



(12) 发明专利

(10) 授权公告号 CN 109563516 B

(45) 授权公告日 2023.10.27

(21) 申请号 201780048589.4

(51) Int.Cl.

(22) 申请日 2017.07.28

C12N 15/12 (2006.01)

(65) 同一申请的已公布的文献号

C12N 15/11 (2006.01)

申请公布号 CN 109563516 A

C07K 14/705 (2006.01)

(43) 申请公布日 2019.04.02

C12Q 1/6883 (2018.01)

(30) 优先权数据

62/367,973 2016.07.28 US

(56) 对比文件

CN 1829502 A, 2006.09.06

(85) PCT国际申请进入国家阶段日

CN 1946685 A, 2007.04.11

2019.02.01

HOWA05.G protein-coupled receptor

(86) PCT国际申请的申请数据

156.《UNIPROT》.2012, 第1页.

PCT/US2017/044321 2017.07.28

LX000688.LX000688.《EBI》.2015, 第1页.

(87) PCT国际申请的公布数据

AAS99192.AAS99192.《EBI》.2005, 第1页.

W02018/022967 EN 2018.02.01

Lisa A. Catapano 等.G Protein-Coupled

(73) 专利权人 雷杰纳荣制药公司

Receptors in Major Psychiatric Disorders.

地址 美国纽约

《Biochim Biophys Acta.》.2007, 第1768卷(第4期), 第976-993页.

审查员 隋亚平

(72) 发明人 克劳迪娅·贡萨加-贾雷吉

权利要求书2页 说明书45页

(74) 专利代理机构 深圳永慧知识产权代理事务

序列表(电子公布) 附图2页

所(普通合伙) 44378

专利代理人 黄鑫

(54) 发明名称

GPR156变体及其用途

(57) 摘要

CN 109563516 B

本公开提供包含在对应于人类G蛋白偶联受体156蛋白(GPR156)的位置533的位置处编码天冬氨酸的改变的核酸,包括cDNA。本公开还提供在对应于位置533的位置处包含天冬氨酸的分离的和重组的人类GPR156蛋白变体。改变为天冬氨酸和编码此改变的基因与单极抑郁症相关联。本公开还提供用于基于编码GPR156的基因(DNA或RNA)中的此类改变的鉴定,确定受试者是否具有单极抑郁症或具有发展单极抑郁症的风险的方法。

1. 一种特异性地与等位基因特异性引物或等位基因特异性探针杂交的核酸分子，其中所述核酸分子编码人类G蛋白偶联受体156 (GPR156) 蛋白，所述蛋白由根据SEQ ID NO:4或SEQ ID NO:5的氨基酸序列组成，其中所述等位基因特异性引物或等位基因特异性探针与核酸分子中的某些核苷酸杂交，所述核苷酸在对应于根据SEQ ID NO:4的位置533的位置处或在对应于根据SEQ ID NO:5的位置529的位置处编码天冬氨酸。

2. 根据权利要求1所述的特异性地与等位基因特异性引物或等位基因特异性探针杂交的核酸分子，其中所述核酸分子包括DNA。

3. 根据权利要求1所述的特异性地与等位基因特异性引物或等位基因特异性探针杂交的核酸分子，其中所述核酸分子是cDNA。

4. 根据权利要求1所述的特异性地与等位基因特异性引物或等位基因特异性探针杂交的核酸分子，其中所述核酸分子是mRNA。

5. 根据权利要求1至4中任一项所述的特异性地与等位基因特异性引物或等位基因特异性探针杂交的核酸分子，其中所述等位基因特异性引物或等位基因特异性探针的长度包含10至15个核苷酸。

6. 一种cDNA，其由SEQ ID NO:16的核酸序列组成，其中所述cDNA编码在对应于根据SEQ ID NO:4的位置533的位置处包含天冬氨酸的GPR156蛋白。

7. 一种cDNA，其由SEQ ID NO:17的核酸序列组成，其中所述cDNA编码在对应于根据SEQ ID NO:4的位置533的位置处包含天冬氨酸的GPR156蛋白。

8. 一种载体，其包含根据权利要求6或7所述的cDNA。

9. 根据权利要求8所述的载体，其中所述载体是质粒。

10. 根据权利要求8所述的载体，其中所述载体是病毒。

11. 一种宿主细胞，其包含根据权利要求6或7所述的cDNA。

12. 一种宿主细胞，其包含根据权利要求8-10中任一项所述的载体。

13. 根据权利要求11或权利要求12所述的宿主细胞，其中所述cDNA可操作地连接到在所述宿主细胞中具有活性的启动子。

14. 根据权利要求13所述的宿主细胞，其中所述启动子是外源启动子。

15. 根据权利要求13所述的宿主细胞，其中所述启动子是诱导型启动子。

16. 一种探针或引物，其包含至少20个核苷酸的与编码GPR156蛋白的核酸序列互补的核酸序列，所述GPR156蛋白由SEQ ID NO:4组成，其中所述探针或引物包含与在对应于根据SEQ ID NO:4的位置533的位置处编码天冬氨酸的核酸密码子互补的核酸序列。

17. 根据权利要求16所述的探针或引物，其中所述探针或引物包括DNA。

18. 根据权利要求16所述的探针或引物，其中所述探针或引物包括RNA。

19. 根据权利要求16至18中任一项所述的探针或引物，其中所述探针或引物包含标记。

20. 根据权利要求19所述的探针或引物，其中所述标记是荧光标记、放射性标记或生物素。

21. 一种支持物，其与根据权利要求16至20中任一项所述的探针或引物偶联。

22. 根据权利要求21所述的支持物，其中所述支持物是微阵列。

23. 一种分子复合物，其包含与等位基因特异性探针或等位基因特异性引物杂交的根据权利要求1-5中任一项的所述的核酸分子或权利要求6或7所述的cDNA，所述探针或引物

包含与编码GPR156蛋白的核酸序列互补的核酸序列，所述GPR156蛋白由SEQ ID NO:4组成，包含天冬氨酸，所述天冬氨酸的位置对应于SEQ ID NO:4的位置533。

GPR156变体及其用途

[0001] 相关申请的交叉引用

[0002] 本申请要求2016年7月28日提交的美国临时申请号62/367,973的优先权,所述申请的内容以引用的方式整体并出于所有目的并入本文。

[0003] 序列表的引用

[0004] 本申请包括作为文本文件电子提交的序列表,所述文本文件命名为18923800702SEQ,于2017年7月28日创建,大小为986千字节。序列表以引用的方式并入本文。

技术领域

[0005] 本公开总体上涉及遗传学的领域。更具体地,本公开涉及与例如单极抑郁症和焦虑病症相关联的G蛋白偶联受体156中的基因改变和多肽变体。

背景技术

[0006] 不同的参考文献(包括专利、专利申请、登录号、技术文献和学术文章)在整个说明书中引用。每个参考文献以引用的方式整体并出于所有目的并入本文。

[0007] G蛋白偶联受体(GPCR)是特征在于7个螺旋跨膜结构域连同N末端胞外结构域和C末端胞内结构域的细胞表面受体的大超家族(Cherezov V等人,Science 318:1258-1265 (2007))。GPCR在各种细胞类型中表达,并且参与跨细胞膜转导胞外信号并将胞外信号转导到细胞内部(Kobilka R,Biochim.Biophys.Acta 1768:794-807 (2007))。GPCR在生物学中发挥重要作用,使得在2012年,诺贝尔化学奖被授予鉴定GPCR如何作用的两位科学家(Clark R,Proc.Natl.Acad.Sci.USA 110:5274-5275 (2013))。

[0008] GPR156(G蛋白偶联受体156)是编码属于代谢型谷氨酸受体超家族的GPCR的人类基因(Calver A等人,Brain Res.Mol.Brain Res.110:305-307 (2003))。用于在文献中引用GPR156的其他名称是GABABL(GABAB样)和PGR28(Vassilatis D,Proc.Natl.Acad.Sci.USA 100:4903-4908 (2003))。在小鼠中,GPR156被称为Gpr156或Gabab1。鉴定新的GPR156变体将有助于继续研究GPR156在人类生理学中的作用以及其在疾病中的潜在作用。本公开提供新型GPR156变体,所述新型GPR156变体将帮助理解GPR156的生理学并且将有助于鉴定调控GPR156和/或其细胞生物通路的潜在治疗剂。

发明内容

[0009] 本公开提供编码GPR156变体多肽的新型核酸分子(即,基因组DNA、mRNA和cDNA)以及已在本文中证明与一系列的心境障碍诸如抑郁症(诸如单极抑郁症)和焦虑病症相关联的GPR156变体多肽。

[0010] 本公开提供包含编码人类GPR156蛋白的核酸序列或核酸序列的补体的分离的核酸分子,其中所述蛋白质在对应于根据SEQ ID NO:4的位置533的位置处包含天冬氨酸。

[0011] 本公开还提供包含编码人类GPR156蛋白的至少一部分的核酸序列或核酸序列的

补体的基因组DNA分子,其中所述部分在对应于根据SEQ ID NO:4的位置533的位置处包含天冬氨酸。

[0012] 本公开还提供包含编码人类GPR156蛋白的至少一部分的核酸序列或核酸序列的补体的cDNA分子,其中所述部分在对应于根据SEQ ID NO:4的位置533的位置处包含天冬氨酸。

[0013] 本公开还提供包含编码人类GPR156蛋白的至少一部分的核酸序列或核酸序列的补体的mRNA分子,其中所述部分在对应于根据SEQ ID NO:4的位置533的位置处包含天冬氨酸。

[0014] 本公开还提供包含本文公开的分离的核酸分子中的任一种的载体。

[0015] 本公开还提供包含本文公开的分离的核酸分子或载体中的任一种和携带体的组合物。

[0016] 本公开还提供包含本文公开的分离的核酸分子或载体中的任一种的宿主细胞。

[0017] 本公开还提供包含人类GPR156蛋白的至少一部分的分离的或重组的多肽,其中所述部分在对应于根据SEQ ID NO:4的位置533的位置处包含天冬氨酸。

[0018] 本公开还提供包含本文公开的分离的或重组的多肽中的任一种和携带体的组合物。

[0019] 本公开还提供一种包含具有至少约5个核苷酸的核酸序列的探针或引物,所述探针或引物与编码在对应于根据SEQ ID NO:4的位置533的位置处包含天冬氨酸的人类GPR156蛋白的核酸序列杂交,或者所述探针或引物与编码在对应于根据SEQ ID NO:4的位置533的位置处包含天冬氨酸的人类GPR156蛋白的核酸序列的补体杂交。

[0020] 本公开还提供包含本文公开的探针中的任一种所杂交的底物的支持物。

[0021] 本公开还提供一种包含与编码GPR156蛋白的核酸序列互补的核酸序列的等位基因特异性探针或引物,其中所述等位基因特异性探针或引物包含与在对应于根据SEQ ID NO:4的位置533的位置处编码天冬氨酸的核酸密码子互补的核酸序列。在一些实施方案中,所述等位基因特异性探针或引物与编码在对应于根据SEQ ID NO:4的位置533的位置处包含天冬氨酸的GPR156蛋白的核酸序列或与其补体特异性地杂交。所述等位基因特异性探针或引物不与编码在对应于根据SEQ ID NO:4的位置533的位置处不包含天冬氨酸的GPR156蛋白的核酸序列杂交。

[0022] 本公开还提供用于在人类受试者中诊断单极抑郁症或检测单极抑郁症的风险的方法,所述方法包括:检测从人类受试者获得的编码GPR156蛋白的核酸分子中的改变,其中所述改变在对应于根据SEQ ID NO:4的位置533的位置处编码天冬氨酸;以及如果所述受试者具有一种或多种抑郁症症状,则将人类受试者诊断为具有单极抑郁症,或者如果所述受试者不具有一种或多种抑郁症症状,则将人类受试者诊断为具有单极抑郁症的风险。

[0023] 本公开还提供用于在人类受试者中诊断焦虑病症或检测焦虑病症的风险的方法,所述方法包括:检测从人类受试者获得的编码GPR156蛋白的核酸分子中的改变,其中所述改变在对应于根据SEQ ID NO:4的位置533的位置处编码天冬氨酸;以及如果所述受试者具有一种或多种焦虑病症症状,则将人类受试者诊断为具有焦虑病症,或者如果所述受试者不具有一种或多种焦虑病症症状,则将人类受试者诊断为具有焦虑病症的风险。

[0024] 本公开还提供用于在确定包含编码人类GPR156蛋白的基因中的改变的人类受试

者中治疗单极抑郁症的抗抑郁药,其中所述改变在对应于根据SEQ ID NO:4的位置533的位置处编码天冬氨酸。

[0025] 本公开还提供用于在确定包含编码人类GPR156蛋白的基因中的改变的人类受试者中治疗焦虑病症的抗抑郁药,其中所述改变在对应于根据SEQ ID NO:4的位置533的位置处编码天冬氨酸。

附图说明

[0026] 并入本说明书且构成其一部分的附图示出若干个方面,并连同描述一起用于解释本公开的原理。

[0027] 图1示出mRor1ss.hGPR156表达载体的载体图。

[0028] 图2示出mRor1ss.hGPR156_E533D表达载体的载体图。

[0029] 本公开的另外的优点将部分陈述在以下描述中,并且部分将从所述描述中是明显的或者可通过实践本文公开的实施方案来得知。本公开的优点将借助所附权利要求书中具体指出的要素和组合来实现并达成。应了解上文一般性描述与下文详述两者均仅具有示例性和说明性且不限制如所要求的实施方案。

[0030] 说明书

[0031] 与本公开的各方面相关的各种术语在整个说明书和权利要求书中使用。除非另外指示,否则此类术语将被赋予其在本领域中的普通含义。其他特定定义的术语将以与本文提供的定义一致的方式来解释。

[0032] 除非另外明确说明,否则决不意图将本文陈述的任何方法或方面解释为要求以特定顺序执行其步骤。因此,在权利要求书或说明书中,当方法权利要求没有确切地说明步骤是限于特定顺序时,在任何方面决不意图推断顺序。这适用于任何可能的用于解释的非表达基础,包括相对于步骤安排或操作流程的逻辑事项、从语法组织或标点符号得到的清晰含义、或者在说明书中描述的方面的数量或类型。

[0033] 除非上下文另外明确指出,否则如本文所用,单数形式“一个(种) (a/an)”和“所述(the)”包括复数对象。

[0034] 如本文所用,术语“受试者”和“患者”可互换使用。受试者可包括任何动物,包括哺乳动物。哺乳动物包括但不限于农场动物(例如,马、牛、猪)、伴侣动物(例如,狗、猫)、实验室动物(例如,小鼠、大鼠、兔)和非人类灵长类动物。在一些实施方案中,受试者是人类。

[0035] 如本文所用,“核酸”、“核酸分子”、“核酸序列”“多核苷酸”或“寡核苷酸”可包括任何长度的核苷酸的聚合物形式,可包括DNA和/或RNA,并且可以是单链的、双链的或多链的。核酸的一条链还是指其补体。

[0036] 如本文所用,当在给定氨基酸或核酸序列或位置的编号的上下文中使用时,短语“对应于”或其语法变体是指当给定氨基酸或核酸序列与参考序列(例如,本文的参考序列是(野生型或全长)GPR156的核酸分子或多肽)相比时,指定参考序列的编号。换言之,给定聚合物的残基(例如,氨基酸或核苷酸)编号或残基(例如,氨基酸或核苷酸)位置相对于参考序列来指定,而非通过残基在给定氨基酸或核酸序列内的实际编号位置来指定。例如,给定氨基酸序列可通过引入缺口以优化两个序列之间的残基匹配来与参考序列对齐。在这些情况下,虽有存在缺口,但是对给定氨基酸或核酸序列中的残基的编号相对于它所对齐的

参考序列来进行。

[0037] 例如,短语“GPR156蛋白,其中所述蛋白质在对应于根据SEQ ID NO:4的位置533的位置处包含天冬氨酸”意指,如果GPR156蛋白的氨基酸序列与SEQ ID NO:4的序列对齐,则GPR156蛋白在对应于SEQ ID NO:4的位置533的位置处具有天冬氨酸。这同样适用于短语“GPR156蛋白,其中所述蛋白质在对应于参考SEQ ID NO:4的位置533的位置处包含天冬氨酸”以及“GPR156蛋白,其中所述蛋白质在对应于SEQ ID NO:4的位置533的位置处包含天冬氨酸”。或者,换言之,这些短语(例如,短语“GPR156蛋白,其中所述蛋白质在对应于根据SEQ ID NO:4的位置533的位置处包含天冬氨酸”)是指GPR156蛋白,其具有与SEQ ID NO:4的位置533处的天冬氨酸同源的天冬氨酸。在本文中,这种蛋白质还被称为“具有E533D突变的GPR156蛋白”或“具有E533D变化的GPR156蛋白”。与此相符,对应的E533D变化还被称为“E533D突变(在GPR156蛋白内)”或“E533D变化(在GPR156蛋白内)”。

[0038] 如上所述,在GPR156蛋白内的对应于SEQ ID NO:4的位置533的位置可易于通过执行给定的GPR156蛋白与SEQ ID NO:4的氨基酸序列之间的序列比对来鉴定。存在可用于执行序列比对以便鉴定对应于SEQ ID NO:4中的位置533的氨基酸位置的各种计算算法。例如,通过使用NCBI BLAST算法(Altschul等人1997Nucleic Acids Res.25:3389-3402)或CLUSTALW软件(Sievers和Higgins 2014Methods Mol.Biol.1079:105-116.),可执行序列比对。然而,序列还可手动地进行比对。

[0039] 根据本公开,已观察到,GPR156中的某些变化与发展心境障碍(诸如单极抑郁症和焦虑病症)的风险相关联。总体上,对此蛋白质的功能还不太了解,关于其生物化学或其相关的表型的公开信息很少。据信,GPR156基因或蛋白的变体与成年人类中的任何心境障碍或神经病不具有任何已知的联系。还据信,GPR156基因或蛋白的变体与特定地在成年人类中的一般抑郁症或单极抑郁症或焦虑病症不具有任何已知的联系。受影响家族成员中的以单极抑郁症的表型分离的GPR156基因中的罕见变体已根据本公开鉴定。例如,已观察到将人类GPR156蛋白(例如,野生型SEQ ID NO:1)中的位置533的氨基酸改变为天冬氨酸的遗传改变指示具有这种改变的人类可发展单极抑郁症。总之,本文所述的遗传分析令人意外地指示GPR156基因和具体地GPR156中的变体与对单极抑郁症和焦虑病症的增加的易感性相关联。因此,具有与单极抑郁症或焦虑病症相关联的GPR156改变的人类受试者可进行治疗,使得单极抑郁症被抑制,其症状减少并且/或者症状的发展被遏制。因此,本公开提供分离的或重组的GPR156变体基因(包括cDNA和mRNA)以及分离的或重组的GPR156变体多肽。另外,本公开提供用于利用受试者中的此类变体的鉴定以对此类受试者的发展单极抑郁症或焦虑病症的风险进行鉴定或分级或将受试者诊断为具有单极抑郁症或焦虑病症、使得具有风险的受试者或具有活动性疾病的受试者可进行治疗的方法。

[0040] 两种全长野生型GPR156蛋白的氨基酸序列在SEQ ID NO:1中列出。参考SEQ ID NO:1,全长野生型GPR156蛋白的位置516是谷氨酸或天冬氨酸。另一种野生型GPR156蛋白的氨基酸序列在SEQ ID NO:2(其在对应于SEQ ID NO:1的位置198至201的位置处含有4个氨基酸的缺失)中列出。此野生型GPR156蛋白的位置512(参考SEQ ID NO:2)是谷氨酸或天冬氨酸。在一些实施方案中,此野生型GPR156蛋白的位置512(参考SEQ ID NO:2)是谷氨酸。在一些实施方案中,此野生型GPR156蛋白的位置512(参考SEQ ID NO:2)是天冬氨酸。包含SEQ ID NO:1的位置310至814的GPR156蛋白的氨基酸序列在SEQ ID NO:3中列出。SEQ ID NO:1

的位置310至814限定GPR156的细胞质结构域。此GPR156蛋白的位置207(参考SEQ ID NO:3)是谷氨酸或天冬氨酸。在一些实施方案中,此GPR156蛋白的位置207(参考SEQ ID NO:3)是谷氨酸。在一些实施方案中,此GPR156蛋白的位置207(参考SEQ ID NO:3)是天冬氨酸。

[0041] 本公开提供编码与单极抑郁症或焦虑病症相关联的GPR156变体蛋白的核酸分子。例如,本公开提供包含编码人类GPR156蛋白的核酸序列或核酸序列的补体的分离的核酸分子,其中所述蛋白质在对应于根据SEQ ID NO:4的位置533(E533D)的位置处包含天冬氨酸。在一些实施方案中,对应于SEQ ID NO:4的位置516的位置是谷氨酸。在一些实施方案中,对应于SEQ ID NO:4的位置516的位置是天冬氨酸。

[0042] 存在GPR156的同种型,其中蛋白质的N末端至中心区中的四个氨基酸缺失。此GPR156同种型的氨基酸序列在SEQ ID NO:5中示出。在此GPR156同种型中,位置529对应于根据SEQ ID NO:4的位置533。在一些实施方案中,分离的核酸分子包含编码人类GPR156蛋白的核酸序列或核酸序列的补体,其中所述蛋白质在对应于根据SEQ ID NO:5的位置529(E529D)的位置处包含天冬氨酸。在一些实施方案中,SEQ ID NO:5的位置512是谷氨酸。在一些实施方案中,对应于SEQ ID NO:5的位置512的位置是天冬氨酸。

[0043] 在一些实施方案中,所述核酸分子包含编码具有与SEQ ID NO:4具有至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的氨基酸序列的人类GPR156蛋白的核酸序列或核酸序列的补体或者由其组成,其中GPR156蛋白在对应于根据SEQ ID NO:4的位置533的位置处包含天冬氨酸。在一些实施方案中,此核酸分子的对应于位置516(参考SEQ ID NO:4)的位置是谷氨酸。在一些实施方案中,此核酸分子的对应于位置516(参考SEQ ID NO:4)的位置是天冬氨酸。在本文中,如果对序列同一性百分比进行参考,则较高的序列同一性百分比相对于较低的序列同一性百分比是优选的。

[0044] 在一些实施方案中,所述核酸分子包含编码具有与SEQ ID NO:5具有至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的氨基酸序列的人类GPR156蛋白的核酸序列或核酸序列的补体或者由其组成,其中GPR156蛋白在对应于根据SEQ ID NO:5的位置529的位置处包含天冬氨酸。在一些实施方案中,此GPR156蛋白的对应于位置512(参考SEQ ID NO:5)的位置是谷氨酸。在一些实施方案中,此GPR156蛋白的对应于位置512(参考SEQ ID NO:5)的位置是天冬氨酸。

[0045] 在一些实施方案中,所述核酸序列包含DNA,并且在对应于位置533的位置处的天冬氨酸由密码子GAT或GAC编码。在一些实施方案中,所述核酸序列包含RNA,并且在对应于位置533的位置处的天冬氨酸由RNA分子中的密码子GAU或GAC编码。在一些实施方案中,在对应于位置533的位置处的天冬氨酸由密码子GAT编码。在一些实施方案中,由所述核酸序列编码的GPR156蛋白包含SEQ ID NO:4或SEQ ID NO:5的氨基酸序列。

[0046] 两种野生型GPR156基因组DNA(即,另选的等位基因)的核酸序列在SEQ ID NO:7中列出(参见,在位置117,166处的“n”,其可以是a、g、t或c)。这些野生型GPR156基因组DNA的位置117,164至117,166(参考SEQ ID NO:7)编码谷氨酸(通过密码子GAA或GAG)或天冬氨酸(通过密码子GAT或GAC)。这些野生型GPR156基因组DNA的位置117,215至117,217(参考SEQ ID NO:7)编码谷氨酸(通过密码子GAA或GAG)。

[0047] 另一种野生型GPR156基因组DNA的核酸序列在SEQ ID NO:8(其在对应于SEQ ID NO:7的位置98,428至98,439的位置处含有12个核苷酸的缺失)中列出。此野生型GPR156基因组DNA的位置117,152至117,154(参考SEQ ID NO:8)编码谷氨酸(通过密码子GAA或GAG)或天冬氨酸(通过密码子GAT或GAC)。此野生型GPR156基因组DNA的位置117,203至117,205(参考SEQ ID NO:8)编码谷氨酸(通过密码子GAA或GAG)。

[0048] 在一些实施方案中,所述核酸分子是基因组DNA。在一些实施方案中,所述基因组DNA包含SEQ ID NO:9的核酸序列(其中位置117,215至117,217是GAT,从而编码天冬氨酸)或SEQ ID NO:10的核酸序列(其中位置117,215至117,217是GAC,从而编码天冬氨酸)或与SEQ ID NO:9具有至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性(并且在对应于在根据SEQ ID NO:4的位置533处编码天冬氨酸的位置的位置处包含GAT密码子)的核酸序列或与SEQ ID NO:10具有至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性(并且在对应于在根据SEQ ID NO:4的位置533处编码天冬氨酸的位置的位置处包含GAC密码子)的核酸序列或者由其组成。在一些实施方案中,这些GPR156基因组DNA的位置117,164至117,166(参考SEQ ID NO:9和SEQ ID NO:10)编码谷氨酸(通过密码子GAA或GAG)。在一些实施方案中,这些GPR156基因组DNA的对应于位置117,164至117,166(参考SEQ ID NO:9和SEQ ID NO:10)的位置编码天冬氨酸(通过密码子GAT或GAC)。

[0049] 在一些实施方案中,所述基因组DNA包含SEQ ID NO:11或SEQ ID NO:12的核酸序列或者由其组成,其中的每个在分别对应于SEQ ID NO:9和SEQ ID NO:10的位置98,428至98,439的位置处含有12个核苷酸的缺失。在一些实施方案中,所述基因组DNA包含在SEQ ID NO:11中列出的核酸序列或者由其组成,其中位置117,203至117,205是GAT,从而编码天冬氨酸。在一些实施方案中,所述基因组DNA包含在SEQ ID NO:12中列出的核酸序列或者由其组成,其中位置117,203至117,205是GAC,从而编码天冬氨酸。在一些实施方案中,所述基因组DNA包含与SEQ ID NO:11具有至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性(并且在对应于在根据SEQ ID NO:5的位置529处编码天冬氨酸的位置的位置处包含GAT密码子)的核酸序列或与SEQ ID NO:12具有至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性(并且在对应于在根据SEQ ID NO:5的位置529处编码天冬氨酸的位置的位置处包含GAC密码子)的核酸序列或者由其组成。在一些实施方案中,这些GPR156基因组DNA的对应于位置117,152至117,154(参考SEQ ID NO:11和SEQ ID NO:12)的位置编码谷氨酸(通过密码子GAA或GAG)。在一些实施方案中,这些GPR156基因组DNA的对应于位置117,152至117,154(参考SEQ ID NO:11和SEQ ID NO:12)的位置编码天冬氨酸(通过密码子GAT或GAC)。

[0050] 在一些实施方案中,所述分离的核酸分子包含少于完整基因组DNA序列。在一些实施方案中,所述分离的核酸分子包含SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11或SEQ ID NO:12的至少约15个、至少约20个、至少约25个、至少约30个、至少约35个、至少约40个、至少约45个、至少约50个、至少约60个、至少约70个、至少约80个、至少约90个、至少约100个、至少约200个、至少约300个、至少约400个、至少约500个、至少约600个、至少约700个、至少约

800个、至少约900个、至少约1000个、至少约2000个、至少约3000个、至少约4000个、至少约5000个、至少约6000个、至少约7000个、至少约8000个、至少约9000个、至少约10000个、至少约11000个、至少约12000个、至少约13000个、至少约14000个、至少约15000个、至少约16000个、至少约17000个、至少约18000个、至少约19000个或至少约20000个连续核苷酸或者由其组成。在一些实施方案中，所述分离的核酸分子包含SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11或SEQ ID NO:12的至少约1000个到至少约2000个连续核苷酸或者由其组成。

[0051] 在一些实施方案中，所述分离的核酸分子包含SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11或SEQ ID NO:12的至少约15个、至少约20个、至少约25个、至少约30个、至少约35个、至少约40个、至少约45个、至少约50个、至少约60个、至少约70个、至少约80个、至少约90个、至少约100个、至少约200个、至少约300个、至少约400个、至少约500个、至少约600个、至少约700个、至少约800个、至少约900个、至少约1000个、至少约1000个、至少约1100个、至少约1200个、至少约1300个、至少约1400个、至少约1500个、至少约1600个、至少约1700个、至少约1800个、至少约1900个、至少约2000个、至少约2100个、至少约2200个、至少约2300个或至少约2400个连续核苷酸或者由其组成。在一些实施方案中，此类连续的核苷酸可与连续核苷酸的其他核酸分子组合以产生本文所述的cDNA分子。

[0052] 此类分离的核酸分子可用于例如表达变体GPR156 mRNA和蛋白或者作为外源供体序列。应理解，群体内的基因序列可由于多态性（诸如SNP）而变化。本文提供的实例仅是示例性序列，并且其他序列也是可能的。

[0053] 在一些实施方案中，所述分离的核酸分子包含变体GPR156小基因，其中编码SEQ ID NO:4或SEQ ID NO:5的一个或多个非必需片段已相对于对应的野生型GPR156基因缺失。在一些实施方案中，缺失的非必需片段包含一个或多个内含子序列。在一些实施方案中，GPR156小基因与SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11或SEQ ID NO:12的一部分具有至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或100%序列同一性，其中所述小基因在对应于根据SEQ ID NO:4的位置533的位置处包含天冬氨酸。

[0054] 两种野生型GPR156 cDNA的核酸序列在SEQ ID NO:13中列出。这些野生型GPR156 cDNA的位置1546至1548（参考SEQ ID NO:13）编码谷氨酸（通过密码子GAA或GAG）或天冬氨酸（通过密码子GAT或GAC）。这些野生型GPR156 cDNA的位置1597至1599（参考SEQ ID NO:13）编码谷氨酸（通过密码子GAA或GAG）。

[0055] 另一种野生型GPR156 cDNA的核酸序列在SEQ ID NO:14（其在对应于SEQ ID NO:13的位置592-603的位置处含有12个核苷酸的缺失）中列出。此野生型GPR156 cDNA的位置1534至1536（参考SEQ ID NO:14）编码谷氨酸（通过密码子GAA或GAG）或天冬氨酸（通过密码子GAT或GAC）。此野生型GPR156 cDNA的位置1585至1587（参考SEQ ID NO:14）编码谷氨酸（通过密码子GAA或GAG）。

[0056] 编码SEQ ID NO:4的位置310至814的GPR156 cDNA的核酸序列在SEQ ID NO:15中列出。此序列对应于GPR156的细胞质结构域。此GPR156 cDNA的位置619至621（参考SEQ ID NO:15）编码谷氨酸（通过密码子GAA或GAG）或天冬氨酸（通过密码子GAT或GAC）。此GPR156 cDNA的位置670至672（参考SEQ ID NO:15）编码谷氨酸（通过密码子GAA或GAG）。

[0057] 本公开还提供cDNA分子。本公开提供包含编码人类GPR156蛋白的至少一部分的核

酸序列的cDNA分子，其中所述部分在对应于根据SEQ ID NO:4的位置533的位置处包含天冬氨酸。在一些实施方案中，所述cDNA分子包含编码人类GPR156蛋白的至少一部分的核酸序列，其中所述部分在对应于根据SEQ ID NO:5的位置529的位置处包含天冬氨酸。

[0058] 在一些实施方案中，cDNA编码在对应于根据SEQ ID NO:4的位置533的位置处包含天冬氨酸的全长人类GPR156蛋白。在一些实施方案中，所述cDNA包含编码具有与SEQ ID NO:4具有至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的氨基酸序列的多肽的核酸序列或者由其组成，其中GPR156蛋白在对应于根据SEQ ID NO:4的位置533的位置处包含天冬氨酸。

[0059] 在一些实施方案中，所述cDNA编码在对应于根据SEQ ID NO:5的位置529的位置处包含天冬氨酸的人类GPR156蛋白。在一些实施方案中，所述cDNA包含编码具有与SEQ ID NO:5具有至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的氨基酸序列的多肽的核酸序列或者由其组成，其中GPR156蛋白在对应于根据SEQ ID NO:5的位置529的位置处包含天冬氨酸。

[0060] 在一些实施方案中，所述cDNA编码对应于SEQ ID NO:4的位置310至814的GPR156蛋白的部分。在一些实施方案中，所述cDNA包含编码具有与对应于SEQ ID NO:4的位置310至814的GPR156蛋白的部分具有至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的氨基酸序列的多肽的核酸序列或者由其组成，其中所述多肽在对应于根据SEQ ID NO:4的位置533的位置处包含天冬氨酸。

[0061] 在一些实施方案中，所述cDNA编码对应于SEQ ID NO:5的位置310至814的GPR156蛋白的部分。在一些实施方案中，所述cDNA包含编码具有与对应于SEQ ID NO:5的位置310至814的GPR156蛋白的部分具有至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的氨基酸序列的多肽的核酸序列或者由其组成，其中所述多肽在对应于根据SEQ ID NO:5的位置533的位置处包含天冬氨酸。

[0062] 在一些实施方案中，所述cDNA包含SEQ ID NO:16的核酸序列或者由其组成，其中位置1597至1599是GAT，从而编码天冬氨酸。在一些实施方案中，所述cDNA包含与SEQ ID NO:16具有至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的核酸序列或者由其组成，并且在对应于在根据SEQ ID NO:4的位置533处编码天冬氨酸的位置的位置处包含GAT密码子)。在一些实施方案中，GPR156 cDNA的对应于位置1546至1548(参考SEQ ID NO:16)的位置编码谷氨酸。在一些实施方案中，GPR156 cDNA的对应于位置1546至1548(参考SEQ ID NO:16)的位置编码天冬氨酸。在一些实施方案中，所述cDNA在对应于GPR156 cDNA的位置1599(参考SEQ ID NO:16)的位置处包含胸腺嘧啶。此胸腺嘧啶对应于在对应于根据SEQ ID NO:4的位置533的位置处编码天冬氨酸的密码子的第三位置)。

[0063] 在一些实施方案中，所述cDNA包含SEQ ID NO:17的核酸序列或者由其组成，其中位置1597至1599是GAC，从而编码天冬氨酸。在一些实施方案中，所述cDNA包含与SEQ ID

NO:17具有至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的核酸序列或者由其组成，并且在对应于在根据SEQ ID NO:4的位置533处编码天冬氨酸的位置处包含GAC密码子)。在一些实施方案中，GPR156 cDNA的对应于位置1546至1548(参考SEQ ID NO:17)的位置编码谷氨酸。在一些实施方案中，GPR156 cDNA的位置1546至1548(参考SEQ ID NO:17)编码天冬氨酸。在一些实施方案中，所述cDNA在对应于GPR156 cDNA的位置1599(参考SEQ ID NO:17)的位置处包含胞嘧啶。此胞嘧啶对应于在对应于根据SEQ ID NO:4的位置533的位置处编码天冬氨酸的密码子的第三位置)。

[0064] 在一些实施方案中，所述cDNA包含SEQ ID NO:18的核酸序列或者由其组成，其在对应于SEQ ID NO:16的位置592至603的位置处含有12个核苷酸的缺失。在一些实施方案中，所述cDNA包含在SEQ ID NO:18中列出的核酸序列或者由其组成，其中位置1585至1587是GAT，从而编码天冬氨酸。在一些实施方案中，所述cDNA包含与SEQ ID NO:18具有至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的核酸序列或者由其组成(并且在对应于在对应于根据SEQ ID NO:5的位置529的位置处编码天冬氨酸的位置处包含GAT密码子)。在一些实施方案中，此GPR156 cDNA的对应于位置1534至1536(参考SEQ ID NO:18)的位置编码谷氨酸(通过密码子GAA或GAG)。在一些实施方案中，此GPR156 cDNA的对应于位置1534至1536(参考SEQ ID NO:18)的位置编码天冬氨酸(通过密码子GAT或GAC)。在一些实施方案中，所述cDNA在对应于GPR156 cDNA的位置1587(参考SEQ ID NO:18)的位置处包含胸腺嘧啶。此胸腺嘧啶对应于在对应于根据SEQ ID NO:5的位置529的位置处编码天冬氨酸的密码子的第三位置)。

[0065] 在一些实施方案中，所述cDNA包含SEQ ID NO:19的核酸序列或者由其组成，其在对应于SEQ ID NO:17的位置592至603的位置处含有12个核苷酸的缺失。在一些实施方案中，所述cDNA包含在SEQ ID NO:19中列出的核酸序列或者由其组成，其中位置1585至1587是GAC，从而编码天冬氨酸。在一些实施方案中，所述cDNA包含与SEQ ID NO:19具有至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的核酸序列或者由其组成(并且在对应于在根据SEQ ID NO:5的位置529处编码天冬氨酸的位置处包含GAC密码子)。在一些实施方案中，此GPR156 cDNA的对应于位置1534至1536(参考SEQ ID NO:19)的位置编码谷氨酸(通过密码子GAA或GAG)。在一些实施方案中，此GPR156 cDNA的对应于位置1534至1536(参考SEQ ID NO:19)的位置编码天冬氨酸(通过密码子GAT或GAC)。在一些实施方案中，所述cDNA在对应于cDNA的位置1587(参考SEQ ID NO:19)的位置处包含胞嘧啶。此胞嘧啶对应于在对应于根据SEQ ID NO:5的位置529的位置处编码天冬氨酸的密码子的第三位置)。

[0066] GPR156的细胞质结构域由GPR156的位置310至814编码。在一些实施方案中，所述cDNA编码SEQ ID NO:4的位置310至814。在一些实施方案中，此cDNA的核酸序列在SEQ ID NO:20中列出，其中位置670至672编码天冬氨酸(通过密码子GAT)。在一些实施方案中，所述cDNA包含与SEQ ID NO:20具有至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的核酸序列或者由其组成，并且在对应于在对应于根据SEQ ID NO:4的位置533的位置处编码天冬

氨酸的位置处包含GAT密码子。在一些实施方案中,GPR156 cDNA的对应于位置619至621(参考SEQ ID NO:20)的位置编码谷氨酸。在一些实施方案中,GPR156 cDNA的对应于位置619至621(参考SEQ ID NO:20)的位置编码天冬氨酸。在一些实施方案中,所述cDNA在对应于cDNA的位置672(参考SEQ ID NO:20)的位置处包含胸腺嘧啶。此胸腺嘧啶对于在对应于根据SEQ ID NO:4的位置533的位置处编码天冬氨酸的密码子的第三位置)。

[0067] 在一些实施方案中,编码SEQ ID NO:4的位置310至814的cDNA的核酸序列在SEQ ID NO:21中列出,其中位置670至672编码天冬氨酸(通过密码子GAC)。在一些实施方案中,所述cDNA包含与SEQ ID NO:21具有至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的核酸序列或者由其组成,并且在对应于在对应于根据SEQ ID NO:4的位置533的位置处编码天冬氨酸的位置处包含GAC密码子。在一些实施方案中,GPR156 cDNA的对应于位置619至621(参考SEQ ID NO:21)的位置编码谷氨酸。在一些实施方案中,GPR156 cDNA的对应于位置619至621(参考SEQ ID NO:21)的位置编码天冬氨酸。在一些实施方案中,所述cDNA在对应于cDNA的位置672(参考SEQ ID NO:21)的位置处包含胞嘧啶。此胞嘧啶对于在对应于根据SEQ ID NO:4的位置533的位置处编码天冬氨酸的密码子的第三位置)。

[0068] 在一些实施方案中,所述分离的核酸分子包含具有本文公开的cDNA或mRNA分子中的任一种的位置1521至1680的核酸序列,从而在对应于位置1599的位置处具有胸腺嘧啶。

[0069] 在一些实施方案中,所述cDNA分子包含少于GPR156的完整cDNA序列。在一些实施方案中,所述cDNA分子包含SEQ ID NO:16、SEQ ID NO:17、SEQ ID NO:18、SEQ ID NO:19、SEQ ID NO:20或SEQ ID NO:21的至少约5个、至少约8个、至少约10个、至少约12个、至少约15个、至少约20个、至少约25个、至少约30个、至少约35个、至少约40个、至少约45个、至少约50个、至少约60个、至少约70个、至少约80个、至少约90个、至少约100个、至少约200个、至少约300个、至少约400个、至少约500个、至少约600个、至少约700个、至少约800个、至少约900个、至少约1000个、至少约1100个、至少约1200个、至少约1300个、至少约1400个、至少约1500个、至少约1600个、至少约1700个、至少约1800个、至少约1900个、至少约2000个、至少约2100个、至少约2200个、至少约2300个或至少约2400个连续核苷酸或者由其组成。在一些实施方案中,所述cDNA分子包含SEQ ID NO:16、SEQ ID NO:17、SEQ ID NO:18、SEQ ID NO:19、SEQ ID NO:20或SEQ ID NO:21的至少约1000个到至少约2000个连续核苷酸或者由其组成。就这一点而言,较长的cDNA分子相对于较短的cDNA分子是优选的。在一些实施方案中,所述cDNA分子包含SEQ ID NO:16、SEQ ID NO:17、SEQ ID NO:18、SEQ ID NO:19、SEQ ID NO:20或SEQ ID NO:21的至少约15个、至少约20个、至少约25个、至少约30个、至少约35个、至少约40个、至少约45个、至少约50个、至少约60个、至少约70个、至少约80个、至少约90个、至少约100个、至少约200个、至少约300个、至少约400个或至少约500个连续核苷酸或者由其组成。就这一点而言,较长的cDNA分子相对于较短的cDNA分子是优选的。在一些实施方案中,此类cDNA分子包含在对应于根据SEQ ID NO:4的位置533的位置处编码天冬氨酸的密码子或在分别对应于根据SEQ ID NO:16或SEQ ID NO:17的位置1599的位置处包含胸腺嘧啶或半胱氨酸。

[0070] 两种野生型GPR156 mRNA的核酸序列在SEQ ID NO:22中列出。这些野生型GPR156 mRNA的位置1546至1548(参考SEQ ID NO:22)编码谷氨酸(通过密码子GAA或GAG)或天冬氨

酸(通过密码子GAT或GAC)。这些野生型GPR156 mRNA的位置1597至1599(参考SEQ ID NO:22)编码谷氨酸(通过密码子GAA或GAG)。

[0071] 另一种野生型GPR156 mRNA的核酸序列在SEQ ID NO:23(其在对应于SEQ ID NO:22的位置592至603的位置处含有12个核苷酸的缺失)中列出。此野生型GPR156 mRNA的位置1534至1536(参考SEQ ID NO:23)编码谷氨酸(通过密码子GAA或GAG)或天冬氨酸(通过密码子GAU或GAC)。此野生型GPR156 mRNA的位置1585至1587(参考SEQ ID NO:23)编码谷氨酸(通过密码子GAA或GAG)。

[0072] 编码SEQ ID NO:4的位置310至814的GPR156 mRNA的核酸序列在SEQ ID NO:24中列出。此GPR156 mRNA的位置619至621(参考SEQ ID NO:24)编码谷氨酸(通过密码子GAA或GAG)或天冬氨酸(通过密码子GAU或GAC)。此GPR156 mRNA的位置670至672(参考SEQ ID NO:24)编码谷氨酸(通过密码子GAA或GAG)。

[0073] 本公开还提供mRNA分子。本公开提供包含编码人类GPR156蛋白的至少一部分的核酸序列的mRNA分子,其中所述部分在对应于根据SEQ ID NO:4的位置533的位置处包含天冬氨酸。在一些实施方案中,所述mRNA分子包含编码人类GPR156蛋白的至少一部分的核酸序列,其中所述部分在对应于根据SEQ ID NO:5的位置529的位置处包含天冬氨酸。

[0074] 在一些实施方案中,mRNA编码在对应于根据SEQ ID NO:4的位置533的位置处包含天冬氨酸的全长人类GPR156蛋白。在一些实施方案中,所述mRNA包含编码具有与SEQ ID NO:4具有至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的氨基酸序列的多肽的核酸序列或者由其组成,其中GPR156蛋白在对应于根据SEQ ID NO:4的位置533的位置处包含天冬氨酸。

[0075] 在一些实施方案中,所述mRNA编码在对应于根据SEQ ID NO:5的位置529的位置处包含天冬氨酸的人类GPR156蛋白。在一些实施方案中,所述mRNA包含编码具有与SEQ ID NO:5具有至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的氨基酸序列的多肽的核酸序列或者由其组成,其中GPR156蛋白在对应于根据SEQ ID NO:5的位置529的位置处包含天冬氨酸。

[0076] 在一些实施方案中,所述mRNA编码对应于SEQ ID NO:4的位置310至814的GPR156蛋白的部分。在一些实施方案中,所述mRNA包含编码具有与对应于SEQ ID NO:4的位置310至814的GPR156蛋白的部分具有至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的氨基酸序列的多肽的核酸序列或者由其组成,其中所述多肽在对应于根据SEQ ID NO:4的位置533的位置处包含天冬氨酸。

[0077] 在一些实施方案中,所述mRNA编码对应于SEQ ID NO:5的位置310至814的GPR156蛋白的部分。在一些实施方案中,所述mRNA包含编码具有与对应于SEQ ID NO:5的位置310至814的GPR156蛋白的部分具有至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的氨基酸序列的多肽的核酸序列或者由其组成,其中所述多肽在对应于根据SEQ ID NO:5的位置533的位置处包含天冬氨酸。

[0078] 在一些实施方案中,所述mRNA包含SEQ ID NO:25的核酸序列或者由其组成,其中位置1597至1599是GAU,从而编码天冬氨酸。在一些实施方案中,所述mRNA包含与SEQ ID NO:25具有至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的核酸序列或者由其组成,并且在对应于根据SEQ ID NO:4的位置533的位置处编码天冬氨酸的位置处包含GAU密码子)。在一些实施方案中,GPR156 mRNA的对应于位置1546至1548(参考SEQ ID NO:25)的位置编码谷氨酸。在一些实施方案中,GPR156 mRNA的对应于位置1546至1548(参考SEQ ID NO:25)的位置编码天冬氨酸。在一些实施方案中,所述mRNA在对应于GPR156 mRNA的位置1599(参考SEQ ID NO:25)的位置处包含尿嘧啶。此尿嘧啶对于在对应于根据SEQ ID NO:4的位置533的位置处编码天冬氨酸的密码子的第三位置)。

[0079] 在一些实施方案中,所述mRNA包含SEQ ID NO:26的核酸序列或者由其组成,其中位置1597至1599是GAC,从而编码天冬氨酸。在一些实施方案中,所述mRNA包含与SEQ ID NO:26具有至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的核酸序列或者由其组成,并且在对应于根据SEQ ID NO:4的位置533的位置处编码天冬氨酸的位置处包含GAC密码子)。在一些实施方案中,GPR156 mRNA的对应于位置1546至1548(参考SEQ ID NO:26)的位置编码谷氨酸。在一些实施方案中,GPR156 mRNA的对应于位置1546至1548(参考SEQ ID NO:26)的位置编码天冬氨酸。在一些实施方案中,所述mRNA在对应于GPR156 mRNA的位置1599(参考SEQ ID NO:26)的位置处包含胞嘧啶。此胞嘧啶对于在对应于根据SEQ ID NO:4的位置533的位置处编码天冬氨酸的密码子的第三位置)。

[0080] 在一些实施方案中,所述mRNA包含SEQ ID NO:27的核酸序列或者由其组成,其在对应于SEQ ID NO:25的位置592至603的位置处含有12个核苷酸的缺失。在一些实施方案中,所述mRNA包含在SEQ ID NO:27中列出的核酸序列或者由其组成,其中位置1585至1587是GAU,从而编码天冬氨酸。在一些实施方案中,所述mRNA包含与SEQ ID NO:27具有至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的核酸序列或者由其组成(并且在对应于根据SEQ ID NO:5的位置529的位置处编码天冬氨酸的位置处包含GAU密码子)。在一些实施方案中,此GPR156 mRNA的对应于位置1534至1536(参考SEQ ID NO:27)的位置编码谷氨酸(通过密码子GAA或GAG)。在一些实施方案中,此GPR156 mRNA的对应于位置1534至1536(参考SEQ ID NO:27)的位置编码天冬氨酸(通过密码子GAU或GAC)。在一些实施方案中,所述mRNA在对应于GPR156 mRNA的位置1587(参考SEQ ID NO:27)的位置处包含尿嘧啶。此尿嘧啶对于在对应于根据SEQ ID NO:5的位置529的位置处编码天冬氨酸的密码子的第三位置)。

[0081] 在一些实施方案中,所述mRNA包含SEQ ID NO:28的核酸序列或者由其组成,其在对应于SEQ ID NO:26的位置592至603的位置处含有12个核苷酸的缺失。在一些实施方案中,所述mRNA包含在SEQ ID NO:28中列出的核酸序列或者由其组成,其中位置1585至1587是GAC,从而编码天冬氨酸。在一些实施方案中,所述mRNA包含与SEQ ID NO:28具有至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的核酸序列或者由其组成(并且在对应于在根

据SEQ ID NO:5的位置529处编码天冬氨酸的位置的位置处包含GAC密码子)。在一些实施方案中,此GPR156 mRNA的对应于位置1534至1536(参考SEQ ID NO:28)的位置编码谷氨酸(通过密码子GAA或GAG)。在一些实施方案中,此GPR156 mRNA的对应于位置1534至1536(参考SEQ ID NO:28)的位置编码天冬氨酸(通过密码子GAU或GAC)。在一些实施方案中,所述mRNA在对应于GPR156 mRNA的位置1587(参考SEQ ID NO:28)的位置处包含胞嘧啶。此胞嘧啶对应于在对应于根据SEQ ID NO:5的位置529的位置处编码天冬氨酸的密码子的第三位置)。

[0082] 在一些实施方案中,所述mRNA编码SEQ ID NO:4的位置310至814。在一些实施方案中,此mRNA的核酸序列在SEQ ID NO:29中列出,其中位置670至672编码天冬氨酸(通过密码子GAU)。在一些实施方案中,所述mRNA包含与SEQ ID NO:29具有至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的核酸序列或者由其组成,并且在对应于在对应于根据SEQ ID NO:4的位置533的位置处编码天冬氨酸的位置的位置处包含GAU密码子。在一些实施方案中,GPR156 mRNA的对应于位置619至621(参考SEQ ID NO:29)的位置编码谷氨酸。在一些实施方案中,GPR156 mRNA的对应于位置619至621(参考SEQ ID NO:29)的位置编码天冬氨酸。在一些实施方案中,所述mRNA在对应于GPR156 mRNA的位置672(参考SEQ ID NO:29)的位置处包含尿嘧啶。此尿嘧啶对应于在对应于根据SEQ ID NO:4的位置533的位置处编码天冬氨酸的密码子的第三位置)。

[0083] 在一些实施方案中,编码SEQ ID NO:4的位置310至814的mRNA的核酸序列在SEQ ID NO:30中列出,其中位置670至672编码天冬氨酸(通过密码子GAC)。在一些实施方案中,所述mRNA包含与SEQ ID NO:30具有至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的核酸序列或者由其组成,并且在对应于在对应于根据SEQ ID NO:4的位置533的位置处编码天冬氨酸的位置的位置处包含GAC密码子。在一些实施方案中,GPR156 mRNA的对应于位置619至621(参考SEQ ID NO:30)的位置编码谷氨酸。在一些实施方案中,GPR156 mRNA的对应于位置619至621(参考SEQ ID NO:30)的位置编码天冬氨酸。在一些实施方案中,所述mRNA在对应于GPR156 mRNA的位置672(参考SEQ ID NO:30)的位置处包含胞嘧啶。此胞嘧啶对应于在对应于根据SEQ ID NO:4的位置533的位置处编码天冬氨酸的密码子的第三位置)。

[0084] 在一些实施方案中,所述分离的核酸分子包含具有本文公开的任何mRNA分子的位置1521至1680的核酸序列,从而在对应于位置1599的位置处具有尿嘧啶。

[0085] 在一些实施方案中,所述分离的核酸分子包含比完整GPR156 mRNA序列少的核苷酸。在一些实施方案中,所述分离的核酸分子包含SEQ ID NO:25、SEQ ID NO:26、SEQ ID NO:27、SEQ ID NO:28、SEQ ID NO:29或SEQ ID NO:30的至少约5个、至少约8个、至少约10个、至少约12个、至少约15个、至少约20个、至少约25个、至少约30个、至少约35个、至少约40个、至少约45个、至少约50个、至少约60个、至少约70个、至少约80个、至少约90个、至少约100个、至少约200个、至少约300个、至少约400个、至少约500个、至少约600个、至少约700个、至少约800个、至少约900个、至少约1000个、至少约1100个、至少约1200个、至少约1300个、至少约1400个、至少约1500个、至少约1600个、至少约1700个、至少约1800个、至少约1900个、至少约2000个、至少约2100个、至少约2200个、至少约2300个或至少约2400个连续核苷酸或者由其组成。在一些实施方案中,所述分离的核酸分子包含SEQ ID NO:25、SEQ ID

NO:26、SEQ ID NO:27、SEQ ID NO:28、SEQ ID NO:29或SEQ ID NO:30的至少约1000个到至少约2000个连续核苷酸或者由其组成。就这一点而言,较长的mRNA分子相对于较短的mRNA分子是优选的。在一些实施方案中,所述分离的核酸分子包含SEQ ID NO:25、SEQ ID NO:26、SEQ ID NO:27、SEQ ID NO:28、SEQ ID NO:29或SEQ ID NO:30的至少约15个、至少约20个、至少约25个、至少约30个、至少约35个、至少约40个、至少约45个、至少约50个、至少约60个、至少约70个、至少约80个、至少约90个、至少约100个、至少约200个、至少约300个、至少约400个、至少约500个、至少约600个、至少约700个、至少约800个、至少约900个或至少约1000个连续核苷酸或者由其组成。就这一点而言,较长的mRNA分子相对于较短的mRNA分子是优选的。在一些实施方案中,此类mRNA分子包含在对应于根据SEQ ID NO:4的位置533的位置处编码天冬氨酸的密码子或包含分别对应于SEQ ID NO:25或SEQ ID NO:26的位置1599的尿嘧啶或半胱氨酸。

[0086] 本公开还提供分离的核酸分子,其与变体GPR156基因组DNA(诸如SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11或SEQ ID NO:12)、变体GPR156小基因、变体GPR156 cDNA(诸如SEQ ID NO:16、SEQ ID NO:17、SEQ ID NO:18、SEQ ID NO:19、SEQ ID NO:20或SEQ ID NO:21)和/或变体GPR156 mRNA(诸如SEQ ID NO:25、SEQ ID NO:26、SEQ ID NO:27、SEQ ID NO:28、SEQ ID NO:29或SEQ ID NO:30)杂交。在一些实施方案中,此类分离的核酸分子包含至少约5个、至少约8个、至少约10个、至少约11个、至少约12个、至少约13个、至少约14个、至少约15个、至少约16个、至少约17个、至少约18个、至少约19个、至少约20个、至少约21个、至少约22个、至少约23个、至少约24个、至少约25个、至少约30个、至少约35个、至少约40个、至少约45个、至少约50个、至少约55个、至少约60个、至少约65个、至少约70个、至少约75个、至少约80个、至少约85个、至少约90个、至少约95个、至少约100个、至少约200个、至少约300个、至少约400个、至少约500个、至少约600个、至少约700个、至少约800个、至少约900个、至少约1000个、至少约2000个、至少约3000个、至少约4000个、至少约5000个、至少约6000个、至少约7000个、至少约8000个、至少约9000个、至少约10000个、至少约11000个、至少约12000个、至少约13000个、至少约14000个、至少约15000个、至少约16000个、至少约17000个、至少约18000个、至少约19000个或至少约20000个核苷酸或者由其组成。在一些实施方案中,所述分离的核酸分子包含至少15个核苷酸或者由其组成。在一些实施方案中,所述分离的核酸分子包含至少15个核苷酸到至少约35个核苷酸或者由其组成。在一些实施方案中,此类分离的核酸分子在严格条件下与变体GPR156基因组DNA(诸如SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11或SEQ ID NO:12)、变体GPR156小基因、变体GPR156 cDNA(诸如SEQ ID NO:16、SEQ ID NO:17、SEQ ID NO:18、SEQ ID NO:19、SEQ ID NO:20或SEQ ID NO:21)和/或变体GPR156 mRNA(诸如SEQ ID NO:25、SEQ ID NO:26、SEQ ID NO:27、SEQ ID NO:28、SEQ ID NO:29或SEQ ID NO:30)杂交。此类核酸分子可例如用作探针、用作引物或用作等位基因特异性引物,如本文所述或示例。

[0087] 在一些实施方案中,所述分离的核酸分子与和变体GPR156基因组DNA(诸如SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11或SEQ ID NO:12)、变体GPR156小基因、变体GPR156 cDNA(诸如SEQ ID NO:16、SEQ ID NO:17、SEQ ID NO:18、SEQ ID NO:19、SEQ ID NO:20或SEQ ID NO:21)和/或变体GPR156 mRNA(诸如SEQ ID NO:25、SEQ ID NO:26、SEQ ID NO:27、SEQ ID NO:28、SEQ ID NO:29或SEQ ID NO:30)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96%、至少约97%、至少约98%、至少约99%或100%相同的核酸分子的至少约15个连续核苷酸杂交。在一些实施方案中，所述分离的核酸分子包含约15个至约100个核苷酸或约15个至约35个核苷酸或者由其组成。在一些实施方案中，所述分离的核酸分子包含约15个至约100个核苷酸或者由其组成。在一些实施方案中，所述分离的核酸分子包含约15个至约35个核苷酸或者由其组成。

[0088] 在一些实施方案中，本文公开的核酸分子、基因组DNA分子、cDNA分子或mRNA分子中的任一种可以是纯化的，例如至少约90%纯的。在一些实施方案中，本文公开的核酸分子、基因组DNA分子、cDNA分子或mRNA分子中的任一种可以是纯化的，例如至少约95%纯的。在一些实施方案中，本文公开的核酸分子、基因组DNA分子、cDNA分子或mRNA分子中的任一种可以是纯化的，例如至少约99%纯的。纯化用人造的纯化技术手动进行。

[0089] 本公开还提供本文公开的分离的核酸分子、基因组DNA分子、cDNA分子或mRNA分子中的任一种的片段。在一些实施方案中，所述片段包含本文公开的核酸序列中的任一种或其任何补体的至少约5个、至少约8个、至少约10个、至少约11个、至少约12个、至少约13个、至少约14个、至少约15个、至少约16个、至少约17个、至少约18个、至少约19个、至少约20个、至少约21个、至少约22个、至少约23个、至少约24个、至少约25个、至少约30个、至少约35个、至少约40个、至少约45个、至少约50个、至少约55个、至少约60个、至少约65个、至少约70个、至少约75个、至少约80个、至少约85个、至少约90个、至少约95个或至少约100个连续残基或者由其组成。就这一点而言，较长的片段相对于较短的片段是优选的。在一些实施方案中，所述片段包含至少约5个、至少约8个、至少约10个、至少约11个、至少约12个、至少约13个、至少约14个、至少约15个、至少约16个、至少约17个、至少约18个、至少约19个、至少约20个、至少约21个、至少约22个、至少约23个、至少约24个、至少约25个、至少约30个、至少约35个、至少约40个、至少约45个或至少约50个连续残基或者由其组成。就这一点而言，较长的片段相对于较短的片段是优选的。在一些实施方案中，所述片段包含至少约20个、至少约25个、至少约30个或至少约35个连续残基或者由其组成。在一些实施方案中，所述片段包含至少约20个连续残基或者由其组成。在一些实施方案中，所述片段包含至少约25个连续残基或者由其组成。在一些实施方案中，所述片段包含至少约30个连续残基或者由其组成。在一些实施方案中，所述片段包含至少约35个连续残基或者由其组成。设想，所述片段包含在对应于具有根据SEQ ID NO:4的氨基酸序列的蛋白质的位置533的位置处编码天冬氨酸或在对应于具有根据SEQ ID NO:5的氨基酸序列的蛋白质的位置529的位置处编码天冬氨酸的核酸分子的部分或者由其组成。此类片段可例如用作探针、用作引物或用作等位基因特异性引物，如本文所述或例示。

[0090] 本公开还提供探针和引物。本公开的探针或引物具有与本文公开的核酸分子中的任一种特异性地杂交的核酸序列或其补体。在一些实施方案中，所述探针或引物在严格条件下与本文公开的核酸分子中的任一种特异性地杂交。本公开还提供具有在中等条件下与本文公开的核酸分子中的任一种杂交的核酸序列或其补体的核酸分子。根据本公开的探针或引物优选地涵盖编码GPR156蛋白内的E533D变化的核酸密码子或其补体。因此，在优选的实施方案中，本公开提供在本文中以上和以下更详细地定义的等位基因特异性引物。

[0091] 根据本公开的探针或引物可用于检测编码GPR156蛋白（例如，根据SEQ ID NO:4）的核酸序列内的E533D变化和/或编码GPR156的同种型（例如，根据SEQ ID NO:5）的核酸分

子内的E529D突变。例如,根据本公开的引物可用于扩增包含E533D变化的GPR156或其片段。根据本公开的引物还可用于扩增包含E529D突变的GPR156同种型(诸如具有SEQ ID NO:5的氨基酸序列的同种型)或其片段。根据本公开的例示的引物在表1中示出。

[0092] 本公开还提供包含以上所述的引物中的一种的一对引物。通过使用这种一对引物,还可检测GPR156内的长度多态性。例如,一对引物可用于将编码典型GPR156蛋白(例如,SEQ ID NO:4的GPR156蛋白)的核酸分子与编码在GPR156蛋白的N末端至中心区中缺少四个氨基酸的GPR156同种型(例如,SEQ ID NO:5的GPR156同种型)的核酸分子区分开。更具体地,如果引物杂交的位置侧接在GPR156同种型中缺少的四个氨基酸的位置,则与对应于GPR156同种型的扩增片段相比,对于典型GPR156蛋白的扩增片段更长。

[0093] 本文公开的核酸分子可包含天然存在的GPR156基因、cDNA或mRNA转录物的核酸序列,或者可包含非天然存在的序列。在一些实施方案中,由于同义突变或不影响编码的GPR156多肽的突变,天然存在的序列可不同于非天然存在的序列。例如,除同义突变或不影响编码的GPR156多肽的突变以外,所述序列可以是相同的。同义突变或取代是在编码蛋白质的基因的外显子中用一个核苷酸取代另一个核苷酸,使得产生的氨基酸序列不被修改。因为遗传密码的简并性,这是可能的,其中一些氨基酸通过多于一个三碱基对密码子编码。同义取代例如在密码子优化的过程中使用。本文公开的核酸分子可以是密码子优化的。

[0094] 本文还提供可与公开的核酸分子相互作用的功能性多核苷酸。功能性多核苷酸是具有特定功能(诸如结合靶分子或催化特定反应)的核酸分子。功能性多核苷酸的实例包括但不限于反义分子、适体、核酶、三链体形成分子以及外部导向序列。功能性多核苷酸可充当靶分子所具有的特定活性的影响剂、抑制剂、调节剂和刺激剂,或者功能性多核苷酸可具有独立于任何其他分子的全新活性。

[0095] 反义分子被设计成通过典型或非典型碱基配对来与靶核酸分子相互作用。反义分子与靶分子的相互作用被设计成通过例如RNA酶-H介导的RNA-DNA杂种降级来促进靶分子的破坏。可替换地,反义分子被设计成干扰正常地发生在靶分子上的处理功能,诸如转录或复制。反义分子可基于靶分子的序列来设计。存在用于通过鉴定靶分子的最可接近区域来优化反义效率的多种方法。示例性方法包括但不限于体外选择实验和使用DMS和DEPC进行的DNA修饰研究。反义分子通常以小于或等于约 10^{-6} 、小于或等于约 10^{-8} 、小于或等于约 10^{-10} 或小于或等于约 10^{-12} 的解离常数(k_d)结合靶分子。协助反义分子的设计和使用的方法和技术的代表性样品可见于美国专利的以下非限制性列表:5,135,917;5,294,533;5,627,158;5,641,754;5,691,317;5,780,607;5,786,138;5,849,903;5,856,103;5,919,772;5,955,590;5,990,088;5,994,320;5,998,602;6,005,095;6,007,995;6,013,522;6,017,898;6,018,042;6,025,198;6,033,910;6,040,296;6,046,004;6,046,319;以及6,057,437。反义分子的实例包括但不限于反义RNA、小的干扰性RNA(siRNA)和短发夹RNA(shRNA)。

[0096] 本文公开的分离的核酸分子可包括RNA、DNA或RNA和DNA两者。所述分离的核酸分子还可连接或融合到异源核酸序列(诸如在载体中)或异源标记。例如,本文公开的分离的核酸分子可在包含分离的核酸分子和异源核酸序列的载体或外源供体序列中。所述分离的核酸分子还可连接或融合到异源标记,诸如荧光标记。标记的其他实例在本文中其他地方公开。

[0097] 所述标记可以是直接可检测的(例如,荧光团)或间接可检测的(例如,半抗原、酶

或荧光团淬灭剂)。此类标记可通过光谱学、光化学、生物化学、免疫化学或化学手段检测。此类标记包括例如放射性标记,其可用辐射计数装置测量;颜料、染料或其他色原,其可用分光光度计在视觉上观察或测量;自旋标记,其可用自旋标记分析仪测量;以及荧光标记(例如,荧光团),其中输出信号通过合适的分子加成物的激励生成,并且其可通过用被染料吸收的光进行的激励可视化或者可用标准荧光计或成像系统测量。所述标记还可以是例如化学发光物质,其中输出信号通过信号化合物的化学修饰生成;含有金属的物质;或酶,其中发生信号的酶依赖性二次生成,诸如从无色底物形成有色产物。术语“标记”还可指代“标签”或半抗原,其可选择性地结合到缀合分子,使得当随后连同底物一起添加时,缀合分子用于生成可检测的信号。例如,可使用生物素作为标签,并且然后使用辣根过氧化物酶(HRP)的抗生物素蛋白或链霉抗生物素缀合物来结合到标签,并且然后使用量热底物(例如,四甲基联苯胺(TMB))或荧光底物来检测HRP的存在。可用作标签来有助于纯化的示例性标记包括但不限于myc、HA、FLAG或3XFLAG、6XHis或聚组氨酸、谷胱甘肽-S-转移酶(GST)、麦芽糖结合蛋白、表位标签或免疫球蛋白的Fc部分。多种标记是已知的,并且包括例如颗粒、荧光团、半抗原、酶及其量热、荧光和化学发光底物和其他标记。

[0098] 公开的核酸分子可包括例如核苷酸或非天然或修饰的核苷酸,诸如核苷酸类似物或核苷酸替代物。此类核苷酸包括含有修饰的碱基、糖或磷酸酯基团或者在其结构中并入非天然部分的核苷酸。非天然核苷酸的实例包括但不限于双脱氧寡核苷酸、生物素化的、氨基化、脱氨基化的、烷基化的、苄基化的和荧光团标记的核苷酸。

[0099] 本文公开的核酸分子还可包括一个或多个核苷酸类似物或取代。核苷酸类似物是含有碱基、糖或磷酸酯部分的修饰的核苷酸。对碱基部分的修饰包括但不限于A、C、G和T/U以及不同的嘌呤或嘧啶碱基(例如像假尿苷、尿嘧啶-5-基、次黄嘌呤-9-基(I)和2-氨基腺嘌呤-9-基)的天然和合成修饰。修饰的碱基包括但不限于5-甲基胞嘧啶(5-me-C)、5-羟甲基胞嘧啶、黄嘌呤、次黄嘌呤、2-氨基腺嘌呤、腺嘌呤和鸟嘌呤的6-甲基和其他烷基衍生物、腺嘌呤和鸟嘌呤的2-丙基和其他烷基衍生物、2-硫代尿嘧啶、2-硫代胸腺嘧啶和2-硫代胞嘧啶、5-卤代尿嘧啶和胞嘧啶、5-丙炔基尿嘧啶和胞嘧啶、6-偶氮尿嘧啶、胞嘧啶和胸腺嘧啶、5-尿嘧啶(假尿嘧啶)、4-硫代尿嘧啶、8-卤代、8-氨基、8-硫代、8-硫代烷基、8-羟基和其他8-取代的腺嘌呤和鸟嘌呤、5-卤代(具体地5-溴)、5-三氟甲基和其他5-取代的尿嘧啶和胞嘧啶、7-甲基鸟嘌呤和7-甲基腺嘌呤、8-氮杂鸟嘌呤和8-氮杂腺嘌呤、7-脱氮鸟嘌呤和7-脱氮腺嘌呤以及3-脱氮鸟嘌呤和3-脱氮腺嘌呤。某些核苷酸类似物(例如像5-取代的嘧啶、6-氮杂嘧啶以及N-2、N-6和O-6取代的嘌呤,包括但不限于2-氨基丙基腺嘌呤、5-丙炔基尿嘧啶、5-丙炔基胞嘧啶和5-甲基胞嘧啶)可增加双链体形成的稳定性。碱基修饰可常常与例如糖修饰(诸如2'-0-甲氧基乙基)组合以便实现独特特性,诸如增加的双链体稳定性。

[0100] 核苷酸类似物还可包括糖部分的修饰。对糖部分的修饰包括但不限于核糖和脱氧核糖的天然修饰以及合成修饰。糖修饰包括但不限于在2'位置处的以下修饰:OH;F;O-、S-或N-烷基;O-、S-或N-烯基;O-、S-或N-炔基;或O-烷基-O-烷基,其中烷基、烯基和炔基可以是取代或未取代的C₁₋₁₀烷基或、C₂₋₁₀烯基和C₂₋₁₀炔基。示例性2'糖修饰还包括但不限于-O[(CH₂)_nO]_mCH₃、-O(CH₂)_nOCH₃、-O(CH₂)_nNH₂、-O(CH₂)_nCH₃、-O(CH₂)_n-ONH₂和-O(CH₂)_nON[(CH₂)_nCH₃]₂,其中n和m是1至约10。

[0101] 2'位置处的其他修饰包括但不限于C₁₋₁₀烷基、取代的低级烷基、烷芳基、芳烷基、

0-烷芳基或0-芳烷基、SH、SCH₃、OCN、Cl、Br、CN、CF₃、OCF₃、SOCH₃、SO₂CH₃、ONO₂、NO₂、N₃、NH₂、杂环烷基、杂环烷芳基、氨基烷基氨基、聚烷基氨基、取代的甲硅烷基、RNA裂解基团、报道基团、嵌入剂、用于改进寡核苷酸的药物代谢动力学特性的基团或用于改进寡核苷酸的药效学特性的基团，以及具有相似特性的其他取代基。相似的修饰还可在糖上的其他位置处进行，特别是3'末端核苷酸上或2'-5'连接型寡核苷酸中的糖的3'位置和3'末端核苷酸的5'位置。修饰的糖还可包括在桥接的环氧处含有修饰（诸如CH₂和S）的那些。核苷酸糖类似物还可具有糖模拟物，诸如环丁基部分代替呋喃戊糖基糖。

[0102] 核苷酸类似物还可在磷酸酯部分处进行修饰。修饰的磷酸酯部分包括但不限于可进行修饰、使得两个核苷酸之间的键联含有硫代磷酸酯、手性硫代磷酸酯、二硫代磷酸酯、磷酸三酯、氨基烷基磷酸三酯、甲基和其他烷基磷酸酯（包括3' - 亚烷基磷酸酯和手性磷酸酯、亚膦酸酯）、氨基磷酸酯（包括3' - 氨基氨基磷酸酯和氨基烷基氨基磷酸酯、硫羰氨基磷酸酯）、硫羰烷基磷酸酯、硫羰烷基磷酸三酯和硼烷基磷酸酯的那些。两个核苷酸之间的这些磷酸酯或修饰的磷酸酯键联可通过3' - 5' 键联或2' - 5' 键联，并且所述键联可包含反转的极性诸如3' - 5' 至5' - 3' 或2' - 5' 至5' - 2'。还包括各种盐、混合盐以及游离酸形式。

[0103] 核苷酸取代物包括具有与核苷酸相似的功能特性、但不含有磷酸酯部分的分子，诸如肽核酸（PNA）。核苷酸取代物包括将以Watson-Crick或Hoogsteen方式来识别核酸、但通过不同于磷酸酯部分的部分连接在一起的分子。在与适当的靶标核酸相互作用时，核苷酸取代物能够符合双螺旋类型结构。

[0104] 核苷酸取代物还包括磷酸酯部分或糖部分已被替代的核苷酸或核苷酸类似物。在一些实施方案中，核苷酸取代物可不包含标准磷原子。磷酸酯的取代物可以是例如短链烷基或环烷基核苷间键联、混合的杂原子和烷基或环烷基核苷间键联、或一个或多个短链杂原子或杂环核苷间键联。这些包括具有以下的那些：吗啉代键联（部分由核苷的糖部分形成）；硅氧烷主链；硫化物、亚砜和砜主链；甲酰基（formacetyl）和硫代甲酰基主链；亚甲基甲酰基和硫代甲酰基主链；含烯烃的主链；氨基磺酸酯主链；亚甲基亚胺基和亚甲基肼基主链；磺酸酯和磺酰胺主链；酰胺主链；以及具有混合的N、O、S和CH₂组成部分的其他主链。

[0105] 还应了解，在核苷酸取代物中，核苷酸的糖部分和磷酸酯部分两者均可通过例如酰胺类型键联（氨基乙基甘氨酸）（PNA）被替代。

[0106] 也可将其他类型分子（缀合物）连接至核苷酸或核苷酸类似物以增强例如细胞吸收。缀合物可化学连接至核苷酸或核苷酸类似物。此类缀合物包括例如脂质部分，诸如胆固醇部分、胆酸、硫醚（诸如己基-S-三苯甲基硫醇基）、硫代胆固醇、脂肪族链（诸如十二烷二醇或十一烷基残基）、磷脂（诸如二-十六烷基-外消旋-甘油或1,2-二-O-十六烷基-外消旋-甘油-3-H-膦酸三乙铵）、聚胺或聚乙二醇链、金刚烷乙酸、棕榈基部分或十八烷基胺或己基氨基-羧基-羟胆固醇部分。

[0107] 本公开还提供包含本文公开的核酸分子中的任一种或多种的载体。在一些实施方案中，所述载体包含本文公开的核酸分子中的任一种或多种和异源核酸。所述载体可以是能够转运核酸分子的病毒或非病毒载体。在一些实施方案中，所述载体是质粒或粘粒（例如，可将额外的DNA区段连接到其中的环状双链DNA）。在一些实施方案中，所述载体是病毒载体，其中额外的DNA区段可连接到病毒基因组中。在一些实施方案中，所述载体可在其被引入的宿主细胞中自主复制（例如，具有细菌复制起点的细菌载体和附加型哺乳动物载

体)。在一些实施方案中,所述载体(例如,非附加型哺乳动物载体)可在引入到宿主细胞中之后整合到宿主细胞的基因组中,并且由此与宿主基因组一起复制。此外,特定载体可指导它们所操作性地连接到的基因的表达。此类载体在本文中被称为“重组表达载体”或“表达载体”。此类载体还可以是靶向载体(即,外源供体序列)。

[0108] 在一些实施方案中,由本文公开的各种遗传变体编码的蛋白质通过将编码公开的遗传变体的核酸分子插入到表达载体中、使得基因操作性地连接到表达控制序列(诸如转录和翻译控制序列)来表达。表达载体包括但不限于质粒、粘粒、逆转录病毒、腺病毒、腺相关病毒(AAV)、植物病毒(诸如花椰菜花叶病毒和烟草花叶病毒)、酵母人工染色体(YAC)、Epstein-Barr(EBV)衍生的附加体以及本领域中已知的其他表达载体。在一些实施方案中,包含公开的遗传变体的核酸分子可连接到载体中,使得载体内的转录和翻译控制序列发挥其调控遗传变体的转录和翻译的预期功能。所选择的表达载体和表达控制序列应可与所使用的表达宿主细胞相容。包含公开的遗传变体的核酸序列可作为变体遗传信息插入到单独的载体中或插入到同一表达载体中。包含公开的遗传变体的核酸序列可通过标准方法(例如,连接包含公开的遗传变体的核酸和载体上的互补限制性位点,或者在不存在限制性位点的情况下进行钝端连接)插入到表达载体中。

[0109] 除包含公开的遗传变体的核酸序列之外,重组表达载体可携带控制遗传变体在宿主细胞中的表达的调控序列。表达载体的设计(包括调控序列的选择)可取决于诸如以下的因素:待转化的宿主细胞的选择、所需蛋白质的表达水平等。用于哺乳动物宿主细胞表达的所需调控序列可包括例如指导哺乳动物细胞中的高水平蛋白质表达的病毒元件,诸如来源于逆转录病毒LTR、巨细胞病毒(CMV)(诸如CMV启动子/增强子)、猿猴病毒40(SV40)(诸如SV40启动子/增强子)、腺病毒(例如,腺病毒主要晚期启动子(AdMLP))的启动子和/或增强子、多瘤病毒和哺乳动物强启动子(诸如天然免疫球蛋白和肌动蛋白启动子)。在细菌细胞或真菌细胞(例如,酵母细胞)中表达多肽的方法也是熟知的。

[0110] 启动子可以是例如组成型活性启动子、条件启动子、诱导型启动子、时间限制的启动子(例如,发育调控的启动子)或空间限制的启动子(例如,细胞特异性或组织特异性启动子)。启动子的实例可见于例如WO 2013/176772。

[0111] 诱导型启动子的实例包括例如化学调控的启动子和物理调控的启动子。化学调控的启动子包括例如醇调控的启动子(例如,醇脱氢酶(alcA)基因启动子)、四环素调控的启动子(例如,四环素应答型启动子、四环素操纵基因序列(tetO)、tet-On启动子或tet-Off启动子)、类固醇调控的启动子(例如,大鼠糖皮质激素受体、雌激素受体的启动子或蜕皮素受体的启动子)或金属调控的启动子(例如,金属蛋白启动子)。物理调控的启动子包括例如温度调控的启动子(例如,热激启动子)和光调控的启动子(例如,光诱导型启动子或光抑制型启动子)。

[0112] 组织特异性启动子可以是例如神经元特异性启动子、神经胶质特异性启动子、肌肉细胞特异性启动子、心脏细胞特异性启动子、肾细胞特异性启动子、骨细胞特异性启动子、内皮细胞特异性启动子或免疫细胞特异性启动子(例如,B细胞启动子或T细胞启动子)。

[0113] 发育调控的启动子包括例如仅在发育的胚胎期过程中或仅在成人细胞中活化的启动子。

[0114] 除包含公开的遗传变体的核酸序列和调控序列之外,重组表达载体可携带额外的

序列,诸如调控载体在宿主细胞中复制的序列(例如,复制起点)和可选择标记物基因。可选择标记物基因有助于载体已引入其中的宿主细胞的选择(参见例如美国专利4,399,216;4,634,665;以及5,179,017)。例如,可选择标记物基因可将对于药物诸如G418、潮霉素或甲氨蝶呤的抗性赋予载体已引入其中的宿主细胞。示例性的可选择标记物基因包括但不限于二氢叶酸还原酶(DHFR)基因(用于使用甲氨蝶呤选择/扩增的dhfr-宿主细胞中)、新基因(用于G418选择)和谷氨酸合成酶(GS)基因。

[0115] 额外的载体描述于例如2016年7月28日提交的美国临时申请号62/367,973,所述临时申请以引用的方式整体并入本文。

[0116] 本公开还提供包含本文公开的分离的核酸分子、基因组DNA分子、cDNA分子或mRNA分子中的任一种或多种的组合物。在一些实施方案中,所述组合物是药物组合物。

[0117] 本公开还提供变体GPR156多肽。本公开提供包含GPR156蛋白的至少一部分的分离的或重组的多肽,其中所述部分在对应于根据SEQ ID NO:4的位置533的位置处包含天冬氨酸。在一些实施方案中,所述分离的或重组的多肽包含GPR156蛋白的至少一部分,其中所述部分在对应于根据SEQ ID NO:5的位置529的位置处包含天冬氨酸。

[0118] 在一些实施方案中,所述分离的或重组的多肽包含与SEQ ID NO:4具有至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的氨基酸序列或者由其组成,其中多肽的对应于位置533(根据SEQ ID NO:4)的位置处的氨基酸是天冬氨酸。在一些实施方案中,所述多肽包含在对应于根据SEQ ID NO:4的位置533的位置处包含天冬氨酸的全长GPR156蛋白。在一些实施方案中,所述全长GPR156蛋白包含与SEQ ID NO:4具有至少约90%、约91%、约92%、约93%、约94%、约95%、约96%、约97%、约98%或约99%序列同一性的氨基酸序列或者由其组成,其中多肽的对应于位置533(根据SEQ ID NO:4)的位置处的氨基酸是天冬氨酸。在一些实施方案中,所述多肽包含SEQ ID NO:4的氨基酸序列。在一些实施方案中,所述多肽由SEQ ID NO:4的氨基酸序列组成。此变体GPR156蛋白的对应于位置516(参考SEQ ID NO:4)的位置是谷氨酸或天冬氨酸。在一些实施方案中,此变体GPR156蛋白的对应于位置516(参考SEQ ID NO:4)的位置是谷氨酸。在一些实施方案中,此变体GPR156蛋白的对应于位置516(参考SEQ ID NO:4)的位置是天冬氨酸。

[0119] 在一些实施方案中,所述分离的或重组的多肽包含与SEQ ID NO:5具有至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的氨基酸序列或者由其组成,其中多肽的对应于位置529(根据SEQ ID NO:5)的位置处的氨基酸是天冬氨酸。在一些实施方案中,所述多肽包含在对应于根据SEQ ID NO:5的位置529的位置处包含天冬氨酸的GPR156蛋白。在一些实施方案中,所述GPR156蛋白包含与SEQ ID NO:5具有至少约90%、约91%、约92%、约93%、约94%、约95%、约96%、约97%、约98%或约99%序列同一性的氨基酸序列或者由其组成,其中多肽的对应于位置529(根据SEQ ID NO:5)的位置处的氨基酸是天冬氨酸。在一些实施方案中,所述多肽包含SEQ ID NO:5的氨基酸序列。在一些实施方案中,所述多肽由SEQ ID NO:5的氨基酸序列组成。此变体GPR156蛋白的对应于位置512(参考SEQ ID NO:5)的位置是谷氨酸或天冬氨酸。在一些实施方案中,此变体GPR156蛋白的对应于位置512(参考SEQ ID NO:5)的位置是谷氨酸。在一些实施方案中,此变体GPR156蛋白的对应于位置512

(参考SEQ ID N0:5)的位置是天冬氨酸。

[0120] 在一些实施方案中,所述多肽包含对应于SEQ ID N0:4的位置310至814的GPR156蛋白的部分。在一些实施方案中,所述多肽包含与SEQ ID N0:4的位置310至814具有至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的氨基酸序列或者由其组成,其中对应于根据SEQ ID N0:4的位置533的位置(例如,位置224)处的氨基酸是天冬氨酸。在一些实施方案中,所述多肽包含SEQ ID N0:6的氨基酸序列。在一些实施方案中,所述多肽由SEQ ID N0:6的氨基酸序列组成。此GPR156蛋白的对应于位置207(参考SEQ ID N0:6)的位置是谷氨酸或天冬氨酸。在一些实施方案中,此GPR156蛋白的对应于位置207(参考SEQ ID N0:6)的位置是谷氨酸。在一些实施方案中,此GPR156蛋白的对应于位置207(参考SEQ ID N0:6)的位置是天冬氨酸。

[0121] 本公开还提供本文公开的多肽中的任一种的片段。在一些实施方案中,所述片段包含编码的多肽的至少约10个、至少约15个、至少约20个、至少约25个、至少约30个、至少约35个、至少约40个、至少约45个、至少约50个、至少约55个、至少约60个、至少约65个、至少约70个、至少约75个、至少约80个、至少约85个、至少约90个、至少约95个、至少约100个、至少约150个、至少约200个、至少约250个、至少约300个、至少约350个、至少约400个、至少约450个、至少约500个、至少约550个、至少约600个、至少约650个、至少约700个、至少约750个或至少约800个连续氨基酸残基。就这一点而言,较长的片段相对于较短的片段是优选的。在一些实施方案中,所述片段包含编码的多肽的至少约10个、至少约15个、至少约20个、至少约25个、至少约30个、至少约35个、至少约40个、至少约45个、至少约50个、至少约55个、至少约60个、至少约65个、至少约70个、至少约75个、至少约80个、至少约85个、至少约90个、至少约95个或至少约100个连续氨基酸残基。就这一点而言,较长的片段相对于较短的片段是优选的。

[0122] 本公开还提供包含具有变体GPR156的分离的多肽的二聚体,其中所述多肽选自本文公开的多肽中的任一种。

[0123] 在一些实施方案中,本文公开的分离的多肽连接或融合到异源多肽或异源分子或标记,其多个实例在本文中其他地方公开。例如,所述蛋白质可融合到异源多肽,从而提供增加的或降低的稳定性。融合的结构域或异源多肽可定位在多肽的N末端、C末端或多肽内。融合配偶体可例如有助于提供T辅助表位(免疫融合配偶体),或者有助于在高于天然重组多肽的产率下表达蛋白质(表达增强子)。某些融合配偶体是免疫和表达增强融合配偶体两者。其他融合配偶体可选择来增加多肽的溶解性或有助于将多肽靶向到所需的细胞内隔室。一些融合配偶体包括亲和力标签,其有助于多肽的纯化。

[0124] 在一些实施方案中,融合蛋白直接融合到异源分子,或通过接头(诸如肽接头)连接到异源分子。合适的肽接头序列可例如基于以下因素来选择:1)采用柔性延伸构象的能力;2)采用可与第一多肽和第二多肽上的功能性表位相互作用的二级结构的抗性;以及3)缺少可与多肽功能性表位反应的疏水或带电荷残基。例如,肽接头序列可含有Gly、Asn和Ser残基。其他近中性氨基酸(诸如Thr和Ala)也可用于接头序列。可有用地用作接头的氨基酸序列包括公开于例如以下中的那些:Maratea等人, Gene, 1985, 40, 39-46; Murphy等人, Proc. Natl. Acad. Sci. USA, 1986, 83, 8258-8262;以及美国专利4,935,233和4,751,180。接

头序列的长度通常可以是例如1个至约50个氨基酸。当第一多肽和第二多肽具有可用于将功能结构域分开并防止位阻影响的非必需N末端氨基酸区域时，通常不需要接头序列。

[0125] 在一些实施方案中，所述多肽可操作地连接到细胞穿透结构域。例如，所述细胞穿透结构域可衍生自HIV-1TAT蛋白，TLM细胞穿透基序衍生自人类乙型肝炎病毒、MPG、Pep-1、VP22，细胞穿透肽衍生自单纯疱疹病毒或聚精氨酸肽序列。参见例如WO 2014/089290。所述细胞穿透结构域可定位在蛋白质的N末端、C末端或蛋白质内的任何位置。

[0126] 在一些实施方案中，所述多肽可操作地连接到易于追踪或纯化的异源多肽，诸如荧光蛋白、纯化标签或表位标签。荧光蛋白的实例包括但不限于绿色荧光蛋白（例如，GFP、GFP-2、tagGFP、turboGFP、eGFP、Emerald、Azami Green、单体Azami Green、CopGFP、AceGFP、ZsGreen1）、黄色荧光蛋白（例如，YFP、eYFP、Citrine、Venus、YPet、PhiYFP、ZsYellow1）、蓝色荧光蛋白（例如，eBFP、eBFP2、Azurite、mKalamal、GFPuv、Sapphire、T-sapphire）、青色荧光蛋白（例如，eCFP、Cerulean、CyPet、AmCyan1、Midoriishi-Cyan）、红色荧光蛋白（例如，mKate、mKate2、mPlum、DsRed单体、mCherry、mRFP1、DsRed-Express、DsRed2、DsRed-单体、HcRed-Tandem、HcRed1、AsRed2、eqFP611、mRaspberry、mStrawberry、Jred）、橙色荧光蛋白（例如，mOrange、mKO、Kusabira-Orange、单体Kusabira-Orange、mTangerine、tdTomato）以及任何其他适合的荧光蛋白。标签的实例包括但不限于谷胱甘肽-S-转移酶(GST)、甲壳质结合蛋白(CBP)、麦芽糖结合蛋白、硫氧还蛋白(TRX)、聚(NANP)、串联亲和纯化(TAP)标签、myc、AcV5、AU1、AU5、E、ECS、E2、FLAG、血凝素(HA)、nus、Softag 1、Softag 3、Strep、SBP、Glu-Glu、HSV、KT3、S、S1、T7、V5、VSV-G、组氨酸(His)、生物素羧基载体蛋白(BCCP)以及钙调蛋白。在一些实施方案中，所述异源分子是免疫球蛋白Fc结构域、肽标签、转导结构域、聚(乙二醇)、聚唾液酸或乙醇酸。

[0127] 在一些实施方案中，所述分离的多肽包含非天然或修饰的氨基酸或肽类似物。例如，存在多种D-氨基酸或具有与天然存在的氨基酸不同的官能取代基的氨基酸。公开天然存在的肽的相反立体异构体以及肽类似物的立体异构体。这些氨基酸可易于通过使tRNA分子带有选择的氨基酸并对利用例如琥珀密码子的遗传构建体进行工程化以便以位点特异性方式将类似物氨基酸插入到肽链中来并入到多肽链中。

[0128] 在一些实施方案中，所述分离的多肽是肽模拟物，其可生产成类似于肽，但是不通过天然肽键联连接。例如，氨基酸或氨基酸类似物的键联包括但不限于-CH₂NH-、-CH₂S-、-CH₂-、-CH=CH-（顺式和反式）、-COCH₂-、-CH(OH)CH₂-和-CHH₂SO-。肽类似物可在键原子之间具有多于一个原子，诸如b-丙氨酸、g氨基丁酸等。氨基酸类似物和肽类似物常常具有增强的或所希望的特性，诸如更经济的生产、更大的化学稳定性、增强的药物特性（半衰期、吸收、效力、功效等）、改变的特异性（例如，宽范围的生物活性）、降低的抗原性以及其他所希望的特性。

[0129] 在一些实施方案中，所述分离的多肽包含D-氨基酸，其可用于生成更稳定的肽，因为D氨基酸不被肽酶识别。用相同类型的D-氨基酸系统性取代共有序列的一个或多个氨基酸（例如，D-赖氨酸代替L-赖氨酸）可用于生成更稳定的肽。半胱氨酸残基可用于将两个或更多个肽环化或附接在一起。这对于将肽限定为特定构型可以是有益的（参见例如Rizo和Gierasch, Ann. Rev. Biochem., 1992, 61, 387）。

[0130] 本公开还提供编码本文公开的多肽中的任一种的核酸分子。这包括与特定多肽序

列相关的所有简并序列(具有编码一个特定多肽序列的序列的所有核酸以及编码蛋白质序列的公开的变体和衍生物的所有核酸,包括简并核酸)。因此,虽然每个特定核酸序列可能未在本文中写出,但是实际上每一个序列通过公开的多肽序列在本文中公开并描述。

[0131] 核酸内的核酸序列或多肽内的氨基酸序列的特定伸长之间的同一性百分比(或互补性百分比)可常规地使用BLAST程序(基本的局部比对搜索工具)和PowerBLAST程序(Altschul等人,J.Mol.Biol.,1990,215,403-410;Zhang和Madden,Genome Res.,1997,7,649-656)或通过使用Gap程序(Wisconsin序列分析包,用于Unix的版本8,Genetics Computer集团,University Research Park,Madison Wis.)使用默认设置(其使用Smith和Waterman的算法(Adv.Appl.Math.,1981,2,482-489))来确定。在本文中,如果对序列同一性百分比进行参考,则较高的序列同一性百分比相对于较低的序列同一性百分比是优选的。

[0132] 本公开还提供包含本文公开的核酸序列中的任一种或多种和/或多肽中的任一种或多种和携带体和/或赋形剂的组合物。在一些实施方案中,所述携带体增加核酸分子和/或多肽的稳定性(例如,在给定的储存条件(例如,-20°C、4°C或环境温度)下延长降解产物保持低于阈值(诸如低于起始核酸或蛋白质的0.5重量%))的时间段;或增加体内稳定性)。携带体的实例包括但不限于聚(乳酸)(PLA)微球、聚(D,L-乳酸-共乙醇酸)(PLGA)微球、脂质体、胶束、反胶束、脂质螺旋物和脂质微管。携带体可包括缓冲盐溶液,诸如PBS、HBSS等。

[0133] 本公开还提供产生本文公开的多肽中的任一种或其片段的方法。此类多肽或其片段可通过任何合适的方法产生。例如,多肽或其片段可由包含编码此类多肽或其片段的核酸分子的宿主细胞(例如,重组表达载体)产生。此类方法可包括在足以产生多肽或其片段的条件下培养包含编码多肽或其片段的核酸分子的宿主细胞(例如,重组表达载体),从而产生多肽或其片段。所述核酸可操作地连接到在宿主细胞中具有活性的启动子,并且所述培养可在表达核酸的条件下实施。此类方法还可包括回收表达的多肽或其片段。所述回收还包括纯化多肽或其片段。

[0134] 用于蛋白质表达的合适的系统的实例包括宿主细胞,例如像:细菌细胞表达系统(例如,大肠杆菌(*Escherichia coli*)、乳酸乳球菌(*Lactococcus lactis*))、酵母细胞表达系统(例如,酿酒酵母(*Saccharomyces cerevisiae*)、巴斯德毕赤酵母(*Pichia pastoris*))、昆虫细胞表达系统(例如,杆状病毒介导的蛋白质表达)以及哺乳动物细胞表达系统。

[0135] 编码多肽或其片段的核酸分子的实例在本文中其他地方更详细地公开。在一些实施方案中,所述核酸分子是针对在宿主细胞中的表达优化的密码子。在一些实施方案中,核酸分子可操作地连接到在宿主细胞中具有活性的启动子。所述启动子可以是异源启动子(例如,不是天然存在的启动子的启动子)。适用于大肠杆菌的启动子的实例包括但不限于阿拉伯糖、lac、tac和T7启动子。适用于乳酸乳球菌的启动子的实例包括但不限于P170和乳酸链球菌肽启动子。适用于酿酒酵母的启动子的实例包括但不限于组成型启动子(诸如醇脱氢酶(ADH1)或烯醇酶(ENO)启动子)或诱导型启动子(诸如PHO、CUP1、GAL1和G10)。适用于巴斯德毕赤酵母的启动子的实例包括但不限于醇氧化酶I(AOX I)启动子、甘油醛3磷酸脱氢酶(GAP)启动子以及谷胱甘肽依赖性甲醛脱氢酶(FLDI)启动子。适用于杆状病毒介导的系统的启动子的实例是晚期病毒强多角体蛋白启动子。

[0136] 在一些实施方案中,所述核酸分子编码具有多肽或其片段的框架中的标签以有助于蛋白质纯化。标签的实例在本文中其他地方公开。此类标签可例如结合到配偶体配体(例如,固定在树脂上),使得标记蛋白可与所有其他蛋白质(例如,宿主细胞蛋白)分开。亲和色谱法、高效液相色谱法(HPLC)和尺寸排阻色谱法(SEC)是可用于改善表达的蛋白质的纯度的方法的实例。

[0137] 其他方法也可用于产生多肽或其片段。例如,两个或更多个肽或多肽可通过蛋白质化学技术连接在一起。例如,肽或多肽可使用Fmoc(9-芴甲氧羰基)或Boc(叔丁氧羰基)化学来化学合成。此类肽或多肽可通过标准化学反应合成。例如,肽或多肽可合成并且不从其合成树脂裂解,而肽或蛋白质的另一个片段可合成并且随后从树脂裂解,从而暴露另一个片段上的被功能性阻断的末端基团。通过肽缩合反应,这两个片段可通过肽键分别在其羧基和氨基末端处共价连接。可替换地,肽或多肽可如本所述在体内独立地合成。一旦分离,这些独立的肽或多肽可通过相似的肽缩合反应连接以形成肽或其片段。

[0138] 在一些实施方案中,克隆或合成肽片段的酶连接允许相对较短的肽片段连接以产生较大肽片段、多肽或全蛋白结构域(Abrahmsen等人,Biochemistry,1991,30,4151)。可替换地,合成肽的天然化学连接可用于从较短的肽片段合成构建较大肽或多肽。此方法可由两步化学反应组成(Dawson等人,Science,1994,266,776-779)。第一步骤可以是去保护合成肽-硫酯与含有氨基末端Cys残基的另一个去保护肽片段的化学选择性反应以得到硫酯连接中间体作为初始共价产物。在反应条件没有变化的情况下,此中间体可经历自发、快速分子内反应以在连接位点处形成天然肽键。

[0139] 在一些实施方案中,去保护肽片段可化学连接,其中由于化学连接而在肽片段之间形成的键是非天然(非肽)键(Schnolzer等人,Science,1992,256,221)。

[0140] 在一些实施方案中,多肽可具有表达后修饰,例如像糖基化、乙酰化和磷酸化以及本领域已知的其他修饰,可以是天然存在的和非天然存在的。多肽可以是完整蛋白质或其序列。

[0141] 本公开还提供产生本文公开的多肽中的任一种的方法,其包括培养包含重组表达载体的宿主细胞,所述宿主细胞包含具有能够编码本文公开的多肽中的一种或多种或其互补的多核苷酸的核酸分子,从而产生多肽。

[0142] 本公开还提供包含核酸分子中的任一种或多种的细胞(例如,重组宿主细胞),其包括包含核酸分子和/或本文公开的多肽中的任一种或多种的载体。所述细胞可以是体外的、离体的或体内的。核酸分子可连接到启动子和其他调控序列,所以它们被表达以产生编码的蛋白质。还提供此类细胞的细胞系。

[0143] 在一些实施方案中,所述细胞是全能细胞或多能细胞(例如,胚胎干(ES)细胞,诸如啮齿动物ES细胞、小鼠ES细胞或大鼠ES细胞)。全能细胞包括可产生任何细胞类型的未分化细胞,并且多能细胞包括具有发育成多于一种分化的细胞类型的能力的未分化细胞。此类多能和/或全能细胞可以是例如ES细胞或类似ES的细胞,诸如诱导的多能干(iPS)细胞。ES细胞包括胚胎衍生的全能或多能细胞,其能够在引入到胚胎之后有利于发育胚胎的任何组织。ES细胞可衍生自胚泡的内细胞团,并且能够分化为三个脊椎动物胚层(内胚层、外胚层和中胚层)中的任一个的细胞。根据本公开,胚胎干细胞可以是非人类胚胎干细胞。

[0144] 在一些实施方案中,所述细胞是原代体细胞或不是原代体细胞的细胞。体细胞可

包括不是配子、生殖细胞、配子体或未分化干细胞的任何细胞。在一些实施方案中，所述细胞也可以是原代细胞。原代细胞包括直接从生物体、器官或组织分离的细胞或细胞的培养物。原代细胞包括既不是转化的也不是无限增殖的细胞。原代细胞包括从生物体、器官或组织获得的先前未在组织培养物中经过或先前已在组织培养物中经过但是不能够无限在组织培养物中经过的任何细胞。此类细胞可通过常规技术分离并且包括例如体细胞、造血细胞、内皮细胞、上皮细胞、成纤维细胞、间充质细胞、角质细胞、黑素细胞、单核细胞(monocyte)、单核细胞(mononuclear cell)、脂肪细胞、前脂肪细胞、神经元、神经胶质细胞、肝细胞、骨骼肌细胞和平滑肌细胞。例如，原代细胞可衍生自结缔组织、肌肉组织、神经系统组织或上皮组织。

[0145] 在一些实施方案中，所述细胞通常可能不是无限增殖的，但是由于突变或改变已避免正常的细胞衰老并且反而可保持进行分裂。此类突变或改变可天然地发生或有意地进行诱导。无限增殖化的细胞的实例包括但不限于中国仓鼠卵巢(CHO)细胞、人类胚胎肾细胞(例如，HEK 293细胞)和小鼠胚胎成纤维细胞(例如，3T3细胞)。多种类型的无限增殖化的细胞是熟知的。无限增殖化或原代细胞包括通常用于培养或用于表达重组基因或蛋白质的细胞。在一些实施方案中，所述细胞是分化细胞，诸如肝细胞(例如，人类肝细胞)。

[0146] 所述细胞可来自任何来源。例如，所述细胞可以是真核细胞、动物细胞、植物细胞或真菌(例如，酵母)细胞。此类细胞可以是鱼类细胞或禽类细胞，或者此类细胞可以是哺乳动物细胞，诸如人类细胞、非人类哺乳动物细胞、啮齿动物细胞、小鼠细胞或大鼠细胞。哺乳动物包括但不限于人类、非人类灵长类动物、猴、猿、猫、狗、马、公牛、鹿、野牛、山羊、啮齿动物(例如，小鼠、大鼠、仓鼠、豚鼠)、家畜(例如，牛物种，诸如奶牛、食用牛等；绵羊物种，诸如绵羊、山羊等；以及猪物种，诸如猪和公猪)。禽类包括但不限于鸡、火鸡、鸵鸟、鹅、鸭等。也包括家养动物和农用动物。术语“非人类动物”排除人类。

[0147] 另外的宿主细胞描述于例如2016年7月28日提交的美国临时申请号62/367,973，所述临时申请以引用的方式整体并入本文。

[0148] 本文公开的核酸分子和多肽可通过任何手段引入到细胞中。转染方案以及用于将核酸或蛋白质引入到细胞中的方案可变化。非限制性转染方法包括使用脂质体、纳米颗粒、钙、树状分子和阳离子聚合物诸如DEAE-葡聚糖或聚乙烯亚胺的基于化学的转染方法。非化学方法包括电穿孔、超音波穿孔和光学转染。基于颗粒的转染包括使用基因枪或磁体辅助的转染。病毒方法也可用于转染。

[0149] 将核酸或蛋白质引入到细胞中也可通过电穿孔、通过胞质内注射、通过病毒转染、通过腺病毒、通过腺相关病毒、通过慢病毒、通过逆转录病毒、通过转染、通过脂质介导的转染或通过核转染来介导。核转染是改进的电穿孔技术，其能够使核酸底物不仅递送到细胞质、并且还通过核膜并递送到核中。另外，在本文公开的方法中使用核转染通常比常规电穿孔需要远远更少的细胞(例如，与通过常规电穿孔需要7百万个细胞相比，核转染仅需要约2百万个细胞)。在一些实施方案中，核转染使用LONZA[®]NUCLEOFECTOR[™]系统来执行。

[0150] 将核酸或蛋白质引入到细胞中也可通过微注射来实现。mRNA的微注射通常是到细胞质中(例如，将mRNA直接递送到翻译机器)，而蛋白质或DNA的微注射通常是到核中。可替换地，微注射可通过注射到核和细胞质两者来实施：首先可将针引入到核中并且可注射第一量，并且在将针从细胞移除时，可将第二量注射到细胞质中。如果将核酸酶剂蛋白质引入

到细胞质中，则所述蛋白质可包含核定位信号以确保递送到核/原核。

[0151] 用于将核酸或蛋白质引入到细胞中的其他方法可包括例如载体递送、颗粒介导的递送、胞外体介导的递送、脂质-纳米颗粒介导的递送、细胞穿透肽介导的递送或可植入装置介导的递送。将核酸或蛋白质施用到受试者以在体内修饰细胞的方法在本文中其他地方公开。将核酸和蛋白质引入到细胞中也可通过流体力学递送(HDD)来实现。

[0152] 用于将核酸或蛋白质引入到细胞中的其他方法可包括例如载体递送、颗粒介导的递送、胞外体介导的递送、脂质-纳米颗粒介导的递送、细胞穿透肽介导的递送或可植入装置介导的递送。在一些实施方案中，核酸或蛋白质可在携带体中引入到细胞中，所述携带体诸如聚(乳酸)(PLA)微球、聚(D,L-乳酸-共乙醇酸)(PLGA)微球、脂质体、胶束、反胶束、脂质螺旋物和脂质微管。

[0153] 本公开还提供探针和引物。探针和引物的实例在以上作为实例公开。本公开提供包含与本文公开的核酸分子中的任一种特异性地杂交的核酸序列的探针和引物。例如，所述探针或引物可包含核酸序列，所述核酸序列与编码在对应于根据SEQ ID NO:4的位置533的位置处包含天冬氨酸的GPR156蛋白的核酸分子杂交或与此核酸分子的补体杂交。在一些实施方案中，本公开的探针或引物包含核酸序列，所述核酸序列与编码根据SEQ ID NO:4的GPR156蛋白的核酸分子杂交或与此核酸分子的补体杂交。在一些实施方案中，所述探针或引物可包含核酸序列，所述核酸序列与编码在对应于根据SEQ ID NO:5的位置529的位置处包含天冬氨酸的GPR156蛋白的核酸分子杂交或与此核酸分子的补体杂交。在一些实施方案中，所述探针或引物包含核酸序列，所述核酸序列与编码根据SEQ ID NO:5的GPR156蛋白的核酸分子杂交或与此核酸分子的补体杂交。所述探针或引物可具有任何合适的长度，所述长度的非限制性实例包括至少约5个、至少约8个、至少约10个、至少约11个、至少约12个、至少约13个、至少约14个、至少约15个、至少约16个、至少约17个、至少约18个、至少约19个、至少约20个、至少约21个、至少约22个、至少约23个、至少约24个或至少约25个核苷酸的长度。在优选的实施方案中，所述探针或引物具有至少约18个核苷酸的长度。所述探针或引物可具有约10个至约35个、约10个至约30个、约10个至约25个、约12个至约30个、约12个至约28个、约12个至约24个、约15个至约30个、约15个至约25个、约18个至约30个、约18个至约25个、约18个至约24个或约18个至约22个核苷酸的长度。在优选的实施方案中，所述探针或引物的长度是约18至约30个核苷酸。

[0154] 本公开还提供等位基因特异性探针和等位基因特异性引物。本公开的等位基因特异性探针或引物与编码在对应于根据SEQ ID NO:4的位置533的位置处包含天冬氨酸的GPR156蛋白的核酸序列或与其补体特异性地杂交。

[0155] 在本公开的上下文中，“特异性地杂交”意指所述探针或引物(例如，等位基因特异性探针或引物)不与编码在对应于根据SEQ ID NO:4的位置533的位置处不包含天冬氨酸的GPR156蛋白的核酸序列杂交。根据本公开的探针或引物优选地涵盖编码GPR156蛋白内的E533D变化n的核酸密码子或其补体。

[0156] 因此，本公开提供一种包含与编码GPR156蛋白的核酸序列互补的核酸序列的等位基因特异性探针或引物，其中所述等位基因特异性探针或引物包含与在对应于根据SEQ ID NO:4的位置533的位置处编码天冬氨酸的核酸密码子互补的核酸序列。

[0157] 因此，根据本公开的等位基因特异性探针或引物包含核酸序列，所述核酸序列与

编码在根据SEQ ID NO:4的位置533处包含天冬氨酸的GPR156蛋白的核酸序列互补。根据本公开的等位基因特异性探针或引物可用于检测GPR156内的E533D变化和/或GPR156的较短同种型内的E529D突变。例如，根据本公开的等位基因特异性引物可用于扩增包含E533D变化的GPR156(例如，具有SEQ ID NO:4的氨基酸序列)或其片段的核酸序列。根据本公开的等位基因特异性引物还可用于扩增包含E529D突变的GPR156同种型(诸如具有SEQ ID NO:5的氨基酸序列的同种型)或其片段的核酸序列。

[0158] 在一些实施方案中，本公开的等位基因特异性探针或引物包含核酸序列，所述核酸序列与编码根据SEQ ID NO:4的GPR156蛋白的核酸序列互补。在一些实施方案中，所述等位基因特异性探针或引物包含核酸序列，所述核酸序列与编码根据SEQ ID NO:5的GPR156蛋白的核酸序列互补。在一些实施方案中，所述等位基因特异性探针或引物包含核酸序列，所述核酸序列与编码在对应于根据SEQ ID NO:4的位置533或根据SEQ ID NO:5的位置529的位置处包含天冬氨酸的GPR156蛋白的核酸序列互补。在一些实施方案中，所述等位基因特异性引物与本文公开的核酸分子中的任一种特异性地杂交，其中所述天冬氨酸由密码子GAT或GAU编码。在一些实施方案中，所述等位基因特异性引物与本文公开的核酸分子中的任一种特异性地杂交，其中所述天冬氨酸由密码子GAC编码。编码天冬氨酸的密码子可在所述等位基因特异性引物的3'末端处。

[0159] 在优选的实施方案中，所述等位基因特异性探针或引物结合到GPR156基因的主要等位基因。因此，所述等位基因特异性引物优选地结合到GPR156基因的主要等位基因并且不结合到GPR156基因的次要等位基因。所述主要等位基因包含在对应于SEQ ID NO:4的位置516的位置处编码天冬氨酸的核酸序列，虽然野生型主要等位基因包含在对应于SEQ ID NO:4的位置533的位置处编码谷氨酸的核酸序列。所述次要等位基因包含在对应于SEQ ID NO:4的位置516的位置处编码谷氨酸的核酸序列，虽然所述次要等位基因包含在对应于SEQ ID NO:4的位置533的位置处编码天冬氨酸的核酸序列。因此，所述等位基因特异性探针或引物可与包含E533D变化的GPR156基因的次要等位基因特异性地杂交。根据本公开，所述等位基因特异性探针或引物不与不包含E533D变化的主要等位基因杂交。

[0160] 加以必要的变更，关于本公开的探针或引物以上所述的长度也适用于本公开的等位基因特异性探针或引物。

[0161] 本公开还提供包含如上所述的等位基因特异性引物中的一种的一对等位基因特异性引物。

[0162] 在一些实施方案中，所述探针或引物(例如，等位基因特异性探针或引物)包括DNA。在一些实施方案中，所述探针或引物(例如，等位基因特异性探针或引物)包括RNA。在一些实施方案中，所述探针(例如，等位基因特异性探针)与本文公开的核酸分子中的任一种杂交，其中所述天冬氨酸(在对应于根据SEQ ID NO:4的位置533的位置处)由密码子GAT(或GAU)编码。在一些实施方案中，所述探针与本文公开的核酸分子中的任一种杂交，其中所述天冬氨酸(在对应于根据SEQ ID NO:4的位置533的位置处)由密码子GAC编码。在一些实施方案中，所述探针(例如，等位基因特异性探针)在严格条件(诸如高严格条件)下与编码人类GPR156蛋白的核酸序列杂交。

[0163] 在一些实施方案中，所述探针包含标记。在一些实施方案中，所述标记是荧光标记、放射性标记或生物素。在一些实施方案中，所述探针的长度在以上描述。可替换地，在一

些实施方案中，所述探针包含至少约20个、至少约25个、至少约30个、至少约35个、至少约40个、至少约45个、至少约50个、至少约55个、至少约60个、至少约65个、至少约70个、至少约75个、至少约80个、至少约85个、至少约90个、至少约95个或至少约100个核苷酸或者由其组成。所述探针(例如，等位基因特异性探针)可例如用于检测本文公开的核酸分子中的任一种。在优选的实施方案中，所述探针具有至少约18个核苷酸的长度。所述探针可具有约10个至约35个、约10个至约30个、约10个至约25个、约12个至约30个、约12个至约28个、约12个至约24个、约15个至约30个、约15个至约25个、约18个至约30个、约18个至约25个、约18个至约24个或约18个至约22个核苷酸的长度。在优选的实施方案中，所述探针的长度是约18个至约30个核苷酸。

[0164] 本公开还提供包含本文公开的探针中的任一种或多种所附接的底物的支持物。固体支持物是分子(诸如本文公开的探针中的任一种)可与其缔合的固态底物或支持物。固体支持物的一种形式是阵列。固体支持物的另一种形式是阵列检测物。阵列检测物是多种不同的探针以阵列、网格或其他组织化模式与其偶联的固体支持物。

[0165] 用于固体支持物中的固态底物可包括分子可与其偶联的任何固体材料。这包括诸如以下的材料：丙烯酰胺、琼脂糖、纤维素、硝化纤维素、玻璃、聚苯乙烯、聚乙烯乙酸乙烯酯、聚丙烯、聚甲基丙烯酸酯、聚乙烯、聚环氧乙烷、聚硅酸酯、聚碳酸酯、聚四氟乙烯、碳氟化合物、尼龙、硅橡胶、聚酐、聚乙醇酸、聚乳酸、聚原酸酯、聚丙基富马酸酯、胶原、糖胺聚糖和聚氨基酸。固态底物可具有任何可用的形式，包括薄膜、膜、瓶、皿、纤维、编织纤维、成形聚合物、颗粒、珠、微颗粒或组合。固态底物和固体支持物可以是多孔的或无孔的。固态底物的一种形式是微量滴定皿，诸如标准96孔类型。在一些实施方案中，可采用通常具有一个阵列/孔的多孔玻璃载片。此特征允许更大地控制测定再现性、增加通量和样品处理并且易于自动化。在一些实施方案中，所述支持物是微阵列。

[0166] 本文公开的多肽中的任一种还可具有除如本文所述的在位置533或529处的天冬氨酸以外的一个或多个取代(诸如保守氨基酸取代)、插入或缺失。插入包括例如氨基或羧基末端融合以及单个或多个氨基酸残基的序列内插入。用于在具有已知序列的DNA中在预先确定的位点处进行取代的技术是熟知的，例如M13引物诱变和PCR诱变。氨基酸取代通常是单个残基，但是也可同时在多个不同位置处发生；插入通常以1至10个氨基酸残基的数量级计；并且缺失的范围是约1至30个残基。缺失或插入可在相邻的对中进行，即2个残基缺失或2个残基插入。取代、缺失、插入或它们的任何组合可进行组合以得到最终的构建体。在一些实施方案中，突变不将序列置于阅读框外，并且不创建可产生二级mRNA结构的互补区。

[0167] 本公开还提供用于制备组合物并利用本文所述的方法的试剂盒。本文所述的试剂盒可包括用于检测受试者的样品中的一个或多个遗传变体的一个或多个测定。

[0168] 在一些实施方案中，用于GPR156变体的人类鉴定的试剂盒利用以上所述的组合物和方法。在一些实施方案中，基本试剂盒可包括具有至少一对针对本文公开的核酸分子中的任一种(例如像，SEQ ID N0:9至12, SEQ ID N0:16至21或SEQ ID N0:25至30)中的基因座的寡核苷酸的容器。试剂盒还可任选地包括用于使用的说明书。试剂盒还可包括其他任选的试剂盒组件，例如像以下中一个或多个：针对扩增的基因座中的每个的等位基因阶梯；用于扩增的足量的酶；用于有助于扩增的扩增缓冲液；用于有助于酶活性的二阶阳离子溶液；用于扩增期间的链延伸的dNTP；用于制备用于电泳的扩增材料的装载溶液；作为模板对照

的基因组DNA;用于确保材料在分离介质中如所预期的进行迁移的大小标记物;以及教导使用者并限制使用错误的方案和手册。试剂盒中的各种试剂的量也可根据多种因素(诸如方法的最佳灵敏度)而变化。在这些教义的范围内的是,提供用于手动应用的测试试剂盒或用于与自动化样品制备、反应设置、检测物或分析仪一起使用的测试试剂盒。

[0169] 在一些实施方案中,试剂盒包括至少一对用于编码天冬氨酸的至少一个基因座的扩增的寡核苷酸引物,其中所述对的一个引物与至少一个基因座的引物结合位点内的SNP核碱基杂交,并且所述对的至少一个引物包含荧光标记。在一些方面,所述基因座包含GPR156基因中的变体。

[0170] 在一些实施方案中,所述试剂盒包括至少一对用于选自本文公开的核酸分子中的任一种(例如像,SEQ ID N0:9至12,SEQ ID N0:16至21或SEQ ID N0:25至30)的至少一个基因座的PCR扩增的寡核苷酸引物,其中所述对的一个引物与至少一个基因座的引物结合位点内的SNP核碱基杂交。在一些实施方案中,试剂盒还包括对应于选自本文公开的核酸分子中的任一种(例如像,SEQ ID N0:9至12,SEQ ID N0:16至21或SEQ ID N0:25至30)的至少一个基因座的等位基因阶梯、方案、酶、dNTP、缓冲液、一种或多种盐以及对照核酸样品中的至少一种。在一些实施方案中,SNP核碱基在选自本文公开的核酸分子中的任一种(例如像,SEQ ID N0:9至12,SEQ ID N0:16至21或SEQ ID N0:25至30)的序列内。

[0171] 在一些方面,本文公开的试剂盒可包括包含3'末端核苷酸的引物,所述引物直接与在本文公开的cDNA或mRNA分子中的任一种(例如像,SEQ ID N0:16至21或SEQ ID N0:25至30)的核苷酸位置1599处的胸腺嘧啶或尿嘧啶杂交。

[0172] 本文还公开试剂盒,其包括:至少一种核酸探针,其被设计来检测本文公开的核酸分子中的任一种中的在对应于SEQ ID N0:4的位置533的位置处编码天冬氨酸的位置处的核苷酸变异,其中检测是基于与核苷酸变异序列的特异性杂交,其中所述核酸探针包含可检测标记、产物和实施退火反应所需的试剂;以及说明书。试剂盒可包括本文所述的探针中的任一种。

[0173] 本文公开试剂盒,其包括至少一对寡核苷酸引物,所述至少一对寡核苷酸引物用于本文公开的核酸分子中的任一种的在对应于SEQ ID N0:4的位置533的位置处编码天冬氨酸的位置处的至少一个基因座的扩增,其中所述对的一个引物与至少一个基因座的引物结合位点内的SNP核碱基杂交,并且所述对的至少一个引物包含荧光标记。试剂盒可包括本文所述的引物中的任一种。

[0174] 本文公开试剂盒,其包括至少一对寡核苷酸引物,所述至少一对寡核苷酸引物用于本文公开的核酸分子中的任一种的在对应于SEQ ID N0:4的位置533的位置处编码天冬氨酸的位置处的至少一个基因座的扩增,其中所述对的一个引物与至少一个基因座的引物结合位点内的SNP核碱基杂交,并且所述对的至少一个引物包含荧光标记;其还包括等位基因阶梯,所述等位基因阶梯对应于本文公开的核酸分子中的任一种的在对应于SEQ ID N0:4的位置533的编码天冬氨酸的位置处的至少一个基因座。试剂盒可包括本文所述的引物中的任一种。

[0175] 本文公开试剂盒,其包括至少一对寡核苷酸引物,所述至少一对寡核苷酸引物用于本文公开的核酸分子中的任一种的在对应于SEQ ID N0:4的位置533的编码天冬氨酸的位置处的至少一个基因座的扩增,其中所述对的一个引物与至少一个基因座的引物结合位

点内的SNP核碱基杂交，并且所述对的至少一个引物包含荧光标记，其中所述引物用于聚合酶链反应(PCR)。试剂盒可包括本文所述的引物中的任一种。

[0176] 本文公开试剂盒，其包括至少一对寡核苷酸引物，所述至少一对寡核苷酸引物用于本文公开的核酸分子中的任一种的在对应于SEQ ID NO:4的位置533的位置处编码天冬氨酸的位置处的至少一个基因座的扩增，其中所述对的一个引物与至少一个基因座的引物结合位点内的SNP核碱基杂交，并且所述对的至少一个引物包含荧光标记；其还包括方案、酶、dNTP、缓冲液、一种或多种盐以及对照核酸样品中的至少一种。试剂盒可包括本文所述的引物中的任一种。

[0177] 表1.引物

	正向引物(5'→3')	SEQ NO:	ID	反向引物(5'→3')	SEQ ID NO
[0178]	TCTCCCCAG		31		
	AGAGGGGTC		32		
	TCTCCCCAT		33		
	AGAGGGGA		34		
	CCAGAGTCTCCTACCA		35	AGAAGGCTGGCGAAG	36
	GATAG			AAAGG	
	CAACGCCAAAGAGAA		37	CTTCCAGGTTGCTGA	38
	GATTG			TGACA	
	AAGATATCTGACTCAA		39	AGAGCCAGGAAAGTG	40
	AAGAC			TGAGC	
	TTTCAGAATGATCCTG		41	ACTGTTCTTCCAGGTT	42
	GCATG			TGCTG	

[0179] 本领域的技术人员应理解，所采用的检测技术通常是非限制性的。相反，各种各样的检测手段在公开的方法和试剂盒的范围内，前提是它们允许待确定的扩增子的存在或不存在。

[0180] 在一些方面，试剂盒可包括本文公开的引物或探针中的一种或多种。例如，试剂盒可包括与公开的遗传变体中的一种或多种杂交的一种或多种探针。

[0181] 在一些方面，试剂盒可包括公开的细胞或细胞系中的一种。在一些方面，试剂盒可包括产生转基因细胞或细胞系所必需的材料。例如，在一些方面，试剂盒可包括细胞和载体，其包含具有公开的遗传变体中的一种或多种的核酸序列。试剂盒还可包括用于细胞培养的培养基。

[0182] 本公开还提供用于检测来自受试者人类的生物样品中的GPR156变体基因、mRNA、cDNA和/或多肽的存在的方法。应理解，群体内的基因序列和由此类基因编码的mRNA和蛋白质可由于多态性(诸如单核苷酸多态性)而变化。本文提供的GPR156基因、mRNA、cDNA和多肽的序列仅是示例性序列。GPR156基因、mRNA、cDNA和多肽的其他序列也是可能的。

[0183] 生物样品可衍生自来自受试者的任何细胞、组织或生物流体。样品可包括任何临上相关的组织诸如骨髓样品、肿瘤活检样品、细针抽吸物，或体液的样品诸如血液、龈沟液、血浆、血清、淋巴、腹水、囊液或尿液。在一些情况下，所述样品包括口腔拭子。用于本文公开的方法的样品根据测定形式、检测方法的性质以及用作样品的组织、细胞或提取物而变化。生物样品可根据所采用的测定进行不同处理。例如，当检测变体GPR156核酸分子时，可采用被设计来分离或富集基因组DNA的样品的初步处理。各种已知的技术可用于此目的。

当检测变体GPR156 mRNA的水平时,不同的技术可用于富集具有mRNA的生物样品。可使用检测mRNA的存在或水平或特定变体基因组DNA基因座的存在的各种方法。

[0184] 在一些实施方案中,本公开提供检测变体GPR156核酸分子的存在或不存在的方法,其包括对生物样品中的核酸的至少一部分进行测序,以确定所述核酸是否包含在对应于根据SEQ ID NO:4的位置533的位置处或在对应于根据SEQ ID NO:5的位置529的位置处编码天冬氨酸的核酸序列。使用本文公开的探针中的任一种可检测本文公开的变体核酸分子中的任一种。

[0185] 在一些实施方案中,检测受试者中的单极抑郁症相关变体GPR156核酸分子(例如,基因、mRNA或cDNA)的存在或不存在的方法包括:在从受试者获得的生物样品上执行测定,所述测定确定生物样品中的核酸分子是否包含在对应于根据SEQ ID NO:4的位置533的位置处或在对应于根据SEQ ID NO:5的位置529的位置处编码天冬氨酸的核酸序列。在一些实施方案中,生物样品包括细胞或细胞裂解物。此类方法还可包括例如从受试者获得包含GPR156基因或mRNA的生物样品,并且如果是mRNA,则任选地将mRNA逆转录为cDNA,并且在生物样品上执行测定,所述测定确定GPR156基因、mRNA或cDNA的一部分是否在对应于变体GPR156多肽的位置533的位置处编码天冬氨酸。此类测定可包括例如确定特定GPR156核酸分子的这些位置的同一性。在一些实施方案中,受试者是人类。

[0186] 在一些实施方案中,检测受试者中的焦虑病症相关变体GPR156核酸分子(例如,基因、mRNA或cDNA)的存在或不存在的方法包括:在从受试者获得的生物样品上执行测定,所述测定确定生物样品中的核酸分子是否包含在对应于根据SEQ ID NO:4的位置533的位置处或在对应于根据SEQ ID NO:5的位置529的位置处编码天冬氨酸的核酸序列。在一些实施方案中,生物样品包括细胞或细胞裂解物。此类方法还可包括例如从受试者获得包含GPR156基因或mRNA的生物样品,并且如果是mRNA,则任选地将mRNA逆转录为cDNA,并且在生物样品上执行测定,所述测定确定GPR156基因、mRNA或cDNA的一部分是否在对应于变体GPR156多肽的位置533的位置处编码天冬氨酸。此类测定可包括例如确定特定GPR156核酸分子的这些位置的同一性。在一些实施方案中,受试者是人类。

[0187] 在一些实施方案中,所述测定包括:对来自受试者的生物样品中的核酸分子的GPR156基因组序列的至少一部分进行测序,其中测序的部分包括对应于在根据SEQ ID NO:4的GPR156蛋白的位置533处或在对应于根据SEQ ID NO:5的位置529的位置处编码天冬氨酸的位置的位置;对来自受试者的生物样品中的核酸分子的GPR156mRNA序列的一部分进行测序,其中测序的部分包括对应于在根据SEQ ID NO:4的GPR156蛋白的位置533处或在对应于根据SEQ ID NO:5的位置529的位置处编码天冬氨酸的位置的位置;或者对从mRNA(从受试者获得的生物样品获得)获得的核酸分子的GPR156 cDNA序列的一部分进行测序,其中测序的部分包括对应于在根据SEQ ID NO:4的GPR156蛋白的位置533处或在对应于根据SEQ ID NO:5的位置529的位置处编码天冬氨酸的位置的位置。

[0188] 在一些实施方案中,所述测定包括:a)将生物样品与引物接触,所述引物与以下杂交:i) GPR156基因组序列的一部分,所述部分接近在对应于根据SEQ ID NO:4的位置533处或对应于根据SEQ ID NO:5的位置529的位置处编码天冬氨酸的位置的位置处的GPR156基因组序列的位置;ii) GPR156 mRNA序列的一部分,所述部分接近对应于在根据SEQ ID NO:4的位置533处或在对应于根据SEQ ID NO:5的位置529的位置处编码天冬氨酸的位置的

GPR156 mRNA的位置;或iii) GPR156 cDNA序列的一部分,所述部分接近对应于在根据SEQ ID NO:4的位置533处或在对应于根据SEQ ID NO:5的位置529的位置处编码天冬氨酸的位置的GPR156 cDNA的位置;b) 将引物延伸至少通过:i) GPR156基因组序列的位置,所述位置对应于超过在根据SEQ ID NO:4的位置533处或在对应于根据SEQ ID NO:5的位置529的位置处编码天冬氨酸的密码子的核苷酸位置;ii) GPR156 mRNA的位置,所述位置对应于超过在对应于根据SEQ ID NO:4的位置533处的位置处或在对应于根据SEQ ID NO:5的位置529的位置处编码天冬氨酸的密码子的核苷酸位置;或iii) GPR156 cDNA的位置,所述位置对应于超过在对应于根据SEQ ID NO:4的位置533处的位置处或在对应于根据SEQ ID NO:5的位置529的位置处编码天冬氨酸的密码子的核苷酸位置;以及c) 确定引物的延伸产物是否包含在对应于根据SEQ ID NO:4的位置533的位置处或在对应于根据SEQ ID NO:5的位置529的位置处编码天冬氨酸的核苷酸。在一些实施方案中,仅分析GPR156基因组DNA。在一些实施方案中,仅分析GPR156mRNA。在一些实施方案中,仅分析从GPR156 mRNA获得的GPR156 cDNA。

[0189] 在一些实施方案中,所述测定包括:a)将生物样品与等位基因特异性引物接触,所述等位基因特异性引物与以下杂交:i) GPR156基因组序列的一部分,所述部分包含在对应于根据SEQ ID NO:4的位置533的位置处或在对应于根据SEQ ID NO:5的位置529的位置处编码氨基酸的核苷酸;ii) GPR156 mRNA序列的一部分,所述部分包含在对应于根据SEQ ID NO:4的位置533的位置处或在对应于根据SEQ ID NO:5的位置529的位置处编码氨基酸的核苷酸;或iii) GPR156 cDNA序列的一部分,所述部分包含在对应于根据SEQ ID NO:4的位置533的位置处或在对应于根据SEQ ID NO:5的位置529的位置处编码氨基酸的核苷酸;b) 使用等位基因特异性聚合酶链反应将引物延伸;以及c) 确定延伸是否发生。等位基因特异性聚合酶链反应技术可用于检测核酸序列中的突变,诸如SNP。因为当与模板存在适配时,DNA聚合酶将不延伸,所以使用等位基因特异性引物。基本的等位基因特异性聚合酶链反应技术的多种变化由技术人员决定。

[0190] 所述等位基因特异性引物可包含核酸序列或核酸序列的补体,所述核酸序列与编码在对应于根据SEQ ID NO:4的位置533的位置处或在对应于根据SEQ ID NO:5的位置529的位置处包含天冬氨酸的GPR156蛋白的核酸序列互补。例如,所述等位基因特异性引物可包含核酸序列或此核酸序列的补体,所述核酸序列与编码SEQ ID NO:4和/或SEQ ID NO:5的核酸序列互补。当核酸序列在对应于根据SEQ ID NO:4的位置533的位置处或在对应于根据SEQ ID NO:5的位置529的位置处通过密码子GAT(或GAU)或GAC编码天冬氨酸时,所述等位基因特异性引物优选地与编码GPR156蛋白的核酸序列特异性地杂交。

[0191] 在一些实施方案中,所述测定包括将生物样品与引物或探针接触,所述引物或探针在严格条件下与变体GPR156基因组序列、mRNA序列或cDNA序列特异性地杂交并且不与对应的野生型GPR156序列杂交;以及确定杂交是否发生。

[0192] 在一些实施方案中,所述测定包括RNA测序(RNA-Seq)。在一些实施方案中,所述测定还包括通过逆转录酶聚合酶链反应(RT-PCR)将mRNA逆转录为cDNA。

[0193] 在一些实施方案中,所述方法利用足够核苷酸长度的探针和引物来结合到靶核酸序列,并且特异性地检测和/或鉴定包含变体GPR156基因、mRNA或cDNA的多核苷酸。杂交条件或反应条件可由操作员决定以实现此结果。此核苷酸长度可以是足以用于所选择的检测

方法(包括本文描述或例示的任何测定)的任何长度。通常,例如,使用具有以下长度的引物或探针:约8个、约10个、约11个、约12个、约14个、约15个、约16个、约18个、约20个、约22个、约24个、约26个、约28个、约30个、约40个、约50个、约75个、约100个、约200个、约300个、约400个、约500个、约600个或约700个核苷酸或更多,或约11个至约20个、约20个至约30个、约30个至约40个、约40个至约50个、约50个至约100个、约100个至约200个、约200个至约300个、约300个至约400个、约400个至约500个、约500个至约600个、约600个至约700个或约700个至约800个或更多个核苷酸。在优选的实施方案中,所述探针或引物具有至少约18个核苷酸的长度。所述探针或引物可具有约10个至约35个、约10个至约30个、约10个至约25个、约12个至约30个、约12个至约28个、约12个至约24个、约15个至约30个、约15个至约25个、约18个至约30个、约18个至约25个、约18个至约24个或约18个至约22个核苷酸的长度。在优选的实施方案中,所述探针或引物的长度是约18个至约30个核苷酸。

[0194] 此类探针和引物可在高严格杂交条件下与靶序列特异性地杂交。探针和引物可与靶序列具有连续核苷酸的完全核酸序列同一性,虽然可通过常规方法设计与靶核酸序列不同并保留特异性地检测和/或鉴定靶核酸序列的能力的探针。因此,探针和引物可与靶核酸分子共享约80%、约85%、约90%、约91%、约92%、约93%、约94%、约95%、约96%、约97%、约98%、约99%或100%序列同一性或互补性。

[0195] 在一些实施方案中,特异性引物可用于扩增变体GPR156基因座和/或GPR156变体mRNA或cDNA以产生扩增子,所述扩增子可用作特异性探针或者自身可被检测以用于鉴定变体GPR156基因座或用于确定生物样品中的特异性GPR156 mRNA或cDNA的水平。GPR156变体基因座可用于指示基因组核酸序列,所述基因组核酸序列包含对应于在根据SEQ ID NO:4的位置533处或在根据SEQ ID NO:5的位置529处编码天冬氨酸的位置的位置。当探针在允许探针与核酸分子结合的条件下与生物样品中的核酸分子杂交时,此结合被检测并允许指示生物样品中变体GPR156基因座的存在或变体GPR156 mRNA或cDNA的存在或水平。已描述结合探针的这种鉴定。特异性探针可包含与变体GPR156基因的特定区域至少约80%、约80%至约85%、约85%至约90%、约90%至约95%以及约95%至约100%相同(或互补)的序列。特异性探针可包含与变体GPR156 mRNA的特定区域至少约80%、约80%至约85%、约85%至约90%、约90%至约95%以及约95%至约100%相同(或互补)的序列。特异性探针可包含与变体GPR156 cDNA的特定区域至少约80%、约80%至约85%、约85%至约90%、约90%至约95%以及约95%至约100%相同(或互补)的序列。

[0196] 在一些实施方案中,为了确定生物样品的核酸补体是否包含在对应于根据SEQ ID NO:4的位置533的位置处或在对应于根据SEQ ID NO:5的位置529的位置处编码天冬氨酸的核酸序列,可将生物样品经受使用引物对的核酸扩增方法以产生诊断在对应于根据SEQ ID NO:4的位置533的位置处或在对应于根据SEQ ID NO:5的位置529的位置处编码天冬氨酸的位置处SNP的存在,所述引物对包含:第一引物,其衍生自与在对应于根据SEQ ID NO:4的位置533的位置处或在对应于根据SEQ ID NO:5的位置529的位置处编码天冬氨酸的位置相邻的5'侧接序列;以及第二引物,其衍生自与在对应于根据SEQ ID NO:4的位置533的位置处或在对应于根据SEQ ID NO:5的位置529的位置处编码天冬氨酸的位置相邻的3'侧接序列。在一些实施方案中,扩增子的长度范围可以是引物对加上一个核苷酸碱基对的组合长度至可通过DNA扩增方案产生的扩增子的任何长度。此距离的范围可以是一个核苷酸碱基对至

扩增反应的极限值或约两万个核苷酸碱基对。任选地，引物对侧接包含在根据SEQ ID NO:4的位置533处或在根据SEQ ID NO:5的位置529处编码天冬氨酸的位置以及在根据SEQ ID NO:4的位置533处或在根据SEQ ID NO:5的位置529处编码天冬氨酸的位置的每侧上的至少1、2、3、4、5、6、7、8、9、10个或更多个核苷酸的区域。相似的扩增子可由mRNA和/或cDNA序列生成。

[0197] 用于制备和使用探针和引物的代表性方法描述于例如Molecular Cloning:A Laboratory Manual,第2版,第1-3卷,Sambrook等人编,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.1989(下文中为“Sambrook等人,1989”);Current Protocols in Molecular Biology,Ausubel等人编,Greene Publishing and Wiley-Interscience,New York,1992(定期更新)(下文中为“Ausubel等人,1992”);以及Innis等人,PCR Protocols:A Guide to Methods and Applications,Academic Press:San Diego,1990)。PCR引物对可例如通过使用意图用于此目的的计算机程序衍生自己知的序列,所述计算机程序诸如Vector NTI版本10的PCR引物分析工具(Informax公司,Bethesda Md.) ;PrimerSelect(DNASTAR公司,Madison,Wis.) ;以及Primer3(版本0.4.0.COPYRGT.,1991,Whitehead Institute for Biomedical Research,Cambridge,Mass.)。另外,所述序列可以视觉的方式扫描,并且引物可使用已知的指导手动鉴定。

[0198] 可使用任何核酸杂交或扩增或测序方法来特异性地检测变体GPR156基因基因座的存在和/或由mRNA产生的变体GPR156 mRNA或cDNA的水平。在一些实施方案中,核酸分子可用作引物以扩增GPR156核酸的区域,或者核酸分子可用作例如在严格条件下与包含变体GPR156基因基因座的核酸分子或包含由mRNA产生的变体GPR156 mRNA或cDNA的核酸分子特异性地杂交的探针。

[0199] 各种技术在本领域中是可获得的,包括例如核酸测序、核酸杂交和核酸扩增。核酸测序技术的示例性实例包括但不限于链终止子(Sanger)测序和染料终止子测序。

[0200] 其他方法包括除了测序以外的核酸杂交方法,包括使用针对纯化的DNA、扩增的DNA和固定细胞制品的标记的引物或探针(荧光原位杂交(FISH))。在一些方法中,靶核酸可在检测之前或与检测同时进行扩增。核酸扩增技术的示例性实例包括但不限于聚合酶链反应(PCR)、连接酶链反应(LCR)、链置换扩增反应(SDA)以及基于核酸序列的扩增反应(NASBA)。其他方法包括但不限于连接酶链反应、链置换扩增反应和嗜热SDA(tSDA)。

[0201] 任何方法可用于检测非扩增或扩增的多核苷酸,包括例如杂交保护测定(HPA)、扩增方法的实时定量评估以及确定样品中初始存在的靶序列的量,但是其并不基于实时扩增。

[0202] 还提供用于鉴定核酸的方法,其不一定需要序列扩增并且基于例如染色体材料的Southern(DNA:DNA)印迹杂交、原位杂交(ISH)和荧光原位杂交(FISH)的已知方法。Southern印迹可用于检测特定核酸序列。在此类方法中,将从样品提取的核酸片段化,在基质凝胶上电泳分离并转移到膜过滤器。将过滤器结合的核酸经受与感兴趣的序列互补的标记的探针进行的杂交。检测结合到过滤器的杂交的探针。在任何此类方法中,所述方法可包括使用本文描述或示例的探针中的任一种进行的杂交。

[0203] 在杂交技术中,可采用严格条件,使得探针或引物特异性地与其靶标杂交。在一些实施方案中,在严格条件下的多核苷酸引物或探针将与其靶序列(例如,变体(例如,E533D

或E529D)GPR156基因基因座、变体(例如,E533D或E529D)GPR156 mRNA或变体(例如,E533D或E529D)GPR156 cDNA)杂交,其程度可检测地大于与其他序列(例如,对应的野生型GPR156基因座、野生型mRNA或野生型cDNA)的杂交,诸如是背景的至少2倍、至少3倍、至少4倍或更多倍,包括是背景的超过10倍。在一些实施方案中,在严格条件下的多核苷酸引物或探针将与其靶序列杂交,其程度可检测地比与其他序列的杂交大至少2倍。在一些实施方案中,在严格条件下的多核苷酸引物或探针将与其靶序列杂交,其程度可检测地比与其他序列的杂交大至少3倍。在一些实施方案中,在严格条件下的多核苷酸引物或探针将与其靶序列杂交,其程度可检测地比与其他序列的杂交大至少4倍。在一些实施方案中,在严格条件下的多核苷酸引物或探针将与其靶序列杂交,其程度可检测地比与其他序列的杂交大相对于背景超过10倍。严格的条件是序列依赖性的并且在不同的环境中将是不同的。通过控制杂交和/或洗涤条件的严格性,可鉴别与探针100%互补的靶序列(同源探测)。可替换地,可调整严格性条件以允许序列中的一些失配,使得检测到较低程度的同一性(异源探测)。

[0204] 促进DNA杂交的适当严格性条件(例如,6X氯化钠/柠檬酸钠(SSC),在约45°C下,接着在50°C下进行2X SSC的洗涤)是已知的,或者可见于Current Protocols in Molecular Biology, John Wiley&Sons, N.Y. (1989), 6.3.1-6.3.6。通常,用于杂交和检测的严格条件将是其中如下所述的那些:在pH 7.0至8.3时的盐浓度低于约1.5M Na离子、通常约0.01至1.0M Na离子浓度(或其他盐),并且温度对于短探针(例如,10至50个核苷酸)而言是至少约30°C,并且对于较长探针(例如,大于50个核苷酸)而言是至少约60°C。还可通过添加去稳定剂如甲酰胺来实现严格条件。示例性低严格性条件包括在37°C下用30%至35%甲酰胺、1M NaCl、1% SDS(十二烷基硫酸钠)的缓冲溶液进行杂交,并且在50°C至55°C下在1X至2X SSC(20X SSC=3.0M NaCl/0.3M柠檬酸三钠)中进行洗涤。示例性中等严格性条件包括在37°C下在40%至45%甲酰胺、1.0M NaCl、1% SDS中进行杂交,并且在55°C至60°C下在0.5X至1X SSC中进行洗涤。示例性高严格性条件包括在37°C下在50%甲酰胺、1M NaCl、1% SDS中进行杂交,并且在60°C至65°C下在0.1X SSC中进行洗涤。任选地,洗涤缓冲液可包含约0.1%至约1% SDS。杂交的持续时间通常少于约24小时,通常约4至约12小时。洗涤时间的持续时间将至少是足以达到平衡的时间长度。

[0205] 在杂交反应中,特异性通常取决于杂交后洗涤、离子强度的关键因素和最终洗涤溶液的温度。对于DNA-DNA杂交体, T_m 可由Meinkoth和Wahl, Anal. Biochem., 1984, 138, 267-284的等式粗略估计: $T_m = 81.5^\circ\text{C} + 16.6(\log M) + 0.41(\% \text{GC}) - 0.61(\% \text{form}) - 500/L$;其中M是单价阳离子的摩尔浓度,%GC是DNA中的鸟苷和胞嘧啶核苷酸的百分比,%form是杂交溶液中的甲酰胺的百分比,并且L是碱基对中的杂交体的长度。 T_m 是50%的互补靶序列与完全匹配探针杂交时所处的温度(在确定离子强度和pH下)。对于每1%的失配, T_m 降低约1°C;因此,可调整 T_m 、杂交和/或洗涤条件以与所需同一性的序列杂交。例如,如果寻求具有≥90%同一性的序列,则 T_m 可降低10°C。通常,在确定的离子强度和pH下,针对特定序列及其补体,严格条件选择为比热解链温度(T_m)低约5°C。然而,非常严格的条件可利用在比热解链温度(T_m)低1°C、2°C、3°C或4°C下进行的杂交和/或洗涤;中等严格的条件可利用在比热解链温度(T_m)低6°C、7°C、8°C、9°C或10°C下进行的杂交和/或洗涤;低严格条件可利用在比热解链温度(T_m)低11°C、12°C、13°C、14°C、15°C或20°C下进行的杂交和/或洗涤。使用所述等式、杂交和洗涤组合物以及所需的 T_m ,本领域的普通技术人员应理解,本质上描述杂交和/或洗涤

溶液的严格性的变化。如果所需的失配程度引起小于45°C(水溶液)或32°C(甲酰胺溶液)的T_m,则最佳的是,增加SSC浓度,使得可使用较高的温度。

[0206] 还提供用于检测生物样品中的变体GPR156多肽的存在或定量其水平的方法,其包括例如蛋白质测序和免疫测定。在一些实施方案中,检测人类受试者中的GPR156Glu533Asp(例如,SEQ ID NO:4或SEQ ID NO:5)存在的方法包括在来自人类受试者的生物样品上执行确定生物样品中的GPR156Glu533Asp(例如,SEQ ID NO:4或SEQ ID NO:5)存在的测定。

[0207] 蛋白质测序技术的例示性非限制性实例包括但不限于质谱和Edman降解。免疫测定的例示性实例包括但不限于免疫沉淀、蛋白质印记、免疫组织化学、ELISA、免疫细胞化学、流式细胞术以及免疫-PCR。使用各种已知的技术(例如,量热、荧光、化学发光或放射性)可检测地标记的多克隆或单克隆抗体适用于免疫测定。

[0208] 本公开还提供用于在人类受试者中诊断单极抑郁症或检测单极抑郁症的风险的方法,所述方法包括:检测从人类受试者获得的编码GPR156蛋白的核酸分子中的改变,其中所述改变在对应于根据SEQ ID NO:4的GPR156蛋白中的位置533的位置处或在对应于根据SEQ ID NO:5的位置529的位置处编码天冬氨酸;以及如果所述受试者具有一种或多种抑郁症症状(包括一种或多种单极抑郁症症状),则将人类受试者诊断为具有单极抑郁症,或者如果所述受试者不具有一种或多种抑郁症症状,则将人类受试者诊断为具有单极抑郁症的风险。在一些实施方案中,所述人类受试者需要这种诊断。在一些实施方案中,所述人类受试者可能具有已诊断为具有单极抑郁症的亲属。所述改变可包括在对应于根据SEQ ID NO:13的位置1599的位置处从G至T(或U)或C的变化。在一些实施方案中,所述改变将GPR156蛋白中的编码对应于位置533的位置的密码子的序列改变为GAT(或GAU)。在一些实施方案中,所述改变将GPR156蛋白中的编码对应于位置533的位置的密码子的序列改变为GAC。

[0209] 由American Psychiatric Association, Washington D.C., 2013发表的Diagnostic and Statistical Manual of Mental Disorders-第五版(DSM-V)是美国心理健康专家的主要诊断参考文献。根据DSM-V,单极抑郁症由以下组成:严重抑郁障碍、心境恶劣障碍、混合型抑郁障碍、具有抑郁心境的适应障碍以及另外未指明的抑郁症(NOS)。抑郁症的DSMIV准则是2周或更长时间的来自以下的五种或更多种症状:i)几乎每天大部分时间感觉抑郁心境;ii)对于所有或几乎所有活动明显减少的兴趣或乐趣;iii)显著的体重减轻或食欲下降;iv)失眠或嗜睡;v)精神运动激动或阻滞;vi)疲劳或能量损耗;vii)感觉无价值或过度内疚;viii)思考或注意力集中的能力下降或优柔寡断;以及ix)重复的死亡想法。心境恶劣障碍的DSMIV准则的特征在于剧烈而又缓慢的抑郁状态,其通过在大部分时间里、为期多天、持续至少2年的抑郁心境表现出来。患有这种病症的人在不多于2月的时间里肯定经历以下症状中的两种或更多种:i)食欲差或过度饱食;ii)失眠或嗜睡;iii)能量低或疲劳;iv)自尊心低;v)注意力差或难以做出决定;以及vi)感觉无助。

[0210] 本公开还提供用于在人类受试者中诊断焦虑病症或检测焦虑病症的风险的方法,所述方法包括:检测从人类受试者获得的编码GPR156蛋白的核酸分子中的改变,其中所述改变在对应于根据SEQ ID NO:4的GPR156蛋白中的位置533的位置处或在对应于根据SEQ ID NO:5的位置529的位置处编码天冬氨酸;以及如果所述受试者具有一种或多种焦虑病症症状(包括一种或多种临床神经精神性焦虑病症症状),则将人类受试者诊断为具有焦虑病症,或者如果所述受试者不具有一种或多种焦虑病症症状,则将人类受试者诊断为具有焦

虑病症的风险。在一些实施方案中,所述人类受试者需要这种诊断。在一些实施方案中,所述人类受试者可能具有已诊断为具有焦虑病症的亲属。所述改变可包括在对应于根据SEQ ID N0:13的位置1599的位置处从G至T(或U)或C的变化。在一些实施方案中,所述改变将GPR156蛋白中的编码对应于位置533的位置的密码子的序列改变为GAT(或GAU)。在一些实施方案中,所述改变将GPR156蛋白中的编码对应于位置533的位置的密码子的序列改变为GAC。在一些实施方案中,焦虑病症是临床神经精神性焦虑病症。

[0211] 焦虑病症的生理症状包括但不限于肌肉紧张、心悸、出汗、眩晕和呼吸短促。情绪症状包括但不限于坐立不安、濒临死亡的感觉、对死亡的恐惧、对尴尬或羞耻的恐惧以及对发生的不好的事情的恐惧。

[0212] 在一些实施方案中,所述方法包括检测从mRNA(从受试者获得的生物样品获得)获得的变体GPR156基因组DNA、mRNA或cDNA的存在。应理解,群体内的基因序列和由此类基因编码的mRNA可由于多态性(诸如单核苷酸多态性(SNP))而变化。本文提供的GPR156基因、mRNA、cDNA和多肽的序列仅是示例性序列,并且其他此类序列(包括除了GPR156蛋白中的在对应于位置533(根据SEQ ID N0:4)/529(根据SEQ I N0:5)的位置处编码天冬氨酸的等位基因以外的GPR156等位基因)也是可能的。

[0213] 在一些实施方案中,检测步骤包括对编码GPR156蛋白的核酸分子的至少一部分进行测序,其中测序的核酸分子编码包含GPR156蛋白的对应于根据SEQ ID N0:4的位置533或根据SEQ ID N0:5的位置529的位置的氨基酸序列。可对本文公开的核酸分子中的任一种(例如,基因组DNA、mRNA或cDNA)进行测序。在一些实施方案中,所述检测步骤包括对完整核酸分子进行测序。

[0214] 在一些实施方案中,所述检测步骤包括:对编码GPR156蛋白的核酸分子的至少一部分进行扩增,其中扩增的核酸分子编码包含对应于根据SEQ ID N0:4的位置533(和/或对应于根据SEQ ID N0:5的位置529)的位置的核酸序列;用可检测的标记对核酸分子进行标记;将标记的核酸与包含探针的支持物接触,其中所述探针包含在严格条件下与在对应于根据SEQ ID N0:4的位置533(和/或对应于根据SEQ ID N0:5的位置529)的位置处编码天冬氨酸的核酸序列杂交的核酸序列;以及检测可检测的标记。本文公开的核酸分子中的任一种可进行扩增。例如,本文公开的基因组DNA、cDNA或mRNA分子中的任一种可进行扩增。在一些实施方案中,所述核酸分子是mRNA,并且所述方法还包括在扩增步骤之前将mRNA逆转录为cDNA。

[0215] 在一些实施方案中,所述检测步骤包括:将核酸分子与包含可检测标记的探针接触,其中所述探针包含在严格条件下与编码氨基酸序列的核酸序列杂交的核酸序列,所述氨基酸序列在对应于根据SEQ ID N0:4的位置533(和/或对应于根据SEQ ID N0:5的位置529)的位置处包含天冬氨酸;以及检测可检测的标记。在一些实施方案中,所述核酸分子存在于从人类受试者获得的细胞内,使得所述检测根据原位杂交技术进行。

[0216] 在一些实施方案中,所述检测步骤包括将核酸分子与等位基因特异性引物接触,以及使用等位基因特异性PCR技术扩增核酸分子。等位基因特异性引物可以是本文所述的任何这种引物,并且对于在对应于根据SEQ ID N0:4的位置533和/或根据SEQ ID N0:5的位置529的位置处编码天冬氨酸的GPR156的等位基因可以是特异性的。

[0217] 可在本文公开的方法中使用的其他测定包括例如逆转录聚合酶链反应(RT-PCR)

或定量RT-PCR (qRT-PCR)。可在本文公开的方法中使用的又其他测定包括例如RNA测序(RNA-Seq)，接着确定生物样品中的变体mRNA或cDNA的存在和量。

[0218] 本公开还提供用于鉴定具有单极抑郁症或具有发展单极抑郁症的风险的人类受试者的方法。所述方法通常包括在从受试者获得的样品中确定GPR156蛋白内的E533D变化的存在或不存在；和/或包含编码GPR156蛋白内的E533D变化的突变的核酸分子的存在或不存在。GPR156蛋白内的E533D变化的存在和/或编码E533D变化的核酸分子的存在指示所述受试者具有单极抑郁症或具有发展单极抑郁症的风险。所述方法可在体外、原位或在体内实施。

[0219] 本公开还提供用于鉴定具有焦虑病症或具有发展焦虑病症的风险的人类受试者的方法。所述方法通常包括在从受试者获得的样品中确定GPR156蛋白内的E533D变化的存在或不存在；和/或包含编码GPR156蛋白内的E533D变化的突变的核酸分子的存在或不存在。GPR156蛋白内的E533D变化的存在和/或编码E533D变化的核酸分子的存在指示所述受试者具有焦虑病症或具有发展焦虑病症的风险。所述方法可在体外、原位或在体内实施。在一些实施方案中，焦虑病症是临床神经精神性焦虑病症。

[0220] 在所述方法的一些实施方案中，所述确定步骤包括对编码GPR156蛋白的核酸分子的至少一部分进行测序。测序的核酸分子可编码包含对应于根据SEQ ID NO:4的位置533的位置的氨基酸序列。所述确定步骤可包括对编码完整GPR156蛋白的核酸分子进行测序。在所述方法的一些实施方案中，所述确定步骤包括：对编码GPR156蛋白的核酸分子的至少一部分进行扩增；用可检测标记对核酸分子进行标记；将标记的核酸与包含探针的支持物接触，其中所述探针包含核酸序列，所述核酸序列包括例如在严格条件下与在对应于根据SEQ ID NO:4的位置533的位置处编码天冬氨酸的核酸序列特异性地杂交；以及检测可检测的标记。扩增的核酸分子优选地编码包含对应于根据SEQ ID NO:4的位置533的位置的氨基酸序列。如果所述核酸包括mRNA，则所述方法还可包括在扩增步骤之前将mRNA逆转录为cDNA。在一些实施方案中，所述确定步骤包括将核酸分子与包含可检测标记的探针接触以及检测可检测标记。所述探针优选地包含核酸序列，所述核酸序列包括例如在严格条件下与编码在对应于根据SEQ ID NO:4的位置533的位置处包含天冬氨酸的氨基酸序列的核酸序列特异性地杂交。所述核酸分子可存在于从人类受试者获得的细胞内。作为所述方法的一部分，GPR156蛋白内的E533D变化可由核酸序列GAT编码，或者可由核酸序列GAC编码。

[0221] 本公开还提供用于在人类受试者中诊断单极抑郁症或检测单极抑郁症的风险的方法，所述方法包括：检测从人类受试者获得的变体GPR156蛋白，诸如包含SEQ ID NO:4或SEQ ID NO:5的蛋白质；以及如果所述受试者具有一种或多种抑郁症症状，则将人类受试者诊断为具有单极抑郁症，或者如果所述受试者不具有一种或多种抑郁症症状，则将人类受试者诊断为具有单极抑郁症的风险。在一些实施方案中，所述人类受试者需要这种诊断。在一些实施方案中，所述人类受试者可能具有已诊断为具有单极抑郁症的亲属。在一些实施方案中，抑郁症或单极抑郁症不是儿童抑郁症和/或焦虑症。

[0222] 本公开还提供用于在人类受试者中诊断焦虑病症或检测焦虑病症的风险的方法，所述方法包括：检测从人类受试者获得的变体GPR156蛋白，诸如包含SEQ ID NO:4或SEQ ID NO:5的蛋白质；以及如果所述受试者具有一种或多种焦虑病症症状，则将人类受试者诊断为具有焦虑病症，或者如果所述受试者不具有一种或多种焦虑病症症状，则将人类受试者

诊断为具有焦虑病症的风险。在一些实施方案中,所述人类受试者需要这种诊断。在一些实施方案中,所述人类受试者可能具有已诊断为具有焦虑病症的亲属。在一些实施方案中,焦虑病症是临床神经精神性焦虑病症。

[0223] 在一些实施方案中,所述方法还包括当在受试者中检测到所述改变并且受试者被诊断为具有单极抑郁症时,用抗抑郁药治疗受试者。

[0224] 在一些实施方案中,抗抑郁药是选择性5-羟色胺再吸收抑制剂(SSRI)。SSRI的合适的实例包括但不限于西酞普兰(Celexa[®])、依他普仑(Lexapro[®], Cipralex[®])、帕罗西汀(Paxil[®], Seroxat[®])、氟西汀(Prozac[®])、氟伏沙明(Luvox[®])以及舍曲林(Zoloft[®], Lustral[®])。

[0225] 在一些实施方案中,抗抑郁药是5-羟色胺-去甲肾上腺素再吸收抑制剂(SNRA)。SNRI的合适的实例包括但不限于去甲文拉法辛(Pristiq[®])、度洛西汀(Cymbalta[®])、左旋米那普伦(Fetzima[®])、米那普仑(Ixel[®], Savella[®])、Tofenacen(Elamol[®], Tofacine[®])以及文拉法辛(Effexor[®])。

[0226] 在一些实施方案中,抗抑郁药是去甲肾上腺素再吸收抑制剂(NRI)。NRI的合适的实例包括但不限于瑞波西汀(Edronax[®])、维洛沙嗪(Vivalan[®])和阿托莫西汀(Strattera)。

[0227] 在一些实施方案中,抗抑郁药是锂。

[0228] 在一些实施方案中,抗抑郁药是5-羟色胺调节剂和刺激剂(SMS)。SMS的合适的实例包括但不限于维拉佐酮(Viibryd[®])和沃替西汀(Trintellix[®])。

[0229] 在一些实施方案中,抗抑郁药是5-羟色胺拮抗剂和再吸收抑制剂(SARI)。SARI的合适的实例包括但不限于依托哌酮(Axiomin[®], Etonin[®])、奈法唑酮(Nefadar[®], Serzone[®])和曲唑酮(Desyrel[®])。

[0230] 在一些实施方案中,抗抑郁药是三环类抗抑郁药(TCA)。TCA的合适的实例包括但不限于阿米替林(Elavil[®], Endep[®])、氯阿米替林(Amioxid[®], Ambivalon[®], Equilibrin[®])、氯米帕明(Anafranil[®])、地昔帕明(Norpramin[®], Pertofrane[®])、二苯西平(Noveril[®], Victoril[®])、二甲他林(Istonil[®])、度硫平(Prothiadene[®])、多虑平(Adapin[®], Sinequan[®])、丙米嗪(Tofranil[®])、洛非帕明(Lomont[®], Gamanil[®])、美利蒽(Dixeran[®], Melixeran[®], Trausabun[®])、硝基氮杂(Nitroxazepine)(Sintamil[®])、去甲替林(Pamelor[®], Aventyl[®])、诺昔替林(Agedal[®], Elronon[®], Nogedal[®])、哌泊非嗪(Azafen[®]/Azaphen[®])、普罗替林(Vivactil[®])以及三甲丙咪嗪(Surmontil[®])。还包括布替林(Evadyne[®])、地美替林(Deparon[®], Tinoran[®])、氯米帕明(Imiprex[®], Elepsin[®])、伊普吲哚(Prondol[®], Galatur[®], Tetran[®])、美他帕明(Timaxel[®])、丙毗西林(Depressin[®], Vagran[®])以及奎纽帕

明(Kinupril[®], Kevopril[®])。还包括奥匹哌醇(Insidon[®])和噻萘普汀(Stablon[®])。

[0231] 在一些实施方案中,抗抑郁药是四环类抗抑郁药(TeCA)。TeCA的合适的实例包括但不限于阿莫沙平(Asendin[®])、马普替林(Ludiomil[®])、米安舍林(Bolvidon[®], Norval[®], Tolvon[®])、米氮平(Remeron[®])以及司普替林(Tecipul[®])。

[0232] 在一些实施方案中,抗抑郁药是单胺氧化酶抑制剂(MAOI)。MAOI的合适的实例包括但不限于异丙烟肼(Marsilid[®])、异卡波肼(Marplan[®])、苯乙肼(Nardil[®])、司来吉兰(Eldepryl[®], Zelapar[®], Emsam[®])、反苯环丙胺(Parnate[®])、美曲吡嗪(Inkazan[®])、吗氯贝胺(Aurorix[®], Manerix[®])、吡唑啉(Pirazidol[®])以及托洛沙酮(Humoryl[®])。其他的包括例如苯莫辛(Neuralex[®])、卡罗沙宗(Surodil[®], Timostenil[®])、异丙氯肼(Sursum[®])、美巴那肼(Actomol[®])、尼拉莫辛(Niamid[®])、奥他莫辛(Ximaol[®])、苯异丙肼(Catron[®])、苯氧丙肼(Drazine[®])、匹戈肼(Tersavid[®])、沙夫肼(Safra[®])、依普贝胺(Befol[®])以及苯哒吗啉(Brantur[®], Cantor[®])。

[0233] 在一些实施方案中,抗抑郁药是非典型的抗精神病药。非典型的抗精神病药的合适的实例包括但不限于氨磺必利(Solian[®])、鲁拉西酮(Latuda[®])以及喹硫平(Seroquel[®])。

[0234] 在一些实施方案中,抗抑郁药是阿戈美拉汀(Valdoxan[®])、二苯美仑(Alnert[®], Celeport[®])、安非他酮(Wellbutrin[®])、氯胺酮(Ketalar[®])、坦度螺酮(Sediel[®])或替尼沙秦(Lucelan[®], Metatone[®])。

[0235] 在一些实施方案中,所述方法还包括当在受试者中检测到所述改变并且受试者被诊断为具有焦虑病症时,用抗焦虑剂治疗受试者。在一些实施方案中,焦虑病症是临床神经精神性焦虑病症。

[0236] 在一些实施方案中,抗焦虑剂是苯并二氮杂,包括但不限于阿普唑仑(Xanax[®])、溴西泮(Lectopam[®], Lexotan[®])、利眠宁(Librium[®])、氯硝西泮(Klonopin[®], Rivotril[®])、氯氮(Tranxene[®])、安定(Valium[®])、氟西泮(Dalmane[®])、劳拉西泮(Ativan[®])、奥沙西泮(Serax[®], Serapax[®])、替马西泮(Restoril[®])、三唑仑(Halcion[®])以及托非索泮(Emandaxin[®], Grandaxin[®])。

[0237] 在一些实施方案中,抗焦虑剂是氨基甲酸酯,包括但不限于氨甲丙二酯(Miltown[®], Equanil[®])。

[0238] 在一些实施方案中,抗焦虑剂是抗组胺剂,包括但不限于羟嗪(Atarax[®])、氯苯那敏(Chlor-Trimeton[®])以及苯海拉明(Benadryl[®])。

[0239] 在一些实施方案中,抗焦虑剂是氮哌酮,包括但不限于丁螺环酮(Buspar[®])和坦度

螺酮(Sediel[®])。

[0240] 在一些实施方案中,抗焦虑剂是SSRI、SNRI、TCA、TeCA或MAOI。

[0241] 在一些实施方案中,抗焦虑剂是Mebicar(Mebicarum[®])、法莫替唑(Afobazole[®])、Selank、布罗曼坦、Emoxypine、普加巴林、异戊酸薄荷醇酯或异戊酸薄荷醇酯(Validol[®])。

[0242] 抗抑郁药或抗焦虑剂的施用可通过任何合适的途径进行,包括但不限于肠胃外、静脉内、口服、皮下、动脉内、颅内、鞘内、腹膜内、局部、鼻内或肌内。用于施用的药物组合物希望是无菌的且基本上等渗的,并且在GMP条件下制造。药物组合物可以单位剂量型(即,单次施用的剂量)提供。药物组合物可使用一种或多种生理上且药学上可接受的携带体、稀释剂、赋形剂或辅助剂来配制。制剂取决于所选的施用途径。术语“药学上可接受的”意指携带体、稀释剂、赋形剂或辅助剂与制剂的其他成分相容,并且对其接受者基本上无害。

[0243] 本公开还提供用于在包含编码人类GPR156蛋白的基因中的改变的人类受试者中治疗单极抑郁症的抗抑郁药,其中所述改变在对应于根据SEQ ID NO:4的位置533或对应于根据SEQ ID NO:4的位置529的位置处编码天冬氨酸。本公开还提供用于在编码人类GPR156蛋白的基因中的改变的人类受试者中治疗单极抑郁症的药物的制造中使用的抗抑郁药,其中所述改变在人类GPR156蛋白中在对应于根据SEQ ID NO:4的位置533或对应于根据SEQ ID NO:5的位置529的位置处编码天冬氨酸。在一些实施方案中,所述改变将编码对应于根据SEQ ID NO:4的位置533或对应于根据SEQ ID NO:5的位置529的位置的密码子的序列改变为GAT(或GAU)。在一些实施方案中,所述改变将编码对应于根据SEQ ID NO:4的位置533或对应于根据SEQ ID NO:5的位置529的位置的密码子的序列改变为GAC。在一些实施方案中,抗抑郁药是选择性5-羟色胺再吸收抑制剂、5-羟色胺-去甲肾上腺素再吸收抑制剂或去甲肾上腺素再吸收抑制剂。在一些实施方案中,抗抑郁药是锂。

[0244] 在用于在人类受试者中治疗单极抑郁症的抗抑郁药的一些实施方案中,人类受试者已被测试对于GPR156蛋白内的E533D变化和/或对于编码GPR156蛋白内的E533D变化的核酸分子是阳性的。在一些实施方案中,所述治疗包括确定人类受试者是否具有携带E533D变化的GPR156蛋白和/或编码携带E533D变化的GPR156蛋白的核酸分子的步骤。在一些实施方案中,根据用于鉴定具有单极抑郁症或具有发展单极抑郁症的风险的人类受试者的方法(包括本文描述或例示的任何这种方法),人类受试者已被鉴定为具有单极抑郁症或具有发展单极抑郁症的风险。E533D变化是指在对应于根据SEQ ID NO:4的位置533的位置处产生天冬氨酸的突变。在一些实施方案中,GPR156蛋白内的E533D变化由密码子GAT编码。在一些实施方案中,GPR156蛋白内的E533D变化由密码子GAC编码。在一些实施方案中,用于在人类受试者中治疗单极抑郁症的抗抑郁药是选择性5-羟色胺再吸收抑制剂、5-羟色胺-去甲肾上腺素再吸收抑制剂或去甲肾上腺素再吸收抑制剂。在一些实施方案中,用于在人类受试者中治疗单极抑郁症的抗抑郁药是锂。

[0245] 本公开还提供用于在包含编码人类GPR156蛋白的基因中的改变的人类受试者中治疗焦虑病症的抗焦虑剂,其中所述改变在对应于根据SEQ ID NO:4的位置533或对应于根据SEQ ID NO:4的位置529的位置处编码天冬氨酸。本公开还提供用于在编码人类GPR156蛋

白的基因中的改变的人类受试者中治疗焦虑病症的药物的制造中使用的抗焦虑剂，其中所述改变在人类GPR156蛋白中在对应于根据SEQ ID N0:4的位置533或对应于根据SEQ ID N0:5的位置529的位置处编码天冬氨酸。在一些实施方案中，所述改变将编码对应于根据SEQ ID N0:4的位置533或对应于根据SEQ ID N0:5的位置529的位置的密码子的序列改变为GAT(或GAU)。在一些实施方案中，所述改变将编码对应于根据SEQ ID N0:4的位置533或对应于根据SEQ ID N0:5的位置529的位置的密码子的序列改变为GAC。在一些实施方案中，抗焦虑剂是苯并二氮杂、氨基甲酸酯、抗组胺剂、氮哌酮、SSRI、SNRI、TCA、TeCA或MAOI。在一些实施方案中，焦虑病症是临床神经精神性焦虑病症。

[0246] 在用于在人类受试者中治疗焦虑病症的抗焦虑剂的一些实施方案中，人类受试者已被测试对于GPR156蛋白内的E533D变化和/或对于编码GPR156蛋白内的E533D变化的核酸分子是阳性的。在一些实施方案中，所述治疗包括确定人类受试者是否具有携带E533D变化的GPR156蛋白和/或编码携带E533D变化的GPR156蛋白的核酸分子的步骤。在一些实施方案中，根据用于鉴定具有焦虑病症或具有发展焦虑病症的风险的人类受试者的方法(包括本文描述或例示的任何这种方法)，人类受试者已被鉴定为具有焦虑病症或具有发展焦虑病症的风险。E533D变化是指在对应于根据SEQ ID N0:4的位置533的位置处产生天冬氨酸的突变。在一些实施方案中，GPR156蛋白内的E533D变化由密码子GAT编码。在一些实施方案中，GPR156蛋白内的E533D变化由密码子GAC编码。在一些实施方案中，用于在人类受试者中治疗焦虑病症的抗焦虑剂是苯并二氮杂、氨基甲酸酯、抗组胺剂、氮哌酮、SSRI、SNRI、TCA、TeCA或MAOI。在一些实施方案中，焦虑病症是临床神经精神性焦虑病症。

[0247] 本公开还提供本文公开的变体GPR156基因、mRNA、cDNA、多肽和杂交核酸分子中的任一种在诊断单极抑郁症或诊断发展单极抑郁症的风险中的用途。

[0248] 本公开还提供本文公开的变体GPR156基因、mRNA、cDNA、多肽和杂交核酸分子中的任一种在诊断焦虑病症或诊断发展焦虑病症的风险中的用途。在一些实施方案中，焦虑病症是临床神经精神性焦虑病症。

[0249] 以上或以下所引用的所有专利文献、网站、其他出版物、登录号等均在犹如每个单独项目均具体地且单独地指示为以引用的方式如此并入一样的相同程度上出于所有目的以引用的方式整体并入。如果不同的序列版本与不同时间时的登录号相关，则与本申请的有效提交日期时的登录号相关的版本是有意义的。有效提交日期意指实际提交日期的较早时间或参考登录号(如果适用)的优先权申请的提交日期。同样，除非另有指示，否则如果不同版本的出版物、网站等在不同时间公布，则最近在申请的有效申请日期公布的版本是有意义的。除非另有具体指示，否则本公开的任何特征、步骤、元件、实施方案或方面均可与任何其他特征、步骤、元件、实施方案或方面组合使用。尽管出于清楚和理解的目的已通过说明和实例详细描述本公开，但将显而易见的是，可在所附权利要求的范围内进行某些变化和修改。

[0250] 本文引用的核苷酸和氨基酸序列使用核苷酸碱基的标准字母缩写和氨基酸的单字母密码示出。核苷酸序列遵循在序列的5'末端处开始并正向前进(即，每行中从左至右)至3'末端的标准惯例。仅示出每一核苷酸序列的一条链，但是互补链应理解为包括在提到的任何所展示的链内。氨基酸序列遵循在序列的氨基末端处开始并正向前进(即，每行中从左至右)至羧基末端的标准惯例。

[0251] 提供以下实施例来更详细地描述实施方案。它们意图说明但不限制所要求保护的实施方案。

实施例

[0252] 提出以下实施例以便为本领域普通技术人员提供本文所要求保护的化合物、组合物、制品、装置和/或方法如何制备和评价的完整公开和描述且旨在仅仅是示例性地并且不意图限制本发明人所认定的发明范围。已努力确保数值(例如,量、温度等)的准确性,但应考虑一些误差和偏差。除非另外指示,否则份数是重量份,温度是以°C为单位或环境温度,并且压力等于或接近大气压。

[0253] 实施例1:患者招募和表型分析

[0254] 在已确定单极抑郁症的遗传评估的门诺族血统的家庭中执行全外显子组测序。通过使用针对DSM-5的结构化临床访问(SCID-5),由神经精神病学团队另外系统性地评估此谱系中的个体。此半结构化访问由经过训练的心理健康专家施用,并被设计来确保DSM-5病症的系统性且准确的诊断。此访问用于确定个体及其亲属是否符合严重抑郁症、泛化性焦虑病症和任何其他共病性精神病状的准则。所述访问通常花费一至两个小时来针对受试者进行实施。除访问之外,汇集附属信息以根据受试者的医疗记录、治疗提供者和在受试者的书面同意书的情况下进行的亲密接触来确认诊断。对涵盖总计50位个体的此大谱系中的全部存在的受影响和不受影响的家庭成员的外显子组进行测序。

[0255] 实施例2:基因组样品

[0256] 从外周血样品提取基因组DNA并转移到Regeneron遗传学中心(RGC)以进行全外显子组测序,并且在-80°C下储存在自动生物库中。执行基于荧光的定量以确保用于测序目的的适当的DNA量和质量。

[0257] 将1μg的DNA剪切至150个碱基对的平均片段长度(Covaris LE220),并且准备来自Kapa Biosystems的定制的试剂盒进行外显子组捕获。使用NimbleGen SeqCap VC Rome 2.1或整合DNA技术xGen外显子组靶设计来捕获样品。将样品进行条形码化、汇集并多重重复以用于在具有v4化学的Illumina HiSeq 2500上使用75bp配对末端测序进行测序。对捕获的片段进行测序以实现在20x或更大覆盖率下覆盖的最少85%的靶碱基。在测序之后,使用在RGC处开发的使用DNAnexus和AWS以针对样品-水平数据产生和分析运行标准工具的基于云的管线对数据进行处理。简而言之,生成序列数据并且使用Illumina's CASAVA软件进行多路分解。对序列读取进行作图并使用BWA-mem与GRCh37/hg19人类基因组参考组件进行比对。在比对之后,使用Picard工具对重复读取进行标记并设置标志,并且使用GATK对indel进行再比对以改善变体识别质量。使用GATK's HaplotypeCaller识别SNP和INDEL变体和基因型,并且应用来自GATK的变体质量分数再校准(VQSR)来评注总体变体质量分数。针对每个样品捕获测序和数据质量计量统计以评估捕获性能、比对性能和变体识别。

[0258] 实施例3:基因组数据分析

[0259] 将最小读取长度(>10)、基因型质量(>30)和等位基因平衡(>15%)的标准质量控制过滤器应用于识别的变体。使用RGC开发的评注和分析管线基于其潜在功能性作用(无论是同义、非同义、剪接、移码或非移码变体)对通过的变体进行分类和评注。通过来自遗传数据的血源同一(IBD)衍生的计量核实家族关系以使用PRIMUS(Staples等人,Amer.J.Human

Genet., 2014, 95, 553-564) 和关于此家族的报告的谱系的交叉参考来推断群组中的亲缘和关系。

[0260] 执行基于谱系的变体分析和分离来根据报告的家庭历史在常染色体显性遗传模式下鉴定候选疾病基因。使用可用于测序的尽可能多的具有信息的受影响个体来执行初级分析以减少由以下产生的潜在干扰因子：病症的不完全外显率、发病的年龄或潜在暴露于可触发疾病的症状性呈现的环境因素。随后通过其在群体控制数据库（诸如dbSNP、1000基因组计划、NHLBI外显子组测序计划、外显子组集成联合数据库(ExAc)以及内部RGC数据库）中观察到的频率来评注并过滤所有受影响的个体之间共享的变体以过滤掉常见的多态性且高频率、可能良性的变体。将用于变体的功能性作用的生物信息预测的算法（诸如LRT、Poly-phen2、SIFT、CADD和Mutation Taster）连同基于多个物种比对（即，GERP、PhastCons、PhyloP）的保守分数作为变体的评注过程的一部分整合并用于提供关于鉴定的候选变体的潜在毒性信息。

[0261] 在基因GPR156中识别到以此谱系的受影响家庭成员的单级抑郁症的表型分离的单个罕见变体(hg19.g.chr3:119886725(C>A); c.1599G>T; p.Glu533Asp,p.E533D)。未包括在初级分析中但也携带p.E533D变体的另外的家庭成员已确认焦虑病症的诊断。为了排除以此家庭中的抑郁症或焦虑病症表型分离的其他可能的变体，基于公布的MERLIN算法(Abecasis等人,Nat.Genet.,2002,30)，使用定制的RG C开发的管线执行单点外显子组宽和多点外显子组宽的连锁分析。简而言之，使用MERLIN(csg.sph.umich.edu/abecasis/merlin/index.html)执行SNP芯片和外显子组衍生的基因型的参数连锁分析。将PLINK格式化的基因型限于至少一个非基本成员中存在的单核苷酸变体(SNV)。在基本成员和谱系的其他非亲属成员中评估等位基因频率。将没有遗传数据的个体指定为缺失/虚拟基因型。使用fcGENE(sourcefor ge.net/projects/fcgene/)将PLINK格式基因型转换为MERLIN格式化的ped文件。使用定制的脚本制备连锁分析所必需的另外的文件(.dat和.map)。使用Pedstats(csg.sph.umich.edu/abecasis/pedstats/index.html)验证谱系结构。使用基因型之间的逐步或网格间距来实施芯片数据的连锁分析，不做关于重组的假设。使用单点连锁来分析基于外显子组的基因型。对于所有的分析，假设不完全外显率模型(0.0001,0.75,1)，其中疾病等位基因频率为大约0.01%。

[0262] 对于外显子组宽和基因组宽的分析两者，在此谱系中鉴定单个连锁峰，达到最大2.98的LOD分数并在人类染色体3中扩展14.4兆碱基(Mb)区域。GPR156基因位于在此连锁区域内，并且通过基于初始罕见变体家族的分离分析鉴定的候选变体(c.1599G>T, p.Glu533Asp)实现2.56的单个标记物LOD分数。

[0263] 如本文所证明，这些遗传分析指示GPR156是与对于单极抑郁症和焦虑病症的增加的家族易感性相关联的基因。

[0264] 另外，将包含野生型和突变体E533D序列的核苷酸片段亚克隆到载体(参见，图1和图2)中以获得用于另外的分析的野生型和变体E533D GPR156构建体。

[0265] 实施例4:检测

[0266] 受试者中某一遗传变体的存在可指示受试者具有增加的具有或发展单极抑郁症的风险。样品（诸如血液样品）可从受试者获得。核酸可使用常见的核酸提取试剂盒从样品分离。在将核酸从受试者获得的样品分离之后，对核酸进行测序以确定是否存在遗传变体。

核酸的序列可与对照序列(野生型序列)进行比较。在从受试者获得的样品获得的核酸与对照序列之间发现差异指示遗传变体的存在。这些步骤可如以上实施例且在整个本公开中所描述的执行。一种或多种遗传变体的存在指示受试者的增加的具有或发展单极抑郁症的风险。

[0267] 本申请中引用的每个参考文献(包括但不限于期刊文章、美国和非美国专利、专利申请公布、国际专利申请公布、基因库登录号等)以引用的方式整体并入本文。

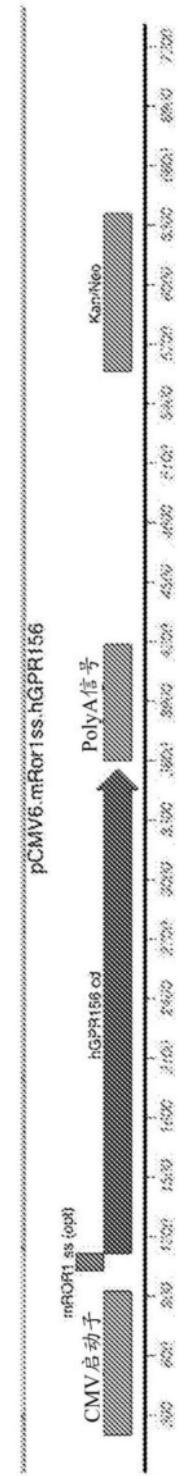


图1

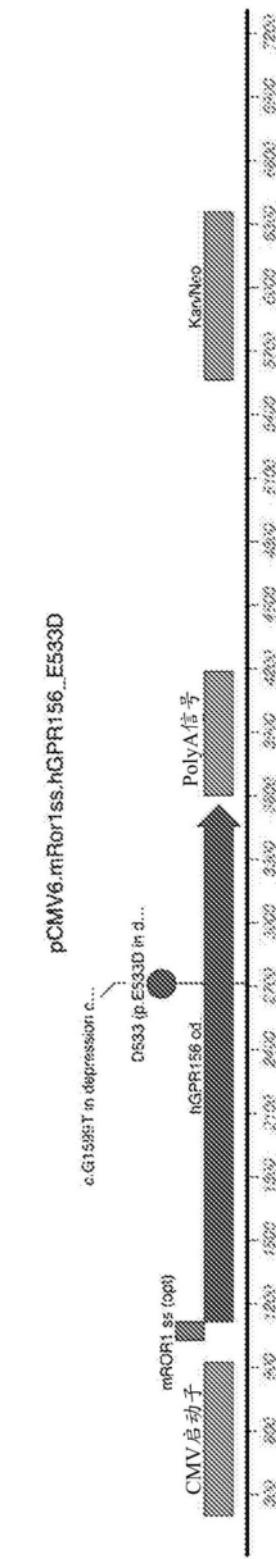


图2