



(12) 发明专利申请

(10) 申请公布号 CN 102224256 A

(43) 申请公布日 2011. 10. 19

(21) 申请号 200980142889. 4

(51) Int. Cl.

(22) 申请日 2009. 08. 27

C12Q 1/68(2006. 01)

(30) 优先权数据

G01N 33/68(2006. 01)

61/092270 2008. 08. 27 US

G01N 33/53(2006. 01)

G06F 19/24(2011. 01)

(85) PCT申请进入国家阶段日

2011. 04. 27

(86) PCT申请的申请数据

PCT/US2009/055144 2009. 08. 27

(87) PCT申请的公布数据

W02010/025216 EN 2010. 03. 04

(71) 申请人 H. 隆德贝克有限公司

地址 丹麦哥本哈根

(72) 发明人 I·安东尼耶维克 J·塔姆

R·阿泰米什恩 C·P·G·格拉尔

J·B·维斯蒂森

(74) 专利代理机构 中国专利代理(香港)有限公

司 72001

代理人 李波 郭文洁

权利要求书 2 页 说明书 69 页

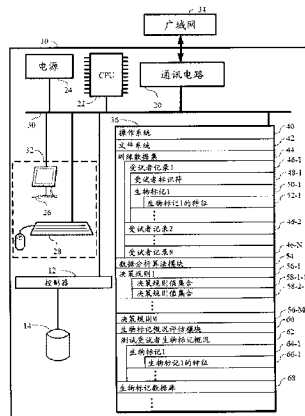
序列表 83 页 附图 10 页

(54) 发明名称

用于测量生物标记概况的系统和方法

(57) 摘要

本发明涉及用于诊断具有情感障碍的患者的方法和系统。该方法也用于预测受试者中对于情感障碍的易感性。



1. 一种诊断测试受试者中的情感障碍的方法,所述方法包括:

评估所述测试受试者的生物标记概况中的多个生物标记的多个特征是否满足值集合,其中满足所述值集合预测所述测试受试者具有所述情感障碍,并且其中所述多个特征是所述多个生物标记的可测量方面,所述多个生物标记包含表 1A 中列出的至少 2 个生物标记。

2. 权利要求 1 的方法,所述方法进一步包括将所述测试受试者是否具有所述情感障碍的诊断输出至用户界面装置、监测器、可触摸计算机可读存储介质、或局部或远程计算机系统;或以用户可读形式显示所述测试受试者是否具有所述情感障碍的诊断。

3. 权利要求 1 的方法,其中所述多个生物标记由表 1A 中列出的 2-29 个生物标记组成。

4. 权利要求 1 的方法,其中所述多个生物标记由表 1A 中列出的 3-20 个生物标记组成。

5. 权利要求 1 的方法,其中所述多个生物标记包含表 1A 中列出的至少 2 个生物标记。

6. 权利要求 1 的方法,其中所述多个生物标记包含表 1A 中列出的至少 3 个生物标记。

7. 权利要求 1 的方法,其中所述多个生物标记包含表 1A 中列出的至少 4 个生物标记。

8. 权利要求 1 的方法,其中所述多个特征由对应于表 1A 中列出的 2-29 个生物标记的 2-29 个特征组成。

9. 权利要求 1 的方法,其中所述多个特征由对应于表 1A 中列出的 3-15 个生物标记的 3-15 个特征组成。

10. 权利要求 1 的方法,其中所述多个生物标记包含对应于表 1A 中列出的至少 2 个生物标记的至少 2 个特征。

11. 权利要求 1 的方法,其中所述多个生物标记包含 ERK1 和 MAPK14。

12. 权利要求 1 的方法,其中所述多个生物标记包含 Gi2 和 IL-1b。

13. 权利要求 1 的方法,其中所述多个生物标记包含 ARRB1 和 MAPK14。

14. 权利要求 1 的方法,其中所述多个生物标记包含 ERK1 和 IL1b。

15. 权利要求 1 的方法,其中所述多个生物标记包含 ARRB1、IL6 和 CD8a。

16. 权利要求 1 的方法,其中所述多个生物标记包含 ARRB1、ODC1 和 P2X7。

17. 权利要求 1 的方法,其中所述多个生物标记中的每个生物标记是核酸。

18. 权利要求 1 的方法,其中所述多个生物标记中的每个生物标记是 DNA、cDNA、扩增的 DNA、RNA 或 mRNA。

19. 权利要求 1 的方法,其中所述测试受试者的生物标记概况中的所述多个生物特征中的特征是所述多个生物标记中的生物标记的可测量方面,并且关于所述特征的特征值使用得自所述测试受试者的生物样品进行测定。

20. 权利要求 19 的方法,其中所述特征是所述生物标记在生物样品中的丰度,并且所述生物样品是全血。

21. 权利要求 1 的方法,所述方法进一步包括在评估步骤前构建所述第一个值集合。

22. 权利要求 21 的方法,其中所述构建步骤包含将数据分析算法应用于得自群体成员的特征。

23. 权利要求 22 的方法,其中所述群体包含来自不具有情感障碍的第一批多个对照受试者的第一批多个生物样品,和来自具有情感障碍的第二批多个受试者的第二批多个生物样品。

24. 权利要求 22 的方法,其中所述数据分析算法是决策树、微阵列预测分析、多重累计

回归树、神经网络、聚类算法、主成分分析、最近邻分析、线性判别分析、二次判别分析、支持向量机、进化法、相关向量机、遗传算法、投影追踪或加权表决。

25. 权利要求 21 的方法,其中所述构建步骤生成决策规则,并且其中所述评估步骤包含将所述决策规则应用于所述多个特征,以便测定它们是否满足所述第一个值集合。

26. 权利要求 25 的方法,其中所述决策规则将所述群体中的受试者分类为 (i) 不具有情感障碍的受试者和 (ii) 的确具有情感障碍的受试者,其准确度 70%或更大。

27. 权利要求 25 的方法,其中所述决策规则将所述群体中的受试者分类为 (i) 不具有情感障碍的受试者和 (ii) 的确具有情感障碍的受试者,其准确度 90%或更大。

28. 权利要求 1 的方法,其中所述情感障碍是双相型障碍 I、双相型障碍 II、心境恶劣障碍或抑郁障碍。

29. 权利要求 1 的方法,其中所述情感障碍是轻度抑郁症、中度抑郁症、重度抑郁症、非典型抑郁症、忧郁抑郁症或边缘型人格障碍。

30. 一种计算机程序产品,其中所述计算机程序产品包含计算机可读存储介质和其中嵌入的计算机程序机制,所述计算机程序机制包含用于执行权利要求 1 的方法的指令。

31. 权利要求 30 的计算机程序产品,其中所述计算机程序机制进一步包含指令用于将所述测试受试者是否具有所述情感障碍的诊断输出至用户界面装置、监测器、可触摸计算机可读存储介质、或局部或远程计算机系统;或以用户可读形式显示所述测试受试者是否具有所述情感障碍的诊断。

32. 一种计算机,其包含:

一个或多个处理器;

与上述一个或多个处理器连接的存储器,所述存储器存储用于执行权利要求 1 的方法的指令。

33. 权利要求 32 的计算机,其中所述存储器进一步包含指令用于将所述测试受试者是否具有所述情感障碍的诊断输出至用户界面装置、监测器、可触摸计算机可读存储介质、或局部或远程计算机系统;或以用户可读形式显示所述测试受试者是否具有所述情感障碍的诊断。

34. 一种测定测试受试者显示出情感障碍症状的可能性的方法,所述方法包括:

评估所述测试受试者的生物标记概况中的多个生物标记的多个特征是否满足值集合,其中满足所述值集合提供了所述测试受试者显示出情感障碍症状的所述可能性,并且其中所述多个特征是所述多个生物标记的可测量方面,所述多个生物标记包含表 1A 中列出的至少 2 个生物标记。

35. 权利要求 34 的方法,所述方法进一步包括将所述测试受试者显示出情感障碍症状的可能性输出至用户界面装置、监测器、可触摸计算机可读存储介质、或局部或远程计算机系统;或以用户可读形式显示所述测试受试者显示出情感障碍症状的可能性。

## 用于测量生物标记概况的系统和方法

[0001] 本申请含有 2009 年 8 月 25 日创建、大小 148,658 字节、以电子形式作为文件名 71021-W0-PCT\_SequenceListing\_ST25.txt 提交的序列列表。序列列表在此整体引入作为参考。

### 1 发明领域

[0002] 本发明提供了通过概况分析且比较相对于患病受试者的在对照受试者中基因的 mRNA 表达水平, 鉴定在患有病症的受试者中的转录概况的方法和组合物。本发明进一步提供了通过测定与此类受试者中的生物标记相关的转录概况, 用于预测和诊断受试者中的病症例如情感障碍的方法和组合物。

### 2 发明背景

[0004] 本申请自始至终通过在括号内引用提及各种出版物。这些出版物的公开内容整体在此引入本申请内作为参考, 以便更全面地描述本发明所属领域状态。

[0005] 目前精神病诊断分类, 特别是用于情感障碍的那些, 缺乏独特的临床描述, 并且不包括生物学特征以描绘一种诊断实体与另一种。尽管今天的分类法允许进一步指定情感障碍例如主要性抑郁障碍的临床特征, 但标准仍是主要讨论的问题且不一定遵循生物学原理 (Parker, 等人 *Am. J. Psychiatry* 2000, 157(8) :1195-1203)。

[0006] 在情感障碍中, 存在许多临床分割, 例如双相型障碍 I 和 II、精神抑郁症和主要性抑郁障碍, 包括精神病性抑郁症、重度与轻度或中度抑郁症比较、忧郁症与非典型抑郁症比较等。像这样, 对于这些分割没有描述独特的生物学标记或生物标记。此外, 缺乏关于特定病症的分割可以具有治疗牵涉。此外, 共病对于无法描绘 2 种病症存在的医生是成问题的。

[0007] 总的来说, 精神病学中的临床评估和非特异性临床诊断标准突出了关于生物学标记的需要, 以便识别共享类似生物学的需要。这看起来是关于情感障碍的特定困境, 因为存在关于显示临床差异和独特生物学特征的亚型的新出现证据 (Gold 和 Chrousos, *Mol. Psychiatry* 2002, 7(3) :254-275)。然而, 迄今为止, 没有生物学标记一致地显示描绘就情感障碍而言的患者群体的分割。

[0008] 先前研究已探索了测量具有抑郁症的受试者与对照受试者比较, 或在治疗前和后的受试者中的生物学改变的测试, 例如地塞米松 / 促肾上腺皮质激素释放激素 (DEX/CRH) 测试。然而, 此类测试已在少量患者中进行检查, 仍未再现, 和 / 或仍未使生物学读数与特定表型联系。(Ising, M. 等人, *Biol. Psychiatry*, 2006 Nov 20, e-pub ahead of print; Kunugi, H. 等人, *Neuropsychopharm.* 2006, 31(1) :212-20)。这是适当的, 因为临床上相关的生物标记必须与特定生物学和特定表型相关, 并且理想地, 应通过治疗回到正常水平。

[0009] 对于糖尿病、阿尔茨海默氏病和癌症已鉴定了蛋白质生物标记 (参见例如, 美国专利号 7, 125, 663 ; 7, 097, 989 ; 7, 074, 576 ; 和 6, 925, 389。) 然而, 用于检测蛋白质生物标记的方法, 例如质谱法和与抗体特异性结合, 通常获得不能再现的数据, 并且这些方法不利于高通量使用。

[0010] 使用微阵列的高流通量表达分析方法已用于评估基因表达改变, 具有混合结果或无关结果 (Brenner, S. 等人 *Nat Biotechnol.* 2000, 18(6) :597-8 ; Schena 等人

Science. 1995, 270 (5235) :467-70 ;Velculescu, V.E. 等人, Science. 1995, 270 (5235) : 484-7)。由于测量的基因表达与受试者数目的大量比,并且考虑到抑郁性病症的异质性,对于微阵列数据预期大量假阳性。(关于综述,参见 Iwamoto K 和 Kato T., Neuroscientist 2006, 12 (4) :349-61 ;Bunney WE, 等人, Am J Psychiatry 2003, 160 (4) : 657-66 ;和 Iga J, Ueno S 和 Ohmori T., Ann Med 2008, 40 (5) :336-42。)Sibille 等人 (Neuropsychopharm. 2004, 29 (2) :351-61) 执行了大规模基因组分析,然而,未发现关于与抑郁症和自杀相关的分子差异证据,并且无法再现关于先前发现与抑郁症相关的基因在表达水平中的改变。因为此类困难,一致概况仍未得到鉴定。

[0011] 关于多重相关基因的聚焦阵列和 qPCR 已用于鉴定应激相关基因,但这些研究仍未鉴定与抑郁症相关的诊断概况 (Rokutan 等人, J. Med. Invest. 2005, 52 (3-4) :137-44 ; Ohmori 等人, J. Med. Invest. 2005, 52 (Suppl) :266-71)。在大鼠脑区域中,特定基因的转录改变已牵涉情绪和焦虑的控制,然而,这些改变与人血样无关 (W02007106685A2)。

[0012] 3 发明概述

[0013] 本发明提供了诊断测试受试者中的情感障碍的方法,该方法包括:评估测试受试者的生物标记概况中的多个生物标记的多个特征是否满足值集合,其中满足该值集合预测测试受试者具有所述情感障碍,并且其中多个特征是多个生物标记的可测量方面,多个生物标记包含表 1A 中列出的至少 2 个生物标记。

[0014] 本发明还提供了计算机程序产品,其中所述计算机程序产品包含计算机可读存储介质和其中嵌入的计算机程序机制,所述计算机程序机制包含用于执行诊断方法的指令。

[0015] 本发明的一个方面提供了计算机,所述计算机包含一个或多个处理器和与一个或多个处理器连接的存储器,所述存储器存储用于执行诊断方法的指令。

[0016] 本发明的另一个方面提供了测定测试受试者显示出情感障碍症状的可能性的方法,该方法包括:评估测试受试者的生物标记概况中的多个生物标记的多个特征是否满足值集合,其中满足该值集合提供了测试受试者显示出情感障碍症状的所述可能性,并且其中多个特征是多个生物标记的可测量方面,所述多个生物标记包含表 1A 中列出的至少 2 个生物标记。

[0017] 在另一个方面,本发明提供了其为关于从多个对照受试者中收集的每种生物样品的转录分析测量的转录概况。例如,本发明提供了其为关于从多个抑郁、严重抑郁或双相型受试者中收集的每种生物样品的转录分析测量的转录概况。本发明进一步提供了其为关于从多个边缘型人格障碍受试者中收集的每种生物样品的转录分析测量的转录概况。本发明还提供了其为关于从多个 PTSD 受试者中收集的每种生物样品的转录分析测量的转录概况。

[0018] 本发明还提供了包含存储在例如数据库中的第一批多个对照受试者的集体测量的转录概况。使用分类算法使包含第二批多个受试者例如患病受试者的集体测量的转录概况与第一批多个对照受试者的转录概况比较。分类算法提供了使受试者各自分类的输出。

[0019] 本发明提供了用于诊断情感障碍的方法,通过鉴定患者中的转录概况,使此类转录概况与对照受试者或对照受试者组的概况比较,从而基于转录概况中改变的存在或不存在诊断患者的情感障碍。

[0020] 本发明的一个方面提供了用于诊断具有情感障碍的受试者的方法,其包括:

- [0021] (a) 从多个对照受试者和多个患病受试者中获得生物样品；
- [0022] (b) 测量在多个对照受试者和多个患病受试者样品中基因的 mRNA 表达水平，其中所述基因选自 ADA、ARRB1、ARRB2、CD8a、CD8b、CREB1、CREB2、DPP4、ERK1、ERK2、Gi2、Gs、GR、IL1b、IL6、IL8、INDO、MAPK14、MAPK8、MKP1、MR、ODC1、P2X7、PBR、PREP、RGS2、S100A10、SERT 和 VMAT2；
- [0023] (c) 收集关于来自多个对照受试者和多个患病受试者的每种基因的 mRNA 表达水平且将其作为 mRNA 数据存储于计算机介质中；
- [0024] (d) 借助于分类算法加工此类 mRNA 数据；和
- [0025] (e) 提供使受试者分类的输出数据，
- [0026] 从而诊断受试者具有情感障碍。
- [0027] 本发明进一步提供了用于预测受试者对情感障碍的易感性的方法，通过使受试者选自下述的基因转录概况与多个对照受试者的基因转录概况比较来进行：ADA、ARRB1、ARRB2、CD8a、CD8b、CREB1、CREB2、DPP4、ERK1、ERK2、Gi2、Gs、GR、IL1b、IL6、IL8、INDO、MAPK14、MAPK8、MKP1、MR、ODC1、P2X7、PBR、PREP、RGS2、S100A10、SERT 和 VMAT2。
- [0028] 本发明的一个方面提供了用于预测受试者显示出情感障碍症状的可能性的方法，其包括：
- [0029] (a) 从多个对照受试者和多个患病受试者中获得生物样品；
- [0030] (b) 测量在多个对照受试者和多个患病受试者样品中基因的 mRNA 表达水平，其中所述基因选自 ADA、ARRB1、ARRB2、CD8a、CD8b、CREB1、CREB2、DPP4、ERK1、ERK2、Gi2、Gs、GR、IL1b、IL6、IL8、INDO、MAPK14、MAPK8、MKP1、MR、ODC1、P2X7、PBR、PREP、RGS2、S100A10、SERT 和 VMAT2；
- [0031] (c) 收集关于来自多个对照受试者和多个患病受试者的每种基因的 mRNA 表达水平且将其作为 mRNA 数据存储于计算机介质中；
- [0032] (d) 借助于分类算法加工此类 mRNA 数据；和
- [0033] (e) 提供使受试者分类的输出数据，
- [0034] 从而预测受试者显示出情感障碍症状的可能性。
- [0035] 4 附图简述
- [0036] 图 1 是依照本发明的一个实施方案的计算机系统举例说明。
- [0037] 图 2A 和 2B。如通过 qPCR 方法经由拷贝 /ng CDR 测量的 ( $p < 0.001$ ；曼-怀二氏 (Mann Whitney) 检验)，分别显示在对照受试者与抑郁受试者比较中 ARRB1 ( $\beta$ -抑制蛋白 1) 和 Gi2 (鸟嘌呤核苷酸结合蛋白  $\alpha i2$ ) 的相对 mRNA 水平的散点图。
- [0038] 图 3A 和 3B。如通过 qPCR 方法经由拷贝 /ng CDR 测量的 ( $p < 0.001$ ；曼-怀二氏检验)，分别显示在对照受试者与抑郁受试者比较中 MAPK14 (p38 丝裂原活化蛋白激酶 14) 和 ODC1 (鸟氨酸脱羧酶 1) 的相对 mRNA 水平的散点图。
- [0039] 图 4A、4B 和 4C。如通过 qPCR 方法经由拷贝 /ng CDR 测量的 ( $p < 0.001$ ；曼-怀二氏检验)，分别显示在对照受试者与严重抑郁受试者比较中 ERK1 (细胞外信号调节激酶 1)、Gi2 (鸟嘌呤核苷酸结合蛋白  $\alpha i2$ ) 和 MAPK14 (p38 丝裂原活化蛋白激酶 14) 的相对 mRNA 水平的散点图。
- [0040] 图 5A、5B 和 5C。如通过 qPCR 方法经由拷贝 /ng CDR 测量的 ( $p < 0.001$ ；曼-怀

二氏检验),分别显示在对照受试者与严重抑郁/双相型受试者比较中 Gi2(鸟嘌呤核苷酸结合蛋白  $\alpha$  i2)、GR( $\alpha$ -糖皮质激素受体)和 MAPK14(p38 丝裂原活化蛋白激酶 14)的相对 mRNA 水平的散点图。

[0041] 图 6A、6B 和 6C。如通过 qPCR 方法经由拷贝/ng CDR 测量的 ( $p < 0.001$ ;曼-怀二氏检验),分别显示在对照受试者与边缘型人格障碍受试者比较中 Gi2(鸟嘌呤核苷酸结合蛋白  $\alpha$  i2)、MAPK14(p38 丝裂原活化蛋白激酶 14)和 MR(盐皮质激素受体)的相对 mRNA 水平的散点图。

[0042] 图 7A、7B 和 7C。如通过 qPCR 方法经由拷贝/ng CDR 测量的 ( $p < 0.001$ ;曼-怀二氏检验),分别显示在 196 个对照受试者与 66 个急性 PTSD 受试者比较中 ARRB2( $\beta$ -抑制蛋白 2)、ERK2(细胞外信号调节激酶 2)和 RGS2(G 蛋白信号调节剂 2)的相对 mRNA 水平的散点图。

[0043] 图 8A 和 8B。图 8A 是执行 SLR 算法的举例说明,所述 SLR 算法执行基因选择和训练,在抑郁受试者与对照比较的分类中评分为 93%的准确度、PPV = 93%和 NPV = 94%。在 RF 基因选择前,支持向量机(Support Vector Machine)(SVM)分类器在抑郁受试者与对照比较的分类中评分为 88%的准确度、PPV = 89%和 NPV = 88%。图 8B 显示基于在抑郁受试者与对照比较的分类中每种方法的统计参数,随机森林(Random Forest)(RF)从表 1A 中选择 14 种基因,并且逐步逻辑回归(Stepwise Logistic Regression)(SLR)选择 17 种基因。在分类过程的选择步骤时通过 RF 和 SLR 方法选择的重叠基因以灰色显示。

[0044] 图 9。图 9 描述了关于其平均表达水平(转录值)在严重抑郁患者和对照之间显著不同 ( $p < 0.05$ ) 的基因。这些基因根据计算的  $-\text{Log}(p)$  值量级进行排序,如表 5A 中可见的。

[0045] 图 10。图 10 表示对于每个受试者根据由 ERK1 和 MAPK14 组成的转录概况的严重抑郁受试者和对照受试者分布。严重抑郁受试者由空心圆圈( $\circ$ )表示,并且对照受试者由实心三角( $\blacktriangle$ )表示。X 和 Y 轴分别描述了关于 ERK1 和 MAPK14 的转录值(拷贝/ng cDNA)。

[0046] 图 11。图 11 表示对于每个受试者根据由 Gi2 和 IL1b 组成的转录概况的严重抑郁受试者和对照受试者分布。严重抑郁受试者由空心圆圈( $\circ$ )表示,并且对照受试者由实心三角( $\blacktriangle$ )表示。X 和 Y 轴分别描述了关于 Gi2 和 IL1b 的转录值(拷贝/ng cDNA)。

[0047] 图 12。图 12 表示对于每个受试者根据由 ERK1 和 IL1b 组成的转录概况的严重抑郁受试者和对照受试者分布。严重抑郁受试者由空心圆圈( $\circ$ )表示,并且对照受试者由实心三角( $\blacktriangle$ )表示。X 和 Y 轴分别描述了关于 ERK1 和 IL1b 的转录值(拷贝/ng cDNA)。

[0048] 图 13。图 13 表示对于每个受试者根据由 ARRB1 和 MAPK14 组成的转录概况的严重抑郁受试者和对照受试者分布。严重抑郁受试者由空心圆圈( $\circ$ )表示,并且对照受试者由实心三角( $\blacktriangle$ )表示。X 和 Y 轴分别描述了关于 ARRB1 和 MAPK14 的转录值(拷贝/ng cDNA)。

[0049] 5 发明详述

[0050] 本发明允许通过评估生物标记概况中的生物标记特征快速且准确诊断情感障碍。这些生物标记概况从受试者的生物样品构建。

[0051] 5.1 定义

[0052] 如本文使用的,“情感障碍”应意指特征在于一致、普遍的情绪改变且影响思维、情绪和行为的障碍。情感障碍的例子包括但不限于抑郁性障碍、焦虑症、双相型障碍、精神抑郁症和分裂情感性障碍。焦虑症包括但不限于广泛性焦虑症、惊恐障碍、强迫性障碍、恐怖症和创伤后应激障碍。抑郁性障碍包括但不限于主要性抑郁障碍 (MDD)、紧张性抑郁症、忧郁性抑郁症、非典型抑郁症、精神病性抑郁症、产后忧郁症、双相型抑郁症和轻度、中度或重度抑郁症。人格障碍包括但不限于偏执型、反社会型和边缘型人格障碍。

[0053] “生物标记”是事实上任何可检测的化合物,例如蛋白质、肽、蛋白聚糖、糖蛋白、脂蛋白、碳水化合物、脂质、核酸(例如 DNA,例如 cDNA 或扩增的 DNA,或 RNA 例如 mRNA)、有机或无机化学制品、天然或合成聚合物、小分子(例如代谢产物),或前述任何的区别分子或区别片段,其存在于或衍生自生物样品,或客观测量且评估为正常生物学过程、致病性过程或对于治疗干预的药理学应答指示剂,或其指示的任何其他特征。参见 Atkinson, A. J., 等人 Biomarkers and Surrogate Endpoints: Preferred Definitions and Conceptual Framework, *Clinical Pharm. & Therapeutics*, 2001 March ;69(3):89-95。如这个内容中使用的“衍生自”指当检测时指示存在于生物样品中的特定分子的化合物。例如,特定 cDNA 的检测可以指示特定 RNA 转录物在生物样品中的存在。作为另一个例子,与特定抗体结合的检测可以指示生物样品中特定抗原(例如,蛋白质)的存在。此处,区别分子或片段是当检测时指示上文鉴定化合物的存在或丰度的分子或片段。

[0054] 生物标记可以例如从生物样品中分离,在生物样品中直接测量,或在生物样品中检测或在生物样品中待测定。生物标记可以例如是功能的、部分功能的或无功能的。在一个实施方案中,分离且使用生物标记,例如以产生可以促进各种诊断测定中的生物标记检测的特异性结合抗体。任何免疫测定可以使用能够结合生物标记分子(例如, Fab、F(ab')<sub>2</sub>、Fv 或 scFv 片段)的任何抗体、抗体片段或其衍生物。此类免疫测定是本领域众所周知的。此外,如果生物标记是蛋白质或其片段,那么它可以进行测序,并且可以使用充分确定的技术克隆其编码基因。

[0055] 如本文使用的,术语“生物标记种类”指本文描述的生物标记的任何区别部分或区别片段,例如本文描述的特定基因的剪接变体(例如,下文表 1A 中列出的基因)。此处,区别部分或区别片段是当检测时指示上文鉴定的转录物、cDNA、扩增核酸或蛋白质的存在或丰度的分子部分或片段。

[0056] “生物标记概况”包含多个一种或多种类型的生物标记(例如, mRNA 分子、cDNA 分子、蛋白质和 / 或碳水化合物或其指示等),连同生物标记的特征例如可测量方面(例如丰度)。生物标记概况包含至少 2 个此类生物标记,其中生物标记可以在相同或不同类别中,例如核酸和碳水化合物。生物标记概况还可以包含至少 2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95 或 100 个或更多个生物标记。在一个实施方案中,生物标记概况包含数百个或甚至数千个生物标记。生物标记概况可以进一步包含一种或多种对照或内部标准。在一个实施方案中,生物标记概况包含充当内部标准的至少一个生物标记。如本文在这个内容中使用的,术语“指示”仅指其中生物标记概况含有关于核酸、mRNA 分子、cDNA 分子、蛋白质和 / 或碳水化合物、或任何其他形式的生物标记,而不是生物标记分子实体其自身的标志、数据、缩写或其他相似标记的情况。例如,本发明的示例性生物标记概况包含表 1A 中的基因名称。

[0057] 生物标记概况中的每个生物标记包括相应“特征”。如本文使用的，“特征”指生物标记的可测量方面。特征可以包括例如如示例性生物标记概况 1 中举例说明的，来自受试者的生物样品中生物标记的存在或不存在：

[0058] 示例性生物标记概况 1

[0059]	生物标记	特征 在样品中的存在
	基因 A 的转录物	存在
	基因 B 的转录物	不存在

[0060] 在示例性生物标记概况 1 中，关于基因 A 的转录物的特征值是“存在”，并且关于基因 B 的转录物的特征值是“不存在”。

[0061] 特征可以包括例如如示例性生物标记概况 2 中举例说明的，来自受试者的生物样品中生物标记的丰度：

[0062] 示例性生物标记概况 2

[0063]	生物标记	特征 以相对单位在样品中的丰度
	基因 A 的转录物	300
	基因 B 的转录物	400

[0064] 在示例性生物标记概况 2 中，关于基因 A 的转录物的特征值是 300 单位，并且关于基因 B 的转录物的特征值是 400 单位。

[0065] 特征还可以是如示例性生物标记概况 3 中举例说明的，生物标记的 2 个或更多个可测量方面的比：

[0066] 示例性生物标记概况 3

[0067]	生物标记	特征 基因 A 的转录物/基因 B 的转录物的丰度比值
	基因 A 的转录物	
	基因 B 的转录物	300/400

[0068] 在示例性生物标记概况 3 中，关于基因 A 的转录物的特征值和关于基因 B 的转录物的特征值是 0.75 (300/400)。

[0069] 在某些实施方案中，如上文示例性生物标记概况 1 中举例说明的，在生物标记概况中的特征和生物标记之间存在一对一对应。在某些实施方案中，如上文示例性生物标记概况 3 中举例说明的，在本发明的生物标记概况中的特征和生物标记之间的关系更复杂。

[0070] 本领域技术人员应当理解可以设计特征的其他计算方法,并且所有此类方法在本发明的范围内。例如,特征可以代表跨越在 2 个或更多个时间点时从受试者中收集的生物样品的生物标记丰度平均值。此外,特征可以是在单个时间点中从受试者中获得的生物样品中 2 个或更多个生物标记丰度的差异或比。生物标记概况还可以包含至少 2、3、4、5、10、20、30 个或更多个特征。在一个实施方案中,生物标记概况包含数百个或甚至数千个特征。

[0071] 在某些实施方案中,生物标记的特征使用定量 PCR(qPCR) 进行测量。qPCR 测量基因转录物丰度的用途是众所周知的。在某些实施方案中,生物标记的特征使用微阵列进行测量。微阵列的构建和用于加工微阵列以便获得丰度数据的技术是众所周知的,并且例如由 Draghici, 2003, *Data Analysis Tools for DNA Microarrays*, Chapman&Hall/CRC, 和国际公开号 WO 03/061564 描述。微阵列包含多种探针。在某些情况下,每种探针识别不同生物标记例如与其结合。在某些情况下,在微阵列上的 2 种或更多种不同探针识别相同生物标记例如与其结合。因此,一般地,在微阵列上的探针点和受试者生物标记之间的关系是二对一对应、三对一对应、或某些其他形式的对应。然而,可以存在这样的情况,在微阵列上的探针和生物标记之间存在独特的一对一对应。

[0072] 如本文使用的,在核酸序列(例如编码本文描述基因的核苷酸序列)内容中的术语“互补的”指由于其氢键合性质,特定含氮碱基之间的化学亲和力。例如,鸟嘌呤(G)仅与胞嘧啶(C)形成氢键,而腺嘌呤在 DNA 的情况下仅与胸腺嘧啶(T)形成氢键,并且在 RNA 的情况下仅与尿嘧啶(U)形成氢键。这些反应描述为碱基配对,并且配对碱基(G与C,或A与T/U)被说成是互补的。因此,2 个核酸序列可以是互补的,如果它们的含氮碱基能够形成氢键。此类序列被称为彼此的“互补物”。此类互补物序列可以是天然存在的,或它们可以通过本领域技术人员已知的任何方法化学合成,如例如在与 DNA 分子或 RNA 分子(例如, mRNA 转录物)的有义链互补的反义核酸分子的情况下。参见例如, Lewin, 2002, *Genes VII*. Oxford University Press Inc., New York, NY。

[0073] 如本文使用的,“数据分析算法”是使用在训练群体中的受试者生物标记概况用于构建决策规则的算法。代表性数据分析算法在下文描述。“决策规则”是数据分析算法的最终产物,并且特征在于一个或多个值集合,其中这些值集合各自指示情感障碍的一个方面,情感障碍的发作、受试者将具有情感障碍的预测、或受试者显示出情感障碍症状的可能性。在一个具体例子中,值集合代表受试者将发生情感障碍的预测。在另一个例子中,值集合代表受试者将不发生情感障碍的预测。

[0074] “决策规则”是用于评估生物标记概况的方法。此类决策规则可以采取本领域已知的一种或多种形式,如 Hastie 等人, 2001, *The Elements of Statistical Learning*, Springer-Verlag, New York 中示例的。检测规则可以用于作用于特征的数据集,来尤其是预测情感障碍的存在、或受试者显示出或具有情感障碍症状的可能性,或显示出发生情感障碍的易感性。可以在本发明的某些实施方案中使用的示例性决策规则在下文进一步详细描述。

[0075] 如本文使用的,术语“内在表型”应意指与病相关的可遗传特征,例如生物标记,哪个特征是否存在个体是有症状的。(关于综述参见 Lenox 等人, 2002, *American Journal of Medical Genetics (Neuropsychiatric Genetics)* 114 :391-406)

[0076] 如本文使用的,术语“基因表达概况”和“转录概况”是通过所选基因的信使核糖

核酸 (mRNA) 水平的相对测量来测定的生物标记概况。转录概况通过来自受试者或患者的生物样品的基因转录分析进行测量。

[0077] 如本文使用的,“健康对照受试者”、“健康对照”和“对照受试者”应意指不具有主要目前医学或精神病学问题,但可能例如患有头痛的受试者。对照受试者优选具有低体重指数 (BMI, 小于 30), 过去 3 个月无药物使用, 和低或零应激得分、家族史得分和症状得分。对照受试者可以不具有任何精神病史、任何物质滥用史、任何精神病家族史、任何早期生活应激物或任何近期应激物, 如通过自填问卷测定的。在获得生物样品前, 对照受试者可以但无需通过医生进一步评估。

[0078] 如本文使用的, 术语“获得”意指“得到”。这可以例如通过从计算机系统中的数据存储中检索数据完成。这还可以例如通过直接测量完成。

[0079] 如本文使用的, 术语“表型”应意指可测量和 / 或可观察的生物学、临床或行为特征, 其是受试者的基因型和环境的结果。

[0080] 如本文使用的, 除非另有说明, 否则术语“蛋白质”、“肽”和“多肽”是可互换的。

[0081] 如本文使用的, “PTSD 对照受试者”应意指未实施极端创伤应激物且已由医生评估为不含任何神经精神性疾病的受试者。本发明的 PTSD 对照受试者是例如来自与显示出病症的受试者相同地理区域且具有相同性别的一般匹配受试者。

[0082] 如本文使用的, 在抗体内容中的术语“特异性”和类似术语指与抗原或片段特异性结合的肽、多肽和抗体或其片段, 并且不与其他抗原或其他片段特异性结合。与抗原特异性结合的肽或多肽可以以较低亲和力与其他肽或多肽结合, 如通过标准实验技术测定的, 例如通过本领域技术人员众所周知的任何免疫测定。此类免疫测定包括但不限于放射性免疫测定 (RIAs) 和酶联免疫吸附测定 (ELISAs)。与抗原特异性结合的抗体或片段可以与相关抗原交叉反应。优选地, 与抗原特异性结合的抗体或其片段不与其他抗原交叉反应。关于就抗原 - 抗体相互作用、特异性和交叉反应性的讨论, 和用于测定所有上述的方法, 参见例如, Paul 编, 2003, *Fundamental Immunology*, 第 5 版, Raven Press, New York 在第 69-105 页。

[0083] 如本文使用的, “受试者”是动物, 优选哺乳动物, 更优选非人灵长类, 且最优选人。术语“受试者”、“个体”、“候选者”和“患者”在本文中可互换使用。在某些实施方案中, 受试者是动物。在其他实施方案中, 受试者是哺乳动物。

[0084] 如本文使用的, “测试受试者”一般是不在于构建决策规则的训练群体中的任何受试者。测试受试者可以任选是怀疑具有情感障碍或发生情感障碍的可能性。

[0085] 如本文使用的, “训练群体”是来自用于构建决策规则的受试者群体的一组样品, 使用数据分析算法, 用于评估处于具有情感障碍危险中的受试者生物标记概况。在一个优选实施方案中, 训练群体包括来自具有情感障碍的受试者和不具有情感障碍的受试者的样品。

[0086] 如本文使用的, “验证群体”是来自用于测定决策规则的准确度或其他性能度量的受试者群体的一组样品。在一个优选实施方案中, 验证群体包括来自具有情感障碍的受试者和不具有情感障碍的受试者的样品。在一个优选实施方案中, 验证群体不包括其为用于训练对于其寻求准确度或其他性能度量的决策规则的训练群体部分的受试者。

[0087] 如本文使用的, “值集合”是关于生物标记概况中的特征的值组合, 或值范围。这个

值集合的性质和其中的值依赖于生物标记概况中存在的特征类型和用于构建指示值集合的决策规则的数据分析算法。为了举例说明,重新考虑示例性生物标记概况 2:

[0088] 示例性生物标记概况 2

[0089]	生物标记	特征 以相对单位在样品中的丰度
	基因 A 的转录物	300
	基因 B 的转录物	400

[0090] 在这个例子中,获得训练群体的每个成员的生物标记概况。每个此类生物标记概况包括关于基因 A 的转录物的测量特征(此处为丰度),和关于基因 B 的转录物的测量特征(此处为丰度)。这些特征值(此处为丰度值)通过数据分析算法用于构建决策规则。在这个例子中,数据分析算法是下文描述的决策树,并且这种数据分析算法的最终产物,决策规则是决策树。决策规则限定值集合。一个此类值集合是情感障碍的预测。其生物标记特征值满足这个值集合的受试者具有情感障碍。这个类别的示例性值集合是示例性值集合 1:

[0091] 示例性值集合 1

[0092]	生物标记	值集合组分 (以相对单位在样品中的丰度)
	基因 A 的转录物	< 400
	基因 B 的转录物	< 600

[0093] 另一个此类值集合预测情感障碍游离状态。其生物标记特征值满足这个值集合的受试者未诊断为具有情感障碍。这个类别的示例性值集合是示例性值集合 2:

[0094] 示例性值集合 2

[0095]	生物标记	值集合组分 (以相对单位在样品中的丰度)
	基因 A 的转录物	> 400
	基因 B 的转录物	> 600

[0096] 在其中数据分析算法是神经网络分析并且这个神经网络分析的最终产物是合适加权神经网络的情况下,一个值集合是生物标记概况特征值的那些范围,这将引起加权神经网络以指出受试者具有情感障碍。另一个值集合是生物标记概况特征值的那些范围,这将引起加权神经网络以指出受试者不具有情感障碍。

[0097] 如本文使用的,在微阵列内容中的术语“探针点”指单链 DNA 分子(例如,单链 cDNA 分子或合成 DNA 寡聚物),在本文中被称为“探针”,其用于测定样品中特定核酸的丰度。例如,探针点可以用于测定来自测试受试者的生物样品(例如,细胞集合)中的 mRNA 水平。在一个具体实施方案中,一般的微阵列包含置于玻璃载玻片(或其他底物)上在格栅上的已知位置中的多个探针点。关于每个探针点的核酸是基因或目的基因序列的单链邻接部分(例如,10 聚体、11 聚体、12 聚体、13 聚体、14 聚体、15 聚体、16 聚体、17 聚体、18 聚体、19 聚体、20 聚体、21 聚体、22 聚体、23 聚体、24 聚体、25 聚体或更大),并且是关于由特定基因或目的基因编码的 mRNA 的探针。每个探针点的特征在于单个核酸序列,并且在促使其仅与其互补 DNA 链或 mRNA 分子杂交的条件下杂交。像这样,在底物上可以存在许多探针点,并且每个可以代表独特基因或目的基因。此外,2 个或更多个探针点可以代表相同基因序列。在某些实施方案中,标记的核酸样品与探针点杂交,并且可以定量与探针点特异性杂交的标记核酸量,以测定特定生物样品中的特异性核酸(例如,特定基因的 mRNA 转录物)水平。探针、探针点和微阵列一般在 Draghici, 2003, Data Analysis Tools for DNA Microarrays, Chapman&Hall/CRC, 第 2 章中描述。

[0098] 5.2 用于筛选受试者的方法

[0099] 本发明允许通过在来自个体的生物样品中检测怀疑具有情感障碍的测试个体的生物标记概况的 2 个或更多个特征来准确、快速预测和 / 或诊断情感障碍。

[0100] 在本发明的具体实施方案中,怀疑具有情感障碍的受试者使用本发明的方法进行筛选。依照这些实施方案,本发明的方法可以用于筛选例如允许精神病看护的受试者和 / 或已经历某类心理创伤的那些受试者。

[0101] 在具体实施方案中,获取生物样品例如血液。在某些实施方案中,生物样品是血液、脑脊髓液、腹膜液、间隙液、红血细胞、白血细胞或血小板。白血细胞(白细胞)包括但不限于:嗜中性粒细胞、嗜碱性粒细胞、嗜酸性粒细胞、淋巴细胞、单核细胞和巨噬细胞。在某些实施方案中,生物样品是全血的某些组分。在一个实施方案中,本发明利用由含有 RNA 稳定剂或防腐剂的现成可用收集管取样的全血。这个方案得到证明且确保非常少的变异性,条件是遵循正确的样品处理程序。本发明提供了可靠且坚固的转录标记,其可以在高通量分析中用于大样品集合。这个可靠方法显示区分对照和患者。在某些实施方案中,在血液的细胞级分或液体(例如血浆或血清级分)内的蛋白质、核酸和 / 或其他分子(例如代谢产物)混合物的某些部分分辨为生物标记概况。这可以通过测量生物标记概况中的生物标记特征来完成。在某些实施方案中,生物样品是全血,但生物标记概况由从全血中分离的白血细胞中表达或以其他方式发现的生物标记分辨。在某些实施方案中,生物样品是全血,但生物标记概况由从全血中分离的红血细胞中表达或以其他方式发现的生物标记分辨。

[0102] 生物标记概况可以包含至少 2 个生物标记,其中生物标记可以在相同或不同类别中,例如核酸和碳水化合物。在某些实施方案中,生物标记概况包含至少 2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、96、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195 或 200 个或更多个生物标记。在一个实施方案中,生物标记概况包括数百个或甚至数千个生物标记。在某些实施方案中,生物标记概况包含至少 2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、45、50 个或更多个生物标记。在一个例子

中,在某些实施方案中,生物标记概况包含至少 2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20 个或更多个选自表 1A 的生物标记。

[0103] 在一般实施方案中,生物标记概况中的每个生物标记由特征代表。换言之,在生物标记和特征之间存在对应。在某些实施方案中,生物标记和特征之间的对应是 1 : 1,意指关于每个生物标记存在特征。在某些实施方案中,对于每个生物标记存在超过一个特征。在某些实施方案中,与生物标记概况中的一个生物标记对应的特征数目不同于与生物标记概况中的另一个生物标记对应的特征数目。像这样,在某些实施方案中,生物标记概况可以包括至少 2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、96、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195 或 200 个或更多个特征,条件是在生物标记概况中存在至少 2、3、4、5、6 或 7 个或更多个生物标记。在某些实施方案中,生物标记概况可以包括至少 2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、45、50 个或更多个特征。与实施方案无关,这些特征可以通过使用任何可再现测量技术或测量技术组合进行测定。此类技术包括本领域众所周知的那些,包括本文描述的任何技术,例如下文 5.4 节中公开的任何技术。一般地,此类技术用于测量特征值,使用在时间中在单个点时得自受试者的生物样品或在时间中在多个点时获得的多个样品。在一个实施方案中,从得自受试者的样品中获得生物标记概况的示例性技术是 cDNA 微阵列(参见例如,下文 5.4.1.2 节)。在另一个实施方案中,从得自受试者的样品中获得生物标记概况的示例性技术是基于蛋白质的阵列或其他形式基于蛋白质的技术,例如 BD Cytometric Bead Array (CBA) Human Inflammation Kit Instruction Manual (BD Biosciences) 中描述的,或美国专利号 5,981,180 中描述的珠阵列,其各自整体引入本文作为参考,并且特别地关于其测定生物样品中的蛋白质浓度的各种方法的教导。在另外一个实施方案中,生物标记概况是混合的,意指它包含其为核酸或其指示的某些生物标记,和其为蛋白质或其指示的某些生物标记。在此类实施方案中,基于蛋白质和基于核酸的技术都用于获得来自得自受试者的一个或多个样品的生物标记概况。换言之,关于其为核酸与生物标记概况中的生物标记相关的特征的特征值通过基于核酸的测量技术(例如,核酸微阵列)获得,并且关于其为蛋白质与生物标记概况中的生物标记相关的特征的特征值通过基于蛋白质的测量技术获得。在某些实施方案中,生物标记概况可以使用试剂盒获得,例如下文 5.3 节中描述的试剂盒。

#### [0104] 5.3 试剂盒

[0105] 本发明还提供了在诊断受试者中的情感障碍中有用的试剂盒。在某些实施方案中,本发明的试剂盒包含至少 2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、96、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195 或 200 个或更多个生物标记和 / 或试剂,以检测此类生物标记的存在或丰度。在其他实施方案中,本发明的试剂盒包含至少 2 个,但多达数百个或更多个生物标记。在某些实施方案中,本发明的试剂盒包含至少 2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20 种或更多种选自表 1A 的生物标记或试剂,以检测此类生物标记的存在或丰度。依照 5.1 节中给出的生物标记定义,在某些情况下,生物标记事实上是例如基因、mRNA 或蛋白质的区别分子,而不是基因、mRNA 或蛋白质其自身。因此,生物标记可以是指示表 1A 中鉴定的特定基因、mRNA 或蛋白质或其片段的存在

或丰度,而不是实际的基因、mRNA 或蛋白质其自身。在某些实施方案中,本发明的试剂盒包含至少 2 个但多达数百个或更多个生物标记。在某些实施方案中,检测生物标记的存在或丰度的至少 25%、至少 30%、至少 35%、至少 40%、至少 60%、至少 80% 生物标记和 / 或试剂选自来自表 1A 的生物标记和 / 或试剂,以检测选自表 1A 的生物标记的存在或丰度。

[0106] 本发明的试剂盒的生物标记可以用于生成根据本发明的生物标记概况。试剂盒化合物类别的例子包括但不限于蛋白质及其片段、肽、蛋白聚糖、糖蛋白、脂蛋白、碳水化合物、脂质、核酸(例如 DNA,例如 cDNA 或扩增的 DNA,或 RNA 例如 mRNA)、有机或无机化学制品、天然或合成聚合物、小分子(例如代谢产物),或前述任何的区别分子或区别片段。在一个具体实施方案中,生物标记具有特定大小(例如,至少 10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、1000、2000、3000、5000、10k、20k、100k 道尔顿或更大)。一个或多个生物标记可以是阵列的部分,或一个或多个可以分开和 / 或个别进行包装。试剂盒还可以包含至少一种内部标准,以用于生成本发明的生物标记概况。同样地,一种或多种内部标准可以是上文描述的化合物类别中的任何。

[0107] 在一个实施方案中,本发明提供了包含探针和 / 或引物的试剂盒,所述探针和 / 或引物可以固定或不固定在底物上的可寻址位置上,例如在微阵列中发现的。在一个特定实施方案中,本发明提供了此类微阵列。

[0108] 在本发明的某些实施方案中,试剂盒可以包含特异性生物标记结合组分,例如适体。如果生物标记包含核酸,那么试剂盒可以提供能够与生物标记或生物标记的互补链形成双链体的寡核苷酸探针。寡核苷酸探针可以可检测地进行标记。在此类实施方案中,探针其自身是属于本发明范围内的生物标记。

[0109] 本发明的试剂盒还可以包括可以用于构建生物标记概况的另外组合物,例如缓冲液。微生物作用的预防可以通过包括各种抗菌剂和抗真菌剂得以确保,例如对羟基苯甲酸酯、三氯叔丁醇、苯酚山梨酸等。很可能希望包括等渗剂例如糖、氯化钠等。

[0110] 本发明的某些试剂盒包含微阵列。在一个实施方案中,这个微阵列包含多个探针点,其中多个探针点中的至少 20% 探针点与表 1A 中的生物标记对应。在某些实施方案中,多个探针点中的至少 25%、至少 30%、至少 35%、至少 40%、至少 60% 或至少 80% 探针点与表 1A 中的生物标记对应,和 / 或检测表 1A 中生物标记的存在或丰度的试剂。此类探针点是在本发明的范围内的生物标记。在某些实施方案中,微阵列由在底物上的约 2 个和约 100 个探针点组成。在某些实施方案中,微阵列由在底物上的约 2 个和约 100 个探针点组成。如这个内容中使用的,术语“约”意指在所述值的 5% 内、在所述值的 10% 内或在所述值的 25% 内。在某些实施方案中,此类微阵列含有使用本领域技术人员已知的技术用于微阵列间校正或用于由其他微阵列例如参考微阵列校正的一个或多个探针点。在某些实施方案中,此类微阵列是核酸微阵列。在某些实施方案中,此类微阵列是蛋白质微阵列。

[0111] 本发明的某些试剂盒作为计算机程序产品实现,其包含嵌入计算机可读存储介质中的计算机程序机制。进一步地,本发明的任何方法都可以在一个或多个计算机或其他形式的仪器中实现。仪器的例子包括但不限于计算机和分光镜测量装置(例如,微阵列阅读器或微阵列扫描仪)。再进一步地,本发明的任何方法都可以在一种或多种计算机程序产品中实现。本发明的某些实施方案提供了编码本文公开的任何或所有方法的计算机程序产

品。此类方法可以存储于 CD-ROM、DVD、磁盘存储产物、或任何其他可触摸计算机可读数据或可触摸程序存储产物上。此类方法还可以嵌入永久存储器中，例如 ROM、一种或多种可编程芯片、或一种或多种专用集成电路 (ASICs)。此类永久存储器可以位于服务器、802.11 访问点、802.11 无线桥/电台、中继器、路由器、移动电话或其他电子装置中。在计算机程序产品中编码的此类方法还可以经由因特网或其他方式电子分布。

[0112] 本发明的某些试剂盒提供了含有一种或多种程序的计算机程序产品，所述一种或多种程序个别或共同执行本发明的任何方法。这些程序模块可以存储在 CD-ROM、DVD、磁盘存储产物、或任何其他可触摸计算机可读数据或可触摸程序存储产物上。程序模块还可以嵌入永久存储器中，例如 ROM、一种或多种可编程芯片、或一种或多种专用集成电路 (ASICs)。此类永久存储器可以位于服务器、802.11 访问点、802.11 无线桥/电台、中继器、路由器、移动电话或其他电子装置中。在计算机程序产品中的软件模块还可以经由因特网或其他方式电子分布。

[0113] 本发明的某些试剂盒包含具有一个或多个处理单元和与一个或多个处理单元偶联的存储器的计算机。存储器存储用于评估处于具有情感障碍危险中的测试受试者的生物标记概况中的多个特征是否满足值集合的指令。在某些实施方案中，满足值集合诊断受试者为具有情感障碍。在某些实施方案中，满足值集合诊断受试者为不具有情感障碍。在一个实施方案中，多个特征与表 1A 中列出的生物标记对应。

[0114] 图 1 详述了支持上述功能性的示例性系统。系统优选是具有下述的计算机系统 10：

[0115] ●中央处理单元 22；

[0116] ●用于存储软件和数据的主要非易失性存储单元 14，例如硬盘驱动器，存储单元 14 通过存储控制器 12 加以控制；

[0117] ●用于存储系统控制程序、数据和应用程序的系统存储器 36，优选高速随机存取存储器 (RAM)，包含从非易失性存储单元 14 装载的程序和数据；系统存储器 36 还可以包括只读存储器 (ROM)；

[0118] ●用户界面 32，包含一种或多种输入装置（例如，键盘 28）和显示器 26 或其他输出装置；

[0119] ●用于与任何有线或无线通讯网络 34（例如广域网，例如因特网）连接的网络接口卡 20；

[0120] ●用于互联系统的上述元件的内部总线 30；和

[0121] ●电源 24 以给上述元件提供动力。

[0122] 计算机 10 的操作主要通过操作系统 40 加以控制，所述操作系统 40 通过中央处理单元 22 执行。操作系统 40 可以存储于系统存储器 36 中。除操作系统 40 外，在一般实施中，系统存储器 36 包括：

[0123] ●用于控制对由本发明使用的各种文件和数据结构的访问的文件系统 42；

[0124] ●用于在构建依照本发明的一种或多种决策规则中使用的训练数据集 44；

[0125] ●用于加工训练数据且构建决策规则的数据分析算法模块 54；

[0126] ●一种或多种决策规则 56；

[0127] ●用于测定测试受试者的生物标记概况中的多个特征是否满足第一个值集合或

第二个值集合的生物标记概况评估模块 60；

[0128] ●包含生物标记 64 和关于每个此类生物标记特征 66 的测试受试者生物标记概况 62；和

[0129] ●本发明的选择生物标记和 / 或关于这些选择生物标记各自的一个或多个特征的数据库 68(例如表 1A)。

[0130] 训练数据集 46 包含关于多个受试者 46 的数据。对于每个受试者 46, 存在受试者标识符 48 和多个生物标记 50。对于每个生物标记 50, 存在至少一个特征 52。尽管图 1 中未显示, 但对于每个特征 52, 存在特征值。对于使用数据分析算法构建的每种决策规则 56, 存在至少一个决策规则值集合 58。

[0131] 如图 1 中举例说明的, 计算机 10 包含软件程序模块和数据结构。计算机 10 中存储的数据结构包括训练数据集 44、决策规则 56、测试受试者生物标记概况 62 和生物标记数据库 68。这些数据结构各自可以包含任何形式的数据存储系统, 包括但不限于平面 ASCII 或二进制文件、Excel 电子表格、关系数据库 (SQL)、或联机分析处理 (OLAP) 数据库 (MDX 和 / 或其变体)。在某些具体实施方案中, 此类数据结构各自以一个或多个包括层次结构 (例如星型架构) 的数据库的形式。在某些实施方案中, 此类数据结构各自以不具有明显层次 (例如, 未层次排列的尺寸表) 的数据库形式。

[0132] 在某些实施方案中, 存储于或易接近系统 10 的数据结构各自是单个数据结构。在其他实施方案中, 此类数据结构事实上包含多个数据结构 (例如, 数据库、文件、档案), 其可以全部或并非全部由相同计算机 10 容纳。例如, 在某些实施方案中, 训练数据集 44 包含存储在计算机 10 和 / 或可由计算机 10 跨越广域网 34 寻址的计算机上的多个 Excel 电子表格。在另一个例子中, 训练数据集 44 包含存储在计算机 10 上或跨越一个或多个计算机分布的数据库, 所述一个或多个计算机可由计算机 10 跨越广域网 34 寻址。

[0133] 应当理解图 1 中举例说明的许多模块和数据结构可以位于一个或多个远程计算机上。例如, 本申请的某些实施方案是网络服务型实现。在此类实施方案中, 生物标记概况评估模块 60 和 / 或其他模块可以位于客户计算机上, 其经由网络 34 与计算机 10 通讯。在某些实施方案中, 例如, 生物标记概况评估模块 60 可以是交互式网页。

[0134] 在某些实施方案中, 图 1 中举例说明的训练数据集 44、决策规则 56 和 / 或生物标记数据库 68 在单个计算机 (计算机 10) 上, 并且在其他实施方案中, 一个或多个此类数据结构和模块由一个或多个远程计算机 (未显示) 容纳。在一个或多个计算机上的在图 1 中举例说明的数据结构和软件模块的任何排列都在本发明的范围内, 只要这些数据结构和软件模块跨越网络 34 或通过其他电子方式可就彼此而言寻址。因此, 本发明完全包括广泛多样的计算机系统。

[0135] 本发明的另外一个实施方案提供了用于测定受试者是否具有情感障碍的图解式用户界面。图解式用户界面包含显示域用于显示在由远程计算机接受的载波上体现的以数字信号编码的结果。多个特征是多个生物标记的可测量方面。多个生物标记包含表 1A 中列出的至少 2 个生物标记。当测试受试者的生物标记概况中的多个特征满足第一个值集合时, 结果具有第一个值。当测试受试者的生物标记概况中的多个特征满足第二个值集合时, 结果具有第二个值。

[0136] 5.4 生物标记概况的生成

[0137] 根据一个实施方案,本发明的方法包括由得自受试者的生物样品生成生物标记概况。生物样品可以是例如外周组织、全血、脑脊髓液、腹膜液、间隙液、红血细胞、白血细胞或血小板。

#### [0138] 5.4.1 检测核酸生物标记的方法

[0139] 在本发明的具体实施方案中,生物标记概况中的生物标记是核酸。例如通过检测本文描述的一种或多种基因(例如,表 1A 中列出的基因)的表达产物(例如,多核苷酸或多肽),可以生成生物标记概况的此类生物标记和相应特征。在一个具体实施方案中,使用本领域技术人员众所周知的任何方法,包括但不限于杂交、微阵列分析、RT-PCR、核酸酶保护测定和 RNA 印迹分析,通过检测和/或分析由本文公开的基因(例如,表 1A 中列出的基因)表达的一种或多种核酸,获得生物标记概况中的生物标记和相应特征。

[0140] 在特定实施方案中,通过本发明的方法和组合物检测和/或分析的核酸包括 RNA 分子,例如所表达的 RNA 分子,其包括信使 RNA (mRNA) 分子、mRNA 剪接变体以及调节 RNA、cRNA 分子(例如,由在体外转录的 cDNA 分子制备的 RNA 分子)及其区别片段。通过本发明的方法和组合物检测和/或分析的核酸还可以包括例如 DNA 分子,例如基因组 DNA 分子、cDNA 分子及其区别片段(例如,寡核苷酸、ESTs、STSs 等)。

[0141] 通过本发明的方法和组合物检测和/或分析的核酸分子可以从样品中分离的天然存在的核酸分子例如基因组或基因组外 DNA 分子,或在生物样品中存在、从生物样品中分离或衍生自生物样品的 RNA 分子例如 mRNA 分子。通过本发明的方法和组合物检测和/或分析的核酸样品包含例如 DNA、RNA 的分子,或 DNA 和 RNA 的共聚物。一般地,这些核酸对应于特定基因或基因的等位基因,或与特定基因转录物(例如,在特定细胞类型中表达的特定 mRNA 序列或衍生自此类 mRNA 序列的特定 cDNA 序列)。通过本发明的方法和组合物检测和/或分析的核酸可以对应于相同基因的不同外显子,例如从而使得可以检测和/或分析基因的不同剪接变体。

[0142] 在具体实施方案中,核酸由在生物样品中存在、或从生物样品中分离或部分分离的核酸体外进行制备。例如,在一个实施方案中,从样品中提取 RNA(例如,总细胞 RNA、聚(A)<sup>+</sup>信使 RNA、其部分),并且从总提取 RNA 中纯化信使 RNA。用于制备总和聚(A)<sup>+</sup>RNA 的方法是本领域众所周知的,并且一般例如在 Sambrook 等人,2001, *Molecular Cloning: A Laboratory Manual*. 第 3 版 Cold Spring Harbor Laboratory Press(Cold Spring Harbor, New York) 中描述。

#### [0143] 5.4.1.1 核酸阵列

[0144] 在本发明的特定实施方案中,通过检测本文描述的任何一种或多种基因(例如,表 1A 中列出的基因)的表达,核酸阵列用于生成生物标记概况中的生物标记特征。在本发明的一个实施方案中,微阵列例如 cDNA 微阵列用于测定生物标记概况中的生物标记特征值。cDNA 阵列的诊断用途是本领域众所周知的。(参见例如, Zou 等人,2002, *Oncogene* 21:4855-4862; 以及 Draghici,2003, *Data Analysis Tools for DNA Microarrays*, Chapman&Hall/CRC)。用于 cDNA 微阵列分析的示例性方法在下文描述。

[0145] 在特定实施方案中,关于生物标记概况中的生物标记特征值通过与阵列可检测标记的核酸杂交获得,所述核酸代表或对应于生物样品中存在的 mRNA 转录物中的核酸序列(例如,由样品合成的荧光标记的 cDNA),与包含一个或多个探针点的微阵列。

[0146] 核酸阵列例如微阵列可以以许多方式制备,其中几个在本文下文中描述。优选地,阵列是可再现的,允许生产给定阵列的多个拷贝且使来自所述微阵列的结果彼此比较。优选地,阵列由在结合(例如,核酸杂交)条件下稳定的材料制备。技术人员将知道用于使测试探针与阵列上的探针点杂交的合适支持物、底物或载体,或将能够通过使用例行实验确定这些。

[0147] 所使用的阵列例如微阵列可以包括一种或多种测试探针。在某些实施方案中,每个此类测试探针包含与待检测的 RNA 或 DNA 子序列互补的核酸序列。每个探针一般具有不同核酸序列,并且在阵列的固体表面上的每个探针位置是通常已知的或可以进行测定。依照本发明有用的阵列可以包括例如寡核苷酸微阵列、基于 cDNA 的阵列、SNP 阵列、剪接变体阵列和能够提供本文所述基因(例如,表 1A 中列出的基因)表达的定性、定量或半定量测量的任何其他阵列。某些类型的微阵列是可寻址阵列。更具体而言,某些微阵列是可定位寻址阵列。在某些实施方案中,阵列的每个探针位于固体载体上的已知、预定位置上,从而使得每个探针的身份(例如,序列)可以由其在阵列上的位置测定(例如,在载体或表面上)。在某些实施方案中,阵列是有序阵列。微阵列一般在 Draghici, 2003, Data Analysis Tools for DNA Microarrays, Chapman&Hall/CRC 中描述。

[0148] 在本发明的某些实施方案中,表达转录物(例如,本文描述基因的转录物)在核酸阵列中表示。在此类实施方案中,一组结合位点可以包括具有不同核酸的探针,所述不同核酸与所表达转录物的不同序列区段互补。属于这个类别内的示例性核酸可以具有 15-200 个碱基、20-100 个碱基、25-50 个碱基、40-60 个碱基或某些其他碱基范围的长度。每个探针序列还可以包含一个或多个接头序列,加上与其靶序列互补的序列。如本文使用的,接头序列是在与其靶序列互补的序列和载体表面之间的序列。例如,本发明的核酸阵列可以包含对于每个靶基因或外显子特异的一个探针。然而,需要时,核酸阵列可以含有对于某些表达转录物(例如,本文描述的例如表 1A 中的基因的转录物)特异的至少 2、5、10、100 或 1000 个或更多个探针。例如,阵列可以含有跨越基因的最长 mRNA 同种型序列倾斜的探针。

[0149] 应当理解当制备与细胞例如生物样品中的细胞的 RNA 互补的 cDNA 且在合适杂交条件下与微阵列杂交时,与阵列中对应于本文描述基因(例如,表 1A 中列出的基因)的位点的杂交水平将反映在细胞中由那种基因转录的 mRNA 或 mRNAs 的普遍。可替代地,在其中待区分由特定基因产生的多种同种型或可变剪接变体的情况下,与总细胞 mRNA 互补的可检测标记(例如用荧光团)的 cDNA 可以与微阵列杂交,并且在阵列上对应于在细胞中在 RNA 剪接过程中未转录或被去除的基因外显子的位点将具有很少的信号或无信号(例如,荧光信号),并且对应于对于其表达编码 mRNA 的外显子普遍的基因外显子的位点将具有相对强的信号。随后通过跨越对于基因监控的整个外显子组的信号强度模式测定通过可变剪接由相同基因产生的不同 mRNAs 的相对丰度。

[0150] 在一个实施方案中,在不同杂交时间时的杂交水平在不同、等同微阵列上分开测量。对于每个此类测量,在杂交时间时当测量杂交水平时,使微阵列短暂洗涤,优选在室温中高至中等盐浓度(例如 0.5-3M 盐浓度)的水溶液中在保留所有结合或杂交核酸同时去除所有未结合核酸的条件下。随后通过方法测量在每个探针上的在其余、杂交核酸分子上的可检测标记,所述方法适合于所使用的特定标记方法。随后使所得到的杂交水平组合,以形成杂交曲线。在另一个实施方案中,杂交水平使用单个微阵列进行实时测量。在这个

实施方案中,允许微阵列不中断地与样品杂交,并且在每个杂交时间时以非侵袭性方式询问微阵列。在另外一个实施方案中,可以使用一个阵列,短时间杂交,洗涤且测量杂交水平,放回相同样品中,杂交另一个时间段,洗涤且再次测量以获得杂交时间曲线。

[0151] 在某些实施方案中,这样选择核酸杂交和洗涤条件,使得待分析的核酸生物标记与阵列的互补核酸特异性结合或特异性杂交,一般与它的互补 DNA 位于其中的特定阵列位点特异性结合或特异性杂交。

[0152] 可以对含有双链探针 DNA 位于其上的阵列实施变性条件,以致使 DNA 在与靶核酸分子接触前成为单链。在与靶核酸分子接触前,含有单链探针 DNA (例如,合成寡聚脱氧核糖核酸) 的阵列可能需要变性,例如以去除由于自身互补序列形成的发夹或二聚体。

[0153] 最佳杂交条件将依赖于探针和靶核酸的长度(例如,寡聚物与超过 200 个碱基的多核苷酸比较)和类型(例如, RNA 或 DNA)。关于核酸的特异性(即严格)杂交条件的一般参数在 Sambrook 等人(同上)和 Ausubel 等人,最近版本, *Current Protocols in Molecular Biology*, Greene Publishing and Wiley-Interscience, New York 中描述。当使用 Shena 等人的 cDNA 微阵列时,一般杂交条件在 5X SSC 加上 0.2% SDS 中在 65°C 下杂交 4 小时,随后在低严格洗涤缓冲液(1X SSC 加上 0.2% SDS)中在 25°C 下洗涤,随后在更高严格性的洗涤缓冲液(0.1X SSC 加上 0.2% SDS)中在 25°C 下 10 分钟(Shena 等人, 1996, *Proc. Natl. Acad. Sci. U. S. A.* 93 :10614)。有用的杂交条件也在例如下述中提供:Tijessen, 1993, *Hybridization With Nucleic Acid Probes*, Elsevier Science Publishers B. V. ; Kricka, 1992, *Nonisotopic DNA Probe Techniques*, Academic Press, San Diego, CA ; 和 Zou 等人, 2002, *Oncogene* 21 :4855-4862 ; 和 Draghici, *Data Analysis Tools for DNA Microanalysis*, 2003, CRC Press LLC, Boca Raton, Florida, 第 342-343 页。

[0154] 在一个具体实施方案中,微阵列可以用于挑选出已通过例如下文在 5.4.1.2 节中描述的方法生成的 RT-PCR 产物。

[0155] 5.4.1.2 RT-PCR

[0156] 在特定实施方案中,为了测定本发明的生物标记概况中的生物标记特征值,通过使用与聚合酶链反应(PCR)组合的逆转录(RT)从样品扩增 RNA,测量本文描述的一种或多种基因(例如,表 1A 中列出的基因)的表达水平。依照这个实施方案,逆转录可以是定量或半定量的。本文教导的 RT-PCR 方法可以与上文例如在 5.4.1.1 节中描述的微阵列方法结合使用。例如,可以执行大批 PCR 反应,可以分辨 PCR 产物且用作微阵列上的探针点。

[0157] 来自样品的总 RNA 或 mRNA 用作模板,并且对于一种或多种基因的转录部分特异的引物用于起始逆转录。将 RNA 逆转录为 cDNA 的方法是众所周知的,并且在 Sambrook 等人, 2001, 同上中描述。引物设计可以基于已知核苷酸序列完成,所述核苷酸序列已得到公开或可从任何可公开获得的序列数据库例如 GenBank 获得。例如,可以对于本文描述的任何基因(参见例如表 1A 中)设计引物。进一步地,引物设计可以通过利用商购可得软件(例如 Primer Designer 1.0, Scientific Software 等)完成。逆转录的产物随后用作模板用于 PCR。

[0158] PCR 提供了用于快速扩增特定核酸序列的方法,通过使用由耐热的 DNA 依赖性 DNA 聚合酶催化的 DNA 复制的多个循环,以扩增目的靶序列。PCR 需要存在待扩增核酸,在待扩增序列侧面的 2 个单链寡核苷酸引物, DNA 聚合酶,脱氧核糖核苷三磷酸,缓冲液和盐。PCR

方法是本领域众所周知的。PCR 例如如 Mullis 和 Faloona, 1987, *Methods Enzymol.* 155 : 335 中所述执行。

[0159] PCR 可以使用模板 DNA 或 cDNA (至少 1fg ;更有用地, 1-1000ng) 和至少 25pmol 寡核苷酸引物执行。一般的反应混合物包括 :2  $\mu$  l DNA、25pmol 寡核苷酸引物、2.5  $\mu$  l 10M PCR 缓冲液 1 (Perkin-Elmer, Foster City, CA)、0.4  $\mu$  l 1.25M dNTP、0.15  $\mu$  l (或 2.5 单位) Taq DNA 聚合酶 (Perkin Elmer, Foster City, CA) 和去离子水至总体积 25  $\mu$  l。覆盖矿物油, 并且使用可编程热循环仪执行 PCR。

[0160] 根据实际上的严格性需求调整 PCR 循环每个步骤的长度和温度以及循环数目。通过对于其引物预期与模板退火的效率和待容忍的错配程度来测定退火温度和时机。最佳化引物退火条件的严格性的能力完全在本领域技术人员知识内。使用 30°C -72°C 之间的退火温度。模板分子的起始变性通常在 92°C -99°C 之间发生 4 分钟, 随后为由变性 (94-99°C 15 秒到 1 分钟)、退火 (如上讨论测定的温度 ;1-2 分钟) 和延伸 (72°C 1 分钟) 组成的 20-40 个循环。最终延伸步骤一般在 72°C 下执行 4 分钟, 并且可以随后为在 4°C 下的不确定 (0-24 小时) 步骤。

[0161] 还可以执行在性质中定量的定量 RT-PCR (“QRT-PCR”), 以提供基因表达水平的定量测量。在 QRT-PCR 中, 可以在 2 个步骤中执行逆转录和 PCR, 或可以同时执行与 PCR 组合的逆转录。用转录物特异性反义探针执行这些技术之一, 对于其存在商购可得的试剂盒, 例如 Taqman (Perkin Elmer, Foster City, California) 或如由 Applied Biosystems (Foster City, California) 提供的。这个探针对于 PCR 产物 (例如, 衍生自基因的核酸片段) 特异, 并且用与寡核苷酸的 5' 末端复合的猝灭剂和荧光报告探针制备。不同荧光标记与不同报告分子附着, 允许在一个反应中测量 2 种产物。当激活 Taq DNA 聚合酶时, 由于其 5' 至 3' 核酸外切酶活性, 它切割与模板结合的探针的荧光报告分子。在不存在猝灭剂的情况下, 报告分子现在发荧光。报告分子中的颜色改变与每种特异性产物的量成比例, 并且通过荧光计进行测量 ;因此, 测量每种颜色的量且定量 PCR 产物。在 96 孔板中执行 PCR 反应, 从而使得同时处理且测量受试者衍生自许多个体的样品。Taqman 系统具有不需要凝胶电泳的另外优点, 并且当与标准曲线一起使用时允许定量。

[0162] 对于定量检测 PCR 产物有用的第二种技术是使用嵌入染料, 例如商购可得的 QuantiTect SYBR Green PCR (Qiagen, Valencia California)。使用 SYBR 绿作为荧光标记执行 RT-PCR, 这在 PCR 阶段过程中掺入 PCR 产物内, 并且产生与 PCR 产物量成比例的荧光。

[0163] Taqman 和 QuantiTect SYBR 系统都可以随后用于 RNA 的逆转录。逆转录可以在与 PCR 步骤 (一步方案) 相同的反应混合物中执行, 或逆转录可以在利用 PCR 扩增前首先执行 (两步方案)。

[0164] 此外, 定量测量 mRNA 表达产物的其他系统是已知的, 包括 MOLECULAR BEACONS<sup>®</sup>, 其使用具有荧光分子和猝灭剂分子的探针, 所述探针能够形成发夹结构, 从而使得在以发夹形式时, 荧光分子被猝灭, 并且在杂交时, 荧光增加, 给出基因表达的定量测量。

[0165] 定量测量 RNA 表达的另外技术包括但不限于聚合酶链反应、连接酶链反应、Q $\beta$  复制酶 (参见例如, 国际申请号 PCT/US87/00880)、等温扩增方法 (参见例如 Walker 等人, 1992, *PNAS* 89 :382-396)、链置换扩增 (SDA)、修复链反应、不对称定量 PCR (参见例如, 美国公开号 US 2003/30134307A1) 和在 Fuja 等人, 2004, *Journal of Biotechnology* 108 :

193-205 中描述的多路微球珠测定。

#### [0166] 5.4.2 检测蛋白质的方法

[0167] 在本发明的具体实施方案中,通过检测蛋白质,例如通过检测本文描述的一种或多种基因(例如,表 1A 中列出的基因)的表达产物(例如,核酸或蛋白质),或此类蛋白质的翻译后修饰或以其他方式修饰或加工的形式,可以获得生物标记概况中的生物标记特征值。在一个具体实施方案中,使用本领域技术人员已知用于检测蛋白质的任何方法,包括但不限于蛋白质微阵列分析、免疫组织化学和质谱法,通过检测和/或分析由本文公开的基因(例如,表 1A 中列出的基因)表达的一种或多种蛋白质和/或其区别片段来生成生物标记概况。

[0168] 标准技术可以用于测定样品中存在的一种或多种目的蛋白质(例如,由表 1A 中列出的基因表达的蛋白质)的量。例如,可以采用标准技术,使用例如免疫测定例如举例来说蛋白质印迹、免疫沉淀随后为十二烷基硫酸钠凝胶电泳(SDS-PAGE)、免疫细胞化学等,以测定样品中存在的一种或多种目的蛋白质的量。用于检测目的蛋白质的一种示例性试剂是能够与目的蛋白质特异性结合的抗体,优选直接或间接地可检测标记的抗体。

[0169] 对于此类检测方法,需要时,使用本领域技术人员众所周知的技术,可以容易地分离来自待分析样品的蛋白质。蛋白质分离方法可以例如是例如 Harlow 和 Lane,1988, *Antibodies :A Laboratory Manual*, Cold Spring Harbor Laboratory Press (Cold Spring Harbor, New York) 中描述的那些。

#### [0170] 5.5 数据分析算法

[0171] 在本发明中鉴定其相应特征值能够诊断情感障碍的生物标记。这些生物标记及其相应特征(例如,表达水平)的身份可以用于开发决策规则,或多个决策规则,其区分具有情感障碍的受试者和不具有情感障碍的受试者。一旦已使用这些示例性数据分析算法或本领域已知的其他技术构建决策规则,决策规则就可以用于将测试受试者分类成 2 个或多个表型类别之一(例如,具有情感障碍,不具有情感障碍)。这通过将决策规则应用于得自测试受试者的生物标记概况来完成。因此,此类决策规则具有众多值作为诊断指示剂。

[0172] 在一个方面,本发明提供了用于评估来自测试受试者的生物标记概况与得自训练群体的生物标记概况。在某些实施方案中,得自训练群体中的受试者以及测试受试者的每个生物标记概况包含关于多个不同生物标记各自的特征。在某些实施方案中,这个比较通过下述完成:(i) 使用来自训练群体的生物标记概况开发决策规则,和(ii) 将决策规则应用于来自测试受试者的生物标记概况。像这样,在本发明的某些实施方案中应用的决策规则用于测定测试受试者是否具有情感障碍。

[0173] 在本发明的某些实施方案中,当决策规则的应用结果指出受试者具有情感障碍时,受试者诊断为“情感障碍”受试者。如果决策规则的应用结果指出受试者不具有情感障碍,那么受试者诊断为“非情感障碍”受试者。因此,在某些实施方案中,上述二元决策情况中的结果具有 4 个可能结果:

[0174] (i) 真正地具有情感障碍,其中决策规则指出受试者具有情感障碍,并且受试者事实上的确具有情感障碍(真阳性, TP);

[0175] (ii) 虚假地具有情感障碍,其中决策规则指出受试者具有情感障碍,但事实上受试者不具有情感障碍(假阳性, FP);

[0176] (iii) 真正地不具有情感障碍,其中决策规则指出受试者不具有情感障碍,并且受试者事实上不具有情感障碍(真阴性, TN);或

[0177] (iv) 虚假地不具有情感障碍,其中决策规则指出受试者不具有情感障碍,并且受试者事实上的确具有情感障碍(假阴性, FN)。

[0178] 应当理解可以作出关于 TP、FP、TN、FN 的其他定义。虽然所有此类可替代定义在本发明的范围内时,但为了容易理解本发明,由上文定义(i)到(iv)给出的关于 TP、FP、TN 和 FN 的定义将在本文中使用的,除非另有说明。

[0179] 如技术人员应当理解的,许多定量标准可以用于联系测试生物标记概况和参考生物标记概况之间作出的比较的性能(例如,决策规则应用于来自测试受试者的生物标记概况)。这些包括正面预测值(PPV)、负面预测值(NPV)、特异性、灵敏度、准确度和确实性。此外,其他构建例如接受者操作曲线(ROC)可以用于评估决策规则性能。如本文使用的:

$$[0180] \quad \text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$[0181] \quad \text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}}$$

$$[0182] \quad \text{特异性} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

$$[0183] \quad \text{灵敏度} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$[0184] \quad \text{准确度} = \text{确实性} = \frac{\text{TP} + \text{TN}}{N}$$

[0185] 此处, N 是比较样品数目(例如,测试样品数目)。例如,考虑到其中存在对于其寻求情感障碍分类的 10 个受试者的情况。对于 10 个测试受试者各自构建生物标记概况。随后,通过应用决策规则评估各个生物标记概况,其中所述决策规则基于得自训练群体的生物标记概况开发。在这个例子中,来自上述等式的 N 等于 10。一般地, N 是样品数目,其中每个样品从群体的不同成员中收集。这个群体事实上可以具有 2 个不同类型。在一个类型中,群体包含其样品和表型数据(例如,生物标记的特征值和受试者是否具有情感障碍的指示)用于构建或改善决策规则的受试者。此类群体在本文中被称为训练群体。在另一个类型中,群体包含未用于构建决策规则的受试者。此类群体在本文中被称为验证群体。除非另有说明,由 N 表示的群体并非唯一地是训练群体或并非唯一地是验证群体,与 2 个群体类型的混合物相反。应当理解与验证群体相反,当它们基于训练群体时时,得分例如准确度将更高(接近于统一)。然而,除非本文另有明确说明,用于评估决策规则(或来自测试受试者的生物标记概况的其他评估形式)性能的包括确定性(准确度)的所有标准指通过与标准对应的决策规则应用于训练群体或验证群体测量的标准。此外,上文定义的关于 PPV、NPV、特异性、灵敏度和准确度的定义还可以在 Draghici, Data Analysis Tools for DNA Microanalysis, 2003, CRC Press LLC, Boca Raton, Florida, 第 342-343 页中找到。

[0186] 在某些实施方案中, N 是超过一个、超过 5 个、超过 10 个、超过 20 个、10-100 个、超过 100 个或小于 1000 个受试者。在某些实施方案中,决策规则(或其他比较形式)可以具有针对训练群体或验证群体至少约 99% 的确定性,或甚至更多。在其他实施方案中,确定性

是针对训练群体或验证群体（并且因此针对其并非训练群体部分的单个受试者，例如临床患者）至少约 97%、至少约 95%、至少约 90%、至少约 85%、至少约 80%、至少约 75%、至少约 70%、至少约 65% 或至少约 60%。确定性的有用程度可以依赖于本发明的具体方法而改变。如本文使用的，“确定性”意指“准确度”。在一个实施方案中，灵敏度和 / 或特异性是针对训练群体或验证群体至少约 97%、至少约 95%、至少约 90%、至少约 85%、至少约 80%、至少约 75% 或至少约 70%。在某些实施方案中，此类决策规则用于以所述准确度预测受试者是否具有情感障碍。在某些实施方案中，此类决策规则用于以所述准确度诊断情感障碍。在某些实施方案中，此类决策规则用于以所述准确度测定受试者具有情感障碍症状的可能性。

[0187] 可以由决策规则以足够准确度用于分类测试受试者的特征数目是 2 个或更多个。在某些实施方案中，它是 3 个或更多个、4 个或更多个、10 个或更多个、或 10-200 个。然而，依赖于寻求的确定性程度，在决策规则中使用的特征数目可以是更多或更少，但在所有情况下是至少 2 个。在一个实施方案中，最佳化可以由决策规则用于分类测试受试者的特征数目，以允许以高确定性分类测试受试者。

[0188] 用于开发决策规则的相关数据分析算法包括但不限于判别分析包括线性、逻辑和更灵活的判别技术（参见例如，Gnanadesikan, 1977, *Methods for Statistical Data Analysis of Multivariate Observations*, New York: Wiley 1977）；基于树的算法例如分类和回归树 (CART) 和变体（参见例如，Breiman, 1984, *Classification and Regression Trees*, Belmont, California: Wadsworth International Group）；广义相加模型（参见例如，Tibshirani, 1990, *Generalized Additive Models*, London: Chapman and Hall）；和神经网络（参见例如，Neal, 1996, *Bayesian Learning for Neural Networks*, New York: Springer-Verlag；和 Insua, 1998, *Feedforward neural networks for nonparametric regression* In: *Practical Nonparametric and Semiparametric Bayesian Statistics*, 第 181-194 页, New York: Springer, 以及下文 5.5.2 节）。

[0189] 在一个实施方案中，执行测试受试者的生物标记概况与得自训练群体的生物标记概况的比较，并且包含应用决策规则。使用数据分析算法例如计算机模式识别算法构建决策规则。用于构建决策规则的其他合适数据分析算法包括但不限于检测特征值分布中的差异的逻辑回归或非参数算法（例如，威氏符号秩和检验 (Wilcoxon Signed Rank Test)（未调整和调整的））。决策规则可以基于与来自 1、2、3、4、5、10、20 个或更多个生物标记的测量可观察量对应的 2、3、4、5、10、20 个或更多个特征。在一个实施方案中，决策规则基于数百个特征或更多个。决策规则还可以使用分类树算法进行构建。例如，来自训练群体的每个生物标记概况可以包含至少 3 个特征，其中特征是分类树算法中的预测物（参见下文 5.5.1 节）。决策规则预测在群体（或类别）内的从属关系，其准确度至少约至少约 70%、至少约 75%、至少约 80%、至少约 85%、至少约 90%、至少约 95%、至少约 97%、至少约 98%、至少约 99% 或约 100%。

[0190] 合适的数据分析算法是本领域已知的，其中某些在 Hastie 等人，同上中综述。在一个具体实施方案中，本发明的数据分析算法包含分类和回归树 (CART；下文 5.5.1 节)、多重累计回归树 (MART)、用于微阵列预测分析 (PAM) 或随机森林分析（下文 5.5.1 节）。此类算法分类来自生物材料例如血样的复合光谱，以将受试者区别为正常或具有特定疾病状

态特有的生物标记表达水平。在其他实施方案中,本发明的数据分析算法包含 ANOVA 和非参数等价物、线性判别分析、逻辑回归分析、最近邻分类器分析、神经网络(下文 5.5.2 节)、主成分分析、二次判别分析、回归分类器和支持向量机(下文 5.5.4 节)、相关向量机和遗传算法(下文 5.5.5 节)。虽然此类算法可以用于构建决策规则和 / 或增加决策规则应用的速度和效率且避免研究者偏差,但本领域普通技术人员将认识到不需要基于计算机的算法来执行本发明的方法。

[0191] 决策规则可以用于评估生物标记概况,与用于生成生物标记概况的方法无关。例如,可以用于评估生物标记概况的合适决策规则使用气相层析生成,如 Harper, “Pyrolysis and GC in Polymer Analysis,” Dekker, New York (1985) 中讨论的。进一步地, Wagner 等人, 2002, *Anal. Chem.* 74 :1824-1835 公开了决策规则,其改善基于通过静态飞行时间二次离子质谱法 (TOF-SIMS) 获得的谱分类受试者的能力。此外, Bright 等人, 2002, *J. Microbiol. Methods* 48 :127-38, 公开了通过分析 MALDI-TOF-MS 谱以高确定性 (79-89% 正确分类率) 区分细菌菌株的方法。Dalluge, 2000, *Fresenius J. Anal. Chem.* 366 : 701-711, 讨论了 MALDI-TOF-MS 和液相层析 - 电喷射电离质谱法 (LC/ESI-MS) 的使用, 以分类复杂生物样品中的生物标记概况。

#### [0192] 5.5.1 决策树

[0193] 可以使用本发明中鉴定的生物标记特征值构建的一类决策规则是决策树。此处,“数据分析算法”是可以构建决策树的任何技术,而最终“决策树”是决策规则。决策树使用训练群体和特异性数据分析算法进行构建。决策树一般由 Duda, 2001, *Pattern Classification*, John Wiley&Sons, Inc., New York. 第 395-396 页描述。基于树的方法将特征空间分隔成一组矩形,并且随后在每一个中拟合模型(如常数)。

[0194] 训练群体数据包括关于本发明的生物标记跨越训练群体的特征(例如,表达值或某些其他可观察量)。可以用于构建决策树的一个具体算法是分类和回归树 (CART)。其他具体决策树算法包括但不限于 ID3、C4.5、MART 和随机森林。CART、ID3 和 C4.5 在 Duda, 2001, *Pattern Classification*, John Wiley&Sons, Inc., New York. 第 396-408 页和第 411-412 页中描述。CART、MART 和 C4.5 在 Hastie 等人, 2001, *The Elements of Statistical Learning*, Springer-Verlag, New York, 第 9 章中描述。随机森林在 Breiman, 1999, “Random Forests-Random Features,” Technical Report 567, Statistics Department, U. C. Berkeley, 1999 年 9 月中描述。

[0195] 在本发明的某些实施方案中,决策树用于分类受试者,使用用于组合本发明生物标记的特征。决策树算法属于监督学习算法类别。决策树目的是由真实世界实例数据诱导分类器(树)。这个树可以用于分类未用于驱动决策树的未见过实例。像这样,决策树衍生自训练数据。示例性训练数据含有关于多个受试者(训练群体)的数据。对于每个分别受试者,分别受试者的类别存在多个特征(例如,具有情感障碍 / 不具有情感障碍)。在本发明的一个实施方案中,训练数据是跨越训练群体用于组合生物标记的表达数据。

[0196] 一般而言,存在许多不同决策树算法,其中许多在 Duda, *Pattern Classification*, 第 2 版, 2001, John Wiley&Sons, Inc 中描述。决策树算法通常需要考虑特征处理、杂质测量、停止标准和修剪。具体决策树算法包括但不限于分类和回归树 (CART)、多变量决策树、ID3 和 C4.5。

[0197] 在一个方法中,当使用决策树时,跨越训练群体关于本发明中描述的选择基因组合的基因表达数据标准化至具有平均零和单位变异。将训练群体成员随机分成训练集和测试集。例如,在一个实施方案中,将训练群体成员的三分之二置于训练集中,并且将训练群体成员的三分之一置于测试集中。用于本发明中所述生物标记的选择组合的表达值用于构建决策树。随后,测定关于决策树正确分类测试集中的成员的能力。在某些实施方案中,对于生物标记的给定组合执行数次这种计算。在每次计算迭代中,训练群体成员随后指定至训练集和测试集中。随后,生物标记组合的性质被视为决策树计算的每次此类迭代平均值。

[0198] 除其中每次分裂基于关于相应生物标记的特征值的单变量决策树外,在本发明的生物标记集合中,或 2 个此类生物标记的相对特征值,多变量决策树可以作为决策规则实现。在此类多变量决策树中,某些或所有决策实际上包含关于本发明多个生物标记的特征值的线性组合。此类线性组合可以使用已知技术进行训练,例如对于分类的梯度下降或通过使用误差平方和标准。为了举例说明此类决策树,考虑表达式:

[0199]  $0.04x_1 + 0.16x_2 < 500$

[0200] 此处,  $x_1$  和  $x_2$  指关于来自本发明的生物标记中的 2 个不同生物标记的 2 个不同特征。为了测验 (poll) 决策规则,特征  $x_1$  和  $x_2$  的值得自从未分类受试者获得的测量。随后将这些值插入等式内。如果计算小于 500 的值,那么获得在决策树中的第一个分支。否则,获得在决策树中的第二个分支。多变量决策树在 Duda, 2001, Pattern Classification, John Wiley&Sons, Inc., New York, 第 408-409 页中描述。

[0201] 可以用于本发明中的另一个方法是多变量适应回归样条 (MARS)。MARS 是关于回归的适当操作,并且完全适合于通过本发明解决的高尺度问题。MARS 可以被视为逐步线性回归的一般化或 CART 方法的修饰,以改善 CART 在回归设置中的性能。MARS 在 Hastie 等人, 2001, The Elements of Statistical Learning, Springer-Verlag, New York, 第 283-295 页中描述。

[0202] 5.5.2 神经网络

[0203] 在某些实施方案中,对于本发明的选择生物标记测量的特征数据(例如, RT-PCR 数据、质谱法数据、微阵列数据)可以用于训练神经网络。神经网络是两阶段回归或分类决策规则。神经网络具有包括通过权层与输出单元连接的输入单元层(和偏差)的分层结构。对于回归,输出单元层一般包括仅一个输出单元。然而,神经网络可以以无缝形式处理多重定量应答。

[0204] 在多层神经网络中,存在输入单元(输入层)、隐匿单元(隐匿层)和输出单元(输出层)。此外,存在与除输入单元外的每个单元连接的单个偏差单元。神经网络在 Duda 等人, 2001, Pattern Classification, Second Edition, John Wiley&Sons, Inc., New York; 和 Hastie 等人, 2001, The Elements of Statistical Learning, Springer-Verlag, New York 中描述。神经网络还在 Draghici, 2003, Data Analysis Tools for DNA Microarrays, Chapman&Hall/CRC; 和 Mount, 2001, Bioinformatics: sequence and genome analysis, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York 中描述。下文公开的那些是神经网络的某些示例性形式。

[0205] 使用神经网络的基本方法是以未经训练的网络开始,将训练模式呈递给输入层,并且使信号经过网且测定在输出层上的输出。这些输出随后与靶值比较;任何差异与误差

对应。这种误差或标准函数是权的某些标量函数,并且当网络输出匹配所需输出时降到最低。因此,调整权以减少这个误差测量。对于回归,这种误差可以是误差平方和。对于分类,这个误差可以是平方误差或交叉熵(偏差)。参见例如,Hastie 等人,2001,The Elements of Statistical Learning, Springer-Verlag, New York。

[0206] 3种常用训练方案是随机、分批和联机。在随机训练中,从训练集中随机选择模式并且对于每个模式呈现更新网络权。通过梯度下降方法例如随机反向传播训练的多层非线性网络在由网络拓扑限定的分类器中执行权值的最大似然估计。在分批训练中,所有模式在学习发生前呈递给网络。一般地,在分批训练中,通过训练数据进行几次经过。在联机训练中,每个模式对网呈递一次并且仅一次。

[0207] 在某些实施方案中,对于关于权的起始值给予考虑。如果权接近零,那么通常在神经网络的隐匿层中使用的S字形的操作部分(参见例如,Hastie 等人,2001,The Elements of Statistical Learning, Springer-Verlag, New York)是大致线性的,并且因此神经网络崩溃成近似线性分类器。在某些实施方案中,关于权的起始值选择为接近零的随机值。因此,分类器开始接近线性,并且随着权增加变得非线性。个别单元定位于方向且在需要时引入非线性。确切零权的使用导致零衍生物和完美对称,并且算法从不移动。可替代地,以大权开始通常导致弱解决方案。

[0208] 因为输入的标定测定底层中权的有效标定,所以它可以对最终解决方案的品质具有大作用。因此,在某些实施方案中,在开始时,使所有表达值标准化以具有平均值零和标准差1。这确保所有输入在正则化过程中同样处理,并且允许选择关于随机起始权的有意义范围。对于标准化输入,一般采取在范围 $[-0.7, +0.7]$ 上的随机一致权。

[0209] 在三层网络使用中的屡发问题是在网络中使用的隐匿单元的最佳数目。三层网络的输入和输出数目由待解决的问题测定。在本发明中,关于给定神经网络的输入数目将等于选自训练群体的生物标记数目。关于神经网络的输出数目一般将是仅一个。然而,在某些实施方案中,使用超过一个输出,从而使得可以由网络限定超过仅2个陈述。例如,多输出神经网络可以用于区分健康表型、情感障碍的各个阶段。如果在神经网络中使用太多的隐匿单元,那么网络将具有太多自由度,并且训练太久,那么存在网络将过度拟合数据的危险。如果存在太少的隐匿单元,那么训练集无法得到学习。然而,一般来说,具有太多的隐匿单元比太少更好。对于太少的隐匿单元,分类器可能不具有足够的灵活性以捕获日期中的非线性;对于太多的隐匿单元,额外权可以朝向零收缩,如果如下所述使用合适正则化或修剪。在一般实施方案中,隐匿单元数目是5-100范围内的某处,其中数目随着输入数目和训练情况数目增加。

[0210] 测定使用的隐匿单元数目的一个一般方法是应用正则化方法。在正则化方法中,构建新标准函数,其不仅依赖于标准训练误差,还依赖于分类器复杂性。具体而言,新标准函数惩罚高度复杂分类器;在这个标准中搜索最小值是平衡关于训练集的误差与关于训练集加上正则化项的误差,这表示解决方案的约束条件或所需性质:

$$[0211] \quad J = J_{\text{pat}} + \lambda J_{\text{reg}}$$

[0212] 调整参数 $\lambda$ 以迫使正则化更强或更弱。换言之,关于 $\lambda$ 的较大值将趋于使权朝向零收缩:一般地与验证集的交叉验证用于估计 $\lambda$ 。这个验证集可以通过留出训练群体的随机子集获得。已提供其他罚分形式,例如权消除罚分(参见例如,Hastie 等人,2001,The Elements of Statistical Learning, Springer-Verlag, New York)。

[0213] 测定使用的隐单元数目的另一种方法是消除 - 修剪 - 最少需要的权。在一个方法中, 消除具有最小量级的权 (设为零)。此类基于量级的修剪可以起作用, 但并非最佳的; 有时具有小量级的权对于学习和训练数据是重要的。在某些实施方案中, 不是使用基于量级的修剪方法, 而是计算 Wald 统计学。在 Wald Statistics 中的基础想法是它们可以用于估计隐单元 (权) 在分类器中的重要性。随后, 消除具有最小重要性的隐单元 (通过将其输入和输出权设为零)。在这点上的 2 个算法是最佳脑损害 (Optimal Brain Damage) (OBD) 和最佳脑外科 (Optimal Brain Surgeon) (OBS) 算法, 其使用二级近似以预测训练误差如何依赖于权, 并且消除导致训练误差中的最小增加的权。

[0214] 最佳脑损害和最佳脑外科共享在权  $w$  时训练网络至局部最小误差的相同基本方法, 并且随后修剪导致训练误差中的最小增加的权。对于全权向量  $\delta w$  中的改变在误差中的预测函数增加是:

$$[0215] \quad \delta J = \left( \frac{\partial J}{\partial w} \right)^t \cdot \delta w + \frac{1}{2} \delta w^t \cdot \frac{\partial^2 J}{\partial w^2} \cdot \delta w + O(\|\delta w\|^3)$$

[0216] 其中  $\frac{\partial^2 J}{\partial w^2}$  是海森 (Hessian) 矩阵。第一个项在误差中的局部最小值时消失; 第三个和更高级的项忽略。考虑删除一个权的约束条件用于使这个函数降到最低的一般解决方案是:

$$[0217] \quad \delta w = - \frac{w_q}{[\mathbf{H}^{-1}]_{qq}} \mathbf{H}^{-1} \cdot u_q \text{ 和 } L_q = \frac{1}{2} - \frac{w_q^2}{[\mathbf{H}^{-1}]_{qq}}$$

[0218] 此处,  $u_q$  是在权空间中沿着  $q$ th 方向的单位向量, 并且  $L_q$  是对于权  $q$  的显著的近似值 - 训练误差中的增加, 如果权  $q$  被修剪, 并且其他权更新  $\delta w$ 。这些等式需要  $\mathbf{H}$  的倒数。计算这个逆矩阵的一个方法是以小值开始,  $\mathbf{H}_0^{-1} = \alpha^{-1} \mathbf{I}$ , 其中  $\alpha$  是小参数 - 有效地权常数。接下来, 矩阵根据下述用每个模式更新

$$[0219] \quad \mathbf{H}_{m+1}^{-1} = \mathbf{H}_m^{-1} - \frac{\mathbf{H}_m^{-1} \mathbf{X}_{m+1} \mathbf{X}_{m+1}^T \mathbf{H}_m^{-1}}{\frac{n}{a_m} + \mathbf{X}_{m+1}^T \mathbf{H}_m^{-1} \mathbf{X}_{m+1}} \text{ 等式 1}$$

[0220] 其中脚注与所示的模式对应, 并且  $a_m$  随着  $m$  减少。在完全训练集已显示后, 逆海森矩阵由  $\mathbf{H}^{-1} = \mathbf{H}_n^{-1}$  给出。在算法形式中, 最佳脑外科方法是:

[0221] 开始初始化  $n_q, w, \theta$

[0222] 训练合理大的网络以使误差降到最低

[0223] 通过等式 1 计算  $\mathbf{H}^{-1}$

[0224]

$$q^* \leftarrow \arg \min_q w_q^2 / (2[\mathbf{H}^{-1}]_{qq}) \quad (\text{显著 } L_q)$$

$$[0225] \quad \mathbf{w} \leftarrow \mathbf{w} - \frac{w_{q^*}}{[\mathbf{H}^{-1}]_{q^*q^*}} \mathbf{H}^{-1} e_{q^*} \quad (\text{显著 } L_q) \quad (\text{显著 } L_q)$$

[0226] 直至  $J(w) > \theta$

[0227] 回到  $w$

[0228] 结束

[0229] 最佳脑损害方法在计算上更简单,因为在行 3 中的逆海森矩阵计算对于对角矩阵特别简单。当误差大于初始化为  $\theta$  的标准时,上述算法终止。另一种方法是当  $J(w)$  中的改变由于权的消除大于某些标准值时,改变行 6 以终止。在某些实施方案中,反向传播神经网络,参见例如,Abdi,1994,“A neural network primer,” J. Biol System. 2, 247-283。

[0230] 5.5.3 聚类

[0231] 在某些实施方案中,关于本发明的选择生物标记的特征用于聚类训练集。例如,考虑其中使用本发明中描述的 10 个特征(与 10 个生物标记对应)的情况。训练群体中的每个成员  $m$  将对于 10 个生物标记各自具有特征值(例如表达值)。来自训练群体中的成员  $m$  的此类值限定向量:

[0232]  $X_{1m} X_{2m} X_{3m} X_{4m} X_{5m} X_{6m} X_{7m} X_{8m} X_{9m} X_{10m}$

[0233] 其中  $X_{im}$  是生物  $m$  中的第  $i$  个生物标记的表达水平。如果训练集中存在  $m$  个生物,那么  $i$  个生物标记的选择将限定  $m$  个向量。应当指出本发明的方法不需要向量中使用的每个单个生物标记的表达值各自在每个单个向量  $m$  中表示。换言之,来自其中未发现第  $i$  个生物标记之一的受试者的数据仍用于聚类。在此类情况下,失去的表达值指定“零”或某些其他标准化值。在某些实施方案中,在聚类前,使特征值标准化,以具有平均值零和单元变异。

[0234] 跨越训练组显示相似表达模式的训练群体的那些成员将趋于聚类在一起。本发明基因的特定组合在本发明的这个方面被视为良好分类器,当向量聚类成训练群体中发现的特性组内。例如,如果训练群体包含类别  $a$ :在研究下不具有情感障碍的受试者,和类别  $b$ :在研究下具有情感障碍的受试者,理想聚类分类器将使群体聚类为 2 个组,其中一个聚类组独特地代表类别  $a$ ,并且另一个聚类组独特地代表类别  $b$ 。

[0235] 聚类在 Duda 和 Hart, Pattern Classification and Scene Analysis, 1973, John Wiley&Sons, Inc., New York, (下文“Duda 1973”) 的第 211-256 页上描述。如 Duda 1973 的 6.7 节中所述,聚类问题描述为在数据集中的发现天然分组之一。为了鉴定天然分组,解决 2 个问题。首先,测定测量 2 个样品之间的相似性(或不同性)的方法。这个度量(相似性量度)用于确保一个聚类中的样品比它们与其他聚类中的样品更彼此相似。其次,测定使用相似性测量用于将数据分隔成聚类的机制。

[0236] 相似性量度在 Duda 1973 的 6.7 节中讨论,其中陈述开始聚类研究的一个方法是限定距离函数且计算数据集中的所有样品对之间的距离矩阵。如果距离是相似性的良好测量,那么相同簇中样品之间的距离将明显小于不同聚类中样品之间的距离。然而,如 Duda 1973 的第 215 页上所述,聚类不需要使用距离度量。例如,非度量相似性函数  $s(x, x')$  可以用于比较 2 个向量  $x$  和  $x'$ 。常规地,  $s(x, x')$  是对称函数,当  $x$  和  $x'$  由于某种原因“相似”时,其值大。非度量相似性函数  $s(x, x')$  的例子在 Duda 1973 的第 216 页上提供。

[0237] 一旦已选择用于测量数据集中的点之间的“相似性”或“不同性”的方法,聚类就需要测量数据的任何分隔的聚类性质的标准函数。使标准函数 extremize 的数据集分隔用于聚类数据。参见 Duda 1973 的第 217 页。标准函数在 Duda 1973 的 6.8 节中讨论。

[0238] 最近以来,已出版了Duda等人,Pattern Classification,第2版,John Wiley&Sons, Inc. New York。第537-563页详细描述了聚类。关于聚类技术的更多信息可以在下述中发现:Kaufman和Rousseeuw,1990,Finding Groups in Data:An Introduction to Cluster Analysis, Wiley, New York, NY;Everitt,1993,Cluster analysis(第3版), Wiley, New York, NY;和Backer,1995,Computer-Assisted Reasoning in Cluster Analysis,Prentice Hall,Upper Saddle River,New Jersey。可以在本发明中使用的具体示例性聚类技术包括但不限于层次聚类(使用最近邻算法、最远邻算法、平均连接算法、质心算法或平方和算法的凝聚聚类)、k-均值聚类、模糊k均值聚类算法和Jarvis-Patrick聚类。

#### [0239] 5.5.4 支持向量机

[0240] 在本发明的某些实施方案中,支持向量机(SVMs)用于分类受试者,使用本发明中描述的基因的特征值。SVMs是相对新类型的学习算法。参见例如,Cristianini和Shawe-Taylor,2000,An Introduction to Support Vector Machines, Cambridge University Press, Cambridge;Boser等人,1992,“A training algorithm for optimal margin classifiers,”in Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory, ACM Press, Pittsburgh, PA,第142-152页;Vapnik,1998,Statistical Learning Theory, Wiley, New York;Mount,2001,Bioinformatics: sequence and genome analysis, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York,Duda,Pattern Classification,第2版,2001, John Wiley&Sons, Inc.;和Hastie,2001,The Elements of Statistical Learning, Springer, New York;和Furey等人,2000,Bioinformatics 16,906-914。当用于分类时,SVMs以超平面分离二元标记数据训练数据的给定集合,所述超平面最大限度距离其。对于其中非线性分离可能的情况,SVMs可以与‘内核(kernels)’技术组合起作用,其自动实现对特征空间的非线性映射。在特征空间中由SVM发现的超平面与输入空间中的非线性决策边界对应。

[0241] 在一个方法中,当使用SVM时,使特征数据标准化,以具有平均值零和单元变异,并且将训练群体成员随机分成训练集和测试集。例如,在一个实施方案中,将训练群体成员的三分之二置于训练集中,并且将训练群体成员的三分之一置于测试集中。关于本发明中描述的基因组合的表达值用于训练SVM。随后测定关于训练SVM正确分类训练集中的成员的能力。在某些实施方案中,对于分子标记的给定组合执行数次这种计算。在每次计算迭代中,训练群体的成员随机指定至训练集和测试集。随后,生物标记组合的性质被视为SVM计算的每次此类迭代平均值。

#### [0242] 5.5.5. 相关向量机和遗传算法

[0243] 相关向量机(RVM)是在回归以及监督多类分类问题中可用的基于内核的贝叶斯(Bayesian)统计模型(Tipping,M;Sparse Bayesian Learning and the Relevance Vector Machine, Journal of Machine Learning Research 1,2001,211-244)。作为分类工具有用,训练RVM作出关于新数据点的类别从属关系的概率预测。在RVM模型中,假定说明变量(即,基因或生物标记)的预定集合通过逻辑连接函数影响类别从属关系概率。为了测定选自许多候选变量的说明变量最佳集合,RVM模型在遗传最佳化算法内操作(Deb, K: Multi-Objective Optimization using Evolutionary Algorithms, Wiley,2001),这评估

对候选变量的不同子集训练且测试的大量 RVMs。每个变量子集的性能通过交叉验证进行评估。

#### [0244] 5.5.6 其他数据分析算法

[0245] 上文描述的数据分析算法仅仅是可以用于构建决策规则用于区别转换器与非转换器的方法类型的例子。此外,可以使用上文描述的技术的组合。某些组合例如决策树和促进的组合的使用已得到描述。然而,许多其他组合也是可能的。此外,在本领域中的其他技术中,例如投影追踪 (Projection Pursuit) 和加权表决 (Weighted Voting),可以用于构建决策规则。

#### [0246] 5.6 生物标记

[0247] 在一个具体实施方案中,生物标记概况包含表 1A 中列出的至少 2 个不同生物标记。生物标记概况进一步包含关于至少 2 个生物标记的分别相应特征。此类生物标记可以是例如 mRNA 转录物、cDNA 或某些其他核酸例如扩增核酸、或蛋白质。一般地,至少 2 个生物标记衍生自至少 2 个不同基因。在其中至少 2 个不同生物标记中的生物标记在表 1A 中列出的情况下,生物标记可以是例如由列出基因制备的转录物、其互补物、或其区别片段或互补物、或其 cDNA、或 cDNA 的区别片段、或与全部或部分转录物或其互补物对应的区别扩增核酸分子、或由基因编码的蛋白质、或蛋白质的区别片段、或上文任何的指示。依照此类实施方案,本发明的生物标记概况可以使用本领域技术人员已知的任何标准测定或在本文描述的测定中获得,以检测生物标记。此类测定能够例如检测特定基因或目的基因的等位基因(例如,表 1A 中公开的基因)的表达产物(例如,核酸和/或蛋白质)。在一个实施方案中,此类测定利用核酸微阵列。

[0248] 在某些实施方案中,生物标记概况具有表 1A 中列出的 2-29 个生物标记。在某些实施方案中,生物标记概况具有表 1A 中列出的 3-20 个生物标记。在某些实施方案中,生物标记概况具有表 1A 中列出的 4-15 个生物标记。在某些实施方案中,生物标记概况具有表 1A 中列出的至少 2 个生物标记。在某些实施方案中,生物标记概况具有表 1A 中列出的至少 3 个生物标记。在某些实施方案中,生物标记概况具有表 1A 中列出的至少 4 个生物标记。在某些实施方案中,生物标记概况具有表 1A 中列出的至少 2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25 个或更多个生物标记。在某些实施方案中,每个此类生物标记是核酸。在某些实施方案中,每个此类生物标记是蛋白质。在某些实施方案中,生物标记概况中的某些生物标记是核酸,并且生物标记概况中的某些生物标记是蛋白质。

#### [0249] 5.7 具体实施方案

[0250] 本发明的一个方面涉及鉴定可能显示出情感障碍症状的受试者的基因转录概况的方法。此类基因转录概况基于来自受试者生物样品的选择基因的转录分析,此类基因选自表 1A。

[0251] 使用本发明,可以鉴定且分析可以聚集成单个概况的个别生物标记丰度(例如表达水平)。此类丰度概况用作用于疾病分类的标记。如下所述,完成转录分析以测定对照受试者和患病受试者的全血样品中的基因表达概况。选自表 1A 的基因丰度在表 4、表 5 和表 6 中例示。表 4、表 5 和表 6 各自是分别关于抑郁受试者、严重抑郁受试者和双相型受试者的基因转录概况的代表性例子,与对照比较。在一个实施方案中,如表 4 中所示具有抑郁症基因转录概况的受试者诊断为具有抑郁症。在另一个实施方案中,如表 5 中所示具有严重

抑郁症基因转录概况的受试者诊断为具有严重抑郁症。在另一个实施方案中,如表 6 中所示具有双相型基因转录概况的受试者诊断为具有双相型障碍。基因转录概况的进一步代表性例子显示于表 4A 和 5B 中。

[0252] 在一个例子中,用于测定基因表达概况的生物标记选自表 1A 中描述的基因。代表性转录生物标记特征集合也在表 1A 中描述。探针集合通过众所周知的方法用于执行定量 PCR(qPCR)。

[0253] 本发明的一个方面提供了如通过选自表 1A 的基因转录分析测定的对于每个受试者的转录概况。

[0254] 通过本领域众所周知的方法可以执行转录分析。例如, RNA 包括信使 RNA (mRNA) 可以从动物体特别是人体的细胞材料或含有细胞材料的液体中分离。应当理解细胞材料含有细胞内容物包括 mRNA。在本发明中使用的生物样品可以例如选自外周组织、全血、脑脊髓液、腹膜液和间隙液。

[0255] 在本发明的其他实施方案中,生物样品选自全血、脑脊髓液和外周组织。本发明还可以使用选自红血细胞、白血细胞和血小板的全血级分 (fractions) 进行。白血细胞 (白细胞) 包括但不限于:嗜中性粒细胞、嗜碱性粒细胞、嗜酸性粒细胞、淋巴细胞、巨噬细胞和单核细胞。

[0256] 为了测量样品中的基因表达,可以对那个样品中的 RNA 或 mRNA 实施逆转录以产生拷贝 DNA,并且随后通过标准方法使用探针或引物序列基于 DNA 序列进行分析。每个单个基因可以通过聚合酶链反应 (PCR)、定量 PCR、原位杂交、RNA 印迹分析、固体载体固定测定例如基于珠的测定或基因治疗和本领域众所周知的其他方法进行分析。

[0257] 依照本文描述的本发明的一个方面,定量 PCR(qPCR) 用于测量 mRNA 水平。一个或多个核酸探针用于测量来自生物样品的 mRNA 水平。探针或引物是与目的基因互补的核苷酸 (nt) 序列,并且通过技术人员众所周知的方法完成此类探针 / 引物的选择和合成。本发明的探针 / 引物并不限于表 1A 中描述的核苷酸序列。

[0258] 本发明进一步提供了对患病受试者分类的方法,其为如从得自受试者的生物样品分析的,通过测定此类受试者的转录概况,与对照受试者比较来进行。

[0259] 本发明提供了通过选自表 1A 的基因的转录分析测定的独特转录概况。如果它测定为类似于已知健康对照受试者或已知患病受试者的转录概况,则此类转录概况测定为在受试者中是明显的 (distinct)。与已知健康对照受试者或已知患者受试者的转录概况的相似性通过分类方法进行测定,例如分类算法,如本文描述的。

[0260] 在某些实施方案中,如本文描述的,从多个对照受试者中收集转录数据。如本文描述的,从患有疾病或病症例如情感障碍的多个受试者中收集转录数据。数据分析算法与作为输入的每个转录数据集合一起使用,以便区分或区别每个转录数据集中含有的分类基因。此类算法一般描述为分类算法,也称为“分类器”。用于执行这个任务的数据分析算法是本领域技术人员众所周知的,并且可以使用下述例子:随机森林 (Breiman, L., 2001, Machine Learning 45(1):5-32)、支持向量机 (SVM) (Cortes, C. 和 Vapnik, V. 1995, Machine Learning, 20(3):273-97)、逐步逻辑回归 (SLR) (Ersbøll, B. K. 和 Conradsen, K. (2005) An Introduction to Statistics. 第 7 版 IMM; Draper, N. 和 Smith, H. (1981) Applied Regression Analysis, 第 2 版, New York: John Wiley & Sons, Inc.)、递归分

隔 (RPART) (James K. E. 等人, 2005, *Statistics in Medicine*, 24(19) :3019-35)、惩罚逻辑回归分析 (Penalized Logistic Regression Analysis) (PELORA) (Dettling, M., 2003, *Proceedings of the 3<sup>rd</sup> International Workshop on Distributed Statistical Computing*, March 20-22, Vienna Austria, Hornick, Leisch 和 Seilis 编)、神经网络、相关向量机 (RVM), LogitBoost (Friedman, J., Hastie, T. 和 Tibshirani, R. 2000, *Annals of Statistics* 28(2) :337-407)、微阵列预测分析 (PAM) 及其他 (参见 V. N. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998)。调整或训练此类分类算法或“分类器”, 以提供基于其转录数据的关于患者分类的输出。

[0261] 分类通过训练分类算法选择的基因或生物标记获得与特定数据集所属类别相关的转录数据的预测测量, 例如与对照数据相关的类别或与疾病数据相关的类别。

[0262] 虽然不希望由任何具体理论束缚, 但随机森林算法被视为总体学习方法, 这基于来自许多决策树的输出使受试者分类。每个决策树对可获得数据的自举 (bootstrap) 样品训练, 并且决策树中的每个结节通过最佳说明变量 (即基因或生物标记) 分裂。随机森林可以提供自动化变量选择, 并且描述选择变量之间的非线性相互作用。

[0263] 逐步逻辑回归 (SLR) 被视为统计模型, 这通过使数据输入与逻辑曲线拟合预测事件发生概率。在逻辑模型中, 假定说明变量 (即基因或生物标记) 的预定集合通过逻辑连接函数影响概率。为了测定选自许多候选变量的说明变量的最佳集合, 以逐步方式由内部模型构建大量逻辑回归模型, 且通过评估 Akaike Information Criteria (AIC) 比较以便测定最准确的模型 (Burnham, K. P. 和 D. R. Anderson, 2002. *Model Selection and Multimodel Inference : A Practical-Theoretic Approach*, 第 2 版 Springer-Verlag)。

[0264] 支持向量机 (SVMs) 被视为属于广义线性分类器家族。考虑到 2 组分类中的输入数据作为在  $n$  维空间中的 2 个向量集合, SVM 通过超平面分离数据, 这使 2 个向量集合之间的边缘达到最大。采取最小距离以使超平面达到最大的向量被称为支持向量。SVM 不提供自动变量 (即基因或生物标记) 选择。

[0265] 相关向量机 (RVMs) 假定说明变量 (即基因或生物标记) 的预定集合通过对数连接函数影响类别从属关系概率。RVMs 寻求测定选自许多候选变量的说明变量的最佳集合。RVM 可以用遗传最佳化算法操作, 这评估且交叉验证许多 RVMs 且选择候选变量 (即基因或生物标记) 的最佳集合。

[0266] 使用上述数据分析算法之一进一步训练用分类算法构建的转录概况。分类误差是关于其训练分类算法预测类别内的从属关系的准确度量度。分类误差可以通过交叉验证方法进行测定, 例如留一法交叉验证 (LOOCV)、 $K$ -折验证、或 10 折验证 (Devijver, P. A. 和 J. Kittler, 1982, *Pattern Recognition : A Statistical Approach*, Prentice-Hall, London)。

[0267] 算法对于指定转录概况的准确度可以通过测定由那个算法在训练过程中预测的真阳性 (TP)、真阴性 (TN)、假阳性 (FP) 和假阴性 (FN) 数目进行测量。准确度测量为 :

[0268] 准确度 =  $(TP+TN) / TP+TN+FP+FN$

[0269] 阳性预测值 (PPV) 或已通过算法阳性评分的患病受试者百分比测量为 :

[0270]  $PPV = TP / TP+FP$

[0271] 阴性预测值 (NPV) 或对照受试者 (其不具有疾病) 且已通过算法阴性评分的百分

比测量为：

[0272]  $NPV = TN / (TN + FN)$

[0273] 分类算法的性能还通过 Jaccard 相似系数 (Jaccard 指数) 进行测定, 这评估分类已如何良好地鉴定正确变量 (即基因)。训练分类算法的准确度可以大于约 60%、65%、70%、75%、80%、85%、90% 或 95%。训练分类算法的 Jaccard 指数可以大于约 60%、65%、70%、75%、80%、85%、90% 或 95%。训练分类算法的 PPV 和 NPV 可以大于约 60%、65%、70%、75%、80%、85%、90% 或 95%。

[0274] 受试者的分类可以用于诊断具有情感障碍或可能显示出情感障碍症状的受试者。用于分类受试者的基因转录概况基于表 1A 中的基因的转录分析。如通过本文描述的方法分析的受试者的转录概况将指示受试者是否属于患病受试者类别。

[0275] 在某些实施方案中, 本发明提供了诊断测试受试者中的情感障碍的方法, 该方法包括评估测试受试者的生物标记概况中的多个生物标记的多个特征满足值集合, 其中满足值集合预测测试受试者具有所述情感障碍, 并且其中多个特征是多个生物标记的可测量方面, 多个生物标记包含表 1A 中列出的至少 2 个生物标记。该方法进一步包括将测试受试者是否具有情感障碍的诊断输出至用户界面装置、监测器、可触摸计算机可读存储介质、或局部或远程计算机系统; 或以用户可读形式显示测试受试者是否具有情感障碍的诊断。

[0276] 在本发明的某些实施方案中, 多个生物标记由表 1A 中列出的 2-29 个生物标记组成。在其他实施方案中, 多个生物标记由表 1A 中列出的 3-20 个生物标记组成。在另外其他的实施方案中, 多个生物标记包含表 1A 中列出的至少 2、3、4 或 5 个生物标记。

[0277] 在某些实施方案中, 多个特征由对应于表 1A 中列出的 2-29 个生物标记的 2-29 个特征组成。在其他实施方案中, 多个特征由对应于表 1A 中列出的 3-15 个生物标记的 3-15 个特征组成。在另外其他的实施方案中, 多个生物标记包含对应于表 1A 中列出的至少 2 个生物标记的至少 2 个特征。

[0278] 在其他实施方案中, 多个生物标记包含 ERK1 和 MAPK14。在其他实施方案中, 多个生物标记包含 Gi2 和 IL-1b。在其他实施方案中, 多个生物标记包含 ARRB1 和 MAPK14。在其他实施方案中, 多个生物标记包含 ERK1 和 IL1b。

[0279] 在本发明的某些方面, 所述多个生物标记中的每个生物标记是核酸。在其他方面, 所述多个生物标记中的每个生物标记是 DNA、cDNA、扩增的 DNA、RNA 或 mRNA。在另外其他方面, 所述多个生物标记中的每个生物标记是蛋白质。

[0280] 在其他实施方案中, 测试受试者的生物标记概况中的所述多个生物特征中的特征是多个生物标记中的生物标记的可测量方面, 并且关于所述特征的特征值使用得自所述测试受试者的生物样品进行测定。在其他实施方案中, 特征是所述生物标记在生物样品中的丰度。在另外其他的实施方案中, 生物样品是外周组织、全血、脑脊髓液、腹膜液、间隙液、红血细胞、白血细胞或血小板。

[0281] 在另一个实施方案中, 所述多个特征中的特征是所述生物标记概况中的生物标记的可测量方面, 并且关于所述特征的特征值使用得自所述测试受试者的样品进行测定。在某些实施方案中, 生物标记概况中的生物标记是核酸的指示或蛋白质的指示。在其他实施方案中, 生物标记概况中的生物标记是 mRNA 分子的指示或 cDNA 分子的指示。在某些实施方案中, mRNA 分子或 cDNA 分子的指示是转录物值例如拷贝 /ng cDNA。在其他实施方案中,

生物标记概况中的第一个生物标记是核酸的指示,并且生物标记概况中的第二个生物标记是蛋白质的指示。

[0282] 在本发明的某些方面,值集合包含如表 4 中所示的生物标记丰度,并且满足表 4 的值集合预测受试者具有抑郁症。在其他方面,值集合包含如表 5 中所示的生物标记丰度,并且满足表 5 的值集合预测受试者具有严重抑郁症。在其他方面,值集合包含如表 6 中所示的生物标记丰度,并且满足表 6 的值集合预测受试者具有双相型障碍。进一步地,本发明提供了如表 4A 中用于诊断抑郁症的值集合和如表 5B 中用于诊断严重抑郁症的值集合。

[0283] 表 4、5 和 6 中所述的值集合由以拷贝 /ng cDNA 的生物标记丰度代表,即生物标记基因的转录物。例如,在表 4 中关于生物标记 ARRBI 的关于抑郁受试者的转录物值范围是  $189062 \pm 62727$  个拷贝 /ngcDNA,这等价于 126335–251789 个拷贝 /ng cDNA 的范围。在表 4 中关于生物标记 CD8a 的关于抑郁受试者的转录物值范围是  $8304 \pm 5825$  个拷贝 /ng cDNA,这等价于 2479–14129 个拷贝 /ng cDNA 的范围。在本发明的某些方面,满足值集合意指对于每个生物标记具有在给定范围内的值。

[0284] 在某些实施方案中,包含在 15148–35504 个拷贝 /ng cDNA 范围内的 ERK1 丰度和在 39241–107071 个拷贝 /ng cDNA 范围内的 MAPK14 丰度的值集合预测受试者具有抑郁症。在其他实施方案中,包含在 61734–168500 个拷贝 /ng cDNA 范围内的 Gi2 丰度和在 15939–43323 个拷贝 /ng cDNA 范围内的 IL1b 丰度的值集合预测受试者具有抑郁症。在其他实施方案中,包含在 126335–251789 个拷贝 /ng cDNA 范围内的 ARRBI 丰度和在 39241–107071 个拷贝 /ng cDNA 范围内的 MAPK14 丰度的值集合预测受试者具有抑郁症。在其他实施方案中,包含在 15148–35504 个拷贝 /ng cDNA 范围内的 ERK1 丰度和在 15939–43323 个拷贝 /ng cDNA 范围内的 IL1b 丰度的值集合预测受试者具有抑郁症。

[0285] 在其他实施方案中,包含 ERK1 丰度除以 MAPK14 丰度比值在 0.25–0.45 范围内的值集合预测受试者具有抑郁症。在其他实施方案中,包含 Gi2 丰度除以 IL1b 丰度比值在 0.16–0.36 范围内的值集合预测受试者具有抑郁症。在其他实施方案中,包含 MAPK14 丰度除以 ARRBI 丰度比值在 0.29–0.49 范围内的值集合预测受试者具有抑郁症。在其他实施方案中,包含 ERK1 丰度除以 IL1b 丰度比值在 0.075–0.95 范围内的值集合预测受试者具有抑郁症。

[0286] 在其他实施方案中,包含 ERK1 丰度除以 MAPK14 丰度比值在 0.19–0.39 范围内的值集合预测受试者具有严重抑郁症。在其他实施方案中,包含 Gi2 丰度除以 IL1b 丰度比值在 0.18–0.38 范围内的值集合预测受试者具有严重抑郁症。在其他实施方案中,包含 MAPK14 丰度除以 ARRBI 丰度比值在 0.32–0.52 范围内的值集合预测受试者具有严重抑郁症。在其他实施方案中,包含 ERK1 丰度除以 IL1b 丰度比值在 0.60–0.80 范围内的值集合预测受试者具有严重抑郁症。

[0287] 在上述方法的其他方面,该方法进一步包括在评估步骤前构建所述生物标记概况。在其他实施方案中,构建步骤包含从所述测试受试者的生物样品中获得所述多个特征。在某些方面,通过用第一个生物标记的特征值除以第二个生物标记的特征值,通过测定生物标记丰度比值来构建生物标记概况。此类生物标记概况可以使用表 4、表 5 或表 6 中所示的值进行构建。

[0288] 在其他实施方案中,样品是外周组织、全血、脑脊髓液、腹膜液、间隙液、红血细胞、

白血细胞或血小板。

[0289] 在上述方法的另外其他方面,该方法进一步包括在评估步骤前构建所述第一个值集合。在其他实施方案中,构建步骤包含将数据分析算法应用于得自群体成员的特征。

[0290] 在某些方面,特征是包含 ERK1 和 MAPK14 的生物标记的可测量方面,并且特征值使用得自所述测试受试者的血样进行测定。

[0291] 在其他实施方案中,群体包含来自不具有情感障碍的第一批多个对照受试者的第一批多个生物样品,和来自具有情感障碍的第二批多个受试者的第二批多个生物样品。在另外其他的实施方案中,数据分析算法是决策树、微阵列预测分析、多重累计回归树、神经网络、聚类算法、主成分分析、最近邻分析、线性判别分析、二次判别分析、支持向量机、进化法、相关向量机、遗传算法、投影追踪或加权表决。

[0292] 在另一个实施方案中,构建步骤生成决策规则,并且其中所述评估步骤包含将所述决策规则应用于多个特征,以便测定它们是否满足第一个值集合。在某些实施方案中,决策规则将所述群体中的受试者分类为 (i) 不具有情感障碍的受试者和 (ii) 的确具有情感障碍的受试者,其准确度 70% 或更大。在其他实施方案中,决策规则将所述群体中的受试者分类为 (i) 不具有情感障碍的受试者和 (ii) 的确具有情感障碍的受试者,其准确度 90% 或更大。

[0293] 在本发明的特定方面,情感障碍是双相型障碍 I、双相型障碍 II、心境恶劣障碍或抑郁障碍。在其他方面,情感障碍是轻度抑郁症、中度抑郁症、重度抑郁症、非典型抑郁症、忧郁抑郁症或边缘型人格障碍。在另外其他方面,情感障碍是 (i) 创伤后应激障碍或 (ii) 无创伤后应激障碍的创伤。在某些方面,情感障碍是急性创伤后应激障碍或复发性创伤后应激障碍。

[0294] 本发明提供了用于诊断测试受试者中的情感障碍的试剂盒,所述试剂盒包含试剂和说明书用于评估测试受试者的生物标记概况中的多个生物标记的多个特征是否满足值集合,其中满足值集合预测测试受试者具有所述情感障碍,并且其中多个特征是多个生物标记的可测量方面,多个生物标记包含表 1A 中列出的至少 2 个生物标记。在某些方面,试剂包含识别选自表 1A 的生物标记的核苷酸序列的探针和 / 或引物。本发明的试剂盒用于生成根据本发明的生物标记概况。在某些方面,本发明的试剂盒提供了用于测试且评估来自多个生物标记的测试受试者生物标记概况的说明书,所述多个生物标记包含表 1A 中列出的至少 2 个生物标记。在其他方面,本发明的试剂盒提供了含有值集合的说明书,以便测定测试受试者的生物标记概况是否满足此类值集合。

[0295] 本发明还提供了计算机程序产品,其中所述计算机程序产品包含计算机可读存储介质和其中嵌入的计算机程序机制,所述计算机程序机制包含用于执行上述方法中的任何的指令。在某些实施方案中,计算机程序机制进一步包含指令用于将测试受试者是否具有情感障碍的诊断输出至用户界面装置、监测器、可触摸计算机可读存储介质、或局部或远程计算机系统;或以用户可读形式显示测试受试者是否具有情感障碍的诊断。

[0296] 本发明还提供了包含下述的计算机:一个或多个处理器;与一个或多个处理器连接的存储器,所述存储器存储用于执行上述方法中的任何的指令。在本发明的某些方面,存储器进一步包含指令用于将测试受试者是否具有情感障碍的诊断输出至用户界面装置、监测器、可触摸计算机可读存储介质、或局部或远程计算机系统;或以用户可读形式显示测试

受试者是否具有情感障碍的诊断。

[0297] 本发明进一步提供了测定测试受试者显示出情感障碍症状的可能性的方法,该方法包括:评估测试受试者的生物标记概况中的多个生物标记的多个特征是否满足值集合,其中满足该值集合提供了测试受试者显示出情感障碍症状的所述可能性,并且其中多个特征是多个生物标记的可测量方面,所述多个生物标记包含表 1A 中列出的至少 2 个生物标记。

[0298] 在某些实施方案中,多个生物标记包含 ERK1 和 MAPK14。在其他实施方案中,多个生物标记包含 Gi2 和 IL-1b。在其他实施方案中,多个生物标记包含 ARRB1 和 MAPK14。在其他实施方案中,多个生物标记包含 ERK1 和 IL1b。

[0299] 在本发明的某些实施方案中,多个生物标记包含 ERK1、PBR 和 MAPK14。在另一个实施方案中,多个生物标记包含 PBR、Gi2 和 IL1b。在其他实施方案中,多个生物标记包含 ERK1、ARRB1 和 MAPK14。在某些实施方案中,多个生物标记包含 MAPK14、ERK1 和 CD8b。在其他实施方案中,多个生物标记包含 MAPK14、ERK1 和 P2X7。在另外其他的实施方案中,多个生物标记包含 ARRB1、IL6 和 CD8a。在特定实施方案中,多个生物标记包含 ARRB1、ODC1 和 P2X7。

[0300] 在另外其他的实施方案中,该方法进一步包括将测试受试者显示出情感障碍症状的可能性输出至用户界面装置、监测器、可触摸计算机可读存储介质、或局部或远程计算机系统;或以用户可读形式显示测试受试者显示出情感障碍症状的可能性。

[0301] 本发明提供了其为关于从多个对照受试者中收集的每种生物样品的转录分析测量的转录概况。本发明提供了其为关于从多个抑郁、严重抑郁或双相型受试者中收集的每种生物样品的转录分析测量的转录概况。本发明进一步提供了其为关于从多个边缘型人格障碍受试者中收集的每种生物样品的转录分析测量的转录概况。本发明提供了其为关于从多个 PTSD 受试者中收集的每种生物样品的转录分析测量的转录概况。

[0302] 本发明还提供了包含第一批多个对照受试者的集体测量的转录概况存储在例如数据库中。使用数据分析算法特别是训练分类算法,使包含第二批多个受试者例如患病受试者的集体测量的转录概况与第一批多个对照受试者的转录概况比较。训练分类算法使每个受试者集合分类。训练分类算法提供了用于诊断且指定分类的预测值。训练分类算法提供了用于预测受试者将显示出病症症状的可能性的预测值。

[0303] 本发明的另一个实施方案涉及诊断或预测受试者对疾病或病症的易感性,或预测显示出病症症状的可能性,基于受试者的独特转录概况与健康对照受试者和患病受试者的转录概况比较。用于诊断用途的基因转录概况基于选自表 1A 的基因的转录分析。

[0304] 本发明的一个方面涉及诊断不同类型的情感障碍,特别是主要性抑郁障碍、双相型障碍、边缘型人格障碍和创伤后应激障碍。

[0305] 本发明的另一个方面涉及通过鉴定转录概况区别患者群体。例如,通常将诊断为主要性抑郁症的患者可以通过经由转录概况分成抑郁症亚型,例如忧郁和非典型抑郁症。存在关于这些抑郁症亚型的差别治疗应答证据。显示出共病 (co-morbidity) 即符合关于超过一个病症的 DSM-IV<sup>®</sup>标准的患者,将获益于转录概况的鉴定。转录概况可以鉴定关于一个病症的共同生物学基础。

[0306] 通过上述方法,在一个实施方案中,本发明提供了其为关于从多个健康对照受试

者中收集的生物样品的转录分析测量的转录概况。本发明还提供了其为关于从多个情感障碍受试者中收集的生物样品的转录分析测量的转录概况。例如,本发明还提供了其为关于从多个抑郁、严重抑郁或双相型受试者中收集的生物样品的转录分析测量的转录概况。本发明提供了如表 4 中其为关于从多个抑郁受试者中收集的生物样品的转录分析测量的转录概况。本发明提供了如表 5 中其为关于从多个严重抑郁受试者中收集的生物样品的转录分析测量的转录概况。本发明提供了如表 6 中其为关于从多个双相型受试者中收集的生物样品的转录分析测量的转录概况。本发明进一步提供了其为关于从多个边缘型人格障碍受试者中收集的生物样品的转录分析测量的转录概况。本发明还提供了其为关于从多个 PTSD 受试者中收集的生物样品的转录分析测量的转录概况。

[0307] 在本发明的一个实施方案中,生物样品是全血。

[0308] 本发明还提供了包含第一批多个对照受试者的集体测量的转录概况存储于例如数据库中。使用分类算法,使包含第二批多个受试者例如患病受试者的集体测量的转录概况与第一批多个对照受试者的转录概况比较。分类算法提供了使受试者各自分类的输出。

[0309] 在本发明的某些方面,转录概况由选自下述的基因的转录分析进行测定:ADA、ARRB1、ARRB2、CD8a、CD8b、CREB1、CREB2、DPP4、ERK1、ERK2、Gi2、Gs、GR、IL1b、IL6、IL8、INDO、MAPK14、MAPK8、MKP1、MR、ODC1、P2X7、PBR、PREP、RGS2、S100A10、SERT 和 VMAT2。

[0310] 在另一个实施方案中,转录概况由选自下述的至少 3 种基因的转录分析进行测定:ADA、ARRB1、ARRB2、CD8a、CD8b、CREB1、CREB2、DPP4、ERK1、ERK2、Gi2、Gs、GR、IL1b、IL6、IL8、INDO、MAPK14、MAPK8、MKP1、MR、ODC1、P2X7、PBR、PREP、RGS2、S100A10、SERT 和 VMAT2。

[0311] 在某些实施方案中,转录概况由选自下述的基因的转录分析进行测定:ARRB1、ARRB2、CD8a、CREB1、CREB2、ERK2、Gi2、MAPK14、ODC1、P2X7 和 PBR。

[0312] 在另一个实施方案中,转录概况由选自 CD8a、ERK1、MAPK14、P2X7 和 PBR 的基因的转录分析进行测定。

[0313] 在另一个实施方案中,转录概况由选自 Gi2、GR 和 MAPK14 的基因的转录分析进行测定。

[0314] 在另一个实施方案中,转录概况由选自 Gi2、GR、MAPK14 和 MR 的基因的转录分析进行测定。

[0315] 在另一个实施方案中,转录概况由选自下述的基因的转录分析进行测定:ARRB1、ARRB2、CD8b、ERK2、IDO、IL-6、MR、ODC1、PREP 和 RGS2。

[0316] 在另一个实施方案中,转录概况由选自下述的基因的转录分析进行测定:ARRB1、CREB1、ERK2、Gs、IL-6、MKP1 和 RGS2。

[0317] 在另一个实施方案中,转录概况由选自 ERK1 和 MAPK14 的基因的转录分析进行测定。在另一个实施方案中,转录概况由选自 Gi2 和 IL1b 的基因的转录分析进行测定。在另一个实施方案中,转录概况由选自 ARRB1 和 MAPK14 的基因的转录分析进行测定。在另一个实施方案中,转录概况由选自 ERK1 和 IL1b 的基因的转录分析进行测定。

[0318] 在另一个实施方案中,转录概况由选自 ERK1、MAPK14 和 P2X7 的基因的转录分析进行测定。在另一个实施方案中,转录概况由选自 Gi2、IL1b 和 PBR 的基因的转录分析进行测定。在另一个实施方案中,转录概况由选自 ARRB1、ODC1 和 P2X7 的基因的转录分析进行测定。在另一个实施方案中,转录概况由选自 ARRB1、CD8a 和 IL6 的基因的转录分析进行测

定。在另一个实施方案中,转录概况由选自 CD8b、ERK1 和 MAPK14 的基因的转录分析进行测定。在另一个实施方案中,转录概况由选自 ARRB1、ERK1 和 MAPK14 的基因的转录分析进行测定。在另一个实施方案中,转录概况由选自 ERK1、MAPK14 和 PBR 的基因的转录分析进行测定。

[0319] 本发明的一个方面提供了用于诊断受试者中的情感障碍的方法,其包括鉴定受试者中的转录概况,并且使此类转录概况与对照受试者或健康对照受试者组的概况比较,从而基于转录概况中的改变或差异的存在或不存在诊断受试者是否显示出情感障碍。

[0320] 在本发明的某些实施方案中,情感障碍选自抑郁症、严重抑郁症、双相型障碍、边缘型人格障碍。在某些实施方案中,情感障碍选自创伤后应激障碍或无创伤后应激障碍的创伤。在其他实施方案中,情感障碍选自急性创伤后应激障碍或复发性创伤后应激障碍。

[0321] 本发明的一个方面提供了用于诊断受试者是否显示出情感障碍的方法,其包括:

[0322] (a) 从怀疑具有情感障碍的受试者中获得生物样品;

[0323] (b) 测量生物样品中的 mRNA 水平,其中所述 mRNA 水平是选自下述的基因的 mRNA 水平:ADA、ARRB1、ARRB2、CD8a、CD8b、CREB1、CREB2、DPP4、ERK1、ERK2、Gi2、Gs、GR、IL1b、IL6、IL8、INDO、MAPK14、MAPK8、MKP1、MR、ODC1、P2X7、PBR、PREP、RGS2、S100A10、SERT 和 VMAT2;

[0324] (c) 收集 mRNA 水平且将其作为 mRNA 数据存储于计算机介质中;

[0325] (d) 经由分类算法加工此类 mRNA 数据,由此所述加工测定了 mRNA 数据与来自健康对照受试者的 mRNA 数据是相同还是不同;和

[0326] (e) 提供使受试者分类的输出数据,

[0327] 从而诊断受试者是否显示出情感障碍。

[0328] 本发明进一步提供了用于预测受试者对情感障碍的易感性的方法,通过使受试者选自下述的基因的转录概况与多个健康对照受试者的所述基因的转录概况比较来进行:ADA、ARRB1、ARRB2、CD8a、CD8b、CREB1、CREB2、DPP4、ERK1、ERK2、Gi2、Gs、GR、IL1b、IL6、IL8、INDO、MAPK14、MAPK8、MKP1、MR、ODC1、P2X7、PBR、PREP、RGS2、S100A10、SERT 和 VMAT2。

[0329] 本发明的一个方面提供了用于预测受试者显示出情感障碍症状的可能性的方法,其包括:

[0330] (a) 从受试者中获得生物样品;

[0331] (b) 测量 mRNA 水平,其中所述 mRNA 水平是选自下述的基因的 mRNA 水平:ADA、ARRB1、ARRB2、CD8a、CD8b、CREB1、CREB2、DPP4、ERK1、ERK2、Gi2、Gs、GR、IL1b、IL6、IL8、INDO、MAPK14、MAPK8、MKP1、MR、ODC1、P2X7、PBR、PREP、RGS2、S100A10、SERT 和 VMAT2;

[0332] (c) 收集 mRNA 水平且将其作为 mRNA 数据存储于计算机介质中;

[0333] (d) 经由分类算法加工此类 mRNA 数据,由此所述加工测定了 mRNA 数据与来自健康对照受试者的 mRNA 数据是相同还是不同;和

[0334] (e) 提供使受试者分类的输出数据,

[0335] 从而预测受试者显示出情感障碍症状的可能性。

[0336] 在另一个实施方案中,该方法可以包含测量选自下述的至少 2 种基因的 mRNA 水平:ADA、ARRB1、ARRB2、CD8a、CD8b、CREB1、CREB2、DPP4、ERK1、ERK2、Gi2、Gs、GR、IL1b、IL6、IL8、INDO、MAPK14、MAPK8、MKP1、MR、ODC1、P2X7、PBR、PREP、RGS2、S100A10、SERT 和 VMAT2。

[0337] 在其他实施方案中,该方法包括测量表 1A 中列出的任何 3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27 或 28 种基因的 mRNA 水平。

[0338] 在其他实施方案中,该方法包括测量选自下述的基因的 mRNA 水平:ARRB1、ARRB2、CD8a、CREB1、CREB2、ERK2、Gi2、MAPK14、ODC1、P2X7 和 PBR。

[0339] 在另一个实施方案中,该方法包括测量选自 CD8a、ERK1、MAPK14、P2X7 和 PBR 的基因的 mRNA 水平。

[0340] 在另一个实施方案中,该方法包括测量选自 Gi2、GR 和 MAPK14 的基因的 mRNA 水平。

[0341] 在另一个实施方案中,该方法包括测量选自 Gi2、GR、MAPK14 和 MR 的基因的 mRNA 水平。

[0342] 在另一个实施方案中,该方法包括测量选自下述的基因的 mRNA 水平:ARRB1、ARRB2、CD8b、ERK2、IDO、IL-6、MR、ODC1、PREP 和 RGS2。

[0343] 在另一个实施方案中,该方法包括测量选自下述的基因的 mRNA 水平:ARRB1、CREB1、ERK2、Gs、IL-6、MKP1 和 RGS2。

[0344] 在另一个实施方案中,该方法包括测量选自 ERK1 和 MAPK14 的基因的 mRNA 水平。在另一个实施方案中,该方法包括测量选自 Gi2 和 IL1b 的基因的 mRNA 水平。在另一个实施方案中,该方法包括测量选自 ARRB1 和 MAPK14 的基因的 mRNA 水平。在另一个实施方案中,该方法包括测量选自 ERK1 和 IL1b 的基因的 mRNA 水平。

[0345] 在另一个实施方案中,该方法包括测量选自 ERK1、MAPK14 和 P2X7 的基因的 mRNA 水平。在另一个实施方案中,该方法包括测量选自 Gi2、IL1b 和 PBR 的基因的 mRNA 水平。在另一个实施方案中,该方法包括测量选自 ARRB1、ODC1 和 P2X7 的基因的 mRNA 水平。在另一个实施方案中,该方法包括测量选自 ARRB1、CD8a 和 IL6 的基因的 mRNA 水平。在另一个实施方案中,该方法包括测量选自 CD8b、ERK1 和 MAPK14 的基因的 mRNA 水平。在另一个实施方案中,该方法包括测量选自 ARRB1、ERK1 和 MAPK14 的基因的 mRNA 水平。在另一个实施方案中,该方法包括测量选自 ERK1、MAPK14 和 PBR 的基因的 mRNA 水平。

[0346] 在本发明的某些实施方案中,情感障碍选自抑郁症、严重抑郁症、双相型障碍、边缘型人格障碍。在某些实施方案中,情感障碍选自创伤后应激障碍或无创伤后应激障碍的创伤(trauma without post traumatic stress disorder)。在其他实施方案中,情感障碍选自急性创伤后应激障碍或复发性创伤后应激障碍。

[0347] 在某些实施方案中,上述方法是计算机辅助方法。

[0348] 5.7 情感障碍

[0349] 本文描述的精神病或精神障碍及其临床表现是从业精神病医师已知的。每种病症的特定症状可以由大多数精神病医师识别。

[0350] 由美国精神病协会(American Psychiatric Association)出版的 Diagnostic and Statistical Manual of Mental Disorders,第 5 版,Text Revision(DSM-IV-TR<sup>®</sup>) (1994 年 10 月,2000 年 5 月文本修订)是在美国由医生用于精神障碍临床分类的标准。关于精神/精神病学病症的症状学和诊断标准在 DSM-IV-TR<sup>®</sup>指导中阐述。

[0351] 5.7.1 抑郁障碍

[0352] DSM-IV-TR<sup>®</sup>列出了用于抑郁症和主要性抑郁障碍(MDD)的具体诊断标准。

[0353] DSM-IV-TR<sup>®</sup>限定了作为综合征的主要性抑郁发作,其中在相同 2 周时间中,下述症状中的至少 5 个呈现且其自身表现为良好功能的先前状态的改变(此外,症状必须包括(1)或(2)):

[0354] 1. 情绪低落

[0355] 2. 兴趣或愉快感缺失

[0356] 3. 明显的重量减轻或增加

[0357] 4. 失眠或睡眠过度

[0358] 5. 精神性运动激动或阻滞

[0359] 6. 疲倦或精力缺乏

[0360] 7. 无价值感

[0361] 8. 思考或集中能力下降;犹豫不决

[0362] 9. 反复思考死亡、自杀观念、自杀企图或关于自杀的特定计划 DSM-IV-TR<sup>®</sup>进一步包括了必须存在于抑郁症的各种亚型中的症状描述。抑郁症可以注意到具有或不具有精神病症状,并且可以具有忧郁或紧张性特征或被分类为非典型抑郁症。

[0363] 依赖于由患者显示出的症状数目和严重性,抑郁发作可以指定为轻度、中度或重度。临床医生还可以测定患者是否患有典型(忧郁)、非典型、紧张或精神病性抑郁症。

[0364] 临床上,抑郁症被视为非常异质的疾病。抑郁患者的基因表达概况可以反映这个异质性。基于本发明,基于基因表达概况可能更好地确定抑郁症的这些亚型,以便更好地分类或诊断患者。随后,药物的开发和施用可以适合于患有抑郁症亚型的患者。

[0365] 通过获得且分析来自对照的病历和症状信息,基因表达概况还用于预测受试者显示出本文描述病症症状的可能性。

[0366] 抑郁障碍、双相型障碍和心境恶劣障碍被视为情绪障碍范畴的部分。

[0367] 本发明提供了指示抑郁障碍例如轻度、中度或重度抑郁症的转录概况的客观测量。本发明还提供了用于分类抑郁障碍亚型的转录概况。本发明进一步提供了用于诊断具有抑郁障碍例如轻度、中度或重度抑郁症的受试者的方法。

[0368] 5.7.2 双相型障碍

[0369] 如对于抑郁症描述的,双相型障碍(BD)是异质疾病并且分成亚类或亚型,包括双相型障碍 I、双相型障碍 II 和躁郁症。双相型障碍也称为躁狂抑郁病,是引起个人的情绪、精力和活动能力中的罕见转移的脑病症。不同于所有个体经历的正常“起伏”,双相型障碍的症状是严重的,并且可以导致关系损害、工作或学习表现差甚至自杀。

[0370] BD 表现为一般在个人生命中复发的躁狂和抑郁的间歇发作。在发作之间,具有双相型障碍的大多数人是无症状的,或可能具有某些残留症状。抑郁发作是通常存在的,并且可能是主要或严重的。躁狂发作的特征在于症状例如明显的情绪紊乱,这足以引起工作中的损伤或对患者或其他人的危险,并且不是物质滥用或医学病状的结果,睡眠需要减少,过度说话或强制言语,和/或思维奔逸或观念奔驰和更多,如根据 DSM-IV-TR<sup>®</sup>描述的。

[0371] 本发明提供了用于诊断具有双相型障碍的受试者的方法。BD 患者将获益于指示双相型障碍的转录概况的客观测量。

[0372] 5.7.3 边缘型人格障碍

[0373] 边缘型人格障碍(BPD)包含自我表象、人际关系和情感的不稳定性模式,具有明

显冲动性。这种不稳定性通常破坏家庭和工作生活以及个体的自我认同。

[0374] DSM-IV-TR®如由下述至少 5 个指出的表征 BPD：

[0375] 1. 特征在于过度理想化和贬值 (devaluation) 的极端之间交替的不稳定和强烈人际关系模式。

[0376] 2. 在潜在自我伤害的至少 2 个区域中的冲动性, 例如开销、性、物质使用、入店行窃、鲁莽驾驶或暴饮暴食。

[0377] 3. 由于情绪显著反应性的情感不稳定性。

[0378] 4. 不适当的、暴怒或怒气缺乏控制, 例如频繁展示脾气、不断怒气或反复搏斗。

[0379] 5. 反复自杀威胁、姿势或行为或自残行为。

[0380] 6. 认同紊乱 ; 显著和持续的不稳定自我表象。

[0381] 7. 长期的空虚感或厌倦。

[0382] 8. 避免实际或想象放弃的狂乱努力。

[0383] 9. 短暂的、压力相关的偏执观念或严重分裂症状。

[0384] 具有 BPD 的患者是在心理疗法中可见的最有挑战性且抗拒治疗的患者中的一种。

[0385] 本发明提供了用于诊断具有 BPD 的受试者的方法。BPD 患者将获益于指示边缘型人格障碍的转录概况的客观测量。

[0386] 5.7.4 创伤后应激障碍 (PTSD)

[0387] DSM-IV-TR®将创伤后应激障碍描述为在暴露于极度创伤应激物后特征性症状的发展, 涉及事件的直接个人经历, 所述事件涉及实际或威胁死亡或严重损伤。个人可能已目睹涉及死亡、损伤或对另一个人的身体完整性危险的事件。个人对事件的应答涉及强烈害怕、无助或恐惧。个人可能具有事件的持续回忆, 包括影像、思维或感觉, 或可能具有事件的反复痛苦梦见。

[0388] 本发明提供了用于诊断具有急性 PTSD、复发性 PTSD 或无 PTSD 创伤的受试者的方法。患者 / 受试者将获益于指示急性 PTSD、复发性 PTSD 或无 PTSD 创伤的转录概况的客观测量。

[0389] 基于通过上述方法鉴定的转录概况, 可以测定、区别和 / 或区分正常或健康受试者和患有情感障碍的受试者。例如, 本发明将通过下文的实验细节得到更好理解。技术人员将容易理解本文讨论的具体方法和结果仅举例说明本发明, 如在其后跟随的权利要求中更全面描述的。6 实验细节

[0390] 总 RNA 分离。将人血收集到 PAXgene™ blood RNA 管 (PreAnalytiX, Hombrechtikon, CH) 内, 通过倒转数次混合, 并且贮存于 -20 或 -80°C, 直至加工用于 RNA 分离。加工通过使样品在室温下温育过夜开始, 随后在 3000xG 下离心 10 分钟。弃去上清液, 并且使团块重悬浮于 5ml 水中, 随后为另一个离心步骤。洗涤和离心步骤重复第二次, 并且使团块重悬浮于管中其余的残留水 (约 100ul)。向这个溶液中加入 941 μl Ambion ToTALLY RNA™ 裂解 / 变性溶液 (Ambion, Austin, TX) 和 59 μl 3M 乙酸钠, pH 5.5 (Ambion), 随后混合。在室温下温育 15 分钟, 加入 770 μl 酸性苯酚 / 氯仿 (Ambion), 并且通过涡旋使管混合。将溶液转移至 2ml 塑料螺旋盖试管并且在室温下温育 5 分钟。使苯酚提取物在微量离心机中以全速旋转 1 分钟 (约 13,000x G), 并且取出水层 (1100 μl) 至含有 550 μl 100% 乙醇的新管。在混合后, 将溶液应用于 Ambion RNAqueous® -96Automated Kit 滤板的一

个孔,并且遵循制造商的方案纯化 RNA。在 RNA 洗脱后,用 DNA 酶 I (Invitrogen, Carlsbad, CA) 将样品处理第二次,以去除残留基因组 DNA。使 RNA 在 1x DNA 酶消化缓冲液中温育,加上 3 个单位酶在室温下 1 小时。通过加入 EDTA 至 13mM 终浓度灭活酶,随后在 68°C 下加热 10 分钟。通过经过 MultiScreen<sup>®</sup> PCR<sub>micro96</sub> 板 (Millipore, Billerica, MA) 使混合物脱盐,并且在 50 μl 水中洗脱。在 Agilent 2100Bioanalyzer (Agilent, Waldbronn, Germany) 上分析 RNA 的 1 μl 等分试样,并且将其余部分贮存于 -80°C。使用由 Bioanalyzer 软件计算的 RIN 值评估 RNA 样品的质量。

#### [0391] cDNA 合成

[0392] 通过使约 1 μg 总 RNA 与 1.5 μl 随机六聚物 (Invitrogen, 500ng/μl) 以 16.5 μl 的终体积混合完成 cDNA 的合成。在 75°C 10 分钟和在 25°C 10 分钟温育后,加入 6 μl 第一链缓冲液 (Invitrogen)、1.5 μl 10mM dNTPs (Invitrogen, 每种 dNTP 10mM)、1.25 μl Superscript II<sup>™</sup> (Invitrogen, 200 单位/μl) 和 4 μl 水。最终反应体积是 30 μl,并且温育在 25°C 下执行 10 分钟,42°C 1 小时和 95°C 10 分钟。使反应冷却至 4°C 直至加入 70 μl 水,随后用 MultiScreen<sup>®</sup> PCR<sub>micro96</sub> 板纯化。用 100 μl 水执行 cDNA 的洗脱,并且使所得到的材料贮存于 -20°C 直至定量。在某些情况下,使 cDNA 反应体积加倍以增加材料的得率。

#### [0393] cDNA 的定量

[0394] 染料嵌入测定用于测定 cDNA 得率。使 5 μl cDNA 与 7 μl 0.5N NaOH、50mM EDTA 在 47 μl 的终体积中混合。使混合物在 65°C 下温育 1 小时,以水解 RNA,并且随后通过加入 10 μl 1M Tris, pH7 中和。根据制造商的说明书,使用 Quant-it<sup>™</sup> Oligreen<sup>®</sup> ssDNA 试剂 (Invitrogen) 测量在水解反应的 25 μl 等分试样中的 cDNA 浓度。使未知样品与使用已知浓度的单链 DNA 生成的标准曲线比较。使用 Fusion<sup>™</sup> α 仪器 (Packard, Meridan, CT) 进行所有荧光读数。对于每个未知 cDNA 样品求得自两份重复水解反应的值的平均值。如果两份重复不在彼此的 15% 内,那么运行第三个样品,与前 2 次测定比较,并且求 2 个最相似值的平均值。

#### [0395] 定量聚合酶链反应 (qPCR)

[0396] 使用表 1A 和 1B 中所示的引物/探针组,在 Applied Biosystems 7900HT Fast Real Time PCR System (Applied Biosystems, Foster City, CA) 或 MX3000P<sup>®</sup> (Stratagene, La Jolla, CA) 上执行所有 qPCR 运行。所有探针用 FAM<sup>™</sup> (Applied Biosystems, Norwalk, CT) 在 5' 末端上和 BHQ-1<sup>®</sup> 猝灭剂在 3' 末端上进行标记,并且通过 Biosearch (Novato, CA) 进行合成。检查每个引物/探针组以确保 PCR 扩增的效率在测定的表达范围上是约 100%。构建含有来自每个人供体的 1ng 或 10ng cDNA/孔的复制平板 (96 孔形式)。板还含有 2 个阴性对照孔 (“NTC”, 仅水) 和 3 个衍生自 10 个个体血液的集合的、商购的 cDNA 孔 (参考 cDNA)。每个 qPCR 反应是 25 μl (终体积) 且含有下述组分: 12.5 μl Brilliant QPCR Master Mix<sup>®</sup> (Stratagene)、400nM 正向引物、400nM 反向引物、50nM 探针和 60nM/300nM ROX<sup>™</sup> (Applied Biosystems, Norwalk, CT) (MX3000P<sup>®</sup> 7900HT 仪器)。循环条件是 95°C, 10 分钟,随后为 40 个循环,每个循环为 95°C, 15 秒和 60°C, 1 分钟。对于每种基因执行两份重复的 qPCR 运行。罕见地,当关于基因的复制板不充分一致时,运行第三块 qPCR 板。依赖于获得的 Ct 值,求来自所有 3 块板的值的平均值或从进一步分析中排除多余板。

[0397] 用于 qPCR 运行的仪器指示预备数据分析步骤。然而,在每种情况下,目的是将接

近扩增曲线中点的扩增阈值设定为用于给定板上的所有样品的相同阈值。对于相同基因的两个重复板运行, 阈值是相似的, 尽管不一定等同。对于 MX3000P<sup>®</sup>, 下述设置用于最初测定阈值: 平滑参数 = 5, 基线计算采用 MX4000 算法, 和使用循环 6 到 14 的基于本底阈值, 具有  $\sigma$  乘数 20。需要时, 手工进行阈值的微小调整, 以将其大致置于扩增曲线图的中间。对于在 7900HT 上运行的板, 仪器的缺省设置用于最初设定阈值。需要时, 其后进行手工调整。

[0398]

表 1A: 关于选择基因/生物标记的引物/探针序列。

基因名称	缩写	基因登记号 (SEQ ID NO:)	代表性引物/探针序列 (5'至 3') <sup>†</sup>
腺苷脱氧酶	ADA	NM_000022 (SEQ ID NO:88)	F = GGTGGTGGAGCTGTGTAAGAAAGTAC (SEQ ID NO:1) R = CTTCTGGGATGGTCTCATCTC (SEQ ID NO:2) P = CAGCAGACCCGTGGTAGCCATTGACCT (SEQ ID NO:3)
$\beta$ -抑制蛋白 1	ARRB1	L04685 (SEQ ID NO:89)	F = AGACACGAACTTGGCCTCTAGC (SEQ ID NO:4) R = TTGTAGGAAACAATGATCCCCAG (SEQ ID NO:5) P = TTGAGGGAAGGTGCCAACCGTGAGAT (SEQ ID NO:6)
$\beta$ -抑制蛋白 2	ARRB2	BC007427 (SEQ ID NO:90)	F = TCTTCCATGCTCCGTCACAC (SEQ ID NO:7) R = CGAATCTCAAAGTCTACGCCG (SEQ ID NO:8) P = AGCCAGGCCAGAGGATACAGGAAA (SEQ ID NO:9)
CD8 $\alpha$	CD8a	M12824 (SEQ ID NO:91)	F = TTCCGCCGAGAGAACGAG (SEQ ID NO:10) R = AAGACCCGGCACGAAAGTGG (SEQ ID NO:11) P = TCGGCCCTGAGCAACTCCATCATGTA (SEQ ID NO:12)
CD8 $\beta$	CD8b	M37601 (SEQ ID NO:92)	F = TGACAGTCACCACGAGTTCCTG (SEQ ID NO:13) R = TCTCCTGTCCACCCTTCCACC (SEQ ID NO:14) P = CTCTGGGATTCCCGCAAAAGGGACTAT (SEQ ID NO:15)
cAMP 应答元件 结合蛋白 1	CREB1	NM_134442 (SEQ ID NO:93)	F = CTGGCTAACAAATGGTACCGATG (SEQ ID NO:16) R = GTGGTCTGTGCATACTGTAGAATGG (SEQ ID NO:17) P = CATGACCAATGCAGCAGCCACTCA (SEQ ID NO:18)

[0399]

[0400]

基因名称	缩写	基因登记号 (SEQ ID NO:)	代表性引物/探针序列 (5'至3') <sup>†</sup>
cAMP 应答元件 结合蛋白 2	CREB2	M86842 (SEQ ID NO:94)	F = CACGTTGGATGACACTTGTGATC (SEQ ID NO:19) R = CTGGGAGATGGCCAAATTGG (SEQ ID NO:20) P = ACTAATAAGCAGCCCCCCCCAGACGGT (SEQ ID NO:21)
二肽基肽酶 IV	DPP4	M74777 (SEQ ID NO:95)	F = GTGTCATTCAGTAAAGAGGCGAAG (SEQ ID NO:22) R = CTCAGCCCTTTATCATTCACGC (SEQ ID NO:23) P = TTCCGGTCCCTGGTCTGCCCCCTCTATA (SEQ ID NO:24)
细胞外信号调节 激酶 1	ERK1	M84490 (SEQ ID NO:96)	F = TGACGGAGTATGTGGTACGC (SEQ ID NO:25) R = CCACAGACCAGATGTCGATGG (SEQ ID NO:26) P = CTGGTACCGGGCCCCAGAGATCAT (SEQ ID NO:27)
细胞外信号调节 激酶 2	ERK2	M84489 (SEQ ID NO:97)	F = TAACGTTCTGCACCGTGACC (SEQ ID NO:28) R = CAGGCCAAAGTCACAGATCTTG (SEQ ID NO:29) P = ACCTGTCTGCTCAACACCACCTGTGAT (SEQ ID NO:30)
鸟嘌呤核苷酸结 合蛋白 $\alpha$ i2	Gi2	X04828 (SEQ ID NO:98)	F = AGGCGTGCTCCCTGATGAC (SEQ ID NO:31) R = GCTCCAGGTCGTTTCAGGTAGTAG (SEQ ID NO:32) P = AGGCCTGCTTTGGCCCGCTCAA (SEQ ID NO:33)
鸟嘌呤核苷酸结 合蛋白 $\alpha$ s (长)	Gs	AF493897 (SEQ ID NO:99)	F = GACTATGTCCGAGCGGATCAG (SEQ ID NO:34) R = GTCCACCTGGAACCTTGGTCTCA (SEQ ID NO:35) P = CTGCTTCGCTGCCGTGTCCTGA (SEQ ID NO:36)
$\alpha$ -糖皮质激素受 体	GR	X03225 (SEQ ID NO:100)	F = TCCCTGGTCGAAACAGTTTTTTC (SEQ ID NO:37) R = TTTGGGAGGTGGTCTCTGTTG (SEQ ID NO:38) P = TGTAAGCTCTCCTCCATCCAGCTCCTCAA (SEQ ID NO:39)

[0401]

基因名称	缩写	基因登记号 (SEQ ID NO:)	代表性引物/探针序列 (5'至3') <sup>†</sup>
白介素 1, $\beta$	IL1b	NM_000576 (SEQ ID NO:101)	F = GATGGCCCTAAACAGATGAAGTG (SEQ ID NO:40) R = CCTGAAGCCCTTGCTGTAGTG (SEQ ID NO:41) P = ATGGCGGCATCCAGCTACGAATCTC (SEQ ID NO:42)
白介素 6	IL6	M14584 (SEQ ID NO:102)	F = AGCCACTCACCTCTTCAGAACG (SEQ ID NO:43) R = CATGTCCTCTTCTCAGGGCTG (SEQ ID NO:44) P = CAAATTTCGGTACATCCTCGACGGCAT (SEQ ID NO:45)
白介素 8	IL8	M28130 (SEQ ID NO:103)	F = CTGCTAGCCAGGATCCACAAG (SEQ ID NO:46) R = CTGTGAGGTAAGATGGTGGCTAATAC (SEQ ID NO:47) P = CTTGTTCCACTGTGCCTTGGTTTCTCCTT (SEQ ID NO:48)
吲哚胺-吡咯 2,3 双加氧酶	INDO	NM_002164 (SEQ ID NO:104)	F = GCTTCGAGAAAGAGTTGAGAAAGTTAAAC (SEQ ID NO:49) R = GACCTTTGCCCCACACATATG (SEQ ID NO:50) P = CTCACAGACCACAAGTCACAGCGCCTT (SEQ ID NO:51)
p38 丝裂原活化 蛋白激酶 14	MAPK14	L35253 (SEQ ID NO:105)	F = CGGCAGGAGCTGAACAAGAC (SEQ ID NO:52) R = AGCAGCACACACAGAGCCATAG (SEQ ID NO:53) P = CCGAGCGTTACCAGAACCTGTCTCCA (SEQ ID NO:54)
丝裂原活化蛋白 激酶 8	MAPK8	AY893269 (SEQ ID NO:106)	F = CCAACACCCGTACATCAATGTC (SEQ ID NO:55) R = CACTCTTCTATTGTGTGTTCCCTTTC (SEQ ID NO:56) P = CACCACCAAAGATCCCTGACACAGCAGTT (SEQ ID NO:57)
map 激酶磷酸酶 1	MKP1	X68277 (SEQ ID NO:107)	F = GCCAGGCAGGCATTTC (SEQ ID NO:58) R = ATGCTTCGCCTCTGCTTCAC (SEQ ID NO:59) P = TCAGCCACCAATCTGCCTTGCTTACCCTT (SEQ ID NO:60)

基因名称	缩写	基因登记号 (SEQ ID NO:)	代表性引物/探针序列 (5'至3') <sup>†</sup>
盐皮质激素受体	MR	M16801 (SEQ ID NO:108)	F = AGCCCAGAGGAAGGACAAC (SEQ ID NO:61) R = TGTGAGCGCTCGTGAGATTG (SEQ ID NO:62) P = CTCCTGCAAAAAGAACCCTCGGTCAACA (SEQ ID NO:63)
鸟氨酸脱羧酶 1	ODC1	NM_002539 (SEQ ID NO:109)	F = CCATGTAGGAAGCGGCTGTAC (SEQ ID NO:64) R = TCAGCCCCCATGTCAAAAAC (SEQ ID NO:65) P = ATCCTGAGACCTTCGTGCAAGCAATCT (SEQ ID NO:66)
嘌呤能受体 P2X7	P2X7	NM_002562 (SEQ ID NO:110)	F = GCTGTCGCTCCCATATTATCC (SEQ ID NO:67) R = CACAATGGACTCGCACTTCTTC (SEQ ID NO:68) P = CTGTCAGCCCTGTGTGGTCAACGAATAC (SEQ ID NO:69)
苯并二氮草 (benzodiazepine)受体 (外周型)	PBR	BC001110 (SEQ ID NO:111)	F = CTGGTCTGGAAAGAGCTGGG (SEQ ID NO:70) R = CAGCAGGAGATCCACCAAGG (SEQ ID NO:71) P = CCCCATCTTCTTTGGTGCCCGAC (SEQ ID NO:72)
脯氨酸内酰胺酶	PREP	D21102 (SEQ ID NO:112)	F = GGGAAATATGACTACTGACC AATG (SEQ ID NO:73) R = GGATCCCTGAAGTCAAATGTTGATC (SEQ ID NO:74) P = CATTCAAGACGAATCGCCAGTCTCCC (SEQ ID NO:75)
G 蛋白信号调节子 2	RGS2	NM_002923 (SEQ ID NO:113)	F = GATTGGAAGACCCGTTTGAGC (SEQ ID NO:76) R = CAGGAGAAGGCTTGATGAAAGC (SEQ ID NO:77) P = CTGGGAAGCCCCAAACC GGCAA (SEQ ID NO:78)
S100 钙结合蛋白 A10 (p11)	S100A10	NM_002966 (SEQ ID NO:114)	F = AGGAGTTCCTGGATTTTGG (SEQ ID NO:79) R = GCCCACTTTGCCATCTACAC (SEQ ID NO:80) P = CAAAAGACCCTCTGGCTGTGGACAAA (SEQ ID NO:81)

[0402]

基因名称	缩写	基因登记号 (SEQ ID NO:)	代表性引物/探针序列 (5'至3') <sup>†</sup>
血清素转运蛋白	SERT	NM_001045 (SEQ ID NO:115)	F = CATGGCTGAGATGAGGAATGAAG (SEQ ID NO:82) R = GCTGGCATGTTGGCTATCG (SEQ ID NO:83) P = ACCGAGGTCCCAGCCTCCTCTTCAT (SEQ ID NO:84)
囊泡单胺转运蛋白 2	VMAT2	L23205 (SEQ ID NO:116)	F = TGGATTGTCATCAATGATGCCCTATC (SEQ ID NO:85) R = ATGCCACATCCGCAATGG (SEQ ID NO:86) P = AGACCTGCGGCACGTTGTCCTCCTA (SEQ ID NO:87)

<sup>†</sup> F = 正向引物序列; R = 反向引物序列; P = 探针序列

[0403] 基因表达的标准化

[0404] 为了有效比较不同样品之间的基因表达概况, 优选控制可以掩蔽任何潜在生物学改变的变量。例如, 在酶促反应效率、仪器性能和吸取中的每日差异将全部影响在给定

日时获得的信号。使这些变量的影响降到最低的优选方法是通过使用多重标准化基因 (Andersen, C.L. 等人, *Cancer Res*, 2004, 64 :5245-5250 ;Jin, P. 等人, *BMC Genomics*, 2004, 5 :55 ;Huggett, J. 等人, *Genes and Immunity*, 2005, 6 :279-284)。理想标准化基因以方便测量的水平表达, 并且通过其为实验设计部分的操作是不改变的。尽管标准化基因的使用是平常的, 但研究者通常未验证他们使用的基因是否在其表达系统中稳定表达。为了避免这个问题, 使用商购可得的软件程序 GeNorm™ (PrimerDesign Ltd., Southampton, UK)。该方法基于由 Vandesompele, J. 等人, *Genome Biol*, 2002, 3(7) : RESEARCH0034.1-0034.11 (Epub 2002年6月18日) 公开的工作, 并且允许测定候选标准化基因是否稳定表达。为了选择标准化基因, 首先扫描文献以鉴定先前已由研究者用于标准化人中的基因表达的基因, 重点在于用血样进行的实验 (Vandesompele, J. 等人 *Genome Biol*, Epub June 18, 2002, 3(7) :RESEARCH0034.1-0034.11, 特别是在第 0034.5 页, 表 3 ; Applied Biosystems Application Note 2006, 公开 127AP08-01, 特别是在第 3 页, 图 1)。从这个搜索中, 鉴定了表 1B 中显示的基因。为了证实这些基因对于本文实验中的标准化有效, 用衍生自不同实验集合的血样用 Genorm™ 分析 7 种基因的表达概况, 包括正常受试者、无药物治疗的抑郁患者和用药物治疗的抑郁患者。在所有集合中, 7 种基因的组合达到良好标准化, 如通过 0.15 或更小的配对变异值 (V) 测定的 (Vandesompele, J. 等人, *Genome Biol*, Epub June 18, 2002, 3(7) :RESEARCH0034.1-0034.11)。

[0405] 尽管 Genorm™ 说明仅需要使用 2 或 3 种最好的基因用于标准化, 但由于几个原因应考虑超过 3 种标准化基因的组合。首先, 使用更多的标准化基因将帮助预测, 考虑到新药物治疗、遗传背景或疾病状态可能影响标准化基因的表达。超过 3 种标准化基因预期通过阻止在特定实验中未稳定表达的任何基因的影响来改善该过程。此外, 通过一致地使用超过 3 种基因以使表达数据标准化, 表达结果可以与随着时间过去进行的所有研究比较。因为临床样品不一直与合适对照匹配, 超过 3 种标准化基因的使用是重要考虑。虽然当跨越不同实验比较基因表达时, 用超过 3 种基因的标准化是优选方法, 但在任何特定实验内使用 2 种或 3 种基因也是有效的, 条件是待比较的所有样品以相同方式进行处理。

[0406] 表 1B : 标准化基因。

[0407]

基因名称	缩写	基因登记号 (SEQ ID NO:)
$\beta$ -肌动蛋白	ACTB	NM_001101 (SEQ ID NO:117)
$\beta$ -2-微球蛋白	B2M	NM_004048 (SEQ ID NO:118)
3-磷酸甘油醛脱氢酶	GAPD	NM_002046 (SEQ ID NO:119)
葡糖醛酸糖苷酶, $\beta$	GUSB	NM_000181 (SEQ ID NO:120)
羟甲基胆色烷合酶	HMBS	NM_000190 (SEQ ID NO:121)
次黄嘌呤磷酸核糖基转移酶I	HPRT1	NM_000194 (SEQ ID NO:122)
磷酸甘油酸酯激酶	PGK1	NM_000291 (SEQ ID NO:123)
肽脯氨酰异构酶A (亲环蛋白A)	PPIA	NM_021130 (SEQ ID NO:124)
核糖体蛋白, 大, P0	RPLP0	NM_001002 (SEQ ID NO:125)
核糖体蛋白L13a	RPL13A	NM_012423 (SEQ ID NO:126)
琥珀酸脱氢酶复合物, 亚单位A	SDHA	NM_004168 (SEQ ID NO:127)
TATA盒结合蛋白 (转录因子IID)	TBP	NM_003194 (M34960) (SEQ ID NO:128)

[0408]

基因名称	缩写	基因登记号 (SEQ ID NO:)
转移受体 (p90, CD71)	TFRC	NM_003234 (SEQ ID NO:129)
遍在蛋白C	UBC	NM_021009 (M26880) (SEQ ID NO:130)
酪氨酸3-单加氧酶/色氨酸5-单加氧酶激活蛋白, $\zeta$ 多肽	YWHAZ	NM_003406 (SEQ ID NO:131)
真核18S核糖体RNA	18S	X03205 (SEQ ID NO:132)

[0409] 如上文 5.4.1.2 节中所述,可以对于本文描述的任何基因设计引物。关于表 1A 和表 1B 中鉴定的基因的可公开获得的序列由基因登记号 (GenBank 数据库) 指出,并且整体引入本文作为参考。关于表 1A 和表 1B 中鉴定的基因的序列公开于附随序列表中,如由表中给出的合适 SEQ ID NO 列出的。

[0410] 转录数据分析

[0411] 对于每种基因测定关于衍生自两份重复 PCR 板的每个未知样品的平均 Ct (循环阈值) 值。在实时 PCR 测定中,通过荧光信号的累积检测阳性反应。Ct 定义为荧光信号越过阈值 (即超过本底水平) 所需的循环数目。Ct 水平与样品中的靶基因量成反比 (即 Ct 水平越低,样品中的靶核酸量越大)。

[0412] 通过  $2^{-\Delta Ct}$  方法 (Livak, K. 和 Schmittgen, T., Methods, 2001, 25 :402-408) 计算关于每个未知 cDNA 样品以及参考 cDNA 的相对表达水平,使用来自 7 种标准化基因的平均 Cts。接下来,将参考 cDNA 的相对表达水平设为 100%,所有其他样品随后表示为参考的百分比。最后,通过将百分比乘以参考 cDNA 中含有的每种基因的拷贝数目,将这些百分比转换为拷贝 /ng cDNA。

[0413] 单变量统计分析和制图

[0414] 使用 R 统计包研究基因表达值和衍生自患者 / 受试者问卷的临床参数之间的关系。需要时使用问卷数据编码,以促进比较。基因表达数据在分析前进行对数转化,并且执行参数和非参数分析。关于显著性的阈值设为  $p < 0.05$ 。参见例如,表 3。单变量检验用于测定特定基因对于给定受试者群体是否一致上或下调。

[0415] 使用 GraphPad Prism4<sup>®</sup> (GraphPad Software, Inc, San Diego, CA) 对于每种基因生成比较对照受试者和抑郁患者之间的表达水平的散布图和相关单变量统计分析。因为基因表达值不一定是正态分布的,所以非参数曼 - 怀二氏检验用于比较组。显著性阈值设为  $p < 0.05$ 。特定基因及其在血液中的相对表达水平在图 2 到 7 中例示。

[0416] 多变量分析

[0417] 为了使患病患者与健康对照受试者区分,使用分类算法。分类算法一般是机器学习算法,通过下述 2 个步骤运行:(1) 从 mRNA 转录数据集合中选择基因子集,其基因表达水

平共同发现是最有益的；(2) 训练且返回对如步骤 (1) 中鉴定的基因子集训练的预先选择类型的分类算法。

#### [0418] (1) 基因的选择

[0419] 在第一个步骤中,来自健康对照受试者和抑郁受试者,或其他患病受试者的 mRNA 转录数据集,共同用作对于随机森林算法的输出 (Breiman, L., 2001, Machine Learning 45(1):5-32)。代表来自每个受试者血样的 mRNA 转录数据的每个数据集基于表 1A 中列出的基因和本文描述的方法。通过成功消除最不重要的基因,随机森林算法返回含有最重要基因的列表,使用泄露 (out-of-bag) (OOB) 误差最小化标准 (Liaw, A 和 Wiener, M. December 2002, Classification and regression by randomForest. R News 第 2/3 卷:18-22)。

#### [0420] (2) 训练和分类

[0421] 在第二个步骤中,使用与如步骤 (1) 中鉴定的最重要基因相关的转录概况调整支持向量机分类算法 (Cortes, C. 和 Vapnik, V. 1995, Machine Learning, 20(3):273-97) 等,且基于交叉验证进行训练。

[0422] 在另一个方法中,逐步逻辑回归用于步骤 (1) 选择最重要或说明的基因,和步骤 (2) 经由交叉验证训练算法用于分类。

[0423] 在其他分析中,使用 RVM 分类器,连同遗传算法。用 RVM 算法训练数据集,并且遗传算法评估大量 RVMs,其对不同候选变量子集训练且测试,以鉴定可能的基因相互作用。每个变量子集的性能通过交叉验证进行评估。

[0424] 在训练步骤过程中,通过算法执行交叉验证方法,例如留一法交叉验证 (LOOCV) 或 10 折交叉验证。交叉验证是使数据样品分离成不同子集的统计实践,从而使得分析最初对单个子集执行,而一个或多个其他子集保留用于在证实且验证最初分析中后续使用。数据的最初子集是训练集;一个或多个其他子集是验证或测试集,这作为未知处理以便测定其分类。

[0425] 例如,来自所有样品 (N) 的数据分裂成 2 个不同子集,其中一个数据子集 (m) 用于验证样品,即子集 m 用作未知集合。其余子集 (N-m) 训练分类算法。重复此类交叉验证 (CV) 方法,直至所有数据集作为未知处理。准确度值和预测值可以基于作为未知处理的样品各自正确分类与否进行计算。

[0426] 在一个此类交叉验证方法中,分类算法用 90% 的样品数据集进行训练,并且其余 10% 的样品数据的分类通过训练算法进行预测。此类 10 折 CV 重复 10 次。交叉验证可以举例说明“操作曲线”,即训练分类算法执行得比某些随机选择过程更好,例如比偶然性更好。为了估计根据上文 (1) 和 (2) 中给出的指示构建的分类算法的分类误差,对于准确度、阳性预测值 (PPV) 和阴性预测值 (PPV) 进行计算,以测定训练分类算法已如何良好地执行。

[0427] 训练分类算法的准确度是在样品总数目中的正确分类出的总数目。

[0428] 通过上述方法,正确评分在“患病”类别中的数据集 (即受试者) 数目给出阳性预测值 (PPV) 的测量。PPV 也称为精确率,或疾病的验后概率,是正确诊断的具有阳性测试结果受试者比例。

[0429] 同样通过上述方法,正确评分在“健康”或“对照”组中的数据集 (即受试者) 数目给出阴性预测值 (NPV) 的测量。阴性预测值是正确诊断的具有阴性测试结果的受试者比例。

[0430] 随机化（交换）数据集的分析。

[0431] 为了测定使用 SLR 或 SVM 获得的分类准确度是否有意义，即比偶然性更好，每个数据集如下进一步分析：

[0432] a) 关于原始数据集的准确度通过上文说明的方法获得。

[0433] b) 产生 3 个新交换数据集，其中关于每个个别样品的指定是随机指定的，同时仍维持与原始数据集中的相同患者百分比。

[0434] 表 6

[0435]

生物标记(基因缩写)	双相型受试者组特征: 丰度 = 生物标记的平均转录物值 ( $\pm$ SD)	对照受试者组特征: 丰度 = 生物标记的平均转录物值 ( $\pm$ SD)
ADA	4775 $\pm$ 1508	4511 $\pm$ 1710
ARRB1	292298 $\pm$ 89272	297143 $\pm$ 91094
ARRB2	111023 $\pm$ 39397	114780 $\pm$ 39962
CD8a	11668 $\pm$ 5573	14693 $\pm$ 8416
CD8b	7998 $\pm$ 3841	8687 $\pm$ 3880
CREB1	62347 $\pm$ 18282	63725 $\pm$ 16022
CREB2	79456 $\pm$ 16778	77059 $\pm$ 15755
DPP4	7618 $\pm$ 3077	7169 $\pm$ 2890
ERK1	34901 $\pm$ 15116	39016 $\pm$ 12900
ERK2	57832 $\pm$ 21427	54137 $\pm$ 18660
Gi2	192417 $\pm$ 98987	226358 $\pm$ 87609
Gs	304202 $\pm$ 171505	303930 $\pm$ 139837
GR	124054 $\pm$ 42231	80610 $\pm$ 26544
IL1b	21577 $\pm$ 13468	21006 $\pm$ 9313
IL6	173 $\pm$ 78	182 $\pm$ 221
IL8	24568 $\pm$ 19226	28024 $\pm$ 19993
INDO	5428 $\pm$ 3847	5596 $\pm$ 4418
MAPK14	66946 $\pm$ 25751	51632 $\pm$ 20341
MAPK8	12584 $\pm$ 3060	12162 $\pm$ 3500
MKP1	501068 $\pm$ 251853	499308 $\pm$ 220665
MR	3409 $\pm$ 1094	2830 $\pm$ 887
ODC1	67672 $\pm$ 50925	58670 $\pm$ 40801
P2X7	1322 $\pm$ 418	1542 $\pm$ 563
PBR	64761 $\pm$ 29660	64439 $\pm$ 29328
PREP	6806 $\pm$ 1677	7072 $\pm$ 2102
RGS2	499864 $\pm$ 264854	477280 $\pm$ 165907
S100A10	42063 $\pm$ 12765	35819 $\pm$ 10568
SERT	1435 $\pm$ 710	1711 $\pm$ 1317
VMAT2	2736 $\pm$ 1050	2792 $\pm$ 1344

[0436] (SD = 标准差)

[0437] c) 随后对于每个随机化数据集计算准确度。

[0438] d) 使用曼-怀二氏检验使 10 个准确度 (来自原始数据集的 10 折 CV) 与 30 个交

换准确度（已经历 10 折 CV 的 3 个随机集合）比较。

[0439] e) 产生小于 0.01 的 p 值的比较解释为意指来自原始数据集的准确度不是由于随机偶然性,即对照和患者组可以分离。产生大于 0.01 的 p 值的比较被视为随机的,意指患者和对照组不是有说服力地分离的。

[0440] 用于转录概况鉴定的患者 / 受试者

[0441] 这些研究的一个目标是定义、联系且连接在正常供体的血液中鉴定的转录概况与亚组,所述亚组可帮助鉴定处于神经精神性病症例如情感障碍危险中的表型。一旦已建立了正常供体的基线转录概况,在正常群体和具有临床诊断抑郁症、严重抑郁症、双相型障碍、BPD 或 PTSD 的患者之间进行比较。这些研究的另一个目标是鉴定可以将受试者分类为正常对照或具有情感障碍的患者的概况,所述情感障碍例如抑郁症、严重抑郁症、双相型障碍、BPD 或 PTSD。

[0442] 为了测定在正常群体内的亚组的存在,例如具有危险概况的受试者,且能够使亚组与全血中的转录概况联系,建立正常志愿者的基线数据库。

[0443] 对照患者 / 受试者 (美国)

[0444] 从在服务东南方 Pennsylvania 和 Delaware 地区的血库捐血的正常志愿者收集 500 个血样。从所有供体获得知情同意书。个人信息不可逆地匿名。

[0445] 供体局限于高加索人,以使群体内的变异降到最低。在群体内,供体在性别之间平均分开。超过由血库对于供体使用的那些不存在另外的排除因子。所有供体需要填写问卷,以帮助表征一般身体状况、医学问题、药物使用和滥用、家族史和精神病问题。问卷的要素基于在公众域中可获得的标准精神病测量。关于问卷的回答是自述的,并且供体未接受医学或精神病评估。问卷覆盖多重因素,包括表 2 中分类的那些因素。

[0446] 表 2

[0447]

人口统计学	一般医学	家族史	精神病史/生活经验	抑郁症状
种族	高度/重量	亲属自杀	过去/近期应激生活事件的存在/严重度	植物性机能 (vegetative functions) 中的改变
性别	目前/过去用药	家族精神病史	精神病的先前诊断	认知功能中的改变
婚姻状况	目前/过去医学问题			焦虑/惊恐发作
雇用状况	手术			
职业	烟草使用			
用餐频率	酒精使用			
	药物滥用			

[0448] 广泛问卷用于获得关于供体的历史或目前医学病状的多重因素的数据,所述因素可能增加其未来精神病病症的危险,并且使独特的转录概况与使用问卷鉴定的特定表型关联。这个数据用于分段正常群体,并且比通过使用目前可获得的方法更可靠且一致地鉴定在抑郁患者内的区段。评估的因素包括(但不限于):近期应激生活事件的严重度、早期生活应激的存在和严重度、精神病病症的家族史和抑郁前营养症状组包括食欲和睡眠形式中的改变。需要时,使来自多组问卷的得分组合,以评估多重负面因素的影响,即症状得分。

[0449] 为了避免共性因子的混淆,例如吸烟或体重指数(BMI),这可以视为可以潜在影响血液转录概况的极端,通过人口统计学、个人或医学属性中的可鉴定模式,问卷数据用于使供体分组。独立地评估这些因素,以评估其对转录概况的作用。供体的鉴定和分段根据非精神病因素,以评估其对转录概况的作用,因为这些可以在抑郁前表型的鉴定中混淆,其中此类因素包括: BMI、吸烟、酒精滥用、药物使用(和滥用)。还评估了其他因素的作用。

[0450] 对照患者/受试者(丹麦)

[0451] 200个受试者选自来自约1000个健康志愿者(对照受试者)的最初血液收集,基于丹麦种族起源(回溯2代)且在地理学上覆盖丹麦。因此,获得关于出生地(以及父母和祖父母的出生地)的数据。最初获得一般健康状况和精神病史。精神病史信息补充有关于抑郁的先前发作的短筛选。200个对照受试者的群导致具有约40岁的平均年龄(范围18-65岁)的男性和女性的相等分布。使每个受试者暴露于较小的体格检查,包括评估高度、重量、测量腹和臀围以及EKG。每个受试者完成详细问卷,其中他们就特定人格特点和更全面的医学和精神病家族史进行表征。(参见表2。)

[0452] 使用如上所述由对照受试者提供的数据,使正常群体分段,且使特定表型与外周血中鉴定的转录概况中的改变关联。参见表3A和3B。

[0453] 对照/患者受试者(英国)

[0454] 从参与英国中的控制临床研究的健康志愿者中收集血样。从所有供体获得知情同意书。男性和女性都包括在研究中。包括的女性如果使用公认的避孕方法（双屏障避孕），已手术不育，或是绝经后的（定义为 2 年内无月经）- 口服避孕药是不允许的。包括的受试者是  $\geq 18$  岁和  $\leq 45$  岁，但小于  $\geq 65$  岁。在研究者看来，研究中包括的每个受试者都处于良好健康中，基于研究前体格检查、医疗史、生命体征、ECG 以及血液生物化学、血液学和血清学测试和尿液分析结果。

[0455] 抑郁患者中的转录概况的鉴定

[0456] 为了评估抑郁患者中的转录概况中的改变，在控制临床研究中获得来自抑郁患者，即患有主要性抑郁障碍（MDD）的患者的血液。从所有供体获得知情同意书。

[0457] 患者选择标准：

[0458] 对于研究合格的患者 / 受试者是门诊病人、男性或女性、患有中度 MDD，在基线就诊时具有 MADRS 总得分  $\geq 26$  和 CGI-S 得分  $\geq 4$ 。MDD 的初步诊断必须根据 DSM-IV-TR<sup>®</sup> 标准。患者年龄是 18-65 岁（包括末端），且从精神病门诊病人诊所和全科医生招募。患有继发共病焦虑症的患者可以包括在研究中，除强迫性障碍（OCD）、创伤后应激障碍（PTSD）或惊恐障碍（PD）（DSM-IV-TR<sup>®</sup> 标准）外。此外，在研究者看来，患者在其他方面是健康的，其基于体格检查、医疗史和生命体征。在研究者看来，不太可能顺从临床研究方案或由于任何原因不适合的患者可能从研究中排除。

[0459] 在抑郁患者中的转录概况的鉴定

[0460] 为了评估患有重度主要性抑郁障碍（SMDD）的患者中的转录概况中的改变，在控制临床研究中获得来自这些患者的血液。从所有供体获得知情同意书。

[0461] 患者选择标准：

[0462] 对于这个研究合格的患者 / 受试者是患有 MDD 的门诊病人、从精神病门诊病人诊所招募、男性或女性、年龄 18-65 岁（包括末端）。在这个研究中包括的所有患者都应具有 MADRS 总得分 30 或以上（即更严重抑郁的患者）。选择的患者患有主要性抑郁发作（MDE）作为根据 DSM-IV-TR<sup>®</sup> 标准的初步诊断（用简明国际神经精神访谈（Mini International Neuropsychiatric Interview）（MINI）评估的目前发作）。报道的目前 MDE 持续时间在基线时是至少 3 个月和小于 12 个月。基于如上文就中度抑郁的患者而言所述的标准，从研究中包括 / 排除患者。在研究者看来，不太可能顺从临床研究方案或由于任何原因不适合的患者可以从研究中排除。

[0463] 在双相型患者中的转录概况的鉴定

[0464] 为了评估双相型患者中的转录概况中的改变，获得来自双相型患者的血液。这些患者已经历通过精神科医生的广泛评估且在医疗护理下。从所有供体获得知情同意书。

[0465] 患者选择标准：

[0466] 在患者 / 受试者可以在这个方案下捐血前，必须达到下述标准：

[0467] a) 根据 DSM-IV-TR<sup>®</sup>，患者已诊断具有中度或中度主要性抑郁症或双相型 I。87% 的患者符合关于双相型 I 障碍的 DSM-IV-TR<sup>®</sup> 标准。

[0468] b) 在血液收集时，患者未服用任何精神药理学药物且未服用任何精神药理学药物至少 2 周。此外，患者无一已用氟西汀、不可逆 MAOI 或储存神经松弛药治疗至少 2 个月。

[0469] c) 患者未患有其他急性精神病症状，例如物质滥用。

[0470] d) 可能时,来自女性患者的血样应在月经开始2周内收集。在任何情况下,应记录末次月经期的第一天的日期。

[0471] e) 患者在过去6个月过程中未服用任何违禁药物/药物滥用。

[0472] f) 患者在过去6个月过程中未滥用酒精。

[0473] g) 女性患者未怀孕且未哺乳。

[0474] h) 患者目前(包括过去一周)未患有任何其他急性一般医学病状(包括较小病状,例如普通感冒)。

[0475] i) 患者目前(包括过去一周)未服用任何常规药剂(包括口服避孕药、草药疗法、营养补充剂、维生素)。

[0476] j) 患者在血液收集前一周内不应服用任何药剂(包括口服避孕药、草药疗法、营养补充剂、维生素)。如果服用药物,例如用于急性头痛,那么血样收集应延迟1周。

[0477] k) 如果患者指出烟草使用,那么需要提供关于平均量/天的信息。

[0478] l) 如果患者指出无滥用的酒精消费,那么需要提供关于平均量/天的信息。

[0479] m) 患者伴随血样收集已返回问卷。

[0480] n) 患者已阅读且理解患者信息。

[0481] o) 患者已签名知情同意书。

[0482] 从在这个方案下捐血的所有患者中,必须获得下述信息:详细精神病和一般医疗史、精神病家族史、目前症状的详细临床描述、关于至少过去3个月的医疗史、和关于至少过去6个月中的违禁和非违禁药物滥用的信息。

[0483] 在边缘型人格障碍患者中的转录概况的鉴定

[0484] 为了评估具有边缘型人格障碍(BPD)患者中的转录概况中的改变,获得来自边缘型人格障碍患者的血液。这些患者已经历通过精神科医生的广泛评估且在医疗护理下。从所有供体获得知情同意书。

[0485] 用于BPD研究的患者/受试者选择标准:

[0486] 在患者可以在这个方案下捐血前,必须达到下述标准:

[0487] a) 根据DSM-IV®,患者已诊断具有边缘型人格障碍。

[0488] b) 对于未治疗患者组,在血液收集时,患者未服用任何精神药理学药物且未服用任何精神药理学药物至少2周。在血液收集前,在过去已用氟西汀、不可逆MAOI或储存神经松弛药治疗的患者未服用这些药剂中的任何至少4周。

[0489] c) 在原发性精神障碍(边缘型人格障碍)的急性精神病恶化过程中,将从小群患者(约25个患者)中收集血样。在血液收集时,所有其他患者将未患有急性精神病恶化。仅在其血液在急性恶化过程中取样的患者中,在缓解过程中将收集第二个血样。医学上可能时,在2个时间点时的处理将是相同的。

[0490] d) 患者未患有其他急性精神病症状,例如物质滥用。

[0491] e) 可能时,来自女性患者的血样应在月经开始2周内收集。在任何情况下,应记录末次月经期的第一天的日期。

[0492] f) 患者在过去6个月过程中未服用任何违禁药物/药物滥用。

[0493] g) 患者在过去6个月过程中未滥用酒精。

[0494] h) 女性患者未怀孕且未哺乳。

[0495] i) 患者目前（包括过去一周）未患有任何其他急性一般医学病状（包括较小病状，例如普通感冒）。

[0496] j) 患者目前（包括过去一周）未服用除处方文拉法辛或度洛西汀外的任何常规药剂（包括口服避孕药、草药疗法、营养补充剂、维生素）。

[0497] k) 如果患者用文拉法辛或度洛西汀治疗，那么治疗必须已以目前剂量给予至少 3 个月。

[0498] l) 患者在血液收集前一周内不应服用任何药剂（包括口服避孕药、草药疗法、营养补充剂、维生素）。如果服用药物，例如用于急性头痛，那么血样收集应延迟 1 周。

[0499] m) 如果患者指出烟草使用，那么需要提供关于平均量 / 天的信息。

[0500] n) 如果患者指出无滥用的酒精消费，那么需要提供关于平均量 / 天的信息。

[0501] o) 患者伴随血样收集已返回问卷。

[0502] p) 患者已阅读且理解患者信息。

[0503] q) 患者已签名知情同意书。

[0504] 从在这个方案下捐血的所有患者中，获得详细精神病史，包括家族史、临床描述以及药剂和药物记录。

[0505] 患者完成为具体解决可以混淆转录概况的因素开发的问卷，所述因素例如药物使用、一般医学病状。患者将问卷返回给研究者。问卷用与血样和其他临床数据相同的编码进行编码，以确保患者的身份对于转录分析场所的个人不公开。问卷连同血样一起转移至转录分析场所。

[0506] 在创伤后应激障碍 (PTSD) 患者中的转录概况

[0507] 为了评估具有 PTSD 患者中的转录概况中的改变，获得来自 PTSD 患者的血液。这些患者已经历通过精神科医生的广泛评估且在医疗护理下。从所有供体获得知情同意书。

[0508] 用于 PTSD 研究的患者 / 受试者选择标准：

[0509] 用于这个研究的受试者是符合下述标准的男性：

[0510] a) 患者已诊断具有急性 PTSD 或复发性 PTSD (根据 DSM-IV®)，或已暴露于创伤且未发展 PTSD 或分类为对照。选择未暴露于创伤且最初来自相同地理区域的对照用于这个研究。

[0511] b) 在血液收集时，患者未服用任何精神药理学药物且未服用任何精神药理学药物至少 2 周。在血液收集前，在过去已用氟西汀、不可逆 MAOI 或储存神经松弛药治疗的患者未服用这些药剂中的任何至少 4 周。

[0512] c) 患者未患有其他急性精神病症状，例如物质滥用。

[0513] d) 患者在过去 6 个月过程中未服用任何违禁药物 / 药物滥用。

[0514] e) 患者在过去 6 个月过程中未滥用酒精。

[0515] f) 患者目前（包括过去一周）未患有任何其他急性一般医学病状（包括较小病状，例如普通感冒）。

[0516] g) 患者在血液收集前一周内不应服用任何药剂（包括口服避孕药、草药疗法、营养补充剂、维生素）。如果服用药物，例如用于急性头痛，那么血样收集应延迟 1 周。

[0517] h) 如果患者指出烟草使用，那么需要提供关于平均量 / 天的信息。

[0518] i) 如果患者指出无滥用的酒精消费，那么需要提供关于平均量 / 天的信息。

[0519] j) 患者目前 (包括过去一周) 未服用任何常规药剂, 包括草药疗法、营养补充剂、维生素)。

[0520] 在将信息转移至转录分析场所 (Lundbeck Research USA, Inc., Paramus, NJ) 前, 在血液收集场所收集如上所述的所有临床和人口数据。在 Lundbeck Research USA 执行临床特征和转录概况之间的任何关系的探索分析, 没有患者身份的了解。

[0521] 结果和讨论

[0522] 在对照受试者中的转录概况的鉴定

[0523] 在来自对照受试者的血样中测量关于表 1A 中列出的 29 种基因的基因表达水平, 包括来自 2 个对照组 (U. S. 和 DK) 的受试者。

[0524] 尽管这些个体都是健康的, 但鉴定了与对于问卷项目的特定应答关联的基因表达趋势。如果鉴定, 那么此类趋势可能在抑郁患者群体中放大。

[0525] 将问卷应答转换成用于统计分析的编码值

[0526] 由美国和丹麦对照受试者填写的自我评估问卷含有相似但并非等同的项目。为了使用来自问卷的信息以搜索应答和基因表达数据之间的可能联系, 在统计分析前必须编码信息。

[0527] 编码策略的例子如下:

[0528] a) 连续变量例如能量和 BMI 如由受试者报告的使用。可替代地, 在分析前使原始得分组合成 2 个或 3 个框 (高、中、低值)。

[0529] b) 使性别转换为二元应答 (0, 1)。

[0530] c) 使关于与抑郁症相关的症状频率的问题, 例如睡眠困难、精力缺乏或情绪低落从文字答案 (从不、有时、大多数天、每一天) 转换为数字值 (0、1、2、3)。

[0531] d) 通过将关于特定症状组合的值相加以产生复合得分, 产生组合症状得分。随后使复合得分框并。

[0532] e) 使关于受试者的抑郁症 / 焦虑家族史的问题从文字答案 (无一、仅二级亲属、一级亲属) 转换为数字值 (0、1、2)。

[0533] f) 使关于受试者的抑郁症 / 焦虑个人史或用于抑郁症 / 焦虑的药理学治疗的问题从文字答案 (无一、一个或多个) 转换为二元应答 (0, 1)。

[0534] 在编码后, 各种统计检验, 包括 Spearman 相关分析、t 检验和 ANOVA, 用于搜索基因表达水平和特异性临床变量之间的关联。

[0535] 合适时, 使用统计检验, 使每种基因的表达与由受试者在自我评估问卷上提供的编码答案比较, 以鉴定关联。因为进行总共 377 个比较 (29 种基因乘以 13 个问卷应答), 所以关于显著性的阈值设为  $p < 0.01$ , 以使 1 型误差的概率降到最低, 同时仍保留大量统计上显著的结果。

[0536] 表 3A 和 3B 显示基于分析的问卷应答, 关于在对照群体内具有显著差异的 29 种基因 (来自表 1A) 中的仅 15 种的关联数据。对于其余基因未检测出显著差异。表 3A 和 3B 显示关于 13 个问卷应答中的 11 个的数据, 然而, 未显示关于 BMI 和年龄的关联数据, 因为它们并无显著不同。与显著基因表达概况相关的某些临床参数是人生经验、终生治疗和症状得分。

[0537] 表 3A 和 3B。在 2 个对照组中临床变量和基因表达之间的关联。

[0538] (\*\* = p < 0.01 标准; \*\*\* = p < 0.00 标准)

[0539] 表 3A1)US 受试者 2)DK 受试者

[0540]

	CREB2	DPP4	ERK1	ERK2	GR	Gs	MAPK8	MAPK14
1) 家族史 (D/A/S)		Inc **					Inc **	
2) 家族史 (D/A/S)								
1) 烟草使用								
2) 烟草使用								
1) 人生经验 (D/A)	Inc ***	Inc ***		Inc **	Inc ***		Inc ***	
2) 人生经验 (D/A)		Inc ***			趋势向上			
1) 终生治疗 (D/A)	Inc **	Inc ***			Inc **		Inc ***	
2) 终生治疗 (D/A)					趋势向上	Inc **		
1) 食欲改变		Inc **						
2) 食欲改变								
1) 睡眠问题		Inc **						
2) 睡眠问题		Inc **						
1) 10 个症状得分 (*)		Inc ***					Inc ***	
2) 10 个症状得分 (*)			Inc **	Inc ***				

[0541]

	CREB2	DPP4	ERK1	ERK2	GR	Gs	MAPK8	MAPK14
1) 植物性症状		Inc **						
2) 植物性症状								
1) 近期应激		Inc **						
2) 近期应激								
1) 早期生活应激								
2) 早期生活应激								
1) 对性的兴趣					Inc **			
2) 对性的兴趣								Inc **

[0542] (D/A/S = 抑郁症 / 焦虑 / 自杀 ; D/A = 抑郁症 / 焦虑)

[0543] 表 3B1) US 受试者 2) DK 受试者

[0544]

	MKP1	MR	PBR	RGS2	S100A10	SERT	VMAT2
1) 家族史 (D/A/S)							
2) 家族史 (D/A/S)							
1) 烟草使用						Dec ***	
2) 烟草使用						趋势向下	
1) 人生经验 (D/A)				Inc **			Inc **
2) 人生经验 (D/A)				趋势向上			
1) 终生治疗 (D/A)	Inc **						
2) 终生治疗 (D/A)	趋势向上						

[0545]

	MKP1	MR	PBR	RGS2	S100A10	SERT	VMAT2
1) 食欲改变						Dec **	
2) 食欲改变						趋势向下	
1) 睡眠问题							
2) 睡眠问题							
1) 10 个症状得分 (*)	趋势 向上		Dec **				
2) 10 个症状得分 (*)	Inc **		Inc **				
1) 植物性症状							
2) 植物性症状							
1) 近期应激							
2) 近期应激							
1) 早期生活应激							
2) 早期生活应激					Inc ***		
1) 对性的兴趣			趋势 向下				
2) 对性的兴趣		Dec **	Inc **				

[0546] (D/A/S = 抑郁症 / 焦虑 / 自杀 ;D/A = 抑郁症 / 焦虑)

[0547] 在分析的 377 个总组合中,23 个组合 (6%) 表明在分析的 2 个对照组之间的显著差异。然而,345 个 (94%) 组合显示出相同概况。这些组合中的 9 个显示对于研究的 2 个对照组以相同方向 (即基因的上或下调) 在基因表达中的改变,如由表 3A 和 3B 中的阴影框指出的。总的来说,分析显示用于分析的 2 个对照组显示非常相似的基因表达趋势或概况。

[0548] 与临床参数相关的基因表达概况还可以通过本文描述的多变量算法进行分析。因此,可以对用转录数据组合的临床变量实施本领域技术人员已知的任何合适算法,例如逐步逻辑回归或 PELORA。

[0549] 在抑郁患者中的转录概况的鉴定。

[0550] 通过单变量方法首先分析得自未接受抗抑郁药治疗的 174 个中度抑郁患者 / 受试

者的血样。测量关于选自表 1A 的基因的转录水平,并且与此类基因在 196 个健康对照受试者中的表达水平比较。与对照比较,在抑郁患者中代表性基因的表达概况显示于图 2A-2B 和 3A-3B 中。

[0551] 使用 RF(选择)和 SVM(训练)中度抑郁患者与对照比较的分类导致 88%的高准确度,如图 8A 中所示( $PPV = 89\%$ ;  $NPV = 88\%$ )。使用 SLR 算法中度抑郁患者与对照比较的分类导致 93%的高准确度,如图 8A 中所示( $PPV = 93\%$ ;  $NPV = 94\%$ ),所述 SLR 算法执行基因选择和训练。

[0552] 如图 8B 中所示,在基于完整数据集选择的基因中,2 种算法都显示出良好一致。随机森林选择了 14 种基因,并且 SLR 选择了 17 种基因,作为用于分类的最重要基因,基于每种方法的统计参数。11 种基因被 2 个方法选择出来,包括 ARRB1、ARRB2、CD8a、CREB1、CREB2、ERK2、Gi2、MAPK14、ODC1、P2X7 和 PBR。

[0553] 使数据集随机化,即使样品作为患者或对照的指定随机化,并且实施与上文相同的多变量分析。在随机化后,2 种分类算法(RF/SVM 和 SLR)都产生在统计上不同于用实际数据获得的那些的准确度值,指出上文列出的值(图 8A)优于偶然性,并且组可在统计上分开。

[0554] 受试者可以进行概况分析,并且如上文所述对其基于表 1A 中的基因的转录数据实施用参数训练的分类算法,以获得中度抑郁症的诊断。

[0555] 基于每个生物标记(即,基因转录物)的丰度,关于选自表 1A 的基因的抑郁受试者的转录概况显示于表 4 中。显示对照受试者转录物值用于比较。

[0556] 表 4

[0557]

生物标记 (基因缩写)	抑郁受试者组特征: 丰度 = 生物标记的平均转录物值 ( $\pm$ SD)	对照受试者组特征: 丰度 = 生物标记的平均转录物值 ( $\pm$ SD)
ADA	4691 $\pm$ 2453	4511 $\pm$ 1710
ARRB1	189062 $\pm$ 62727	297143 $\pm$ 91094
ARRB2	84195 $\pm$ 31728	114780 $\pm$ 39962
CD8a	8304 $\pm$ 5825	14693 $\pm$ 8416
CD8b	8145 $\pm$ 4394	8687 $\pm$ 3880
CREB1	71743 $\pm$ 20237	63725 $\pm$ 16022
CREB2	63732 $\pm$ 14463	77059 $\pm$ 15755
DPP4	6649 $\pm$ 2331	7169 $\pm$ 2890
ERK1	25326 $\pm$ 10178	39016 $\pm$ 12900
ERK2	58338 $\pm$ 18813	54137 $\pm$ 18660
Gi2	115117 $\pm$ 53383	226358 $\pm$ 87609
Gs	262885 $\pm$ 112989	303930 $\pm$ 139837
GR	73224 $\pm$ 23517	80610 $\pm$ 26544
IL1b	29631 $\pm$ 13692	21006 $\pm$ 9313
IL6	348 $\pm$ 523	182 $\pm$ 221
IL8	45487 $\pm$ 106224	28024 $\pm$ 19993
INDO	6031 $\pm$ 10133	5596 $\pm$ 4418
MAPK14	73156 $\pm$ 33915	51632 $\pm$ 20341
MAPK8	12906 $\pm$ 3836	12162 $\pm$ 3500
MKP1	525383 $\pm$ 268053	499308 $\pm$ 220665
MR	2565 $\pm$ 1110	2830 $\pm$ 887
ODC1	71892 $\pm$ 32249	58670 $\pm$ 40801
P2X7	1095 $\pm$ 432	1542 $\pm$ 563
PBR	70854 $\pm$ 30278	64439 $\pm$ 29328
PREP	6715 $\pm$ 2072	7072 $\pm$ 2102
RGS2	632976 $\pm$ 262593	477280 $\pm$ 165907
S100A10	32173 $\pm$ 9530	35819 $\pm$ 10568
SERT	1400 $\pm$ 1164	1711 $\pm$ 1317
VMAT2	3469 $\pm$ 1602	2792 $\pm$ 1344

[0558] (SD = 标准差)

[0559] 还通过使关于抑郁受试者的转录物值与对照受试者比较的比值来评估 2 基因组合。在抑郁受试者和对照受试者之间可见特定生物标记的丰度比值中的明显差异, 如表 4A 中。

[0560] 表 4A

[0561]

生物标记	关于抑郁受试者组的转录物丰度比值	关于对照组的转录物丰度比值
ERK1	0.35	0.76
MAPK14		
IL1b	0.26	0.09
Gi2		
MAPK14	0.39	0.17
ARRB1		
ERK1	0.85	1.86
IL1b		

[0562] 为了评估更严重抑郁的患者群体中的转录概况中的改变,获得来自 120 个严重抑郁患者的血液,并且对于选自表 1A 的基因测量基因表达。通过单变量方法在统计上分析基因表达数据。使患者转录数据与 196 个对照的转录数据比较,并且关于个别基因数据的代表性散布图显示于图 4A-4C 中。

[0563] 使用 RF/SVM 的分类导致 92% 的高准确度 (PPV = 89% ;NPV = 94%)。SLR 算法的分类导致 93% 的高准确度 (PPV = 91% ;NPV = 95%),所述 SLR 算法执行基因选择和训练。

[0564] 在基于完整数据集选择的基因中,2 种算法都显示出良好一致。基于每种方法的统计参数,随机森林分类选择了总共 7 种基因并且 SLR 选择了总共 12 种基因作为用于分类的最重要基因。5 种基因被 2 个方法选择出来,包括 CD8a、ERK1、MAPK14、P2X7 和 PBR。

[0565] 在患者 / 对照指定的随机化后,2 种分类算法 (RF/SVM 和 SLR) 都产生在统计上不同于用实际数据获得的那些的准确度值,指出上文列出的值优于偶然性,并且组可在统计上分开。

[0566] 受试者可以进行概况分析,并且如上文所述对其基于表 1A 中包括的基因的转录数据实施训练分类算法,以获得重度抑郁症的诊断。

[0567] 基于每个生物标记的丰度 (即,基因转录物),关于选自表 1A 的基因的严重抑郁受试者的转录概况显示于表 5 中。显示对照受试者转录物值用于比较。

[0568] 表 5

[0569]

生物标记(基因缩写)	严重抑郁受试者组特征: 丰度 = 生物标记的平均转录物值 ( $\pm$ SD)	对照受试者组特征: 丰度 = 生物标记的平均转录物值 ( $\pm$ SD)
ADA	3812 $\pm$ 1365	4511 $\pm$ 1710
ARRB1	161284 $\pm$ 47341	297143 $\pm$ 91094
ARRB2	79487 $\pm$ 22860	114780 $\pm$ 39962
CD8a	7666 $\pm$ 4603	14693 $\pm$ 8416
CD8b	6897 $\pm$ 3320	8687 $\pm$ 3880
CREB1	64463 $\pm$ 18736	63725 $\pm$ 16022
CREB2	71534 $\pm$ 12311	77059 $\pm$ 15755
DPP4	5873 $\pm$ 2194	7169 $\pm$ 2890
ERK1	19389 $\pm$ 7612	39016 $\pm$ 12900
ERK2	48236 $\pm$ 17894	54137 $\pm$ 18660
Gi2	97344 $\pm$ 42195	226358 $\pm$ 87609
Gs	185642 $\pm$ 82731	303930 $\pm$ 139837
GR	75411 $\pm$ 24542	80610 $\pm$ 26544
IL1b	27643 $\pm$ 12046	21006 $\pm$ 9313
IL6	153 $\pm$ 100	182 $\pm$ 221
IL8	38817 $\pm$ 29253	28024 $\pm$ 19993
INDO	5735 $\pm$ 5467	5596 $\pm$ 4418
MAPK14	67519 $\pm$ 29094	51632 $\pm$ 20341
MAPK8	11446 $\pm$ 3231	12162 $\pm$ 3500
MKP1	615915 $\pm$ 307961	499308 $\pm$ 220665
MR	2023 $\pm$ 893	2830 $\pm$ 887
ODC1	55085 $\pm$ 30043	58670 $\pm$ 40801
P2X7	769 $\pm$ 331	1542 $\pm$ 563
PBR	67863 $\pm$ 24974	64439 $\pm$ 29328
PREP	5186 $\pm$ 1620	7072 $\pm$ 2102
RGS2	571284 $\pm$ 270572	477280 $\pm$ 165907
S100A10	21812 $\pm$ 7985	35819 $\pm$ 10568
SERT	795 $\pm$ 553	1711 $\pm$ 1317
VMAT2	3073 $\pm$ 1715	2792 $\pm$ 1344

[0570] (SD = 标准差)

[0571] 在严重抑郁患者和对照之间显著不同 ( $p < 0.05$ ) 的平均表达水平 (转录物值) 的基因是: ADA、ARRB1、ARRB2、CD8a、CD8b、CREB2、DPP4、ERK1、Gi2、Gs、IL1b、IL8、MAPK14、MKP1、MR、P2X7、PREP、RGS2、S100A10 和 SERT (表 5A)。

[0572] 表 5A :基于 p 值 ( $p < 0.05$ ), 与对照受试者比较, 在严重抑郁受试者中显著不同的基因。

[0573]

生物标记 (基因缩写)	p 值
ADA	$3.2673 \times 10^{-6}$
ARRB1	$4.40419 \times 10^{-60}$
ARRB2	$1.61434 \times 10^{27}$
CD8a	$1.92916 \times 10^{38}$
CD8b	$3.13307 \times 10^8$
CREB2	0.0000507671
DPP4	$1.25015 \times 10^7$
ERK1	$1.12946 \times 10^{-72}$
Gi2	$3.27538 \times 10^{-64}$
Gs	$1.98625 \times 10^{35}$
IL1b	$2.13924 \times 10^{-11}$
IL8	$2.00073 \times 10^{-6}$
MAPK14	$5.2042 \times 10^{-15}$
MKP1	$1.25421 \times 10^{-6}$
MR	$1.73784 \times 10^{-23}$
P2X7	$3.7121 \times 10^{-67}$
PREP	$2.72022 \times 10^{-26}$
RGS2	0.0000152985
S100A10	$2.3756 \times 10^{-53}$
SERT	$4.36216 \times 10^{-26}$

[0574] 这些基因根据计算的  $-\text{Log}(p)$  值的量级进行排序 (图 9), 从而指出关于数种基因在患者转录物值和对照值之间的明显差异, 所述基因例如 ERK1、P2X7、Gi2、ARRB1 和

S100A10。

[0575] 为了搜索转录物值之间的线性和非线性相互作用,执行相关向量机 (RVM) 分类算法,随后使用遗传算法,以便通过可能的基因间相互作用的间隔搜索,并且选择最有力和有意义的相互作用。单基因解决方案还通过这种算法集合进行检查,并且证实单基因解决方案用于使患者与对照分离的有效性。ARRB1 (准确度 = 0.86) 和 ERK1 (准确度 = 0.85) 测定为在单基因分析中高度有益的,随后为 P2X7 (准确度 = 0.82) 和 Gi2 (准确度 = 0.81)。还参见例如图 2 到 5,其中描述了关于中度抑郁、重度抑郁和双相型患者与对照比较的有益基因表达数据。

[0576] 已鉴定了几个 2 基因解决方案用于分类抑郁患者和对照,具有 90% 或更大的准确度。显示 ERK1 和 MAPK14 转录物值,以分类抑郁患者与对照比较,具有 92% 的准确度。图 10 描述了仅基于 ERK1 和 MAPK14 的转录物值,严重抑郁受试者和对照的分布。抑郁受试者 (具有如表 4 中的概况) 的分类与严重抑郁受试者的结果一致。图 11、12 和 13 描述了基于其他 2 基因转录概况的转录物值,分别为 IL1b/Gi2、MAPK14/ARRB1 和 ERK1/IL1b,严重抑郁受试者和对照的分布。还通过比较关于严重抑郁受试者与对照受试者比较的转录物值的比值来评估 2 基因组合。在严重抑郁受试者和对照受试者之间的丰度比值中的明显差异在表 5B 中可见。

[0577] 表 5B

[0578]

生物标记 (基因缩写)	关于严重抑郁受试者组的转录物丰度比值	关于对照组的转录物丰度比值
ERK1	0.29	0.76
MAPK14		
IL1b	0.28	0.09
Gi2		
MAPK14	0.42	0.17
ARRB1		
ERK1	0.70	1.86
IL1b		

[0579] 在具有双相型障碍患者中的转录概况的鉴定。

[0580] 为了评估具有双相型障碍患者中的转录概况中的改变,获得来自 23 个抑郁患者 (根据 DSM-IV 标准,20 个患者被决定性诊断具有双相型障碍) 的血液,并且对于选自表 1A 的基因测量基因表达。通过单变量方法在统计上分析基因表达数据。使患者转录数据与 196 个对照的转录数据比较,并且对于个别基因数据的代表性散布图显示于图 5A-5C 中。

[0581] 使用 RF/SVM 的分类导致 94% 的高准确度 (PPV = 86% ;NPV = 95%)。SLR 算法

的分类导致 97% 的高准确度 (PPV = 90%; NPV = 99%), 所述 SLR 算法执行基因选择和训练。

[0582] 在基于完整数据集选择的基因中, 2 种算法都显示出良好一致, 其中基于每种方法的统计参数, 随机森林分类选择了总共 3 种基因并且 SLR 选择了总共 5 种基因作为用于分类的最重要基因。3 种基因被 2 个方法选择出来, 包括 Gi2、GR 和 MAPK14。

[0583] 在患者 / 对照指定的随机化后, 2 种分类算法 (RF/SVM 和 SLR) 都产生在统计上不同于用实际数据获得的那些的准确度值, 指出上文列出的值优于偶然性, 并且组可在统计上分开。

[0584] 受试者可以进行概况分析, 并且如上文所述对其基于表 1A 中包括的基因的转录数据实施训练分类算法, 以获得双相型障碍的诊断。

[0585] 基于每个生物标记的丰度 (即, 基因转录物), 关于选自表 1A 的基因的双相型受试者的转录概况显示于表 6 中。显示对照受试者转录物值用于比较。

[0586] 在具有边缘型人格障碍患者中的转录概况的鉴定。

[0587] 为了评估具有边缘型人格障碍患者中的转录概况中的改变, 获得来自 21 个边缘型人格障碍患者的血液, 并且对于选自表 1A 的基因测量基因表达。通过单变量方法在统计上分析基因表达数据。使患者转录数据与 196 个对照的那种比较, 并且对于个别基因数据的代表性散布图显示于图 6A-6C 中。

[0588] 使用 RF (选择) 和 SVM (训练) 的分类导致 97% 的高准确度 (PPV = 87%; NPV = 98%)。SLR 算法的分类导致 98% 的高准确度 (PPV = 90%; NPV = 100%), 所述 SLR 算法执行基因选择和训练。

[0589] 在基于完整数据集选择的基因中, 2 种算法都显示出良好一致, 其中基于每种方法的统计参数, 随机森林分类选择了总共 5 种基因并且 SLR 选择了总共 4 种基因作为用于分类的最重要基因。4 种基因被 2 个方法选择出来, 包括 Gi2、GR、MAPK14 和 MR。

[0590] 在患者 / 对照指定的随机化后, 2 种分类算法 (RF/SVM 和 SLR) 都产生在统计上不同于用实际数据获得的那些的准确度值, 指出上文列出的值优于偶然性, 并且组可在统计上分开。

[0591] 受试者可以进行概况分析, 并且如上文所述对其基于表 1A 中包括的基因的转录数据实施训练分类算法, 以获得边缘型人格障碍的诊断。

[0592] 在具有 PTSD 患者中的转录概况的鉴定。

[0593] 在具有急性 PTSD 患者、具有复发性 PTSD 患者和已实施创伤事件而不发展 PTSD 的个体组中评估转录概况。这些组的组合评估呈现机会, 以鉴定与急性 PTSD 相关的表达改变, 以及确定可以与从疾病恢复或对疾病抵抗相关的差异。通过单变量方法在统计上分析基因表达数据。来自具有急性 PTSD 的 66 个患者的患者转录数据与 196 个对照的转录数据比较, 并且对于个别基因数据的代表性散布图显示于图 7A-7C 中。

[0594] 使用 RF (选择) 和 SVM (训练) 的急性 PTSD 患者与对照受试者比较分类导致 77% 的准确度 (PPV = 64%; NPV = 82%)。用 SLR 算法的分类导致 84% 的准确度 (PPV = 77%; NPV = 87%), 所述 SLR 算法执行基因选择和训练。使用这个测试数据集, SLR 算法优于 SVM 算法。每个分类算法与数据集的随机化 (交换) 形式比较, 并且使用交换数据集, SLR 产生 73% 的准确度值 (PPV = 39%; NPV = 75%)。统计分析指出用实际与随机化数据比

较获得的 SLR 准确度值是不同的,指出组可分开。

[0595] 使用交换数据集,SVM 产生 73%的准确度值 (PPV = 10%;NPV = 75%),指出关于交换(随机化)数据的向下趋势。应当指出在 SVM 算法中使用实际数据的 PPV(阳性预测具有疾病的患者的能力)优于 60%,与用交换数据的 10%预测比较,指出使用实际数据训练的算法优于随机预测。

[0596] 基于急性 PTSD 患者与对照比较的完整数据集,SLR 选择了总共 10 种基因作为用于分类的最重要基因:ARRB1、ARRB2、CD8b、ERK2、IDO、IL-6、MR、ODC1、PREP 和 RGS2。

[0597] 受试者可以进行概况分析,并且如上文所述对其基于表 1A 中包括的基因的转录数据实施训练分类算法,以获得急性 PTSD 的诊断。

[0598] 使用 RF(选择)和 SVM(训练)的复发性 PTSD 患者与对照受试者比较的分类导致 81%的准确度 (PPV = 59%;NPV = 85%)。SLR 算法的分类导致 80%的准确度 (PPV = 33%;NPV = 86%),所述 SLR 算法执行基因选择和训练。然而,当对这个数据集的随机化形式运行分类算法时,SVM 和 SLR 分别产生 82%和 81%的准确度值。这些值与用实际数据获得的那些并无统计上不同,指出算法无法可靠地分离这些组。因为缺乏分离,对于这个比较未报告基因列表。从临床观点来看,算法无法区分对照和复发性患者是预期的,由于缺乏这些组之间的生物学差异。因为复发性患者不再显示病的症状,所以假定其基因表达水平已回到正常水平是合理的,从而阻止算法使组有效分离。

[0599] 使用 RF(选择)和 SVM(训练)的受创伤但未发展 PTSD 受试者与对照受试者比较的分类导致 74%的准确度 (PPV = 61%;NPV = 79%)。SLR 算法的分类导致 73%的准确度 (PPV = 59%;NPV = 80%),所述 SLR 算法执行基因选择和训练。当对随机化数据集执行分类算法时,RF/SVM 和 SLR 分类算法都产生在统计上不同于用实际数据获得的那些的准确度值,指出如上报告的值优于偶然性,并且组可分开。

[0600] 基于每种方法的统计参数并且使用来自创伤患者和对照的完整数据集,随机森林分类选择了总共 14 种基因并且 SLR 选择了总共 13 种基因作为用于分类的最重要基因。7 种基因被 2 个方法选择出来,包括 ARRB2、CREB1、ERK2、Gs、IL-6、MKP1 和 RGS2。

[0601] 尽管这些个体未诊断具有 PTSD,但算法仍可以使其与对照区分,虽然具有比本文呈现的某些其他比较更低的准确度、PPV 和 NPV 值。有趣的是,在来自急性 PTSD 患者的 SLR 基因列表上的 6 种基因匹配在关于无 PTSD 创伤患者的相应列表上的那些 (ARRB2、CD8b、ERK2、MR、IL-6 和 RGS2)。虽然受创伤患者仍未发展该病,但他们与已发展该病的患者共享某些基因表达概况,指出他们可能处于危险中。

[0602] 受试者可以进行概况分析,并且如上文所述对其基于表 1A 中包括的基因的转录数据实施训练分类算法,以获得无 PTSD 创伤的诊断。

[0603] 7 引用的参考文献

[0604] 本文引用的所有参考文献整体且为了所有目的引入本文作为参考,其程度与每个个别出版物或专利或专利申请特别且个别指出为了所有目的整体引入本文作为参考相同。

[0605] 8 修改

[0606] 如对于本领域技术人员显而易见的,可以进行本发明的许多修改和变化,而不背离其精神和范围。本文描述的具体实施方案仅提供作为例子,并且本发明仅被附加权利要求,连同此权利要求所要求保护的等价物的全部范围限制。

[0001]

## 序列表

- <110> H. Lundbeck A/S  
Antonijevic, Irina  
Tamm, Joseph  
Artymyshyn, Roman  
Gerald, Christophe P.G.
- <120> 用于测量生物标记概况的系统和方法
- <130> 71021-WO-PCT
- <150> US 61/092, 270  
<151> 2008-08-27
- <160> 132
- <170> PatentIn version 3.4
- <210> 1  
<211> 25  
<212> DNA  
<213> 智人
- <400> 1  
ggtggtggag ctgtgtaaga agtac 25
- <210> 2  
<211> 22  
<212> DNA  
<213> 智人
- <400> 2  
cttcctggga tggctcctc tc 22
- <210> 3  
<211> 26  
<212> DNA  
<213> 智人
- <400> 3  
cagcagaccg tggtagccat tgacct 26
- <210> 4  
<211> 22  
<212> DNA  
<213> 智人
- <400> 4  
agacacgaac ttggcctcta gc 22
- <210> 5

[0002]

<211> 23		
<212> DNA		
<213> 智人		
<400> 5		
ttgtaggaaa caatgatccc cag		23
<210> 6		
<211> 26		
<212> DNA		
<213> 智人		
<400> 6		
ttgaggggaag gtgccaaccg tgagat		26
<210> 7		
<211> 20		
<212> DNA		
<213> 智人		
<400> 7		
tcttccatgc tccgtcacac		20
<210> 8		
<211> 21		
<212> DNA		
<213> 智人		
<400> 8		
cgaatctcaa agtctacgcc g		21
<210> 9		
<211> 25		
<212> DNA		
<213> 智人		
<400> 9		
agccaggccc agaggataca ggaaa		25
<210> 10		
<211> 18		
<212> DNA		
<213> 智人		
<400> 10		
ttccgccgag agaacgag		18
<210> 11		
<211> 18		
<212> DNA		
[0003]		

<213> 智人	
<400> 11 aagaccggca cgaagtgg	18
<210> 12 <211> 26 <212> DNA <213> 智人	
<400> 12 tcggccctga gcaactccat catgta	26
<210> 13 <211> 22 <212> DNA <213> 智人	
<400> 13 tgacagtcac cacgagttcc tg	22
<210> 14 <211> 22 <212> DNA <213> 智人	
<400> 14 tctcctgttc cacctcttca cc	22
<210> 15 <211> 26 <212> DNA <213> 智人	
<400> 15 ctctgggatt cgc metaagg gactat	26
<210> 16 <211> 22 <212> DNA <213> 智人	
<400> 16 ctggctaaca atggtaccga tg	22
<210> 17 <211> 25 <212> DNA <213> 智人	

[0004]

<400> 17		
gtggtctgtg catactgtag aatgg		25
<210> 18		
<211> 24		
<212> DNA		
<213> 智人		
<400> 18		
catgaccaat gcagcagcca ctca		24
<210> 19		
<211> 23		
<212> DNA		
<213> 智人		
<400> 19		
cacgttggat gacacttgtg atc		23
<210> 20		
<211> 19		
<212> DNA		
<213> 智人		
<400> 20		
ctgggagatg gccaatgg		19
<210> 21		
<211> 26		
<212> DNA		
<213> 智人		
<400> 21		
actaataagc agccccccca gacggt		26
<210> 22		
<211> 24		
<212> DNA		
<213> 智人		
<400> 22		
gtgtcattca gtaaagaggc gaag		24
<210> 23		
<211> 22		
<212> DNA		
<213> 智人		
<400> 23		
ctcagccctt tatcattcac gc		22
[0005]		

<210> 24	
<211> 26	
<212> DNA	
<213> 智人	
<400> 24	
ttccggtcct ggtctgcccc tctata	26
<210> 25	
<211> 21	
<212> DNA	
<213> 智人	
<400> 25	
tgacggagta tgtggctacg c	21
<210> 26	
<211> 21	
<212> DNA	
<213> 智人	
<400> 26	
ccacagacca gatgtcgatg g	21
<210> 27	
<211> 24	
<212> DNA	
<213> 智人	
<400> 27	
ctggtaccgg gccccagaga tcat	24
<210> 28	
<211> 20	
<212> DNA	
<213> 智人	
<400> 28	
taacggtctg caccgtagcc	20
<210> 29	
<211> 22	
<212> DNA	
<213> 智人	
<400> 29	
caggccaaag tcacagatct tg	22

[0006]

<210> 30		
<211> 26		
<212> DNA		
<213> 智人		
<400> 30		
acctgctgct caacaccacc tgtgat		26
<210> 31		
<211> 19		
<212> DNA		
<213> 智人		
<400> 31		
aggcgtgctc cctgatgac		19
<210> 32		
<211> 23		
<212> DNA		
<213> 智人		
<400> 32		
gctccaggtc gttcaggtag tag		23
<210> 33		
<211> 21		
<212> DNA		
<213> 智人		
<400> 33		
aggcctgctt tggccgctca a		21
<210> 34		
<211> 21		
<212> DNA		
<213> 智人		
<400> 34		
gactatgtgc cgagcgatca g		21
<210> 35		
<211> 22		
<212> DNA		
<213> 智人		
<400> 35		
gtccacctgg aacttggctct ca		22
<210> 36		
<211> 22		
[0007]		

<212> DNA	
<213> 智人	
<400> 36	22
ctgcttcgct gccgtgtcct ga	
<210> 37	
<211> 22	
<212> DNA	
<213> 智人	
<400> 37	22
tccttggtcg aacagttttt tc	
<210> 38	
<211> 20	
<212> DNA	
<213> 智人	
<400> 38	20
tttgggaggt ggtcctgttg	
<210> 39	
<211> 29	
<212> DNA	
<213> 智人	
<400> 39	29
tgtaagctct cctccatcca gtcctcaa	
<210> 40	
<211> 23	
<212> DNA	
<213> 智人	
<400> 40	23
gatggcccta aacagatgaa gtg	
<210> 41	
<211> 21	
<212> DNA	
<213> 智人	
<400> 41	21
cctgaagccc ttgctgtagt g	
<210> 42	
<211> 25	
<212> DNA	
<213> 智人	

[0008]

<400> 42 atggcggcat ccagctacga atctc	25
<210> 43 <211> 22 <212> DNA <213> 智人	
<400> 43 agccactcac ctcttcagaa cg	22
<210> 44 <211> 22 <212> DNA <213> 智人	
<400> 44 catgtctcct ttctcagggc tg	22
<210> 45 <211> 26 <212> DNA <213> 智人	
<400> 45 caaattcggg acatcctcga cggcat	26
<210> 46 <211> 21 <212> DNA <213> 智人	
<400> 46 ctgctagcca ggatccacaa g	21
<210> 47 <211> 26 <212> DNA <213> 智人	
<400> 47 ctgtgaggta agatggtggc taatac	26
<210> 48 <211> 29 <212> DNA <213> 智人	
<400> 48	

[0009]

ctgttcac tgtgccttg tttctcctt	29
<210> 49	
<211> 28	
<212> DNA	
<213> 智人	
<400> 49	
gcttcgagaa agagttgaga agttaaac	28
<210> 50	
<211> 21	
<212> DNA	
<213> 智人	
<400> 50	
gacctttgcc ccacacatat g	21
<210> 51	
<211> 27	
<212> DNA	
<213> 智人	
<400> 51	
ctcacagacc acaagtcaca ggcctt	27
<210> 52	
<211> 20	
<212> DNA	
<213> 智人	
<400> 52	
cggcaggagc tgaacaagac	20
<210> 53	
<211> 22	
<212> DNA	
<213> 智人	
<400> 53	
agcagcacac acagagccat ag	22
<210> 54	
<211> 26	
<212> DNA	
<213> 智人	
<400> 54	
ccgagcgtta ccagaacctg tctcca	26
[0010]	

<210> 55	
<211> 22	
<212> DNA	
<213> 智人	
<400> 55	
ccaacaccgg tacatcaatg tc	22
<210> 56	
<211> 26	
<212> DNA	
<213> 智人	
<400> 56	
cactcttcta ttgtgtgttc cctttc	26
<210> 57	
<211> 28	
<212> DNA	
<213> 智人	
<400> 57	
caccaccaaa gatccctgac aagcagtt	28
<210> 58	
<211> 17	
<212> DNA	
<213> 智人	
<400> 58	
gccaggcagg catttcc	17
<210> 59	
<211> 20	
<212> DNA	
<213> 智人	
<400> 59	
atgcttcgcc tctgcttcac	20
<210> 60	
<211> 27	
<212> DNA	
<213> 智人	
<400> 60	
tcagccacca tctgccttgc ttacctt	27
<210> 61	
[0011]	

<211> 20	
<212> DNA	
<213> 智人	
<400> 61	
agcccagagg aagggacaac	20
<210> 62	
<211> 20	
<212> DNA	
<213> 智人	
<400> 62	
tgtgagcgct cgtgagattg	20
<210> 63	
<211> 27	
<212> DNA	
<213> 智人	
<400> 63	
ctcctgcaaa agaaccctcg gtcaaca	27
<210> 64	
<211> 21	
<212> DNA	
<213> 智人	
<400> 64	
ccatgtagga agcggctgta c	21
<210> 65	
<211> 20	
<212> DNA	
<213> 智人	
<400> 65	
tcagcccca tgtcaaaaac	20
<210> 66	
<211> 27	
<212> DNA	
<213> 智人	
<400> 66	
atcctgagac cttcgtgcag gcaatct	27
<210> 67	
<211> 22	
<212> DNA	
[0012]	

<213>	智人	
<400>	67	
	gctgtcgctc ccatatttat cc	22
<210>	68	
<211>	22	
<212>	DNA	
<213>	智人	
<400>	68	
	cacaatggac tcgcaattet tc	22
<210>	69	
<211>	28	
<212>	DNA	
<213>	智人	
<400>	69	
	ctgtcagccc tgtgtggtca acgaatac	28
<210>	70	
<211>	20	
<212>	DNA	
<213>	智人	
<400>	70	
	ctggtctgga aagagctggg	20
<210>	71	
<211>	20	
<212>	DNA	
<213>	智人	
<400>	71	
	cagcaggaga tccaccaagg	20
<210>	72	
<211>	23	
<212>	DNA	
<213>	智人	
<400>	72	
	ccccatcttc tttggtgccc gac	23
<210>	73	
<211>	24	
<212>	DNA	
<213>	智人	
[0013]		

<400> 73 gggaatatga ctacgtgacc aatg	24
<210> 74 <211> 24 <212> DNA <213> 智人	
<400> 74 ggatccctga agtcaatggt gatc	24
<210> 75 <211> 26 <212> DNA <213> 智人	
<400> 75 cattcaagac gaatcgccag tctccc	26
<210> 76 <211> 21 <212> DNA <213> 智人	
<400> 76 gattggaaga cccgtttgag c	21
<210> 77 <211> 22 <212> DNA <213> 智人	
<400> 77 caggagaagg cttgatgaaa gc	22
<210> 78 <211> 22 <212> DNA <213> 智人	
<400> 78 ctgggaagcc caaaaccggc aa	22
<210> 79 <211> 21 <212> DNA <213> 智人	
<400> 79 aggagttccc tggatttttg g	21

[0014]

<210> 80	
<211> 22	
<212> DNA	
<213> 智人	
<400> 80	
gcccactttg ccactctctac ac	22
<210> 81	
<211> 28	
<212> DNA	
<213> 智人	
<400> 81	
caaaaagacc ctctggctgt ggacaaaa	28
<210> 82	
<211> 23	
<212> DNA	
<213> 智人	
<400> 82	
catggctgag atgaggaatg aag	23
<210> 83	
<211> 19	
<212> DNA	
<213> 智人	
<400> 83	
gctggcatgt tggctatcg	19
<210> 84	
<211> 25	
<212> DNA	
<213> 智人	
<400> 84	
acgcaggtcc cagcctctctc ttcat	25
<210> 85	
<211> 23	
<212> DNA	
<213> 智人	
<400> 85	
tggattcgtc aatgatgcct atc	23

[0015]

<210> 86  
 <211> 18  
 <212> DNA  
 <213> 智人

<400> 86  
 atgccacatc cgcaatgg 18

<210> 87  
 <211> 24  
 <212> DNA  
 <213> 智人

<400> 87  
 agacctgcgg cacgtgtecg tcta 24

<210> 88  
 <211> 1566  
 <212> DNA  
 <213> 智人

<400> 88  
 ggccccgttaa gaagagcgtg gccggcccgc gccaccgctg gccccaggga aagccgagcg 60  
 gccaccgagc cggcagagac ccaccgagcg gcggcggagg gacgagcgcc ggggcgcacg 120  
 agggcaccat ggcccagacg cccgccttcg acaagcccaa agtagaactg catgtccacc 180  
 tagacggatc catcaagcct gaaaccatct tatactatgg caggaggaga gggatcggcc 240  
 tcccagctaa cacagcagag gggctgctga acgtcattgg catggacaag ccgtcacc 300  
 ttccagactt cctggccaag ttgactact acatgcctgc tatcggggc tgccgggagg 360  
 ctatcaaaag gatcgcctat gagttttag agatgaaggc caaagagggc gtggtgtatg 420  
 tggagggtgc gtacagtcg cacctgctgg ccaactcaa agtggagcca atcccctgga 480  
 accaggctga aggggacctc accccagacg aggtggtggc cctagtgggc cagggcctgc 540  
 aggaggggga gcgagacttc ggggtcaagg cccggtccat cctgtgctgc atgcgccacc 600  
 agcccaactg gtcccccaag gtggtggagc tgtgtaagaa gtaccagcag cagaccgtgg 660  
 tagccattga cctggctgga gatgagacca tcccaggaag cagcctcttg cctggacatg 720  
 tccaggccta ccaggaggct gtgaagagcg gcattcaccg tactgtccac gccggggagg 780  
 tgggctcggc cgaagtagta aaagaggctg tggacatact caagacagag cggctgggac 840  
 acggctacca cacctggaa gaccaggccc ttataacag gctgcggcag gaaaacatgc 900  
 acttcgagat ctccccctgg tccagctacc tcaactggtgc ctggaagccg gacacggagc 960

[0016]



aaccgagaga agcggggcct cgccttggac gggaaagtca agcacgaaga cacgaacttg	900
gcctctagca ccctgttgag ggaaggtgcc aaccgtgaga tcctggggat cattgtttcc	960
tacaaagtga aagtgaagct ggtggtgtct cggggcggcc tgttgggaga tcttgcattc	1020
agcgacgtgg ccgtggaact gcccttcacc ctaatgcacc ccaagcccaa agaggaaccc	1080
ccgcatcggg aagtccaga gaacgagacg ccagtagata ccaatctcat agaacttgac	1140
acaaatgatg acgacattgt atttgaggac tttgctgcc agagactgaa aggcatggag	1200
gatgacaagg aggaagagga ggatggtacc ggctctccgc ggctcaacga caga	1254

<210> 90  
 <211> 1770  
 <212> DNA  
 <213> 智人

<400> 90	
tggcagcggg cgaggaggct gcgagcgcgc gcggaaccga gcgggcggcg ggcgcgcgca	60
ccatggggga gaaaccggg accagggtct tcaagaagtc gagccctaac tgcaagctca	120
ccgtgtactt gggcaagcgg gacttcgtag atcacctgga caaagtggac cctgtagatg	180
gcgtggtgct tgtggaccct gactacctga aggaccgcaa agtgtttgtg accctcacct	240
gcgccttccg ctatggcgt gaagacctgg atgtgctggg cttgtccttc cgcaaagacc	300
tgttcatcgc cacctaccag gccttcccc eggtgcccaa cccaccccg cccccaccc	360
gcctgcagga ccggctgctg aggaagctgg gccagcatgc ccacccttc ttcttacca	420
tacccagaa tcttccatgc tccgtcacac tgcagccagg cccagaggat acaggaaagg	480
cctgcggcgt agactttgag attcgagcct tetgtgctaa atcactagaa gagaaaagcc	540
acaaaaggaa ctctgtgcgg ctggtgatcc gaaaggtgca gttcgcccc gagaaaccg	600
gccccagcc ttcagccgaa accacacgcc acttctcat gtctgaccgg tcctgcacc	660
tcgaggcttc cctggacaag gagctgtact accatgggga gccctcaat gtaaagtcc	720
acgtcaccaa caactccacc aagaccgtca agaagatcaa agtctctgtg agacagtacg	780
ccgacatctg cctcttcagc accgcccagt acaaggttcc tgtggtcaa ctcgaacaag	840
atgaccaggt atctcccagc tccacattct gtaaggtgta caccataacc ccaactgtca	900
gtgacaaccg ggagaagcgg ggtctcgccc tggatgggaa actcaagcac gaggacacca	960
acctggcttc cagcaccatc gtgaaggagg gtgccaacaa ggaggtgctg ggaatcctgg	1020

[0018]

tgtcctacag ggtcaagggtg aagctgggtg tgtctcgagg cgggatgtc tctgtggagc	1080
tgccttttgt tcttatgcac cccaageccc acgaccacat cccctcccc agaccccagt	1140
cagccgctcc ggagacagat gtccctgtgg acaccaacct cattgaattt gataccaact	1200
atgccacaga tgatgacatt gtgtttgagg actttgcccg gcttcggctg aaggggatga	1260
aggatgacga ctatgatgat caactctgct aggaagcggg gtgggaagaa gggaggggat	1320
ggggttggga gaggtgaggg caggattaag atccccactg tcaatggggg attgtcccag	1380
cccctcttcc cttcccctca cctggaaget tcttcaacca atcccttcac actctctccc	1440
ccatcccccc aagatacaca ctggaccctc tcttgcagaa tgtgggcatt aattttttga	1500
ctgcagctct gcttctccag ccccgccgtg ggtggcaagc tgtgttcata cctaaatttt	1560
ctggaagggg acagtgaaaa gaggagtgc aggagggaaa gggggagaca aaactcctac	1620
tctcaacctc acaccaacac ctcccattat cactctctct gccccattc cttcaagagg	1680
agaccctttg gggacaaggc cgtttctttg tttctgagca taaagaagaa aataaatctt	1740
ttactaagca tgaaaaaaaa aaaaaaaaaa	1770

<210> 91  
 <211> 1975  
 <212> DNA  
 <213> 智人

<400> 91	
actttcccc ctcggegecc caccggetcc cgcgcgcctc cctcgcgcc cgagcttcca	60
gccaaagcagc gtcctgggga gcgcgtcatg gccttaccag tgaccgcctt gctcctgccg	120
ctggccttgc tgctccacgc cggcaggccg agccagttcc ggggtgctgcc gctggatcgg	180
acctggaacc tgggcgagac agtggagctg aagtgccagg tgctgctgtc caaccgcagc	240
tegggtctgt cgtggetctt ccagccgcgc ggcgcgcccg ccagtcccac cttctctcta	300
tacctctccc aaaacaagcc caaggcggcc gaggggctgg acaccagcg gttctcgggc	360
aagaggttgg gggacacctt cgtcctcacc ctgagcgact tccgccgaga gaacgagggc	420
tactatctct gctcggccct gagcaactcc atcatgtact tcagccactt cgtgccggtc	480
ttcctgccag cgaagcccac cagcagcca gcgcgcgac caccaacacc ggcgccacc	540
atcgcgtcgc agcccctgtc cctgcgcca gaggcgtgcc ggccagcggc ggggggcgca	600
gtgcacacga gggggctgga cttegcctgt gatatctaca tctgggcgcc cctggccggg	660
acttggtggg tccttctcct gtcactggtt atcacccttt actgcaacca caggaaccga	720

[0019]

agacgtgttt gcaaatgtcc cggcctgtg gtcaaatcgg gagacaagcc cagcctttcg	780
gcgagatag tctaacctg tgcaacagcc actacattac ttcaaactga gatccttct	840
tttgaggag caagtcttc cctttcattt ttccagctc tctcctgt gtattcattt	900
tcatgattat tatttttagtg ggggcggggt gggaaagatt actttttctt tatgtgtttg	960
acgggaaca aaactaggta aaatctacag tacaccacaa gggtcacaat actgtttgtc	1020
gcacatcgcg gtagggcgtg gaaaggggca ggccagagct acccgcagag ttctcagaat	1080
catgctgaga gagctggagg cacccatgcc atctcaacct ctccccgcc cgttttacia	1140
agggggaggc taaagcccag agacagcttg atcaaaggca cacagcaagt cagggttgg	1200
gcagtagctg gagggacctt gtctcccagc tcagggtctt ttctccaca ccattcaggt	1260
ctttcttctc gaggccctg tctcagggtg aggtgcttga gtctccaacg gcaagggaac	1320
aagtacttct tgatacctgg gatactgtgc ccagagctc gaggaggtaa tgaattaaag	1380
aagagaactg cctttggcag agttctataa tgtaacaat atcagacttt tttttttat	1440
aatcaagcct aaaattgtat agacctaaaa taaaatgaag tggtagctt aaccctggaa	1500
aatgaatccc tctatctcta aagaaaatct ctgtgaaacc cctacgtgga ggcggaattg	1560
ctctcccagc ccttgcttg cagaggggcc catgaaagag gacaggctac ccctttacia	1620
atagaatttg agcatcagtg aggttaaact aaggccctct tgaatctctg aatttgagat	1680
acaacatgt tcttgggatc actgatgact tttatactt tgtaaagaca attgttggag	1740
agcccctcac acagccctgg cctccgctca actagcagat acagggatga ggcagacctg	1800
actctcttaa ggagctgag agcccaaact gctgtcccaa acatgcactt ccttgcttaa	1860
ggtatggtac aagcaatgcc tgcccattgg agagaaaaaa cttaagtaga taaggaaata	1920
agaaccactc ataattcttc accttaggaa taatctctg ttaatatggt gtaca	1975

<210> 92  
 <211> 1411  
 <212> DNA  
 <213> 智人

<400> 92	
gcgactgtct ccgccgagcc cccggggcca ggtgtcccgg gcgcgccacg atgcggccgc	60
ggctgtggct cctcctggcc ggcagctga cagttctcca tggcaactca gtcctccagc	120
agaccctgc atacataaag gtgcaaacca acaagatggt gatgctgtcc tgcgaggcta	180

[0020]

aaatctcct cagtaacatg cgcatctact ggctgagaca ggccaggca cggagcagtg	240
acagtcacca cgagttcctg gccctctggg attccgcaaa agggactatc cacggtgaag	300
aggtggaaca ggagaagata gctgtgtttc gggatgcaag ccggttcatt ctcaatctca	360
caagcgtgaa gccggaagac agtggcatct acttctgcat gatcgtcggg agccccgagc	420
tgacctcgg gaagggaact cagctgagtg tggttgattt cttcccacc actgcccagc	480
ccaccaagaa gtccaccctc aagaagagag tgtgccggtt acccaggcca gagaccaga	540
agggeccact ttgtagcccc atcaccttg gcctgctggt ggctggcgtc ctggttctgc	600
tggtttcct gggagtggcc atccacctgt gctgccggcg gaggagagcc cggttcggt	660
tcatgaaaca atttacaaa taagcagaga atacggtttt ggtgtcctgc taaaaaaga	720
catcggtcag taatgagcac gatgtgaaa aatgagagaa gggacacatt caaccctgga	780
gagttcaatg gctgctgaag ctgcctgctt ttcactgctg caaggccttt ctgtgtgtga	840
tgtgcatggg agcaactgt tegtgggtca tcgggaatac tagggagaag gtttcattgc	900
ccccaggga cttcacagag tgtgctggag gactgagtaa gaaatgctgc ccatgccacc	960
gcttccggt cctgtgctt ccctgaactg ggaccttag tggtgccat ttagccacca	1020
tctttgcagg ttgctttgcc ctggtagggc agtaacattg ggtcctgggt ctttcatggg	1080
gtgatgctgg gctggctccc tcttggctt cccaggctgg ggctgacctt cctcgcagag	1140
aggccagtg caggttggga atgaggctt ctgagagggg ctgtccagtt cccagaagc	1200
atatcagtct ctgaggctt cctttggggc cgggaacttg cgggtttgag gataggagtt	1260
cacttcatct tctcagctc catttctact cttaagtttc tcccatttc tactcttaag	1320
tttctcagct ccattteta ctctccatg gcttcatget tctttcattt ttctgtttgt	1380
tttatacaaa tgtgttagtt gtacaaataa a	1411

<210> 93

<211> 9794

<212> DNA

<213> 智人

<400> 93

tgtttccgtg cgcggccgct gcgcactcgg cactgggcgg cgctggctgg ctccctggct	60
gcggctcctc agtcggcggc ggtgctgct gcctgtggcc cgggcggctg ggagaagcgg	120
agtgttggtg agtgacgcgg cggaggtgta gttgacgcg gtgtgttacg tgggggagag	180
aataaaactc cagcgagatc cgggccgtga acgaaagcag tgacggagga gcttgtacca	240

[0021]

ccggtaacta aatgaccatg gaatctggag ccgagaacca gcagagtgga gatgcagctg	300
taacagaagc tgaaaaccaa caaatgacag ttcaagccca gccacagatt gccacattag	360
cccaggtatc tatgccagca gtcctatgcaa catcatctgc tcccaccgta actctagtac	420
agctgcccac tgggcagaca gttcaagtcc atggagtcac tcaggcggcc cagccatcag	480
ttattcagtc tccacaagtc caaacagttc agtcttctg taaggactta aaaagacttt	540
tctccggaac acagatttca actattgcag aaagtgaaga ttcacaggag tcagtggata	600
gtgtaactga ttcccaaaag cgaagggaaa ttctttcaag gaggccttcc tacaggaaaa	660
ttttgaatga cttatcttct gatgcaccag gagtgccaa gattgaagaa gagaagtctg	720
aagaggagac ttcagcacct gccatcacca ctgtaacggt gccaaactcca atttaccaaa	780
ctagcagtg acagtatatt gccattacc agggaggagc aatacagctg gctaacaatg	840
gtaccgatgg ggtacagggc ctgcaaacat taacctgac caatgcagca gccactcagc	900
cgggtactac cattctacag tatgcacaga ccaactgatg acagcagatc ttagtgccca	960
gcaaccaagt tgtttgtcaa gctgcctctg gagacgtaca aacataccag attcgcacag	1020
cacccactag cactattgcc cctggagttg ttatggcacc ctcccagca ctctctacac	1080
agcctgctga agaagcagca cgaaagagag aggtccgtct aatgaagaac agggaagcag	1140
ctcgagagtg tcgtagaaag aagaaagaat atgtgaaatg tttagaaaac agagtggcag	1200
tgettgaaaa tcaaaacaag acattgattg aggagctaaa agcacttaag gacctttact	1260
gccacaaatc agattaattt gggattttaa tttcacctg ttaaggtgga aatggactg	1320
gcttgccac aacctgaaag acaaaaataa cattttattt tctaaacatt tcttttttc	1380
tatgcgcaaa actgcctgaa agcaactaca gaatttcatt catttgtgct tttgcattaa	1440
actgtgaatg ttccaacacc tgctccact tctcccctca agaaattttc aacgccagga	1500
atcatgaaga gacttctgct ttcaacccc caccctctc aagaagtaat aatttgttta	1560
cttgtaaatt gatgggagaa atgaggaaaa gaaaatcttt ttaaaaatga tttcaaggtt	1620
tgtgctgagc tccttgattg ccttagggac agaattacc cagcctcttg agctgaagta	1680
atgtgtgggc cgcctgcata aagtaagtaa ggtgcaatga agaagtgtg attgccaat	1740
tgacatgttg tcacattctc attgtgaatt atgtaaagtt gtaagagac ataccctcta	1800
aaaagaact ttagcatggt attgaaggaa ttagaaatga atttgagtg ctttttatgt	1860
atgttgctct ctccaatact gaaaatttgt ccttggttct taaaagcatt ctgtactaat	1920

[0022]

acagctcttc	catagggcag	ttgttgettc	ttaattcagt	tctgtatgtg	ttcaacattt	1980
ttgaatacat	taaaagaagt	aaccaactga	acgacaaagc	atggtatttg	aattttaaat	2040
taaagcaaag	taaataaaag	tacaaagcat	attttagtta	gtactaaatt	cttagtaaaa	2100
tgctgatcag	taaaccaate	ccttgagtta	tataacaaga	tttttaaata	aatgttattg	2160
tcctcacctt	caaaaatatt	tatattgtca	ctcatttacg	taaaaagata	tttctaattt	2220
actgttgccc	attgcactta	cataccacca	ccaagaaagc	cttcaagatg	tcaaataaag	2280
caaagtgata	tatatttgtt	tatgaaatgt	tacatgtaga	aaaatactga	ttttaaatat	2340
tttccatatt	aacaatttaa	cagagaatct	ctagtgaatt	ttttaaatga	aagaagttgt	2400
aaggatataa	aaagtacagt	gtagatgtg	cacaaggaaa	gttattttca	gacatatttg	2460
aatgactgct	gtactgcaat	atgttgattg	tcattcttac	aaaacatttt	ttgtttctct	2520
tgtaaaaaga	gtagttatta	gttctgcttt	agctttccaa	tatgctgtat	agcctttgtc	2580
attttataat	tttaattcct	gattaaaaca	gtctgtattt	gtgtatatca	tacattgttt	2640
tcaataccac	ttttaattgt	tactcatttt	attcactaag	ctcgataaat	ctaacagtta	2700
ctcttaaaaa	aaaaaaaaaa	agactaaggt	ggattttaaa	aattggaaac	tgacataatg	2760
ttaggttata	atctctcatt	tggagccggg	cgcagtggtc	cacgcctgta	atcccagcac	2820
tttgggaggc	caaggtgggt	ggatcacctg	tggtcaagag	ttcaagacca	gcctggccat	2880
catggtgaaa	ccccatctct	actaaaaata	caaaaattag	ccaggcgtgg	tggtggcgcg	2940
ctgtaatccc	agctactcag	gaggttgagg	cagcagaatt	gcttgaacce	aggaggcaga	3000
gggttgcaat	gagccgagat	agcaccattg	cactccagcc	tgggcgactc	catctcaaaa	3060
aataaaaaata	aaaaaaatgt	ctcatttggg	aaggaaattc	cttttaaaaa	agagttgaga	3120
cacttagaaa	actaatgttt	tatatttagt	caagagttat	ttaagaaagt	caagcttggt	3180
taacaacaaa	atatgaagat	ttaagtgtta	attgctggat	ccattttaaa	ataagatttt	3240
aattaacatt	tgtaaattgt	atattttcgt	ttgtaacaaa	ccattgtcct	ttttcaagga	3300
tgaacagagt	ttatgaagga	gcacattctt	aagaattaag	tgatgtagtc	tttatgtttg	3360
gacagttcac	cagatttctca	agaaggcttt	caaacaacta	taaagtttga	tgtttgcctt	3420
gctgagctaa	tggggaaagt	tatagcataa	aaattgtgta	accgcataga	tatgtcattt	3480
ttaaaaactg	gtttaacaga	aatcaagcaa	agtcacaaat	atgttcacaa	gttggaatta	3540
tttattgagt	caaatgtcgc	aatcgaacat	tttgaatgaa	gtaagtgtta	taaatgaaaa	3600

[0023]

attgcctgat gtttagcagt ttgtattctc taaagctttt ttcaaaagt tcaggctttc	3660
tacttactgg gaagttgggtg gtccctcttag tcctgataa atcaaggcaa tcacattcat	3720
gtgagctgga tgaatttata agttataaag acctatcct tcataccttg aggatgattg	3780
cactggtttt gaagtcagtt gcttaatgat gaggtgagaa atgtatcctg ttgctaaatc	3840
tgtcttagac ccttggtgaa acttgaagat ttcagtttat aaagataaaa tcaagcatct	3900
tttgtgcagt tttctttttt taatgcaaga atgggtggga ggtttgtttg taagcatgaa	3960
accttgagaa tctttattaa gaaaatgaca taatttttaa aaaccttgta gccagaaca	4020
tatgtggcca cattaccagt aataaatggt tttctcttta tattggccaa aagggaataa	4080
aatgtcatc ataggaattt gtacatatgc tactgatttg cctagaaaat agcaagtttg	4140
atattgctca ctttgcaaat ataggccat gtggcacttt tatctatagg acagattaat	4200
aaaaatgaag tggggagggg tttatttttg atatattact cttatgagtt ttcaagcttt	4260
gataatgttt aactgaaaag tggcttagaa agggctagat ccaatgtgtt cattattaaa	4320
taattgctat cagatacaat ttaagtcca tctttttca aactcaagta ccatattggc	4380
aaccataata ttgtcatagg tgctctcttc attagatat tcttgggggg ggtggcattt	4440
gtataatata tgtgtacata tatatatata tatatatata tacatacagt atataatcta	4500
aagctctgag agctcttaag tcaggaatgc tgagtattat agtatattga ggtcagatga	4560
aattttacat ttttgtgtgt tctgttgeat tcttctgggt agtttctatg actgcattac	4620
tccagcactc atgattgatt ttatctteta atttcttcc aagtatttta tttttatta	4680
gttttctttg gcttgatact tttaaatatg ttactagtca cttgaaagcc tctccccaa	4740
aagtatttgg tttgtatgct ttgtctgtgg cagctataac agtggtgaaga acattttgaa	4800
gatagctttt taaaggaacc actgattttt tcaaaaatca tcctggggga ggaattttgg	4860
catttcattt gagcagggat tttgtcagaa aatgtgtttt gatggtaggt cagcagcagt	4920
gctagtctct gaaagcacia taccagtcag gcagcctatc ccatcagatg tcatctggct	4980
gaagtttata tctgtctctc aggataaatc cctgtaggac aaatccctac tatcatttct	5040
accttttggg gtgacatgtg gaatcataca aaggcttagg aagaaatacg tttgtttaa	5100
ccaggatgct ttacttactt gaagtgactt caatctagat ttcttttaat atttaacaaa	5160
tttttaattc tatgatcagc cacagtcagc tattaccata aattggctctc tgtttatttt	5220
gaagatcacg gctgcttcat tttgcaggat taagtagggc taatgtatct taaagttaag	5280

[0024]

atcttgaatt aaagtgagtt ttagaaatag tgttacatac cttttcagtt gttttcaaga	5340
ggctttatntt ttgttgcctt tgtagccctg aaagctgttg gtatatnttt tccctcatgg	5400
acccaataga aaagttgtat atnttatnttg attatatntta cattctgtcc tttgtaaagt	5460
tttggtgtaa ctngcactntt tntaaatgac ccagnttggg tattagcaac ttaagaaant	5520
ccctcatcaa gtaattctca actntnttagt ctttctctc tcttcaaate atgtgactnt	5580
ttaaattgaa gntntntcatt gattaaaata tnttagcacc taaaagctag ccttaaaaac	5640
agctgtaaaa gaaaaacatc aggaaattag atatgactag cccagntaat taaaagacgg	5700
gctcaaacct tgtntntatc tntntcatct tggatgaaga ttgaaggga aataactcaa	5760
gtgcataata tntatntntca atntntaatg agactnttate ctcatcacia cattaact	5820
gtacatagta tgccaaaata tccattaant tgtctagaat agtacaagac tntntaaagc	5880
aattgtctc acagagacca catgtaatat actgaaatat gntcattntt aatggctntg	5940
ttaacatcaa agaaatgctg cctaaattg atntcagatg aggaaggaga aagtaaagt	6000
tgcatagtaa ggctgtaggt gaagagnttgt gagataaata gntcactcag ttgtacaaag	6060
cacaactaga actntntgntt gggaggctta catacatctt gaatattctt aatgtaataa	6120
tgttgactat taagntggct acacagtcac tgtatgtact aggaactgg ttccttgaca	6180
ttctagaate aatggctagg agaggcatta atctntgagg ggctgaacat atcatgaagc	6240
tgagtcagta tggaaaatnt tcaaataaac aggggtctga agntccatct gtctcatctg	6300
cttatgataa gntctntatg attagtgaat gtagcttaag cctntgtatg tgtctcagg	6360
gggcagaccg actnttaagag ggaccagata acgntntgaat ggagggatta tntntcaggt	6420
gntnttagctt gaaatnttatt tnttaaaaaa agaaaaatnt aaaaaatata taaataaant	6480
agaacaaagc cggtgatgca agntgatatt ataaacaggc agntnttagca cagaaagaaa	6540
atactgacct gtctgcattc tggtagcgtg ggtgcaggtc ccagctgggt atgacatgat	6600
acattntntaa ttattctcac cagcaagtaa aaggaaaatg aacaatctnt tggaaatntc	6660
tttgaaaagg atcaaagagt aggaaattca cattngacct aacattactt gcctatagaa	6720
gtatggcatt tccaagctnt tgtctgagga gcactcaga gaagtgagag taaatctgag	6780
ttagctntaa aatntgtagg gaggaagaaa atctctgcaa ataatgatt tatgntntgnt	6840
ggccaagtga aatgatctat cattngntt gggagntntt atntntctat gntntntaaa	6900
ttggtaaant ctttatagat gtatntntat ccaagtgcca ctccaatntg tgtatgtaat	6960

[0025]

aaaattat	ttt	atattaaaag	tg	ggaaataa	ttgtcaacat	tttttttgag	tatagattta	7020
ttaggggtg	g	caaagaagag	tgctagttag	cagttttcca	tgtaaagttg	tccttgactg		7080
at	ttgtccac	atgtcagttg	taactcccc	actccctgca	aaaggaatta	tttctaacc		7140
agatgtatca	cttgaaactt	tttagaagca	aaataatcag	ggaagttcct	agaaaggtgt			7200
ttggcttttt	ggtttttgag	ggttgggta	aagaagactt	ccccacaac	tgtcagcaca			7260
aaacagggtta	ttgattttta	actctgatgt	ttctattgga	gttgaatact	aaataaataa			7320
ctataatgag	ggaaatacat	ttctaataaa	attccctaca	ttctagaaac	atccctgttt			7380
taat	tttttt	atctaaatct	ttt	gtgctt	tatgtgtaaa	gaaaaaaatg	tactgagtta	7440
caatgcattt	tattaacact	atgtacataa	tagctgcttt	gtgttcagaa	tagtagcagt			7500
tgctttgtat	attaagtg	tccttgta	ttt	gtgaaat	attgtcataa	agtgttttt		7560
cttactgtaa	tctttgtggt	atcaactgtc	ataatgctct	ttttacacaa	acatttatgt			7620
gcagtcacat	aaacatgctt	ttaaaaactc	tgtaagtctc	ttttttggg	atgggatctc			7680
tatattttgt	tggtttttt	ttgctagtag	tgtgaagcca	tgttttattg	gacttaaagt			7740
tacaatatat	tacaagcttg	tg	ttggaagg	cagcaaaact	aattcagaca	acaacatgtc		7800
ttcagttact	ggatccctaa	tttcaggac	aaaacctggt	tttcaataag	attgaacagt			7860
gcctat	ttt	gtg	ttt	ggag	atg	tactgt		7920
tgtgtctgat	gtcataaaac	acgtgttcac	tgaaaggaca	ataagactat	ataccttctc			7980
aggtccccctt	gcaattctaa	aactctgtga	tcatataaat	tggaaggaaa	ggggagggga			8040
tatggttaat	ctttgcttaa	gctgtaagaa	taaaaagtt	atctcctata	ctattaactt			8100
ctgaaataag	ttctgagacg	agacatctga	aaataagcag	ctgcattatt	tgtatgtttc			8160
ttcactgcca	agatgtgttc	aagcctgcta	tacctgcat	tgtattgga	ggcttaatga			8220
at	ttcattta	ttttctgcaa	caacgattac	agaatttatt	gcacaaaatg	agacattttg		8280
agagtgat	at	taattacatg	agggacaata	ggcatgaact	aggattgttc	taagcaaatc		8340
ggaatcgggt	cacctgcca	cgttcaggtg	cttggacctt	caggaaaaga	ttgcccattc			8400
tg	tcatttga	ccaggcactg	aagtgacaag	accatccttg	agaagtcaca	tccaaagata		8460
aaattctgat	ccatttctag	ttttagtggt	tcgccactga	agacttaaca	tatgtctttt			8520
acactcaggt	tgcaaaacac	aggcccaaga	caaacttaac	ttctcccca	aatcttcctt			8580
ccgctggttt	ttccatctcg	taagtgggtgc	cactatccat	ctgttaaatt	gtttagggga			8640

[0026]

aacctagaaa agcactacct taatcagtg tatecttctt cttactgtg cgtcctaatt	8700
tctccacatc tttcttaagt gcagtgacca aaccggatga gaattctaac acgggcctga	8760
catcaaatgg aaaggaagga taatgtccag gagttggaat gttatccttg tttttaatta	8820
agatgcaatt cacataaatt aactttttaa gtgaacaatt aagtggtagt acatccacaa	8880
tgggtgatac ccaccacttc tatctagctc caaacattc tcatcactcc aaaagtaaag	8940
tcccgttact ctccattttc tctcccacc gcccttgtec ctggcaacca ccaatctgct	9000
tctgttttct ttggatttac atccgggtat ttcattgtgag actcatacac tgtgtattac	9060
ttctttctgc tagctttaat gtgttgttga ggttgatcca ttgtaacatg ttatcactac	9120
ttcattcctt tttatagcta agtatacttt ttatagtaag tatgccattg tagatatata	9180
ccacaagttt atcgattcat ccagttgagt tgtttctact gtttgctaa tgttcatagt	9240
gctgttatga atgttctgtt acaagtattt gagtccgtgt tttcaattat ttgggtata	9300
tgccctgggag tggagttgct gggctatggt gaaatcgac atttaacttt ttgaggaact	9360
gtcaaacctt cctcagcag ctgtaccgtt ttacctcca ccattgatgt atgagggttc	9420
caatttctcc acaccttcac caaaccttat tttgccattt taaaattat agccatcctc	9480
atgggtgtgg tctctcattg tggttttgat ttgcatttcc ctgattacta atgatgtgga	9540
gcactttttg ttgtctttgg ccatctgcgt atcttctttg aagaaatgct tgttgaggtc	9600
ctttgttcat tgaattttg ttgttgggtt ctgagttcct tatatattct gggactagg	9660
cccttataat attttgcct ataagtttt gctttataat gtcctcattg ttttcaact	9720
tactttatgt aatatgtaca cttctaaaaa aaagaaacat ggaaaagggc aaactgtaa	9780
aaaaaaaaaaaa aaaa	9794

<210> 94  
 <211> 1241  
 <212> DNA  
 <213> 智人

<400> 94	
gaattcggc cgcgccttc tcacggcatt cagcagcagc gttgctgtaa ccgacaaaga	60
caccttcgaa ttaagcacat tctcagatc cagcaaagca ccgcaacatg accgaaatga	120
gcttctctgag cagcaggtg ttgggtggggg acttgatgct ccccttcgac ccgtcgggtt	180
tgggggctga agaaagccta ggtctcttag atgattacct ggagggtggc aagcacttca	240

[0027]

aacctcatgg gttctccagc gacaaggcta aggcgggctc ctccgaatgg ctggctgtgg	300
atgggttggc cagtcctcc aacaacagca aggaggatgc cttctccggg acagattgga	360
tgttgagaa aatggatttg aaggagtctg acttgatgc cctgttgggt atagatgacc	420
tggaaacctat gccagatgac cttctgacca cgttgatga cacttgatg ctctttgccc	480
ccctagtcca ggagactaat aagcagcccc cccagacggg gaaccaatt ggccatctcc	540
cagaaagttt aacaaaacc gaccaggttg ccccttccac cttcttaca cctcttcccc	600
tttccccagg ggtcctgtcc tccactccag atcattcctt tagtttagag ctgggcagtg	660
aagtggatat cactgaagga gataggaagc cagactacac tgcttacgtt gccatgatcc	720
ctcagtgcac aaaggaggaa gacaccctt cagataatga tagtggcatc tgtatgagcc	780
cagagtccta tctgggtct cctcagcaca gccctctac caggggtct ccaaatagga	840
gctcccatc tccagggtgt cttctgtgggt ctgccgtcc caaacctac gatcctcctg	900
gagagaagat ggtagcagca aaagtaaagg gtgagaaact ggataagaag ctgaaaaaaaa	960
tggagcaaaa caagacagca gccactaggt accgccagaa gaagaggcgg gagcaggagg	1020
ctcttactgg tgagtgcaaa gagctggaag agaagaacga ggctctaaaa gagagggcgg	1080
attccctggc caaggagatc cagtacctga aagatttgat agaagaggtc cgcaaggcaa	1140
gggggaagaa aagggtcccc tagttgagga tagtcaggag cgtcaatgtg cttgtacata	1200
gagtgctgta gctgtgtgtt ccaataaatt atttttagg g	1241

<210> 95  
 <211> 2924  
 <212> DNA  
 <213> 智人

<400> 95	
gacgccgacg atgaagacac cgtggaaggt tcttctggga ctgctgggtg ctgctgcgct	60
tgtcaccatc atcacctgac ccgtggttct gctgaacaaa ggcacagatg atgctacagc	120
tgacagtcgc aaaacttaca ctctaactga ttacttaaaa aatacttata gactgaagtt	180
atactcctta agatggattt cagatcatga atatctctac aaacaagaaa ataatatctt	240
ggtattcaat gctgaatatg gaaacagctc agttttcttg gagaacagta catttgatga	300
gtttggacat tctatcaatg attattcaat atctctgat gggcagttta ttctcttaga	360
atacaactac gtgaagcaat ggaggcattc ctacacagct tcatatgaca tttatgattt	420
aaataaaagg cagctgatta cagaagagag gattccaac aacacacagt gggtcacatg	480

[0028]

gtcaccagtg ggtcataaat tggcatatgt ttggaacaat gacatttatg ttaaaattga	540
accaaattta ccaagttaca gaatcacatg gacggggaaa gaagatataa tatataatgg	600
aataactgac tgggtttatg aagaggaagt cttcagtgcc tactctgctc tgtggtggtc	660
tccaaacggc acttttttag catatgceca atttaacgac acagaagtcc cacttattga	720
atactccttc tactctgatg agtcactgca gtacccaaag actgtacggg ttccatatcc	780
aaaggcagga gctgtgaatc caactgtaaa gttctttggt gtaaatacag actctctcag	840
ctcagtcacc aatgcaactt ccatacaaat cactgctcct gcttctatgt tgatagggga	900
tcactacttg tgtgatgtga catgggcaac acaagaaaga atttctttgc agtggctcag	960
gaggattcag aactattcgg tcatggatat ttgtgactat gatgaatcca gtggaagatg	1020
gaactgctta gtggcacggc aacacattga aatgagtact actggctggg ttggaagatt	1080
taggccttca gaacctcatt ttaccettga tggtaatagc ttctacaaga tcatcagcaa	1140
tgaagaaggt tacagacaca tttgctatth ccaaatagat aaaaaagact gcacatttat	1200
tacaaaaggc acctgggaag tcatcgggat agaagctcta accagtgatt atctatacta	1260
cattagtaat gaatataaag gaatgccagg aggaaggaat cttataaaa tccaacttag	1320
tgactataca aaagtgcacat gcctcagttg tgagctgaat ccggaaaggt gtcagtacta	1380
ttctgtgtca ttcagtaaag aggcgaagta ttatcagctg agatgttccg gtctgtgtct	1440
gcccctctat actctacaca gcagcgtgaa tgataaaggg ctgagagtcc tggaagacaa	1500
ttcagctttg gataaaatgc tgcagaatgt ccagatgccc tccaaaaaac tggacttcat	1560
tatthttgaat gaaacaaaat tttggtatca gatgatcttg cctcctcatt ttgataaatc	1620
caagaaatat cctctactat tagatgtgta tgcaggccca tgtagtcaaa aagcagacac	1680
tgtcttcaga ctgaactggg ccacttacct tgcaagcaca gaaaacatta tagtagctag	1740
ctttgatggc agaggaagtg gttaccaagg agataagatc atgcatgcaa tcaacagaag	1800
actgggaaca tttgaagttg aagatcaaat tgaagcagcc agacaatttt caaaaatggg	1860
atthgtggac aacaaacgaa ttgcaatttg gggctggtca tatggagggt acgtaacctc	1920
aatggctctg ggatcaggaa gtggcgtgth caagtgtgga atagccgtgg cgcctgtatc	1980
ccggtgggag tactatgact cagtgtacac agaacgttac atgggtctcc caactccaga	2040
agacaacctt gaccattaca gaaattcaac agtcatgagc agagctgaaa atthtaaaca	2100
agthgagtac ctcttattc atggaacagc agatgataac gthcactthc agcagtcagc	2160

[0029]

tcagatctcc aaagccctgg tcgatgttgg agtggatttc caggcaatgt ggtatactga	2220
tgaagacat ggaatagcta gcagcacagc acaccaacat atatataccc acatgagcca	2280
cttcataaaa caatgtttct ctttacctta gcacctcaaa ataccatgcc atttaaagct	2340
tattaaact catttttgtt ttcattatct caaaactgca ctgtcaagat gatgatgatc	2400
tttaaatac aactcaaat caagaaactt aaggttacct ttgttcccaa atttcatacc	2460
tatcatctta agtagggact tctgtcttca caacagatta ttaccttaca gaagtttgaa	2520
ttatceggtc gggttttatt gtttaaaatc atttctgcat cagetgctga aacaacaaat	2580
aggaattgtt tttatggagg ctttgcatag attccctgag caggatttta atctttttct	2640
aactggactg gttcaaatgt tgttctcttc tttaaagga tggcaagatg tgggcagtga	2700
tgctactagg gcagggacag gataagaggg attagggaga gaagatagca gggcatggct	2760
gggaacccaa gtccaagcat accaacacga ccaggctact gtcagctccc ctcgagaaaa	2820
actgtgcagt ctgcgtgta acagctcttc tccttagag cacaatggat ctcgagggat	2880
cttcatacc taccagttct gcgcctcgag gccgcgactc taga	2924

<210> 96  
 <211> 1745  
 <212> DNA  
 <213> 智人

<400> 96	
ccccgtagaa ccgagggggg gggcccgggg gtcccggggg aggtggagat ggtgaagggg	60
cagccgtteg acgtgggccc gcgctacacg cagttgcagt acatcggcga gggcgcgtac	120
ggcatggtea gctcggecta tgaccacgtg cgcaagactc gcgtggccat caagaagatc	180
agcccctteg aacatcagac ctactgccag cgcacgctcc gggagatcca gatcctgctg	240
cgcttccgcc atgagaatgt catcggcatc cgagacattc tgcgggcgtc caccctggaa	300
gcatgagag atgtctacat tgtgcaggac ctgatggaga ctgacctgta caagttgctg	360
aaaagccagc agctgagcaa tgaccatata tgctacttcc tctaccagat cctgcggggc	420
ctcaagtaca tccactccgc caacgtgctc caccgagatc taaagccctc caacctgctc	480
atcaacacca cctgcgacct taagatttgt gatttcggcc tggcccggat tgccgatcct	540
gagcatgacc acaccggett cctgacggag tatgtggcta cgcgctggtc ccgggcccga	600
gagatcatgc tgaactccaa gggtataacc aagtccatcg acatctggtc tgtgggctgc	660

[0030]

attctggctg agatgctctc taaccggccc atcttcctg gcaagcacta cctggatcag	720
ctcaaccaca ttctgggcat cctgggctcc ccatcccagg aggacctgaa ttgtatcate	780
aacatgaagg cccgaaacta cctacagtct ctgccctcca agaccaaggt ggcttgggcc	840
aagcttttcc ccaagtcaga ctccaaagcc ctggacctgc tggaccggat gttaaccttt	900
aaccccaata aacggatcac agtggaggaa gcgctggctc acccctacct ggagcagtac	960
tatgaccgga cggatgagcc agtggccgag gagcccttca ccttcgccat ggagctggat	1020
gacctaccta aggagcggct gaaggagctc atcttcagg agacagcacg cttccagccc	1080
ggagtgctgg aggcccccta gccagacag acatctctgc accctggggc ctggacctgc	1140
ctcctgcctg cccctctccc gccagactgt tagaaaatgg aactgtgcc cagccccgac	1200
cttggcagcc caggccgggg tggagcatgg gcctggccac ctctctcctt tgctgaggcc	1260
tccagcttca ggcaggccaa ggcttctcct ccccaccgc cctccccacg ggcctcggga	1320
cctcaggtgg gccagttca atctcccgt gctgctgctg cgccttacc ttccccagcg	1380
tcccagtctc tggcagtttt ggaatggaag ggttctggct gcccacact gctgaagggc	1440
agaggtggag ggtggggggc gctgagtagg gactcacggc catgcctgcc cccctcatct	1500
cattcaaacc ccaccctagt ttcctgaag gaacattcct tagtctcaag ggctagcate	1560
cctgaggagc caggccgggc cgaatcccct cctgtcaaa gctgtcactt cgcgtgcct	1620
cgtgcttct gtgtgtggtg agcagaagtg gagctggggg gcgtggagag cccggctgcc	1680
cctgccacct cctgaccg tctaataat aaatatagag atgtgtctat ggctgaaaaa	1740
aaaaa	1745

<210> 97  
 <211> 1611  
 <212> DNA  
 <213> 智人

<400> 97	
acataatttc tggagccctg taccaacgtg tggccacata ttctgtcagg aacctgtgt	60
gatcatggtc tggatctgca acacgggcca ggccaaagtc acagatcttg agatcacagg	120
tgggtgtttag cagcaggcag gcaggcaatc ggtccgagtg gctgtcggct cttcagctct	180
ccgctcggcg tcttcttcc tctcccgtc agcgtcggcg gctgcaccgg cggcgggcag	240
tcctgcggga ggggcgacaa gagctgaggc gcggccgccg agcgtcagc tcagcgcggc	300
ggaggcggcg gcggcccggc agccaacatg gcggcggcgg_cggcggcggg cgcgggcccg	360

[0031]

gagatggtcc gcgggcaggt gttcgacgtg gggccgcgct acaccaacct ctcgtacatc	420
ggcgagggcg cctacggcat ggtgtgctct gcttatgata atgtcaacaa agttcgagta	480
gctatcaaga aaatcagccc ctttgagcac cagacctact gccagagaac cctgagggag	540
ataaaaatct tactgcgctt cagacatgag aacatcattg gaatcaatga cattattcga	600
gcaccaacca tcgagcaaat gaaagatgta tatatagtac aggacctcat ggaaacagat	660
ctttacaagc tcttgaagac acaacacctc agcaatgacc atatctgcta tttctctac	720
cagatcctca gagggttaaa atatatccat tcagctaacg ttctgcaccg tgacctcaag	780
ccttccaacc tgctgctcaa caccacctgt gatctcaaga tctgtgactt tggcctggcc	840
cgtgttgag atccagacca tgatcacaca gggttcctga cagaatatgt ggccacacgt	900
tggtagagg ctccagaaat tatgttgaat tccaagggt acaccaagtc cattgatatt	960
tggctctgtag gctgcattct ggcagaaatg ctttccaaca ggcccatctt tccaggaag	1020
cattatcttg accagctgaa tcacattttg ggtattcttg gatccccatc acaagaagac	1080
ctgaattgta taataaattt aaaagctagg aactatttgc tttctcttcc acacaaaaat	1140
aaggtagcat ggaacaggct gttcccaaat gctgactcca aagctctgga cttattggac	1200
aaaatgttga cattcaacc acacaagagg attgaagtag aacaggctct gggccacca	1260
tatctggagc agtattacga cccgagtgc gagcccatcg ccgaagcacc attcaagttc	1320
gacatggaat tggatgactt gcctaaggaa aagctaaaag aactaatttt tgaagagact	1380
gctagattcc agccaggata cagatcttaa atttgtcagg acaagggtc agaggactgg	1440
acgtgctcag acatcgggtg tcttcttccc agttcttgac ccctggtcct gtctccagcc	1500
cgtcttggct tatccacttt gactcctttg agccgtttgg agggcggtt tctgtagtt	1560
gtggctttta tgctttcaaa gaatttcttc agtccagaga attcactggc c	1611

<210> 98  
 <211> 1702  
 <212> DNA  
 <213> 智人

<400> 98	
ccggcagtc cgagtgcctc ccgcagaggg ctggtggtgg gagcggagtg gagtcgggcg	60
gggccgaagc cgggccgtgg gcgtagatgg gggccgggcg gcggcggagc ggcggaacgc	120
gggatgggct gcaccgtgag cgccgaggac aaggcggcgg ccgagcgctc taagatgate	180

[0032]

gacaagaacc tgcgggagga cggagagaag gcggcgcggg aggtgaagtt gctgctgttg	240
ggtgctgggg agtcaggga gagcaccatc gtcaagcaga tgaagatcat ccacgaggat	300
ggctactccg aggaggaatg cgggcagtac cgggcggttg tctacagcaa caccatccag	360
tccatcatgg ccattgtcaa agccatggga aacctgcaga tgcactttgc cgaccctcc	420
agagcggacg acgccaggca gctatttgca ctgtcctgca ccgccgagga gcaaggcgtg	480
ctccctgatg acctgtccgg cgtcatccgg aggctctggg ctgaccatgg tgtgcaggcc	540
tgctttggcc gctcaaggga ataccagctc aacgactcag ctgcctacta cctgaacgac	600
ctggagcgtg ttgcacagag tgactacatc cccacacagc aagatgtgct acggaccgcg	660
gtaaagacca cggggatcgt ggagacacac ttcaccttca aggacctaca cttcaagatg	720
tttgatgtgg gtggtcagcg gtctgagcgg aagaagtgga tccactgctt tgaggcgtc	780
acagccatca tcttctgcgt agccttgagc gcctatgact tgggtgctagc tgaggacgag	840
gagatgaacc gcatgcatga gagcatgaag ctattcgata gcatctgcaa caacaagtgg	900
ttcacagaca cgtccatcat cctcttctc aacaagaagg acctgtttga ggagaagatc	960
acacacagtc cctgaccat ctgcttccct gagtacacag gggccaacaa atatgatgag	1020
gcagccagct acatccagag taagtttgag gacctgaata agcgcaaaga caccaaggag	1080
atctacacgc acttcacgtg cgccaccgac accaagaacg tgcagttcgt gtttgacgcc	1140
gtcaccgatg tcatcatcaa gaacaacctg aaggactgcg gcctcttctg aggggcagcg	1200
gggcctggcg ggatgggcca ccgccgaatt tgtaccccc aaccctgag gaagatgggg	1260
gcaagaagat cacgtcccc gcctgttccc ccgccgcttt tctctcttt cctctctttg	1320
ttctcagctc ccctgtccc ctccagctcca aacgtagggg aggggttcgc acaggcctcc	1380
ctgtttgaag cctgcccttg tctgagatgc tggtaatggc catggtacc cttctgggc	1440
atctgttctg gtttttaacc attgtcttgt tctgtgatga ggggagggg gcacatgctg	1500
agtctcccaa ggctgcgtct ggaggggcc ctgcttctcc agcctggacc ccagctttg	1560
cccaacacca gccctgccc cagcccaagt ccaaatgttt acgggagcct cctgcccagt	1620
cccccaacc cagccgctcg gaggcccaa aggaaaaagc acaagaagcg tgagacgcca	1680
ccattctgga aaaccacagt cc	1702

<210> 99  
 <211> 1185  
 <212> DNA

[0033]

<213> 智人

<400> 99

```

atgggctgcc tcgggaacag taagaccgag gaccagcgca acgaggagaa ggcgcagcgt    60
gaggccaaca aaaagatcga gaagcagctg cagaaggaca agcaggtcta ccgggccacg    120
caccgcctgc tgctgctggg tgctggagaa tctggtaaaa gcaccattgt gaagcagatg    180
aggatcctgc atgttaatgg gttaaatgga gagggcggcg aagaggaccc gcaggctgca    240
aggagcaaca gcgatggtga gaaggcaacc aaagtgcagg acatcaaaaa caacctgaaa    300
gaggcgattg aaaccattgt ggccgccatg agcaacctgg tgccccctgt ggagctggcc    360
aaccgccaga accagttcag agtggactac atcctgagtg tgatgaacgt gcctgacttt    420
gacttcctc ccgaattcta tgagcatgcc aaggctctgt gggaggatga aggagtgcgt    480
gcctgctacg aacgctcaa cgagtaccag ctgattgact gtgcccagta cttcctggac    540
aagatcgacg tgatcaagca ggctgactat gtgccgagcg atcaggacct gcttcgctgc    600
cgtgtcctga cttctggaat ctttgagacc aagttccagg tggacaaagt caacttccac    660
atgtttgacg tgggtggcca gcgcgatgaa cgccgcaagt ggatccagtg cttcaacgat    720
gtgactgcca tcacttctgt ggtggccagc agcagctaca acatggatcat ccgggaggac    780
aaccagacca accgctgca ggaggctctg aacctcttca agagcatctg gaacaacaga    840
tgctgcgca ccatctctgt gatcctgttc ctcaacaagc aagatctgct cgctgagaaa    900
gtccttgctg ggaaatcga gattgaggac tactttccag aatttgctcg ctacactact    960
cctgaggatg ctactcccga gcccgagag gacccacgcg tgaccgggc caagtacttc   1020
attcgagatg agtttctgag gatcagcact gccagtggag atgggcgtca ctactgctac   1080
cctcatttca cctgcgctgt ggacactgag aacatccgcc gtgtgttcaa cgactgccgt   1140
gacatcattc agcgcgatga ccttcgtcag tacgagctgc tctaa                       1185

```

<210> 100

<211> 4788

<212> DNA

<213> 智人

<400> 100

```

tttttagaaa aaaaaaatat atttcctcc tgctcctct gcgttcacaa gctaagttgt    60
ttatctcggc tgcggcgga actgcggacg gtggcgggag agcggctcct ctgccagagt   120
tgatattcac tgatggactc caagaatca ttaactcctg gtagagaaga aaaccccagc   180

```

[0034]

agtgtgcttg ctcaggagag gggagatgtg atggacttct ataaaaccct aagaggagga	240
gctactgtga aggtttctgc gtcttcaccc tcaactggctg tcgcttctca atcagactcc	300
aagcagcgaa gacttttggg tgattttcca aaaggctcag taagcaatgc gcagcagcca	360
gatctgtcca aagcagtttc actctcaatg ggactgtata tgggagagac agaaacaaaa	420
gtgatgggaa atgacctggg attcccacag cagggccaaa tcagccttc ctcgggggaa	480
acagacttaa agcttttggg agaaagcatt gcaaacctca ataggctgac cagtgttcca	540
gagaacccca agagttcagc atccactgct gtgtctgctg ccccccacaga gaaggagttt	600
ccaaaaactc actctgatgt atcttcagaa cagcaacatt tgaagggcc aactggcacc	660
aacggtggca atgtgaaatt gtataccaca gaccaaagca cctttgacat tttgcaggat	720
ttggagtttt cttctgggtc ccaggtaaa gagacgaatg agagtcttg gagatcagac	780
ctgttgatag atgaaaactg tttgctttct cctctggcgg gagaagacga ttcattcctt	840
ttggaaggaa actcgaatga ggactgcaag cctctcattt taccggacac taaacccaaa	900
attaaggata atggagatct ggttttgtca agccccagta atgtaacact gccccagtg	960
aaaacagaaa aagaagattt catcgaactc tgcaccctg gggtaattaa gcaagagaaa	1020
ctgggcacag tttactgtca ggcaagcttt cctggagcaa atataattgg taataaaatg	1080
tctgccattt ctgttcatgg tgtgagtacc tctggaggac agatgtacca ctatgacatg	1140
aatacagcat ccctttctca acagcaggat cagaagccta tttttaatgt cattccacca	1200
attcccgttg gttccgaaaa ttggaatagg tgccaaggat ctggagatga caacttgact	1260
tctctgggga ctctgaactt ccctggtcga acagtttttt ctaatggcta ttcaagcccc	1320
agcatgagac cagatgtaag ctctctcca tccagctcct caacagcaac aacaggacca	1380
ctcccaaac tctgcctggt gtgctctgat gaagcttcag gatgtcatta tggagtctta	1440
acttgtggaa gctgtaaagt tttcttcaa agagcagtgg aaggacagca caattaccta	1500
tgtgctggaa ggaatgattg catcatcgat aaaattcgaa gaaaaactg cccagcatgc	1560
cgctatcgaa aatgtcttca ggctggaatg aacctggaag ctcgaaaaac aaagaaaaaa	1620
ataaaaggaa ttcagcaggc cactacagga gtctcacaag aaacctctga aaatcctggt	1680
aacaaaaca tagttctgc aacgttacca caactcacc ctaccctggt gtcactgttg	1740
gaggttattg aacctgaagt gttatatgca ggatatgata gctctgttcc agactcaact	1800
tggaggatca tgactacgct caacatgtta ggaggcggc aagtgattgc agcagtgaaa	1860

[0035]

tgggcaaagg caataccagg tttcaggaac ttacacctgg atgaccaa	gaccctactg	1920
cagtactcct ggatgtttct tatggcattt gctctggggg ggagatcata	tagacaatca	1980
agtgcaaacc tgctgtgttt tgctcctgat ctgattatta atgagcagag	aatgactcta	2040
ccctgcatgt acgaccaatg taaacacatg ctgtatgttt cctctgagtt	acacaggctt	2100
caggatatctt atgaagagta tctctgtatg aaaaccttac tgcttctctc	ttcagttcct	2160
aaggacggtc tgaagagcca agagctatct gatgaaatta gaatgaccta	catcaaagag	2220
ctaggaaaag ccattgtcaa gagggaagga aactccagcc agaactggca	gcggttttat	2280
caactgacaa aactcttgga ttctatgcat gaagtgggtg aaaatctcct	taactattgc	2340
ttccaaacat ttttggataa gaccatgagt attgaattcc ccgagatgtt	agctgaaatc	2400
atcaccaatc agatacaaaa atattcaaat ggaaatatca aaaaacttct	gtttcatcaa	2460
aagtgactgc ctaataaga atggttgccct taaagaaagt cgaattaata	gcttttattg	2520
tataaactat cagtttgcc tgtagaggtt ttgttgtttt attttttatt	gttttcatct	2580
gttgttttgt tttaaatacg cactacatgt ggtttataga gggccaagac	ttggcaacag	2640
aagcagttga gtcgcatca cttttcagtg atgggagagt agatggtgaa	atttattagt	2700
taatatatcc cagaaattag aaaccttaat atgtggacgt aatctccaca	gtcaaagaag	2760
gatggcacct aaaccaccag tgcccaaagt ctgtgtgatg aactttctct	tcatactttt	2820
tttcacagtt ggctggatga aattttctag actttctggt ggtgtatccc	ccccctgtat	2880
agttaggata gcatttttga tttatgcatg gaaacctgaa aaaaagtta	caagtgtata	2940
tcagaaaagg gaagttgtgc cttttatagc tattactgtc tggttttaac	aatttccttt	3000
atatttagtg aactacgctt gctcattttt tcttacataa ttttttattc	aagttattgt	3060
acagctgttt aagatgggca gctagttcgt agctttccca aataaactct	aacattaat	3120
caatcatctg tgtgaaaatg ggttgggtgct tctaacctga tggcacttag	ctatcagaag	3180
accacaaaaa ttgactcaaa tctccagtat tcttgtcaaa aaaaaaaaaa	aaaagctca	3240
tattttgtat atatctgctt cagtggagaa ttatataggt tgtgcaaatt	aacagtccta	3300
actggtatag agcacctagt ccagtgacct gctgggtaaa ctgtggatga	tggttgcaaa	3360
agactaattt aaaaaataac taccaagagg ccctgtctgt acctaacgcc	ctatttttgc	3420
aatggctata tggcaagaaa gctggtaaac tatttgtctt tcaggacctt	ttgaagtagt	3480
ttgtataact tcttaaaagt tgtgattcca gataaccagc tgtaacacag	ctgagagact	3540

[0036]

tttaatcaga caaagtaatt cctctcacta aactttaccc aaaaactaaa tctctaatat	3600
ggcaaaaatg gctagacacc cattttcaca tteccatctg tcaccaattg gttaatcttt	3660
cctgatggta caggaaagct cagctactga tttttgat ttagaactgt atgtcagaca	3720
tccatgtttg taaaactaca catcccta atgtgtgccata gagtttaaca caagtcctgt	3780
gaatttcttc actgttgaaa attattttaa acaaaaataga agctgtagta gccctttctg	3840
tgtgcacctt accaactttc tgtaaaactca aaacttaaca tatttactaa gccacaagaa	3900
atgtgatttc tattcaaggt ggccaaatta tttgtgtaat agaaaactga aatctaata	3960
ttaaaaatat ggaacttcta atatatTTTT atatttagtt atagtttcag atatatatca	4020
tattggattt cactaatctg ggaagggag ggctactgca gctttacatg caatttatta	4080
aatgattgt aaaatagctt gtatagtgt aaataagaat gatttttaga tgagattgtt	4140
ttatcatgac atgttatata tttttgtag gggtaaaga aatgctgatg gataacctat	4200
atgatttata gttgtacat gcattcatac aggcagcagat ggtctcagaa accaaacagt	4260
ttgctctagg ggaagaggg gatggagact ggtcctgtgt gcagtgaagg ttgctgaggc	4320
tctgaccag tgagattaca gaggaagta tcctctgct cccattctga ccacccttct	4380
cattcaaca gtgagtctgt cagcgcaggt ttagtttact caatctccc ttgcactaaa	4440
gtatgtaaag tatgtaaaca ggagacagga aggtggtgct tacatcetta aaggcaccat	4500
ctaatagcgg gttactttca catacagccc tccccagca gttgaatgac aacagaagct	4560
tcagaagttt ggcaatagtt tgcatagagg taccagcaat atgtaaatag tgcagaatct	4620
cataggttgc caataataca ctaattcctt tctatcctac aacaagagtt tatttccaaa	4680
taaaatgagg acatgttttt gttttctttg aatgcttttt gaatgttatt tgttattttc	4740
agtattttgg agaaattatt taataaaaa acaatcattt gctttttg	4788

<210> 101  
 <211> 1498  
 <212> DNA  
 <213> 智人

<400> 101	
accaaacctc ttcgaggcac aaggcacaac aggtctctct gggattctct tcagccaatc	60
ttcattgctc aagtgtctga agcagccatg gcagaagtac ctgagctcgc cagtgaaatg	120
atggcttatt acagtggcaa tgaggatgac ttgttctttg aagctgatgg ccctaaacag	180
atgaagtgct ccttccagga cctggacctc tgcctctgg atggcggcat ccagctacga	240

[0037]

atctccgacc accactacag caagggcttc aggcaggccg cgtcagttgt tgtggccatg	300
gacaagctga ggaagatgct ggttccctgc ccacagacct tccaggagaa tgacctgagc	360
accttctttc ccttcatctt tgaagaagaa cctatcttct tcgacacatg ggataacgag	420
gcttatgtgc acgatgcacc tgtacgatea ctgaaactgca cgctccggga ctcacagcaa	480
aaaagcttgg tgatgtctgg tccatattgaa ctgaaagctc tccacctcca gggacaggat	540
atggagcaac aagtgggtgtt ctccatgtcc tttgtacaag gagaagaaaag taatgacaaa	600
atacctgtgg ccttgggcct caaggaaaag aatctgtacc tgtcctgcgt gttgaaagat	660
gataagccca ctctacagct ggagagtgtg gatcccaaaa attacccaaa gaagaagatg	720
gaaaagcgat ttgtcttcaa caagatagaa atcaataaca agctggaatt tgagtctgcc	780
cagttcccca actggtacat cagcacctct caagcagaaa acatgcccgt cttcctggga	840
gggaccaaag gcggccagga tataactgac ttcaccatgc aatttgtgtc ttcctaaaga	900
gagctgtacc cagagagtcc tgtgtgaat gtggactcaa tccctagggc tggcagaaaag	960
ggaacagaaa ggTTTTTgag tacggctata gcctggactt tctgtttgtc tacaccaatg	1020
cccaactgcc tgccttaggg tagtgetaag aggatctctt gtccatcagc caggacagtc	1080
agctctctcc tttcagggcc aatccccagc cttttgttg agccaggcct ctctcacctc	1140
tectactcac ttaaagcccg cctgacagaa accacggcca catttggttc taagaaacc	1200
tctgtcattc gctcccacat tctgatgagc aaccgcttc ctatttattt atttatttgt	1260
ttgtttgttt tattcattgg tctaatttat tcaaaggggg caagaagtag cagtgtctgt	1320
aaaagagcct agtttttaat agctatggaa tcaattcaat ttggactggg gtgctctctt	1380
taaatcaagt cctttaatta agactgaaaa tatataagct cagattattt aaatgggaat	1440
atttataaat gagcaaatat catactgttc aatggttctg aaataaactt cactgaag	1498

<210> 102  
 <211> 1128  
 <212> DNA  
 <213> 智人

<400> 102	
attctgcctt cgagcccacc gggaacgaaa gagaagctct atctcccctc caggagccca	60
gctatgaact ccttctccac aagcgccttc ggtccagttg ccttctcctt ggggctgctc	120
ctgggtgttg ctgctgcctt ccctgccccca gtacccccag gagaagattc caaagatgta	180

[0038]

gccgccccac acagacagcc actcacctct tcagaacgaa ttgacaaaca aattcggtag 240  
 atcctcgacg gcattctcagc cctgagaaag gagacatgta acaagagtaa catgtgtgaa 300  
 agcagcaaag aggcactggc agaaaacaac ctgaaccttc caaagatggc tgaaaaagat 360  
 ggatgcttcc aatctggatt caatgaggag acttgectgg tgaaaatcat cactggctctt 420  
 ttggagtttg aggtatacct agagtacctc cagaacagat ttgagagtag tgaggaacaa 480  
 gccagagctg tccagatgag tacaaaagtc ctgatccagt tectgcagaa aaaggcaaag 540  
 aatctagatg caataaccac ccttgaccca accacaaatg ccagcctgct gacgaagctg 600  
 caggcacaga accagtggct gcaggacatg acaactcctc tcattctgcg cagctttaag 660  
 gagttcctgc agtccagcct gagggctctt cggcaaagt agcatgggca cctcagattg 720  
 ttgttgtaa tgggcattcc ttcttctggt cagaaacctg tccactgggc acagaactta 780  
 tgttgttctc tatggagaac taaaagtatg agcgtagga cactatttta attatttta 840  
 atttattaat atttaaataat gtgaagctga gtttaattat gtaagtcata tttatattt 900  
 ttaagaagta ccacttgaaa cattttatgt attagttttg aaataataat ggaaagtggc 960  
 tatgcagttt gaatatecct tgtttcagag ccagatcatt tcttgaaag ttaggctta 1020  
 cctcaaataa atggctaact ttatacatat ttttaaagaa atatttatat tgtatttata 1080  
 taatgtataa atggttttta taccaataaa tggcatttta aaaaattc 1128

<210> 103  
 <211> 5191  
 <212> DNA  
 <213> 智人

<400> 103  
 gaattcagta acccagcat tattttatcc tcaagtccta ggttggttg agaaagataa 60  
 caaaaagaaa catgattgtg cagaaacaga caaacctttt tggaaagcat ttgaaaatgg 120  
 cattccccct ccacagtgtg ttcacagtgt gggcaaattc actgctctgt cgtactttct 180  
 gaaaatgaag aactgttaca ccaaggtgaa ttatttataa attatgtact tgcccagaag 240  
 cgaacagact ttactatca taagaacctc tcttggtgt gctctttatc tacagaatcc 300  
 aagaccttc aagaaaggtc ttgattcct ttcttcagga cactaggaca taaagccacc 360  
 tttttatgat ttgtgaaat ttctactcc atcccttttg ctgatgatca tgggtectca 420  
 gaggtcagac ttgggtcct tggataaaga gcatgaagca acagtggctg aaccagagtt 480  
 ggaaccaga tgctcttcc actaagcata caactttcca ttagataaca cctccctccc 540

[0039]

accccaacca agcagctcca gtgcaccact ttctggagca taaacatacc ttaactttac	600
aacttgagtg gccttgaata ctgttcctat ctggaatgtg ctgtttcttt tcatcttctt	660
ctattgaagc cctcctattc ctcaatgcct tgctccaact gcctttggaa gattctgctc	720
ttatgcctcc actggaatta atgtcttagt accacttgtc tattctgcta tatagtcagt	780
ccttacattg ctttcttctt ctgatagacc aaactcttta aggacaagta cctagtctta	840
tctattttcta gateccccac attactcaga aagttactcc ataaatgttt gtggaactga	900
tttctatgtg aagacatgtg ccccttcaact ctgttaacta gcattagaaa aacaaatctt	960
ttgaaaagtt gtagtatgcc cctaagagca gtaacagttc ctagaaactc tctaaaatgc	1020
ttagaaaaag atttatttta aattacctcc ccaataaaat gattggctgg cttatcttca	1080
ccatcatgat agcatctgta attaactgaa aaaaaataat tatgccatta aaagaaaatc	1140
atccatgata ttgttctaac acctgccact ctagtactat atctgtcaca tggctatga	1200
taaagttatc tagaaataaa aaagcataca attgataatt caccaaattg tggagcttca	1260
gtattttaaa tgtatattaa aattaaatta ttttaaagat caaagaaaac tttcgtcata	1320
ctccgtattht gataaggaac aaataggaag tgtgatgact caggtttgcc ctgaggggat	1380
gggccatcag ttgcaaatcg tggaaattcc tctgacataa tgaaaagatg aggggtgcata	1440
agttctctag tagggatgatg atataaaaag ccaccggagc actccataag gcacaaactt	1500
tcagagacag cagagcacac aagcttctag gacaagagcc aggaagaaac caccggaagg	1560
aaccattctc actgtgtgta aacatgactt ccaagctggc cgtggctctc ttggcagcct	1620
tcttgatttc tgcagctctg tgtgaaggta agcacatctt tctgacctac agcgttttcc	1680
tatgtctaaa tgtgatcctt agatagcaaa gctattcttg atgctttggg aacaaacatc	1740
ctttttattc agaaacagaa tataatctta gcagtcaatt aatgttaaata tgaagattta	1800
gaaaaaacta tatataacac ttaggaaata taaaggtttg atcaatatag atattctgct	1860
tttataattht ataccaggta gcatgcatat atttaacgta aataagtaat ttatagtatg	1920
tcctattgag aaccacggtt acctatatta tgtattaata ttgagttgag caaggtaact	1980
cagacaattc cactccttgt agtatttcat tgacaagcct cagatttgct attaattcct	2040
gtctggttta aagataccct gattatagac caggcatgta taacttattt atatatttct	2100
gttaattctt tctgaaggca atttctatgc tggagagtct tagcttgctt actataaata	2160
acactgtggg atcacagagg attatgcaat attgaccaga taaaaatacc atgaagatgt	2220

[0040]

tgatattgta caaaaagaac tctaactctt atataggaag ttgttcaatg ttgtcagtta 2280  
tgactgtttt ttaaaacaaa gaactaactg aggtcaaggg ctaggagata ttcaggaatg 2340  
agttcactag aaacatgatg ccttccatag tctccaaata atcatattgg aattagaagg 2400  
aagtagctgg cagagctgtg cctgttgata aatcaatcc ttaatcactt tttccccaa 2460  
caggtgcagt tttgccaagg agtgctaaag aacttagatg tcagtgcata aagacatact 2520  
ccaaaccttt ccaccccaaa tttatcaaag aactgagagt gattgagagt ggaccacact 2580  
gcgccaacac agaaattatg taagtacttt aaaaaagatt agatattttg ttttagcaaa 2640  
cttaaaatta aggaaggtgg aatatttag gaaagttcca ggtgtagga ttacagtagt 2700  
aatgaaaca aaacaaaata aaaatatttg tctacatgac atttaaatat ggtagcttcc 2760  
acaactacta taaatgttat tttggactta gactttatgc ctgacttaag gaatcatgat 2820  
ttgaatgcaa aactaaata ttaatctgaa ccatttcttt cttatttcag tgtaaagctt 2880  
tctgatggaa gagagctctg tctggacccc aaggaaaact ggggtgcagag ggttgtggag 2940  
aagtttttga agaggtaagt tatatatttt ttaatttaa tttttcattt atcctgagac 3000  
atataatcca aagtcagcct ataaatttct ttctgttgct aaaaatcgtc attaggtatc 3060  
tgcctttttg gttaaaaaaa aaggaatagc atcaatagtg agtttgttgt acttatgacc 3120  
agaaagacca tacatagttt gcccaggaaa ttctgggttt aagcttgtgt cctatactct 3180  
tagtaaagtt ctttgtcact cccagtagtg tctatttta gatgataatt tctttgatct 3240  
ccctatttat agttgagaat atagagcatt tctaacacat gaatgtcaaa gactatattg 3300  
acttttcaag aaccctactt tccttcttat taaacatagc tcacttttat atttttaatt 3360  
ttatttttagg gctgagaatt cataaaaaaa ttcatctctt gtggtatcca agaactcagt 3420  
aagatgccag tgaaacttca agcaaatcta cttcaacact tcatgtattg tgtgggtctg 3480  
ttgtagggtt gccagatgca atacaagatt cctgggttaa tttgaatttc agtaaacaat 3540  
gaatagtttt tcattgtacc atgaaatata cagaacatac ttatatgtaa agtattattt 3600  
atttgaatct acaaaaaaca acaataaatt tttaaataa aggattttcc tagatattgc 3660  
acgggagaat atacaaatag caaaattggg ccaagggcc aagagaatata cgaacttta 3720  
tttcaggaat tgaatgggtt tgctagaatg tgatatttga agcatcacat aaaaatgatg 3780  
ggacaataaa ttttgccata aagtcaaatt tagctggaaa tcctggattt ttttctgtta 3840  
aatctggcaa ccctagtctg ctagccagga tccacaagtc cttgttccac tgtgccttgg 3900

[0041]

tttctccttt atttctaagt ggaaaaagta ttagccacca tcttacctca cagtgatggt	3960
gtgaggacat gtggaagcac tttaagtttt ttcatacataa cataaattat tttcaagtgt	4020
aacttattaa cctatttatt atttatgtat ttatttaagc atcaaatatt tgtgcaagaa	4080
tttgaaaaa tagaagatga atcattgatt gaatagtat aaagatgtta tagtaaattt	4140
attttatttt agatattaaa tgatgtttta ttagataaat ttcaatcagg gtttttagat	4200
taaacaaaca aacaattggg taccaggtta aattttcatt tcagatatac aacaaataat	4260
tttttagtat aagtacatta ttgtttatct gaaattttta ttgaactaac aatcctagtt	4320
tgatactccc agtcttgtca ttgccagctg tgttgtagt gctgtgttga attacggaat	4380
aatgagttag aactattaaa acagccaaaa ctccacagtc aatattagta atttcttgct	4440
ggttgaaact tgtttattat gtacaaatag attcttataa tattatttaa atgactgcat	4500
ttttaaatac aaggtttat atttttaact ttagtgtttt tatgtgctct ccaaattttt	4560
tttactgitt ctgattgtat ggaaatataa aagtaaatat gaaacattta aatataatt	4620
tgttgcaaaa gtaatcaagt gtttgtcttt ttttagttt tagcttattg ggattctctt	4680
tgtttatatt taaaattata ctttgattta gaaaacataa atgcttcccc ttagcatttt	4740
gttatggaaa attacaaact tttattttta gaaaacagaa ctctttcca gaaatagggt	4800
acaaacagta gtgtcttcca cagaatgttg gaaatgttt caactcccca ctgtatacta	4860
tcttgetaat aagtctgtct tcagatttcg attaaccggt ttgtatgtct gtgcacttta	4920
gcatagctgg acattaaaga ggaaagagag tacatattat aagttgctta tcagtaactg	4980
aggagtaaaa ctgataaatg tgaggcaaag aagtttaaaa tatggttaaa gcctaagcat	5040
attgcaaac aatcaaaaca atactctgag aagtaaaaac ataattattt aattaacaaa	5100
tttcagtgga taaattttat aacaaattag acacagttga aaataaaatt agaaaactag	5160
aaaatagaac aaaagaaact tctggaattc a	5191

<210> 104  
 <211> 1572  
 <212> DNA  
 <213> 智人

<400> 104	
aatttctcac tgcccctgtg ataaactgtg gtcactggct gtggcagcaa ctattataag	60
atgctctgaa aactcttcag aactgaggg gcaccagagg agcagactac aagaatggca	120

[0042]

cacgctatgg aaaactcctg gacaatcagt aaagagtacc atattgatga agaagtgggc	180
tttgctctgc caaatccaca ggaaaatcta cctgattttt ataatgactg gatgttcatt	240
gctaaacatc tgccatgatct catagagtct ggccagcttc gagaaagagt tgagaagtta	300
aacatgctca gcattgatca tctcacagac cacaagtcac agcgccttgc acgtctagtt	360
ctgggatgca tcaccatggc atatgtgtgg ggcaaaggtc atggagatgt ccgtaaggtc	420
ttgccaagaa atattgctgt tccttactgc caactctcca agaaactgga actgcctcct	480
atdddggddd atgcagactg tgtcttggca aactggaaga aaaaggatcc taataagccc	540
ctgacttatg agaacatgga cgttttgttc tcatttcgtg atggagactg cagtaaagga	600
ttcttcctgg tctctctatt ggtggaaata gcagctgctt ctgcaatcaa agtaattcct	660
actgtattca aggcaatgca aatgcaagaa cgggacactt tgctaaagge gctgttggaa	720
atagcttctt gcttggagaa agcccttcaa gtgtttcacc aatccacga tcatgtgaac	780
ccaaaagcat ttttcagtgt tcttcgcata tatttgtctg gctggaaagg caacccccag	840
ctatcagacg gtctgggtgta tgaagggttc tgggaagacc caaaggagtt tgcagggggc	900
agtgcaaggc aaagcagcgt ctttcagtgc tttgacgtcc tgctgggcat ccagcagact	960
gctgggtggag gacatgctgc tcagttctc caggacatga gaagatatat gccaccagct	1020
cacaggaact tcctgtgctc attagagtca aatccctcag tccgtgagtt tgtcctttca	1080
aaaggtgatg ctggcctgcg ggaagcttat gacgcctgtg tgaaagctct ggtctccctg	1140
aggagctacc atctgcaaat cgtgactaag tacatcctga ttctgcaag ccagcagcca	1200
aaggagaata agacctctga agacccttca aaactggaag ccaaaggaac tggaggcact	1260
gatttaatga atttctgaa gactgtaaga agtacaactg agaaatccct tttgaaggaa	1320
ggttaatgta acccaacaag agcacatddd atcatagcag agacatctgt atgcattcct	1380
gtcattacc c attgtaacag agccacaaac taatactatg caatgtttta ccaataatgc	1440
aatacaaaag acctcaaat acctgtgcat ttctttagg aaaacaacaa aaggttaatta	1500
tgtgtaatta tactagaagt tttgtaatct gtatcttacc attggaataa aatgacattc	1560
aataaataaa aa	1572

<210> 105  
 <211> 1539  
 <212> DNA  
 <213> 智人

[0043]

<400> 105  
 ggaattccgg gcccggtctt tcctcccgcc gccgccggcc tggccccggg gactggcctc 60  
 cacgtccgac tcgtccgagc tgaagcccag cagcactttg ctgccagccg cgggggcggc 120  
 ggaggcgccc ccgggccctc ccaggaggct ctctgggcca gaggccgaga ttcggcacag 180  
 gccccagga gtccgtaagt aggagaggtc gcccgagacc ggccggaccc ccatccccgc 240  
 ggccgcccgc gccgctggtc ccgcggtgc gaccgtggcg gctgccgctg gaaaatgtct 300  
 caggagagge ccacgttcta ccggcaggag ctgaacaaga caatctggga ggtgcccgag 360  
 cgttaccaga acctgtctcc agtgggctct ggccctatg gctctgtgtg tgctgctttt 420  
 gacacaaaaa cggggttacg tgtggcagtg aagaagctct ccagaccatt tcagtccatc 480  
 attcatgcga aaagaacctc cagagaactg cggttactta aacatatgaa acatgaaaat 540  
 gtgattggtc tgttgacgt ttttacacct gcaaggtctc tggaggaatt caatgatgtg 600  
 tatctggtga cccatctcat gggggcagat ctgaacaaca ttgtgaaatg tcagaagctt 660  
 acagatgacc atgttcagtt ccttatctac caaattctcc gaggtctaaa gtatatacat 720  
 tcagctgaca taattcacag ggacctaaaa cctagtaatc tagctgtgaa tgaagactgt 780  
 gagctgaaga ttctggattt tggactggct cggcacacag atgatgaaat gacaggctac 840  
 gtggccacta ggtggtacag ggctcctgag atcatgctga actggatgca ttacaaccag 900  
 acagttgata tttggtcagt gggatgcata atggccgagc tgttgactgg aagaacattg 960  
 tttctggta cagaccatat tgatcagttg aagctcattt taagactcgt tggaacccca 1020  
 ggggctgagc ttttgaagaa aatctcctca gactctgcaa gaaactatat tcagtctttg 1080  
 actcagatgc cgaagatgaa ctttgcgaat gtatttattg gtgccaatcc cctggctgtc 1140  
 gacttgctgg agaagatgct tgtattggac tcagataaga gaattacagc ggcccaagcc 1200  
 cttgcacatg cctactttgc tcagtaccac gatcctgatg atgaaccagt ggccgatcct 1260  
 tatgatcagt cctttgaaag caggacctc cttatagatg agtggaaaag cctgacctat 1320  
 gatgaagtca tcagctttgt gccaccacc cttgaccaag aagagatgga gtccctgagca 1380  
 cctggtttct gttctgttga tcccacttca ctgtgagggg aaggcctttt cacgggaact 1440  
 ctccaaatat tattcaagtg cctcttgttg cagagatttc ctccatggtg gaagggggtg 1500  
 tgcgtgcgtg tgcgtgcgtg ttagtgtgtg tgeatgtg 1539

<210> 106

<211> 1155

[0044]

<212> DNA  
<213> 智人

<400> 106

```

atgagcagaa gcaagcgtga caacaat ttt tatagtgtag agattggaga ttctacattc      60
acagtcctga aacgatata caaattaaaa cctataggct caggagctca aggaatagta      120
tgccgagctt atgatgcat tcttgaaaga aatgttgcaa tcaagaagct aagccgacca      180
tttcagaatc agactcatgc caagcgggcc tacagagagc tagttcttat gaaatgtggt      240
aatcacaaaa atataattgg cttttgaaat gttttcacac cacagaaatc cctagaagaa      300
tttcaagatg tttacatagt catggagctc atggatgcaa atctttgcca agtgattcag      360
atggagctag atcatgaaag aatgtcctac cttctctatc agatgctgtg tggaatcaag      420
caccttcatt ctgctggaat tattcatcgg gacttaaagc ccagtaatat agtagtaaaa      480
tctgattgca ctttgaagat tcttgacttc ggtctggcca ggactgcagg aacgagtttt      540
atgatgacgc cttatgtagt gactcgctac tacagagcac ccgaggatc ccttgcatg      600
ggctacaagg aaaacgtgga tttatggtct gtggggtgca ttatgggaga aatggtttgc      660
cacaaaatcc tctttccagg aagggactat attgatcagt ggaataaagt tattgaacag      720
cttggaacac catgtcctga attcatgaag aaactgcaac caacagtaag gacttacgtt      780
gaaaacagac ctaaatatgc tggatatagc tttgagaaac tcttccctga tctccttttc      840
ccagctgact cagaacacaa caaacttaa gccagtcagg caagggattt gttatccaaa      900
atgctggtaa tagatgcatc taaaaggatc tctgtagatg aagctctcca acaccctac      960
atcaatgtct ggtatgatcc ttctgaagea gaagctccac caccaaagat ccctgacaag     1020
cagttagatg aaagggaaaca cacaatagaa gagtggaaag aattgatata taaggaagtt     1080
atggacttgg aggagagaa caagaatgga gttatacggg ggcagccctc tcctttagca     1140
caggtgcage agtga                                                    1155

```

<210> 107  
<211> 2000  
<212> DNA  
<213> 智人

<400> 107

```

tttgggctgt gttgogacg cgggtcggag ggcagtcgg gggaaccgag aagaagccga      60
ggagcccgga gcccgcgtg acgtctctct ctcagtccaa aagcggcttt tggttcggcg      120
cagagagacc cgggggtcta gcttttctct gaaaagegcc gccctgccct tggccccgag      180

```

[0045]

aacagacaaa gagcaccgca gggccgatca cgctgggggc gctgaggccg gccatggtca	240
tggaagtggg caccctggac gctggaggcc tgcgggcgct gctgggggag cgagcggcgc	300
aatgcctgct gctggactgc cgctccttct tcgctttcaa cgccggccac atcgccggct	360
ctgtcaacgt gcgcttcagc accatcgtgc ggcgcggggc caagggcgcc atgggcctgg	420
agcacatcgt gcccaacgcc gagctccgcg gccgcctgct ggccggcgcc taccacgccg	480
tggtgttgct ggacgagcgc agcgcgccc tggacggcgc caagcgcgac ggcaccctgg	540
ccctggcggc cggcgcgctc tgccgcgagg cgcgcgccc gcaagtctc ttctcaaag	600
gaggatacga agcgttttcg gcttcctgcc cggagctgtg cagcaaacag tcgacccca	660
tggggctcag ccttcccctg agtactagcg tccctgacag cgcggaatct ggggtcagtt	720
cctgcagtac cccactctac gatcagggtg gcccggtgga aatcctgcc tttctgtacc	780
tgggcagtgc gtatcacgct tcccgcaagg acatgctgga tgccttgggc ataactgcct	840
tgatcaacgt ctcagccaat tgtcccaacc attttgaggg tcactaccag tacaagagca	900
tccctgtgga ggacaaccac aaggcagaca tcagctcctg gttcaacgag gccattgact	960
tcatagactc catcaagaat gctggaggaa ggggtgtttgt ccaactgccag gcaggcattt	1020
cccggtcagc caccatctgc cttgettacc ttatgaggac taatcgagtc aagctggacg	1080
aggcctttga gtttgtgaag cagaggcgaa gcatcatctc tcccaacttc agcttcatgg	1140
gccagctgct gcagtttgag tcccaggtgc tggctccgca ctgttcggca gaggctggga	1200
gccccgccat ggctgtgctc gaccgaggca cctccaccac caccgtgtc aacttccccg	1260
tctccatccc tgtccactcc acgaacagtg cgctgagcta ccttcagagc ccattacga	1320
cctctcccag ctgctgaaag gccacgggag gtgaggctct tcacatcca ttgggactcc	1380
atgctccttg agaggagaaa tgcaataact ctgggagggg ctcgagaggg ctggtcctta	1440
tttatttaac ttcacccgag ttctctggg tttctaagca gttatgggtga tgacttagcg	1500
tcaagacatt tgctgaactc agcacattcg ggaccaatat atagtgggta catcaagtcc	1560
atctgacaaa atggggcaga agagaaagga ctcagtgtgt gatccggtt ctttttgctc	1620
gcccctgttt tttgtagaat ctcttcatgc ttgacatacc taccagtatt attcccagcg	1680
acacataac atatgagaat ataccttatt ttttttgg taggtgtctg ctttcacaaa	1740
tgtcattgct tactcctaga agaaccaaat acctcaattt ttgttttga gtactgtact	1800
atcctgtaaa tatacttaa gcaggtttgt tttcagcact gatggaaaat accagtgttg	1860

[0046]

ggtttttttt tagttgccaa cagttgtatg tttgctgatt atttatgacc tgaataata 1920  
 tattttcttct tctaagaaga cattttgta cataaggatg acttttttat acaatggaat 1980  
 aaattatggc atttctattg 2000

<210> 108  
 <211> 5749  
 <212> DNA  
 <213> 智人

<400> 108  
 cgcgggagcc aacttcaggc tgctcagagg aagcccgtgc agtcagtcac ctgggtgcaa 60  
 gagcgttgct gcctcgggct ctcccgtgc agggagagcg gcactcgctg gcctggatgt 120  
 ggttgatttt aggggggctc cgcagcaggg gtttcgtggc ggtggcaagc gctgcaacag 180  
 gtagacggcg agagacggac cccggccgag gcagggatgg agaccaaagg ctaccacagt 240  
 ctccctgaag gtctagatat ggaaagacgg tggggtaag tttctcagge tgtggagcgt 300  
 tcttccctgg gacctacaga gaggaccgat gagaataact acatggagat tgtcaacgta 360  
 agctgtgttt ccggtgctat tccaacaac agtactcaag gaagcagcaa agaaaaaca 420  
 gaactactcc cttgccttca gcaagacaat aatcggcctg ggattttaac atctgatatt 480  
 aaaactgagc tggaatctaa ggaactttca gcaactgtag ctgagtccat gggtttatat 540  
 atggattctg taagagatgc tgactattcc tatgagcagc agaaccaaca aggaagcatg 600  
 agtccagcta agatttatca gaatgttgaa cagctgggtga aattttaca aggaaatggc 660  
 catcgtcctt ccaacttaag ttgtgtgaac acgcccttga gatcatttat gtctgactct 720  
 gggagctccg tgaatgggtg cgtcatgccc gccattgtta aaagccctat catgtgtcat 780  
 gagaaaagcc cgtctgtttg cagccctctg aacatgacat cttcggtttg cagccctgct 840  
 ggaatcaact ctgtgtctc caccacagcc agctttggca gttttccagt gcacagccca 900  
 atcaccagg gaactcctct gacatgctcc cctaatgctg aaaatcgagg ctccaggctg 960  
 cacagccctg cacatgctag caatgtgggc tctcctctct caagtccgtt aagtagcatg 1020  
 aaatcctcaa tttccagccc tccaagtcac tgcagtgtaa aatctccagt ctccagtccc 1080  
 aataatgca ctctgagatc ctctgtgtct agccctgcaa atattaaca ctcaaggtgc 1140  
 tctgtttcca gcccttcgaa cactaataac agatccacgc tttccagtcc ggcagccagt 1200  
 actgtgggat ctatctgtag ccctgtaaac aatgccttca gctacactgc ttctggcacc 1260

[0047]

tctgctggat	ccagtacatt	gcgggatgtg	gttcccagtc	cagacacgca	ggagaaaggt	1320
gctcaagagg	tcccttttcc	taagactgag	gaagtagaga	gtgccatctc	aaatgggtgtg	1380
actggccagc	ttaatattgt	ccagtacata	aaaccagaac	cagatggagc	ttttagcagc	1440
tcatgtctag	gaggaaatag	caaaataaat	tcgattctt	cattctcagt	accaataaag	1500
caagaatcaa	ccaagcattc	atgttcaggc	acctctttta	aagggaatcc	aacagtaaac	1560
ccgtttccat	ttatggatgg	ctcgtatfff	tcctttatgg	atgataaaga	ctattattcc	1620
ctatcaggaa	ttttaggacc	acctgtgccc	ggctttgatg	gtaactgtga	aggcagcggg	1680
ttcccagtg	gtattaaaca	agaaccagat	gacgggagct	attaccaga	ggccagcatc	1740
ccttcctctg	ctattgttgg	ggtgaattca	ggtggacagt	ccttccacta	caggattgg	1800
gctcaaggta	caatatcttt	atcacgatcg	gctagagacc	aatctttcca	acacctgagt	1860
tcctttctc	ctgtcaatac	tttagtggag	tcatggaaat	cacacggcga	cctgtcgtct	1920
agaagaagt	atgggtatcc	ggtcttagaa	tacattccag	aaaatgtatc	aagctctact	1980
ttacgaagt	tttctactgg	atcttcaaga	ccttcaaaaa	tatgtttgg	gtgtggggat	2040
gaggcttcag	gatgccatta	tgggtagtc	acctgtggca	gctgcaaagt	tttcttcaaa	2100
agagcagtg	aagggcaaca	caactattta	tgtgctggaa	gaaatgattg	catcattgat	2160
aagattcgac	gaaagaattg	tcttgettgc	agacttcaga	aatgtcttca	agctggaatg	2220
aatttaggag	cacgaaagtc	aaagaagttg	ggaaagttaa	aagggttca	cgaggagcag	2280
ccacagcagc	agcagcccc	acccccacc	ccacccccgc	aaagcccaga	ggaagggaca	2340
acgtacatcg	ctcctgcaaa	agaaccctcg	gtcaacacag	cactggttcc	tcagctctcc	2400
acaatctcac	gagcgtcac	acctcccc	gttatgggtcc	ttgaaaacat	tgaacctgaa	2460
attgtatatg	caggctatga	cagctcaaaa	ccagatacag	ccgaaaatct	gctctccacg	2520
ctcaaccgct	tagcaggcaa	acagatgatc	caagtcgtga	agtgggcaaa	ggtacttcca	2580
ggatttaaaa	acttgctct	tgaggaccaa	attaccctaa	tccagtattc	ttggatgtgt	2640
ctatcatcat	ttgccttgag	ctggagatcg	tacaaacata	cgaacagcca	atttctctat	2700
tttgaccag	acctagtctt	taatgaagag	aagatgcac	agtctgcat	gtatgaacta	2760
tgccagggga	tgcaccaa	cagccttcag	ttcgttcgac	tgcagctcac	cttgaagaa	2820
tacaccatca	tgaagtttt	gctgctacta	agcacaattc	caaaggatgg	cctcaaaagc	2880
caggctgcat	ttgaagaaat	gaggacaaat	tacatcaaag	aactgaggaa	gatggtaact	2940

[0048]

aagtgtccca	acaattctgg	gcagagctgg	cagaggttct	accaactgac	caagctgctg	3000
gactccatgc	atgacctggt	gagcgacctg	ctggaattct	gcttctacac	cttccgagag	3060
tcccatgctc	tgaaggtaga	gttccccgca	atgctggtgg	agatcatcag	cgaccagctg	3120
cccaaggtgg	agtcggggaa	cgccaagccg	ctctacttcc	accggaagtg	actgcccgct	3180
gcccagaaga	actttgcctt	aagtttcctt	gtgttgttcc	acaccagaa	ggaccaaga	3240
aaacctgttt	ttaacatgtg	atggttgatt	cacacttggt	caacagtttc	tcaagtttaa	3300
agtcatgtca	gaggtttgga	gccgggaaag	ctgtttttcc	gtggatttgg	cgagaccaga	3360
gcagtctgaa	ggattcccca	cctccaatcc	cccagcgctt	agaaacatgt	tctgttccct	3420
cgggatgaaa	agccatatct	agtcaataac	tctgattttg	atattttcac	agatggaaga	3480
agttttaact	atgccgtgta	gtttctggta	tcgttcgctt	gttttaaaag	ggttcaagga	3540
ctaacgaacg	ttttaagct	tacccttggg	ttgcacataa	aacgtatagt	caatatgggg	3600
cattaatatt	cttttgttat	taaaaaaca	caaaaaata	ataaaaaaat	atatacagat	3660
tctgtttgtg	taataacaga	actcgtggcg	tggggcagca	gctgcctctg	agccctcgtc	3720
cgtccacggg	cttctgcctc	actggtatac	acactcgtta	gcgtccattt	cttatttaat	3780
tagaatgat	aagatgatgt	taaatgcctt	ggtttgattt	ctagtatcta	ttgtgttggc	3840
tttacaata	atTTTTTgca	gtcttttgct	gtgctgtaca	ttactgtatg	tataaattat	3900
gaaggacctg	aaataaggta	taaggatctt	ttgtaaatga	gacacataca	aaaaaaatct	3960
ttaatggtta	ataggatgaa	tgggaaagta	tttttgaaag	aattctattt	tgctggagac	4020
tatttaagta	ctatctttgt	ctaaacaagg	taattttttt	ttgtaaagtg	caatgtcctg	4080
catgcataat	gaaccgttta	cagtgtattt	aagaaagggg	aagctgtgcc	tttttttagct	4140
tcatatctaa	tttaccatta	ttttacagtc	tctgttgtaa	ataaccacac	tgaaacctct	4200
tcggttgtct	tgaaaccttt	ctactttttc	tgtacttttt	gttttgttct	tggtctcccg	4260
cttggggcat	ttgtgggact	ccagcacggt	ttctggcttc	tgcttcatcc	tgtccatcg	4320
gggaatgaca	cactgcggtg	tctgcagctc	ctggaaggtg	tcatttgaca	acacatgtgg	4380
gagaggaggt	ccttggagtg	ctgcagcttt	gggaaagcct	gcctcgtttc	ccttttcctc	4440
tagaagcaga	accagctcta	cgagagtggg	actgggaact	tgatggctca	gagagcatct	4500
tttctccca	ttttagaaaa	tcagattttc	tctgtgggga	aaaaaaaatt	ccatgcactc	4560
tctctctgtt	aaagatcagc	tattcccttc	tgatcttggg	aagaggttct	gcactcctgg	4620

[0049]

aaccggtcac aggaacgcac agatcatggc aggatgcgct gggacggccc atcttggcaa	4680
ggttcagtct gaatggcatg gagaccgga gatagagggg ttttagattt ttaaaaggta	4740
ggttttaaaa ataagtttta tacataaaca gttttggaga aaaattacag atcatataag	4800
caagacagtg gcaactaaaat gtttaattca ttaatctggt tgtttggcac tgatgcaatg	4860
tatggctttt ctcttgcccc aatcacaaa catatgtatc tttggggaaa ctaacaatat	4920
gattgcacta aataaactac tttgaataga ggccaaatta atcttttaa aatgatgata	4980
atcatcaggt ttactcagtg aatcatatt aattattttc caaaatctaa aagctgtagc	5040
tggagaagcc catggccacg aggaagcagc aattaattag atcaacactt ttctccaggg	5100
ttcaccatgc aggcaacatt accttgtctt tcaaaagaca cctgccttag tgcaagggga	5160
aacctgtgaa agctgcactc agagggagga gtctttctta cataatttgc aatttcagga	5220
atttaattta taggcagatc tttaaataca gtcaacttac ggtgcacagt aatatgaaag	5280
ccacactttg aaggtaataa atacacagca tgcagactgg gagttgctag caaacaaatg	5340
gettacttac aaaagcagct tttagttcag acttagtttt tataaaatga gaattctgac	5400
ttacttaacc aggtttggga tggagatggt ctgcatcagc tttttgtatt aacaaagta	5460
ctggtctttt gtgtgtctcc aggtaacttt gcttgattaa acagcaaagc catattctaa	5520
attcactggt gaatgcctgt cccagtccaa attgtctgtc tgctcttatt tttgtacat	5580
attgctctta aaaatcttgg tttggtacag ttcataattc accaaaaagt tcatataatt	5640
taaagaaca ctaaattagt ttaaaatgaa gcaatttata tctttatgca aaaacatatg	5700
tctgtctttg caaaggactg taagcagatt acaataaatc ctttacttt	5749

<210> 109  
 <211> 2062  
 <212> DNA  
 <213> 智人

<400> 109	
gtcagtcct cctgtagccg cggccggcgc cgcccggcgc ccctctgcca gcagctccgg	60
cgccacctcg ggccggcgtc tccggcgggc gggagccagg cgctgacggg cgcggcgggg	120
gcggccgagc gctcctgagg ctgcgactca ggctccggcg tctgcgcttc cccatggggc	180
tggcctgagg cgcttggggc ctctgagatt gtcactgctg ttccaagggc acacgcagag	240
ggatttgaa ttcttgaga gttgcctttg tgagaagctg gaaatattc tttcaattcc	300
atctcttagt tttccatagg aacatcaaga aatcatgaac aactttgta atgaagagtt	360

[0050]

tgactgccac	ttcctcgatg	aaggttttac	tgccaaggac	attctggacc	agaaaattaa	420
tgaagtttct	tcttctgatg	ataaggatgc	cttctatgtg	gcagacctgg	gagacattct	480
aaagaaacat	ctgaggtggt	taaaagctct	ccctcgtgtc	acccccittt	atgcagtcaa	540
atgtaatgat	agcaaagcca	tctgtaagac	ccttgctgct	accgggacag	gatttgactg	600
tgctagcaag	actgaaatac	agttggtgca	gagtctgggg	gtgcctccag	agaggattat	660
ctatgcaaat	ccttgtaaac	aagtatctca	aattaagtat	gctgctaata	atggagtcca	720
gatgatgact	tttgatagtg	aagttgagtt	gatgaaagtt	gccagagcac	atcccaaagc	780
aaagttggtt	ttgcggattg	ccactgatga	ttccaaagca	gtctgtctgc	tcagtgtgaa	840
attcgggtgcc	acgctcagaa	ccagcaggct	ccttttggaa	cgggcgaaag	agctaaatat	900
cgatgttggt	ggtgtcagct	tccatgtagg	aagcggctgt	accgatcctg	agacctcgt	960
gcaggcaatc	tctgatgcc	gctgtgtttt	tgacatgggg	gctgaggttg	gtttcagcat	1020
gtatctgctt	gatattggcg	gtggctttcc	tggatctgag	gatgtgaaac	ttaaatttga	1080
agagatcacc	ggcgtaatca	accagcgtt	ggacaaatac	tttccgtcag	actctggagt	1140
gagaatcata	getgagcccg	gcagatacta	tgttgcatca	gctttcacgc	ttgcagttaa	1200
tatcattgcc	aagaaaattg	tattaaagga	acagacgggc	tctgatgacg	aagatgagtc	1260
gagtgagcag	acctttatgt	attatgtgaa	tgatggcgtc	tatggatcat	ttaattgcat	1320
actctatgac	cacgcacatg	taaagcccct	tctgcaaaag	agacctaac	cagatgagaa	1380
gtattattca	tccagcatat	ggggaccaac	atgtgatggc	ctcgatcgga	ttgttgagcg	1440
ctgtgacctg	cctgaaatgc	atgtgggtga	ttggatgctc	tttgaaaaca	tgggcgctta	1500
cactgttgct	gctgccteta	cgttcaatgg	cttccagagg	ccgacgatct	actatgtgat	1560
gtcagggcct	gcgtggcaac	tcatgcagca	attccagaac	cccgacttcc	cacccgaagt	1620
agaggaacag	gatgccagca	cctgcctgt	gtcttgtgcc	tgggagagtg	ggatgaaacg	1680
ccacagagca	gcctgtgctt	cgctagtat	taatgtgtag	atagcactct	ggtagctgtt	1740
aactgcaagt	ttagcttgaa	ttaagggtt	tggggggacc	atgtaactta	attactgcta	1800
gttttgaaat	gtctttgtaa	gagtagggtc	gcatgatgc	agccatatgg	aagactagga	1860
tatgggtcac	acttatctgt	gttctatgg	aaactatttg	aatatttggt	ttatatggat	1920
ttttattcac	tcttcagaca	cgctactcaa	gagtgccct	cagctgctga	acaagcattt	1980
gtagcttgta	caatggcaga	atgggccaaa	agcttagtgt	tgtgacctgt	ttttaaaata	2040

[0051]

aagtatcttg aaataattag gc	2062
<210> 110	
<211> 3155	
<212> DNA	
<213> 智人	
<400> 110	
gtttcatitt gcagttactg ggagggggct tgctgtggcc ctgtcaggaa gagtagagct	60
ctggtccagc tccgcgcagg gagggaggct gtcaccatgc cggcctgctg cagctgcagt	120
gatgttttcc agtatgagac gaacaaagtc actcggatcc agagcatgaa ttatggcacc	180
attaagtggg tcttccacgt gatcatcttt tctacgttt gctttgctct ggtgagtgc	240
aagctgtacc agcggaaaga gctgtcacc agttctgtgc acaccaaggt gaaggggata	300
gcagaggtga aagaggagat cgtggagaat ggagtgaaga agttggtgca cagtgtcttt	360
gacaccgcag actacacctt ccctttgcag gggaactctt tcttcgtgat gacaaacttt	420
ctcaaaacag aaggccaaga gcagcggttg tgtcccagat atcccaccg caggacgctc	480
tgttcctctg accgaggtg taaaaagga tggatggacc cgcagagcaa aggaattcag	540
accggaaggt gtgtagtgta tgaaggaac cagaagacct gtgaagtctc tgctggtgc	600
cccatcgagg cagtggaaga ggcccccg cctgctctct tgaacagtgc cgaaaacttc	660
actgtgctca tcaagaacaa tatcgacttc cccggccaca actacaccac gagaacatc	720
ctgccagggt taaacatcac ttgtacctc cacaagactc agaatccaca gtgtcccatt	780
ttccgactag gagacatctt ccgagaaaca ggcgataatt tttcagatgt ggcaattcag	840
ggcggataaa tgggcattga gatctactgg gactgcaacc tagaccgtg gttccatcac	900
tgccgtccca aatacagttt ccgtcgcctt gacgacaaga ccaccaacgt gtccttgtag	960
cctggctaca acttcagata cgccaagtac tacaaggaaa acaatgttga gaaacggact	1020
ctgataaaag tcttcgggat ccgttttgac atcctggttt ttggcaccgg aggaaaattt	1080
gacattatcc agctggttgt gtacatcggc tcaaccctct cctacttcgg tctggccgct	1140
gtgttcatcg acttctcat cgacacttac tccagtaact gctgtcgtc ccatatztat	1200
ccctggtgca agtgetgtca gccctgtgtg gtcaacgaat actactacag gaagaagtgc	1260
gagtccattg tggagccaaa gccgacatta aagtatgtgt cctttgtgga tgaatcccac	1320
attaggatgg tgaaccagca gctactaggg agaagtctgc aagatgtcaa gggccaagaa	1380

[0052]

gtcccaagac ctgcatgga cttcacagat ttgtccagge tgcccctggc cctccatgac	1440
acacccccga ttcttgaca accagaggag atacagctgc ttagaaagga ggcgactcct	1500
agatccaggg atagccccgt ctggtgccag tgtggaagct gcctcccatc tcaactcct	1560
gagagccaca ggtgcctgga ggagctgtgc tgccggaaaa agccgggggc ctgcatcacc	1620
acctcagagc tgttcaggaa gctggctctg tccagacacg tctgcagtt cctcctgctc	1680
taccaggagc ccttctgctg gctggatgtg gattccacca acagccggtc gcggcactgt	1740
gcctacaggt gctacgccac ctggcgcttc ggctcccagg acatggctga ctttgccaac	1800
ctgcccagct gctgccgctg gaggatccgg aaagagtttc cgaagagtga agggcagtac	1860
agtggcttca agagtcctta ctgaagccag gcaccgtggc tcacgtctgt aatcccagcg	1920
ctttgggagg ccgaggcagg cagatcacct gaggtcggga gttggagacc cgcctggcta	1980
acaaggcgaa atcctgtctg tactaaaaat acaaaaatca gccagacatg gtggcatgca	2040
cctgcaatcc cagctactcg ggaggtgag gcacaagaat cacttgaacc cgggaggcag	2100
aggtttagt gagcccagat tgtgccactg ctctccagcc tgggaggcac agcaaactgt	2160
ccccaaaaa aaaaaagag tccttaccaa tagcaggggc tgcagtagcc atgttaacat	2220
gacatttacc agcaactga acttcacctg caaagctctg tggccacatt ttcagccaaa	2280
gggaaatatg ctttcatctt ctgttctct ctgtgtctga gagcaaagt acctggttaa	2340
acaaaccaga atccctctac atggactcag agaaaagaga ttgagatgta agtctcaact	2400
ctgtccccag gaagttgtgt gaccctagcc ctctcacctc tgtgcctctg tctccttgtt	2460
gcccactac tatctcagag atattgtgag gacaaatga gacagtgcac atgaactgtc	2520
ttttaatgtg taaagatcta catgaatgca aaacatttca ttatgaggtc agactaggat	2580
aatgtccaac taaaacaaa cccttttcat cctggctgga gaatgtggag aactaaaggt	2640
ggccacaaat tctttgacac tcaagtcccc caagacctaa gggttttatc tcctcccctt	2700
gaatatgggt ggctctgatt gctttatcca aaagtggaag tgacattgtg tcagtttcag	2760
atcctgatct taagaggctg acagcttcta ctgctgtcc cttggaactc ttgctatcgg	2820
ggaagccaga cgccatttaa aagtctgcct atcctggcca ggtgtggtgg ctcacacctg	2880
taatcccagc actttgggag accaaggcgg gcggatcact taaagtcagg agtccaagac	2940
cagactcgcc aacatgggtga aaccgtatct ctaataaaaa taaaaaatt agctgggcat	3000
ggtgcgggca cctgtagtcc tagctatcaa gaggctgaga caggagaaac acttgaacct	3060

[0053]

gggaggtgga ggttgcattg agctgagatc gtgccactgc actccaggct gggtagacaga 3120  
 gcgagactcc atctcaaaaa aaaaaaaaaag aaaaa 3155

<210> 111  
 <211> 871  
 <212> DNA  
 <213> 智人

<400> 111  
 ctgccaggca gtgcccttcc cggagcgtgc cctcgccgct gagctcccct gaacagcagc 60  
 tgcagcagcc atggccccgc cctgggtgcc cgccatgggc ttcacgctgg cggccagcct 120  
 ggggtgcttc gtgggctccc gctttgtcca cggcgagggt ctccgctggt acgccggcct 180  
 gcagaagccc tcgtggcacc cgccccactg ggtgctgggc cctgtctggg gcacgctcta 240  
 ctcagccatg gggtagcgtt cctacctggt ctggaaagag ctgggaggct tcacagagaa 300  
 ggtgtgtggt cccctgggcc tctacactgg gcagctggcc ctgaactggg catggcccc 360  
 catcttcttt ggtgccccgac aaatgggctg ggccctggtg gatctcctgc tggtcagtgg 420  
 ggcggcggca gccactaccg tggcctggta ccaggtgagc ccgctggccg cccgcctgct 480  
 ctacccttac ctggcctggc tggccttgcg gaccacactc aactactgcg tatggcggga 540  
 caaccatggc tggcatgggg gacggcggct gccagagtga gtgcccggcc caccaggac 600  
 tgcagctgca ccagcaggtg ccatcacgct tgtgatgtgg tggccgtcac gctttcatga 660  
 ccaactgggc tgctagtctg tcagggcctt ggcccagggg tcagcagagc ttcagagggt 720  
 gccccactg agccccacc cgggagcagt gtctgtgct ttctgcatgc ttagagcatg 780  
 ttcttgaac atggaatttt ataagctgaa taaagttttt gacttccttt aaaaaaaaaa 840  
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa a 871

<210> 112  
 <211> 2133  
 <212> DNA  
 <213> 智人

<400> 112  
 atgctgtcct tccagtacc cgcagtgtac cgcgacgaga ccgccataca ggattatcat 60  
 ggtcataaaa tttgtgacc ttacgcctgg cttgaagacc ccgacagtga acagactaag 120  
 gcctttgtgg aggcccagaa taagattact gtgccatttc ttgagcagtg tccatcaga 180  
 ggtttataca aagagagaat gactgaacta tatgattatc ccaagtatag ttgccacttc 240

[0054]

aagaaaggaa aacggtat	ttat	tttttac	aatacaggt	tgcagaacca	gcgagtatta	300
tatgtacagg attccttaga	gggtgaggcc	agagtgtcc	tggaccecaa	catactgtct		360
gacgatggca cagtggcact	ccgaggttat	gcgttcagcg	aagatgggtga	atattttgcc		420
tatggctctga gtgccagtgg	ctcagactgg	gtgacaatca	agttcatgaa	agttgatgg		480
gccaaagagc ttccagatgt	gcttgaaaga	gtcaagttca	gctgtatggc	ctggacccat		540
gatgggaagg gaatgttcta	caactcatac	cctcaacagg	atggaaaaag	tgatggcaca		600
gagacatcta ccaatctcca	ccaaaagctc	tactaccatg	tcttgggaac	cgatcagtc		660
gaagatattt tgtgtgctga	gtttctgat	gaacctaaat	ggatgggtgg	agctgagtta		720
tctgatgatg gctgctatgt	cttgttatca	ataagggag	gatgtgatcc	agtaaaccga		780
ctctggtaact gtgacctaca	gcaggaatcc	agtggcatcg	cgggaatcct	gaagtgggta		840
aaactgattg acaactttga	aggggaatat	gactacgtga	ccaatgaggg	ggcgggtgtc		900
acattcaaga cgaatcgcca	gtctcccaac	tatcgctga	tcaacattga	cttcagggat		960
cctgaagagt ctaagtggaa	agtacttggt	cctgagcatg	agaaagatgt	cttagaatgg		1020
atagcttggtg tcaggtccaa	cttcttggtc	ttatgctacc	tccatgacgt	caagaacatt		1080
ctgcagctcc atgacctgac	tactggtgct	ctccttaaga	ccttcccgt	cgatgtcggc		1140
agcattgtag ggtacagcgg	tcagaagaag	gacactgaaa	tcttctatca	gtttacttcc		1200
tttttatctc caggtatcat	ttatcactgt	gatctgacca	aagaggagct	ggagccaaga		1260
gtcttccgcg aggtgaccgt	gaaaggaatt	gatgcttctg	actaccagac	agtccagctt		1320
ttctacccta gcaaggatgg	tacgaagatt	ccaatgttca	ttgtgcataa	aaaaagcata		1380
aaattggatg gctctcatcc	agctttctta	tatggctatg	gcggcttcaa	catatccatc		1440
acaccaact acagtgttcc	caggcttatt	tttgtgagac	acatgggtgg	tatcctggca		1500
gtggccaaca tcagaggagg	tggcgaatat	ggagagacgt	ggcataaagg	tggtatcttg		1560
gccaacaac aaaactgctt	tgatgacttt	cagtgtgctg	ctgagtatct	gatcaaggaa		1620
ggttacacat ctcccaagag	gctgactatt	aatggagggt	caaatggagg	cctcttagtg		1680
gctgcttggtg caaatcagag	acctgacctc	tttggttgtg	ttattgccca	agttggagta		1740
atggacatgc tgaagtttca	taaataatcc	atcggccatg	cttgaccac	tgattatggg		1800
tgctcggaca gcaacaaca	ctttgaatgg	cttgtcaaat	actctccatt	gcataatgtg		1860
aagttaccag aagcagatga	catccagtac	ccgtccatgc	tgctcctcac	tgctgaccat		1920

[0055]

gatgaccgcg tggteccgct teactccctg aagttcattg ccacccttca gtacatcgtg 1980  
 ggccgcagca ggaagcaaag caacccccctg cttatccacg tggacaccaa ggcggggccac 2040  
 ggggcgggga agcccacagc caaagtgata gaggaagtct cagacatggt tgcgttcate 2100  
 gcgcggtgcc tgaacatcga ctggattccg taa 2133

<210> 113  
 <211> 1375  
 <212> DNA  
 <213> 智人

<400> 113  
 gcaaacagcc ggggctccag cgggagaacg ataatgcaaa gtgctatggt cttggctggt 60  
 caacacgact gcagacccat ggacaagagc gcaggcagtg gccacaagag cgaggagaag 120  
 cgagaaaaga tgaacggac ccttttaaaa gattggaaga cccgtttgag ctacttetta 180  
 caaaattcct ctactcctgg gaagcccaaa accggcaaaa aaagcaaaca gcaagcttcc 240  
 atcaagcctt ctctgagga agcacagctg tggtcagaag catttgacga gctgctagcc 300  
 agcaaatatg gtcttgctgc attcagggtt tttttaaaagt cggaattctg tgaagaaaat 360  
 attgaattct ggctggcctg tgaagacttc aaaaaaacca aatcacccca aaagctgtcc 420  
 tcaaaagcaa ggaaaatata tactgacttc atagaaaagg aagctccaaa agagataaac 480  
 atagattttc aaacaaaaac tctgattgcc cagaatatac aagaagctac aagtggctgc 540  
 tttacaactg ccagaaaag ggtatacagc ttgatggaga acaactctta tcctcgtttc 600  
 ttggagtcag aattctacca ggacttgtgt aaaaagccac aatcaccac agagcctcat 660  
 gctacatgaa atgtaaaagg gagcccagaa atggaggaca tttcattctt tttcctgagg 720  
 ggaaggactg tgacctgcca taaagactga cttgaattc agcctgggtg ttcaggaaac 780  
 atcactcaga actattgatt caaagttggg tagtgaatca ggaagccagt aactgactag 840  
 gagaagctgg tatcagaaca gttccctca ctgtgtacag aacgcaagaa gggaataggt 900  
 ggtctgaacg tgggtgtctca ctctgaaaag caggaatgta agatgatgaa agagacaatg 960  
 taatactggt ggtccaaaag catttaaaat caatagatct gggattatgt ggccttaggt 1020  
 agctgggtgt acatctttcc ctaaategat ccatgttacc acatagtagt tttagtttag 1080  
 gattcagtaa cagtgaagtg tttactatgt gcaacggtat tgaagttctt atgaccacag 1140  
 atcatcagta ctgttgctc atgtaatgct aaaactgaaa tggtcctgtg ttgcattggt 1200  
 aaaaatgatg tgtgaaatag aatgagtgct atggtgttga aaactgcagt gtccgttatg 1260

[0056]

agtgcacaaa atctgtcttg aaggcagcta cactttgaag tggctttga atacttttaa 1320  
 taaatttatt ttgataaata atattgaaca aaaaaaaaaa aaaaaaaaaa aaaaa 1375

<210> 114  
 <211> 1069  
 <212> DNA  
 <213> 智人

<400> 114  
 aagtaattcc tagaccgta ggtggccgca gagccggtta cctctggttc tgcgccagcg 60  
 tgccccaccc gcaggacggc cgggttcttt gatttgata ctttctaaaa ccaaacccga 120  
 gaggaagggc aggctcaggg tggggatgcc ctgaaatatt cgagagcagg accgtttcta 180  
 ctgaagagaa gtttacaaga acgctctgtc tggggcgggc gaggcctctg cgaggcgggt 240  
 ccgggagcga gggcagggcg tgggcccgcg gcccggggtc gggggagtcg ggggcaggaa 300  
 gagggggagg agacagggct gggggagcgc cctgccgagc gcccgccagg ctctcccgc 360  
 tcccgcgccg cctcccteta cccaccgcc gcacgtacta aggaaggcgc acagcccgcc 420  
 gcgctcgctt ctccgccccg cgtccagctc gccagctcg cccagcgtcc gccgcgctc 480  
 ggccaaggct tcaacggacc acacaaaaat gccatctcaa atggaacacg ccatggaaac 540  
 catgatgttt acatttcaca aattcgctgg ggataaaggc tacttaaca aggaggacct 600  
 gagagtactc atgaaaagg agttccctgg atttttggaa aatcaaaaag accctctggc 660  
 tgttgacaaa ataatgaagg acctggacca gtgtagagat ggcaaagtgg gcttccagag 720  
 cttcttttcc ctaattgcgg gcctcaccat tgcattcaat gactatcttg tagtacacat 780  
 gaagcagaag ggaaagaagt aggcagaaat gagcagttcg ctctccctg ataagagttg 840  
 tcccaaaggg tcgcttaagg aatctgcccc acagcttccc ccatagaagg atttcatgag 900  
 cagatcagga cacttagcaa atgtaaaaat aaaatctaac tctcatttga caagcagaga 960  
 aagaaaagtt aaataccaga taagcttttg atttttgtat tgtttgcatc cccttgcctt 1020  
 caataaataa agttcttttt tagttccaaa ttgaaaaaa aaaaaaaaa 1069

<210> 115  
 <211> 4535  
 <212> DNA  
 <213> 智人

<400> 115  
 ggtccctccc ctctggctc tggggctggg cgcgcacccc gccccgtagc gggcccctc 60

[0057]

cctggcgagc gcaaccccat ccagcgggag cgcggagccg cggccgcggg gaagcattaa	120
gtttattgc ctcaaagtga cgcaaaaatt ctcaagagc tctttggcgg cggtatcta	180
gagatcagac catgtgaggg cccgcgggta caaatacggc cgcgccggcg cccctccgca	240
cagccagcgc cgccgggtgc ctcgagggcg cgaggccagc ccgctgccc agcccgggac	300
cagcctcccc gcgcagcctg gcaggctctc tggaggcaag gcgaccttgc ttgccctctc	360
ttgcagaata acaaggggct tagccacagg agttgctggc aagtggaaag aagaacaaat	420
gagtcaatcc cgacgtgtca atcccagca tagagagctc ggaggtgatc cacaaatcca	480
agcaccaga gatcaattgg gatccttggc agatggacat cagtgtcatt tactaaccag	540
caggatggag acgacgcct tgaattctca gaagcagcta tcagcgtgtg aagatggaga	600
agattgtcag gaaaacggag ttctacagaa gttgttccc accccagggg acaaagtgga	660
gtccgggcaa atatccaatg ggtactcagc agttccaagt cctggtgcgg gagatgacac	720
acggcactct atcccagcga ccaccaccac cctagtggct gagcttcac aaggggaacg	780
ggagacctgg ggcaagaagg tggatttct tctctcagt attggctatg ctgtggacct	840
gggcaatgtc tggcgcttc cctacatatg ttaccagaat ggaggggggg cattcctcct	900
cccctacacc atcatggcca tttttggggg aatcccgtc ttttcatgg agctcgcact	960
gggacagtac caccgaaatg gatgcatttc aatatggagg aaaatctgcc cgattttcaa	1020
agggattggt tatgccatct gcacattgc cttttacatt gttcctact acaacacat	1080
catggcctgg gcgctatact acctcatctc ctcttcacg gaccagctgc cctggaccag	1140
ctgcaagaac tcttgaaca ctggcaactg caccaattac ttctccgagg acaacatcac	1200
ctggaccctc cattccagct cccctgctga agaattttac acgcgccacg tctgcagat	1260
ccaccggtct aaggggctcc aggacctggg gggcatcagc tggcagctgg ccctctgcat	1320
catgctgatc ttcactgtta tctacttcag catctggaaa ggcgtcaaga cctctggcaa	1380
ggtggtgtgg gtgacagcca cttccctta tatcatcctt tctgtcctgc tggtagggg	1440
tgccaccctc cctggagcct ggaggggtgt tctcttctac ttgaaacca attggcagaa	1500
actcctggag acaggggtgt ggatagatgc agccgctcag atctttctct ctcttggtcc	1560
gggctttggg gtctgctgg cttttgctag ctacaacaag ttcaacaaca actgctacca	1620
agatgccctg gtgaccagcg tggatgaactg catgacgagc ttcgtttcgg gatttgcac	1680
cttcacagtg ctcggttaca tggctgagat gaggaatgaa_gatgtgtctg aggtggccaa	1740

[0058]

agacgcaggt cccagcctcc tcttcatcac gtatgcagaa gcgatagcca acatgccagc	1800
gtccactttc tttgccatca tcttctttct gatgttaate acgctgggct tggacagcac	1860
gtttgcaggc ttggaggggg tgatcacggc tgtgctggat gagttcccac acgtctgggc	1920
caagcgccgg gagcggttcg tgctcgccgt ggtcatcacc tgcttctttg gatccctggt	1980
cacctgact tttggagggg cctacgtggt gaagctgctg gaggagtatg ccacggggcc	2040
cgcagtgctc actgtcgcgc tgatcgaagc agtcgctgtg tcttggttct atggcatcac	2100
tcagttctgc agggacgtga aggaaatgct cggcttcagc ccgggggtggt tctggaggat	2160
ctgctgggtg gccatcagcc ctctgtttct cctgttcate atttgcagtt ttctgatgag	2220
cccgccaaa ctacgacttt tccaatataa ttatccttac tggagtatca tcttgggtta	2280
ctgcatagga acctcatctt tcatttgcac cccacatat atagcttate ggttgatcat	2340
cactccaggg acatttaaag agcgtattat taaaagtatt accccagaaa caccaacaga	2400
aattccttgt ggggacatcc gcttgaatgc tgtgtaacac actcaccgag aggaaaaagg	2460
cttctccaca acctcctcct ccagttctga tgaggcacgc ctgccttctc ccctccaagt	2520
gaatgagttt ccagctaagc ctgatgatgg aagggccttc tccacagga cacagtctgg	2580
tgcccagact caaggcctcc agccacttat ttccatggat tcccctggac atattcccat	2640
ggtagactgt gacacagctg agctggccta ttttggacgt gtgaggatgt ggatggaggt	2700
gatgaaaacc acctatcat cagttaggat taggtttaga atcaagtctg tgaagctctc	2760
ctgtatcatt tcttggatg atcattggta tctgatatct gtttgcttct aaaggttca	2820
ctgttcatga atacgtaaac tgcgtaggag agaacagga tgctatctcg ctagccatat	2880
atcttctgag tagcatatat aattttattg ctggaatcta ctagaacctt ctaatccatg	2940
tgctgctgtg gcatcaggaa aggaagatgt aagaagctaa aatgaaaaat agtgtgtcca	3000
tgcaagcttg tgagtctgtg tatattgttg tttcagtgta ttcttatctc tagtccaata	3060
ttttgggccc attacaaata tatgaattcc ccaaattttt cttacattaa caaattctac	3120
caactcaatt gtgtatggag gttattattt gaagggtaca atcactacaa catgctctgc	3180
caccactcc ttttccagtg aactacttg agccacacac tttcctttac aggccagcct	3240
ctggcgtttg ctgcacctca ttgccacctt cctgtctctc tgtgctaaac attcaggaca	3300
gtgttccaca ggcagatctg gcctatttca ttagtcacca tggttggct gtgaagtacg	3360
ttgaaggtgg atcttgtcac atgccccttc agtgttcacc tggccctctg gtttaagtte	3420

[0059]

tgtctgcctt acgtgactga gtttgactgt ccaggttgct ttgctcggtg aagagaggag	3480
ggtaaactcg attctcggtt agcactgggt tatacagatc tggcacccta acctaaacca	3540
aggcatcttc actccaagag cagttggaga gtctgggtta gccttacgtg gacctcgccg	3600
ctcgcctggcg gtcacgattg tgagccctcc agataatfff taaggttgag tctaagtaag	3660
gctgcttggg aaatggtcag ctaagtaa at cacctttcat ttcacataag gcccttaata	3720
tagataagta aatttggcct ttgggtgtctc gtgactctca gaggcgtagg tagaggagca	3780
aattaatatt tgcagcatgg gaattcetta tcagaatfff gaggggaata aatcctcatc	3840
agagacaaaa ggacttaate atctggccac ctatcacttc agttctctgt ataaatgaaa	3900
tttaattcta acaaccttat aaaaagaagg tccagacagc agaggaaaca tctgtccaa	3960
ttctaggfff tctcctctg gctccttcc cccagcattg tctaccctgg cccacttct	4020
gcattctccc catgccctgc tatttctgat tctttgcttc tcttagcgag atactttct	4080
tatatgatag ctgctgagaa gtttcccaga actgctagag gaaaagaagt ggggaattta	4140
ggaaatatcc ctactgacc taactccatt atcttactc tttcttctt cctgccacct	4200
catgcccatt ctctttactg tctagcatgc tgaagaagg aagtgatcta aatgccagcg	4260
tgttcagtg taaatattag ttggtgcaaa agaaaaacca tgattacttt tgcactaacc	4320
taatagcttt gcaaatttta agaacttgc ttatgaagat attcggatat ggattctccc	4380
caccccat acttagacat tgttcaata tactactttt aaaaaaacac cttttcaaac	4440
agaattagcg ttttgccaag tctggtatta atggaattgt acaggagctt tgaagtttt	4500
caactttat taaactaaaa aaaaaaatc gaaaa	4535

<210> 116  
 <211> 1800  
 <212> DNA  
 <213> 智人

<400> 116	
actgcgacc ggagccgcc ggactgacgg agcccactgc ggtgcggcg ttggcgcg	60
cacggaggac ccgggcaggc atcgcaagcg accccgagcg gagccccgga gccatggccc	120
tgagcgagct ggcgctggtc cgctggctgc aggagagccg ccgctcgcg aagctcatcc	180
tgttcatcgt gttcctggcg ctgctgctgg acaacatgct gctcactgtc gtggccccca	240
tcatcccaag ttatctgtac agcattaagc atgagaagaa tgctacagaa atccagacgg	300

[0060]

ccaggccagt gcacactgcc tccatctcag acagcttcca gagcatcttc tcctattatg	360
ataactcgac tatggteacc gggaatgcta ccagagacct gacacttcat cagaccgcca	420
cacagcacat ggtgaccaac gcgtccgctg ttccttccga ctgtcccagt gaagacaaag	480
acctcctgaa tgaanaactg caagttggtc tgttgtttgc ctcgaaagcc accgtccagc	540
tcateaccaa cctttcata ggactactga ccaacagaat tggtatcca attccatat	600
ttgcgggatt ctgcatcatg tttgtctcaa caattatggt tgcttctcc agcagctatg	660
ccttctgct gattgccagg tcgctgcagg gcatcggtc gtctgtctcc tctgtggctg	720
ggatgggcat gcttgcaggt gtctacacag atgatgaaga gagaggcaac gtcattggaa	780
tcgccttggg aggctggcc atgggggtct tagtgggccc cccctcggg agtgtgtctt	840
atgagtttgt ggggaagacg gctccgttcc tgggtgctggc cgccctggta ctcttgatg	900
gagctattca gctctttgtg ctccagccgt cccgggtgca gccagagagt cagaagggga	960
caccctaac cacgtgctg aaggaccgt acatcctcat tgctgcaggc tccatctgct	1020
ttgcaaacat gggcatgcc atgctggagc cagccctgcc catctggatg atggagacca	1080
tgtgttcccg aaagtggcag ctgggcgttg ccttcttggc agctagtatc tcttatctca	1140
ttggaaccaa tatttttggg atacttgac acaaaatggg gaggtggctt tgtgtcttc	1200
tgggaatgat aattgttga gtcagcattt tatgtattcc atttgcaaaa aacatttatg	1260
gactcatagc tccgaacttt ggagttggtt ttgcaattgg aatggtggat tcgtcaatga	1320
tgctatcat gggctacctc gtagacctgc ggcacgtgtc cgtctatggg agtgtgtacg	1380
ccattgcgga tgtggcattt tgtatggggt atgctatagg tccttctgct ggtggtgcta	1440
ttgcaaaggc aattggattt ccatggctca tgacaattat tgggataatt gatattcttt	1500
ttgcccctct ctgctttttt cttcgaagtc cacctgcca aagaagaaaa atggctattc	1560
tcattgatca caactgcctt attaaaacaa aatgtacac tcagaataat atccagtcac	1620
atccgatagg tgaagatgaa gaatctgaaa gtgactgaga tgagatcctc aaaaatcacc	1680
aaagtgttta attgtataaa acagtgttcc cagtgcaca actcatccag aactgtctta	1740
gtcataccat ccatccctgg tgaagagta aaaccaaagg ttattatttc ctttccatgg	1800

- <210> 117
- <211> 1852
- <212> DNA
- <213> 智人

[0061]

<400> 117  
 accgccgaga ccgcgtccgc cccgcgagca cagagcctcg cctttgccga tccgccgccc 60  
 gtccacaccc gccgccagct caccatggat gatgatatcg ccgcgctcgt cgtcgacaac 120  
 ggctccggca tgtgcaaggc cggcttcgcg ggcgacgatg cccccgggc cgtcttcccc 180  
 tccatcgtgg ggcgccccag gcaccagggc gtgatggtgg gcatgggtca gaaggattcc 240  
 tatgtgggcg acgaggccca gagcaagaga ggcacccctca cctgaagta ccccatcgag 300  
 cacggcatcg tcaccaactg ggacgacatg gagaaaatct ggcaccacac cttctacaat 360  
 gagctgcgtg tggctcccga ggagcaccac gtgctgctga ccgaggcccc cctgaacccc 420  
 aaggccaacc gcgagaagat gaccagatc atgtttgaga cttcaacac cccagccatg 480  
 tacgttgcta tccaggctgt gctatccctg tacgcctctg gccgtaccac tggcatcgtg 540  
 atggactccg gtgacggggt caccacact gtgccatct acgaggggta tgccctcccc 600  
 catgccatcc tgcgtctgga cctggetgge cgggacctga ctgactacct catgaagatc 660  
 ctcaccgagc gcggctacag cttcaccacc acggccgagc gggaaatcgt gcgtgacatt 720  
 aaggagaagc tgtgetacgt cgcctggac ttcgagcaag agatggccac ggctgcttcc 780  
 agctcctccc tggagaagag ctacgagctg cctgacggcc aggtcaccac cattggcaat 840  
 gagcggttcc gctgcctga ggcactctc cagccttct tctgggcat ggagtctgt 900  
 ggcateccag aaactacett caactccate atgaagtgtg acgtggacat ccgcaaagac 960  
 ctgtacgcca acacagtgt gtctggcggc accaccatgt accctggcat tgccgacagg 1020  
 atgcagaagg agatcactgc cctggcacc agcacaatga agatcaagat cattgctcct 1080  
 cctgagcgca agtactcctg gtggatcggc ggctccatcc tggcctcgt gtccacctc 1140  
 cagcagatgt ggatcagcaa gcaggagtat gacgagtcg gccctccat cgtccaccgc 1200  
 aaatgcttct aggcggacta tgacttagtt gcgttacacc cttcttgac aaaacctaac 1260  
 ttgcgcagaa aacaagatga gattggcatg gctttattg tttttttgt tttgttttg 1320  
 ttttttttt ttttttgct tgactcagga tttaaaact ggaacgggta aggtgacagc 1380  
 agtcggttg agcgagcacc ccccaaagt cacaatgtgg ccgaggactt tgattgcaca 1440  
 ttgttgttt tttaatagtc attccaata tgagatgcgt ttttacagga agtcccttgc 1500  
 catcctaaaa gccacccac ttctctctaa ggagaatgga ccagtcctc cccaagtcca 1560  
 cacaggggag gtgatagcat tgctttcgtg taaattatg aatgcaaaat tttttaatc 1620  
 ttgccttaa tactttttta ttttgttta ttttgaatga tgagccttcg tgccccctc 1680  
 [0062]

tccccctttt ttgtcccca acttgagatg tatgaaggct tttggtctcc ctgggagtgg	1740
gtggaggcag ccagggtta cctgtacact gacttgagac cagttgaata aaagtgcaca	1800
ccttaaaaat gaaaaaaaa aaaaaaaaa aaaaaaaaa aaaaaaaaa aa	1852

<210> 118  
 <211> 987  
 <212> DNA  
 <213> 智人

<400> 118	
aatataagtg gaggcgtcgc gctggcgggc attcctgaag ctgacagcat tggggccgag	60
atgtctcgct ccgtggcctt agctgtgctc gcgctactct ctctttctgg cctggaggct	120
atccagcgta ctcaaagat tcaggtttac tcacgtcatc cagcagagaa tggaaagtca	180
aatttcctga attgctatgt gtctgggttt catccatccg acattgaagt tgacttactg	240
aagaatggag agagaattga aaaagtggag cattcagact tgtctttcag caaggactgg	300
tctttctatc tcttgtacta cactgaattc acccccactg aaaaagatga gtatgcctgc	360
cgtgtgaacc atgtgacttt gtcacagccc aagatagtta agtgggatcg agacatgtaa	420
gcagcatcat ggaggtttga agatgccgca tttggattgg atgaattcca aattctgctt	480
gcttgctttt taatattgat atgcttatac acttacactt tatgcacaaa atgtagggtt	540
ataataatgt taacatggac atgatcttct ttataattct actttgagtg ctgtctccat	600
gtttgatgta tctgagcagg ttgctccaca ggtagctcta ggagggttg caacttagag	660
gtggggagca gagaattctc ttatccaaca tcaacatctt ggtcagattt gaactcttca	720
atctcttgca ctcaaagctt gttaagatag ttaagcgtgc ataagttaac ttccaattta	780
catactctgc ttagaatttg ggggaaaatt tagaaatata attgacagga ttattggaaa	840
tttgtataa tgaatgaaac attttgtcat ataagattca tatttacttc ttatacattt	900
gataaagtaa ggcattggtg tggttaatct ggtttatfff tgttccacaa gttaaataaa	960
tcataaaact tgatgtgtta tctctta	987

<210> 119  
 <211> 1310  
 <212> DNA  
 <213> 智人

<400> 119	
aaattgagcc cgcagcctcc cgcttegctc tctgctcctc ctgttcgaca gtcagccgca	60

[0063]

tcttcttttg	cgtcgccagc	cgagccacat	cgctcagaca	ccatggggaa	ggtgaaggtc	120
ggagtcaacg	gatttggtcg	tattgggcgc	ctggtcacca	ggctgcttt	taactctggt	180
aaagtggata	ttgttgccat	caatgacccc	ttcatlgacc	tcaactacat	ggtttacatg	240
ttccaatag	attccaccca	tggcaaattc	catggcaccg	tcaaggctga	gaacgggaag	300
cttgtcatca	atggaaatcc	catcaccatc	ttccaggagc	gagatccctc	caaaatcaag	360
tggggcgatg	ctggcgctga	gtacgtcgtg	gagtccactg	gcgtcttcac	caccatggag	420
aaggctgggg	ctcatttgca	ggggggagcc	aaaagggtca	tcatctctgc	cccctctgct	480
gatgccccca	tgttcgtcat	gggtgtgaac	catgagaagt	atgacaacag	cctcaagatc	540
atcagcaatg	cctcctgcac	caccaactgc	ttagcacccc	tggccaaggt	catccatgac	600
aactttggta	tcgtggaagg	actcatgacc	acagtccatg	ccatcactgc	caccagaag	660
actgtggatg	gccccctcgg	gaaactgtgg	cgtgatggcc	gcggggctct	ccagaacatc	720
atccctgcct	ctactggcgc	tgccaaaggct	gtgggcaagg	tcatccctga	gctgaacggg	780
aagctcactg	gcatggcctt	ccgtgtcccc	actgccaacg	gtcagtggt	ggacctgacc	840
tgccgtctag	aaaaacctgc	caaatatgat	gacatcaaga	aggtggtgaa	gcaggcgteg	900
gagggcccc	tcaagggcat	cctgggtac	actgagcacc	aggtggtctc	ctctgacttc	960
aacagcgaca	cccactcctc	caccttgac	gctggggctg	gcattgccct	caacgaccac	1020
tttgtcaagc	tcatttctg	gtatgacaac	gaatttggt	acagcaacag	ggtggtggac	1080
ctcatggccc	acatggcctc	caaggagtaa	gaccctgga	ccaccagccc	cagcaagagc	1140
acaagaggaa	gagagagacc	ctcactgctg	gggagtccct	gccacactca	gtccccacc	1200
acactgaatc	tccccctcctc	acagttgcca	tgtagacccc	ttgaagaggg	gaggggccta	1260
gggagccgca	ccttgtcatg	taccatcaat	aaagtaccct	gtgctcaacc		1310

<210> 120  
 <211> 2245  
 <212> DNA  
 <213> 智人

<400> 120						
acgcgacccg	ccctacgggc	acctcccgcg	cttttcttag	cgccgcagac	ggtggccgag	60
cgggggaccg	ggaagcatgg	cccgggggtc	ggcggttgc	tgggcggcgc	tcgggcccgtt	120
gttgtggggc	tgccgcctgg	ggctgcaggg	cgggatgctg	taccccagg	agagcccgtc	180

[0064]

gcgggagtgc aaggagctgg acggcctctg gagcttccgc gccgacttct ctgacaaccg	240
acgccggggc ttcgaggagc agtgggtaccg gcggccgctg tgggagtcag gccccaccgt	300
ggacatgcca gttccctcca gcttcaatga catcagccag gactggcgtc tgcggcattt	360
tgtcggctgg gtgtgggtacg aacgggaggt gatcctgccg gagcgatgga cccaggacct	420
gcgcacaaga gtgggtgctga ggattggcag tgcccattcc tatgccatcg tgtgggtgaa	480
tggggctcgc acgctagagc atgagggggg ctacctcccc ttcgaggccg acatcagcaa	540
cctggctccag gtggggcccc tgccctcccg gctccgaate actatcgcca tcaacaacac	600
actaccccc accaccctgc caccaggac catccaatac ctgactgaca cctccaagta	660
tccaagggt tactttgtcc agaacacata ttttgacttt ttcaactag ctggactgca	720
gcggctctgta cttctgtaca cgacaccac cacctacatc gatgacatca ccgtcaccac	780
cagcgtggag caagacagtg ggctgggtgaa ttaccagatc tctgtcaagg gcagtaacct	840
gttcaagttg gaagtgcgtc ttttggatgc agaaaacaaa gtcgtggcga atgggactgg	900
gaccagggc caacttaagg tgccagggtg cagcctctgg tggccgtacc tgatgcacga	960
acgccctgcc tatctgtatt cattggaggt gcagctgact gcacagacgt cactggggcc	1020
tgtgtctgac ttctacacac tcctgtggg gatccgcact gtggctgtca ccaagagcca	1080
gttcctcatc aatgggaaac ctttctatct ccacgggtgc aacaagcatg aggatgcgga	1140
catccgaggg aagggttcg actggccgct gctgggtgaag gacttcaacc tgcttcgctg	1200
gcttgggtgcc aacgcttcc gtaccagcca ctaccctat gcagaggaag tgatgcagat	1260
gtgtgaccgc tatgggattg tggatcatga tgagtgtccc ggctggggcc tggcgctgcc	1320
gcagttcttc aacaacgttt ctctgcatca ccacatgcag gtgatggaag aagtgggtgcg	1380
tagggacaag aaccaccccg cggctgtgat gtggctctgtg gccaacgagc ctgcgtccca	1440
cctagaatct gctggctact acttgaagat ggtgatcgct cacaccaaact ccttggacct	1500
ctccccgct gtgaccttg tgagcaactc taactatgca gcagacaagg gggctccgta	1560
tgtggatgtg atctgtttga acagctacta ctcttggat cacgactacg ggcacctgga	1620
gttgattcag ctgcagctgg ccaccagtt tgagaactgg tataagaagt atcagaagcc	1680
cattattcag agcgagtatg gagcagaaac gattgcaggg tttcaccagg atccacctct	1740
gatgttcaact gaagagtacc agaaaagtct gctagagcag taccatctgg gtctggatca	1800
aaaacgcaga aaatacgtgg ttggagagct catttggat tttgccgatt tcatgactga	1860

[0065]

acagtcaccg acgagagtgc tggggaataa aaaggggatc ttcactcggc agagacaacc	1920
aaaaagtgca gcgttccttt tgcgagagag atactggaag attgccaatg aaaccaggta	1980
tccccactca gtagccaagt cacaatgttt ggaaaacagc ccgtttactt gagcaagact	2040
gataccacct gcgtgtccct tctccccga gtcagggcga cttccacagc agcagaacaa	2100
gtgcctcctg gactgttcac ggcagaccag aacgtttctg gcctgggttt tgtggtcatc	2160
tattctagca gggaaacta aaggtggaat taaaagattt tctattatgg aaataaagag	2220
ttggcatgaa agtggctact gaaaa	2245

<210> 121  
 <211> 1526  
 <212> DNA  
 <213> 智人

<400> 121	
ccggaagtga cgcgaggctc tgcggagacc aggagtcaga ctgtaggacg acctcgggtc	60
ccacgtgtcc ccggtactcg ccggccggag cccccggctt cccggggccg ggggacctta	120
gcggcaccca cacacagcct actttccaag cggagccatg tctggtaacg gcaatgcggc	180
tgcaacggcg gaagaaaaca gcccaaagat gagagtgatt cgcgtgggta cccgcaagag	240
ccagcttgct cgcatacaga cggacagtgt ggtggcaaca ttgaaagcct cgtaccctgg	300
cctgcagttt gaaatcattg ctatgtccac cacaggggac aagattcttg atactgcact	360
ctctaagatt ggagagaaaa gcctgtttac caaggagctt gaacatgcc tggagaagaa	420
tgaagtggac ctggttgctt actccttgaa ggacctgcc actgtgcttc ctctggctt	480
caccatcgga gccatctgca agcgggaaaa ccctcatgat gctgttgtct ttcacccaaa	540
attgttggg aagaccctag aaaccctgcc agagaagagt gtggtgggaa ccagctccct	600
gcgaagagca gccagctgc agagaaagtt cccgcatctg gagttcagga gtattcgggg	660
aaacctcaac acccggttc ggaagctgga cgagcagcag gagttcagtg ccatcatcct	720
ggcaacagct ggctgcagc gcatgggctg gcacaaccgg gtggggcaga tctgcaccc	780
tgaggaatgc atgtatgctg tgggccaggg ggccttgggc gtggaagtgc gagccaagga	840
ccaggacatc ttggatctgg tgggtgtgct gcacgatccc gagactctgc ttcgctgcat	900
cgtgaaagg gccttctga ggcacctgga aggaggctgc agtgtgccag tagccgtgca	960
tacagctatg aaggatggc aactgtacct gactggagga gtctggagtc tagacggctc	1020
agatagcata caagagacca tgcaggctac catccatgct cctgcccagc atgaagatgg	1080

[0066]

cctgaggat gaccacagt tggtaggcat cactgctcgt aacattccac gagggcccca	1140
gttgctgcc cagaacttgg gcatcagcct ggccaacttg ttgctgagca aaggagccaa	1200
aaacatcctg gatgttgac ggcagcttaa cgatgccat taactggttt gtggggcaca	1260
gatgcctggg ttgctgctgt ccagtgccta catcccgggc ctcagtgcc cattctcact	1320
gctatctggg gagtgattac cccgggagac tgaactgcag ggttcaagcc ttccagggat	1380
ttgcctcacc ttggggcctt gatgactgcc ttgcctcctc agtatgtggg ggcttcatct	1440
ctttagagaa gtccaagcaa cagccttga atgtaaccaa tctactaat aaaccagttc	1500
tgaaggtgta aaaaaaaaa aaaaaa	1526

<210> 122

<211> 1435

<212> DNA

<213> 智人

<400> 122

ggcggggcct gcttctctc agcttcagc ggctgcgac agccctcagg cgaacctctc	60
ggctttcccg cgcggcgccg cctcttgctg cgctccgcc tctcctctg ctccgccacc	120
ggcttctctc tctgagcag tcagcccgcg cgcggccgg ctccgttatg gcgaccgca	180
gccctggcgt cgtgattagt gatgatgaac caggttatga ccttgattta tttgcatac	240
ctaatcatta tgctgaggat ttgaaaggg tgtttattcc tcatggacta attatggaca	300
ggactgaacg tcttgctcga gatgtgatga aggagatggg aggccatcac attgtagccc	360
tctgtgtgct caagggggc tataaattct ttgctgacct gctggattac atcaaagcac	420
tgaatagaaa tagtgataga tccattccta tgactgtaga ttttatcaga ctgaagagct	480
attgtaatga ccagtcaaca gggacataa aagtaattgg tggagatgat ctctcaactt	540
taactgaaa gaatgtcttg attgtggaag atataattga cactggcaaa acaatgcaga	600
ctttgctttc cttggtcagg cagtataatc caaagatggg caaggtcgca agcttgctgg	660
tgaaaaggac cccacgaagt gttggatata agccagactt tgttggattt gaaattccag	720
acaagtttgt tgtaggatat gcccttgact ataatgaata cttcagggat ttgaatcatg	780
tttgtgtcat tagtgaaact ggaaaagcaa aatacaaagc ctaagatgag agttcaagtt	840
gagtttgaa acatctggag tctattgac atcgccagta aaattatcaa tgttctagtt	900
ctgtggccat ctgcttagta gagcttttg catgtatctt ctaagaattt tatctgtttt	960

[0067]

gtactttaga	aatgtcagtt	gctgcattcc	taaactgttt	atttgacta	tgagcctata	1020
gactatcagt	tcctttggg	cggattgttg	ttaacttgt	aatgaaaaa	attctcttaa	1080
accacagcac	tattgagtga	aacattgaac	tcatatctgt	aagaaataaa	gagaagatat	1140
attagttttt	taattggtat	ttaattttt	atatatgcag	gaaagaatag	aagtgattga	1200
atattgttaa	ttataccacc	gtgtgtaga	aaagtaagaa	gcagtcaatt	ttcacatcaa	1260
agacagcatc	taagaagttt	tgttctgtcc	tgaattatt	ttagtagtgt	ttcagtaatg	1320
ttgactgtat	ttccaactt	gttcaaatta	ttaccagtga	atctttgtca	gcagttcctt	1380
tttaaatgca	aatcaataaa	ttcccaaaaa	tttaaaaaaa	aaaaaaaaaa	aaaaa	1435

<210> 123  
 <211> 2439  
 <212> DNA  
 <213> 智人

<400> 123						
gagagcagcg	gccgggaagg	ggcggtcgg	gaggcggggt	gtggggcgg	agtgtgggccc	60
ctgttcctgc	ccgcgcggtg	ttccgcattc	tgcaagcctc	cggagcgcac	gtcggcagtc	120
ggctccctcg	ttgaccgaat	caccgacctc	tctccccagc	tgtatttcca	aatgtctgct	180
ttctaacaag	ctgacgctgg	acaagctgga	cgttaaagg	aagcgggtcg	ttatgagagt	240
cgacttcaat	gttcctatga	agaacaacca	gataacaaac	aaccagagga	ttaaggctgc	300
tgtcccaagc	atcaaattct	gcttggacaa	tggagccaag	tcggtagtcc	ttatgagcca	360
cctaggccgg	cctgatggtg	tgcccatgcc	tgacaagtac	tccttagagc	cagttgctgt	420
agaactcaaa	tctctgctgg	gcaaggatgt	tctgttcttg	aaggactgtg	taggcccaga	480
agtggagaaa	gcctgtgcca	accagctgc	tgggtctgtc	atcctgctgg	agaacctccg	540
ctttcatgtg	gaggaagaag	ggaagggaaa	agatgcttct	gggaacaagg	ttaaagccga	600
gccagccaaa	atagaagctt	tccgagcttc	actttccaag	ctaggggatg	tctatgtcaa	660
tgatgctttt	ggcactgctc	acagagccca	cagctccatg	gtaggagtca	atctgccaca	720
gaaggctggt	gggtttttga	tgaagaagga	gctgaactac	tttgcaaagg	ccttgagagag	780
cccagagcga	cccttcttgg	ccatcctggg	cggagctaaa	gttgacagaca	agatccagct	840
catcaataat	atgctggaca	aagtcaatga	gatgattatt	ggtggtggaa	tggcttttac	900
cttccttaag	gtgetcaaca	acatggagat	tggcacttct	ctgtttgatg	aagagggagc	960
caagattgtc	aaagacctaa	tgtccaaagc	tgagaagaat	ggtgtgaaga	ttaccttgcc	1020

[0068]

tgttgacttt gtcactgctg acaagtttga tgagaatgcc aagactggcc aagccactgt	1080
ggcttctggc atacctgctg gctggatggg cttggactgt ggtcctgaaa gcagcaagaa	1140
gtatgctgag gctgtcactc gggctaagca gattgtgtgg aatggtcctg tgggggtatt	1200
tgaatgggaa gcttttgccc ggggaaccaa agctctcatg gatgaggtgg tgaagccac	1260
ttctaggggc tgcacacca tcataggtgg tggagacact gccacttgct gtgccaaatg	1320
gaacacggag gataaagtca gccatgtgag cactgggggt ggtgccagtt tggagctcct	1380
ggaaggtaaa gtccttctg ggggtgatgc tctcagcaat atttagtact ttctgcctt	1440
ttagttctg tgcacagccc ctaagtcaac ttagcatttt ctgcatctcc acttggcatt	1500
agctaaaacc ttccatgtca agattcagct agtggccaag agatgcagtg ccaggaacct	1560
ttaaacagtt gcacagcacc tcagctcacc ttcactgcac cctggatttg catacattct	1620
tcaagatccc atttgaattt tttagtact aaaccattgt gcattctaga gtgcatatat	1680
ttatattttg cctgttaaaa agaaagtgag cagtgttagc ttagttctct tttgatgtag	1740
gttattatga ttagctttgt cactgtttca ctactcagca tggaaacaag atgaaattcc	1800
atgttaggt agtgagacaa aattgatgat ccattaagta aacaataaaa gtgtccattg	1860
aaaccgtgat tttttttttt ttctgtcat actttgttag gaagggtgag aatagaatct	1920
tgaggaacgg atcagatgtc tatattgctg aatgcaagaa gtggggcagc agcagtgagg	1980
agatgggaca attagataaa tgtccattct ttatcaaggg cctactttat ggcagacatt	2040
gtgctagtgc ttttattcta acttttattt ttatcagtta cacatgatca taatttaaaa	2100
agtcaagget tataacaaaa aagccccagc ccattcctcc cattcaagat tcccactccc	2160
cagaggtgac cactttcaac tcttgagttt ttcaggtata tacctccatg tttctaagta	2220
atatgettat attgttcaact tctttttttt ttatttttta aagaaatcta tttcatacca	2280
tggaggaagg ctctgttcca catatatttc cacttcttca ttctctcggg atagttttgt	2340
cacaattata gattagatca aaagtctaca taactaatac agctgagcta tgtagtatgc	2400
tatgattaaa ttacttatg taaaaaaaaa aaaaaaaaaa	2439

<210> 124  
 <211> 2276  
 <212> DNA  
 <213> 智人

<400> 124

[0069]

gaacgtggta	taaaaggggc	gggaggccag	gctcgtgccg	ttttgcagac	gccaccgccg	60
aggaaaaccg	tgtactatta	gccatggtea	accccaccgt	gttcttcgac	attgccgtcg	120
acggcgagcc	cttggggccgc	gtctccittg	agctgtttgc	agacaaggte	ccaaagacag	180
cagaaaatth	tctgtctctg	agcactggag	agaaaggatt	tggttataag	ggttcctgct	240
ttcacagaat	tattccaggg	tttatgtgtc	agggtggatg	cttcacacgc	cataatggca	300
ctgggtggca	gtccatctat	ggggagaaat	ttgaagatga	gaacttcata	ctaaagcata	360
cgggtcctgg	catcttgtcc	atggcaaatg	ctggacccaa	cacaaatggt	tcccagtttt	420
tcatctgcac	tgccaagact	gagtggttgg	atggcaagca	tgtgggtgtt	ggcaaagtga	480
aagaaggcat	gaatattgtg	gaggccatgg	agcgccttgg	gtccaggaat	ggcaagacca	540
gcaagaagat	caccattgct	gactgtggac	aactcgaata	agtttgactt	gtgttttatc	600
ttaaccacca	gatcattcct	tctgtagctc	aggagagcac	ccctccacce	catttgctcg	660
cagtatccta	gaatctttgt	gctctcgctg	cagttccctt	tgggttccat	gttttccttg	720
ttccctccca	tgccatagctg	gattgcagag	ttaaagttat	gattatgaaa	taaaactaa	780
ataacaattg	tcctcgtttg	agttaagagt	gttgatgtag	gctttattht	aagcagtaat	840
gggttacttc	tgaacatca	cttgtttgct	taattctaca	cagtacttag	atthttttta	900
ctttccagtc	ccaggaagtg	tcaatgtttg	ttgagtggaa	tattgaaaat	gtaggcagca	960
actgggcatg	gtggctcact	gtctgtaatg	tattacctga	ggcagaagac	cacctgaggg	1020
taggagtcaa	gatcagcctg	ggcaacatag	tgagacgctg	tctctacaaa	aaataattag	1080
cctggcctgg	tgggtcatgc	ctagtcctag	ctgatctgga	ggctgacgtg	ggaggattgc	1140
ttgagcctag	agtgagctat	tatcatgcca	ctgtacagcc	tgggtgttca	cagatcttgt	1200
gtctcaaagg	taggcagagg	caggaaaagc	aaggagccag	aattaagagg	ttgggtcagt	1260
ctgcagtgag	ttcatgcatt	tagaggtgtt	cttcaagatg	actaatgtca	aaaattgaga	1320
catctgttgc	ggthtttttt	ttthtttttt	cccctggaat	gcagtggcgt	gatctcagct	1380
cactgcagcc	tcgcctcct	gggttcaagt	gattctagtg	cctcagcctc	ctgagtagct	1440
gggataatgg	gctgtgtcca	ccatgccccag	ctaathtttg	tathtttagt	atagatgggg	1500
tttcatcatt	ttgaccaggc	tggctcmeta	ctcttgacct	cagctgatgc	gctgccttg	1560
gcctccmeta	ctgctgagat	tacagatgtg	agccaccgca	cctacctca	ttttctgtaa	1620
caaagctaag	cttgaacact	gttgatgttc	ttgagggag	catattgggc	tttaggctgt	1680

[0070]

aggccaagtt	tatacatctt	aattatggtg	gaattcctat	gtagagtcta	aaaagccagg	1740
tacttgggtgc	tacagtcagt	ctccctgcag	agggttaagg	cgagactac	ctgcagtgag	1800
gaggtactgc	ttgtagcata	tagagcctct	ccctagcttt	ggttatggag	gctttgaggt	1860
tttgcaaacc	tgaccaatth	aagccataag	atctgggtcaa	agggataccc	ttcccactaa	1920
ggacttgggtt	tctcaggaaa	ttatatgtac	agtgcttgct	ggcagttaga	tgtcaggaca	1980
atctaagctg	agaaaacccc	ttctctgccc	accttaacag	acctctaggg	ttcttaaccc	2040
agcaatcaag	tttgccctatc	ctagaggtgg	cggatttgat	catttgggtg	gttgggcaat	2100
ttttgtttta	ctgtctgggt	ccttctgcgt	gaattaccac	caccaccact	tgtgcatctc	2160
agtcttgggt	gttgtctgg	tacgtattcc	ctgggtgata	ccattcaatg	tcttaatgta	2220
cttgtggctc	agacctgagt	gcaaggtgga	aataaacatc	aaacatcttt	tcatta	2276

<210> 125  
 <211> 1229  
 <212> DNA  
 <213> 智人

<400> 125

gtctgacggg	cgatggcgca	gccaatagac	aggagcgcta	tccgcggttt	ctgattggct	60
actttgttcg	cattataaaa	ggcagcgcg	ggcgcgaggc	ccttctctcg	ccaggcgctc	120
togtgggaag	gacatcgtct	ttaaaccctg	cgtggcaatc	cctgacgcac	cgccgtgatg	180
cccaggggaag	acagggcgac	ctggaagtcc	aactacttcc	ttaagatcat	ccaactattg	240
gatgattatc	cgaaatgttt	cattgtggga	gcagacaatg	tgggctcaa	gcagatgcag	300
cagatccgca	tgtcccttcg	cgggaaggct	gtgggtctga	tgggcaagaa	cacatgatg	360
cgcaaggcca	tccgagggca	cctggaaaac	aaccagctc	tggagaaact	gctgcctcat	420
atccggggga	atgtgggctt	tgtgttcacc	aaggaggacc	tactgagat	caggacatg	480
ttgctggcca	ataagtgcc	agctgctgcc	cgtgctgggtg	ccattgcccc	atgtgaagtc	540
actgtgccag	cccagaacac	tggctctggg	cccgagaaga	cctccttttt	ccaggcttta	600
ggtatcacca	ctaaaatctc	caggggcacc	attgaaatcc	tgagtgatgt	gcagctgatc	660
aagactggag	acaaagtggg	agccagcgaa	gccacgctgc	tgaacatgct	caacatctcc	720
cccttctct	ttgggctggt	catccagcag	gtgttcgaca	atggcagcat	ctacaaccct	780
gaagtgett	atatcacaga	ggaaactctg	cattctcgtct	tcttgagggg	tgtccgcaat	840
gttgccagtg	tctgtctgca	gattggctac	ccaactgttg	catcagtagc	ccattctatc	900

[0071]

atcaacgggt acaaacgagt cctggccttg tctgtggaga cggattacac cttcccactt	960
gctgaaaagg tcaaggcctt cttggctgat ccactctgcct ttgtggctgc tgcccctgtg	1020
gctgctgcca ccacagctgc tcctgctgct gctgcagccc cagctaaggt tgaagccaag	1080
gaagagtcgg aggagtcgga cgaggatag ggatttggtc tctttgacta atcaccaaaa	1140
agcaaccaac ttagccagtt ttatttgcaa aacaaggaaa taaaggctta cttctttaa	1200
aagtaaaaaa aaaaaaaaaa aaaaaaaaaa	1229

<210> 126  
 <211> 1142  
 <212> DNA  
 <213> 智人

<400> 126	
cttttccaag cggctgccga agatggcggg ggtgcaggtc ctggtgcttg atggtcgagg	60
ccatctctcg ggccgcctgg cggccatcgt ggctaaacag gtactgctgg gccggaaggt	120
ggtggtcgta cgctgtgaag gcatcaacat ttctggcaat ttctacagaa acaagttgaa	180
gtacctggct ttctctcgca agcggatgaa caccaacct tcccagggcc cctaccactt	240
ccgggcccc agccgcctct tctggcggac cgtgcgaggt atgctgcccc acaaaaacaa	300
gcgaggccag gccgctctgg accgtctcaa ggtgtttgac ggcattccac cgcctacga	360
caagaaaaag cggatggtgg ttctgctgc cctcaaggtc gtgcgtctga agcctacaag	420
aaagtttgc tatctggggc gcctggctca cgaggttggc tggaagtacc aggcagtgc	480
agccaccctg gaggagaaga ggaaagagaa agccaagatc cactaccgga agaagaaaca	540
gctcatgagg ctacggaaac aggccgagaa gaacgtggag aagaaaattg acaaatacac	600
agaggtctc aagaccacg gactcctggt ctgagcccaa taaagactgt taattcctca	660
tgcgttgctt gcccttctc cattgttgc ctggaatgta cgggaccag gggcagcagc	720
agtccaggtg ccacaggcag ccctgggaca taggaagctg ggagcaagga aagggtctta	780
gtcactgcct cccgaagttg cttgaaagca ctcggagaat tgtgcaggtg tcatttatct	840
atgaccaata ggaagagcaa ccagttacta tgagtgaag ggagccagaa gactgattgg	900
agggccctat cttgtgagtg gggcatctgt tggactttcc acctggtcat atactctgca	960
gctgttagaa tgtgcaagca cttggggaca gcatgagctt gctgtttgtac acagggtatt	1020
tctagaagca gaaatagact gggaagatgc acaaccaagg ggttacaggc atcgcccatg	1080

[0072]

ctcctcacct gtattttgta atcagaaata aattgctttt aaagaaaaaa aaaaaaaaaa	1140
aa	1142
<210> 127	
<211> 2405	
<212> DNA	
<213> 智人	
<400> 127	
tccggcgtgg tgcgcaggcg cggatatccc cctccccgc cagctcgacc ccggtgtggt	60
gcgcaggcgc agtctgcgca gggactggcg ggactgcgcg gcggcaacag cagacatgtc	120
gggggtccgg ggccctgtgc ggctgctgag cgctcggcgc ctggcgctgg ccaaggcgtg	180
gccaacagtg ttgcaaacag gaaccogagg ttttcacttc actgttgatg ggaacaagag	240
ggcatctgct aaagtttcag attccatttc tgctcagtat ccagtagtgg atcatgaatt	300
tgatgcagtg gtggtaggcg ctggaggggc aggcttgcca gctgcatttg gcctttctga	360
ggcagggttt aatacagcat gtgttaccac gctgtttcct accaggtcac aactgttgc	420
agcacagggg ggaatcaatg ctgctctggg gaacatggag gaggacaact ggaggtggca	480
tttctacgac accgtgaagg gctccgactg gctgggggac caggatgcca tccactacat	540
gacggagcag gccccgcg ccgtggtcga gctagaaaat tatggcatgc cgttttagcag	600
aactgaagat gggaagattt atcagcgtgc atttggtgga cagagcctca agtttgaaa	660
gggcggggcag gcccatcggg gctgctgtgt ggctgatcgg actggccact cgctattgca	720
caccttatat ggaaggtctc tgcgatatga taccagctat tttgtggagt attttgcctt	780
ggatctcctg atggagaatg gggagtgccg tgggtgtcatc gcactgtgca tagaggacgg	840
gtccatccat cgcataagag caaagaacac tgttgttgcc acaggaggct acgggcgcac	900
ctacttcage tgcacgtctg cccacaccag cactggcgac ggcacggcca tgatcaccag	960
ggcaggcctt ccttgccagg acctagagtt tgttcagtc caccctacag gcatatatgg	1020
tgtgtgtgtg ctattacgg aaggatgtcg tggagaggga ggcattctca ttaacagtca	1080
aggcгааagg tttatggagc gatacgcgcc tgtcgcgaag gacctggcgt ctagagatgt	1140
ggtgtctcgg tccatgactc tggagatccg agaaggaaga ggctgtggcc ctgagaaaga	1200
tcacgtctac ctgcagctgc accacctacc tccagagcag ctggccacgc gctgcctgg	1260
catttcagag acagccatga tcttcgctgg cgtggacgtc acgaaggagc cgatccctgt	1320
cctccccacc gtgcattata acatgggagg cattccccacc aactacaagg ggcaggtcct	1380

[0073]

gaggcacgtg aatggccagg atcagattgt gcccgccctg tacgcctgtg gggaggccgc	1440
ctgtgcctcg gtacatgggtg ccaaccgcct cggggcaaac tcgctcttgg acctggttgt	1500
ctttggtcgg gcatgtgcc tgagcatcga agagtcatgc aggcctggag ataaagtccc	1560
tccaattaaa ccaaacgctg gggaagaatc tgtcatgaat cttgacaaat tgagatttgc	1620
tgatggaagc ataagaacat cggaactgcg actcagcatg cagaagtcaa tgcaaatca	1680
tgctgccgtg ttccgtgtgg gaagcgtgtt gcaagaaggt tgtgggaaaa tcagcaagct	1740
ctatggagac ctaaagcacc tgaagacgtt cgaccgggga atggtctgga acacggacct	1800
ggtggagacc ctggagctgc agaacctgat gctgtgtgcg ctgcagacca tctacggagc	1860
agaggcacgg aaggagtcac ggggcgcgca tgccagggaa gactacaagg tgcggattga	1920
tgagtacgat tactccaagc ccatccaggg gcaacagaag aagccctttg aggagcactg	1980
gaggaagcac accctgtcct atgtggacgt tggcactggg aaggtcactc tggaatatag	2040
accctgatac gacaaaactt tgaacgaggc tgactgtgcc accgtcccgc cagccattcg	2100
ctcctactga tgagacaaga tgtggtgatg acagaatcag cttttgtaat tatgtataat	2160
agctcatgca tgtgtccatg tcataactgt cttcatacgc ttctgcactc tggggaagaa	2220
ggagtacatt gaaggagat tggcacctag tggctgggag cttgccagga acccagtggc	2280
cagggagcgt ggcacttacc tttgtccctt gcttcattct tgtgagatga taaaactggg	2340
cacagctctt aaataaaata taaatgaaca aactttcttt tatttccaaa aaaaaaaaaa	2400
aaaaa	2405

<210> 128  
 <211> 1867  
 <212> DNA  
 <213> 智人

<400> 128	
ggttcgctgt ggcgggccc tgggccgccc gctgtttaac ttcgcttccg ctggcccata	60
gtgatctttg cagtgaccca gcagcatcac tgtttcttgg cgtgtgaaga taacccaagg	120
aattgaggaa gttgctgaga agagtgtgct ggagatgctc taggaaaaaa ttgaatagtg	180
agacgagttc cagcgaagg gtttctggtt tgccaagaag aaagtgaaca tcatggatca	240
gaacaacagc ctgccacctt acgctcaggg cttggcctcc cctcagggtg ccatgactcc	300
cggaatccct atctttagtc caatgatgcc ttatggcact ggactgacct cacagcctat	360
[0074]	

tcagaacacc aatagtctgt ctatittgga agagcaacaa aggcagcagc agcaacaaca	420
acagcagcag cagcagcagc agcagcaaca gcaacagcag cagcagcagc agcagcagca	480
gcagcagcag cagcagcagc agcagcagca gcaacaggca gtggcagctg cagccgttca	540
gcagtcaacg tcccagcagg caacacaggg aacctcagge caggcaccac agcttctcca	600
ctcacagact ctcaaacctg cacccttgcc gggcaccact ccaactgtatc cctcccccat	660
gactccccatg accccccatca ctcttgccac gccagcttcg gagagtctg ggattgtacc	720
gcagctgcaa aatattgtat ccacagtga tcttggttgt aaacttgacc taaagaccat	780
tgcacttcgt gcccgaaacg ccgaatataa tccaagcgg tttgctgcgg taatcatgag	840
gataagagag ccacgaacca cggcactgat tttcagttct gggaaaatgg tgtgcacagg	900
agccaagagt gaagaacagt ccagactggc agcaagaaaa tatgctagag ttgtacagaa	960
gttgggtttt ccagctaagt tcttggaact caagattcag aatatggtgg ggagctgtga	1020
tgtgaagttt cctataaggt tagaaggcct tgtgctcacc caccaacaat ttagtagtta	1080
tgagccagag ttatttctg gtttaacta cagaatgatc aaaccagaa ttgttctct	1140
tatttttggt tctgaaaag ttgtattaac aggtgctaaa gtcagagcag aaatttatga	1200
agcatttgaa aacatctacc ctattctaaa gggattcagg aagacgacgt aatggctctc	1260
atgtaccctt gcctccccca ccccttctt ttttttttt taaacaaatc agtttgtttt	1320
ggtaccttta aatggtggtg ttgtgagaag atggatgtg agttgcaggg tgtggcacca	1380
ggtgatgccc ttctgtaagt gccaccgcg ggatgccggg aaggggcatt atttgtgcac	1440
tgagaacacc ggcagcgtg actgtgagtt gctcataccg tgctgctatc tgggcagcgc	1500
tgcccattta tttatatgta gattttaaac actgctgttg acaagttggt ttgagggaga	1560
aaactttaag tgtaaagcc acctctataa ttgattggac ttttaattt taatgttttt	1620
ccccatgaac cacagttttt atatttctac cagaaaagta aaaatctttt ttaaaagtgt	1680
tgtttttcta atttataact cctaggggtt atttctgtgc cagacacatt ccacctctcc	1740
agtattgcag gacagaatat atgigttaat gaaatgaat ggctgtacat attttttct	1800
ttcttcagag tactctgtac aataaatgca gttataaaa gtgttaaaaa aaaaaaaaaa	1860
aaaaaaaa	1867

<210> 129  
 <211> 5241  
 <212> DNA

[0075]

<213> 智人

<400> 129

```

agagcgtcgg gatatcgggt ggcggctcgg gacggaggac gcgctagtgt gagtgcgggc      60
ttctagaact acaccgacce tcgtgtcctc ccttcactct gcggggctgg ctggagcggc      120
cgctccgggtg ctgtccagca gccatagga gccgcacggg gagcgggaaa gcggtcgcgg      180
ccccaggcgg ggcggccggg atggagcggg gccgcagacc tgtggggaag gggctgtggc      240
ggcgcctcga gcggtcgcag gttcttctgt gtggcagttc agaatgatgg atcaagctag      300
atcagcattc tetaacttgt ttggtggaga accattgtca tatacccggt tcagcctggc      360
tcggcaagta gatggcgata acagtcatgt ggagatgaaa cttgctgtag atgaagaaga      420
aaatgctgac aataacacaa aggccaatgt cacaaaacca aaaaggtgta gtggaagtat      480
ctgctatggg actattgctg tgatcgtctt tttcttgatt ggatttatga ttggctactt      540
gggctattgt aaagggtag aacaaaaaac tgagtgtgag agactggcag gaaccgagtc      600
tccagtgagg gaggagccag gagaggactt ccttcagca cgtcgcttat attgggatga      660
cctgaagaga aagttgtcgg agaaactgga cagcacagac ttcaccgca ccatcaagct      720
gctgaatgaa aattcatatg tcctctgta ggctggatct caaaaagatg aaaatcttgc      780
gttgtatggt gaaaatcaat ttcgtgaatt taaactcagc aaagtctggc gtgatcaaca      840
ttttgttaag attcaggtca aagacagcgc tcaaaactcg gtgatcatag ttgataagaa      900
cggtagactt gtttacctgg tggagaatcc tgggggttat gtggcgata gtaaggctgc      960
aacagttact ggtaaactgg tccatgctaa ttttggact aaaaaagatt ttgaggattt     1020
atacactcct gtgaatggat ctatagtgat tgtcagagca gggaaaatca ctttgcaga     1080
aaaggttgca aatgctgaaa gcttaaatgc aattggtgtg ttgatataca tggaccagac     1140
taaatttccc attgttaacg cagaactttc attctttgga catgctcacc tggggacagg     1200
tgacccttac acacctgat tcccttcctt caatcacact cagtttccac catctcggtc     1260
atcaggattg cctaataatac ctgtccagac aatctccaga gctgctgcag aaaagctggt     1320
tgggaatatg gaaggagact gtccctctga ctggaaaaca gactctacat gtaggatggt     1380
aacctcagaa agcaagaatg tgaagctcac tgtgagcaat gtgctgaaag agataaaaat     1440
tcttaacatc tttggagtta ttaaaggctt tgtagaacca gatcactatg ttgtagttgg     1500
ggcccagaga gatgcatggg gccctggagc tgcaaaatcc ggtgtaggca cagctctcct     1560
attgaaactt gccagatgt tctcagatat ggtcttaaaa gatgggtttc agcccagcag     1620

```

[0076]

aagcattatc tttgccagtt ggagtgctgg agactttgga tcggttggtg ccaactgaatg 1680  
gctagagggga tacctttcgt ccctgcattt aaaggettcc acttatatta atctggataa 1740  
agcggttctt ggtaccagca acttcaaggt ttctgccage ccaactgttg atacgcttat 1800  
tgagaaaaca atgcaaatg tgaagcatcc ggttactggg caatttctat atcaggacag 1860  
caactgggcc agcaaagttg agaaactcac tttagacaat gctgctttcc ctttccttgc 1920  
atattctgga atcccagcag tttctttctg tttttgcgag gacacagatt atccttattt 1980  
gggtaccacc atggacacct ataaggaact gattgagagg attcctgagt tgaacaaaagt 2040  
ggcacgagca gctgcagagg tcgctggtea gttcgtgatt aaactaaccc atgatgttga 2100  
attgaacctg gactatgaga ggtacaacag ccaactgctt tcatttgtga gggatctgaa 2160  
ccaatacaga gcagacataa aggaaatggg cctgagtta cagtggtctg attctgctcg 2220  
tggagacttc ttccgtgcta cttccagact aacaacagat ttcgggaatg ctgagaaaac 2280  
agacagattt gtcataaga aactcaatga tcgtgcatg agagtggagt atcacttcct 2340  
ctctccctac gtatctcaa aagagtctcc tttccgacat gtcttctggg gctccggctc 2400  
tcacacgtg ccagctttac tggagaactt gaaactgcgt aaacaaaata acggtgcttt 2460  
taatgaaacg ctgttcagaa accagttgce tctagctact tggactatc agggagctgc 2520  
aaatgcctc tctggtgacg tttgggacat tgacaatgag ttttaaatgt gataccata 2580  
gcttccatga gaacagcagg gtagtctggt ttctagactt gtgctgatcg tgctaaattt 2640  
tcagttaggc taaaaacct gatgttaaaa ttccatccca tcactttggt actactagat 2700  
gtcttttagc agcagctttt aatacagggt agataacctg tacttcaagt taaagtgaat 2760  
aaccacttaa aaaatgtcca tgatggaata ttccctatc tctagaattt taagtgttt 2820  
gtaatgggaa ctgcctcttt cctgttggtt ttaatgaaa tgcagaaac cagttatgtg 2880  
aatgatctct ctgaatccta agggctggtc tctgctgaag gttgtaagtg gtcgcttact 2940  
ttgagtgate ctccaacttc atttgatgct aaataggaga taccaggttg aaagacctc 3000  
tccaatgag atctaagcct ttccataagg aatgtagctg gtttctcat tcctgaaaga 3060  
aacagttaac tttcagaaga gatgggcttg ttttcttgc aatgaggtct gaaatggagg 3120  
tccttctgct ggataaaatg aggttcaact gttgattgca ggaataaggc cttaatatgt 3180  
taacctcagt gtcatttatg aaaagagggg accagaagcc aaagacttag tatattttct 3240  
ttctctgtgt ccttcccc ataagcctcc atttagttct ttgttatttt tgtttcttc 3300

[0077]

aaagcacatt gaaagagaac cagtttcagg tgtttagttg cagactcagt ttgtcagact 3360  
ttaaagaata atatgctgcc aaattttggc caaagtgtta atcttagggg agagctttct 3420  
gtccttttgg cactgagata tttattgttt atttatcagt gacagagttc actataaatg 3480  
gtgttttttt aatagaatat aattatcgga agcagtgcct tccataatta tgacagttat 3540  
actgtcgggt ttttttaaat aaaagcagca tctgctaata aaaccaaca gatactggaa 3600  
gttttgcatt tatggccaac acttaagggt ttagaaaac agccgtcagc caaatgtaat 3660  
tgaataaagt tgaagctaag atttagagat gaattaaatt taattagggg ttgctaagaa 3720  
gcgagcactg accagataag aatgctgggt tcctaaatg cagtgaattg tgaccaagtt 3780  
ataaatcaat gtcacttaa ggctgtggta gtactcctgc aaaattttat agctcagttt 3840  
atccaagggt taactcta tccattttg caaaatttc agtaccttg tcacaatcct 3900  
aacacattat cgggagcagt gtcttcata atgtataaag aacaaggtag tttttaccta 3960  
ccacagtgtc tgtatcggag acagtgatct ccatatgtta cactaagggt gtaagtaatt 4020  
atcgggaaca gtgtttccca taattttctt catgcaatga catcttcaa gcttgaagat 4080  
cgttagtatc taacatgtat cccaactcct ataattcct atcttttagt tttagttgca 4140  
gaaacatfff gtggcatta agcattgggt gggtaaattc aaccactgta aatgaaatt 4200  
actacaaaat ttgaaattta gcttgggttt ttgttacctt tatggtttct ccaggctctc 4260  
tacttaatga gatagtagca tacatttata atgtttgcta ttgacaagtc attttaactt 4320  
tatcacatta tttgcatgtt acctcctata aacttagtgc ggacaagttt taatccagaa 4380  
ttgacctttt gacttaaagc agagggactt tgtatagaag gtttgggggc tgtggggaag 4440  
gagagtcccc tgaaggctctg acacgtctgc ctaccattc gtggtgatca attaaatgta 4500  
ggatgaata agttcgaagc tccgtgagtg aaccatcatt ataaacgtga tgatcagctg 4560  
tttgcataag ggcagttgga aacggcctcc tagggaaaag ttcataagggt ctcttcaggt 4620  
tcttagtgc acttacctag atttacagcc tcaattgaat gtgtcactac tcacagtctc 4680  
tttaatcttc agttttatct ttaatctcct cttttatctt ggactgacat ttagcgtagc 4740  
taagtgaaaa ggtcatagct gagattcctg gttcgggtgt tacgcacacg tacttaaatg 4800  
aaagcatgtg gcatgttcat cgtataacac aatatgaata caggcatgc atttgcagc 4860  
agtgagtctc ttcagaaaac cttttctac agttagggtt gagttacttc ctatcaagcc 4920  
agtacgtgct aacaggctca atattcctga atgaaatac agactagtga caagctcctg 4980

[0078]

gtcttgagat gtcttctcgt taaggagatg ggccttttgg aggtaaagga taaaatgaat 5040  
gagttctgtc atgattcact attctagaac ttgcatgacc tttactgtgt tagctctttg 5100  
aatgttcttg aaattttaga ctttctttgt aaacaaatga tatgtcctta tcattgtata 5160  
aaagctgtta tgtgcaacag tgtggagatt ccttgtctga ttaataaaa tacttaaaca 5220  
ctgaaaaaaaa aaaaaaaaaa a 5241

<210> 130  
<211> 2602  
<212> DNA  
<213> 智人

<400> 130  
gagccgcggc taaggaacgc gggccgcccc cccgctcccg gtgcagcggc ctccgcgccg 60  
ggttttggcg cctcccgcgg gcgccccct cctcacggcg agcgctgcca cgtcagacga 120  
agggcgcagc gagegtcctg atccttccgc ccggacgctc aggacagcgg cccgctgctc 180  
ataagactcg gccttagaac ccagtatca gcagaaggac attttaggac gggacttggg 240  
tgactctagg gcactggttt tctttccaga gagcggaaca ggcgaggaaa agtagtcctt 300  
tctcggcgat tctgcggagg gatctccgtg gggcggtgaa cgccgatgat tatataagga 360  
cgcgccgggt gtggcacagc tagttccgtc gcagccggga tttgggtcgc agttcttgtt 420  
tgtggatcgc tgtgatcgtc acttgacaat gcagatcttc gtgaagactc tgactggtaa 480  
gaccatcacc ctcgaggttg agcccagtga caccatcgag aatgtcaagg caaagatcca 540  
agataaggaa ggcacccctc ctgaccagca gaggctgata tttgctggaa aacagctgga 600  
agatgggcgc accctgtctg actacaacat ccagaaagag tccaccctgc acctggtgct 660  
ccgtctcaga ggtgggatgc aaatcttctg gaagacactc actggcaaga ccatcacctt 720  
tgaggtcgag cccagtgaca ccatcgagaa cgtcaaagca aagatccagg acaaggaagg 780  
cattcctcct gaccagcaga ggttgatctt tgccggaaag cagctggaag atgggcgcac 840  
cctgtctgac tacaacatcc agaaagagtc taccctgcac ctggtgctcc gtctcagagg 900  
tgggatgcag atcttctgta agaccctgac tggtaagacc atcacctcag aggtggagcc 960  
cagtgcacc atcgagaatg tcaaggcaaa gatccaagat aaggaaggca ttccttctga 1020  
tcagcagagg ttgatctttg ccggaaaaca gctggaagat ggtcgtaccc tgtctgacta 1080  
caacatccag aaagagtcca ccttgcacct ggtactccgt ctgagaggtg ggatgcaaat 1140

[0079]

cttcgtgaag	acactcactg	gcaagacat	cacccttgag	gtcgagccca	gtgacactat	1200
cgagaacgtc	aaagcaaaga	tccaagacaa	ggaaggcatt	cctcctgacc	agcagaggtt	1260
gatctttgcc	ggaaagcagc	tggaagatgg	gcgcacctg	tctgactaca	acatccagaa	1320
agagtctacc	ctgcacctgg	tgctccgtct	cagaggtggg	atgcagatct	tcgtgaagac	1380
cctgactggg	aagaccatca	ctctcgaagt	ggagccgagt	gacaccattg	agaatgtcaa	1440
ggcaaagatc	caagacaagg	aaggcatccc	tctgaccag	cagaggttga	tctttgccgg	1500
aaaacagctg	gaagatggtc	gtaccctgtc	tgactacaac	atccagaaag	agtccacctt	1560
gcacctgggtg	ctccgtctca	gaggtgggat	gcagatcttc	gtgaagacc	tgactggtaa	1620
gaccatcact	ctcgaggtgg	agccgagtga	caccattgag	aatgtcaagg	caaagatcca	1680
agacaaggaa	ggcatccctc	ctgaccagca	gaggttgatc	tttgctggga	aacagctgga	1740
agatggacgc	accctgtctg	actacaacat	ccagaaagag	tccaccctgc	acctgggtgct	1800
ccgtcttaga	ggtgggatgc	agatcttcgt	gaagaccctg	actggtaaga	ccatcactct	1860
cgaagtggag	ccgagtgaca	ccattgagaa	tgtcaaggca	aagatccaag	acaaggaagg	1920
catecctcct	gaccagcaga	ggttgatctt	tgctgggaaa	cagctggaag	atggacgcac	1980
cctgtctgac	tacaacatcc	agaaagagtc	caccctgcac	ctgggtctcc	gtcttagagg	2040
tgggatgcag	atcttcgtga	agaccctgac	tggttaagacc	atcactctcg	aagtggagcc	2100
gagtgacacc	attgagaatg	tcaaggcaaa	gatccaagac	aaggaaggca	tcctcctga	2160
ccagcagagg	ttgatctttg	ctgggaaaca	gctggaagat	ggacgcaccc	tgtctgacta	2220
caacatccag	aaagagtcca	ccctgcacct	ggtgctccgt	ctcagaggtg	ggatgcaaat	2280
cttcgtgaag	accctgactg	gtaagacat	caccctcgag	gtggagccca	gtgacacat	2340
cgagaatgtc	aaggcaaaga	tccaagataa	ggaaggcatc	cctcctgac	agcagaggtt	2400
gatctttgct	gggaaacagc	tggaagatgg	acgcaccctg	tctgactaca	acatccagaa	2460
agagtccact	ctgcacttgg	tctgcgctt	gagggggggt	gtctaagttt	ccccttttaa	2520
ggtttcaaca	aatttcattg	cactttcctt	tcaataaagt	tgttgcatte	ccaaaaaaaa	2580
aaaaaaaaaa	aaaaaaaaaa	aa				2602

<210> 131  
 <211> 3003  
 <212> DNA  
 <213> 智人

[0080]

<400> 131	
ctttctcctt ccccttcttc cgggctcccc tcccggctca tcacccggcc tgtggccac	60
tcccaccgcc agctggaacc ctggggacta cgacgtccct caaaccttgc ttctaggaga	120
taaaaagaac atccagtcac ggataaaaaat gagctggttc agaaggccaa actggccgag	180
caggctgagc gatatgatga catggcagcc tgcataaagt ctgtaactga gcaaggagct	240
gaattatcca atgaggagag gaatcttctc tcagttgctt ataaaaatgt tgtaggagcc	300
cgtaggtcat cttggagggt cgtctcaagt attgaacaaa agacggaagg tgctgagaaa	360
aacacgcaga tggctcgaga atacagagag aaaattgaga cggagctaag agatatctgc	420
aatgatgtac tgtctctttt ggaaaagttc ttgatcccca atgcttcaca agcagagagc	480
aaagtcttct atttgaaaat gaaaggagat tactaccgtt acttggctga ggttgccgt	540
ggtgatgaca agaaagggat tgtcgatcag tcacaacaag cataccaaga agcttttgaa	600
atcagcaaaa aggaaatgca accaacacat cctatcagac tgggtctggc ccttaacttc	660
tctgtgttct attatgagat tctgaactcc ccagagaaaag cctgctctct tgcaaagaca	720
gcttttgatg aagccattgc tgaacttgat acattaagtg aagagtcata caaagacagc	780
acgctaataa tgcaattact gagagacaac ttgacattgt ggacatcgga tacccaagga	840
gacgaagctg aagcaggaga aggaggggaa aattaaccgg ccttccaact tttgtctgcc	900
tcattctaaa atttacacag tagaccattt gtcattcatg ctgtcccaca aatagttttt	960
tgtttacgat ttatgacagg tttatgttac ttctatttga atttctatat ttccatgtg	1020
gtttttatgt ttaatattag gggagtagag ccagttaaca tttagggagt tatctgtttt	1080
catcttgagg tggccaatat ggggatgtgg aatttttata caagttataa gtgtttggca	1140
tagtactttt ggtacattgt ggcttcaaaa gggccagtgt aaaactgctt ccatgtctaa	1200
gcaaagaaaa ctgcctacat actggtttgt cctggcgggg aataaaaggg atcattggtt	1260
ccagtcacag gtgtagtaat tgtgggtact ttaaggttg gagcacttac aaggctgtgg	1320
tagaatcata ccccatggat accacatatt aaaccatgta tatctgtgga atactcaatg	1380
tgtacacctt tgactacagc tgcagaagtg ttcctttaga caaagtttg acccatttta	1440
ctctggataa gggcagaac ggttcacatt ccattatttg taaagttacc tgctgttagc	1500
tttcattatt tttgctacac tcattttatt tgtatttaaa tgttttaggc aacctaagaa	1560
caaatgtaaa agtaaagatg caggaaaaat gaattgcttg gtattcatta cttcatgtat	1620
atcaagcaca gcagtaaac aaaaacccat gtatttaact ttttttagg atttttgctt	1680

[0081]

ttgtgatttt ttttttttg atacttgcct aacatgcatg tgctgtaaaa atagttaaca	1740
gggaaataac ttgagatgat ggctagcttt gtttaatgtc ttatgaaatt ttcatgaaca	1800
atccaagcat aattgttaag aacacgtgta ttaaattcat gtaagtggaa taaaagtttt	1860
atgaatggac ttttcaacta ctttctctac agcttttcat gtaaattagt cttggttctg	1920
aaacttctct aaaggaaatt gtacattttt tgaaatttat tccttattcc ctcttggcag	1980
ctaattgggct cttaccaagt ttaaacacaa aatttatcat aacaaaaata ctactaatat	2040
aactactgtt tccatgtccc atgatcccct ctcttctec ccacctgaa aaaaatgagt	2100
tcctattttt tctgggagag ggggggattg attagaaaaa aatgtagtgt gttccattta	2160
aaattttggc atatggcatt ttctaactta ggaagccaca atgttcttgg cccatcatga	2220
cattgggtag cattaactgt aagttttgtg cttccaaate actttttggt ttttaagaat	2280
ttcttgatac tcttatagcc tgccttcaat ttgtatcett tattctttct atttgtcagg	2340
tgcacaagat taccttctg ttttagcctt ctgtcttgc accaaccatt cttacttgg	2400
ggccatgtac ttggaaaaag gccgcatgat ctttctgget ccaactcagt tctaaggcac	2460
cctgcttct ttgcttgcac cccacagact atttccctca tectatttac tgcagcaaat	2520
ctctcttag ttgatgagac tgtgtttatc tccctttaa accctaccta tcctgaatgg	2580
tctgtcattg tctgccttta aaatccttcc tctttctcc tectctattc tctaaataat	2640
gatggggcta agttataccc aaagctcact ttacaaaata tttctcagt actttgcaga	2700
aacacccaaa caaaaatgcc attttaaaaa aggtgtattt tttcttttag aatgtaagct	2760
cctcaagagc agggacaatg ttttctgtat gttctattgt gcctagtaca ctgtaaatgc	2820
tcaataaata ttgatgatgg gaggcagtga gtcttgatga taagggtgag aaactgaaat	2880
cccaaacact gttttgttgc ttgttttatt atgacctcag attaaattgg gaaatattgg	2940
cccttttgaa taattgtccc aaatattaca ttcaataaaa agtgcaatgg agaaaaaaaa	3000
aaa	3003

<210> 132  
 <211> 1869  
 <212> DNA  
 <213> 智人

<400> 132  
 tacctggttg atcctgccag tagcatatgc ttgtctcaaa gattaagcca tgcattgcta 60

[0082]

agtacgcacg gccggtacag tgaactgcg aatggctcat taaatcagtt atggttcctt	120
tggtcgctcg ctctctccc acttgataa ctgtggtaat tctagagcta atacatgccg	180
acgggcgctg acccccttcg cgggggggat gcgtgcattt atcagatcaa aaccaaccg	240
gtcagccctt ctccggcccc ggccgggggg cgggcgccgg cggctttggt gactctagat	300
aacctcgggc cgatcgcacg cccccgtgg cggcgacgac ccattcgaac gtctgccta	360
tcaactttcg atggtagtcg ccgtgcctac catggtgacc acgggtgacg gggaatcagg	420
gttcgattcc ggagagggag cctgagaaac ggctaccaca tccaaggaag gcagcaggcg	480
cgcaaattac ccaactccga cccggggagg tagtgacgaa aaataacaat acaggactct	540
ttcgaggccc tgtaattgga atgagtccac tttaaatcct ttaacgagga tccattggag	600
ggcaagtctg gtgccagcag ccgcggtaat tccagctcca atagcgtata ttaaagttgc	660
tgcagttaaa aagctcgtag ttggatcttg ggagcgggcg ggcggtccgc cgcgaggcga	720
gccaccgccc gtccccgccc cttagcctctc ggcgccccct cgatgctctt agctgagtgt	780
cccgcggggc ccgaagcggt tactttgaaa aaattagagt gttcaaagca ggcccagacc	840
gcctggatac cgcagctagg aataatggaa taggaccgcg gttctatctt gttggttttc	900
ggaactgagg ccatgattaa gagggacggc cgggggcatt cgtattgcgc cgctagaggt	960
gaaattcttg gaccggcgca agacggacca gagcgaaagc atttgccaag aatgttttca	1020
ttaatcaaga acgaaagtcg gaggttcgaa gacgatcaga taccgtcgta gttccgacca	1080
taaacgatgc cgaccggcga tgcggcggcg ttattcccat gaccgcccgg gcagcttccg	1140
ggaaacaaaa gtctttgggt tccgggggga gtaggttgc aaagctgaaa cttaaaggaa	1200
ttgacggaag ggcaccacca ggagtggagc ctgcggctta atttgactca acacgggaaa	1260
cctcaccgcg ccgggacacg gacaggattg acagattgat agctctttct cgattccgtg	1320
ggtggtggtg catggccggt cttagttggt ggagcgattt gtctggttaa ttccgataac	1380
gaacgagact ctggcatgct aactagttac gcgacccccg agcggtcggc gtcccccaac	1440
ttcttagagg gacaagtggc gttcagccac ccgagattga gcaataacag gtctgtgatg	1500
cccttagatg tccggggctg cacgcgcgct aactgactg gctcagcgtg tgcctaccct	1560
acgccggcag gcgcgggtaa cccggtgaac ccattcgtg atggggatcg gggattgcaa	1620
ttattcccca tgaacgagga attcccagta agtgccggctc ataagcttgc gttgattaag	1680
tccctgcctt ttgtacacac cgcccgtcgc tactaccgat tggatggttt agtgaggccc	1740

[0083]

---

tcggatcggc cccgccgggg tcggcccacg gccctggcgg agcgctgaga agacggtcga	1800
acttgactat ctagaggaag taaaagtcgt aacaaggttt ccgtaggtga acctgcggaa	1860
ggatcatta	1869

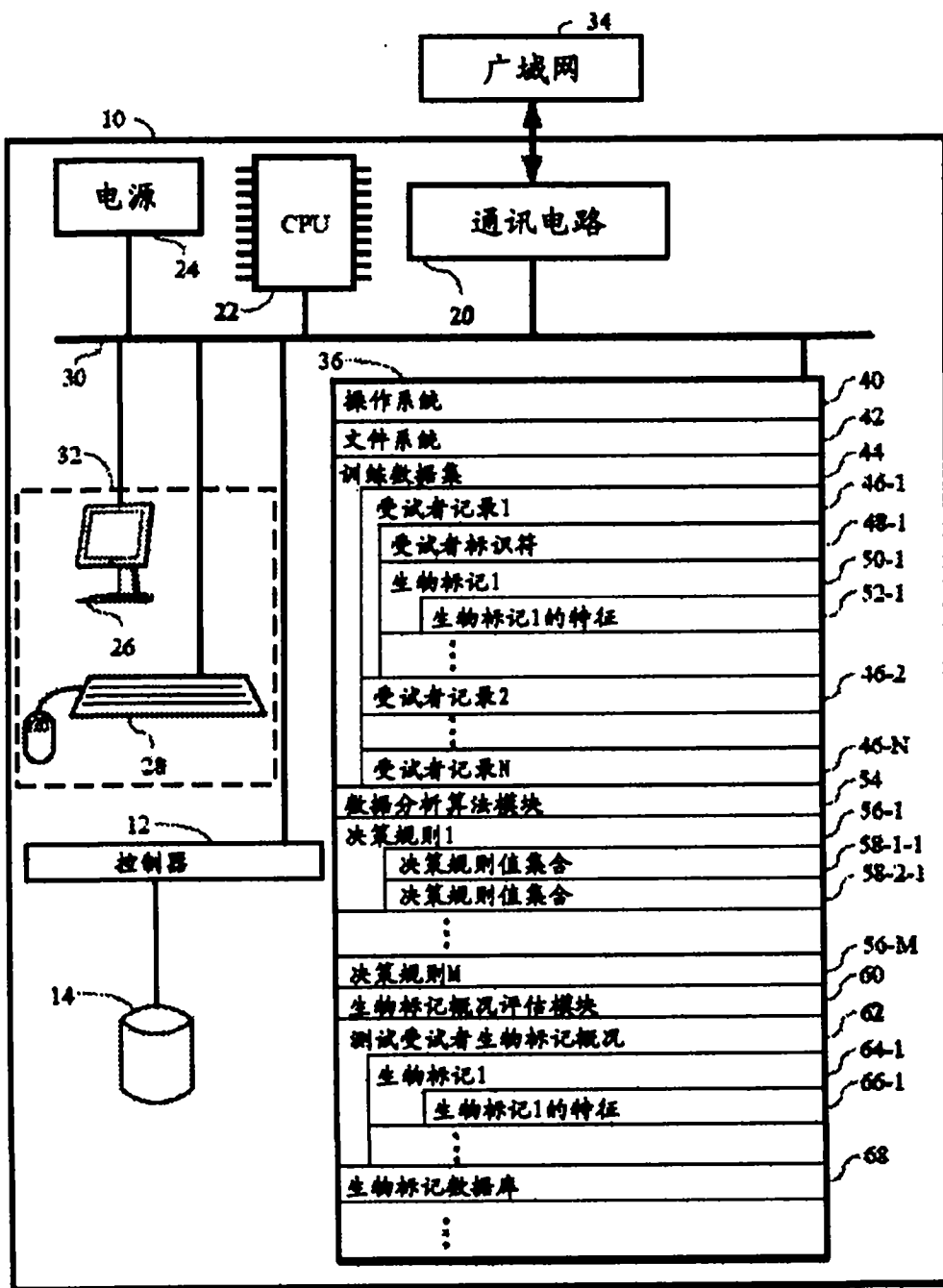


图 1

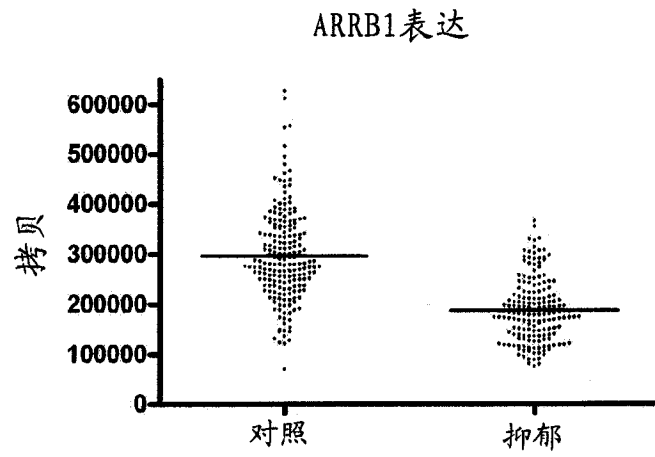


图 2A

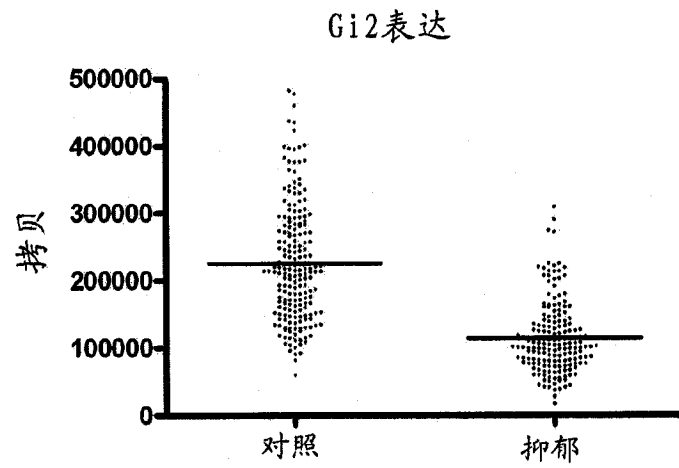


图 2B

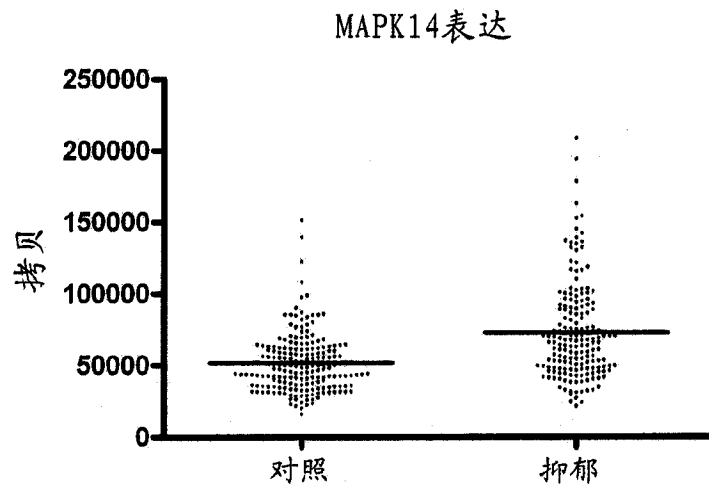


图 3A

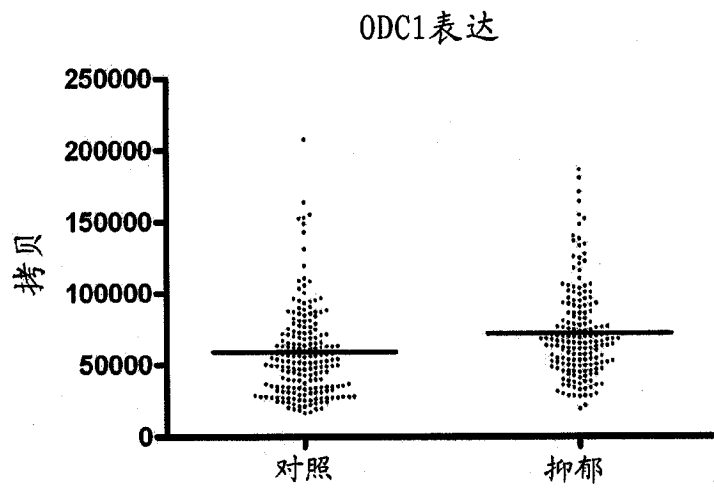


图 3B

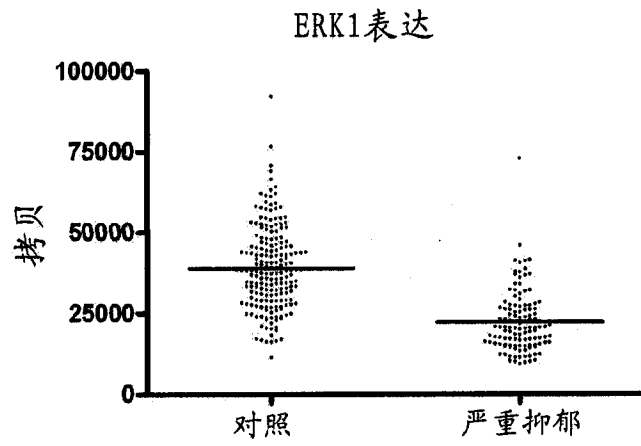


图 4A

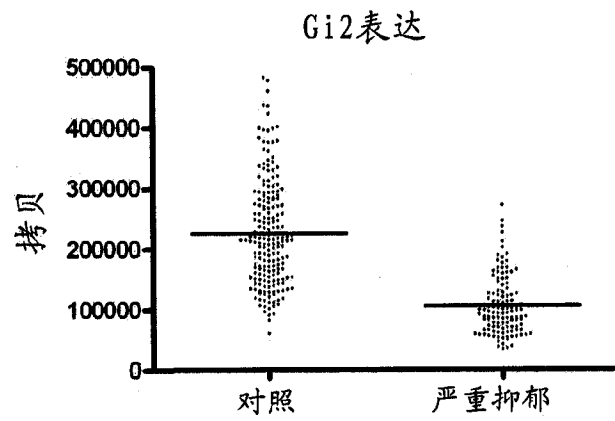


图 4B

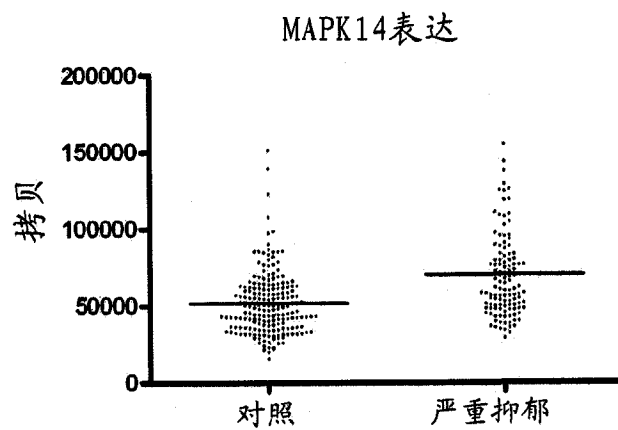


图 4C

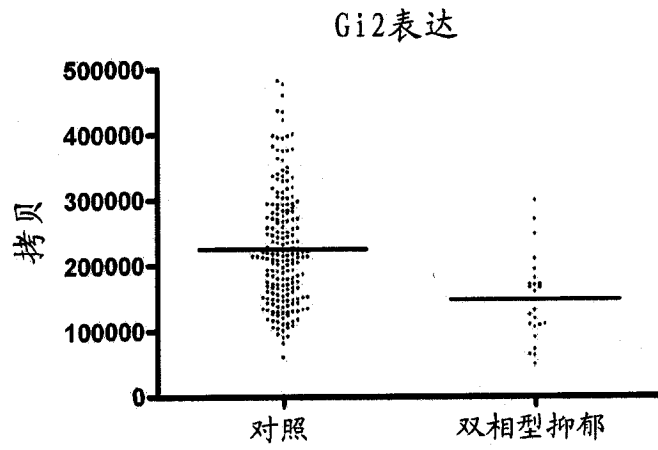


图 5A

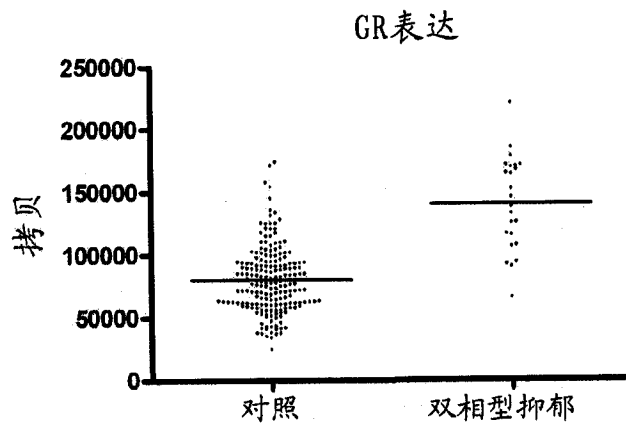


图 5B

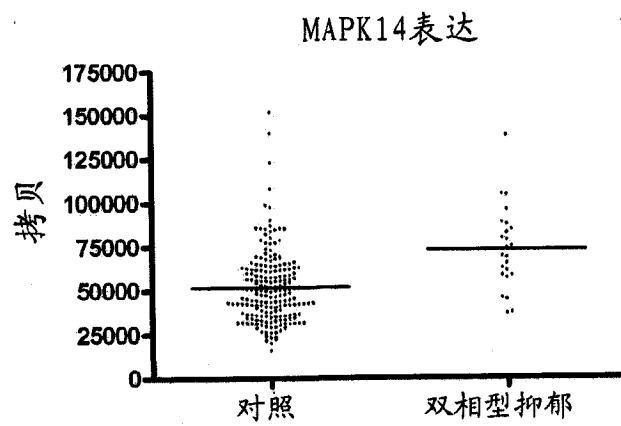


图 5C

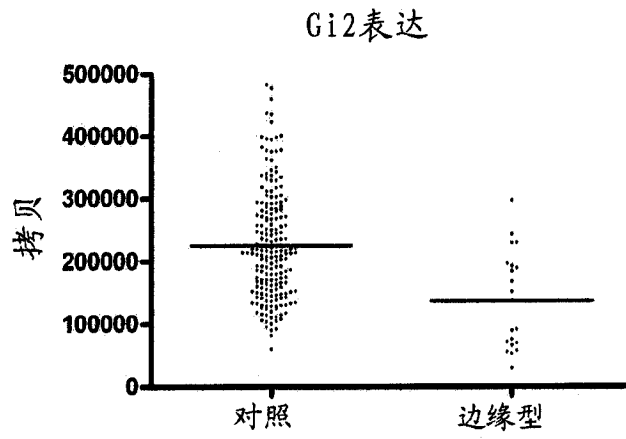


图 6A

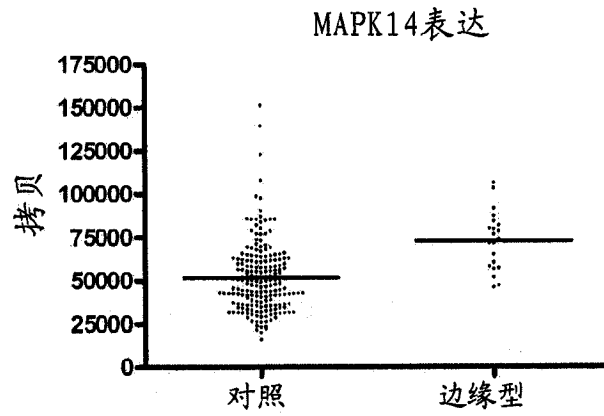


图 6B

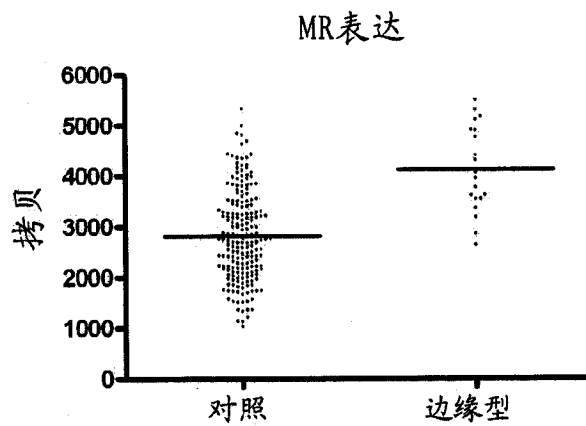


图 6C

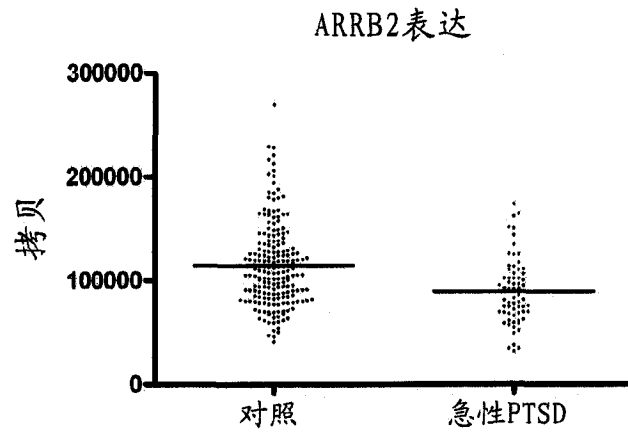


图 7A

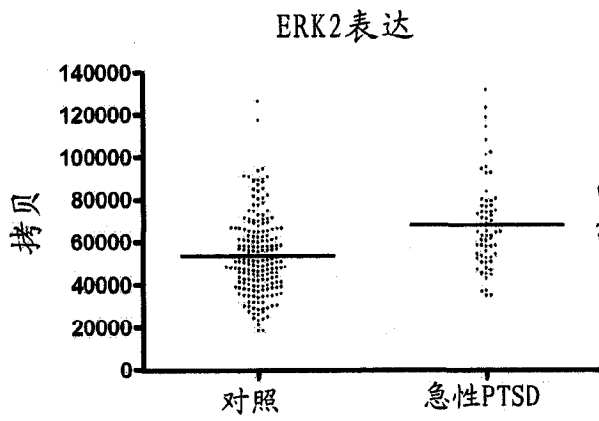


图 7B

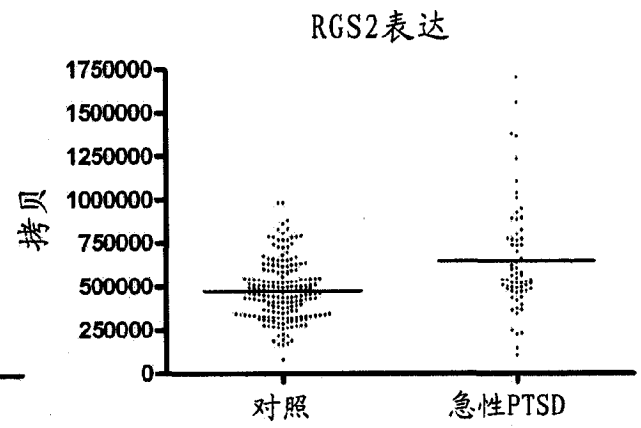


图 7C

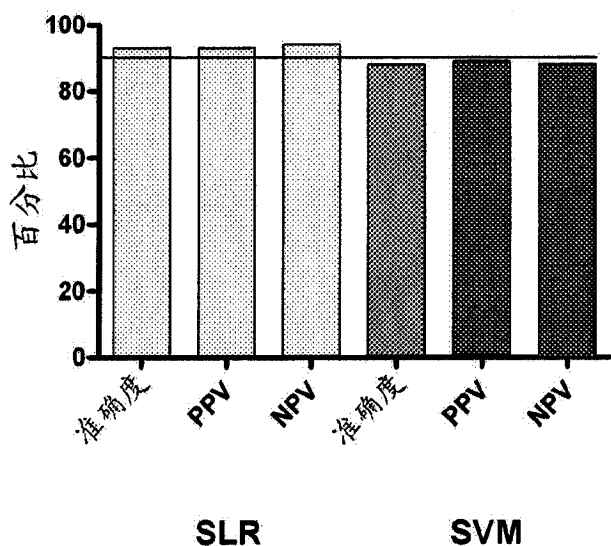


图 8A

SLR	RF
ARRB1	ARRB1
ARRB2	ARRB2
CD8a	CD8a
CREB1	CREB1
CREB2	CREB2
ERK2	ERK2
Gi2	Gi2
MAPK14	MAPK14
ODC1	ODC1
P2X7	P2X7
PBR	PBR
ADA	ERK1
CD8b	IL-1b
IDO	RGS2
MAPK8	
SERT	
S100A10	

图 8B

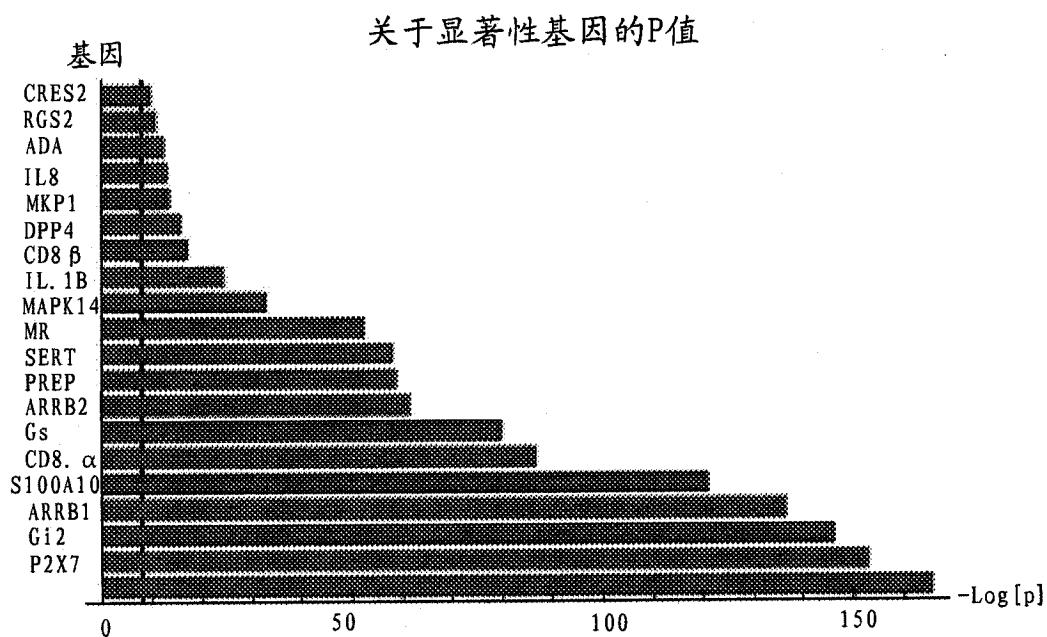


图 9

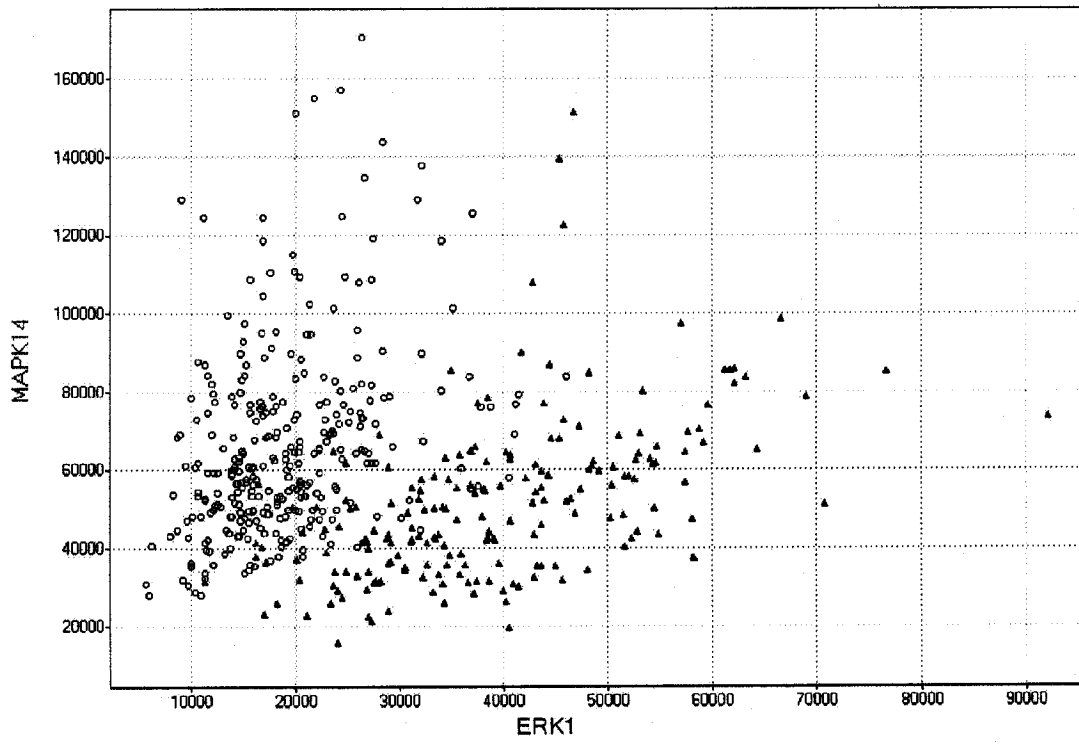


图 10

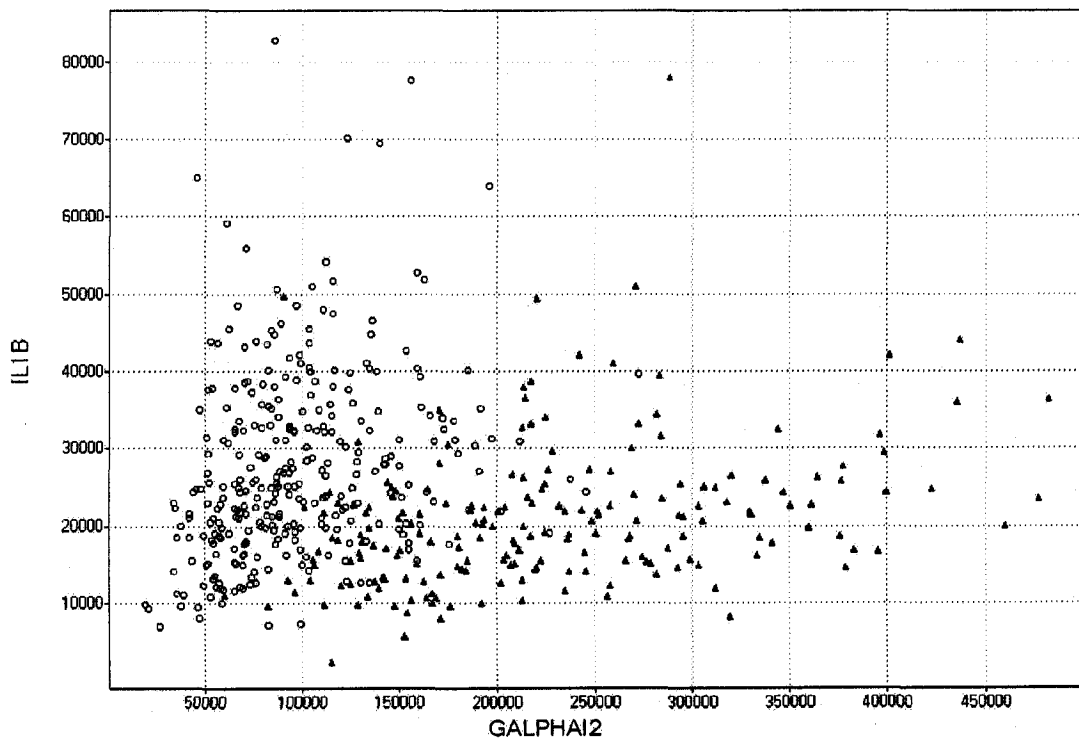


图 11

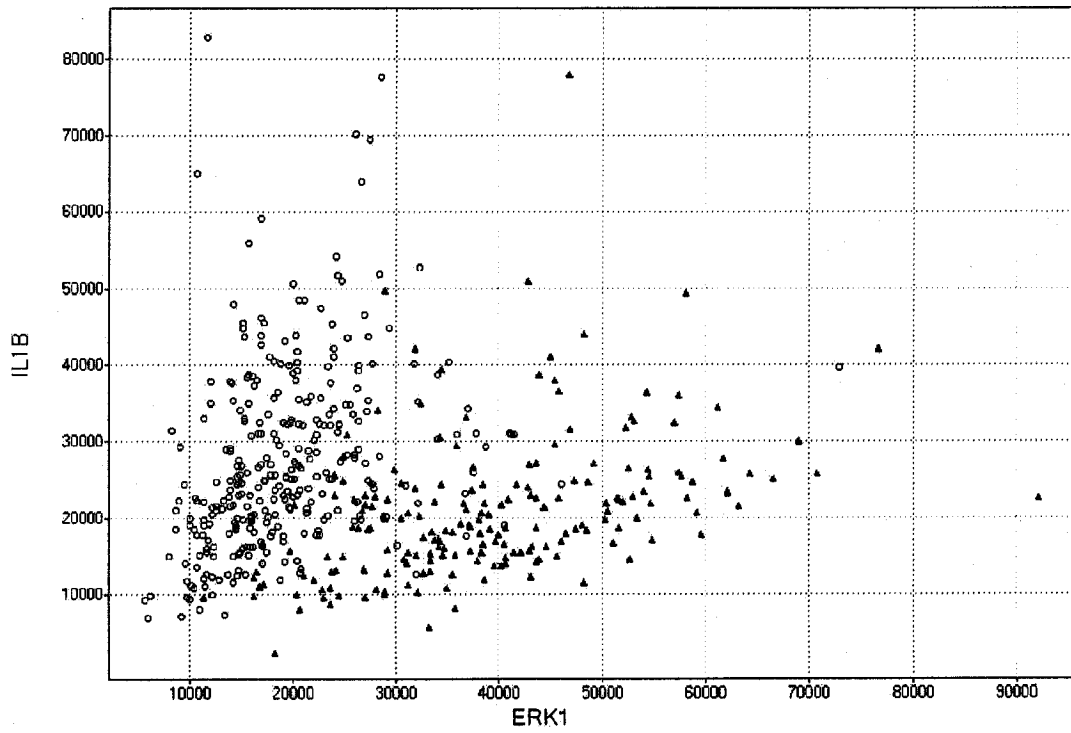


图 12

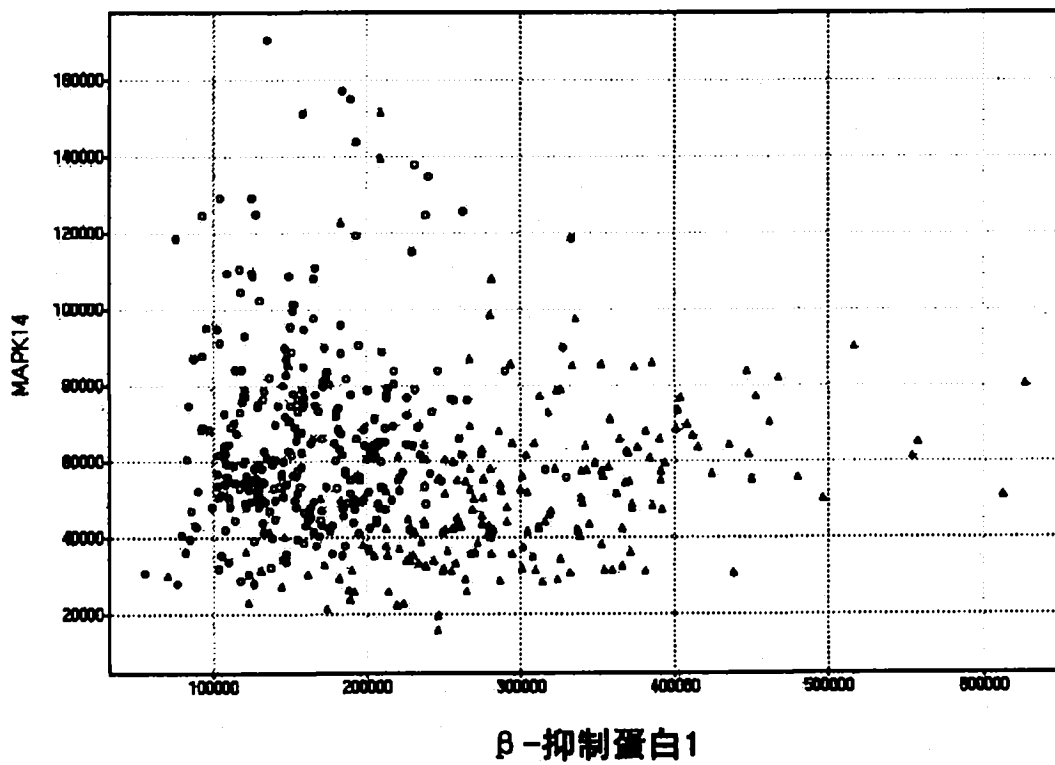


图 13

专利名称(译)	用于测量生物标记概况的系统和方法		
公开(公告)号	<a href="#">CN102224256A</a>	公开(公告)日	2011-10-19
申请号	CN200980142889.4	申请日	2009-08-27
[标]申请(专利权)人(译)	朗德贝克公司		
申请(专利权)人(译)	H.隆德贝克有限公司		
当前申请(专利权)人(译)	H.隆德贝克有限公司		
[标]发明人	I 安东尼耶维克 J 塔姆 R 阿泰米什恩 C P G 格拉尔 J B 维斯蒂森		
发明人	I·安东尼耶维克 J·塔姆 R·阿泰米什恩 C·P·G·格拉尔 J·B·维斯蒂森		
IPC分类号	C12Q1/68 G01N33/68 G01N33/53 G06F19/24		
CPC分类号	C12Q1/6883 C12Q2600/158 G16B40/00		
代理人(译)	李波 郭文洁		
优先权	61/092270 2008-08-27 US		
外部链接	<a href="#">Espacenet</a> <a href="#">SIPO</a>		

摘要(译)

本发明涉及用于诊断具有情感障碍的患者的方法和系统。该方法也用于预测受试者中对于情感障碍的易感性。

