQ515834. GenBank. 1999,

(73) 专利权人 国家公共卫生学院

审查员 李振鹏

地址 芬兰赫尔辛基

(72) 发明人 莱纳·佩尔托宁 纳比尔·埃纳塔

伊尔马·耶尔韦莱 蒂莫·萨希

埃尔基·萨维拉赫蒂

权利要求书2页 说明书26页

序列表10页 附图9页

(54) 发明名称

鉴定与成人型肠乳糖酶缺乏相关的 DNA 变异体

(57) 摘要

本发明涉及一种引起或指示成人型肠乳糖酶缺乏并包含肠乳糖酶-根皮苷水解酶(LPH)基因的5'部分的核酸分子,该核酸分子选自:(a)具有或包含SEQ ID NO:1的核酸序列的核酸分子,SEQ ID NO:1的序列也描述在图4中并包含于图8描述的序列中;(b)具有或包含SEQ ID NO:2的核酸序列的核酸分子,SEQ ID NO:2的序列也描述在图5中并包含于图9描述的序列;(c)至少20个核苷酸的核酸分子,该核酸序列的互补链在严格条件下与(a)或(b)的核酸分子杂交,其中所述的多核苷酸/核酸分子在相应于LPH基因5'-13910位点的位置上有一个胞嘧啶残基;和(d)至少20个核苷酸的核酸分子,该核酸分子的互补链在严格条件下与(a)或(b)的核酸分子杂交,其中所述的多核苷酸/核酸分子在相应于LPH基因5'-22018位点的位置上有一个鸟嘌呤残基。本发明还涉及以包含于上述核酸分子中的SNP的分析为基础,测验成人型肠乳糖酶缺乏的存在或易感性的方法。此外,本发明涉及用于检测成人型肠乳糖酶缺乏的存在或易感性的诊断组合物和试剂盒。

1. 一种指示成人型肠乳糖酶缺乏的核酸分子,其位于肠乳糖酶-根皮苷水解酶(LPH)基因的上游,该核酸分子选自:

(a) 由 SEQ ID NO:1 表示的核酸序列;

(b) 由 SEQ ID NO:2 表示的核酸序列。

2. 一种用于说明某人没有患有成人型肠乳糖酶缺乏因此也没有易感性的核酸分子,其位于肠乳糖酶-根皮苷水解酶(LPH)基因的上游,该核酸分子选自:

(a) 由 SEQ ID NO:3 表示的核酸序列;

(b) 由 SEQ ID NO:4 表示的核酸序列。

3. 如权利要求 1 或 2 所述的核酸分子,该核酸分子为基因组 DNA。

4. 如权利要求 3 所述的核酸分子,其中所述的基因组 DNA 为基因的一部分。

5. 一种核酸分子,其为权利要求 1 或 2 所述核酸分子的片段,其具有至少 17 个核苷酸,该片段包含 LPH 基因的核苷酸位点 -13910 或核苷酸位点 -22018。

6. 一种核酸分子,该核酸分子与权利要求 1、3 至 5 任一项所述的核酸分子互补。

7. 一种核酸分子,该核酸分子与权利要求 2 至 5 任一项所述的核酸分子互补。

8. 一种载体,该载体包含权利要求 1、3 至 5 任一项所述的核酸分子。

9. 一种载体,该载体包含权利要求 2 至 4 任一项所述的核酸分子。

10. 一种由权利要求 6 所述的载体转化的非人类宿主细胞。

11. 一种由权利要求 7 所述的载体转化的非人类宿主细胞。

12. 如权利要求 10 或 11 所述的非人类宿主细胞,该非人类宿主细胞为细菌、昆虫细胞、真菌细胞、哺乳动物细胞或植物细胞。

13. 如权利要求 10 或 11 所述的非人类宿主细胞,该非人类宿主细胞为酵母细胞、转基因动物细胞或转基因植物细胞。

14. 一种药物组合物,该药物组合物包含如权利要求 2、3、4 所述的野生型核酸分子或权利要求 9 所述的载体,其中所述的野生型核酸分子在相应于 LPH 基因 -13910 位点的位置上有一个胸苷和 / 或在相应于 -22018 位点的位置上有一个腺苷。

15. 一种诊断组合物,该诊断组合物包含如权利要求 1 至 7 任一项所述的核酸分子和 / 或权利要求 8 或 9 所述的载体。

16. 权利要求 1、3 至 6 任一项所述的核酸分子在制备用于测验成人型肠乳糖酶缺乏的存在或易感性的诊断组合物中的应用,该应用包括测验来自预期患者或被疑具有这样的易感性的人的样品中纯合或杂合状态下如权利要求 1、3 至 6 任一项所述的核酸分子的存在。

17. 权利要求 2 至 5、7 任一项所述的核酸分子在制备用于测验成人型肠乳糖酶缺乏或相关特性的存在或易感性的诊断组合物中的应用,该应用包括测验来自预期患者或被疑具有这样的易感性的人的样品中纯合或杂合状态下如权利要求 2 至 5、7 任一项所述的核酸分子的存在。

18. 如权利要求 16 或 17 所述的应用,其中待实施的所述测验包括在严格的条件下,将权利要求 6 所述的与引起或指示成人型肠乳糖酶缺乏的核酸分子互补的互补核酸分子或如权利要求 7 所述的与野生型序列互补的核酸分子作为探针,与包含于所述样品中的核酸分子杂交,并检测所述杂交。

19. 如权利要求 16 或 18 任一项所述的应用,该应用还包括使用限制性内切酶消化所述

的杂交产物或将所述的杂交用限制性内切酶消化,并分析所述的消化产物。

20. 如权利要求 18 所述的应用,其中所述的探针具有可检测标记。

21. 如权利要求 16 或 17 所述的应用,其中所述的测验包括测定如权利要求 1 至 7 任一项所述的核酸分子的至少一部分核酸序列,所述的部分包含 LPH 基因的核苷酸位点 -13910 和 / 或核苷酸位点 -22018。

22. 如权利要求 21 所述的应用,其中由固相小型测序法确定所述的核酸分子的序列。

23. 如权利要求 21 所述的应用,该应用还包括在确定所述核酸分子的序列前,扩增至少所述核酸分子的所述部分。

24. 如权利要求 23 所述的应用,其中所述的扩增由聚合酶链反应(PCR)实现或所述的扩增即聚合酶链反应。

25. 如权利要求 16 至 24 任一项所述的应用,其中所述的样品为血液、血清、血浆、唾液、尿液、粘膜组织、粘液、阴道组织、皮肤、毛发、毛囊或其他人类组织。

26. 如权利要求 16 至 25 任一项所述的应用,其中将所述的样品中的所述核酸分子固定于一种固体支持物上。

27. 如权利要求 26 所述的应用,其中所述的固体支持物为芯片、氧化硅晶片、珠子或微量滴定板。

28. 权利要求 1 至 7 任一项所述的核酸分子在制备用于测验成人型肠乳糖酶缺乏的存在或易感性的诊断组合物中的应用。

29. 一种试剂盒,该试剂盒包含在一个或一个以上容器中的如权利要求 1 至 7 任一项所述的核酸分子、权利要求 8 或 9 所述的载体。

## 鉴定与成人型肠乳糖酶缺乏相关的 DNA 变异体

### 技术领域

[0001] 本发明涉及一种引起或指示成人型肠乳糖酶缺乏并包含肠乳糖酶-根皮苷水解酶(LPH)基因5'部分的核酸分子,该核酸分子选自:(a)具有或包含SEQ ID NO:1的核酸序列的核酸分子,SEQ ID NO:1的序列也描述在图4中并包含于图8描述的序列中;(b)具有或包含SEQ ID NO:2的核酸序列的核酸分子,SEQ ID NO:2的序列也描述在图5中并包含于图9描述的序列中;(c)至少20个核苷酸的核酸分子,该核酸分子的互补链在严格条件下与(a)或(b)的核酸分子杂交,其中所述的多核苷酸/核酸分子在相应于LPH基因5'-13910位点的位置上有一个胞嘧啶残基;和(d)至少20个核苷酸的核酸分子,该核酸分子的互补链在严格条件下与(a)或(b)的核酸分子杂交,其中所述的多核苷酸/核酸分子在相应于LPH基因5'-22018位点的位置上有一个鸟嘌呤残基。本发明还涉及在对包含在上述的核酸分子中的SNP的分析的基础上,检验成人型肠乳糖酶缺乏的存在或易感性的方法。此外,本发明还涉及用于检测成人型肠乳糖酶缺乏的存在或易感性的诊断组合物和试剂盒。

### [0002] 背景技术

[0003] 本说明书引用了许多文献。这些文献公开的内容,包括厂家手册和目录,在此引入作为参考。

[0004] 乳糖酶-根皮苷水解酶(LPH)只由肠上皮细胞表达,其将乳糖,奶中的糖分解为葡萄糖和半乳糖<sup>1</sup>。在哺乳动物的断奶期,乳糖不再是饮食中必不可少的一部分,LPH酶的表达急剧降至非常低的水平。在人类中,这种情况被认为是成人型肠乳糖酶缺乏或乳糖酶非持续性,其影响大多数人群并由于乳糖的不耐受性而严重限制了成人对新鲜牛奶的利用。乳糖酶非持续性状况的起始年龄在人群中不同,由1-2岁(泰国人)到10-20岁(芬兰人)<sup>2-3</sup>。然而,在北欧和一些其他种族群体中,LPH活性在大多数成人中终生持续,这种情况被认为是乳糖酶的持续性。已显示乳糖酶的持续性/非持续性基因型由遗传决定,相对于非持续状态,持续状态是显性的<sup>4-6</sup>。

[0005] 本领域诊断成人型肠乳糖酶缺乏的依据是乳糖耐受性试验(LTT)。过夜禁食(10小时)后,按1克乳糖/千克体重给予12.5%的乳糖溶液,最大剂量为50克。摄入乳糖前及20和30分钟后采集毛细血管血样。采用葡萄糖氧化酶法测定葡萄糖浓度(Hjelm和de Verdier,1963)。记录LTT当天的腹部症状。将血糖浓度上升的最大值等于或大于1.1毫摩尔/升(mmol/l)作为乳糖吸收不良的标志(Gudman-Hoyer和Harnum 1968,Jussila 1970,Sahi 1972)。LTT包含的假阳性和假阴性诊断率为10%,即LTT的敏感性和特异性约为90%(Isokoski等1972,Newcomer等1975,Sahi 1983)。

[0006] 给予0.3克乙醇/千克体重抑制肝内半乳糖的代谢(Tygstrup和Lundqvist 1962),15分钟后按1克乳糖/千克体重给予12.5%的乳糖溶液可增加LTT的精确度。

[0007] 在首次或重复LTT中血糖上升最大值低于0.2mg/100ml的儿童须通过消化内镜检查进行小肠活检。这种侵入性措施需要专业技术,通常只能由大学医院的胃肠病学专家操作。活检样品在解剖显微镜下进行组织学检查,并测定其粘膜麦芽糖酶、蔗糖酶和乳糖酶的

活性 (Launiala 等,1964)。如果肠活检组织学正常,而乳糖酶活性低于 20U/g 蛋白且乳糖酶 / 蔗糖酶的比率低于 0.30,或使用乙醇的 LTT 中血糖浓度最大的上升值低于 20mg/100ml 且半乳糖浓度等于或小于 5mg/100ml,则诊断儿童肠乳糖酶缺乏是正确的 (Sahi 等,1972)。如上所述,目前诊断成人型肠乳糖酶缺乏的方法很繁琐。LTT 不准确,因此在确诊前需行侵入性方法即消化内镜检查。由于成人型肠乳糖酶缺乏十分常见并且是非特异性腹部症状的主要病因 (三分之一的患者抱怨腹痛),所以明显地需要改进这种常见健康问题的诊断。

[0008] 然而,目前没有发展出易掌握且同时提供快速而精确的结果的生化测验法。在基因组 DNA/ 表达水平澄清该病的病因也未成功。因此,成人中 LPH 基因的编码区和启动子区域的测定没有揭示与乳糖酶的持续性 / 非持续性相关的 DNA 变异,也没有与该特征相关的来自拼接变异体或 mRNA 编辑变异体的证据<sup>7-8</sup>。过去的研究表明乳糖酶的持续性 / 非持续性特点可能由位于乳糖酶基因内或旁的顺式作用元件控制,并在跨跃乳糖酶基因的 70kb 单倍型中观察到强连锁不平衡 (LD)<sup>9,10</sup>。一些研究报告了 LPH 基因表达的控制主要在转录调控水平上进行的证据<sup>11-13</sup>。然而,提示影响 LPH 基因表达的转录和转录后控制的变异可与成人型肠乳糖酶缺乏的病因相关<sup>14-15</sup>。

## 发明内容

[0009] 鉴于以上内容,本发明的主要技术问题是提供能够准确而方便地诊断成人型肠乳糖酶缺乏或该疾病的易感性的手段和方法。

[0010] 通过具有权利要求书所限定的技术特征的实施方式可以获得解决所述技术问题的技术方案。

[0011] 因此,本发明涉及一种引起或指示成人型肠乳糖酶缺乏的核酸分子,该核酸分子包含肠乳糖酶 - 根皮苷水解酶 (LPH) 基因的 5' 部分,该核酸分子选自:(a) 具有或包含 SEQ ID NO:1 的核酸序列的核酸分子,SEQ ID NO:1 的序列也描述在图 4 中并包含于图 8 描述的序列中;(b) 具有或包含 SEQ ID NO:2 的核酸序列的核酸分子,SEQ ID NO:2 的序列也描述在图 5 中并包含于图 9 描述的序列中;(c) 至少 20 个核苷酸的核酸分子,该核酸分子的互补链在严格条件下与 (a) 或 (b) 的核酸分子杂交,其中所述的多核苷酸 / 核酸分子在相应于 LPH 基因 5' -13910 位点的位置上有一个胞嘧啶残基;和 (d) 至少 20 个核苷酸的核酸分子,该核酸分子的互补链在严格条件下与 (a) 或 (b) 的核酸分子杂交,其中所述的多核苷酸 / 核酸分子在相应于 LPH 基因 5' -22018 位点的位置上有一个鸟嘌呤残基。

[0012] 本发明中术语“肠乳糖酶 - 根皮苷水解酶 (LPH) 基因”表示编码具有将乳糖水解为葡萄糖和半乳糖成分的活性的酶的基因。该酶由 E. C. 3. 2. 1. 23. 62 表征。

[0013] 术语“成人型肠乳糖酶缺乏”指一种也被认为是乳糖不耐受的疾病,是一种由于在很大一部分世界人群中肠细胞的乳糖酶 - 根皮苷水解酶 (LPH) 活性的“生理性”降低引起的常染色体隐性疾病。

[0014] 短语“引起或指示成人型肠乳糖酶缺乏”指发现的 SNP 和相应的核酸分子指示并可能引起该疾病这一事实。相应地,需要该短语中列举的 5' 位置指示该疾病。另一方面,所述的短语无需该 5' 部分导致或引起该疾病。然而所述的短语不排除 SNP 之一或两种 SNP 起导致或协助作用。

[0015] 术语“在严格条件下杂交”指本领域技术人员熟知或可根据传统程序建立的杂交

条件。该术语最优先地指高度严格的条件。每个序列的适当的严格条件可根据熟知的参数例如温度、核酸分子的组成和盐条件等建立：见例如 Sambrook 等《分子克隆：实验室手册 (Molecular Cloning, A Laboratory Manual)》；冷泉港实验室出版社，冷泉港，1989 或 Higgins 和 Hames (编辑)《核酸杂交，实用方法 (Nucleic acid hybridization, a practical approach)》，IRL 出版社，牛津 1985 (参考文献 54)，具体见 Britten & Davidson 的“杂交策略 (Hybridization Strategy)”章节，3 至 15。典型 (高度严格) 的条件包括在 65°C 下、0.5×SSC 和 0.1% SDS 中杂交或在 42°C 下、50% 甲酰胺、4×SSC 和 0.1% SDS 中杂交。杂交后通常进行洗涤以去除非特异性信号。洗涤条件包括例如 65°C 下，0.2×SSC 和 0.1% SDS 或 25°C 至 65°C 下，2×SSC 和 0.1% SDS 或 0.3×SSC 和 0.1% SDS。

[0016] 作为本文上述公开，本发明还涉及一种至少 20 个核苷酸的杂交核酸分子，见上述 (c) 和 (d)。然而，本发明还涉及至少 50，至少 100，至少 150 或至少 200 个核苷酸的核酸分子。优选地，所述的杂交片段包含至少 25，至少 50，或至少 75 个核苷酸，至少 100 个核苷酸，如上述 (c) 定义的 -13910 位点的 5' 和 3' 或如上述 (d) 定义的 -22018 位点的 5' 和 3'。

[0017] 术语“核酸分子”指天然和非天然存在的核酸分子。非天然存在的核酸分子包括 cDNA 及其衍生物如 PNA。

[0018] 在本发明中，术语“包含 SEQ ID NO: 核酸序列的核酸分子 [...]”均指至少比 SEQ ID NO 定义的核酸分子长至少 1 个核苷酸的核酸分子。同时，这些核酸分子最多延伸超出本发明定义的 (例如 SEQ ID NO: 2 或 1、3 或 4) 核酸分子的 5' 和 / 或 3' 末端 30000 个核苷酸。

[0019] 令人吃惊地，发现根据本发明，距离 LPH 基因相当远的两个与肠乳糖酶缺乏相关的变异体位于 MCM6 基因不同的内含子中。MCM6 是基因家族 (MCM2-7) 的一员，为确保在细胞周期中 DNA 复制起始只发生一次所必须<sup>31</sup>。MCM6 不同于 LPH，其组织分布不受限制，MCM6 和 LPH 转录本的水平没有相关性<sup>18</sup>。这些发现提示这两种基因不共用任何提供组织特异性或调控发育的、功能重要性的顺式作用元件<sup>18</sup>。鉴定的变异体很可能具有不同的表达 LPH 和 MCM6 基因的功能重要性。更令人吃惊地，基于完全和肠乳糖酶缺乏相关，这两种基因 (或其中之一) 与肠上皮中 LPH 基因转录水平的依赖年龄的下调相关，但对 MCM6 的转录几乎或完全没有影响。

[0020] 根据实验，对 9 个扩展的芬兰家族利用连锁、等位相关和扩展单倍型分析，将成人型肠乳糖酶缺乏的位点限制在 2q21 上一段 47kb 的区间。对该区域的序列分析揭示一种单核苷酸多态性 (SNP)，即 C/T-13910，其在所有芬兰家族中和一套来自 4 个不同人群的 236 个个体的样本中，与成人型肠乳糖酶缺乏完全共分离。另一 SNP，即 G/A-22018，在与 C/T-13910 距离 8kb 的端粒处，其与除了 7 个病例以外的所有个体的特征相关。C/T-13910 SNP 在 1047 个 DNA 样品中的发病率反映了在三个不同人群中报告的成人型肠乳糖酶缺乏的发病率，为这一特征的重要性提供了附加的证据。

[0021] 上述令人惊奇的发现首次在对引用的 LPH 基因上游的单核苷酸多态性的分子分析的基础上建立测验系统。尽管两种 SNP 均为诊断成人型肠乳糖酶缺乏或其易感性提供了坚实的基础，优选地可单独分析核苷酸位置 -13910，或结合核苷酸位置 -22018 分析。这是由于位置 -13910 的 SNP 与分析的成人型肠乳糖酶缺乏病例 100% 相关，而位置 -22018 的 SNP 与成人型肠乳糖酶缺乏的所有病例仅 98% 相关。然而单独分析位置 -22018 也常为诊

断成人型肠乳糖酶缺乏或其易感性提供可信的基础。

[0022] 由于已建立多种评估 SNP 存在的方法,目前可方便地在短时间内、花费较低的成本、高度精确且不会给受测者带来明显麻烦的情况下诊断成人型肠乳糖酶缺乏的遗传易感性。

[0023] 本发明还涉及一种包含肠乳糖酶-根皮苷水解酶(LPH)基因5'部分的核酸分子,该核酸分子选自:(a)具有或包含SEQ ID NO:3的核酸序列的核酸分子,SEQ ID NO:3的序列也描述在图6中;(b)具有或包含SEQ IDNO:4的核酸序列的核酸分子,SEQ ID NO:4的序列也描述在图7中;(c)一种核酸分子,该核酸分子的互补链在严格条件下与(a)或(b)的核酸分子杂交,其中所述的多核苷酸/核酸分子在相应于LPH基因-13910位点的位置上有一个胸苷残基;和(d)一种核酸分子,该核酸分子的互补链在严格条件下与(a)或(b)的核酸分子杂交,其中所述的多核苷酸/核酸分子在相应于LPH基因-22018位点的位置上有一个腺苷残基。

[0024] 本发明的实施方式可方便地用于说明某人没有患有成人型肠乳糖酶缺乏因此也没有易感性。进一步地,这些核酸分子反映LPH基因上游的位置-13910或-22018的“野生型”状态可作为实验中测验成人型肠乳糖酶缺乏的易感性的对照手段。本说明书中描述的所有测验方法均可使用。

[0025] 在本发明一个优选实施方式中,核酸分子是基因组DNA。

[0026] 本发明优选实施方式反映通常以分离自受测者体液、细胞或组织的基因组DNA为基础进行分析这一事实。

[0027] 在本发明的核酸分子的另一优选实施方式中,所述的基因组DNA是基因的一部分。

[0028] 本发明优选分析含有相应于LPH基因的-13910或-22018位置的MCM6基因的至少一个内含子。

[0029] 此外,本发明涉及具有至少14个核苷酸的上述核酸分子的片段,其中所述片段包含LPH基因(上游)的核苷酸位置-13910或核苷酸位置-22018。

[0030] 本发明的片段可为天然的或(半)合成的。因此该片段可为例如根据传统有机化学程序合成的核酸分子。重要地,本发明的核酸片段包含LPH基因上游的核苷酸位置-13910或核苷酸位置-22018。这些位置中,该片段可为野生型核苷酸或引起或指示成人型肠乳糖酶缺乏的核苷酸(也被称为“突变的”序列)。因此,本发明的片段可用于分析例如区别野生型和突变序列。

[0031] 本发明的片段进一步优选由至少17个核苷酸,更优选至少21个核苷酸,最优选至少25个核苷酸如30个核苷酸组成。

[0032] 进一步地,本发明涉及与上述的核酸分子互补的核酸分子。

[0033] 本发明的实施方式包含至少14个核苷酸并至少包含LPH基因上游序列的位点-13910或位点-22018,该实施方式对分析杂交试验中列举位置的基因结构尤其有用。因此例如与野生型序列(即位点-13910上的胸苷或位点-22018上的腺苷)或引起或指示成人型肠乳糖酶缺乏的变异体(即位点-13910上的胞苷或位点-22018上的鸟苷)精确互补的15聚体可用于区分多态性变异体。这是因为如果选择适合的杂交和洗涤条件,不能与所分析样品中的DNA精确互补的标记有可检测标记的核酸分子将不能产生可检测的信号。

[0034] 在这点上,需注意本发明的核酸分子及其片段和互补的核酸分子可标记有可检测标记。可检测标记包括放射性标记如  $^3\text{H}$  或  $^{32}\text{P}$  或荧光标记。标记核酸的方法为本领域人员熟知并在例如 Sambrook 等,上述引文中描述。

[0035] 此外,本发明还涉及包含上述核酸分子的载体。本发明的载体含有包含野生型序列的核酸分子或含有包含突变序列的核酸分子。

[0036] 具体地,载体可为包含本发明的核酸分子并传统地用于基因工程中的质粒、粘粒、病毒或噬菌体。优选地,所述的载体是表达载体和/或基因转移或目标载体。来自病毒的表达载体如逆转录载体,痘苗病毒、腺相关病毒、疱疹病毒或牛乳头瘤病毒,可用于将本发明的核酸分子传递至目标细胞群中。可利用本领域技术人员已知的方法构建重组病毒载体;见例如 Sambrook 等,上述引文和 Ausubel 等,《Current protocols in MolecularBiology》,Green Publishing Associates and Wiley Interscience,纽约(1989)中描述的技术。可替换地,本发明的核酸分子和载体可被重组到脂质体中并传递给目标细胞。含有本发明的核酸分子的载体可根据宿主细胞的类型由不同的已知方法转移进宿主细胞。例如,氯化钙转染常用于原核细胞,而例如磷酸钙或 DEAE-葡聚糖介导的转染或电穿孔可用于其他细胞宿主;见上述的 Sambrook。

[0037] 这样的载体可进一步包含例如允许在适合的宿主细胞和适合条件下选择所述的载体的标记基因。优选地,本发明的核酸分子可操作性地与表达调控序列连接,实现在原核或真核细胞中的表达。所述多核苷酸的表达包括将多核苷酸转录成可翻译的 mRNA。本领域人员已知确保在真核细胞中,优选哺乳动物细胞中表达的调控元件。它们通常包括确保转录起始的调控序列,可选择地包括确保转录终止和转录本稳定的多腺苷信号和/或进一步增强所述多核苷酸表达的内含子。附加的调控元件可包括转录增强子和翻译增强子,和/或天然相关或异源性启动子区域。可使在原核宿主细胞中表达的调控元件包括例如大肠杆菌的 PL、lac、trp 或 tac 启动子,使在真核宿主细胞中表达的调控元件的例子为酵母的 AOX1 或 GAL1 启动子或哺乳动物或其他动物细胞的 CMV 启动子、SV40 启动子、RSV 启动子(Rous 肉瘤病毒)、CMV 增强子、SV40 增强子或球蛋白内含子。除了负责转录起始外,这些调控元件还包括多核苷酸下游的转录终止信号,如 SV40 多聚腺苷位点或 tk 多聚腺苷位点。可选择地,所述异源序列可编码一种包含赋予所需特性,例如稳定表达的重组产物或简化其纯化的 C 或 N 末端鉴定肽的融合蛋白。文中本领域已知的适合的表达载体例如 Okayama-Berg cDNA 表达载体 pcDV1 (Pharmacia)、pCDM8、pRc/CMV、pcDNA1、pcDNA3、Echo™ 克隆系统 (Invitrogen)、pSPORT1 (GIBCO BRL) 或 pRevTet-On/pRevTet-Off 或 pCI (Promega)。优选地,表达调控序列是能够转化或转染真核宿主细胞的载体中的真核启动子系统,但也可使用原核宿主的控制序列。

[0038] 如上述,本发明的载体还可为基因转移载体或基因目标载体。基因治疗的基础是通过活体外或体内技术将治疗性基因引入细胞内,这是基因转移的一个最主要的应用。文献中描述了体内或活体外基因治疗的适合的载体和方法并为本领域技术人员已知;见例如:Giordano, Nature Medicine 2(1996),534-539;Schaper, Circ. Res. 79(1996),911-919;Anderson, Science 256(1992),808-813;Isner, Lancet 348(1996),370-374;Muhlhauser, Circ. Res. 77(1995),1077-1086;Wang, Nature Medicine 2(1996),714-716;WO 94/29469;WO 97/00957, Schaper, Current Opinion in Biotechnology 7(1996),

635-640 或 Kay 等 (2001) *Nature Medicine*, 7, 33-40) 及本文中引用的参考文献。本发明的多核苷酸和载体可设计为直接引入细胞或由脂质体或病毒载体 (例如腺病毒, 逆转录病毒) 引入细胞。优选地, 所述的细胞是生殖系细胞, 胚胎细胞或卵细胞及其衍生形式, 最优选地所述细胞是干细胞。基因治疗只设想野生型核酸分子。

[0039] 本发明还涉及引物或引物对, 其中的引物或引物对在 (高度) 严格的条件下与上述包含 LPH 基因的核苷酸位点 -13910 或 -22018 的核酸或其互补链杂交。

[0040] 优选地, 本发明的引物长度至少为 14 个核苷酸例如 17 或 21 个核苷酸。更优选地引物最长 24 个核苷酸。在合适的条件下引物与包含位点 -13910 或位点 -22018 的基因组序列发生杂交或未杂交并联合适合的检测方法例如延伸反应或扩增反应可用于区别多态性变异体, 然后得出例如受测者的成人型肠乳糖酶缺乏的易感性的结论。本发明设想两种类型的引物 / 引物对。一种与包含突变序列的序列杂交。换言之, 该引物与包含位点 -13910 的胞苷或位点 -22018 位点的鸟苷的序列或其互补链精确地互补。另一种类型的引物与包含位点 -13910 的胸苷或位点 -22018 的腺苷的序列或其互补链精确地互补。由于杂交条件优选足够严格, 将例如与突变序列精确互补的引物和野生型等位基因接触, 由于形成错配不产生有效的杂交。洗涤后, 由于引物被除去不能检测到信号。

[0041] 此外, 本发明涉及由本发明上述载体转化的非人类宿主。该宿主可携带突变或野生型序列。通过孵育等, 该宿主可成为 SNP 之一或二者的杂合子或纯合子。

[0042] 本发明的宿主可携带本发明瞬时或稳定地整合入基因组中的载体。本领域人员熟知产生非人类宿主的方法。例如, 利用传统的转染程序 (见 Sambrook 等, 上述引文) 可产生转化的细菌 (如大肠杆菌) 或转化的酵母菌。利用本发明的非人类宿主可例如阐明成人型肠乳糖酶缺乏的发病。

[0043] 本发明的一个优选实施方式中, 非人类宿主是细菌、酵母细胞、昆虫细胞、真菌细胞、哺乳类细胞、植物细胞、转基因动物或转基因植物。

[0044] 其中优选的细菌为大肠杆菌 (*E. coli*), 优选的酵母细胞为酿酒酵母 (*S. cerevisiae*) 或巴斯德毕赤酵母 (*Pichia pastoris*) 细胞。优选的真菌细胞是曲霉 (*Aspergillus*) 细胞, 优选的昆虫细胞包括秋粘虫卵巢 (*Spodoptera frugiperda*) 细胞。优选的哺乳动物细胞为显示 LPH 酶表达的结肠癌细胞系并包括 CaCo2 细胞。

[0045] 产生转基因非人类动物例如转基因小鼠的方法, 包括将前述的多核苷酸或目标载体引入生殖细胞、胚胎细胞、干细胞或卵细胞或其衍生细胞中。可根据本发明所述的筛选方法使用所述非人类动物。产生转基因胚胎并进行筛选, 可按例如 A. L. Joyner 编辑的《基因目标导向, 实用的方法 (Gene Targeting, A Practical Approach)》中的描述, (1993), 牛津大学出版社。如上述, 可使用例如适合的互补核酸分子由 DNA 印迹法分析胚胎的胚膜 DNA。本领域已有常用的产生转基因非人类动物的方法的描述, 见例如 WO 94/24274。产生转基因非人类生物 (包括同源的目标非人类动物) 优选胚胎干细胞 (ES 细胞)。鼠 ES 细胞, 例如 AB-1 系可用于同源基因目标导向, 所述 AB-1 系在有丝分裂不活跃的 SNL76/7 细胞滋养层 (McMahon 和 Bradley, *Cell* 62:1073-1085 (1990)) 中生长, 基本如 Robertson, E. J. (1987) 的《Teratocarcinomas and Embryonic Stem Cells: A Practical Approach》E. J. Robertson 编辑, 牛津: IRL 出版社, p. 71-112 中的描述。其他合适的 ES 系包括但不限于: E14 系 (Hooper 等, *Nature* 326:292-295 (1987)), D3 系 (Doetschman 等, *J. Embryol.*

Exp. Morph. 87 :27-45 (1985), CCE 系 (Robertson 等, Nature 323 :445-448 (1986)), AK-7 系 (Zhuang 等, Cell 77 :875-884 (1994))。由带有特定目标突变的 ES 细胞产生的小鼠系的成功依赖于该 ES 细胞的多能性 (即其进入宿主的发育胚胎例如胚泡或桑葚胚后参与胚胎形成的能力及对产生的动物的生殖细胞的贡献能力)。包含注入的 ES 细胞的胚泡在雌性非人假孕动物的子宫中发育, 并出生形成嵌合小鼠。由于回交具有所需核酸分子的细胞、并以 PCR 或 DNA 印迹分析筛选其后代的尾组织活检 DNA 中正确的目标转基因的存在, 以鉴定杂合了本发明的核酸分子的转基因小鼠, 因此产生的转基因小鼠是嵌合的。

[0046] 转基因非人动物可为例如转基因小鼠, 大鼠, 仓鼠, 狗, 猴子 (猿), 兔子, 猪或牛。优选所述的转基因非人类动物是小鼠。本发明的转基因动物尤其可用于研究本发明的核酸分子和载体的基因型表达 / 结果。进一步地, 本发明的转基因动物可用于研究 LPH 酶的发育表达, 例如在啮齿类肠中。进一步设想, 本发明的非人类转基因动物可用于测验可能改善成人型肠乳糖酶缺乏的治疗制剂 / 组合物或其他可行的疗法。

[0047] 此外, 本发明涉及与本发明的突变核酸分子特异性结合但不相应的野生型核酸分子特异性结合的抗体或适体或噬菌体。

[0048] 抗体可用于测验结合并可用于本领域任何已知的血清学技术中, 例如试管、凝胶、固相中的凝集反应和有或没有第二抗体的捕获技术, 或用于在有或没有免疫荧光增强的流式细胞仪中 (技术描述见例如, Harlow 和 Lane, "Antibodies, A Laboratory Manual", 冷泉港实验室出版社, 冷泉港, 美国, 1988 (见参考文献 53))。

[0049] 根据本发明, 抗体特异性识别包含位点 -13910 (其中核苷酸为胞嘧啶核苷酸) 或位点 -22018 (其中核苷酸为鸟嘌呤核苷酸) 的表位。该抗体不或基本不与包含位点 -13910 的胸苷的表位或 -22018 位的腺苷的表位发生交联反应。根据标准程序制备的抗体的特异性可通过与携带野生型序列和突变序列的 DNA 分子接触, 由例如 ELISA 分析法进行检测。只有那些在突变序列背景下产生信号而在野生型序列背景下不产生信号的抗体被选中。

[0050] 本发明的抗体可为单抗或来自或包含于多抗血清中的抗体。根据本发明的术语“抗体”还包括所述抗体的片段例如 Fab, F(ab')<sub>2</sub>, Fv 或 scFv 片段; 见例如 Harlow 和 Lane<sup>53</sup>, 上述引文。抗体或其片段可由天然或由 (半) 合成产生。这样的合成产物还包括与本发明的抗体相同或基本相同的结合特异性的非蛋白质作为半蛋白质。这些产物可自例如肽模拟物获得。

[0051] 本领域人员熟知术语“适体”, 其定义见例如 Osborne 等, Curr. Opin. Chem. Biol. 1 (1997), 5-9 (见参考文献 51) 或 Stall 和 Szoka, Pharm. Res. 12 (1995), 465-483 (见参考文献 52)。

[0052] 此外, 本发明涉及一种抗体或适体或噬菌体, 其与上述野生型核酸分子特异性结合但不与引起或指示成人型肠乳糖酶缺乏的相应突变序列特异性结合。这里关于与突变序列特异性结合的抗体的特异性等的描述已作必要修正。

[0053] 进一步地, 本发明涉及一种包含上述野生型核酸分子的药物组合物。

[0054] 本发明的药物组合物可用于基因治疗方法, 尤其是体基因治疗。

[0055] 上述包含于本发明的药物组合物中的野生型核酸分子可与药物上可接受的载体和 / 或稀释剂组合。

[0056] 本领域熟知适合的药物载体的例子, 其包括磷酸盐缓冲的盐溶液、水、乳剂例如油

/ 水乳剂、各种湿润剂、无菌溶液等。包含这种载体的组合物可由熟知的传统方法制备。可将适量的这些药物组合物施用于个体。可采用多种途径施用适合的组合物,例如,通过静脉内、腹膜内、皮下、肌内、外用、皮内、鼻内或支气管内施用。剂量范围由住院医师和临床因素决定。正如医学界熟知,给予任一患者的剂量依赖许多因素,包括患者的体重、体表面积、年龄和使用的特定化合物、性别、时间和给药途径、一般健康状况和其他同时使用的药物。典型的剂量可为例如 0.001 至 1000  $\mu\text{g}$  促进表达或抑制表达的核酸;然而,也可想象使用该示例性范围以下或以上的剂量,尤其是考虑到上述因素的时候。剂量会有所不同但静脉内施用的优选 DNA 剂量约为  $10^6$  到  $10^{12}$  个 DNA 分子的拷贝。周期性评定进展。本发明的组合物可用于局部或全身。给药方式通常胃肠外施用,如静脉内施用;也可直接将 DNA 施用于靶点,例如通过生物导弹递送施用于内部或外部靶点或由导管施用于动脉内的位点。胃肠外给药制剂包括无菌水或非水溶液、悬液和乳剂。非水溶剂的例子为丙二醇、聚乙二醇、植物油如橄榄油和可注射的有机酯如油酸乙酯。水载体包括水、酒精/水溶液、乳剂或悬液,包括盐类和缓冲介质。胃肠外载体包括氯化钠溶液、林格氏右旋糖、右旋糖和氯化钠、乳酸林格氏液或固定油。静脉内施用的载体包括液体和营养补充液、电解质补充液(例如以林格氏右旋糖为基础的液体)等等。也可存在防腐剂和其他添加剂例如抗微生物剂、抗氧化剂、螯合剂和惰性气体等等。

[0057] 此外,本发明涉及包含上述核酸分子,上述的载体,上述引物或引物对,和/或上述抗体、适体和/或噬菌体的诊断组合物。

[0058] 诊断组合物用于评估具有出现成人型肠乳糖酶缺乏的易感性或诊断急性期疾病的人的基因状况。诊断组合物的多种可能成分可置于一个或一个以上的瓶中、溶剂中或其他形式例如冻干形式。如果将诊断组合物溶于溶剂中,优选冷却到至少 +8°C 到 +4°C。其他情况优选冷冻。

[0059] 本发明还涉及测验成人型肠乳糖酶缺乏或相关特征的存在或易感性的方法,包括测验自预期患者或被疑具有易感性的人获得的样品的纯合或杂合状态的上述核酸分子的出现。在不同的实施方式中,可测试野生型序列或突变序列的存在。

[0060] 本发明的方法用于检测所述的人/患者的基因结构并得到患者患的病情是否是成人型肠乳糖酶缺乏的适当的结论。可替换地,可评估没有该病情的人是否具有成人型肠乳糖酶缺乏的易感性。对于 LPH 基因上游的 -13910 位点,只有发现纯合状态的胞苷,才能将病情诊断为成人型肠乳糖酶缺乏或证实相应的易感性。另一方面,如果发现纯合状态的胸苷或该个体是杂合子(C/T),则可推断患者的病情与患有成人型肠乳糖酶缺乏无关,进一步地,该患者没有发展成该病情的倾向性。然而,如果携带胞苷残基的染色体与来自父母中另一方的相应染色体配对,则可断定携带杂合基因型的人的孩子可能出现该病情。

[0061] 对位点 -22018 中的 SNP 分析与上述情况类似,其结论也基本相同。存在纯合的鸟苷残基标志成人型肠乳糖酶缺乏的易感性或出现急性成人型肠乳糖酶缺乏。杂合 G/A 状态很可能不发生该病情。认为携带纯合状态的腺苷的个体不会出现此病情。相似地,出现病情的患者不能诊断为患有成人型肠乳糖酶缺乏。

[0062] 在本发明方法的一个优选实施方式中,所述的测验包括将与上述引起或指示成人型肠乳糖酶缺乏的核酸分子互补的核酸分子或与上述野生型序列互补的核酸分子作为探针在(高度)严格的条件下,与所述样品中的核酸分子杂交,并检测所述的杂交。

[0063] 根据使用的核酸分子探针,将再次检测出野生型或突变序列(即引起或指示成人型肠乳糖酶缺乏的序列)。应理解所选的杂交条件使与野生型序列互补的核酸分子不会或基本不会与突变序列杂交。相似地,与突变序列互补的核酸分子不会或基本不会与野生型序列杂交。例如可分别监测/检测杂交后的检测信号的强度/密度来区分本发明的杂交方法中由纯合和杂合基因型得到的结果。为了区分本发明的杂交方法中的野生型纯合、杂合和/或突变纯合等位基因,在分析中包括相应基因型样品的内部对照。

[0064] 在进一步优选实施方式中,本发明的方法还包括以限制性内切酶消化所述的杂交产物或将所述的杂交产物用限制性内切酶消化,并分析所述的消化产物。

[0065] 本发明的这个优选实施方式使通过方便的方法区分有效杂交和非有效杂交。例如,如果邻近-13910位点或-22018位点的DNA序列包含限制性内切酶位点,杂交产物可通过有效的杂交由适合的限制性内切酶切开,而未杂交则不会产出双链产物或不包含可识别的限制性位点,因此并不会被切开。具体的,对DNA变异体C/T<sub>-13910</sub>序列特异的限制性内切酶是CviJI,对DNA变异体G/A<sub>-22018</sub>特异的限制性内切酶是HhaI和AciI。利用Webcutter程序发现所述的限制性内切酶切割rg/cy。可由传统方法分析消化产物,例如凝胶电泳,其可选择性地与核酸染色例如溴化乙锭联合。还可预见与其他的技術例如DNA印迹法联合。

[0066] 例如通过抗DNA双链抗体或使用标记的寡核苷酸实现对所述杂交的检测。方便地,本发明的方法与印迹技术例如DNA印迹或RNA印迹及相关技术一同使用。通过例如标准程序可实现标记,其包括使用放射性标记、荧光、磷光、化学发光、酶标记等标记(也见上述)。

[0067] 根据上述内容,在本发明的方法的另一优选实施方式中,所述的探针具有可检测的标记,例如通过上述的方法或标记。

[0068] 在本发明的方法的另一个优选实施方式中,所述的测验包括测定上述核酸分子至少一部分的核酸序列,所述的部分包含LPH基因的核苷酸位点-13910和/或核苷酸位点-22018。

[0069] 通过传统程序之一例如Sanger或Maxam/Gilbert规则(进一步的指导见Sambrook等,上述引文)实现核酸分子的测定。

[0070] 在本发明的方法的另一个优选实施方式中,通过固相小型测序实现核酸分子的测定。固相小型测序以溶液中野生型和突变型核苷酸的定量分析为基础。首先,包含突变的基因组区域通过PCR由一种生物素化的和未生物素化的引物扩增,其中生物素化的引物连接于链霉抗生素(SA)包被的板。PCR产物变性成单链形式以使小型测序引物恰好在突变位点前与该链结合。将氘(H3)或荧光标记的突变和野生型核苷酸与未标记的dNTPs一起加入小型测序反应,并使用Taq聚合酶测序。结果以β计数器或荧光仪测定反应中野生型和突变核苷酸的量为根据,并表示为R比率。也见Syvänen AC, Sajantila A, Lukka M. *Am J Hum Genet* 1993;52,46-59和Suomalainen A和Syvanen AC. *Methods Mol Biol* 1996;65:73-79。

[0071] 本发明的方法的一个优选实施方式还包括在测定所述的核酸序列前,扩增至少所述核酸分子的所述部分。

[0072] 优选地,由聚合酶链反应(PCR)扩增。也可使用其它的扩增方法例如连接酶链反应。

[0073] 在本发明的方法一个优选实施方式中,所述的测验包括进行扩增反应分析扩增产物,其中至少一种用于所述的扩增反应中的引物是上述的引物或属于上述的引物对。在该实施方式中,根据研究者/医生想获得的信息,可使用与野生型或与突变序列杂交的引物。

[0074] 如果所述的目标序列携带与杂交的引物精确互补的序列,本发明的方法只引起目标序列的扩增。这是由于在优选的(高度)严格的条件下,依赖使用的引物类型,寡核苷酸引物不会与野生型/突变序列杂交(结果未获得扩增产物),而只与精确配对的序列杂交。自然地,可使用与两种 SNP 杂交的引物对的联合。在这种情况下,分析预期的扩增产物(如果第二种未区分的引物对每个位点均相同,扩增产物可为 0、1、2、3 或 4 种)将提供 -13910 和 -22018 两个位点的基因状态的信息。

[0075] 在本发明的方法一个优选实施方式中,所述的扩增可由聚合酶链反应(PCR)实现或所述的扩增即 PCR。

[0076] 本领域人员已掌握 PCR。用于本发明的典型条件包括例如共 35 个循环,总体积为 50 $\mu$ l,示范性的变性步骤为在 93 $^{\circ}$ C 下变性 3 分钟;退火步骤为 55 $^{\circ}$ C 下退火 30 秒;延伸步骤为 72 $^{\circ}$ C 下延伸 75 秒,最后在 72 $^{\circ}$ C 下延伸 10 分钟。

[0077] 本发明进一步涉及测验成人型肠乳糖酶缺乏的存在或易感性的方法,该方法包括分析自人类获得的样品与上述的抗体或适体或噬菌体的特异性结合。在本说明书中,如果使用的抗体适当,与纯合野生型对照样品(包含两个持续的等位基因)相比,表明本发明的抗原存在的较弱的染色表示杂合野生型(一种持续的等位基因和一种肠乳糖酶缺乏的等位基因),而纯合的肠乳糖酶缺乏的个体却不被染色。优选地,本发明的方法在对照样品(作为内部对照)存在下进行,该对照样品对应于所有三种可能的等位基因的结合。可使用对野生型或突变序列特异的抗体等进行测验。

[0078] 测验结合可再次涉及使用标准技术例如 ELISAs;见例如 Harlow 和 Lane<sup>53</sup>,上述引文。

[0079] 在本发明的方法的优选实施方式中,所述的抗体或适体或噬菌体由可检测标记进行标记。

[0080] 其中适体优选由上述的 <sup>3</sup>H 或 <sup>32</sup>P 或荧光标记进行放射性标记,噬菌体或抗体可由相应的方式(<sup>131</sup>I 为优选的放射性标记)或由例如 His 标签、FLAG 标签或 myc 标签标记。

[0081] 在本发明的方法进一步优选实施方式中,所述测验为免疫分析。

[0082] 在本发明的方法另一个优选实施方式中,所述的样品为血液、血清、血浆、胎儿组织、唾液、尿液、粘膜组织、粘液、阴道组织和自阴道、皮肤、毛发、毛囊获得的胎儿组织或其他人类组织。

[0083] 在本发明的方法另一优选实施方式中,样品中所述的核酸分子固定于一种固体支持物。

[0084] 将核酸分子固定于固体支持物可容易地控制测验方法并进一步地,至少有些固体支持物例如芯片,氧化硅晶片或微量滴定板允许同时分析大量样品。理想地,该固体支持物允许应用例如机器人系统的自动测验方法。

[0085] 在本发明的方法特别优选的实施方式中,所述的固体支持物为芯片,氧化硅晶片、珠子或微量滴定板。

[0086] 进一步地,本发明涉及使用上述的核酸分子分析成人型肠乳糖酶缺乏的存在或易

感性。

[0087] 如上详细所述,所述核酸分子同时允许不存在成人型肠乳糖酶缺乏的病情或其易感性的分析。

[0088] 此外,本发明涉及一种试剂盒,该试剂盒包含在一个或一个以上容器中的上述核酸分子,上述的引物或引物对,上述的载体,和 / 或上述的抗体、适体和 / 或噬菌体。

[0089] 本发明还涉及上述的核酸分子或载体在基因治疗中的应用。

[0090] 上述结合本发明的载体讨论的基因治疗方法,这里同样适用。值得注意的是,根据本发明,上述的核酸分子片段也可用于基因治疗方法中,具体地,SEQ ID NO :3 至 4 描述了上述片段。所述的片段包含上述 (c) 中定义的位点 -13910 的核苷酸 (也显示于 SEQ ID NO :3 中) 或上述 (d) 中定义的位点 -22018 的核苷酸 (也显示于 SEQ ID NO :4 中)。优选地,所述的片段包含至少 200,至少 250,至少 300,至少 400 和最优选至少 500 个核苷酸。

[0091] 在本发明的应用的一个优选实施方式中,所述的基因治疗方法用于治疗或预防成人型肠乳糖酶缺乏。

## 附图说明

[0092] 图 1 :研究的芬兰人成人型肠乳糖酶缺乏家族。黑体符号表示肠乳糖酶缺乏的个体,星号 (\*) 表示没有获得样品,问号 (?) 表示未知疾病状态。↑ 表示鉴定 SNP 的测序中使用的个体 (表 2)。

[0093] 图 2 :成人型肠乳糖酶缺乏位点的物理图谱。图中显示 BAC 克隆位于水平线上。粗黑箭头表示 LPH、MCM6 和 DARS 三个基因,箭头的尖端指向上述黑盒子上的基因的 3' 端。图中显示使用十个多态性微卫星标记精确定位位点。水平线中的倒斜线表示毗连群序列中的缺口。通过将自以前描述的 PAC 库分离的 PAC 106020<sup>40</sup> 与该缺口桥连确定标记 D2S2169 的位置。图中显示 MCM6 基因的结构包括与乳糖酶持续性基因型相关的变异体的位置在位于 LPH 的第一个 ATG 的 5' 的 13.9kb 和 22kb 的内含子 9 和 13 内。

[0094] 图 3 :使用七种极为相似的微卫星标记对来自芬兰成人型肠乳糖酶缺乏家族的持续性染色体进行的扩展单倍型分析。隐藏了代表祖辈创始持续性染色体的单倍型。只显示了也存在于持续性染色体中的非持续性染色体的单倍型。在祖辈重组的基础上,可将成人型肠乳糖酶缺乏位点限于标记 LPH1 和 AC3 中的 47kb 的区域。

[0095] 图 4 :包含于 MCM6 基因 (3220bp) 的内含子 13 的序列中并包含 -13910 位点上的 SNP 的序列,其中对乳糖酶持续性特异的胸苷由胞苷取代。所述的位置由小字母表示。本序列为 SEQ ID NO :1。

[0096] 图 5 :包含于 MCM6 基因 (1295bp) 的内含子 9 的序列中并包含 -22018 位点上的 SNP 的序列,其中对乳糖酶持续性序列特异的腺苷由鸟苷取代。所述的位置由小字母表示。本序列为 SEQ ID NO :2。

[0097] 图 6 :包含于 MCM6 基因 (3220bp) 的内含子 13 中并包含 -13910 位点上的胸苷的序列。其中所述的位置由小字母表示。本序列为 SEQ ID NO :3。

[0098] 图 7 :包含于 MCM6 基因 (1295bp) 的内含子 9 的序列中并包含 -22018 位点上的腺苷。所述的位置由小字母表示。序列为 SEQ ID NO :4。

[0099] 图 8 :包含于 MCM6 基因 (3220bp) 的内含子 13 中并包含 -13910 位点上的 SNP 的

序列,其中对乳糖酶持续性序列特异的胸苷由胞苷取代。所述的位置由小字母表示。本序列为 SEQ ID NO :5。

[0100] 图 9 :包含于 MCM6 基因 (1295bp) 的内含子 9 中并包含 -22018 位点上的 SNP 的序列,其中对乳糖酶持续性序列特异的腺苷由鸟苷取代。所述的位置由小字母表示。本序列为 SEQ ID NO :6。

## 具体实施方式

[0101] 下列实施例说明本发明。

[0102] 实施例 1 :连锁和连锁不平衡分析

[0103] 分析九个扩展的芬兰肠乳糖酶缺乏家族的位于 2q21 上 LPH 基因侧翼 D2S114 和 D2S2385 之间的七个多态性微卫星标记 (图 1)。发现与标记 D2S314、D2S442、D2S2196 和 D2S1334 连锁的重要证据,由 D2S2196 标记获得的最大优势对数评分值在  $\theta = 0$  时为 7.67 (表 1)。由标记 D2S114 (家族 B, IV3) 检测专性重组事件,从而确定乳糖酶持续性 / 非持续性位点的着丝粒边界;由标记 D2S2385 (家族 B, IV17) 检测专性重组事件 (图 1, 表 1), 从而确定该位点的端粒边界。为精确定位该关键区域,分析另外九个多态性标记 (表 1)。根据所检测到的连锁,以等位基因频率和重组分数作为不良参数监测该区域的连锁不平衡 (LD)<sup>16-17</sup>。跨越约 200kb 间隔的九个标记中的六个 (LPH13、LPH2、LPH1、AC3、AC4 和 AC10) 显示 LD ( $p < 10^{-4}$ ) 的高度重要的证据,而 LPH 基因 3' 的标记没有显示任何 LD 的证据 (表 1)。LPH2 和 AC3 两种标记显示乳糖酶持续性等位基因的最显著的连锁不平衡 ( $p < 10^{-7}$ )。

[0104] 家族研究对象由九个最早由 Sahi 研究的扩展的芬兰家谱构成<sup>5</sup>。19 世纪 70 年代针对成人型肠乳糖缺乏酶测验了所有家族研究对象。通过收集年轻一代家族成员的 DNA 扩展了本研究的家族研究对象。本研究中的家族研究对象共由 194 个构成 (图 1)。使用乙醇的乳糖耐受试验 (LTTE)<sup>4-5</sup> 确定了除 49 个人外所有家族成员的基因型状态。由血清 IgA 抗组织转谷氨酰胺酶法<sup>45</sup> 排除所有患有麦胶性肠病的患者。经允许后根据标准程序收集所有受试家庭成员的血样<sup>46</sup>,并提取其中的 DNA。作为病例 - 对照研究,对赫尔辛基大学医院的自空肠活检标本分离并测定空肠活检标本中双糖酶活性<sup>47</sup> 的 196 个随机 DNA 样本进行测序。根据标准程序分离肠活检标本的 DNA<sup>46</sup>。这一系列的样品包括 137 个乳糖酶持续性样品和 59 个乳糖酶非持续性样品。此外,对由那不勒斯大学 M. Rossi 惠赠的 9 个意大利人 DNA 样本、由波恩大学 M. Lentze 惠赠的 9 个德国人 DNA 样本和由汉城国立大学 J. K. Seo 惠赠的 22 个韩国人肠活检样本进行分析 (在表中 :23 个韩国人,9 个意大利人和 7 个德国人,其中一个德国人是韩裔)。诊断的依据是双糖酶活性的测定。最后,分析来自芬兰东部和西部小型教区的 938 个匿名芬兰献血者的 DNA 和 CEPH 家族的 109 个父辈的 DNA<sup>19</sup>,确定 C/T<sub>-13910</sub> 变异体在芬兰人群中的频率。此外采用标准程序<sup>48</sup>,分析了狒狒 (*Papio hamedryas ussinus*) 的肝活检样品的基因组 DNA。该研究经赫尔辛基大学医院伦理委员会 (Ethical Committees of the Helsinki University Hospital) 和芬兰红十字输血站 (Finnish Red Cross Blood Transfusion Service) 的许可。

[0105] 实施例 2 :扩展的单倍型分析

[0106] 第一阶段,如别处所述分析了 2q21 上 LPH 基因侧翼的十个高度多态性微卫星标记<sup>40,55</sup>。简言之,采用如下基因距离,分析了来自 Généthon ResourceCenter<sup>55</sup> 的乳糖酶基因附

近的 2q 上十个高度多态性微卫星标记 :cen-D2S114-1cM-D2S1334-0cM-D2S2196-0cM-D2S442-2cM-D2S314-2cM-D2S2385-1cM-D2S2288-1cM-D2S397-1cM-D2S150-1cM-D2S132。大多数上述标记的顺序主要自 Généthon 图谱提供的染色体 2 (Chumakov 等, 1995<sup>56</sup>) 的物理 YAC 毗连群获得。在总体积为 15  $\mu$  l 并包含 12ng 模板 DNA、5pmol 引物、0.2mM 每种核苷酸、20mM TrisHCl (pH 8.8)、15mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>、1.5mM MgCl<sub>2</sub>、0.1% 吐温 20、0.01% 明胶和 0.25U Taq 聚合酶 (Dynazyme, Finnzymes) 的溶液中进行 PCR。其中一种引物 5' 端由 <sup>32</sup>P- $\gamma$  ATP 进行放射性标记。该反应在多孔微量滴定板中共进行 35 个循环 :94°C 下变性 30 秒, 不同的引物在不同温度下退火 30 秒, 并在 72°C 下延伸 30 秒 ;将变性定为 3 分钟而最后的延伸定为 5 分钟。扩增的片段在 6% 聚丙烯酰胺凝胶上分离并进行自显影。

[0107] 第二阶段, 使用重复序列过滤程序 (Repeat Masker program, <http://ftp.genome.washington.edu/cgi-bin/RepeatMasker>) 将 LPH 基因上构建的毗连群内另外 9 个微卫星标记自公开的 BACs 基因组序列 (NH034L23、NH0318L13、NH0218L22 和 RP11-329I1) 中鉴定出来。合成重复序列侧翼的引物。PCR 条件如在他处<sup>40</sup> 描述的那样。该扩增的片段在 6% 聚丙烯酰胺凝胶上分离并进行自显影。

[0108] 使用连锁程序包 (LINKAGE program package) 的 MLINK 选项<sup>49</sup> 计算成对优势对数评分值。假定成人型肠乳糖酶缺乏的常染色体隐性遗传完全外显, 重组分数无性别差异并且该疾病的等位基因频率为 0.4。由于芬兰人群中显示该病情的年龄为 20 岁以上<sup>5-6</sup>, 因此本研究只包括 20 岁以上的个体。认为未被 LTTE 确定的个体的疾病状态为未知。使用参数连锁分析的 Downfreq 程序<sup>49</sup> 评估来自家族研究对象的标记的等位基因频率和杂合性。此外, 假定遗传模式为常染色体隐性的, 进行假标记连锁和连锁不平衡分析<sup>16</sup>。以等位基因频率和重组分数作为不良参数, 根据检测到的连锁检验 LD<sup>16, 49</sup>。表 1 显示这些分析的 P 值。使用如下顺序的微卫星标记人工地构建单倍型 :LPH1-LPH2-LPH13-AC7-AC3-AC4-AC5 (图 3)。家族研究对象中共 54 个非持续性染色体和 33 个持续性染色体可用于单倍型分析。

[0109] 将所述关键区域的四个 BAC 克隆 NH0034L23、NH0218L22、NH0318L13 和 329I10 组装成一个不间断的序列部分以确定紧密连锁的标记的顺序。该毗连群自标记 AC8 延伸至天门冬氨酰-tRNA 合酶 (DARS) 基因外显子 10, 共覆盖 222,5kb (图 2)。根据该连锁区域的物理图谱, 构建具有七个标记的单倍型, 其覆盖 150kb 间隔 (cen-LPH13-LPH2-LPH1-AC7-AC3-AC4-AC5-tel) (图 3)。一种主要的单倍型存在于 20 个持续性等位基因 (60%) 而存在于 3 个非持续性等位基因 (5%) 中, 而在非持续性等位基因中观察到大量的单倍型。非持续性等位基因中余下的 40% 的单倍型与祖传单倍型不同, 其与由历史重组事件造成的单倍型损坏一致。在保存的单倍型分析的基础上, 可将乳糖酶持续性位点限于标记 LPH1 和 AC3 之间 47kb 的间隔中 (图 3)。

[0110] 实施例 3 :成人型肠乳糖酶缺乏位点的序列分析

[0111] 在来自九个肠乳糖酶缺乏家族的一些成员的基因组 DNA 的重叠 PCR 片段中扩增标记 LPH1 和 AC3 之间 47kb 的区域, 并进行测序。该区域包括覆盖所述关键性 47kb 区域的 36kb 的微型染色体保持 (MCM6) 基因<sup>18</sup> (图 2)。除了共 52 个变异体外, 在 MCM6 基因编码区中未检测到变异 ;在该关键性 47kb 区域中鉴定了 43 种 SNP 和 9 种缺失 / 插入多态性 (表 2)。只有两种变异体 (C/T<sub>-13910</sub> 和 G/A<sub>-22018</sub>) 与芬兰家族的乳糖酶持续性 / 非持续特征相关 (表 2 和 3)。第一个相关的变异体 C/T<sub>-13910</sub> 在 MCM6 基因的内含子 13 内, 位于距离 LPH 基

因的第一个 ATG 密码子的 -13910 位点。第二个相关的变异体 G/A<sub>-22018</sub> 在 MCM6 基因的内含子 9 内,位于距离 LPH 基因第一个 ATG 密码子的 -22018 位点(图 2)。这两种变异体相距 8kb,与九个扩展的芬兰家族中的成人型肠乳糖酶缺乏完全共分离。所有肠乳糖酶缺乏(非持续性)家族成员均为纯合 C<sub>-13910</sub> 和 G<sub>-22018</sub>(表 3)。有趣地是,两种变异体均位于重复元件中,C/T<sub>-13910</sub> 位于 L2 衍生元件中而 G/A<sub>-22018</sub> 位于 Alu 元件中。

[0112] 实验性地,在第一阶段,对来自家族研究对象并在关键区域内共用一个相似的单倍型的 3 个非持续性个体、2 个纯合持续性个体以及 2 个杂合持续性个体进行测序(图 1)。使用软件 Sequencher 4(Gene Codes Corporation)将覆盖成人型肠乳糖酶缺乏关键区域的公开的 BACs 初稿基因组序列:NH0034L23、NH0218L22、NH0318L23 和 RP-329I10 组装成一个毗连群。设计位于标记 LPH1 和 AC2 之间跨越该关键区域的寡核苷酸引物(以下描述一系列寡核苷酸引物)。在具有基因组 DNA(100ng)、引物(各 20ng)、dNTPs(200 μM)、标准缓冲液中的 0.5U Taq 聚合酶(Dynazyme, Finnzymes)的体积为 50 μl 溶液中进行 PCR 扩增。多数 PCR 在如下 PCR 循环条件下扩增:94℃ 下进行起始变性 3 分钟,然后在 94℃,30 秒;55℃,30 秒;72℃,1.25 分钟下循环 35 次,最后在 72℃ 下延伸 10 分钟,当 PCR 产物的大小长于 1kb 时,我们使用 Dynazyme 延伸试剂盒(条件见下述)。使用 BigDye terminator chemistry(PE Biosystems)对纯化的 PCR 产物(15-40ng)进行循环测序。由 ABI 序列分析仪 3.3(PE Biosystems)和 Sequencher 4.1(Gene Codes)分析数据。

[0113] 由测序法检测乳糖酶变异体

[0114] 在具有基因组 DNA(100ng)、引物(各 20ng)、dNTPs(200 μM)、标准缓冲液中的 0.5U Taq 聚合酶(Dynazyme, Finnzymes)的体积为 50 μl 溶液进行 PCR 扩增。在如下 PCR 循环条件下扩增两种 PCRs:94℃ 下进行起始变性 3 分钟,然后在 94℃,30 秒;55℃,30 秒;72℃,1.25 分钟下循环 35 次,最后在 72℃ 延伸 10 分钟。由酶反应纯化 PCR。使用 BigDye terminator chemistry(PE Biosystems)对纯化的 PCR 产物(15-40ng)进行循环测序。由 ABI 序列分析仪 3.3(PE Biosystems)和 Sequencher 4.1(Gene Codes)分析数据。

[0115] 通过固相小型测序法筛选乳糖酶变异体

[0116] 使用一种生物素化的引物(5'-Bio-CCTCGTTAATACCCACTGACCTA-3')和一种未经生物素化的引物(5'-GTCACCTTGATATGATGAGAGCA-3')扩增跨越 C/T<sub>-13910</sub> 变异体的 DNA 片段。在上述条件下使用一种生物素化的引物(5'-Bio-TGCTCAGGACATGCTGATCAA-3')和一种未经生物素化的引物(5'-CTACCCTATCAGTAAAGGCCTA-3')扩增跨越 G/A<sub>-22018</sub> 变异体的 DNA 片段。将 10 μl 的 PCR 产物捕获在链霉抗生素包被的微量滴定孔中(Lab systems, 芬兰)。经洗孔后,按照如 Syvänen 等(Am J Hum Genet. (1993), 52, 46-59)和 Syvänen 与 Landegren(Hum Mutat. (1994), 3, 172-9)所述的方法将结合的 DNA 变性。在每个孔中加入含有 10 皮摩尔(pmoles) C/T<sub>-13915</sub>(5'-GGCAATACAGATAAGATAATGTAG-3')、G/A<sub>-22018</sub>(5'-AAAAACAGCATTCTCAGCTGGGC-3')的小型测序引物和 0.1 μl 与乳糖酶非持续性等位基因对应的 H-dCTP、H-dGTP(115Ci/mmol; Amersham, 英国)或与乳糖酶持续性等位基因对应的 H-dTTP、H-sATP 和缓冲液中的 0.05 单位的 DNA 聚合酶(Dynazyme II, Finnzymes)的 50 μl 小型测序反应混合物。在 50℃ 下孵育微量滴定板 20 分钟,并洗涤各孔。洗脱检测引物,在液体闪烁计数器中测定洗脱的放射性(Rackbeta 1209, Wallac, 芬兰)。对每种 PCR 产物均进行两次平行的 每种 PCR 产物均进行两次平行的小型测序反应。

[0117] C/T<sub>-13910</sub> 变异体的 PCR 引物和检测引物为：

[0118] 正向 PCR 引物： GTCACCTTGATATGATGAGAGCA Tm 58 SEQ ID :8

[0119] 检测引物 GGCAATACAGATAAGATAATGTAG Tm 58 SEQ ID :10

[0120] 生物素化反向引物 Bio-CCTCGTTAATACCCACTGACCTA Tm 62 SEQ ID :9

[0121] 或 Bio-TAGGTCAGTGGGTATTAACGAGGT SEQ ID :7

[0122] G/A<sub>-22018</sub> 变异体的 PCR 引物和检测引物为：

[0123] 正向 PCR 引物： CTACCCTATCAGTAAAGGCCTA Tm 58 SEQ ID :12

[0124] 检测引物 AAAAACAGCATTCTCAGCTGGGC Tm 62 SEQ ID :14

[0125] 生物素化的反向引物 Bio-TGCTCAGGACATGCTGATCAA Tm 62 SEQ ID :13

[0126] 或 Bio-TTGATCAGCATGTCTGAGCA SEQ ID :11

[0127] 实施例 4 :监测病例 / 对照研究样本的 DNA 变异体

[0128] 将双糖酶活性分析作为肠乳糖酶缺乏的诊断性测验,在自 196 个分析了双糖酶活性的肠活检样本分离的 DNA 样品中分析 C/T<sub>-13910</sub> 和 G/A<sub>-22018</sub> 变异体的频率。共 59 个病例显示初级乳糖酶缺乏。59 个病例中的 6 个(表 3)为 G/A<sub>-22018</sub> 变异体的杂合 GA,余下的 53 个为纯合的 G 等位基因。所有 59 个样品均为变异体 C/T<sub>-13910</sub> 的纯合 C 等位基因。在 137 个显示乳糖酶持续性的个体中,发现 C/T<sub>-13910</sub> 和 G/A<sub>-22018</sub>,分别为 74 个是纯合的等位基因 T 和 A,63 个是杂合的 CT 和 GA 等位基因,没有 C 和 G 纯合等位基因(表 3)。

[0129] 为分析其他人群的这些变异体,自 40 个确诊为双糖酶缺乏的非芬兰人病例中分离 DNA 样品并进行测序:23 个病例来自韩国,9 个来自意大利、8 个来自德国。一个意大利病例为 G/A<sub>-22018</sub> 的杂合 GA,其余所有 39 个病例分别为 C/T<sub>-13910</sub> 和 G/A<sub>-22018</sub> 的纯合 CC 和 GG(表 3)。在表 7 中显示一项扩展的研究数据,其中代表 6 个不同人群中的 400 个个体的 C/T<sub>-13910</sub> 变异体的数据与生化证实的乳糖酶缺乏(乳糖酶非持续性)完全相关。在 401 个病例中,400 个病例中的 G/A<sub>-22018</sub> 变异体与乳糖酶非持续性相关。

[0130] 实施例 5 :乳糖酶持续性变异体 C/T<sub>-13910</sub> 的分子流行病学

[0131] 为监控芬兰人群中与乳糖酶缺乏相关的变异体的发病率,使用固相小型测序法<sup>19,20</sup>筛选来自芬兰西部早期定居区或东部晚期定居区的 938 个匿名芬兰献血者的 DNA 样品(表 4)。实验性地,跨越 C/T<sub>-13910</sub> 变异体的 DNA 片段使用一种生物素化的引物(5'-CCTCGTTAATACCCCTGACCTA-3')和一种未生物素化的引物(5'-GTCACCTTGATATGATGAGAGCA-3')进行扩增。对于 G/A<sub>-22018</sub>,我们使用一种生物素化的引物(5'-AGTCTGTGGCATGTGTCTTCATG-3')和一种未生物素化的引物(5'-TGCTCAGGACATGCTGATCAACT-3')在上述条件下进行扩增。将 10 μl 的 PCR 产物捕获在链霉抗生物素包被的微量滴定孔中(Lab systems,芬兰)。经洗孔后,如以前描述的方法<sup>19,20</sup>将结合的 DNA 变性。将含有 10pmoles G/A<sub>-22005</sub>(5'-GACAAAGGTGTGAGCCACCG-3')、G/A<sub>-13915</sub>(5'-GGCAATACAGATAAGATAATGTAG-3')的小型测序引物和 0.1 μl 与乳糖酶非持续性等位基因对应的 H-dCTP(115Ci/mmol;Amersham,英国)或与乳糖酶持续性等位基因对应的 H-dTTP 和缓冲液中的 0.05 单位 DNA 聚合酶(Dynazyme II,Finzymes)的 50 μl 小型测序反应混合物加入孔中。将微量滴定板在 50°C 下孵育 20 分钟,并洗涤各孔。洗脱检测引物,并在液体闪烁计数器中测定洗脱的放射性(Rackbeta 1209,Wallac,芬兰)。对每种 PCR 产物均进行两次平行的小型测序反应。推定的乳糖酶缺乏基因型 CC<sub>-13910</sub> 的总发病率(170 个)

为 18.1%，西部样品的发病率比东部样品低 (16.8% 比 18.9%) (表 4)。这些值与流行病学研究报告的讲芬兰语的芬兰人的 17% 的发病率 (由西向东递增)<sup>2</sup> 相符。对于 G/A<sub>-22018</sub> 多态性, 同样一套样品也经过基因型分析, 这两种 SNPs 间的 LD 由 D' 统计<sup>21</sup> 监测。发现它们在几乎完全的 LD 中 ( $D' = 0.98, p = 7.62 \times 10^{-11}$ , 表 5)。

[0132] 已知不同人群的肠乳糖酶缺乏的发病率差别很大, 从小于 5% 到几乎 100%<sup>3,6</sup>。为确定乳糖酶缺乏的发病率的这些变化是否与基因型 CC<sub>-13910</sub> 的分布有关, 对 CEPH 家族的父辈 DNA<sup>22</sup> 进行分析。CEPH 家族主要来自法国 (报告的乳糖酶缺乏的发病率为约 37%<sup>23</sup>) 和来自北欧的犹他人群 (报告的乳糖酶缺乏的发病率低于 5%<sup>24</sup>)。CEPH 家族的父辈的基因型显示法国家族中 41.2% (17 个样品中有 7 个) 具有基因型 CC, 而犹他家族中只有 7.6% (92 个样品中有 7 个) 具有基因型 CC (表 4)。尽管分析样品的数量较小, 但这些数字再次与这些人群的乳糖酶缺乏的流行病学研究中的数字<sup>23,24</sup> 相符。

[0133] 表 8 表明观察到的变异体发病率与所述人群的乳糖不耐受的频率很好地相符。

[0134] 实施例 6 : 乳糖酶持续性变异体 C/T<sub>-13910</sub> 的谱系学

[0135] 芬兰的芬兰家族单倍型的分析表明大多数芬兰人 (如果不是全部) 乳糖酶持续性等位基因来自同一祖先。使用连锁不平衡估计将该持续性等位基因引入芬兰人群的时间<sup>25</sup>。假设 20 年产生一代, 估计芬兰人群奠基突变的时间是约 9000-11400 年前 (表 6)。这与约 8000-9000 年前最早的芬兰大陆定居迹象<sup>26</sup> 较好地相符, 并与公元前 8000-10000 的乳业开始时间<sup>27</sup> 相当符合。更重要地, 不同人群中出现同样的持续性等位基因 DNA 变异体提示该变异体更古老, 而且该突变在分析的人群分化前已出现。

[0136] 为获得乳糖酶等位基因的亲缘遗传起源, 对狒狒 (*Papio Hamadryas*) 的 MCM6 基因的内含子 9 和部分内含子 13 的序列进行测序。在 G/A<sub>-22018</sub> 和 C/T<sub>-13910</sub> 狒狒的 DNA 中存在基因型 GG 和 CC。这提示代表非持续性等位基因 G 和 C 分别反映了该祖先基因的出现, 以及突变将该等位基因转化成持续性等位基因。LD 的鉴定和持续性等位基因中的共用单倍型相对于非持续性等位基因中的高度多样化的等位基因支持这一假设。

[0137] 实施例 7 : C/T 和 G/A 变异体的成对 LD

[0138] 使用 D' 统计法<sup>21</sup> 估计 C/T<sub>-13910</sub> 和 G/A<sub>-22018</sub> 的成对 LD。使用 EH 程序估计单倍型频率的最大可能性<sup>50</sup>。计算 D' 由 ( $D/D_{\text{最大}}, D/D_{\text{最小}}$ ) 的最大值计算: 其中不平衡测量  $D = h_{pq} - pq$ , 其中  $h_{pq}$  是每个位点具有稀有等位基因的单倍型的频率, p 和 q 是位点 1 和 2 的稀有等位基因的频率, 如果  $D > 0$ ,  $D_{\text{最大}} = \min p(1-p), q(1-q)$ ; 如果  $D < 0$ ,  $D_{\text{最小}} = -\min pq,$

$(1-p)(1-q)$ 。D' 与 0 的偏差的意义由具有 1df 的  $\chi^2$  分布的统计参数  $D^2 \sqrt{\frac{N}{P(1-p)q(1-q)}}$

测定<sup>21</sup>。

[0139] 基因登记号: 对 BACs NH0218L22、N0034L34、NH0318L13 和 RP11-329I10 分别为 AC012551、AC011893、AC011999 和 AC016516。人类多态性的登记号为: GenBank AF395607-AF395615。

[0140] 参考文献

[0141] 1. Flatz, G. & Rotthauwe, H. 人乳糖酶多态性: 乳糖吸收和吸收不良的生理学和遗传学 (The human lactase polymorphism: physiology and genetics of lactose

absorption and malabsorption.)Prog. Med. Genet. 2, 205-249 (1977)。

[0142] 2. Sahi, T., Isokoski, M., Jussila, J. & Launiala, K. 芬兰学龄儿童中的乳糖吸收不良 (Lactose malabsorption in Finnish children of school age.)Acta Paediatr Scand. 61, 11-16 (1972)。

[0143] 3. Wang, Y. 等, 儿童中乳糖酶的基因程控下调 (The genetically programmed down-regulation of lactase in children.)Gastroenterology. 114:1230-1236 (1998)。

[0144] 4. Sahi, T., Isokoski, M. Jussila, J., Launiala, K. & Pyörälä, K. 成人型乳糖酶吸收不良的隐性遗传 (Recessive inheritance of adult-type lactose malabsorption.) Lancet, 823-826 (1973)。

[0145] 5. Sahi, T. 选择性成人型乳糖酶吸收不良的遗传 (The inheritance of selective adult-type lactose malabsorption.)Scand. J. Gastroenterol. suppl. 30, 1-73 (1974)。

[0146] 6. Sahi, T. 成人型肠乳糖酶缺乏的遗传学和流行病学 (Genetics and epidemiology of adult-type hypolactasia)Scand. J. Gastroenterol. suppl. 202, 7-20 (1994)。

[0147] 7. Boll, W, Wagner, P. & Mantei, N. 成人型肠乳糖酶缺乏或乳糖酶持续性的人群中编码乳糖酶-根皮苷水解酶的染色体基因和 cDNAs 的结构 (Structure of the chromosomal gene and cDNAs coding for lactase-phlorizin hydrolase in human with adult-type hypolactasia or persistence of lactase.)Am. J. Hum. Genet. 48:889-902 (1991)。

[0148] 8. Mantei, N 等, 人和兔乳糖酶-根皮苷水解酶的完整一级结构: 酶的生物合成、膜锚定和进化含义 (Complete primary structure of human and rabbit lactase-phlorizin hydrolase: implications for biosynthesis, membrane anchoring and evolution of the enzyme.)EMBO J. 7, 2705-2613 (1998)。

[0149] 9. Wang, Y. 等, 乳糖酶持续性 / 非持续性多态性由顺式作用元件控制 (The lactase persistence/non-persistence polymorphism is controlled by a cis-acting element.)Hum. Mol. Genet. 4, 657-662 (1995)。

[0150] 10. Harvey, C. B., Pratt, W. S., Islam, I., Whitehouse, D. B. & Swallow, D. M. 乳糖酶基因的 DNA 多态性: 70kb 区域的连锁不平衡 (DNA polymorphisms in the lactase gene: linkage disequilibrium across the 70kb region.)Eur J. Hum. Genet. 3, 27-41 (1995)。

[0151] 11. Escher, J C 等, 成人中乳糖酶水平的分子基础 (Molecular basis of lactase levels in adult humans.)J. Clin. Invest. 89, 480-483 (1992)。

[0152] 12. Lloyd, M 等, 成人肠乳糖酶缺乏中肠乳糖酶的调控 (Regulation of intestinal lactase in adult hypolactasia.)J. Clin. Invest. 89, 524-529 (1992)。

[0153] 13. Fajardo, O., Naim, H. Y. & Lacey, S. W. 成人中乳糖酶多态性表达在信使 RNA 水平调控 (The polymorphic expression of lactase in adults is regulated at the messenger RNA level.)Gastroenterology 106, 1233-

[0154] 14. Luigi, M. 等, 成人型乳糖酶的嵌合性调控 (Mosaic regulation of lactase in human adult-type.)Gastroenterology 112, 1506-1514 (1997)。

[0155] 15. Rossi, M. 等, 成人中乳糖酶持续性相对于降低: 多因子事件参与断奶后的下调 (Lactase persistence versus decline in human adults: Multifactorial

events are involved in down-regulation after weaning.)Gastroenterology 112, 1506-1514(1997)。

[0156] 16. Göring, H. H. H. & Terwilliger, J. D. 存在错误 IV 时的连锁分析: 当遗传模式不明确时对家谱和单态的混合物上连锁和 / 或连锁不平衡的假标记共分析 (Linkage analysis in the presence of errors IV: Joint pseudomarker analysis of linkage and/or linkage disequilibrium on a mixture of pedigrees and singletons when mode of inheritance cannot be accurately specified.) Am. J. Hum. Genet. 66, 1310-1327 (2000)。

[0157] 17. Terwilliger, J. D. & Göring, H. H. H. 20 和 21 世纪的基因图谱: 统计学方法、数据分析和试验设计 (Gene mapping in the 20th and 21st centuries: Statistical methods, data analysis, and experimental design.) Hum. Biol. 72, 63-132 (2000)。

[0158] 18. Harvey, C. B. 等, 乳糖酶 - 根皮苷水解酶、LCT 在染色体 2q21 上的区域定位 (Regional localization of the lactase-phlorizin hydrolase, LCT, to chromosome 2q21.) Ann. Hum. Genet. 57, 179-185 (1993)。

[0159] 19. Syvänen, A-C., Sajantila, A., Lukka, M. 通过 PCR 和固相小型测序由双等位基因 DNA 标记分析鉴定个体 (Identification of individuals by analysis of biallelic DNA markers, using PCR and solid-phase minisequencing.) Am. J. Hum. Genet. 52, 46-59 (1993)。

[0160] 20. Syvänen, A-C & Landegren, U. 通过固相方法检测点突变 (Detection of point mutations by solid-phase methods.) Hum. Mutat. 3, 172-179 (1994)。

[0161] 21. Thompson, E. A., Deeb, S., Walker, D. & Motulsky, A. G. 检测紧密连锁标记间的连锁不平衡: AI-CIII 载脂蛋白基因的 RFLPs (The detection of Linkage disequilibrium between closely linked markers: RFLPs at the AI-CIII Apolipoprotein genes.) Am. J. Hum. Genet. 42, 113-124 (1998)。

[0162] 22. Dausset, J. 等, 人类多态性研究中心 (CEPH): 人类基因组协作性基因图谱 (Centre d'étude du polymorphisme humain (CEPH): Collaborative genetic mapping of human genome.) Genomics 6, 575-577 (1990)。

[0163] 23. Cuddenec, Y., 等, Delbrück, H & Flatz, G. 成人型乳糖酶基因型的分布 - 在一群 131 个新兵中乳糖吸收者和吸收不良者 (Distribution of the adult lactase phenotypes - lactose absorber and malabsorber - in a group of 131 army recruit) Gastroenterol. Clin. Biol. 6, 776-779 (1982)。

[0164] 24. McLellan, T., Jorde, L. B. & Skolnick, M. H. 犹他摩门教徒和相关人群的基因距离 (Genetic distance between the Utah Mormons and related populations.) Am. J. Hum. Genet. 36, 836-857 (1984)。

[0165] 25. Terwilliger, J. D. 一种有力的特性位点和一种或一种以上多态性标记位点的连锁不平衡的可能性分析法 (A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci.) Am. J. Hum. Genet. 56, 777-787 (1995)。

[0166] 26. Nunez, M. G. 早期芬兰定居的模型 (A model of the early settlement of

Finland.) *Fennoscandia archaeologica* IV, 3-18 (1997)。

[0167] 27. Simoons, F. J. 最初的成人乳糖非耐受和哺乳习惯: 一个生物学和文化相互关系的问题 II 一种文化历史假说 (Primary adult lactose intolerance and the milking habit: a problem in biological and cultural interrelations. II. A cultural historical hypothesis.) *Am. J. Dig. Dis.* 16, 695-710 (1970)。

[0168] 28. Varilo, T 等, 人类突变的年龄: 芬兰人群 CLN5 突变的家谱分析和连锁不平衡分析 (The age of human mutation: genealogical and linkage disequilibrium analysis of the CLN5 mutation in the Finnish population.) *Am. J. Hum. Genet.* 58, 506-512 (1996)。

[0169] 29. Hästbacka, J. 等, 隔离的奠基人群中的连锁不平衡图谱: 发育不良的芬兰人 (Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland.) *Nature genet.* 2: 204-211 (1992)。

[0170] 30. Harvey C. B. 等. 高加索人乳糖酶单倍型频率: 与乳糖酶持续性 / 非持续性多态性的联系 (Lactase haplotype frequencies in Caucasians: association with the lactase persistence/non persistence polymorphism.) *Ann Hum Genet* 62, 215-223 (1998)。

[0171] 31. Ohtani, K 等, 转录因子 E2F 介导的哺乳动物 MCM5 和 MCM6 基因的细胞生长调控表达 (Cell growth-regulated expression of mammalian MCM5 and MCM6 genes mediated by the transcription factor E2F.) *Oncogene* 18, 2299-2309 (1999)。

[0172] 32. Smith A. F. A. 人类基因组中分散重复的起源 (The origin of interspersed repeats in the human genome.) *Curr. Opin. Genet. Dev.* 6, 743-748 (1996)。

[0173] 33. Kazazian, H. H. & Moran J. V. L1 反转录转座子对人类基因组的影响 (The impact of L1 retrotransposons on the human genome.) *Nature Genet.* 19, 19-24 (1998)。

[0174] 34. Moran, J. V., DeBerardinis, R. J. & Kazazian, H. H. L1 反转录转座打乱的外显子 (Exon shuffling by L1 retrotransposon.) *Science* 283, 1530-1534 (1999)。

[0175] 35. Wei, W. 等, 人类 L1 反转录转座: 顺式优选性对反式互补 (Human L1 retrotransposition: cis preference versus trans complementation.) *Mol. Cell Biol.* 21, 1429-1439 (2001)。

[0176] 36. Donnelly, S. R., Hawkins T. E. & Moss, S. E. 一种具有调控哺乳动物基因的保守性核元件 (A conserved nuclear element with a role in mammalian gene regulation.) *Hum. Mol. Genet.* 卷 8, 9, 1723-1728 (1999)。

[0177] 37. Boeke, J. D. LINEs 和 Alus- 多腺嘌呤连接 (LINEs and Alus- the poly A connection.) *Nature Genet.* 16, 6-7 (1997)。

[0178] 38. Jurka, J. 序列模式表示一种酶参与哺乳动物反转录转座子的整合 (Sequence patterns indicate an enzymatic involvement in integration of mammalian retrotransposons.) *Proc. Natl. Acad. Sci. U. S. A.* 94, 1872-1877 (1997)。

[0179] 39. Savilahti E, Launiala K, Kuitunen P. 先天性乳糖酶缺乏 (Congenital lactase deficiency.) *Arch. Dis. Child.* 58, 246-252 (1983)。

[0180] 40. Järvelä, I 等, 将先天性乳糖酶缺乏的位点指定在 2q21 上, 位于肠乳糖

酶-根皮昔水解酶附近但与之分离 (Assignment of the locus for congenital lactase deficiency to 2q21, in the vicinity of but separate from the lactase-phlorizin hydrolase gene.) *Am. J. Hum. Genet.* 63, 1078-1085 (1998)。

[0181] 41. Simoons, F. J. 地理假设和乳糖吸收不良:对证据的衡重 (The geographic hypothesis and lactose malabsorption. A weighing of the evidence.) *Am. J. Dig. Dis.* 23, 963-980 (1978)。

[0182] 42. Flatz, G. & Rotthauwe, H. W. 人乳糖酶多态性:乳糖吸收和吸收不良的生理学和遗传学 (The human lactase polymorphism: physiology and genetics of lactose absorption and malabsorption.) *Prog. Med. Genet.* 2, 205-249 (1997)。

[0183] 43. McCracken, R. D. 乳糖酶缺乏:饮食演化的例子 (Lactase deficiency: an example of dietary evolution.) *Curr. Anthropol.* 12, 479-517 (1971)。

[0184] 44. Arola, H 等, 诊断肠乳糖酶缺乏和乳糖吸收不良 (Diagnosis of hypolactasia and lactose malabsorption.) *Scand. J. Gastroenterol. Suppl.* 202, 26-35 (1994)。

[0185] 45. Sulkanen, S 等, 检测乳糜泻的组织转谷氨酸酶自身抗体酶联免疫吸附测定 (Tissue transglutaminase autoantibody enzyme-linked immunosorbent assay in detecting celiac disease.) *Gastroenterology* 115 (6), 1322-1328 (1998)。

[0186] 46. Sambrook, J., Fritsch, E. F & Maniatis, T. 分子克隆:实验室手册 (Molecular cloning: a laboratory manual.) (第二版) 冷泉港实验室出版社, 冷泉港, 纽约 (1989)。

[0187] 47. Messer, M. & Dahlqvist A. 分析肠双糖酶的一步超微法 (A one-step ultramicro method for the assay of intestinal disaccharidases) *Anal. Biochem.* 14 (3), 376-92 (1996)。

[0188] 48. Cottingham, Jr. R. W., Idury R. M. & Schaffer, A. A. 快速序列基因连锁计算 (Faster sequential genetic linkage computations.) *Am. J. Hum. Genet.* 53, 252-263 (1993)。

[0189] 49. Göring, H. H. H. & Terwilliger, J. D. 存在错误 III 的连锁分析:以标记位点及其图谱为不良参数 (Linkage analysis in the presence of errors III: Marker loci and their map as nuisance parameters). *Am. J. Hum. Genet.* 66, 1298-1309 (2000)。

[0190] 50. Terwilliger, J. D & Ott, J. 人类基因分析手册 (Hand book of human genetic analysis.) 约翰霍普金斯大学出版社, 巴尔地摩, (1994)。

[0191] 51. Osborne 等, *Curr. Opin. Chem. Biol.* 1 (1997), 5-9。

[0192] 52. Stall 和 Szoka, *Pharm. Res.* 12 (1995), 465-483。

[0193] 53. Harlow 和 Lane “抗体, 实验室手册 (Antibodies, A Laboratory Manual)”, 冷泉港实验室出版社, 冷泉港, 美国, 1998。

[0194] 54. Higgins 和 Hames (编辑) “核酸杂交, 一种实际方法 (Nucleic acid hybridization, a practical approach)” IRL Press, 牛津 1985。

[0195] 55. Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, Weissenbach J. 基于 5264 个微卫星的人类基因组综合基因图谱 (A comprehensive genetic map of the human genome based on 5,264 microsatellites.) *Nature.* 1996 年 3 月 14 日; 380 (6570): 152-4。

[0196] 56. Chumakov IM, Rigault P, Le Gall I, Bellanne-Chantelot C, Billault A, Guillous S, Soularue P, Guasconi G, Poullier E, Gros I 等, 人类基因组 YAC 毗连群图谱 (A YAC contig map of the human genome.) Nature. 1995 年 9 月 28 日 ;377(6547 增刊) : 175-297。

[0197] 表 1. 成人型肠乳糖酶缺乏家族的连锁和连锁不平衡分析 (黑体字显示精确定位标记)

[0198]

标记	$\theta$ 的优势对数评分值 (Z)					P值 <sup>a</sup>
	0.0	0.1	0.2	0.3	0.4	
D2S114	$-\infty$	2.44	1.92	1.13	<b>0.41</b>	0.87195
P6112	2.76	2.20	1.45	0.75	<b>0.22</b>	0.66207
D2S1334	3.15	2.45	1.61	0.84	0.25	0.91039
AC8	<b>2.26</b>	1.99	1.36	0.71	0.21	0.53670
LPH13	3.67	2.94	1.96	1.03	0.31	$4 \times 10^{-6}$
LPH2	4.09	3.07	2.00	1.00	0.26	$5.7 \times 10^{-7}$
LPH1	5.91	4.52	2.96	1.53	0.46	$5 \times 10^{-6}$
AC7	3.63	2.60	1.66	0.83	0.23	0.03471
AC3	6.63	4.88	3.16	1.61	0.44	$3.2 \times 10^{-8}$
AC4	3.07	2.22	1.42	0.71	0.19	$4 \times 10^{-5}$
AC5	5.33	4.10	2.72	1.39	0.39	0.02166
AC10	6.60	4.99	3.25	1.65	0.46	$1 \times 10^{-5}$
D2S2196	7.67	5.62	3.62	1.85	0.54	0.00010
D2S442	3.81	3.08	2.08	1.03	0.27	0.22805
D2S314	4.22	3.61	2.50	1.37	0.45	0.27535
D2S2385	$-\infty$	2.79	1.92	1.01	0.28	0.46457

[0199] a :p 值由连锁不平衡测验给出的连锁产生<sup>16,49</sup>

[0200] 表 2 :芬兰家族的成人型肠乳糖酶缺乏位点中鉴定的变异体

[0201]

位点 <sup>a</sup>	变异体	乳糖酶持续性 (纯合)		乳糖酶持续性 (杂合)		乳糖酶非持续性		
		BIV4	AIV3	BIV8	CIV3	BIV9	DIV4	EIII2 <sup>b</sup>
-694	A→G	AA	AA	AG	AA	GG	N <sup>c</sup>	AA
-1640/50	T <sub>13</sub> →T <sub>12</sub>	T <sub>13/13</sub>	T <sub>13/13</sub>	T <sub>13/13</sub>	T <sub>13/13</sub>	T <sub>13/13</sub>	T <sub>12/12</sub>	T <sub>12/12</sub>
-2131	C→T	CC	CC	CT	CC	TT	CT*	TT
-3058/72	T <sub>15</sub> →T <sub>16</sub>	T <sub>15/15</sub>	T <sub>15/15</sub>	T <sub>15/15</sub>	T <sub>15/15</sub>	T <sub>15/15</sub>	T <sub>16/16</sub>	T <sub>16/16</sub>
-3075	G→T	GG	GG	GG	GG	GG	GG	TT
-4480	T→A	TT	TT	TA	TT	AA	TT	TT
-5440	C→T	CC	CC	CT	CC	TT	CC	CC
-5926	A→T	AA	AA	AA	AA	AA	TA	TT
-8540	G→A	GG	GG	GA	GA	AA	AG	AA
-8630	C→G	CC	CC	CG	CG	GG	GC	GG
-13495	T→C	TT	TT	TC	TT	CC	CT	CC
-13910	T→C	TT	TT	TC	TC	CC	CC	CC
-15239	G→A	GG	GG	GA	GG	AA	AG	AA
-15862	T→C	CC	CC	CT	CC	TT	TC	TT
-16568/79	T <sub>11</sub> →T <sub>12</sub>	T <sub>11/11</sub>	T <sub>11/11</sub>	T <sub>11/12</sub>	T <sub>11/11</sub>	T <sub>12/12</sub>	T <sub>11/11</sub>	T <sub>12/12</sub>
-16888	A→G	AA	AA	GA	AA	GG	GA	GG
-17300	C→T	CC	CC	CC	CC	CC	CT	TT
-19044	T→C	TT	TT	TC	TT	CC	CT	CC
-19519	T→C	TT	TT	TC	TT	CC	TT	TT
-20077	C→G	CC	CC	CG	CC	GG	GC	GG
-20486	G→A	GG	GG	GA	GG	AA	GG	GG
-21721/28	A <sub>7</sub> →A <sub>6</sub>	A <sub>7/7</sub>	A <sub>7/7</sub>	A <sub>7/7</sub>	A <sub>7/7</sub>	A <sub>7/7</sub>	A <sub>7</sub> ,A <sub>6</sub>	A <sub>7/7</sub>
-21731	A→C	AA	AA	AA	AA	AA	CC	AA
-21736/43	A <sub>9</sub> →A <sub>8</sub>	A <sub>9/9</sub>	A <sub>9/9</sub>	A <sub>9</sub> ,A <sub>8</sub>	A <sub>9/9</sub>	A <sub>8/8</sub>	A <sub>8/8</sub>	A <sub>8/8</sub>
-22018	G→A	AA	AA	AG	AG	GG	GG	GG
-22741	C→T	CC	CC	CC	CC	CC	N	TT
-22788	A→G	AA	AA	AG	AA	GG	N	GG
-23069	A→G	AA	AA	AG	AA	GG	N	GG
-23442	A→G	AA	AA	AA	AA	AA	N	GG
-23771	T→C	TT	TT	TT	TT	TT	N	CC
-25093/23	△30bp	△△	△△	△△	△△	△△	N	II
-27310	A→G	AA	AA	AG	AA	GG	GA	GG
-27480	G→A	GG	GG	GA	GG	AA	AG	AA
-27807	A→C	AA	AA	AA	AA	AA	AC	CC
-30183	A→G	AA	AA	AG	AA	GG	AA	AA
-31268	A→G	AA	AA	AG	AA	GG	AA	AA

[0202]

-31342	T→C	TT	TT	TT	TT	TT	CT	CC
-33645	C→T	CC	CC	CT	CC	TT	CC	CC
-35176	T→C	TT	TT	TC	TT	CC	CT	CC
-36254	C→T	CC	CC	CT	CC	TT	TC	TT
-36296	G→T	TT	TT	TG	TT	GG	TG	N
-36501	A→T	AA	AA	AT	AA	TT	AT	N
-36506/14	△9bp	△△	△△	△I	△△	II	△I	N
-36671/77	T <sub>7</sub> →T <sub>6</sub>	T <sub>7/7</sub>	T <sub>7/7</sub>	T <sub>7/6</sub>	T <sub>7/7</sub>	T <sub>6/6</sub>	T <sub>7/7</sub>	T <sub>7/7</sub>
-37565	T→G	TT	TT	TG	TT	GG	GG	TG
-38276	G→C	GG	GG	GC	GG	CC	GG	GG
-39036	G→C	GG	N	GC	N	CC	N	N
-40608	G→C	GG	GG	GG	GG	GG	GC	CC
-41590	T→C	TT	TT	TC	TT	CC	CT	CC
-42081/82	△AG	AG	AG	AG/△	AG	△△	AG	AG
-42618	T→C	TT	TT	TC	TT	CC	TT	TT
-42893	G→A	GG	GG	GA	GG	AA	GG	GG

[0203] a:该数字来自使用 BACs NH034L23、NH0218L22、NH0318L13 和 RP11-329I10 经编辑后的基因组序列的 LPH 基因的起始翻译密码子 (ATG); b:自研究的芬兰家族测序并在图 1 中以箭头表示的个体; c:未确定

[0204] 表 3. C/T<sub>-13910</sub> 和 G/A<sub>-22018</sub> 基因型在乳糖酶持续性 / 非持续性等位基因中的分布

[0205]

	基因型	C/T <sub>-13910</sub>			G/A <sub>-22018</sub>			总数
		CC	CT	TT	GG	GA	AA	
家庭成员	乳糖酶非持续性	45	0	0	45	0	0	45
	乳糖酶持续性	0	32	13	0	32	13	45
病例-对照样本 芬兰人	乳糖酶非持续性	59	0	0	53	6	0	59
	乳糖酶持续性	0	63	74	0	63	74	137
非芬兰人 <sup>a</sup>	乳糖酶非持续性	40	0	0	39	1	0	40
	乳糖酶持续性	0	5	0	0	5	0	5
总数	乳糖酶非持续性						0	144
	乳糖酶持续性							187

[0206] a:由 23 个韩国人,9 个意大利人和 7 个德国人的组成的非芬兰人样本

[0207] 表 4 :C/T-13910 变异体在人群样本中的发病率

[0208]

分析的DNA样品	基因型			总数	等位基因频率 (%)		(CC) 基因型%
	CC	CT	TT		C	T	
I. 芬兰人群							
1. 东部区	108	287	176	571	0.440	0.560	18.9%
2. 西部区	62	159	146	367	0.385	0.615	16.8%
总数	170	446	322	938	0.418	0.582	18.1%
II. CEPH父辈							
1. 犹他家族	7	33	52	92	0.255	0.745	7.6%
2. 法国家族	7	9	1	17	0.676	0.324	41.2%

[0209] 共有来自芬兰东部和西部小教区的芬兰匿名献血者的938个DNA样品和来自CEPH父辈的109个DNA样品。CC等位基因的基因型频率反映人群中的肠乳糖酶缺乏的发病率。

[0210] 表5:随机芬兰样品中C/T-13910和G/A-22018变异体间的LD

[0211]

	C/T-13910 基因型			总数	D'	$\chi^2(1df)$	p 值
	CC	CT	TT				
G/A-22018 基因型							
GG	162	2	1	165			
GA	6	440	3	449			
AA	2	4	318	324			
总数	170	446	332	938	0.984	42.41	$7.62 \times 10^{-11}$

[0212] 使用D'统计法<sup>18</sup>计算LD, p值是如方法<sup>18</sup>中描述的D'距离0的显著性。

[0213] 表6:使用DISLAMB程序对C/T-13910变异体引入芬兰人群的估计

[0214]

标记 等位基因	AC3		LPH2	
	乳糖酶持续性	乳糖酶非持续性	乳糖酶持续性	乳糖酶非持续性
1	0	1	0	1
2	31	10	0	20
3	0	1	0	14
4	2	9	32	15
5	0	31	0	2
$\lambda^a$	0.838		0.999	
$\theta^b$	0.00031(0.000038-0.00099)		0.0000(0.00000-0.00052)	
$n^c$	570		450	

[0215] a:  $\lambda$  是疾病染色体中一种特定等位基因(乳糖酶持续性等位基因)相对于其人群频率(0.60)的增长比例。b:  $\theta$  是重组分数,假定1cM = 1Mb,由突变与最近的标记的距离反映。c: n是利用公式  $\lambda = \infty(1-\theta)^n$  自从奠基突变引入人群后的世代数。d: 当  $\theta = 0$  和  $\infty = 1$ , 假设的用于计算的等位基因。

[0216] 表7生化确定样本的乳糖酶不耐受变异体的发病率

[0217]

人群	数目	C/T <sub>13910</sub>			G/A <sub>22018</sub>		
		CC	CT	TT	GG	GA	AA
1. 芬兰人							
乳糖酶持续性	182	0	95	87	0	95	87
乳糖酶非持续性	116	116	0	0	110	6	0
2. 意大利人							
乳糖酶持续性	7	0	7	0	0	7	0
乳糖酶非持续性	23	23	0	0	22	1	0
3. 德国人							
乳糖酶持续性	0	0	0	0	0	0	0
乳糖酶非持续性	8	8	0	0	8	0	0
4. 索马里人							
乳糖酶持续性	0	0	0	0	0	0	0
乳糖酶非持续性	42	42	0	0	42	0	0
6. 韩国人							
乳糖酶持续性	0	0	0	0	0	0	0
乳糖酶非持续性	23	23	0	0	23	0	0
总数	401	212	102	87	205	109	87

[0218] 表 8 不同人群样品中乳糖耐受变异体的发病率

[0219] 人群	[0220] 数量	[0221] 基因型						[0222] 乳糖酶持续性等位基因	[0223] 发病率%
		C/T13910			G/A22018				
		CC	CT	TT	GG	GA	AA		
[0223] 韩国人	23	23	0	0	23	0	0	0	*
[0224] 法国人	17	7	9	1	6	10	1	59	*
[0225] 巴斯克人	85	7	44	34	13	35	37	92	*
[0226] 意大利南部人	100	89	11	0	88	12	0	11	*
[0227] 索马里人	79	74	5	0	78	1	0	6	
[0228] 犹他人	92	7	33	52	7	30	55	92	*
[0229] 非洲裔美国人	96	76	15	5	78	12	5	21	*
[0230] 摩洛哥人 (Marroccans)	90	62	25	3	65	22	3	31	*
[0231] 非洲萨拉魁人	57	29	26	2	28	26	3	49	*
[0232] (Sarawhi, African)									
[0233] 萨米人 (Saami)	30	20	10	0	21	9	0	33	*
[0234] 西藏人	23	23	0	0	23	0	0	0	
[0235] 芬兰东部人	571	108	287	176	107	288	176	81	*
[0236] 芬兰西部人	367	62	159	146	58	161	148	83	*
[0237] 芬兰 - 乌戈尔族									
[0238] (Finn-ugrian tribes)									

[0239]	Xan	20	19	1	0	19	1	0	5	
[0240]	Xm	20	19	1	0	19	1	0	5	
[0241]	曼西人 (Mansi)	22	20	2	0	20	2	0	9	
[0242]	拉口米人 (Lkomi)	10	7	3	0	7	3	0	30	
[0243]	爱若扎人 (Erza)	30	17	10	3	19	9	2	43	
[0244]	莫克沙人 (Moksa)	30	13	17	0	14	16	0	57	*
[0245]	Udmort	30	12	16	2	11	15	4	60	*
[0246]	巴基斯坦族									
[0247]	(Pakistanian tribes)									
[0248]	卡拉什人 (Kalash)	30	30	0	0	28	2	0	0	
[0249]	布鲁受 (Burusho)	30	29	1	0	27	3	0	3	
[0250]	哈扎拉人 (Hazara)	14	13	1	0	11	3	0	7	
[0251]	克什米尔人 (Kashmiri)	20	15	5	0	14	6	0	25	
[0252]	莫克兰尼俾路支人	29	19	10	0	19	8	1	34	
[0253]	(Makrani Baluch)									
[0254]	布拉灰人 (Brahui)	30	17	10	3	16	11	3	43	
[0255]	莫克兰尼人, 准黑人	29	16	10	3	16	10	3	45	
[0256]	(Makrani, Negroid)									
[0257]	帕坦人 (Pathan)	29	12	16	1	13	14	2	59	*
[0258]	印度人	29	11	13	5	10	12	5	62	*
[0259]	总数	2032								

[0260] \*乳糖酶持续性等位基因的发病率与报告的乳糖酶持续性等位基因的发病率 (Simoons Fj. 地理假说和乳糖消化不良 (The geographic hypothesis and lactose malabsorption) Am J Dig Dis 1978;23(11):963-80) 很好相符。

[0001]

## 序 列 表

&lt;110&gt; 国家公共卫生学院

&lt;120&gt; 鉴定与成人型肠乳糖酶缺乏相关的DNA变异体

&lt;130&gt; P04DE012901

&lt;150&gt; EP 01 11 9377.8

&lt;151&gt; 2001-08-10

&lt;150&gt; EP 01 11 9528.6

&lt;151&gt; 2001-08-14

&lt;150&gt; US 60/315,955

&lt;151&gt; 2001-08-31

&lt;160&gt; 14

&lt;170&gt; PatentIn version 3.1

&lt;210&gt; 1

&lt;211&gt; 180

&lt;212&gt; DNA

&lt;213&gt; 智人 (Homo sapiens)

&lt;400&gt; 1

acctttcatt caggaaaaat gtacttagac cctacaatgt actagtaggc ctctgcgctg 60

gcaatacaga taagataatg tagcccctgg cctcaaagga actctctctcc ttaggttgca 120

ttgtataat gtttgatttt tagattgttc tttgagccct gcattccacg aggataggtc 180

&lt;210&gt; 2

&lt;211&gt; 180

&lt;212&gt; DNA

&lt;213&gt; 智人

&lt;400&gt; 2

taagaacatt ttacactctt cagtataaag aagtcagaat acccctaccc tatcagtaaa 60

ggcctataag ttaccattaa aaagatgtcc ttaaaaacag cattctcagc tgggcgcggt 120

ggctcacacc tttgtcccag tactttggga agccgaggig ggtggatcac ctgaggtcag 180

&lt;210&gt; 3

&lt;211&gt; 3213

[0002]

&lt;212&gt; DNA

&lt;213&gt; 智人

&lt;400&gt; 3

```

atcagagtca ctttgatatg atgagagcag agataaacag atttgttgca tgtttttaat 60
ctttggtatg ggacatacta gaattcactg caaatacatt tttatgtaac tgttgaatgc 120
tcatacgacc atggaattct tccctttaa gagcttggta agcatttgag tgtagttggt 180
agacggagac gatcacgtca tagtttatag agtgcataaa gacgtaagtt accatttaat 240
acctttcatt caggaaaaat gtacttagac cctacaatgt actagtaggc ctctgcgctg 300
gcaatacaga taagataatg tagtccctgg cctcaaagga actctcctcc ttaggttga 360
tttgtataat gtttgatttt tagattgttc tttgagccct gcattccacg aggataggtc 420
agtgggtatt aacgaggtaa aaggggagta gtacgaaagg gcattcaagc gtcccatctt 480
cgcttcaacc aaagcagccc tgcgttttcc tagttttatt aataggtttg atgtaaggtc 540
gtctttgaaa agggggtttg gctttttttt acagtgtgac tgaggataaa ttataaaaa 600
gggaaatgta tggcatggtg agttttttca catacatcct tgtgaatacc cagctcaaga 660
tccaaaacat ttccataatt tcagaaagtt ccaaaccctt gcctcttttc agtcttagcc 720
ctcttcccct gaagtaacca ctgttccgac ttcaatcact acttttatcc cacagggtta 780
ttttttggct tttttccact aaattttcaa attcittgat atggtacttt actattgacg 840
aagtactttc aactaggtt atttaatat ctttgattca cccaatattt agggaacacc 900
tgtaggggac aaaaaatgaa tgagagcccc tgccttccat tgctgctaata ctggtgggaa 960
cgagacatgt atttaattaa gcatgtaaaa aatagagtgg gtgatgaaat aatctatata 1020
ctaaatcccc atgacacaca gtttacctat gtaacaaacc tgcatgtgta cccccgaacc 1080
taaaatataa gttggaaatt aaaaaaaaaac gagagggaga atagagcatic acaaccagag 1140
tgctgagatg aattacttta ttaccaaaga aggaggagga ctcagggagg tgccgacggt 1200
taaaccagtc cactgaaggg tgtgcagaat ttggataggc aagataccct gggacaaggt 1260
cattctaaaa ccatgctaac atttgtactt ttttttcat tgtgatagtt cctgaaatga 1320
gttgcataaa actggtacat gtcttagggc agtctctaata tgatttttat tttgttctat 1380

```

[0003]

ttttaaaaat tagtcttcaa atagcagatt cacatgatat taaaatataat gcacataaat	1440
tatatacaca aatataat tttt ctgaatgaaa tttagtatct gcatataat aagagctatt	1500
tctgtctcat atgttcataa tcttcatcca ttaaaaaaac ttttgttagg ctttctcac	1560
tctaagatta taaaaaatc tcccattatt tacctagcta gtttctagt tgttccaaaa	1620
ccatttattg aacaatccat ctttttgaca ctggtttggc atgcctaat tataatattct	1680
tgtgtgtgtt aggatctcct tttggacttt ccattctgtt cattgagtct taccagctcc	1740
tcttacattg gtacatgat gttttaatct atggggcttt gtagtttaaa tgtagggcta	1800
gttccagcgc attgttctct atcagctgtt aggaacttag aatcagctt gctctgtttt	1860
aaagaaaaac ctggtat tttatcagta taacattcta tttatattaa ctgaagaat	1920
tgaaaacatc tatgat tttt cctattcagt aacgtatcac ttagaatagg ttaggttgta	1980
ctactataaa atctcagctg cataaaaca ttttttttg cttgtgctac acatccatta	2040
ggatcatcaag ggactcacct tgtcaagtta ctacagagatt caggctgata taaaggtttg	2100
atcttgacat acgctttcat gatgacagaa agcagggaa agaaaggtggt gagccatgtg	2160
ctttctcccc cttctatcca gaaatgacac atactcacat ttcattcgcc agagaaatta	2220
acatggcccc tcctaagttc aaatggatag agaaatgcct tctaccagg tgcccagaat	2280
tagaagagca aacatttggtg aacagttctg agtaccacaa ataccgttat ctttccactt	2340
aagtcttctg tttcactcag tagtgcttta aacttttctt catatgtttt tcagtgtttc	2400
ttgttgaatt tcttgatatt ttatcatgtt tgttcgtact gggagtagcc ttttttcca	2460
tttcattttc tggctggttt cattgctggt tgttttttg ttttgtttg tttttgagat	2520
ggagtctcac tctgtcgccc aggctggagt gcagtgtcac aatctcggct cactgcaacc	2580
tctgcctccc aggttcaagc gattcttctt tctcagcctc ctgagtagct gggattacag	2640
gcatgtgcca ccatgcccag ctaat ttttt atat ttttag tagagatggg gtttctccat	2700
gttggtcagg ctggtctcaa actccaatc tcaggtgatc cgctgcctc tgccttccaa	2760
agtgcctggga ttatagacat gagccaccgt gcctggccta gttcttatgg gatgtatatg	2820

[0004]

tctttggatt catatgatat gtatatatgt ttatatttct acaagtacat acctaggagt 2880  
 ggaattgttg ggtcataggt taatgcatgt ttttctgcc aacagttgtg tcaatttctg 2940  
 tttcaccgc tgfgaatgag agttgttcta ctttcttgac aacacttgat attgtcagtc 3000  
 attttagcca ttctggtgaa tttatagtg c ttttctgtg tgtgtaagag agagaatgag 3060  
 agagggtgtt tgtgagaaaa ccaaagcaac actgtgagag tgtgtgtgtt tgtgagaaaa 3120  
 ccaaaataca tactactgtg atttcattgg gagaaaatct gtttggata tcaaaaaaag 3180  
 tagcttaatt acttcatcat tattggttta ggt 3213

<210> 4  
 <211> 1296  
 <212> DNA  
 <213> 智人

<400> 4  
 taagaacatt ttacactctt cagtataaag aagtcagaat acccctaccc tatcagtaaa 60  
 ggccataaag ttaccattaa aaagatgtcc ttaaaaacag cattctcagc tgggcacggg 120  
 ggctcacacc tttgtccag tactttggga agccgaggtg ggtggatcac ctgaggtcag 180  
 gagttcgaga ccagcctggc caacatggcg aaaaccatt ttcttacta aaaatacaaa 240  
 aattagccgg gcatgggtggc ggggtcttgt ggtcccagct actcaagagg ctgaggtggg 300  
 aggatcactg agcccaggag gtggaggctg cattgagcca agattgtgcc actgcactcc 360  
 agcctgggtg acagagcgag actctgtctc aaaaaacca aaacaaaaaa aaccagcat 420  
 tctttagtaa ataattcata gttttcttca tctagaattt aaaattgtga tagttgatca 480  
 gcatgtcctg agcacgtgtg tttgctgtta ctagttttaga tcggtagatg tgtatataag 540  
 ttataggtat aaaatcaatc ctgagttgac acaaggtttt gatgttgagt acaagtacag 600  
 taagtgtata tttttagtta tgctcttagt ttttaagtcaa ttgtgtgggt ctttctagct 660  
 ttaggatctg ttgaattatc ttccttagaa aaggaggtta agaatttca cttacctatc 720  
 ttctacttgt ttggagaata gaagagtccc tgtggtagca gactttgtga gtttacttgt 780  
 aattttccat ctgaaagact gttcttgttt ttcgtgatga agtcttgctc tgtcgcccag 840

[0005]

gctggagtgc agtgggtgcaa ccttggctca ctgcaacctc tgccctcccgg gttcaagcaa	900
ttctcctgcc tcagcctccc gagtatctgg gattacaggt gcacaccacc acacctggct	960
aatTTTTgta ttttcagtag agacgggggtt tcaccatggt ggccaggctg gtctcgaact	1020
cttgacctca tgatcagccc acctcagcct tccaaagtgc tgggattaca ggtgtgagcc	1080
cccacactcg gccgttgttg ttttttaaga gacagggctc cactctgtca cctaacctgg	1140
agtacagtgg caatcatggc tcaactgtaac ctcaaatgcc cggccttagt gaagcgttct	1200
tcctgccttg gcctcccaaa gtgctgggat tacaagtgtg agccatgcat ccagcttgaa	1260
agacagcttc ttaggcttga tttgtttggt tacagg	1296

<210> 5  
 <211> 3213  
 <212> DNA  
 <213> 智人

<400> 5	
atcagagtca ctttgatatg atgagagcag agataaacag atttgttgca tgTTTTtaat	60
ctttggtatg ggacatacta gaattcactg caaatacatt tttatgtaac tgttgaatgc	120
tcatacgacc atggaattct tccctttaaa gagcttggta agcatttgag tgtagttggt	180
agacggagac gatcacgtca tagtttatag agtgcataaa gacgtaagtt accatttaat	240
acctttcatt caggaaaaat gtacttagac cctacaatgt actagtaggc ctctgcgctg	300
gcaatacaga taagataatg tagcccctgg cctcaaagga actctcctcc ttaggttgca	360
tttgtataat gtttgatttt tagattgttc tttgagccct gcattccacg aggataggtc	420
agtgggtatt aacgaggtaa aaggggagta gtacgaaagg gcattcaagc gtcccatctt	480
cgcttcaacc aaagcagccc tgcgttttcc tagttttatt aataggtttg atgtaaggtc	540
gtctttgaaa agggggtttg gctttttttt acagtgtgac tgaggataa tttataaaaa	600
gggaaatgta tggcatggtg agttttttca catacatcct tgtgaatacc cagctcaaga	660
tccaaaacat ttccataatt tcagaaagtt ccaaaccctt gcctcttttc agtcttagcc	720
ctcttcccct gaagtaacca ctgttccgac ttcaatcact acttttatcc cacaggttaa	780

[0006]

ttttttggct tttttccact aaatittcaa attctttgat atggtacttt actattgacg	840
aagtactttc acactagggt atttaatat ctttgattca cccaatattt agggaacacc	900
tgtaggggac aaaaaatgaa tgagagcccc tgccttccat tgctgctaata ctggtgggaa	960
cgagacatgt atttaattaa gcatgtaaaa aatagagtgg gtgatgaaat aatctatata	1020
ctaaatcccc atgacacaca gtttacctat gtaacaaacc tgcatgtgta cccccgaacc	1080
taaaatataa gttggaaatt aaaaaaaaaac gagagggaga atagagcatc acaaccagag	1140
tgctgagatg aattacttta ttaccaaaga aggaggagga ctgaggagg tgccgacgtt	1200
taaaccaggt cactgaagggt tgtgcagaat ttggataggc aagataccct gggacaaggt	1260
cattctaaaa ccatgctaac atttgtactt ttttttcat tggatagtt cctgaaatga	1320
gttgcataaa actggtacat gtcttagggc agtctctaata tgatttttat tttgttctat	1380
ttttaaaaat tagtcttcaa atagcagatt cacatgatata taaaatataat gcacataaat	1440
tatatacaca aatataatttt ctgaatgaaa tttagtatct gcatatattt aagagctatt	1500
tctgtctcat atgttcataa tcttcatcca ttaaaaaaac ttttgtagg ctttctcac	1560
tctaagatta taaaaaatc tcccattatt tacctagcta gtttctagt tgttccaaaa	1620
ccatttattg aacaatccat ctttttgaca ctggtttggc atgccttaat tatatattct	1680
tgtgtgtgtt aggatctcct tttggacttt ccattctgtt cattgagtct tatcagctcc	1740
tcttacattg gtaccatgat gttttaatct atggggcttt gtagtttaaa tgtagggcta	1800
gttccagcgc attgttctct atcagctgtt aggaacttag aatcagctt gctctgtttt	1860
aaagaaaaac ctggtatttt tttatcagta taacattcta tttatattaa cttgaagaat	1920
tgaaaacatc tatgattttt cctattcagt aacgtatcac ttagaatagg ttaggttgta	1980
ctactataaa atctcagctg cataaaacaa ttttttttg cttgtgctac acatccatta	2040
ggtcatcaag ggactcacct tgcaagta ctgagagatt caggctgata taaaggtttg	2100
atcttgacat acgctttcat gatgacagaa agcaggggaag agaaggtggt gagccatgtg	2160
ctttctcccc cttctatcca gaaatgacac atactcacat ttcatcgc agagaaatta	2220
acatggcccc tcctaagttc aatggatag agaaatgcct tcctaccagg tgcccagaat	2280

[0007]

tagaagagca aacatttgtg aacagttctg agtaccacaa ataccgttat cttccactt	2340
aagtcttctg tttcactcag tagtgcttta aacttttctt catatgtttt tcagtgtttc	2400
ttgttgaatt tcttgatatt ttatcatggt tgttcgtact gggagtagcc ttttttcca	2460
tttcattttc tggctggttt cattgctggt tgttttttg ttttgtttg tttttgagat	2520
ggagtctcac tctgtcgcc aggctggagt gcagtgtcac aatctcggct cactgcaacc	2580
tctgcctccc aggttcaagc gattcttctt tctcagcctc ctgagtagct gggattacag	2640
gcatgtgccca ccatgccag ctaattttt atatttttag tagagatggg gtttctccat	2700
gttggtcagg ctggctcaa actcccaatc tcaggtgatc cgcctgcctc tgccttccaa	2760
agtgtggga ttatagacat gagccaccgt gcctggccta gttcttatgg gatgtatatg	2820
tctttgatt cataigatat gtatatatgt ttatatttct acaagtacat acctaggagt	2880
ggaattgttg ggtcataggt taatgcatgt ttttctgcc aacagttgtg tcaatttctg	2940
ttttaccgc tgtgaatgag agttgttcta ctttcttgac aacacttgat attgtcagtc	3000
attttagcca ttctggtgaa tttatagtgc tatttctgtg tgtgtaagag agagaatgag	3060
agagggtgtt tgtgagaaaa ccaaagcaac actgtgagag tgtgtgtgtt tgtgagaaaa	3120
ccaaaataca tactactgtg atttcattgg gagaaaatct gtttgggtata tcaaaaaaag	3180
tagcttaatt acttcatcat tattggttta ggt	3213

<210> 6  
 <211> 1296  
 <212> DNA  
 <213> 智人

<400> 6	
taagaacatt ttacactctt cagtataaag aagtcagaat acccctacc tatcagtaaa	60
ggcctataag ttaccattaa aaagatgtcc ttaaaaacag cattctcagc tgggcgcggt	120
ggctcacacc ttigtcccag tactttggga agccgaggtg ggtggatcac ctgaggtcag	180
gagttcgaga ccagcctggc caacatggcg aaaaccatt ttcttacta aaaatacaaa	240
aattagccgg gcatgggtggc ggggtcctgt ggtcccagct actcaagagg ctgaggtggg	300

[0008]

aggatcactg agcccaggag gtggaggctg cattgagcca agattgtgcc actgcactcc	360
agcctgggtg acagagcgag actctgtctc aaaaaacca aaacaaaaa aaccagcat	420
tcttagtaa ataattcata gtttcttca tctagaattt aaaattgtga tagttgatca	480
gcatgtcctg agcacgtgtg tttgctgta ctagtttaga tcggtagatg tgtatataag	540
ttataggtat aaaatcaatc ctgagttgac acaaggtttt gatgttgagt acaagtacag	600
taagtgtata ttttagtta tgctcttagt ttttaagtcaa ttgtgtggtt ctttctagct	660
ttaggatctg ttgaattatc ttccttagaa aaggagtta agaattctca cttacctatc	720
ttctacttgt ttggagaata gaagagtccc tgtggtagca gactttgtga gtttacttgt	780
aattttccat ctgaaagact gttcttgttt ttcgtgatga agtcttgctc tgtcgccag	840
gctggagtgc agtggtgcaa ccttggctca ctgcaacctc tgccctccgg gttcaagcaa	900
ttctctgccc tcagcctccc gagtatctgg gattacaggt gcacaccacc acacctggct	960
aatTTTTgta ttttcagtag agacggggtt tcaccatggt ggccaggctg gtctcgaact	1020
cttgacctca tgatcagccc acctcagcct tccaaagtgc tgggattaca ggtgtgagcc	1080
cccacactcg gccgttggtg ttttttaaga gacagggtct cactctgtca cctaacctgg	1140
agtacagtgg caatcatggc tcaactgtaac ctcaaagcc cggccttagt gaagcgttct	1200
tcctgccttg gcctcccaa gtgctgggat tacaagtgtg agccatgcat ccagcttgaa	1260
agacagcttc ttaggcttga tttgtttggt tacagg	1296

<210> 7

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 7

taggtcagtg ggtattaacg aggt

24

<210> 8

[0009]

<211> 23	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 8	
gtcactttga tatgatgaga gca	23
<210> 9	
<211> 23	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 9	
cctcgttaat acccactgac cta	23
<210> 10	
<211> 24	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 10	
ggcaatacag ataagataat gtag	24
<210> 11	
<211> 21	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 11	
ttgatcagca tgtcctgagc a	21
<210> 12	
<211> 22	

[0010]

---

<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 12	
ctaccctatc agtaaaggcc ta	22
<210> 13	
<211> 21	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 13	
tgctcaggac atgctgatca a	21
<210> 14	
<211> 23	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 14	
aaaaacagca ttctcagctg ggc	23

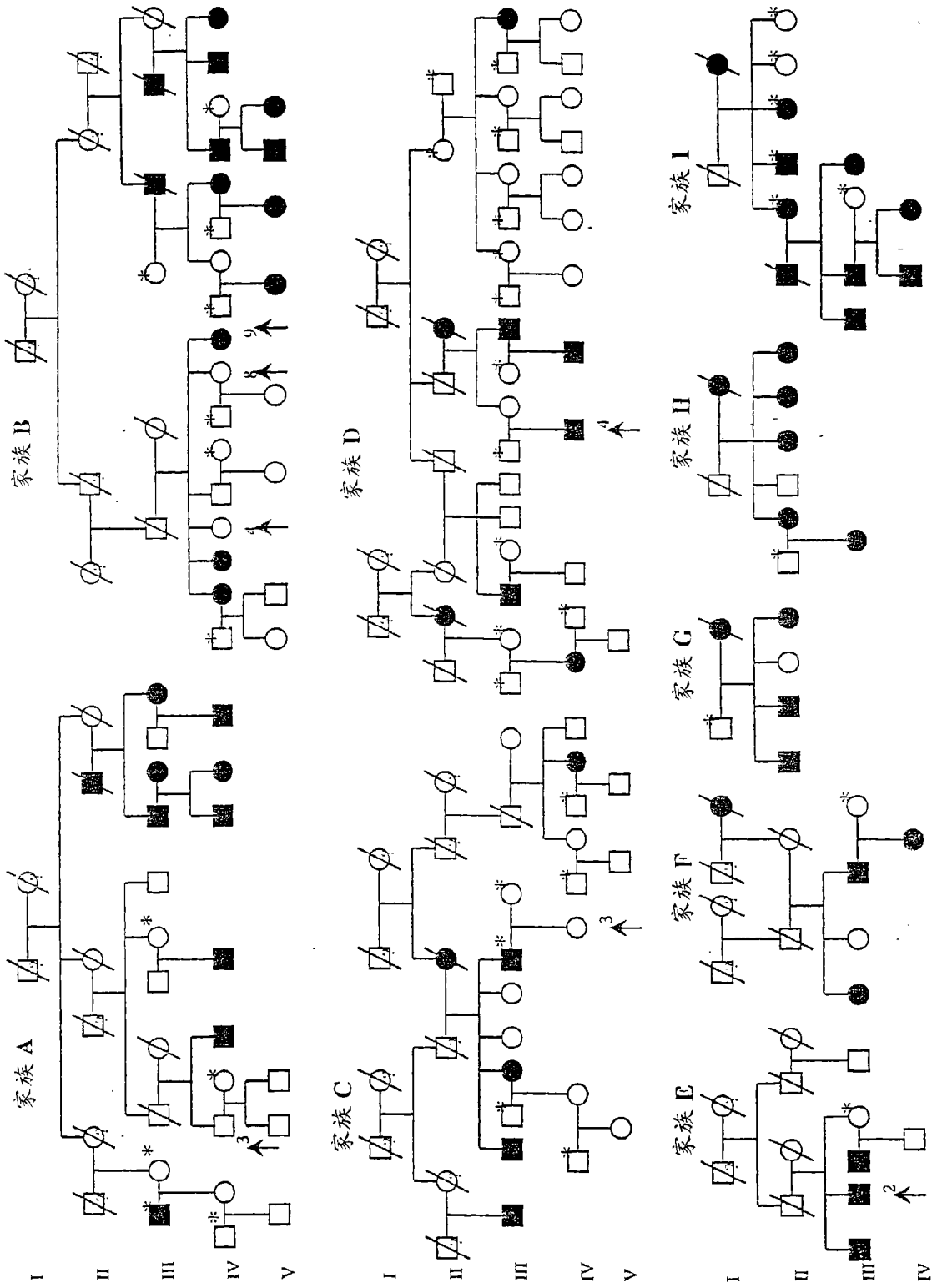


图 1

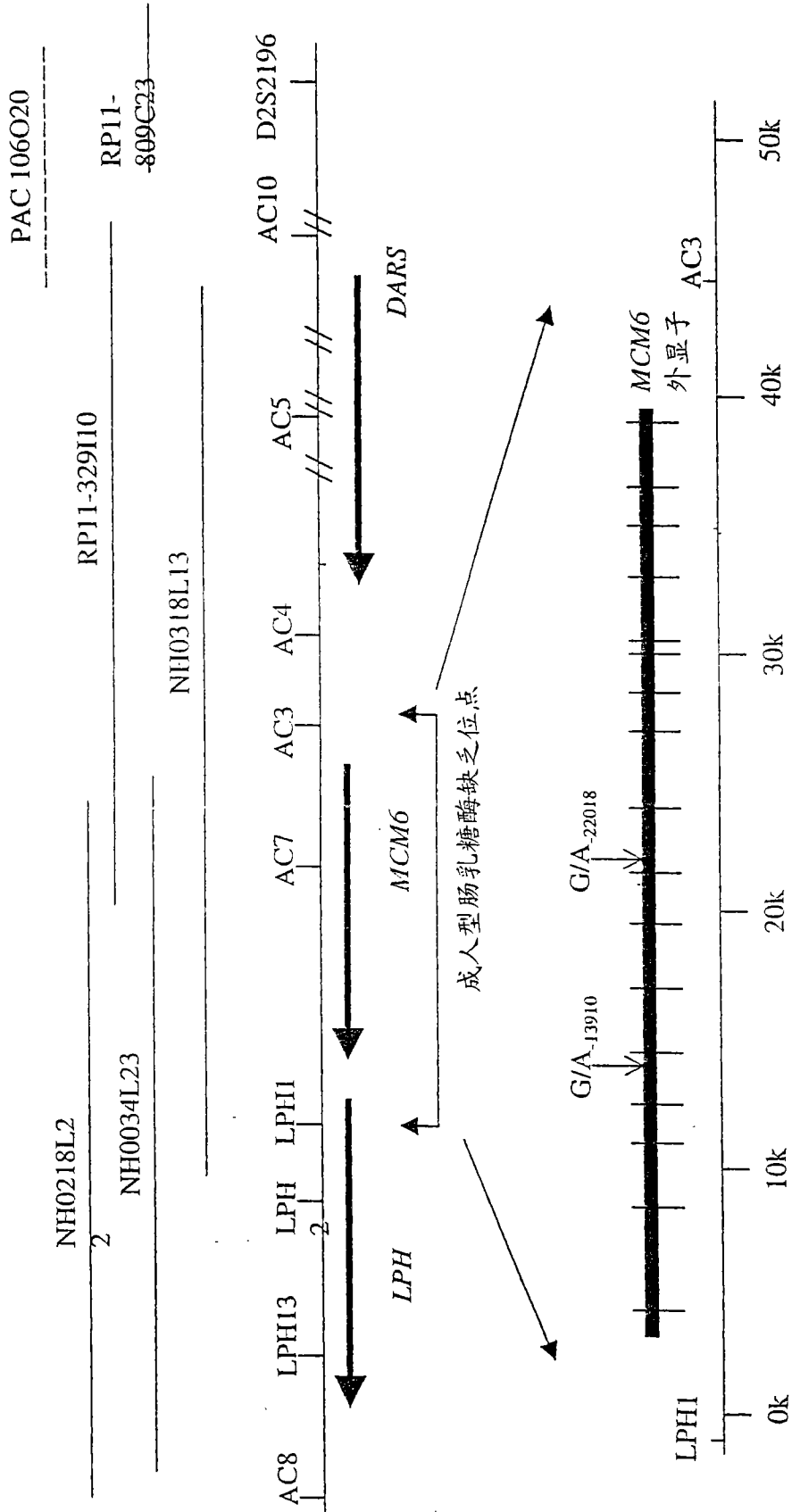


图 2

标记	单倍型										总数	
LPH13	6	6	6	6	6	6	6	6	6	6	6	6
LPH2	4	4	4	4	4	4	4	4	4	4	4	4
LPH1	3	2	2	2	2	1	1	4	4	4	4	4
AC7	5	5	5	5	5	5	5	5	5	5	5	5
AC3	2	2	2	2	4	2	2	2	2	2	2	2
AC4	5	5	5	5	5	5	5	5	5	5	5	5
AC5	6	6	5	5	6	6	6	6	6	6	6	3
乳糖酶持续性等位基因	20	4	1	2	2	2	1	1	1	1	1	33
乳糖酶非持续性等位基因	3	2	0	0	0	0	0	2	0	0	2	54

图 3

**SEQ ID NO:1**

ACCTTTCATTCAGGAAAAATGTA  
CTTAGACCCTACAATGTA  
CTAGTAGGCCTCTGCGCTG  
GCAATACAGATAAGATAATGTAG  
CCCCTGGCCTCAAAGGAACTCT  
CCTCCTTAGGTTGCA  
TTTGTATAATGTTTGATTTT  
TAGATTGTTCTTTGAGCCCTGC  
ATTCCACGAGGATAGGTC

图 4

**SEQ ID NO:2**

TAAGAACATTTTACACTCTTCAG  
TATAAAGAAGTCAGAATACCCCT  
ACCCTATCAGTAAA  
GGCCTATAAGTTACCATTAAAA  
AGATGTCCTTAAAAACAGCATTC  
TCAGCTGGGCgCGGT  
GGCTCACACCTTTGTCCCAGTACT  
TTGGGAAGCCGAGGTGGGTGGAT  
CACCTGAGGTCAG

图 5

SEQ ID NO:3

ATCAGAGTCACTTTGATATGATGAGAGCAGAGATAAACAGATTTGTTGCATGTTTTTAAT  
CTTTGGTATGGGACATACTAGAATTCCTGCAAAATACATTTTTATGTAACGTGTTGAATGC  
TCATACGACCATGGAATCCTCCCTTTAAAGAGCTTGGTAAGCATTGAGTGTAGTTGTT  
AGACGGAGACGATCACGTCATAGTTTATAGAGTGCATAAAGACGTAAGTTACCATTTAAT  
ACCTTTCATTCAGGAAAAATGTACTTAGACCCTACAATGTACTAGTAGGCCTCTGCGCTG  
GCAATACAGATAAGATAATGTAGtCCCTGGCCTCAAAGGAACTCTCCTCCTTAGGTTGCA  
TTTGTATAATGTTTGATTTTTAGATTGTTCTTTGAGCCCTGCATTCCACGAGGATAGGTC  
AGTGGGTATTAACGAGGTAAGGAGGGGAGTAGTACGAAAGGGCATTCAAGCGTCCCATCTT  
CGCTTCAACCAAAGCAGCCCTGCGTTTTCTAGTTTTTATTAATAGGTTTGATGTAAGGTC  
GTCTTTGAAAAGGGGGTTTTGGCTTTTTTTTACAGTGTGACTGAGGTATAATTTATAAAAA  
GGGAAATGTATGGCATGGTGAGTTTTTTTCACATACATCCTTGTGAATACCCAGCTCAAGA  
TCCAAAACATTTCCATAATTTAGAAAGTTCCAAACCCCTGCCTCTTTTCAGTCTTAGCC  
CTCTTCCCCTGAAGTAACCACTGTTCCGACTTCAATCACTACTTTTATCCCACAGGTTAA  
TTTTTTGGCTTTTTTCCACTAAATTTTCAAATTCCTTTGATATGGTACTTTACTATTGACG  
AAGTACTTTCACACTAGGTTATTTAATATTCCTTTGATTCACCCAATATTTAGGGAACACC  
TGTAGGGGACAAAAAATGAATGAGAGCCCTGCCTTCCATTGCTGCTAATCTGGTGGGAA  
CGAGACATGTATTTAATTAAGCATGTAAAAAATAGAGTGGGTGATGAAATAATCTATATA  
CTAAATCCCCATGACACACAGTTTACCTATGTAACAAACCTGCATGTGTACCCCGAACC  
TAAATATAAGTTGGAAATTAAAAAAAAACGAGAGGGAGAATAGAGCATCACAAACCAGAG  
TGCTGAGATGAATTACTTTATTACCAAAGAAGGAGGAGGACTCAGGGAGGTGCCGACGTT  
TAAACCCAGTCACTGAAGGGTGTGCAGAATTTGGATAGGCAAGATACCCTGGGACAAGGT  
CATTCATAAAACCATGCTAACATTTGTACTTTTTTTTTTCATTGTGATAGTTCCTGAAATGA  
GTTGCATAAAACTGGTACATGTCTTAGGGCAGTCTCTAATTGATTTTTATTTTGTCTAT  
TTTTAAAAATTAGTCTTCAAATAGCAGATTCACATGATATTAATATATATGCACATAAAT  
TATATACACAAATATATTTTCTGAATGAAATTTAGTATCTGCATATATTTAAGAGCTATT  
TCTGTCTCATATGTTTATAATCTTCATCCATTAAAAAAATTTTTGTTAGGCCTTTCTCAC  
TCTAAGATTATAAAAAATCTCCCATTTATTTACCTAGCTAGTTTTCTAGTTGTTCCAAAA  
CCATTTATTGAACAATCCATCTTTTTGACACTGGTTTGGCATGCCTTAATTATATATTCT  
TGTGTGTGTTAGGATCTCCTTTTGGACTTTCCATTCTGTTTCATTGAGTCTTATCAGCTCC  
TCTTACATTGGTACCATGATGTTTTAATCTATGGGGCTTTGTAGTTTAAATGTAGGGCTA  
GTTCCAGCGCATTGTTCTCTATCAGCTGTTAGGAACTTAGAAATCAGCTTGCTCTGTTTT  
AAAGAAAAACCTGGTATTTTTTTTATCAGTATAACATTCATTTATATTAACCTGAAGAAT  
TGAAAACATCTATGATTTTTTCTTATTCAGTAACGTATCACTTAGAATAGGTTAGGTTGTA  
CTACTATAAAATCTCAGCTGCATAAAACAATTTTTTTTTGCTTGTGCTACACATCCATTA  
GGTCATCAAGGGACTCACCTTGTCAAGTTACTCAGAGATTCAGGCTGATATAAAGGTTTG  
ATCTTGACATACGCTTTCATGATGACAGAAAGCAGGGAAGAGAAGGTGGTGAGCCATGTG  
CTTCTCCCCCTTCTATCCAGAAATGACACATACTCACATTTTCATTGCGCCAGAGAAATTA  
ACATGGCCCCTCCTAAGTTCAAATGGATAGAGAAATGCCTTCCTACCAGGTGCCCAGAAT  
TAGAAGAGCAAACATTTGTGAACAGTCTGAGTACCACAAAATACCGTTATCTTTCCACTT  
AAGTCTTCTGTTTCACTCAGTAGTCTTTAAACTTTTCTTCATATGTTTTTTCAGTGTTC  
TTGTTGAATTTCTTGATATTTTATCATGTTTGTTCGTAAGTGGGAGTAGCCTTTTTTTTCCA

图 6

TTTCATTTTCTGGCTGGTTCATTGCTGGTTGTTTTTTGTTTTGTTTTGTTTTGAGAT  
GGAGTCTCACTCTGTGCGCCAGGCTGGAGTGCAGTGCACAATCTCGGCTCACTGCAACC  
TCTGCCTCCCAGGTTCAAGCGATTCTTCTTTCTCAGCCTCCTGAGTAGCTGGGATTACAG  
GCATGTGCCACCATGCCAGCTAATTTTTTATATTTTAGTAGAGATGGGGTTTCTCCAT  
GTTGGTCAGGCTGGTCTCAAACCTCCCAATCTCAGGTGATCCGCCTGCCTCTCCCTTCCAA  
AGTGCTGGGATTATAGACATGAGCCACCGTGCCTGGCCTAGTTCTTATGGGATGTATATG  
TCTTTGGATTCATATGATATGTATATATGTTTATATTTCTACAAGTACATACCTAGGAGT  
GGAATTGTTGGGTCATAGGTTAATGCATGTTTTTCTGCCAAACAGTTGTGTCAATTTCTG  
TTTTACCGCTGTGAATGAGAGTTGTTCTACCTTCTTGACAACACTTGATATTGTCAGTC  
ATTTTAGCCATTCTGGTGAATTTATAGTGCTATTTCTGTGTGTGTAAGAGAGAGAATGAG  
AGAGGGTGTGTTGTGAGAAAACCAAAGCAACACTGTGAGAGTGTGTGTGTTTGTGAGAAA  
CCAAAATACATACTACTGTGATTTTATTGGGAGAAAATCTGTTTGGTATATCAAAAAAAG  
TAGCTTAATTACTTCATCAATTATTGGTTTAGGT

图 6 续

## SEQ ID NO:4

TAAGAACATTTTACACTCTTCAGTATAAAGAAGTCAGAATACCCCTACCCTATCAGTAAA  
GGCCTATAAGTTACCATTAAAAAGATGTCCTTAAAAACAGCATTCTCAGCTGGGCaCGGT  
GGCTCACACCTTTGTCCCAGTACTTTGGGAAGCCGAGGTGGGTGGATCACCTGAGGTCAG  
GAGTTCGAGACCAGCCTGGCCAACATGGCGAAAACCCATTTTCTCTACTAAAAATACAAA  
AATTAGCCGGGCATGGTGGCGGGTGCCTGTGGTCCCAGCTACTCAAGAGGCTGAGGTGGG  
AGGATCACTGAGCCCAGGAGGTGGAGGCTGCATTGAGCCAAGATTGTGCCACTGCACTCC  
AGCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAACCAAAACAAAAAAAACCCAGCAT  
TCTTTAGTAAATAATTCATAGTTTCTTCATCTAGAATTTAAAATTGTGATAGTTGATCA  
GCATGTCCTGAGCACGTGTGTTTGCTGTACTAGTTTAGATCGGTAGATGTGTATATAAG  
TTATAGGTATAAAATCAATCCTGAGTTGACACAAGGTTTTGATGTTGAGTACAAGTACAG  
TAAGTGTATATTTTGTAGTTATGCTCTTAGTTTTAAGTCAATTGTGTGGTTCTTTCTAGCT  
TTAGGATCTGTTGAATTATCTTCCTTAGAAAAGGGAGTTAAGAATCTTCACTTACCTATC  
TTCTACTTGTGGAGAATAGAAGAGTCCCTGTGGTAGCAGACTTTGTGAGTTTACTTGT  
AATTTTCCATCTGAAAGACTGTTCTTGTTTTTTCGTGATGAAGTCTTGCTCTGTGCCCCAG  
GCTGGAGTGCAGTGGTGCACCTTGGCTCACTGCAACCTCTGCCTFCCCGGGTTC AAGCAA  
TTCTCCTGCCTCAGCCTCCCGAGTATCTGGGATTACAGGTGCACACCACCACACCTGGCT  
AATTTTTGTATTTTCAGTAGAGACGGGGTTTACCATGTTGGCCAGGCTGGTCTCGAACT  
CTTGACCTCATGATCAGCCCACCTCAGCCTTCCAAAGTGCTGGGATTACAGGTGTGAGCC  
CCCACACTCGGCCGTTGTTGTTTTTTAAGAGACAGGGTCTCACTCTGTCACCTAACCTGG  
AGTACAGTGGCAATCATGGCTCACTGTAACCTCAAATGCCCGGCCTTAGTGAAGCGTTCT  
TCCTGCCTTGGCCTCCCAAAGTGCTGGGATTACAAGTGTGAGCCATGCATCCAGCTTGAA  
AGACAGCTTCTTAGGCTTGATTTGTTGGTTACAGG

图 7

SEQ ID NO:5

ATCAGAGTCACTTTGATATGATGAGAGCAGAGATAAACAGATTTGTTGCATGTTTTTAAT  
 CTTTGGTATGGGACATACTAGAATTCAGTCAAATACATTTTTATGTAAGTGTGAATGC  
 TCATACGACCATGGAATTCCTCCCTTTAAAGAGCTTGGTAAGCATTGAGTGTAGTTGTT  
 AGACGGAGACGATCACGTCATAGTTTATAGAGTGCATAAAGACGTAAGTTACCATTTAAT  
 ACCTTTTCATTGAGGAAAATGTACTTAGACCCTACAATGTACTAGTAGGCCTCTGCGCTG  
 GCAATACAGATAAGATAATGTAGCCCTGGCCTCAAAGGAACTCTCCTCCTTAGGTTGCA  
 TTTGTATAATGTTTGATTTTTTAGATTGTTCTTTGAGCCCTGCATTCCACGAGGATAGGTC  
 AGTGGGTATTAACGAGGTAAAAGGGGAGTAGTACGAAAGGGCATTCAAGCGTCCCATCTT  
 CGCTTCAACCAAAGCAGCCCTGCGTTTTCTAGTTTTATTAATAGGTTTGATGTAAGGTC  
 GTCTTTGAAAAGGGGGTTTGGCTTTTTTTTACAGTGTGACTGAGGTATAATTTATAAAAA  
 GGGAAATGTATGGCATGGTGAGTTTTTTTACATACATCCTTGTGAATACCCAGCTCAAGA  
 TCCAAAACATTTCCATAATTCAGAAAGTTCCAAACCCCTGCCTCTTTTCAGTCTTAGCC  
 CTCTTCCCCTGAAGTAACCACTGTTCCGACTTCAATCACTACTTTTATCCCACAGGTAA  
 TTTTTTGGCTTTTTTCCACTAAATTTTCAAATTCCTTGATATGGTACTTTACTATTGACG  
 AAGTACTTTCACACTAGGTTATTTAATATTCTTTGATTCACCCAATAATTTAGGGAACACC  
 TGTAGGGGACAAAAAATGAATGAGAGCCCTGCCTTCCATTGCTCCTAATCTGGTGGGAA  
 CGAGACATGTATTTAATTAAGCATGTAAAAAATAGAGTGGGTGATGAAAATAATCTATATA  
 CTAAATCCCCATGACACACAGTTTACCTATGTAACAAACCTGCATGTGTACCCCCGAACC  
 TAAAATATAAGTTGGAAATTAAAAAAAGAGAGGGAGAATAGAGCATCACAACCAGAG  
 TGCTGAGATGAATTACTTTATTACCAAAGAAGGAGGAGGACTCAGGGAGGTGCCGACGTT  
 TAAACCCAGTCACTGAAGGGTGTGCAGAATTTGGATAGGCAAGATACCCTGGGACAAGGT  
 CATTCTAAAACCATGCTAACATTTGTACTTTTTTTTTTCAATCTGTGATAGTTCTCTGAAATGA  
 GTTGATAAAACTGGTACATGTCTTAGGGCAGTCTCTAATTCATTTTTATTTGTTCTAT  
 TTTTAAAAATTAGTCTTCAAATAGCAGATTCACATGATATTAATAATATATGCACATAAAT  
 TATATACACAAATATATTTTCTGAATGAAATTTAGTATCTGCATATATTTAAGAGCTATT  
 TCTGTCTCATATGTTTATAATCTTCATCCATTAAAAAAATTTTGTAGGCCTTTCTCAC  
 TCTAAGATTATAAAAAATCTCCCATTATTTACCTAGCTAGTTTTCTAGTTGTTCCAAA  
 CCATTTATTGAACAATCCATCTTTTTGACACTGGTTTGGCATGCCTTAATTATATATCT  
 TGTGTGTGTTAGGATCTCCTTTTGGACTTTCCATTCTGTTTCAATTGAGTCTTATCAGCTCC  
 TCTTACATTGGTACCATGATGTTTTAATCTATGGGGCTTTGTAGTTTAAATGTAGGGCTA  
 GTTCCAGCGCATTGTTCTCTATCAGCTGTTAGGAACCTAGAAATCAGCTTGCTCTGTTTT  
 AAAGAAAACCTGGTATTTTTTTTATCAGTATAACATTCATTTATATTAACCTGAAGAAT  
 TGAAAACATCTATGATTTTTCTTATTCAGTAACGTATCACTTAGAATAGGTTAGGTTGTA  
 CTACTATAAAATCTCAGCTGCATAAAACAATTTTTTTTTTGCTTGTGCTACACATCCATTA  
 GGTCAATCAAGGACTCACCTTGTCAAGTACTCAGAGATTCAGGCTGATATAAAGGTTTG  
 ATCTTGACATACGCTTTCATGATGACAGAAAGCAGGGAAGAGAAGGTGGTGAGCCATGTG  
 CTTTCTCCCCCTTCTATCCAGAAATGACACATACTCACATTTCAATTCGCCAGAGAAATTA  
 ACATGGCCCCCTCCTAAGTTCAAATGGATAGAGAAATGCCTTCCCTACCAGGTGCCCAGAAT  
 TAGAAGAGCAAACATTTGTGAACAGTTCTGAGTACCACAAATACCGTTATCTTTCCACTT  
 AAGTCTTCTGTTTCACTCAGTAGTGCTTTAACTTTTCTTCATATGTTTTTTCAGTGTTC  
 TTGTTGAATTTCTTGATATTTTATCATGTTTGTTCGTACTGGGAGTAGCCTTTTTTTCCA

图 8

TTTCATTTTCTGGCTGGTTTCATTGCTGGTTGTTTTTTGTTTTGTTTTGTTTTGAGAT  
 GGAGTCTCACTCTGTTCGCCCAGGCTGGAGTGCAGTGTACAATCTCGGCTCACTGCAACC  
 TCTGCCTCCCAGGTTCAAGCGATTCTTCTTTCTCAGCCTCCTGAGTAGCTGGGATTACAG  
 GCATGTGCCACCATGCCAGCTAATTTTTTATATTTTTTAGTAGAGATGGGGTTTCTCCAT  
 GTTGGTCAGGCTGGTCTCAAACCTCCAATCTCAGGTGATCCGCCTGCCTCTGCCTTCCAA  
 AGTGCTGGGATTATAGACATGAGCCACCGTGCCTGGCCTAGTTCTTATGGGATGTATATG  
 TCTTTGGATTCATATGATATGTATATATGTTTATATTTCTACAAGTACATAACCTAGGAGT  
 GGAATTGTTGGGTCATAGGTTAATGCATGTTTTTCTGCCAAACAGTTGTGTCAATTTCTG  
 TTTTCACCGCTGTGAATGAGAGTTGTTCTACCTTCTTGACAACACTTGATATTGTTCAGTC  
 ATTTTAGCCATTCTGGTGAATTTATAGTGCTATTTCTGTGTGTGTAAGAGAGAGAATGAG  
 AGAGGGTGTGTTGTGAGAAAACCAAAGCAACACTGTGAGAGTGTGTGTGTTTGTGAGAAAA  
 CCAAATACATACTACTGTGATTTTCATTGGGAGAAAATCTGTTTGGTATATCAAAAAAAG  
 TAGCTTAATTACTTCATCATTATTGGTTTAGGT

图 8 续

**SEQ ID NO:6**

TAAGAACATTTTACACTCTTCAGTATAAAGAAGTCAGAATACCCCTACCCTATCAGTAAA  
 GGCCTATAAGTTACCATTAAAAAGATGTCCTTAAAAACAGCATTCTCAGCTGGGCgCGGT  
 GGCTCACACCTTTGTCCCAGTACTTTGGGAAGCCGAGGTGGGTGGATCACCTGAGGTGAG  
 GAGTTCGAGACCAGCCTGGCCAACATGGCGAAAACCCATTTTCTCTACTAAAAATACAAA  
 AATTAGCCGGGCATGGTGGCGGGTGCCTTGTGGTCCCAGCTACTCAAGAGGCTGAGGTGGG  
 AGGATCACTGAGCCCAGGAGGTGGAGGCTGCATTGAGCCAAGATTGTGCCACTGCACTCC  
 AGCCTGGGTGACAGAGCGGAGACTCTGTCTCAAAAAAACCAAACAAAAAAAACCCAGCAT  
 TCTTTAGTAAATAATTCATAGTTTTCTTCATCTAGAATTTAAAATTGTGATAGTTGATCA  
 GCATGCTCTGAGCACGTGTGTTTGCTGTACTAGTTTAGATCGGTAGATGTGTATATAAG  
 TTATAGGTATAAAATCAATCCTGAGTTGACACAAGGTTTTGATGTTGAGTACAAGTACAG  
 TAAGTGTATATTTTTAGTTATGCTCTTAGTTTTAAGTCAATTGTGTGGTTCTTTCTAGCT  
 TTAGGATCTGTTGAATTATCTTCCTTAGAAAAGGGAGTTAAGAATCTTCACTTACCTATC  
 TTCTACTTGTTTGGAGAATAGAAGAGTCCCTGTGGTAGCAGACTTTGTGAGTTTACTTGT  
 AATTTTCCATCTGAAAGACTGTTCTTGTTTTTCGTGATGAAGTCTTGCTCTGTGCGCCAG  
 GCTGGAGTGCAGTGGTGCAACCTTGGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGCAA  
 TTCTCCTGCCTCAGCCTCCCGAGTATCTGGGATTACAGGTGCACACCACCACACCTGGCT  
 AATTTTTGTATTTTCAGTAGAGACGGGGTTTCACCATGTTGGCCAGGCTGGTCTCGAACT  
 CTTGACCTCATGATCAGCCCACCTCAGCCTTCCAAAGTGCTGGGATTACAGGTGTGAGCC  
 CCCACACTCGGCCGTTGTTGTTTTTTAAGAGACAGGGTCTCACTCTGTCACCTAACCTGG  
 AGTACAGTGGCAATCATGGCTCACTGTAACCTCAAATGCCCGGCCTTAGTGAAGCGTTCT  
 TCCTGCCTTGGCCTCCCAAAGTGCTGGGATTACAAGTGTGAGCCATGCATCCAGCTTGAA  
 AGACAGCTTCTTAGGCTTGATTTGTTTGGTTACAGG

图 9