



(21) 申请号 201880074086.9	C12Q 1/6883 (2006.01)
(22) 申请日 2018.10.01	(56) 对比文件
(65) 同一申请的已公布的文献号 申请公布号 CN 111344415 A	WO 2017001568 A1, 2017.01.05 WO 2016207653 A1, 2016.12.29 WO 2009147386 A1, 2009.12.10
(43) 申请公布日 2020.06.26	C.R. Lim.EPIGENETIC SIGNATURES AND EARLY DETECTION OF NEURODEGENERATIVE DISEASES: DEVELOPMENT OF STRATIFYING BIOMARKERS FOR AMYOTROPHIC LATERAL SCLEROSIS IN ASIAN COHORTS.《WCN2017》 .2017,第1页.
(30) 优先权数据 62/566,756 2017.10.02 US	C.R. Lim.EPIGENETIC SIGNATURES AND EARLY DETECTION OF NEURODEGENERATIVE DISEASES: DEVELOPMENT OF STRATIFYING BIOMARKERS FOR AMYOTROPHIC LATERAL SCLEROSIS IN ASIAN COHORTS.《WCN2017》 .2017,第1页.
(85) PCT国际申请进入国家阶段日 2020.05.15	A Krause.JUNCTOPHILIN 3 (JPH3) EXPANSION MUTATIONS CAUSING HUNTINGTON DISEASE LIKE 2 (HDL2) ARE COMMON IN SOUTH AFRICAN PATIENTS WITH AFRICAN ANCESTRY AND A HUNTINGTON DISEASE PHENOTYPE.《Am J Med Genet B Neuropsychiatr Genet》.2015,第 168卷(第7期),第573-585页.
(86) PCT国际申请的申请数据 PCT/GB2018/052808 2018.10.01	审查员 滕瑞芳
(87) PCT国际申请的公布数据 W02019/069067 EN 2019.04.11	权利要求书2页 说明书44页 附图5页
(73) 专利权人 牛津生物动力有限责任公司 地址 英国牛津	
(72) 发明人 亚历山大·阿库利特切夫 阿鲁尔·塞尔瓦姆·拉玛达斯 伊万·亨特 马修·萨尔特	
(74) 专利代理机构 北京康信知识产权代理有限 责任公司 11240 专利代理师 杜升	
(51) Int.Cl. C12Q 1/6827 (2006.01)	

(54) 发明名称

生物标志物

(57) 摘要

一种分用于析与ALS和亨廷顿舞蹈症有关的染色体区域和相互作用的方法。

1. 至少4个探针序列用于制备用于检测代表群体中的疾病亚组的人类个体中的染色体状态的试剂盒的用途,包括确定基因组的限定区域内存在或不存在与所述染色体状态有关的染色体相互作用,其中所述疾病亚组是肌萎缩性侧索硬化症(ALS)亚组;其中分型至少4种染色体相互作用;并且其中进行所述染色体状态的检测以诊断ALS或确定ALS的预后;

其中所述检测包括:

(i) 存在于来自所述个体的样品中的表观遗传染色体相互作用的交联,以制备交联的DNA,

(ii) 使所述交联的DNA经受剪切,

(iii) 连接剪切的所述交联的DNA的末端,和

(iv) 确定结合所述探针序列的连接的DNA的存在或不存在,以确定特定染色体相互作用的存在或不存在;

其中所述至少4个探针序列:

(a) 选自探针序列TCACCACACATCACCCCTTGCTCCTCGAGTCTTGGTGACCACAACAGGGTGCCACC、GAGGTGGGTGAATCATGAGGTCAAGGGTTCGACAATAGTTGAGAATCTCCAACCACCTGG、GGCCTTATAGTCAGCTGATCAGGTGAAATCGATTGGTCCTTAGGATCAGCTACCATTTGC、GAGGCAGGCGGATCACAAAGTCAAAA GATCGATAACTTCAATAATAGTTACAGATGCAAA、AGCACCATATCTGGGATGTAGCTATTGCTCGAGATTGCAGT GAGCTGTGATCACACCTCT、TCTTCCCTCTTTTTAAAACCACCATTCATCGACCCACACATCCTGTGCCACTCTA CTGC、TAACCATTATGCATCACTAACATAGCATTTCGATATGATATGCTCAGTTTAGTTAGGGAAA、GGCTCAGGAA GAGAACTATTTGTCTCTTTGACACGCACATGCAGGACACTCACACGTAG、GTTGGGTGGATCCCTTGAGCTCAGG AATTGCAAGAATGATTTTTGAGCCCGTGTGGAAGG、ATCAAAAAGAAAATAGATACTTGTCTTACTCGAGTTGAATA AAATCCTCAGCTTTCTGTCC、AAAAGAACTGTGAAAAGTTGTCACATTTGATTAAATCCAAAAAGGTCTTCTAT GAGGC、TTAAAAGTATAGTAGTTGGCATTACATTCGACCTTTTTCTGTTTCAGTAACCAACCCAG、CATCAACTA ATAGTTAAACATTATAATATCGACTGAAGACCTTTCATACTGTAAGATTCA、CATCAACTAATAGTTAAACATTAT AATATCGAGTCTGCAGTGAGCTGAGATCACACTGCC、TTATTCCTTTCCAAAATAGTTAAAATTATTCGAAACTTTT AAGAATCAATATAAAAATTTCC、CATAATTATAAATTAAAAAATGACACTATCGATTATGTCCAGTGTTCCTTGGTT GGTGTC、和CAGAGCACTAAGATAGACTTCTAAGGTTTCGAGGCATATAGCTCCAGCTGTATTGAGGTA,

(b) 选自探针序列TATATTTAAAAATACATACTGGTATACATCGATCTCATGACTTTGCTATTATGCATA GTG、CCCCAGCCCAGCAACCTGGCTCACCTGATCGAGTACATCTTCAAGCCATCCTGTGTGCCC、TATAAATAATA CAGCTCTATTTGCCTACTCGATTAAAGAATCATATTATATCCTTAATTC、CACTATGCATAATAGCAAAGTCATGA GATCGAAAATGTTTGTCAAGCAGTAGGTTTTGGG、GATTTTAGAATCTCTAACAAGGCTGCAATCGAGGTTAGCTG CTGCAGAAAAGAAGAGAAAA、GGTGAATAAATGAGATTGCATTTTCTTTGACCATTTGGCCAGCATGCCAAACAC TGGT、TAACCTCTCCTTTCTTAGGTTCTCCATATCGATAGAAAATTGTCTGCAGCCCTTAATGCC、AGCTCACGGT CAGTGCCGTTCCGTTTGTCTGAATAAAGAACAAGGACCTTAAAAAATAGA、和AAAAGTCATACA ACTACTATGTAA GATATTCGAATACCTGTTAGAATAGGTGAAGGTTTAT,或

(c) 选自探针序列ATTTTCTTTGACCATTTGGCCAGCA、 TTCTAACAGGTATTTCGAATATCTTACATAGT、AAGTAACATTCGACAAACCTCTGGTAA、 TATTCTCTTTGATCGAGGCATATAGCT、AAAGTCATGAGATCGATGTATACCAGTAT、和 TTGACAAACATTTTCGATCTCATGACTT。

2. 根据权利要求1所述的用途,其中,在所述染色体状态的检测中分型以下染色体相互

作用的特定组合：

(i) 包含由权利要求1中限定的(a)或(b)中的探针代表的所有染色体相互作用；或

(ii) 包含由权利要求1中限定的(a)或(b)中的探针代表的至少5、6或7种染色体相互作用；或

(iii) 包含由权利要求1中限定的(c)所示的任何探针代表的所有染色体相互作用；或

(iv) 包含由权利要求1中限定的(c)中的探针代表的至少5、6或7种染色体相互作用。

3. 根据权利要求1所述的用途,其中,在所述染色体状态的检测中分型至少5至9种不同的染色体相互作用。

4. 根据权利要求1所述的用途,其中,确定连接的DNA的存在或不存在包括通过使用能够扩增所述连接的DNA的引物的定量PCR(qPCR)特异性检测所述连接的DNA,并且其中所述探针在PCR反应期间结合连接位点,并且其中所述探针包含:

共价连接至所述探针的5'端的荧光团,和/或

共价连接至所述探针的3'端的淬灭剂。

5. 根据权利要求4所述的用途,其中,所述荧光团选自HEX、德克萨斯红和FAM。

生物标志物

技术领域

[0001] 本发明涉及检测染色体相互作用。

背景技术

[0002] 生物标记允许鉴定疾病特征。目前常用的生物标志物包括RNA表达模式和蛋白质标志物。

发明内容

[0003] 由于与病理或治疗相关的调节性表观遗传控制设置,位点处的特异性染色体构象特征(CCS)可能存在或不存在。CCS具有温和的解离速率(off-rate),并且当代表特定的表型或病理时,它们将只会随着生理性信号传导的向新表型的转变或由于外部干预而变化。此外,这些事件的测量是二元(binary)的,并因此该读数与DNA甲基化、组蛋白修饰和大多数非编码RNA的变化的水平的连续读数形成鲜明对比。迄今为止,大多数分子生物标志物使用的连续读数对数据分析提出了挑战,因为特异性生物标志物的变化幅度随患者大幅变化,这在将其用于对患者群体进行分层次时会给分类统计带来问题。这些分类统计数据更适合使用无大小之分并仅提供表型差异的“是或否”的二元评分的生物标志物—表明染色体构象(EpiSwitch™)生物标志物是潜在的诊断、预后和预测性生物标志物的极佳资源。

[0004] 本发明人已经使用允许鉴定群体亚组的方法鉴定出染色体相互作用与肌萎缩性侧索硬化症(ALS)或亨廷顿舞蹈症有关的基因组区域。因此,本发明提供了用于检测代表群体亚组的染色体状态的方法,该方法包括确定在限定的基因组的疾病相关区域内是否存在或不存在染色体相互作用,其中所述疾病是ALS或亨廷顿舞蹈症。通过确定哪些染色体相互作用与对应于该群体的ALS或亨廷顿舞蹈症亚组的染色体状态相关的方法,染色体相互作用可以可选地鉴定或是可鉴定的(或可得出的),该方法包括使来自具有不同染色体状态的亚组的第一组核酸与第二组索引核酸接触,并允许互补序列杂交,其中第一和第二组核酸中的核酸代表包含来自在染色体相互作用中聚集在一起的两个染色体区域的序列的连接产物,并且其中第一和第二组核酸之间的杂交模式允许确定哪种染色体相互作用对ALS或亨廷顿舞蹈症亚组是特异性的。ALS或亨廷顿舞蹈症亚组可能与诊断(ALS或亨廷顿舞蹈症的存在)或预后(例如ALS或亨廷顿舞蹈症的进展速率)有关。本文描述的任何特异性的相关染色体相互作用(标志物)都可以用作本发明的基础,包括标志物的组合。

[0005] 本发明提供了一种用于检测代表群体中疾病亚组的染色体状态的方法,该方法包括确定在基因组的限定区域内存在或不存在与该染色体状态有关的染色体相互作用,其中所述疾病亚组是肌萎缩性侧索硬化症(ALS)亚组;并且

[0006] -其中所述染色体相互作用通过确定哪些染色体相互作用与对应于该群体的ALS亚组的染色体状态相关的方法可选地鉴定,该方法包括使来自具有该染色体的不同状态的亚组的第一组核酸与第二组索引核酸接触,和允许互补序列杂交,其中该第一和第二组核酸中的核酸代表包含来自在染色体相互作用中聚集在一起的两个染色体区域的序列的连

接产物,并且其中第一组和第二组核酸之间的杂交模式允许确定哪些染色体相互作用对ALS亚组具有特异性;并且

[0007] -其中该染色体相互作用:

[0008] (i) 存在于表1或表5所示的区域或基因中的任何一种中;和/或

[0009] (ii) 对应于由表1或5中所示的任何探针代表的任何一种染色体相互作用,和/或

[0010] (iii) 对应于表10或11中所示的任何一种染色体相互作用,和/或

[0011] (iv) 存在于包含侧接(i),(ii)或(iii)的4,000个碱基的区域中。

[0012] 本发明还提供了一种用于检测代表群体中的疾病亚组的染色体状态的方法,该方法包括确定在基因组的限定区域内存在或不存在与该染色体状态有关的染色体相互作用,其中所述疾病亚组是亨廷顿舞蹈症亚组;并且

[0013] -其中所述染色体相互作用通过确定哪些染色体相互作用与对应于该群体的亨廷顿舞蹈症亚组的染色体状态相关的方法可选地鉴定,该方法包括使来自具有该染色体的不同状态的亚组的第一组核酸与第二组索引核酸接触,和允许互补序列杂交,其中该第一和第二组核酸中的核酸代表包含来自在染色体相互作用中聚集在一起的两个染色体区域的序列的连接产物,并且其中第一组和第二组核酸之间的杂交模式允许确定哪些染色体相互作用对亨廷顿舞蹈症亚组具有特异性;并且

[0014] -其中该染色体相互作用:

[0015] (i) 存在于表12所示的区域或基因中的任何一种中;和/或

[0016] (ii) 对应于由表12中所示的任何探针代表的任何一种染色体相互作用,和/或

[0017] (iii) 对应于表12中代表的任何一种染色体相互作用,和/或

[0018] (iv) 存在于包含或侧接(i),(ii)或(iii)的4,000个碱基的区域中。

附图说明

[0019] 图1显示了如何可以检测染色体相互作用。

[0020] 图2显示了在亨廷顿舞蹈症工作中研究的基因组区域的图形概览。研究了跨越HTT位点的4号染色体上~225kb的区域。锚定点(迹线4中的“锚定”)定义为跨越HTT的外显子1中的CAG重复序列(tract)的~42kb区域(图顶部的紫色箭头)。我们基于与EpiSwitch位点(迹线3)、与HD(迹线5)或其他疾病(轨迹6)相关的SNP的重叠而定义了五个区域(迹线4中的区域1-5),并观察了HC和HD之间的甲基化和乙酰化(H3K4me3,H3K36me3和H3K27Ac)差异(轨迹7至12)。

[0021] 图3显示了在七个独立HD-Sym样品中的六个中观察到三种条件性相互作用(I5,I6和I7)中至少一种的存在。I5是跨越包含rs362331SNP的区域的相互作用,在最大数量的样品中被观察到(6/7)。

[0022] 图4提供了驱动亨廷顿舞蹈症的症状发展的表观遗传变化的模型。这显示了与HD进展相关的染色体构象变化的概览。随着患者从症状前阶段发展到有症状的疾病,在HTT位点观察到基因组结构的离散性、可测量和区分性的变化。

[0023] 图5显示了所有已鉴定的亨廷顿舞蹈症相互作用的总结。

具体实施方式

[0024] 本发明的方法

[0025] 本发明的方法包括用于检测与ALS或亨廷顿舞蹈症有关的染色体相互作用的分型系统。这种分型可以使用本文提到的基于在染色体相互作用中聚集在一起的染色体的交联区域的EpiSwitch™系统,使染色体DNA裂解并随后连接存在于交联实体中的核酸以衍生具有来自形成染色体相互作用的两个区域的序列的连接核酸进行。该连接核酸的检测允许检测特定染色体相互作用的存在或不存在。

[0026] 该染色体相互作用可以使用上述方法鉴定,其中使用了第一和第二核酸的群体。这些核酸也可以使用EpiSwitch™技术生成。

[0027] 与本发明有关的表观遗传相互作用

[0028] 如本文所用,术语“表观遗传”和“染色体”相互作用通常是指染色体的远端区域之间的相互作用,相互作用是动态的,并且取决于染色体区域的状态而改变,形成或破坏。

[0029] 在本发明的具体方法中,通过首先产生包含来自相互作用的部分的染色体的两个区域的序列的连接核酸检测染色体相互作用。在这样的过程中,该区域可以通过任何合适的方式交联。在优选的实施方式中,该相互作用是使用甲醛的交联,但也可以通过任何醛或D-生物素基(Biotinoyl)-e-氨基己酸-N-羟基琥珀酰亚胺酯或洋地黄毒苷-3-O-甲基羰基-e-氨基己酸-N-羟基琥珀酰亚胺酯交联。多聚甲醛可以交联相距4埃的DNA链。优选地,染色体相互作用发生于同一染色体上,并可选地相隔2至10埃。

[0030] 该染色体相互作用可以反映该染色体区域的状态,例如,是否响应于生理条件的变化而被转录或抑制。据发现,对本文限定的亚组具有特异性的染色体相互作用是稳定的,因此提供了测量该两个亚组之间的差异的可靠手段。

[0031] 此外,与特征(如疾病状况)具有特异性的染色体相互作用,例如,与其他表观遗传标志物如甲基化或组蛋白结合的变化相比,通常会在生物过程的早期发生。因此,本发明的方法能够检测生物学过程的早期阶段。这允许进行因此可能更有效的早期干预(例如,治疗)。此外,同一亚组内个体之间的相关染色体相互作用的变化极小。检测染色体相互作用提供非常有用的信息,其中每个基因至多有50种可能的相互作用,因此本发明的方法可以查询500,000个不同的相互作用。

[0032] 优选的标志物集

[0033] 本文中的术语“标志物”或“生物标志物”是指可以在本发明中检测到的特异性染色体相互作用。本文公开了特异性标志物,其中的任何标志物均可用于本发明。另外的标志物集,例如,以本文公开的组合或编号使用。本文表中公开的标志物是优选的。这些可以通过任何合适的方法,例如,本文公开的基于PCR或基于探针的方法,包括qPCR方法分型。这些标志物在本文中通过定位或通过探针和/或引物序列进行定义。

[0034] 表观遗传相互作用的位置和原因

[0035] 表观遗传的染色体相互作用可能重叠并包括显示为编码相关或未描述的基因的染色体区域,但同样可能处于基因间区域内。还应当注意的是,本发明人已经发现,在确定染色体位点状态时,所有区域中的表观遗传学相互作用同等重要。这些相互作用不一定在位于位点的特定基因的编码区中,而可能在基因间区域中。

[0036] 在本发明中检测到的染色体相互作用可能是由于基础DNA序列的变化、环境因素、

DNA甲基化、非编码反义RNA转录物、非诱变致癌物、组蛋白修饰、染色质重塑和特异性局部DNA相互作用引起的。导致染色体相互作用的变化可以由基础核酸序列的变化引起的,其本身并不直接影响基因产物或基因表达方式。这样的改变可以是,例如,基因内和/或外的SNP,基因融合和/或基因间DNA、microRNA和非编码RNA的缺失。例如,已知约20%的SNP处于非编码区域内,并因此描述的过程在非编码情况下也是有益的。在一个实施方式中,染色体聚集在一起形成相互作用的区域在同一染色体上相距小于5kb、3kb、1kb、500个碱基对或200个碱基对。

[0037] 检测到的染色体相互作用优选处于表1或表5中提及的任何基因内。然而,其也可以在处于该基因的上游或下游,例如,距离该基因或距离编码序列上游或下游至多达50,000,至多达30,000,至多达20,000,至多达10,000或至多达5000个碱基。

[0038] 亚组,诊断和个性化治疗

[0039] 本发明的目的是允许检测与ALS或亨廷顿舞蹈症亚组有关的染色体相互作用。因此,该方法可以或可以不用于诊断ALS或亨廷顿舞蹈症。该方法可以或可以不用于ALS或亨廷顿舞蹈症的预后。

[0040] 当该方法用于诊断ALS时,与表1有关的标志物的分型是优选的(即,公开的特异性标志物以及表1中公开的基因、区域和侧接区域中的那些)。在与诊断有关的一个实施方式中,仅与表1有关的标志物被分型,而其他标志物没有分型。在与诊断有关的其他实施方式中,与表5有关(例如,如由所列出的探针序列代表)的至少1、2、3、4、5种或更多标志物未分型。

[0041] 当该方法用于预后时,与表5有关的标志物的分型是优选的(即,公开的特异性标志物以及表5中公开的基因、区域和侧接区域中的那些)。在与预后有关的一个实施方式中,仅与表5有关的标志物被分型,而其他标志物未分型。在与预后有关的其他实施方式中,与表1有关(例如,如由所列出的探针序列代表)的至少1、2、3、4、5种或更多标志物未进行分型。

[0042] 通常,“预后”与ALS的进展相关,并允许将个体分成进展速率亚组。进展情况可以使用ALS-FRS-R评分(ALS功能分级量表)测量,并且可以将个体分类为高于或低于某值,例如,高于或低于每30天0.5点下降。这允许执行预测性预后。

[0043] 如本文所用,“亚组”优选是指群体亚组(群体中的亚组),更优选特定动物如特定真核生物或哺乳动物(例如,人,非人,非人灵长类,或啮齿类动物,例如,小鼠或大鼠)群体亚组。最优选的是,“亚组”是指人类群体的亚组。

[0044] 本发明包括检测和治疗群体中的特定亚组。本发明人已经发现,染色体相互作用在给定群体的子集(例如,至少两个子集)之间是不同的。鉴定出这些差异将使医生可以将其患者分类为该方法中描述的群体一个子集的部分。因此,本发明为医师提供了对于患者基于其表观遗传染色体相互作用的个性化药物的方法。

[0045] 在一个实施方式中,亚组(例如,与预后有关)由ALS功能评分级量表定义,例如,如Cedarbaum, J. M., Stambler, N., Malta, E., Fuller, C., Hilt, D., Thurmond, B. and Nakanishi, A. (1999) The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. BDNF ALS Study Group (Phase III), Journal of the Neurological Sciences. 169(1-2), 13-21中描述的。

[0046] 在一个实施方式中,亚组(例如,与预后有关)由用力肺活量(forced vital capacity) (FVC)定义,例如,如Talakat,N.S.,Pradhan,C.,Nalini,A.,Thennarasu,K.and Raju T.R. (2009) Assessment of Pulmonary Function in Amyotrophic Lateral Sclerosis, Indian J Chest Dis Allied Sci. 51(2):87-91中描述的。

[0047] 生成连接的核酸

[0048] 本发明的某些实施方式利用连接的核酸,特别是连接的DNA。这些包含来自在染色体相互作用中聚集在一起的两个区域的序列,并因此提供有关相互作用的信息。本文描述的EpiSwitch™方法使用此类连接核酸的生成而检测染色体相互作用。

[0049] 因此,本发明的方法可以包括通过以下步骤产生连接的核酸(例如,DNA)的步骤(包括包含这些步骤的方法):

[0050] (i) 存在于染色体位点上的表观遗传染色体相互作用的交联,优选在体外;

[0051] (ii) 可选地从所述染色体位点分离交联的DNA;

[0052] (iii) 通过例如用至少将其切割一次的酶(特别是在所述染色体位点内至少切割一次的酶)进行限制性消化而使所述交联的DNA经受切割;

[0053] (iv) 连接所述剪切的交联的DNA末端(具体地以形成DNA环);和

[0054] (v) 可选地,特别是使用如PCR(聚合酶链反应)的技术,鉴定所述连接的DNA和/或DNA环的存在,以鉴定特异性染色体相互作用的存在。

[0055] 对于本文提到的任何实施方式,可以执行这些步骤以检测染色体相互作用,如用于确定该个体是否是ALS或亨廷顿舞蹈症亚组的部分。还可以进行这些步骤以产生本文提及的第一和/或第二组核酸。

[0056] PCR(聚合酶链反应)可以用于检测或鉴定出连接的核酸,例如,产生的PCR产物的尺寸可以指示特异性染色体相互作用的存在,并因此可以用于鉴定该位点的状态。在优选的实施方式中,PCR反应中使用了至少1、2、3、4、5、6、7或8种表2或7中所示的引物或引物对。在其他优选的实施方式中,PCR反应中使用了至少1、2、3、4、5、6、7或8种任何其他表中所示的引物或引物对。本领域技术人员将意识到许多限制性内切酶都可以用于切割目标染色体位点内的DNA。显而易见的是,使用的具体酶将取决于研究的位点和位于其中的DNA的序列。如本发明可以用于切割DNA的限制酶的非限制性实例是TaqI。

[0057] 如EpiSwitch™技术的实施方式

[0058] EpiSwitch™技术还涉及微阵列EpiSwitch™标志物数据在检测表型特异性的表观遗传构象特征中的用途。以本文描述的方式利用连接的核酸的实施方式如EpiSwitch™,具有多个优点。它们具有低水平的随机噪声,例如,因为来自本发明的第一组核酸的核酸序列与第二组核酸杂交或未能杂交。这提供了二元的结果,允许在表观遗传水平上采用相对简单的方法测量复杂的机制。EpiSwitch™技术还具有快速的处理时间和低成本。在一个实施方式中,该处理时间为3小时至6小时。

[0059] 样品及样品处理

[0060] 本发明的方法通常将在样品上进行。该样品通常包含个体的DNA。其通常包含细胞。在一个实施方式中,样品通过微创手段获得,并且可以,是例如血液样品。DNA可以进行提取,并用标准限制酶切割。这可以预先确定保留哪些染色体构象,并用EpiSwitch™平台检测。由于组织和血液之间的染色体相互作用(包括水平转移)是同步的,血液样本可以用于

检测组织,如与疾病相关的组织中的染色体相互作用。对于某些疾病,如癌症,由于突变引起的遗传噪声可以影响相关组织中的染色体相互作用“信号”,因此使用血液是有利的。

[0061] 本发明核酸的性质

[0062] 本发明涉及某些核酸,如本文中描述为在本发明的方法中使用或产生的连接核酸。这些可以与本文提到的第一和第二核酸相同或具有其任何性质。本发明的核酸通常包含两个部分,各自包含来自在染色体相互作用中聚集在一起的染色体的两个区域之一的序列。通常,每个部分的长度为至少8、10、15、20、30或40个核苷酸,例如,长度为10至40个核苷酸。优选的核酸包含来自任何表中提及的任何基因的序列。通常优选的核酸包含表1或表5中提及的特异性探针序列;或此序列的片段和/或同源物。其他优选的核酸包含表10、11或12中提及的特异性探针序列;或此序列的片段和/或同源物。优选地,核酸是DNA。应该理解的是,在提供特异性序列的情况下,本发明可以使用具体实施方式中所需的互补序列。

[0063] 表2或7中所示的引物也可以用于本文提及的本发明中。在一个实施方式中,使用的引物包含以下的任一种:表2或7中所示的序列;或表2或7中所示任何序列的片段和/或同源物。

[0064] 第二组核酸-“索引”序列

[0065] 第二组核酸序列具有作为一组索引序列的功能,并且实质上是适合于鉴定亚组特异性序列的一组核酸序列。它们可以代表“背景”染色体相互作用,并且可能以某种方式被选择或未被选择。它们通常是所有可能的染色体相互作用的子集。

[0066] 该第二组核酸可以通过任何合适的方法得出。它们可以通过计算得出,或它们可以基于个体中的染色体相互作用得出。它们通常代表比第一组核酸更大的群体组。在一个具体实施方式中,第二组核酸代表一组特异性基因中所有可能的表观遗传染色体相互作用。在另一个具体实施方式中,第二组核酸代表存在于本文描述的群体中的所有可能的表观遗传染色体相互作用的大部分。在一个具体实施方式中,第二组核酸代表至少20、50、100或500个基因,例如,20至100或50至500个基因中的至少50%或至少80%的表观遗传染色体相互作用。

[0067] 第二组核酸通常代表至少100个修饰、调节或以任何方式介导群体中的疾病状态/表型的可能的表观遗传染色体相互作用。第二组核酸可以代表影响物种中疾病状态(通常与诊断或预后相关)的染色体相互作用。第二组核酸通常包含代表与ALS亚组相关和不相关的表观遗传相互作用的序列。

[0068] 在一个具体实施方式中,第二组核酸至少部分源自群体中的天然序列,并通常通过计算机方法获得。与天然核酸中存在的核酸相应部分相比,核酸可以进一步包含单个或多个突变。突变包括一个或多个核苷酸碱基对的缺失、取代和/或添加。在一个具体实施方式中,第二组核酸可以包含代表与天然物种中存在的核酸的相应部分具有至少70%序列同一性的同源物和/或直接同源物(orthologue)的序列。在另一个具体实施方式中,提供了与天然物种中存在的核酸的相应部分的至少80%的序列同一性或至少90%的序列同一性。

[0069] 第二组核酸的性质

[0070] 在一个具体实施方式中,在第二组核酸中存在至少100个不同的核酸序列,优选至少1000、2000或5000个不同的核酸序列,具有至多达100,000、1,000,000或10,000,000个不同的核酸序列。典型的数目是100至1,000,000,如1,000至100,000个不同的核酸序列。全部

或至少90%或至少50%或这些将对应于不同的染色体相互作用。

[0071] 在一个具体实施方式中,第二组核酸代表至少20个不同位点或基因,优选至少40个不同位点或基因,而更优选至少100,至少500,至少1000或至少5000个不同的位点或基因,如100到10,000个不同的位点或基因中的染色体相互作用。第二组核酸的长度适合于它们根据Watson Crick碱基配对与第一组核酸特异性杂交,以允许鉴定亚组特异性的染色体相互作用。通常,第二组核酸将包含在序列上对应于在染色体相互作用中聚集在一起的两个染色体区域的两个部分。第二组核酸通常包含长度至少为10个碱基,优选20个碱基,并且还优选30个碱基(核苷酸)的核酸序列。在另一个实施方式中,该核酸序列的长度可以是最多500,优选最多100,并且优选最多50个碱基对。在优选的实施方式中,第二组核酸包含17至25个碱基对的核酸序列。在一个实施方式中,第二组核酸序列的至少100%、80%或50%具有如上描述的长度。优选不同核酸不具有任何重叠序列,例如,至少100%,90%,80%或50%的核酸在至少5个连续核苷酸内不具有相同序列。

[0072] 假设第二组核酸充当“索引”,则同一组第二核酸可以与不同组的代表具有不同特征的亚组的第一核酸一起使用,即第二组核酸可以代表可以用于识别与不同特征相关的染色体相互作用的“普适”的核酸集。

[0073] 第一组核酸

[0074] 第一组核酸通常来自与ALS或亨廷顿舞蹈症的诊断或预后相关的亚组。第一核酸可以具有本文提及的第二组核酸的任何特征和性质。第一组核酸通常衍生自经过本文描述的处理和加工,特别是EpiSwitch™交联和切割步骤的个体的样品。通常,该第一组核酸代表采集自个体的样品中存在的染色体相互作用的全部或至少80%或50%。

[0075] 通常,与第二组核酸代表的染色体相互作用相比,第一组核酸代表横跨第二基因组所代表的位点或基因的染色体相互作用的较小的群体,即第二组核酸代表限定的位点或基因中的相互作用的背景或索引集。

[0076] 核酸库

[0077] 本文提及的任何类型的核酸群体可以以包含至少200,至少500,至少1000,至少5000或至少10000种该类型的不同核酸的库的形式存在,如“第一”或“第二”核酸。这种库可以是以结合至阵列的形式。

[0078] 杂交

[0079] 本发明需要使得来自第一组核酸和第二组核酸的全部或部分互补核酸序列杂交的方法。在一个实施方式中,在单个试验中,即在单个杂交步骤中,所有第一组核酸与所有第二组核酸接触。然而,可以使用任何合适的试验。

[0080] 标记的核酸和杂交方式

[0081] 本文提及的核酸可以优选使用有助于检测成功杂交的独立标记如荧光团(荧光分子)或放射活性标记进行标记。某些标记可以在UV光下检测。例如,在本文描述的阵列上的杂交模式代表了两个亚组之间表观遗传染色体相互作用的差异,因此提供了比较表观遗传染色体相互作用并确定哪些表观遗传染色体相互作用对本发明的群体中的亚组具有特异性的方法。

[0082] 术语“杂交模式”广泛涵盖第一和第二组核酸之间的杂交的存在和不存在,即第一组的哪些特定核酸与第二组的哪些特定核酸杂交,因此不限于任何具体的试验或技术,或

具有在其上可以检测“模式”的表面或阵列的需要。

[0083] 选择具有特定特征的亚组

[0084] 本发明提供了一种方法,该方法包括检测染色体相互作用的存在或不存在,通常5至20或5至500个这种相互作用,优选20至300或50至100个相互作用,以确定与个体中的ALS或亨廷顿舞蹈症相关的特征的存在或不存在。优选地,染色体相互作用是本文提到的任何基因中的那些。在一个实施方式中,分型的染色体相互作用是由表1或5中的核酸代表的那些。优选地,分型的染色体相互作用是由表10、11或12中的核酸代表的那些。表中标题为“检测的环”的栏代表每个探针检测到哪个亚组(ALS或对照)。可以看出作为该测试的部分,本发明的方法可以检测到ALS亚组和/或对照亚组(非ALS)。

[0085] 测试的个体

[0086] 测试的个体的物种的实例来自本文中提到的那些。另外,本发明的方法中测试的个体可以以某种方式选择。例如,该个体可能易患ALS或亨廷顿舞蹈症。

[0087] 优选的基因区域、位点、基因和染色体相互作用

[0088] 对于本发明的所有方面,优选的基因区域、位点、基因和染色体相互作用在表中,例如,在表1和5中都提及。通常,在本发明的方法中,从表1或表5中列出的至少1、2、3、4、5、6、7或8个相关基因中都检测到染色体相互作用。优选地,检测通过表1或表5中的探针序列表示的至少1、2、3、4、5、6、7或8个相关特异性染色体相互作用的存在或不存在。该染色体相互作用可以处于本文提及的任何基因的上游或下游,例如,上游50kb或下游20kb,例如,从该编码序列开始。

[0089] 优选地,对表1中的至少5、8、10、15种或所有染色体相互作用进行分型。

[0090] 通常实施方式中,对表5中至少5、7、8种或所有染色体相互作用进行分型。

[0091] 优选地,对表10中至少4、6或所有染色体相互作用进行分型。

[0092] 通常,对表11中至少1、2种或所有染色体相互作用进行分型。

[0093] 优选地,对表12中至少4、6种或所有染色体相互作用进行分型。

[0094] 染色体相互作用可以进行分型以确定疾病的存在。染色体相互作用可以进行分型以确定该个体病情进展是快还是慢。

[0095] 表中公开的特定探针和引物序列(或包括片段和同源物的衍生物)可以用于本文公开的任何分型方法。提供了它们在诊断或预后方法中的用途。

[0096] 在一个实施方式中,位点(包括检测到染色体相互作用的基因和/或位置)可以包含CTCF结合位点。这是能够结合转录阻遏物CTCF的任何序列。该序列可以包含可以存在于位点处的1、2或3个拷贝中的序列CCTCC或由其组成。该CTCF结合位点序列可以包含序列CCGCGNGGNGCAG(按照IUPAC表示法)。该CTCF结合位点可以处于染色体相互作用的至少100、500、1000或4000个碱基之内,或处于表1或5所示的任何染色体区域之内。

[0097] 在一个实施方式中,检测的染色体相互作用存在于表1或5所示的任何基因区域。在该方法中检测到连接核酸的情况下,则表1或5中的任何探针序列所示的序列都可能检测到。

[0098] 因此,通常可以检测到来自探针的两个区域(即,来自染色体相互作用的两个位点)的序列。在优选的实施方式中,在该方法中使用的探针包含与任何表中所示的探针相同或互补的序列或由其组成。在一些实施方式中,使用的探针包含与表中所示的任何探针序

列同源的序列。

[0099] 本文中提供的表格

[0100] 表1和表5显示了代表与ALS相关的染色体相互作用的探针 (EpiSwitch™标志物) 数据和基因数据。该探针序列显示了可以用于检测由在染色体相互作用中聚集在一起的基因区域的两个位点产生的连接产物的序列, 即, 该探针将包含与连接产物中的序列互补的序列。前两组开始-结束位置显示探针位置, 且后两组开始-结束位置显示相关的4kb区域。以下信息提供于该探针的数据表中:

[0101] -HyperG_Stats: 基于超几何富集的参数在位点中发现该数目的显著性 EpiSwitch™标志物的概率的p值

[0102] -探针总数: 在位点上测试的EpiSwitch™构象的总数

[0103] -显著探针数 (Probe Count Sig): 在位点发现具有统计学显著性的EpiSwitch™构象数

[0104] -FDR HyperG: 多重测试 (错误发现率) 校正的超几何p值

[0105] -显著百分比 (percent Sig): 相对于位点上测试的标志物数, 显著性EpiSwitch™标志物的百分比

[0106] -logFC: 表观遗传比率 (FC) 的2为底的对数

[0107] -AveExpr: 探针在所有阵列和通道内的平均log2-表达式

[0108] -T: 温和的t-统计量 (moderated t-statistic)

[0109] -p值: 原始p值

[0110] -adj.p值: 调节后的p值或q值

[0111] -B-B统计量 (单项 (lods) 或B) 是该基因表达不同的对数几率 (log-odds)。

[0112] -FC- 非对数倍数变化

[0113] -FC_1- 以0为中心的非对数倍数变化

[0114] -LS- 二进制值, 其与FC_1值有关。如果FC_1值低于-1.1, 则设置为-1, 且如果FC_1值高于1.1, 则设置为1。如果在这些值之间, 则该值为0

[0115] 表1和表5显示了发现出现相关染色体相互作用的基因。其他表显示相似的数据。位点表中的p值与HyperG_Stats相同 (基于超几何富集参数在位点中发现该数目的显著性 EpiSwitch™标志物的概率)。

[0116] 这些探针设计成距TaqI位点30bp。在PCR的情况下, PCR引物也设计成检测连接产物, 但它们距TaqI位点的位置是不同的。

[0117] 探针位置:

[0118] 起始1-片段1上的TaqI位点上游30个碱基

[0119] 末端1-片段1上的TaqI限制位点

[0120] 起始2-片段2上的TaqI限制位点

[0121] 末端2-片段2的TaqI位点下游30个碱基

[0122] 4kb序列位置:

[0123] 起始1-片段1的TaqI位点上游4000个碱基

[0124] 末端1-片段1上的TaqI限制位点

[0125] 起始2-片段2上的TaqI限制位点

- [0126] 末端2-片段2上TaqI位点下游4000个碱基
- [0127] GLMNET值与拟合整个套索或弹性网正则化的过程有关(λ 设置为0.5(弹性网))。
- [0128] 表1至4涉及ALS的诊断,而表5至9涉及ALS的预后,并且在一个实施方式中,基于表1和4进行与诊断有关的检测,且基于表5至9则进行与预后有关的检测。
- [0129] 表1-LS栏显示1或-1。1代表在ALS病例存在,且-1代表ALS病例不存在。
- [0130] 表5-LS栏显示1或-1。1代表标志物存在于快速进展者中且慢进展者中不存在,且-1代表标志物存在于慢进展者中且快进展者中不存在。
- [0131] 表10和11涉及ALS的预后。表10包括先前表中所示的标志物。表11的“检测到环”栏是指标志物在快速进展者中存在,而且在慢进展者中不存在。
- [0132] 标志物在这些表中参照EpiSwitch 3C方法的产物的相关探针唯一鉴定。在水解探针的情况下,这些位于Taq位点(TCGA)的顶部,并且它们在EpiSwitch相互作用中覆盖了两个基因组位点。其测量与60个碱基阵列探针相同的连接点(序列标记每侧30个碱基),但在两侧均跨过定制长度的序列。表19提供了这个的实例。
- [0133] 样品制备和染色体相互作用检测的优选实施方式
- [0134] 本文描述了制备样品和检测染色体构象的方法。可以使用这些方法的优化版本(非常规版本),例如,如本节描述的。
- [0135] 通常,样品含有至少 2×10^5 个细胞。样品可以含有至多达 5×10^5 个细胞。在一个实施方式中,样品含有 2×10^5 至 5.5×10^5 个细胞。
- [0136] 本文描述了存在于染色体位点的表观遗传染色体相互作用的交联。这可以在细胞裂解发生之前进行。细胞裂解可以进行3至7分钟,如4至6分钟或约5分钟。在一些实施方式中,细胞裂解进行至少5分钟且小于10分钟。
- [0137] 本文描述了用限制酶消化DNA。通常,DNA限制酶切在约 55°C 至约 70°C (例如,约 65°C)下进行约10至30分钟,如约20分钟。
- [0138] 优选地,使用频繁切割限制酶,产生平均片段尺寸至多达4000个碱基对的连接DNA片段。可选地,限制酶导致连接DNA的片段具有约200至300个碱基对,例如,约256个碱基对的平均片段尺寸。在一个实施方式中,典型的片段尺寸为200个碱基对至4,000个碱基对,例如,400至2,000个或500至1,000个碱基对。
- [0139] 在EpiSwitch方法的一个实施方式中,在DNA限制酶切消化步骤和DNA连接步骤之间不进行DNA沉淀步骤。
- [0140] DNA连接描述于本文中。通常,DNA连接进行5至30分钟,例如,约10分钟。
- [0141] 样品中的蛋白质可以酶消化,例如,使用蛋白酶,可选地蛋白酶K。蛋白质可以酶消化约30分钟至1小时,例如,约45分钟的时间段。在一个实施方式中,在蛋白质消化后,例如,蛋白酶K消化之后,没有交联逆转或酚DNA提取步骤。
- [0142] 在一个实施方式中,PCR检测能够检测连接的核酸的单个拷贝,优选地具有连接核酸的存在/不存在的二进制读数。
- [0143] 本发明的方法和用途
- [0144] 本发明的方法可以以不同的方式描述。其可以描述为制备连接核酸的方法,该方法包括:(i)在染色体相互作用中聚集在一起的染色体区域的体外交联;(ii)对所述交联的DNA进行切割或限制性消化切割;(iii)连接所述交联的切割的DNA末端以形成连接的核酸,

其中连接的核酸的检测可以用于确定位点上的染色体状态,并且其中优选:

[0145] -位点可以是表1或5中提到的任何位点、区域或基因,

[0146] -和/或其中染色体相互作用可以是本文提及的任何染色体相互作用或对应于表1或5中公开的任何探针,和/或

[0147] -其中连接产物可以具有或包含 (i) 与表1或5中公开的任何探针序列相同或同源的序列;或 (ii) 与 (ii) 互补的序列。

[0148] 本发明的方法可以描述为检测代表群体中不同亚组的染色体状态的方法,包括:确定在基因组的限定的表观遗传活性 (疾病相关) 区域内存在或不存在染色体相互作用,其中优选:

[0149] -亚组由存在或不存在ALS或与ALS相关的特征 (如预后或进展) 定义,和/或

[0150] -染色体状态可以处于表1或表5提及的任何位点、区域或基因;和/或

[0151] -染色体相互作用可以是表1或5中提到的任何相互作用或对应于该表中公开的任何探针。

[0152] 本发明包括检测表1或表5中提及的任何位点、基因或区域的染色体相互作用。本发明包括本文提及的核酸和探针在检测染色体相互作用中的用途,例如,使用至少1、2、4、6或8种此类核酸或探针检测至少1、2、4、6或8个不同位点或基因中的染色体相互作用。本发明包括使用表2或表7中列出的任何引物或引物对或使用如本文的这些引物的变体 (包含引物序列或包含引物序列的片段和/或同源物的序列) 检测染色体相互作用。

[0153] 当分析染色体相互作用是否于所定义的基因、区域或位置“内”发生时,在相互作用中聚集至一起的染色体的两个部分都处于所定义的基因、区域或位置内,或在一些实施方式中,仅染色体的一部分处于所定义的基因、区域或位置内。

[0154] 本发明的方法确定新疗法的用途

[0155] 染色体相互作用的知识可以用于确定疾病的新疗法。本发明提供了本文定义的染色体相互作用鉴定或设计ALS的新治疗剂 (包括与预后有关的治疗) 的方法和用途。

[0156] 同源物

[0157] 本文涉及多核苷酸/核酸 (例如DNA) 序列的同源物。此类同源物通常,例如,在至少10、15、20、30、100个或更多个连续核苷酸的区域内,或跨越来自染色体参与染色体相互作用的区域的核酸部分,具有至少70%的同源性,至少80%,至少85%,至少90%,至少95%,至少97%,至少98%或至少99%的同源性。该同源性可以基于核苷酸同一性 (有时称为“硬同源性”) 计算。

[0158] 因此,在具体实施方式中,多核苷酸/核酸 (例如DNA) 序列的同源物在本文中参照百分比序列同一性。通常,此类同源物,例如,在至少10、15、20、30、100或更多个连续核苷酸的区域内,或跨越来自染色体参与染色体相互作用的区域的核酸部分,具有至少70%的序列同一性,优选至少80%,至少85%,至少90%,至少95%,至少97%,至少98%或至少99%的序列同一性。

[0159] 例如,UWCG软件包提供了BESTFIT程序,该程序可以用于计算同源性和/或百分比序列同一性 (例如,以其默认设置使用) (Devereux et al. (1984) Nucleic Acids Research 12, p387-395)。PILEUP和BLAST算法可以用于计算同源性和/或百分比序列同一性和/或排列序列 (如鉴定等价或对应序列 (通常以其默认设置)), 例如,如Altschul S.F. (1993) J

Mol Evol 36:290-300;Altschul,S.F.et al.(1990)J Mol Biol 215:403-10中描述的。

[0160] 进行BLAST分析的软件通过国家生物技术信息中心(National Center for Biotechnology Information)公开获得。该算法涉及首先通过在查询序列中确定与数据库序列中相同长度的单词比对时匹配或满足某个正值阈值评分T的长度为W的短单词而确定高分序列对(HSP)。T被称为邻近单词分数阈值(Altschul et al.,同上)。这些最初的邻近单词匹配充当启动搜索以查找包含它们的HSP的种子。单词匹配沿着每个序列在两个方向上延伸,直到可以增加累积比分为止。在以下情况下,将停止在每个方向上单词命中的扩展:累积比分数比其最大实现值减少X的量;由于一个或多个负得分残基比对的累积,累积得分变为零或更低;或到达任一序列的末尾。BLAST算法参数W T和X决定该比对的灵敏度和速度。BLAST程序默认使用11的字长(W),BLOSUM62评分矩阵(请参见Henikoff and Henikoff(1992)Proc.Natl.Acad.Sci.USA 89:10915-10919),比对(B)为50,期望值(E)为10,M=5,N=4,和两条链的比对。

[0161] BLAST算法对两个序列之间的相似性进行统计分析。参见,例如,Karlin and Altschul(1993)Proc.Natl.Acad.Sci.USA 90:5873-5787。BLAST算法提供的一种相似性度量是最小总和概率(P(N)),其提供两个多核苷酸序列之间偶然发生匹配的概率。例如,如果第一序列与第二序列比对的最小总和概率小于约1,优选小于约0.1,更优选小于约0.01,而最优选小于约0.001,则认为该序列与另一个序列相似。

[0162] 同源序列通常相差1、2、3、4个或更多个碱基,如少于10、15或20个碱基(其可以是核苷酸的取代、缺失或插入)。关于计算同源性和/或百分比序列同一性,这些变化可以在上述任何区域中测量。

[0163] “一对引物”的同源性,例如,可以通过将两个序列视为单个序列(好像两个序列连接在一起)计算,其目的是随后与另一个又被认为是单个序列的引物对相比较。

[0164] 阵列

[0165] 第二组核酸可以结合至阵列,并且在一个实施方式中,存在至少15,000、45,000、100,000或250,000个结合至阵列的不同的第二核酸,其优选代表至少300、900、2000或5000个位点。在一个实施方式中,一个或多个或所有的不同的第二核酸群体结合至阵列的一个以上的独立区域,实际上在阵列上重复,允许进行错误检测。该阵列可以基于Agilent SurePrint G3 Custom CGH微阵列平台。第一核酸与该阵列的结合的检测可以通过双色系统进行。

[0166] 治疗剂

[0167] 本文提及了治疗剂。本发明提供了用于在某些个体中预防或治疗ALS的试剂,例如,通过本发明的方法确定的那些。这可以包括向需要的个体给予治疗有效量的试剂。本发明提供了该试剂在制备用于预防或治疗某些个体中的ALS的药物中的用途。

[0168] ALS的优选治疗剂为:

[0169] 利鲁唑(Riluzole)(Rilutek):该药物通常以丸剂的形式服用,并且其通过降低大脑中的信使(谷氨酸)水平而减缓疾病的进展。谷氨酸以较高含量存在于ALS患者。

[0170] 依达拉奉(Edaravone)(Radicava):该药物会减少与ALS相关的日常表现的下降。通常每月一次连续10-14天通过静脉输注将该药物给予患者。

[0171] 阿瑞莫罗(Arimoclomol):该药物在运动神经元中起热休克反应诱导剂的作用,并

防御神经元障碍症和细胞死亡。

[0172] 他仑帕奈(Talampanel):该药物会降低肌肉力量下降和症状发展的速度。

[0173] β -内酰胺抗生素:这些抗体如青霉素和头孢菌素可以通过上调GLT1(神经胶质谷氨酸转运蛋白)的水平维持肌肉稳定性并延长寿命。

[0174] 溴隐亭(Bromocriptine):这种药物是一种自由基清除剂,抑制应激诱导的氧化性细胞死亡。

[0175] 普拉克索(Pramipexole)和右旋普拉克索(Dexpramipexole):这些药物起多巴胺激动剂的作用,并且它们具有自由基清除功能。这些药物参与线粒体功能障碍。右旋普拉克索是普拉克索的光学对映体。

[0176] 干细胞疗法:干细胞生长会减少神经元疾病的进展或替代运动神经元。干细胞具有产生脊髓运动神经元、扩展轴突、接收和与肌肉产生突触的潜能。源自成体干细胞的间充质干细胞(MSC)释放营养因子、抗炎细胞因子和免疫调节趋化因子以延缓疾病进展。

[0177] 免疫疗法:抗体疗法,如通过ICV途径输注D3H5抗体,可以使体重保持更长的时间,并延长ALS的转基因小鼠模型的生命。

[0178] 以下是亨廷顿舞蹈症的治疗方法列表。这些可以帮助减轻一些运动和精神障碍症的症状。

[0179] 运动障碍症的药物

[0180] • 丁苯那嗪(Terabenazine)(Xenazine)

[0181] • 抗精神病药,如氟哌啶醇(haloperidol)(Haldol),氯丙嗪(chlorpromazine),利培酮(risperidone)(Risperdal)和喹硫平(quetiapine)(Seroquel)

[0182] • 其他医药包括金刚烷胺,左乙拉西坦(levetiracetam)(Keppra,其他)和氯硝西泮(clonazepam)(Klonopin)

[0183] 用于精神疾病的药物

[0184] • 抗抑郁药包括西酞普兰(citalopram)(Celexa),依西酞普兰(escitalopram)(Lexapro),氟西汀(flouxetine)(Prozac,Sarafem)和舍曲林(sertraline)(Zoloft)

[0185] • 抗精神病药包括喹硫平(Seroquel),利培酮(Risperdal)和奥氮平(alanzapine)(Zyprexa)

[0186] • 情绪稳定药物包括丙戊酸盐(valproate)(Depacon),卡马西平(carbamazepine)(Carbatrol,Epitol,Tegretol)和拉莫三嗪(lamotrigine)(Lamictal)

[0187] 亨廷顿舞蹈症的早期诊断有助于在疾病过程中管理症状的治疗。

[0188] 药物的制剂(用于ALS或亨廷顿舞蹈症)将取决于药物的性质。制剂将以包含药剂和药用载体或稀释剂的药物组合物的形式提供。合适的载体和稀释剂包括等渗盐溶液,例如,磷酸盐缓冲盐水。典型的口服剂量组合物包括片剂、胶囊剂、液体溶液和液体混悬剂。该药物可以配制用于肠胃外、静脉内、肌内、皮下、透皮或口服给药。

[0189] 制剂的剂量可以根据各种参数,尤其是根据使用的物质;被治疗者的年龄、体重和病症;给药途径;和所需的方案确定。医生将可以确定任何具体药物所需的给药途径和剂量。然而,合适的剂量可以是每天1至3次,例如,0.1至100mg/kg体重,例如,1至40mg/kg体重。

[0190] 本文提到的物质的形式

[0191] 本文提及的任何物质,如核酸或治疗剂,可以是纯化或分离的形式。它们可以以与自然界中发现的不同的形式,例如,它们可以以与自然界中未出现的其他物质组合而存在。核酸(包括本文定义的序列的部分)可以具有与自然界发现的那些不同的序列,例如,在同源性部分中描述的序列中具有至少1、2、3、4或更多个核苷酸变化。该核酸可以在5'或3'末端具有异源序列。核酸在化学上可以与自然界中发现的那些不同,例如,它们可以以某种方式修饰,但优选仍然能够进行沃森-克里克(Watson-Crick)碱基配对。在适当情况下,核酸将以双链或单链形式提供。本发明以单链或双链形式提供本文提及的所有特异性核酸序列,并因此包括所公开的任何序列的互补链。

[0192] 本发明还提供了用于实施本发明的任何方法的试剂盒,包括检测与ALS相关的染色体相互作用(例如,与ALS的诊断或预后有关)。这样的试剂盒可以包括可以检测相关染色体相互作用的特异性结合剂,如可以检测由本发明的方法产生的连接核酸的试剂。试剂盒中存在的优选试剂包括可以与该连接核酸或引物对杂交的探针,例如,如本文中,能够在PCR反应中扩增该连接核酸的探针。

[0193] 本发明还提供了一种可以检测相关染色体相互作用的装置。该装置优选包括可以检测这些染色体相互作用的任何特异性结合剂、探针或引物对,如本文的任何此类结合剂、探针或引物对。

[0194] 检测方法

[0195] 在一个实施方式中,使用了在PCR反应期间活化时可检测的探针对与染色体相互作用有关的连接序列进行定量检测,其中所述连接序列包含来自在表观遗传染色体相互作用中聚集在一起的两个染色体区域的序列,其中所述方法包括在PCR反应期间使连接序列与探针接触,和检测探针的活化程度,并且其中所述探针结合连接位点。该方法通常允许使用双重标记的荧光水解探针以符合MIQE的方式检测特定的相互作用。

[0196] 探针通常用可检测的标记进行标记,该标记具有失活和活化状态,使得仅在活化时才能检测到。活化程度将与PCR反应中存在的模板(连接产物)的程度有关。检测可以在全部或一些PCR期间进行,例如进行至少50%或80%的PCR周期。

[0197] 探针可以包含共价连接于寡核苷酸的一个末端的荧光团,以及连接于该核苷酸的另一末端的猝灭剂,使得荧光团的荧光被该猝灭剂猝灭。在一个实施方式中,荧光团连接至寡核苷酸的5'末端,且猝灭剂共价连接至寡核苷酸的3'末端。本发明的方法中可以使用的荧光团包括FAM、TET、JOE、亚基马(Yakima)黄、HEX、花青3、ATTO 550、TAMRA、ROX、德克萨斯(Texas)红、花青3.5、LC610、LC 640、ATTO 647N、花青5、花青5.5和ATTO680。与合适荧光团可以一起使用的猝灭剂包括TAM、BHQ1、DAB、Eclip、BHQ2和BBQ650,可选地其中荧光团选自HEX、德克萨斯红和FAM。荧光团和猝灭剂的优选组合包括BHQ1+FAM和BHQ2+德克萨斯红。

[0198] 探针在qPCR测试中的用途

[0199] 本发明的水解探针通常是用浓度匹配的阴性对照优化的温度梯度。优选地,单步PCR反应被优化。更优选地,计算出标准曲线。使用跨越该连接序列的连接点结合的特异性探针的优点是无需使用巢式PCR方法就可以实现对该连接序列的特异性。本文描述的方法允许对低拷贝数靶标进行准确而精确的定量。在温度梯度优化之前可以对该靶连接的序列进行纯化,例如,凝胶纯化。靶连接序列可以进行测序。优选使用约10ng,或5至15ng,或10至20ng,或10至50ng,或10至200ng的模板DNA进行PCR反应。设计正向和反向引物使得一个引

物与连接的DNA序列中代表的一个染色体区域的序列结合,且另一个引物与连接的DNA序列中代表的另一个染色体区域结合,例如,通过与该序列互补。

[0200] 连接DNA靶标的选择

[0201] 本发明包括选择适用于本文定义的PCR方法的引物和探针,包括基于其结合和扩增连接序列的能力选择引物,和选择其会结合的靶序列的基于探针序列的特性,特别是靶标序列的曲率。

[0202] 探针通常设计/选择为结合跨越限制位点的并列限制片段的连接序列。在本发明的一个实施方式中,例如,使用本文提及的特定算法,计算出与具体染色体相互作用有关的可能的连接序列的预测曲率。该曲率可以表达为每螺旋的度数,例如,每螺旋 10.5° 。选择连接的序列用于靶向,其中连接的序列具有的曲率倾向峰值得分为每螺旋至少 5° ,通常为每螺旋至少 10° , 15° 或 20° ,例如,每螺旋 5° 至 20° 。优选对至少20、50、100、200或400个碱基,如对该连接位点上游和/或下游的20至400个碱基,计算出每螺旋的曲率倾向得分。因此,在一个实施方式中,该连接产物中的靶序列具有任何这些曲率水平。靶序列还可以基于最低的热力学结构自由能进行选择。

[0203] 优选的ALS实施方式

[0204] 段落1.一种用于检测代表群体中疾病亚组的染色体状态的方法,包括确定在基因组的限定区域内存在或不存在与该染色体状态有关的染色体相互作用,其中疾病亚组是肌萎缩性侧索硬化症(ALS)亚组;并且

[0205] -其中染色体相互作用可选地通过确定哪些染色体相互作用与对应于群体的ALS亚组的染色体状态相关的方法确定,包括使来自具有不同染色体状态的亚组的第一组核酸与第二组索引核酸接触,和使互补序列杂交,其中该第一和第二组核酸中的核酸代表包含来自在染色体相互作用中聚集至一起的两个染色体区域的序列的连接产物,并且其中该第一和第二组核酸之间的杂交模式允许确定哪些染色体相互作用对ALS亚组是特异性的;并且

[0206] -其中染色体相互作用:

[0207] (i) 存在于表1或表5所列的任何区域或基因中的任何一种中;和/或

[0208] (ii) 对应于表1或5中所示的任何探针代表的任何一种染色体相互作用,和/或

[0209] (iii) 存在于包含或侧接(i)或(ii)的4,000个碱基的区域中。

[0210] 段落2.根据段落1的方法,进行该方法以诊断ALS或确定ALS的预后。

[0211] 段落3.根据段落1或2的方法,其中染色体相互作用的特异性组合分型为:

[0212] (i) 包含表1或表5中的探针代表的所有染色体相互作用;或

[0213] (ii) 包含表1或5中的探针代表的至少4、5、6或7个染色体相互作用;或

[0214] (iii) 一起存在于表1或5中所列的至少4、5、6或7个区域或基因中;或

[0215] (iv) 至少4、5、6或7个染色体相互作用分型为存在于包含或侧接表1或5中探针代表的染色体相互作用的4,000个碱基的区域中。

[0216] 段落4,根据前述段落中任一项的方法,其中染色体相互作用:

[0217] -在来自个体的样品中分型,和/或

[0218] -通过检测染色体相互作用的位点上DNA环的存在或不存在来分型,和/或

[0219] -通过检测在染色体构象内聚集在一起的染色体的远侧区域的存在或不存在来分

型,和/或

[0220] -通过检测在所述分型期间产生的并且其序列包含各自对应于在染色体相互作用中聚集在一起的染色体区域的两个区域的连接的核酸的存在来分型,其中所述连接的核酸的检测优选通过使用(i)与表1或表5中提及的任何特定探针序列具有至少70%的同一性的探针,和/或(ii)与表2或表7中的任何引物对具有至少70%的同一性的引物对。

[0221] 段落5.根据前述段落中任一项的方法,其中:

[0222] -第二组核酸来自比第一组核酸更大的个体组;和/或

[0223] -第一组核酸来自至少8个个体;和/或

[0224] -第一组核酸来自第一亚组的至少4个个体和优选与第一亚组不重叠的第二亚组的至少4个个体;和/或

[0225] -进行所述方法以选择用于医学治疗的个体。

[0226] 段落6.根据前述段落中任一项的方法,其中:

[0227] -第二组核酸代表未选择的组;和/或

[0228] -其中第二组核酸在限定的位置上结合至阵列;和/或

[0229] -其中第二组核酸代表至少100种不同基因的染色体相互作用;和/或

[0230] -其中第二组核酸包含至少1,000个不同的核酸,代表至少1,000种不同染色体相互作用;和/或

[0231] -其中第一组核酸和第二组核酸包含至少100个长度为10至100个核苷酸碱基的核酸。

[0232] 段落7.根据前述段落中任一项的方法,其中第一组核酸可以通过包括以下步骤的方法获得:

[0233] (i)在染色体相互作用中聚集在一起的染色体区域的交联;

[0234] (ii)可选地通过用酶的限制性消化切割使所述交联的区域裂解;和

[0235] (iii)连接所述交联的裂解的DNA末端以形成第一组核酸(特别包含连接的DNA)。

[0236] 段落8.根据前述段中任一项的方法,其中优选在5至9个不同的区域或基因中分型至少5至9种不同的染色体相互作用。

[0237] 段落9.根据前述段落中任一项的方法,其中基因组的限定区域:

[0238] (i)包含单核苷酸多态性(SNP);和/或

[0239] (ii)表达微RNA(miRNA);和/或

[0240] (iii)表达非编码RNA(ncRNA);和/或

[0241] (iv)表达编码至少10个连续氨基酸残基的核酸序列;和/或

[0242] (v)表达调节元件;和/或

[0243] (vii)包含CTCF结合位点。

[0244] 段落10.通过选择可以引起染色体相互作用改变从而引起治疗效果的试剂来确定或设计治疗ALS的治疗剂的方法,

[0245] -其中该染色体相互作用由表1或5中的任何探针代表;和/或

[0246] -该染色体相互作用存在于表1或表5所列的任何区域或基因中;

[0247] 并且其中可选地:

[0248] -该染色体相互作用通过确定哪些染色体相互作用与段落1中限定的染色体状态

相关的方法确定,和/或

[0249] -染色体相互作用的变化使用(i)与表1或表5中提及的任何探针序列具有至少70%同一性的探针,和/或(ii)通过与表1或表5中的任何引物对具有至少70%的同一性的引物对监测;和/或

[0250] -候选试剂与细胞接触,并监测该细胞中的染色体相互作用,以确定该候选试剂是否能够治疗ALS。

[0251] 段落11.以下各项确定或设计ALS治疗剂的用途

[0252] (i)检测染色体相互作用,其中:

[0253] -该染色体相互作用通过表1或5中的探针代表,和/或

[0254] -该染色体相互作用存在于表1或表5中提到的任何区域或基因中;或

[0255] (ii)与表1或5中提到的任何探针序列具有至少70%的同一性的探针,或

[0256] (iii)与表2或表7中鉴定的任何引物对具有至少70%同一性的引物对;

[0257] 段落12.根据段落11的确定治疗剂的用途,包括给予候选试剂,和使用染色体相互作用、所述探针或所述引物对的所述检测,以检测是否存在染色体的变化,由此确定所述候选试剂是否是治疗剂,其中所述用途可选地在体外,优选在细胞中进行。

[0258] 段落13.一种用于ALS的治疗剂,用于预防或治疗通过根据段落1至9中任一项的方法确定为需要该治疗剂的个体中的ALS的方法。

[0259] 段落14.根据任何前述段落的过程、方法或用途,其中所述分型或所述检测包括通过使用能够扩增所述连接产物的引物和PCR反应期间结合连接位点的探针的定量PCR(qPCR)特异性检测所述连接产物,其中所述探针包含与来自在染色体相互作用中聚集在一起的所述染色体区域的每个的序列互补的序列,其中优选地,所述探针包含:

[0260] 与连接产物特异性结合的寡核苷酸,和/或

[0261] 共价附连至寡核苷酸5'端的荧光团,和/或

[0262] 共价附连至寡核苷酸3'末端的淬灭剂,并且

[0263] 可选地

[0264] 荧光团选自HEX、德克萨斯红和FAM;和/或

[0265] 探针包含长度为10至40个核苷酸碱基,优选长度为20至30个核苷酸碱基的核酸序列。

[0266] 具体的实施方式

[0267] 在一个实施方式中,仅染色体内相互作用进行分型/检测,而不进行染色体外相互作用(在不同染色体之间)分型。

[0268] 出版物

[0269] 本文提及的所有出版物的内容通过引用结合于本说明书中,并可以用于进一步定义与本发明有关的特征。

[0270] 特定的实施方式

[0271] EpiSwitch™平台技术检测位点上正常和异常状况之间调节变化的表观遗传调节特征。EpiSwitch™平台确定和监测与人类染色体的调控的高阶结构(也称为染色体构象特征)相关的基因调控的基本表观遗传水平。染色体特征是基因去调控的级联反应中独特的主要步骤。它们是高阶生物标志物,相对于利用后期表观遗传和基因表达生物标志物如DNA

甲基化和RNA分析的生物标志物平台,具有独特的优势。

[0272] EpiSwitch™阵列测试

[0273] 该定制EpiSwitch™阵列筛选平台具有4种密度的15K、45K、100K和250K独特染色体构象,每个嵌合片段在阵列上重复4次,分别产生60K、180K、400K和1百万的有效密度。

[0274] 定制设计的EpiSwitch™阵列

[0275] 15K EpiSwitch™阵列可以筛选整个基因组,包括使用EpiSwitch™生物标志物发现技术查询的约300个位点。EpiSwitch™阵列构建于Agilent SurePrint G3定制CGH微阵列平台上;该技术提供60K、180K、400K和1百万的4种密度的探针。由于每个EpiSwitch™探针以一式四份出现,则每个阵列的密度降低至15K、45K、100K和250K,从而允许对可重复性进行统计评价。每个遗传位点查询的平均潜在EpiSwitch™标志物数为50;因此,可以研究的位点数量为300、900、2000和5000。

[0276] EpiSwitch™定制阵列管道

[0277] EpiSwitch™阵列是一种双色系统,在生成EpiSwitch™库后,一组样品标记于Cy5中,而另一组样品(对照)在Cy3中待比对/分析。该阵列使用Agilent SureScan扫描仪扫描,并使用Agilent Feature Extraction软件提取所得特征。然后使用R中的EpiSwitch™阵列处理脚本处理数据。使用R:Limma*中的Bioconductor中的标准双色软件包对该阵列进行处理。该阵列的标准化使用Limma*中的normalisedWithinArrays函数完成,并且该操作也对芯片上的Agilent阳性对照和EpiSwitch™阳性对照进行。基于Agilent标志调用(Flag call),对数据进行过滤,移除Agilent对照探针,并对技术性重复探针进行平均,以使用Limma*对其进行分析。这些探针基于两个方案之间的差异比较建模,并随后通过使用错误发现率(False Discovery Rate)校正。变异系数(Coefficient of Variation) (CV) $\leq 30\%$ 且 ≤ -1.1 或 ≥ 1.1 且通过 $p \leq 0.1$ FDR p值的探针用于进一步筛选。为了减少探针集,使用R中的FactorMineR软件包进行进一步的多因素分析。

[0278] *注意:LIMMA是用于评价微阵列实验中差异性表达的线性模型和经验贝叶斯过程。Limma是一种R软件包,用于分析由微阵列或RNA-Seq生成的基因表达数据。

[0279] 首先基于调整的p-值、FC和CV $<30\%$ (任意截止点) 参数选择探针池以进行最终挑选。进一步的分析和最终列表仅基于前两个参数(adj.p-值;FC)得出。

[0280] 本发明通过以下非限制性实施例说明。

[0281] 实施例1

[0282] 统计管道

[0283] 使用R中的EpiSwitch™分析软件包对EpiSwitch™筛选阵列进行处理,以选择高值EpiSwitch™标志物翻译至EpiSwitch™ PCR平台。

[0284] 步骤1

[0285] 基于校正的p值(错误发现率,FDR)选择探针,这是修改的线性回归模型的产物。选择低于p-值 ≤ 0.1 的探针,然后进一步降低其表观遗传比(ER),探针ER必须为 ≤ -1.1 或 ≥ 1.1 ,以选择用于进一步分析。最后一个过滤器是变异系数(CV),探针必须低于 ≤ 0.3 。

[0286] 步骤2

[0287] 对统计列表中的前40个标志物基于其选择作为PCR翻译的标志物的ER进行选择。带有最大负ER负载的前20个标志物和具有最大正ER负载的前20个标志物构成该列表。

[0288] 步骤3

[0289] 来自步骤1的所得标志物(具有统计意义的探针)构成使用超几何富集(HE)的富集分析的基础。该分析使得标志物从显著性探针列表中减少,并与步骤2中的标志物一起构成翻译至EpiSwitch™PCR平台的探针列表。

[0290] 通过HE对统计探针进行处理,以确定哪些遗传位置具有统计显著性探针的富集,表明哪些遗传位置是表观遗传差异的中心。

[0291] 基于校正的p值的最显著性富集位点选出用于生成探针列表。选择出p值低于0.3或0.2的遗传位置。映射至这些遗传位置的统计探针与来自步骤2的标志物一起构成EpiSwitch™PCR翻译的高值标志物。

[0292] 阵列设计与处理

[0293] 阵列设计

[0294] 1.使用SII软件(当前为3.2版)处理遗传位点以:

[0295] a.在这些特异性遗传位点上提取基因组的序列(上游50kb且下游20kb的基因序列)

[0296] b.定义此区域内的序列参与CC的概率

[0297] c.使用特异性RE切割该序列

[0298] d.确定哪些限制片段可能在某方向上相互作用

[0299] e.对不同CC一起相互作用的可能性进行排序。

[0300] 2.确定阵列尺寸,并因此确定可用的探针位置数(x)

[0301] 3.抽出x/4的相互作用。

[0302] 4.为每个相互作用定义部分1的限制性酶切位点的30bp序列和部分2限制性酶切位点的30bp序列。检查那些区域是否重复,如果不重复,则排除,并在列表上记录下一相互作用。将两个30bp连接在一起以定义探针。

[0303] 5.创建x/4探针加限定的的对照探针的列表,并复制4以创建要在阵列上创建的列表

[0304] 6.将探针列表上传到定制CGH阵列的Agilent Sure设计网站上。

[0305] 7.使用探针组设计Agilent定制CGH阵列。

[0306] 阵列处理

[0307] 1.使用EpiSwitch™标准操作程序(SOP)处理样品用于模板生产。

[0308] 2.通过阵列处理实验室采用乙醇沉淀净化。

[0309] 3.按照Agilent SureTag完整DNA标记试剂盒-用于血液、细胞或组织的基因组DNA分析酶标记的Agilent寡核苷酸阵列基CGH对样品进行处理。

[0310] 4.使用Agilent特征提取软件使用AgilentC扫描仪进行扫描。

[0311] EpiSwitch™技术概述

[0312] EpiSwitch™平台为与异常和反应性基因表达相关的主要疾病的筛选、早期检测、伴随诊断、监测和预后分析提供了高效的手段。这种方法的主要优点是其无创且快速,并依赖于高度稳定的基于DNA的靶标作为染色体特征的部分,而不是不稳定的蛋白质/RNA分子。

[0313] EpiSwitch™生物标志物特征在复杂疾病表型分级中证实具有高度的鲁棒性、敏感性和特异性。该技术利用了表观遗传学科学中的最新突破,监测和评价作为高度信息性表

观遗传生物标志物的染色体构象特征。在学术环境中部署的当前研究方法需要3到7天才能对细胞材料进行生化处理,以检测CCS。这些工作程序的敏感性和可重复性是有限的;而且,在设计阶段无法享受EpiSwitch™分析软件包提供的有针对性的洞悉的受益。

[0314] 计算机标志物鉴定中的EpiSwitch™阵列

[0315] 通过EpiSwitch™阵列直接对来自测试群体的临床样本进行基因组上的CCS位点评价,以鉴定所有相关的分级先导生物标志物。EpiSwitch™阵列平台由于其高通量能力,并且快速筛选大量位点的能力,而因此适用于标志物鉴定。使用的阵列是Agilent定制CGH阵列,其允许查询通过计算机软件鉴定出的标志物。

[0316] EpiSwitch™ PCR

[0317] 然后,通过EpiSwitch™ PCR或DNA测序仪(即Roche 454,Nanopore MinION等)验证由EpiSwitch™阵列确定的潜在标志物。选择具有统计学显著性并显示出最佳重现性的顶级PCR标志物用于进一步减少为最终的EpiSwitch™特征集,并在独立的样品组中验证。EpiSwitch™ PCR可以由训练有素的技术人员按照已建立的标准化操作规程进行。试剂的所有操作流程和制造均按照ISO 13485和9001认证进行,以确保工作质量和传输操作流程的能力。EpiSwitch™ PCR和EpiSwitch™阵列生物标志物平台与全血和细胞系的分析相容。这些测试足够灵敏,才能使用少量血液检测极低拷贝数中的异常。

[0318] ALS队列分析

[0319] 本发明人已经使用表观遗传染色体相互作用作为确定用作ALS中伴随诊断方法的生物标志物的基础。EpiSwitch™生物标志物发现平台是由本发明人开发的,用于检测表观遗传学调控特征变化,如驱动ALS中隐含的表型变化。EpiSwitch™生物标志物发现平台确定CCS,其定义了将环境线索整合到表观遗传和转录机制中的初始调控过程。因此,CCS是一系列基因调控的关键步骤。EpiSwitch™生物标志物发现平台分离出的CCS具有许多已充分证明的优点:严格的生化和生理稳定性;它们的二元性质和读数;和其在基因调控的真核级联反应中的关键位置。

[0320] 在基因表达的系统表观遗传控制中检测干扰的能力既允许早期诊断ALS也允许确立患者的预后。从健康和患病患者的血液样品对细胞调节基因组架构进行的比较研究揭示了两个ALS相关的表观遗传学特征:一个具有诊断潜力,而另一个具有预后预测能力。对于这项前瞻性研究,使用EpiSwitch™(由Oxford BioDynamics开发的高通量平台)分析了牛津大学纳菲尔德分校临床神经科学系(Nuffield Department of Clinical Neurosciences (NDCN) at the University of Oxford) (NDCN) 的牛津运动神经元疾病诊所(Oxford Motor Neuron Disorders Clinic)的临床研究小组收集的样品,用于监测染色体构象特征。这项研究比较了ALS-FRS-R、用力肺活量(FVC)和其他临床观察的临床注释,以分配ALS进行性亚型,并分析了表观遗传学特征。

[0321] 共有100名患者进入牛津运动神经元疾病诊所,参加了该研究,并被要求在3和6个月后返回。收集对照(n=100)。在每次访问期间,参与者进行ALS-FRS-R和FVC测试,并提供血液样品。分析样品以确定ALS疾病相关的诊断特征或与预后疾病相关的特征(在3和6个月时)。在0、3和6个月时将临床评价结果与EpiSwitch™分析比较。每月减少ALS-FRS-R得分0.5分的临界值用于将ALS患者分为进展亚型。

[0322] 基于ALS-FRS-R分数的下降速率,3个月样品和6个月样品的初步结果表明,表观遗

传学特征选择的较快(>0.5)和较慢(<0.5)进展的ALS患者具有80%的敏感性和特异性。迄今为止,该结果表明,预后特征对于跨时间选择ALS的亚型具有鲁棒性。表中所示的结果与75个ALS和75个对照样品(n=150)的诊断和预后分析有关,并具有独立的样品队列(n=50)。

[0323] 进一步的结果(包括表11)基于100名患有ALS的患者,这些患者在基线、3个月和6个月生存期通过ALSFRS-R评分和FVC(用力肺活量)的标准惯例进行观察和评价。基于基线和3个月的测量,这些患者分为慢或快ALS(标准分类)。仅使用我们基线处的三种标志物读数,我们也将所有患者分为慢或快(EpiSwitch分类)。然后这些患者基于6个月生存率进行评价,并比较慢速和快速组的生存率。在慢组中,标准分类具有显著数量的致命结局,而在统计学上,慢组和快组之间的生存率差异无统计学显著性, $p=0.052$ 。通过EpiSwitch分类,6个月时慢和快之间的差异是高度显著的, $p=0.0097$,且大多数死亡发生于快组中。

[0324] 实施例2

[0325] HTT位点的基因组结构差异是亨廷顿舞蹈症的症状和症状前病例的基础。

[0326] 亨廷顿舞蹈症(HD)是一种进行性神经退行性疾病,是一种遗传性病症,导致成人大脑中神经元大量变性,最终导致死亡。HD的根本原因是“亨廷顿基因”(HTT)中三核苷酸胞嘧啶-腺嘌呤-鸟嘌呤(CAG)重复序列增加,这导致产生具有形成细胞内聚集体并导致神经元死亡的趋势的突变亨廷顿蛋白。重要的是,CAG重复的数量与疾病的发作和严重程度相关,并且具有39个以上CAG重复的患者将在其生命中的某个时刻发展为HD,个体内疾病的发作可能会有数十年的差异,并且对这种症状前阶段的了解知之甚少。随着表观遗传学方法的进展以及检测DNA甲基化、组蛋白修饰、非编码RNA和染色体构象特征的表观遗传学调控标志物的新技术作为了解基因组结构的一部分的出现,我们对围绕HTT基因周围的结构及其与疾病表现的关系分析基因组中的系统调控变化感兴趣。本文中,我们检查了症状前HD患者和有症状HD患者与未患病对照相比在HTT位点上的可检测的系统性无创表观遗传学差异。

[0327] 方法:使用来自HD患者和健康对照的血液样品,我们使用了EpiSwitch™,经过验证的高分辨率工业平台,进行染色体构象检测,以评价HTT基因紧邻区域的染色质结构。我们评价了横跨225kb的HTT位点内的20个相互作用位点上的疾病性稳定染色质构象的不存在或存在,并比较了健康对照组、验证的有症状HD患者(CAG, $n>39$)和遗传验证的CAG扩展而未表现出HD临床症状的患者之间所产生的染色体构象。

[0328] 结果:在外周血样品中测试时,在整个患者组中观察到一致且稳定的染色体构象。我们发现HD中存在两种基本相互作用(在所有患者组和对照中均发生)和七个疾病性相互作用,而在健康对照中则没有。最重要的是,我们观察到三种疾病性相互作用仅在表现出临床症状的HD患者(有症状的病例)中存在,而在症状前病例中没有。有症状HD组中85%(7名患者中的6名)的患者表现出至少一种与有症状HD相关的特异性染色体构象。

[0329] 结论:我们的结果是第一个表明HD患者中HTT位点的调控染色质结构发生了系统性改变的证据。此外,HTT位点的染色质结构还显示出有症状和症状前HD患者临床阶段之间的疾病性差异。鉴于具有分子工具评件HD疾病进展的高度临床实用性,这些结果强烈表明,对系统性染色体构象特征(CCS)的无创评价可以是对HD患者预后评价的宝贵补充。

[0330] 概述

[0331] 亨廷顿舞蹈症 (HD) 是一种神经退行性疾病,其细胞特征在于基底神经节中神经元的丢失,并且临床上特征在于失控的运动、情绪问题和认知缺失。HD是常染色体显性遗传疾病,并且尽管患病率在很大程度上取决于地理和种族,但认为仅在美国和欧洲就影响了超过50,000人。潜在的遗传原因是亨廷顿基因 (HTT) 中的三核苷酸CAG扩增,这是1983年由麻萨诸塞州总医院 (Massachusetts General Hospital) 的James Gusella作为遗传标志物发现的,这种扩增会导致具有毒性聚谷氨酰胺 (polyQ) 序列的突变亨廷顿蛋白 (mHTT) 的产生。然而,尽管进行了数十年的研究和临床试验,但尚未开发出成功的疗法。HD的“典型”发作在40至50岁之间,但高达15%的病例发病很晚,直到60岁以后才显示出临床症状。在最近一项研究迟发性HD (LoHD) 病例的研究的统合分析中,超过90%的患者具有 ≤ 44 的CAG重复长度。在HD中更感兴趣的观察结果之一是,尽管polyQ重复序列的长度与疾病的发作和严重程度之间存在众所周知的相关性,但各个体患者之间存在很大的差异。例如,对于中程重复长度 (此处定义为40到50) 的患者,任何各个患者的疾病发作可能相差60年。这意味着,许多易患疾病并属于poly-Q道携带者的患者可以在“症状前”状态下生活数十年。目前尚不清楚控制临床症状发作的因素,并使HD患者的预后评价复杂化。

[0332] 尽管在历史上认为是单基因疾病,但对HD的基本病理学的广泛研究表明,导致疾病发作和进展的机制比原先想象的要更复杂。许多不同的技术已经用于研究HD疾病进展发生的分子变化,这些技术包括基因表达、蛋白质组学、代谢组学、网络分析、基因组学和单核苷酸多态性 (SNP) 分析。由于HD被认为是以表观遗传失调为特征的疾病的典型,近来表观遗传方法已经成为评价病理学相关变化的有前途的新工具。HD的大多数表观遗传研究都集中于研究与HD相关的特定位点上的基因组宽组蛋白修饰 (乙酰化,甲基化) 或组蛋白修饰。尽管这些方法对疾病提供了感兴趣的洞悉,但它们产生的结果往往相互矛盾,并显示出小鼠模型与人类疾病之间的矛盾。因此,使用组蛋白修饰读数的HD表观遗传失调的共识图片尚未实现。然而,并非所有与表观遗传调控相关的分子机制都在HD的背景下进行了评价。表观遗传调控的一个重要方面是3维 (3D) 基因组结构水平。

[0333] 基因组的3D组织化反映了外部环境线索和输入的异质效果,并可以通过评价染色体构象或在同时测量几个构象时的染色体构象特征 (CCS) 进行经验性测量。CCS可以看作是分子条形码,可以提供给定细胞群体的表观遗传学情况的读取结果。鉴于mHTT在HD发展中的核心作用,我们假设,患病的个体与健康的未患病的对照之间可能存在HTT位点的基因组结构上的调控差异。我们使用EpiSwitch这种已建立的用于监控CCS的专属工业平台评价有症状和症状前HD患者与健康的未患病个体之间的染色质结构差异。EpiSwitch读取结果提供CCS的高分辨率、可靠而高通量的检测,同时满足质量控制的行业标准的高标杆。

[0334] 方法

[0335] 样品采集

[0336] 所有样品均获自National BioService, LLC。本研究总共使用了20个样品;10个健康对照 (HC) 样品 (CAG重复, $n < 35$) 和10个HD样品 (CAG重复, $n > 39$)。对于HD样品,7例来自有症状患者 (HD-Sym), 而3例来自诊断为HD而尚未显示任何临床症状 (HD-Pre) 的症状前患者。一名HD患者服用四苯那嗪, 一名患者服用舍曲林 (sertraline)。所有样品对于人免疫缺陷病毒、乙型肝炎病毒、丙型肝炎病毒和梅毒均为阴性 (表14)。

[0337] 研究设计

[0338] 我们想要确定健康对照(低CAG)、症状前HD患者(高CAG,无疾病表现)和有症状HD患者(高CAG,疾病表现)之间存在差异的染色体构象。我们专注于HTT位点周围的~225kb区域(chr4:在hg38中标注为3,033,588至3,258,170,)用于我们的分析。使用HTT外显子1(chr4:3,054,162至3,095,930)中的CAG重复扩增序列作为锚定点(“锚定”),我们定义了围绕该锚定的五个基因组区域,以观察样品组之间变化的染色体构象(图2和表15)。这些区域入选是基于:潜在的EpiSwitch锚定位点的存在;已知疾病相关的SNP的存在(HD和其他疾病);和GWASdv2数据库(<http://jjwanglab.org/gwasdb>)中发现的HD中的已知组蛋白修饰位点(H3K4me3,H3K36me3和H3K27ac)的富集(图2)。

[0339] 染色体构象同一性

[0340] 在2018年2月对NCBI:GEO数据库进行了先前报告的HD表观遗传数据的搜索。获得了来自12个(6个HD和6个对照样品)死后额叶前皮质大脑样本(床位形式)的H3K4me3峰调用ChIP-seq数据(GSE68952)。此外,还下载了源自HD iPSC的神经细胞系和对照细胞系的H3K27ac和H3K36me3的ChIP-seq数据的Bigwig轨迹(GSE95342)。数据轨迹与EpiSwitch和参考序列注释一起加载到整合基因组浏览器(IGV)[41]中。在HTT位点上进行了视觉和程序化(BEDtools)比较,以确定5个关注的区域。

[0341] 使用合适的软件确定高概率染色质折叠相互作用,其中一个“末端”存在于邻近CAG重复序列的锚区域中,而另一个则存在于所关注的5个区域中的任何一个区域中。共61个相互作用匹配这些标准,并且出于实际原因,选择了20个相互作用覆盖锚定位点和所有关注的区域之间的相互作用。自动化引物设计应用程序用于设计在进行染色体构象捕获分析测定时扩增由该相互作用引起的预期DNA序列的寡核苷酸对。

[0342] 3C和PCR

[0343] 进行染色体构象捕获和PCR检测。按照生产商的说明书(Oxford BioDynamics Plc),使用EpiSwitch试验提取来自每个患者样品的50 μ L血液样品中的具有完整染色体构象的染色质。所有样品的质量控制采用对MMP1位点上的染色质环的检测(3C分析的历史内部对照)完成。生成每种样品类型的汇总3C库,以为每个样品亚组提供通用的群体样品。采用SYBR绿使用CFX-96(Bio-Rad)机器进行实时PCR,以确定样品类型之间具有不同PCR产物检测模式的相互作用。在对照模板上测试寡核苷酸,以确证每个引物组是否工作正常。根据用于PCR检测的Royal Forensic协议,对每个样品进行一式三份的最终巢式PCR,以获取有关各个HD患者的随访数据。此过程允许以更高的准确性检测有限的拷贝数模板。使用LabChip DNA 1K版本2试剂盒(Perkin Elmer),在Perkin Elmer的LabChip[®]GX上监测所有PCR扩增产物,并使用荧光染料根据制造商的规程将内部DNA标志物装载到DNA芯片上。荧光通过激光检测,并使用仪器软件将电泳图读数转换为凝胶图片上的模拟带。仪器的检测阈值由制造商设置为30个荧光单位及以上。

[0344] 统计分析

[0345] 数据分析在R(用于统计计算的语言和环境)中进行。这包括用于t检验和R平方分析的统计数据和dplyr软件包,以及用于箱线图和回归图的ggplot2软件包。

[0346] 结果

[0347] 患者临床特征

[0348] HC和HD样品是年龄(HC平均值为36.9,HD平均值为35.3)和性别匹配(1/2男性和1/

2女性),且大多数(70%)HD病例有症状(表16)。所有样品均来自非西班牙裔或拉丁裔白人。HC的平均CAG重复长度为25.7,且HD的平均CAG重复长度为44.2(表16)。HC和HD样品在年龄上没有表现出统计学差异。HC患者与HD-Sym在年龄上无统计学差异。HD-Pre患者比HD-Sym患者(平均年龄=39.6)更年轻(平均年龄=25.3)($p=0.02$)。疾病持续时间与CAG重复大小之间呈中等程度的负相关,而诊断时的年龄与CAG重复尺寸之间呈中等程度的正相关,但两者均无统计学显著性。

[0349] 相对于HC,HD患者的CAG重复长度有统计学上的显著性增加($p=1.08E^{-7}$)。在HC与HD-Pre($p=3.43E^{-6}$)和HD-Sym($p=9.50E^{-8}$)之间,CAG重复长度有统计学上的显著增加。HD-Pre和HD-Sym之间的CAG重复长度没有统计学差异($p=0.09$)。

[0350] HD样拼诊断时的平均年龄为35.3岁,平均病程为3.8年,而每10名患者中有7名报告烦躁、舞蹈症或两者兼有的症状(表16)。

[0351] HC、HD-Pre和HD-Sym的染色体构象

[0352] 在评价的20个相互作用中(表17),我们鉴定出9种提供有用信息的相互作用。我们鉴定出HD中存在的两个基本相互作用和七个疾病性相互作用,但健康对照却没有。七个疾病性相互作用中有三个仅存在于HD-Sym中,而在HD-Pre中不存在。

[0353] 基本构象

[0354] 所有样品均通过了针对MMP1相互作用的内部QC分析。确定两个基本(在所有样品中均已鉴定出)染色质环。两个环在锚定点和区域2之间,第一个环跨度28kb,且第二个环跨度34kb。在本研究中观察到所有患者(HD-Sym,HD-Pre和HC)中都出现了两种基本相互作用。两种相互作用(CC1和CC2)在锚定点和区域2之间,CC1跨度为28kb,而CC2跨度为34kb。

[0355] 疾病性构象

[0356] 我们鉴定出七个可以区分本研究中评价的不同患者亚组的疾病性染色体构象。具体地,我们鉴定出了HC中存在而在所有HD样品中均不存在的两种染色体相互作用。第一相互作用(I1)跨越锚定点和区域4,覆盖了77kb,且第二相互作用(I2)跨越锚定点和区域3,覆盖140kb。我们还鉴定出了HC和HD-Pre中存在的两种染色体相互作用,而HD-Sym样品中却不存在的。两种相互作用(I3和I4)跨越了锚定点和区域4,分别覆盖了92kb和104kb。最后,我们鉴定出了HD-Sym样品中存在且HD-Pre和HC样品中不存在的三种染色体相互作用。第一相互作用(I5)跨越了锚定点和区域3,覆盖了122kb。感兴趣的是,这种相互作用包括一个已知是易患HD的因素的SNP(rs362331)。第二和第三相互作用(I6和I7)跨越了锚定点和区域1,分别覆盖了185kb和174kb。最后,我们测试了个体HD样品中所有疾病性相互作用的存在或不存在。在HD-Sym样品中,我们发现七个样品中的六个样品存在至少一个疾病性标志物(I5, I6和I7)(图3)。本研究中评价的所有相互作用的总结如图4和图5中所示。表18给出了优势比(odds ratio)。表13显示了在亚组之间没有任何差异的染色体相互作用。

[0357] 讨论问题陈述和结果总结

[0358] 虽然众所周知,CAG重复大于39的个体会罹患HD,但个体患者的疾病临床发作差异很大,影响疾病何时临床表现的因素表征还不太充分。在本文中,我们使用EpiSwitch这种评价染色质结构的工业平台评价了HD患者和健康的未患病对照中HTT位点的表观基因组学情况。我们确定了七种相互作用的集合,当将它们作为CCS一起使用时,可以将HD与未患病对照区分开,更重要的是,可以区分症状前HD患者和有症状HD患者。这些对有症状HD特异性

的相互作用之一包含一个显示与一个易感疾病单倍群 (haplogroup) 相关的 SNP (rs362331)。综上,这些结果表明,评价CCS的简单无创血液测试可以作为评价HD疾病进展的替代生物标志物。

[0359] 生物学相关性

[0360] 尽管众所周知,poly-Q重复序列的扩增和mHTT的产生是HD的根本原因,但导致临床症状发展的分子事件却表征不太充分。几项研究已将HTT位点内的SNP视为导致疾病发作的潜在贡献者。最近一项对HD患者的SNP基因分型研究在至少30%的患者中发现了~41个SNP杂合,包括HTT基因第50外显子的rs362331C/T SNP。或许更生物学上相关的是,当使用反义寡核苷酸、siRNA或miRNA等位基因选择性敲除rs362331SNP时,无论在体内还是体外,mHTT蛋白水平均显著降低,这间接表明该SNP及其周围环境基因组状况在调节mHTT水平中起着重要作用。在这项研究中,我们观察到仅在HD-Sym患者中存在而在HD-Asy和与rs362331SNP重叠的HC中不存在的染色体构象(I5)。这间接表明,通过rs362331SNP的存在,poly-Q道增加和HD早期发展的遗传易感性患者中神经毒性mHTT的产生可以以较高的染色质结构水平受到调节。HD的另一个突出问题是在父母双方均未得到诊断的情况下该疾病是如何遗传的。两种主要的普遍假设是:1) 携带者父母在疾病发作之前可能已经脱离了另一个因素;2) “不稳定的”CAG重复道随着每次代传而扩增。还存在第三种可能性,即在中等范围(35-50)重复时,个体可以是携带者而没有该疾病的表现,但他们的后代可能无法通过不确定的机制补偿遗传缺陷,并将发展该疾病。在这项研究中评价的HD患者均在该中间范围内具有CAG重复,从而增加了疾病发展中潜在的补偿机制可能通过基因组结构差异介导的可能性。

[0361] 临床相关性

[0362] HD的独特之处在于,存在一个简单的测试用于明确诊断该疾病,HTT基因测序和测量CAG重复数。对于临床护理和临床试验,也有几种测量疾病严重程度的测试,如统一的亨廷顿病评定量表(the Unified Huntington's Disease Rating Scale) (UHDRS)、Shoulson-Fahn量表和迷你精神状态检查(Mini-Mental State Examination) (MMSE)。尽管这些评价度量HD患者的身心健康状况的不同要素为疾病严重程度的替代指标,但它们本质上都是主观的,并且大多数并非HD特异性的。缺少的是监测疾病进展的具体分子工具。

[0363] 目前,在临床前和临床开发的不同阶段中,有22种治疗HD的治疗剂,其中一半处于2期或3期。一旦进一步验证,此处报道的CCS可以在临床试验中用作替代结果生物标志物,以评价有关药物的治疗功效。除了在临床试验中监测有症状患者对特定疗法的反应外,此处方法的另一个优点还在于可以为症状前患者获得信息。对于大多数HD患者,症状前期可以持续数十年。本文中鉴定出的七种(i3-i7)相互作用中有五种清楚地将症状前HD患者与有症状HD患者区分开,并且当进一步验证时,可以充当症状前携带者的HD症状发作的“预警”指标测试。

[0364] 这项研究提供了染色质结构中HD表现特异性的并与已知疾病单倍型相关的可检测疾病性差异的第一个证据。这项研究的关键优点在于独特的方法,该方法基于了解基因组架构的调控作用的最新进展。尽管有数项HD的历史研究旨在基于临床、成像和分子手段开发疾病进展生物标记,但据我们所知,这是首次在临床上可触及的生物流体中评价高级染色质结构应用于HD。

探针	位点	探针总 计数	显著探 针数
C6orf58_6_127480771_127483471_127600017_127604343_FR	C6orf58	10	5
NEFH_22_29442588_29445314_29482081_29484217_RR	NEFH	16	10
IL1A_2_112765786_112772711_112810765_112813086_RR	IL1A	23	15
FARP1_13_98271575_98282700_98346930_98348486_FR	FARP1	26	26
PRKCA_17_66441276_66447067_66475597_66481312_RF	PRKCA	30	16
RNU6-1264P_17_6162286_6163870_6195952_6199184_FR	RNU6-1264P	31	16
PON2_7_95405100_95420940_95465337_95474032_FR	PON2	40	19
CAPN9_1_230738572_230739927_230752057_230757333_RR	CAPN9	50	23
ATXN7L1_7_105654123_105657510_105741521_105750599_FR	ATXN7L1	130	61
CNTNAP2_7_146728706_146734820_146785878_146792823_RF	CNTNAP2	144	60
CTNNA3_10_66299269_66302507_66496211_66513003_FF	CTNNA3	151	46
ZFPM2_8_105632010_105638904_105814873_105824107_FR	ZFPM2	186	68
ALDH1A2_15_58325151_58334051_58485549_58488054_FR	ALDH1A2	185	70
ALDH1A2_15_58325151_58334051_58538695_58540885_FF	ALDH1A2	185	70
MAGI2_7_79009346_79018304_79275810_79284623_RF	MAGI2	186	80
FER1L6_8_123963222_123969450_124085753_124093275_FR	FER1L6	98	13
UBQLN2_X_56536168_56538402_56570114_56575112_FR	UBQLN2	10	7

[0365]

[0366] 表1a

[0367] 表1b

[0368]

HyperG_Stats	FDR_HyperG	%显著	logFC	AveExpr	t
0.192504768	0.761792372	50	-0.206940905	-0.206940905	-7.657670128
0.012576275	0.174966022	62.5	0.155885782	0.155885782	12.07017614
0.001213793	0.03415703	65.22	0.130559454	0.130559454	3.67899153
0.000000000000171	0.000000000079	100	0.254822215	0.254822215	14.73928749
0.013706774	0.176284348	53.33	0.24251793	0.24251793	7.451573283
0.019964247	0.210078331	51.61	-0.152642021	-0.152642021	-9.03898498
0.032185744	0.276690843	47.5	-0.178366996	-0.178366996	-5.268810519
0.029727242	0.269876723	46	-0.224588757	-0.224588757	-8.238208751
0.000354202	0.018221718	46.92	0.189559955	0.189559955	6.597539924
0.011284771	0.170730832	41.67	-0.440614155156995	-0.440614155156995	-7.4296031698139
0.713857099	1	30.46	-0.436584531378388	-0.436584531378387	-14.2285764776972
0.122432344	0.622894915	36.56	-0.425339105073479	-0.425339105073479	-10.493629902935
0.063739285	0.421589843	37.84	0.34586199024656	0.34586199024656	13.4772416929847
0.063739285	0.421589843	37.84	0.339861526563037	0.339861526563037	7.22282802700005
0.00139007	0.035755702	43.01	0.34944598213495	0.34944598213495	10.5772811324938
0.999996196	1	13.27	0.350542987884941	0.350542987884941	5.66944350076527
0.016370015	0.18948292	70	0.353892612578078	0.353892612578078	2.69374295077173

[0369] 表1c

[0370]

P.值	adj.P.值	B	FC	FC_1	LS	检测到的环
0.000000338	8.58E-06	6.769090112	0.866372344	-1.154238137	-1	对照
0.000000000262	7.50E-08	13.89074006	1.114105444	1.114105444	1	ALS
0.001613314	0.006000711	-1.70129901	1.094718133	1.094718133	1	ALS
0.0000000000858	1.48E-08	17.19553875	1.193188696	1.193188696	1	ALS
0.0000005	1.15E-05	6.37489208	1.183055643	1.183055643	1	ALS
0.0000000281	1.46E-06	9.26023757	0.899601504	-1.111603299	-1	对照
0.000045	0.000343598	1.852676493	0.883702704	-1.131602286	-1	对照
0.000000115	3.91E-06	7.847763841	0.85583895	-1.168444134	-1	对照
0.0000027	3.98E-05	4.678642967	1.140415818	1.140415818	1	ALS
5.21452107638594E-07	1.18E-05	6.332521031	0.736820877	-1.357181957	-1	对照
1.58303956750015E-11	1.70E-08	16.61028384	0.738881785	-1.353396471	-1	对照
2.63975825723201E-09	2.70E-07	11.61504783	0.744663678	-1.342888111	-1	对照
4.03123106207005E-11	2.85E-08	15.71064142	1.270910107	1.270910107	1	ALS
7.77879601584934E-07	1.61E-05	5.930445637	1.265635109	1.265635109	1	ALS
2.32055637317047E-09	2.48E-07	11.7426843	1.27407127	1.27407127	1	ALS
0.0000188398197223922	0.000175073	2.726108027	1.275040425	1.275040425	1	ALS
0.0144551346098748	0.035577015	-3.813817915	1.278004231	1.278004231	1	ALS

[0371] 表1d

[0372]

探针序列 60 mer	探针位置	
	Chr	起始 1
TCACCACACATCACCCCTTGCTCCTCGAGTCTTGGTGACCACAACAGGGTGCCACC	6	127483440
GAGGTGGGTGAATCATGAGGTCAAGGGTTCGACAATAGTTGAGAATCTCCAACCACCTGG	22	29442590
GGCCTTATAGTCAGCTGATCAGGTGAAATCGATTGGTCTTAGGATCAGCTACCATTTGC	2	112765788
GAGGCAGGCGGATCACAAAGTCAAAAGATCGATAACTTCAATAATAGTTACAGATGCAAA	13	98282669
AGCACCATATCTGGGATGTAGCTATTGCTCGAGATTGCAGTGAGCTGTGATCACACCTCT	17	66441278
TCTTCCCTCTTTTAAACCACCATTTCATCGACCCACACATCCTGTGCCACTCTACTGC	17	6163839
TAACCATTATGCATCACTAACATAGCATTCGATATGATATGCTCAGTTTAGTTAGGGAAA	7	95420909
GGCTCAGGAAGAGAAGTATTTGTCTCTTTTCGACACGCACATGCAGGACACTCACACGTAG	1	230738574
GTTGGGTGGATCCCTTGAGCTCAGGAATTCGAAGAATGATTTTTCAGCCCGTGTGGAAGG	7	105657479
ATCAAAAGAAAAATAGATACTTGTCTTACTCGAGTTGAATAAAATCCTCAGCTTTCTGTCC	7	146728708
AAAAGAACTGTGAAAAGTTGTCACATTCGATTAATCCAAAAGGTCTTCTATGAGGC	10	66302476
TTAAAAGTATAGTAGTTGGCATAACATTCGACCTTTTTCTGTTTCAGTAACCAACCAG	8	105638873
CATCAACTAATAGTTAAACATTATAATATCGACTGAAGACCTTTCATACTGTAAGATTCA	15	58334020
CATCAACTAATAGTTAAACATTATAATATCGAGTCTGCAGTGAGCTGAGATCACACTGCC	15	58334020
TTATTCCTTTCCAAATAGTTAAAATTATTCGAACTTTTAAAGATCAATATAAAATTTCC	7	79009348
CATAATTATAAATTAATAAATGACACTATCGATTATGTCCAGTGTTCCTGGTTGGTGTC	8	123969419
CAGAGCACTAAGATAGACTTCTAAGGTTTCGAGGCATATAGCTCCAGTGTATTGAGGTA	X	56538371

[0373] 表1e

探针位置			4 kb 序列位置				
末端 1	起始 2	末端 2	Chr	起始 1	末端 1	起始 2	末端 2
127483469	127600019	127600048	6	127479470	127483469	127600019	127604018
29442619	29482083	29482112	22	29442590	29446589	29482083	29486082
112765817	112810767	112810796	2	112765788	112769787	112810767	112814766
98282698	98346932	98346961	13	98278699	98282698	98346932	98350931
66441307	66481281	66481310	17	66441278	66445277	66477311	66481310
6163868	6195954	6195983	17	6159869	6163868	6195954	6199953
95420938	95465339	95465368	7	95416939	95420938	95465339	95469338
230738603	230752059	230752088	1	230738574	230742573	230752059	230756058
105657508	105741523	105741552	7	105653509	105657508	105741523	105745522
146728737	146792792	146792821	7	146728708	146732707	146788822	146792821
66302505	66512972	66513001	10	66298506	66302505	66509002	66513001
105638902	105814875	105814904	8	105634903	105638902	105814875	105818874
58334049	58485551	58485580	15	58330050	58334049	58485551	58489550
58334049	58540854	58540883	15	58330050	58334049	58536884	58540883
79009377	79284592	79284621	7	79009348	79013347	79280622	79284621
123969448	124085755	124085784	8	123965449	123969448	124085755	124089754
56538400	56570116	56570145	X	56534401	56538400	56570116	56574115

[0374]

[0375]

表2a

[0376]

探针	PCR-引物1_ID	PCR_引物1
C6orf58_6_127480771_127483471_127600017_127604343_FR	OBD112-021	AAGCACTTCATTCTCCCCTCACC
NEFH_22_29442588_29445314_29482081_29484217_RR	OBD112-057	ACTGAGCAATGATGGCAACAAC
IL1A_2_112765786_112772711_112810765_112813086_RR	OBD112-077	CTAGGCCTGCGTTTCTCCGT
FARF1_13_98271575_98282700_98346930_98348486_FR	OBD112-093	AGTTCTCTCTAAGAACTCAAGGA
PRKCA_17_66441276_66447067_66475597_66481312_RF	OBD112-101	CCCCAGGCACTCACACCTTA
RNU6-1264P_17_6162286_6163870_6195952_6199184_FR	OBD112-105	GGGCACTCAACACCTTTTGT
PON2_7_95405100_95420940_95465337_95474032_FR	OBD112-125	GATGGGAATCAAGGGCAAGGG
CAPN9_1_230738572_230739927_230752057_230757333_RR	OBD112-145	GCATTAGCCAGCAAGCATACTT
ATXN7L1_7_105654123_105657510_105741521_105750599_FR	OBD112-177	CACCGCCTGATGCAAGTCTT
CNTNAP2_7_146728706_146734820_146785878_146792823_RF	OBD112-217	GGCACTGTTGGTCTGAAGCAC
CTNNA3_10_66299269_66302507_66496211_66513003_FF	OBD112-229	GGTCTAGATGTCAGTCTTTCC
ZFPM2_8_105632010_105638904_105814873_105824107_FR	OBD112-245	GACTATAAATCTCTCCTTGTCAGC
ALDH1A2_15_58325151_58334051_58485549_58488054_FR	OBD112-317	CCTTACCCGACACCAGGTAGC
ALDH1A2_15_58325151_58334051_58538695_58540885_FF	OBD112-309	CCCGACACCAGGTAGCATTC
MAGI2_7_79009346_79018304_79275810_79284623_RF	OBD112-329	GGGACCCCACTAGACCAC
FER1L6_8_123963222_123969450_124085753_124093275_FR	OBD112-333	ATTCTCTCCCTGGTAAATCCTGGT
UBQLN2_X_56536168_56538402_56570114_56575112_FR	OBD112-337	CGCCAGCTCAGCAGCAATAA

[0377]

表2b

[0378]

PCR-引物2_ID	PCR_引物2	GLMNET
OBD112-023	ATACTCCCATCCCCTAGGCC	0.799327478
OBD112-059	CCCTACGACTGGCAAACCCA	-0.323299991
OBD112-079	CCCTGGCATTACATCACCGA	0.422829791
OBD112-095	GTCAAGCAACTGTGTCTGGGG	0.471919696
OBD112-103	GGATCCACGATCTCCCTCCAC	-0.379552353
OBD112-107	AGGTCAGGATGGGTACCGTTG	0.259971774
OBD112-127	CTGGGATGATTCTCTGGACTTCT	-1.132792925

OBD112-147	GGTGGGCCTGGGTTAGATGC	0.980123681
OBD112-179	CAGCTGGCCGATCCATCACC	-0.201292203
OBD112-219	GAGAACGACGACCTGGCACT	-0.613932219
OBD112-231	TCATCCTATCCTCTCCTAGC	-0.36233865
OBD112-247	ACTGTAGGCCAACCAGAAG	0.541529302
OBD112-319	CTGGTCCAGTGTGACGCGTGT	-0.562433398
OBD112-311	GTGACTCTGCACGCACTGTT	0.486039187
OBD112-331	AGGCTCAGCAGGTTTCTGCC	-0.759217304
OBD112-335	CTGGGCAGGTCATCCAGACAG	-1.154344187
OBD112-339	TGCTAGGGCCGAGTAATCATC	0.146973789

[0379] 表3

[0380]

标志物	GLMNET
OBD112-021/023	0.799327478
OBD112-057/059	-0.323299991
OBD112-077/079	0.422829791
OBD112-093/095	0.471919696
OBD112-101/103	-0.379552353
OBD112-105/107	0.259971774
OBD112-125/127	-1.132792925
OBD112-145/147	0.980123681
OBD112-177/179	-0.201292203
OBD112-217/219	-0.613932219
OBD112-229/231	-0.36233865
OBD112-245/247	0.541529302
OBD112-317/319	-0.562433398
OBD112-309/311	0.486039187
OBD112-329/331	-0.759217304
OBD112-333/335	-1.154344187
OBD112-337/339	0.146973789

[0381] 表4a

[0382]

基因	全名
C6orf58	染色体6开放阅读框58
NEFH	神经丝重
IL1A	白介素1 α
FARP1	FERM、ARH/RhoGEF和普列克底物蛋白域蛋白1
PRKCA	蛋白激酶C α
RNU6-1264P	RNA,U6小核1264,假基因
PON2	对氧磷酶2
CAPN9	钙蛋白酶9

ATXN7L1	共济失调蛋白7样1
CNTNAP2	接触蛋白相关蛋白-样2
CTNNA3	接触蛋白 α 3
ZFPM2	锌指蛋白,FOG家族成员2
ALDH1A2	醛脱氢酶1家族成员A2
MAGI2	膜相关鸟苷酸激酶,WW和PDZ域包含2
FER1L6	Fer-1样家庭成员6
UBQLN2	泛素2

[0383] 表4b

活性
蛋白质编码基因
神经元口径 (neuronal caliber) 的维持, 细胞内向轴突和树突的运输, 神经元的损伤以及对肌萎缩性侧索硬化症 (ALS) 的敏感性
各种免疫反应, 炎症过程, 造血作用和细胞损伤, 激活细胞凋亡, 与类风湿关节炎和阿尔茨海默氏病有关
神经元中的树突状生长, 细胞骨架与细胞膜之间的联结
多样的细胞信号传递通路, 佛波酯的受体, 细胞过程如细胞黏附、细胞转化、细胞周期检查点和细胞体积控制
假基因, 与 snRNA 类有关
细胞抗氧化剂, 人体组织和膜结合, 对病原菌的防御反应, 保护细胞免受氧化应激
与胃癌相关的神经变性过程和消化道
蛋白质编码基因
调节神经元和神经胶质之间的相互作用, 与语音和语言发育, 神经系统中的细胞粘附相关
肌肉细胞中的细胞-细胞粘附, 与致心律失常的右心室发育不良有关
调节哺乳动物的造血作用和心脏发育,
发育成人组织, 后器官发育, 预防脊柱裂
齿状核红核苍白球下丘脑萎缩 (dentatorubral and pallidolusian atrophy)
蛋白编码基因
体内蛋白质降解

[0384]

[0385] 表5a

[0386] 3个月样品的阵列数据

[0387]

探针	位点	探针计数总数	显著探针数
ZFPM2_8_104964347_104973135_105209447_105218050_FF	ZFPM2	187	6
VEGFA_6_43733863_43737741_43777557_43780533_FR	VEGFA	42	1
ERBB4_2_212287088_212294815_212317329_212325591_FR	ERBB4	193	10
ZFPM2_8_105209447_105218050_105302264_105310138_FF	ZFPM2	187	6
PASD1_X_151600201_151608969_151687203_151692271_FR	PASD1	41	14
GRM7_3_7259138_7267165_7394377_7401227_RF	GRM7	174	7
GRIK2_6_102047545_102055341_102078831_102089392_FR	GRIK2	188	5
SORCS2_4_7219417_7226457_7241158_7248673_FF	SORCS2	84	6
LINGO2_9_28314155_28333777_28371887_28374860_FR	LINGO2	194	3

[0388] 表5b

[0389]

HyperG_Stats	FDR_HyperG	%_特征	logFC	AveExpr	t	P.值	adj.P.值
0.067854625	0.99237389	3.21	0.272052749	7.772607282	3.665534368	0.001024476	0.999953189
0.475978534	1	2.38	-0.142175177	14.62826333	-2.054497284	0.049386319	0.999953189

0.000818098	0.03828697	5.18	0.383778852	7.597586859	3.176083618	0.003622402	0.999953189
0.067854625	0.99237389	3.21	0.229570464	9.577318093	2.801385891	0.00913346	0.999953189
7.57E-16	3.54E-13	34.15	0.341166309	9.148011765	2.62030276	0.01404442	0.999953189
0.01784403	0.37959119	4.02	0.306594489	9.301524803	2.334933566	0.026957376	0.999953189
0.16129386	1	2.66	0.286167252	9.345860096	2.824849044	0.008630812	0.999953189
0.001827031	0.0712542	7.14	-0.222042969	10.03535881	-2.171963051	0.038499943	0.999953189
0.570146657	1	1.55	0.40947654	6.9839738	2.811551759	0.008912387	0.999953189

[0390] 表5c

[0391]

B	FC	FC_1	LS	检测到的环
-1.464031868	1.207524741	1.207524741	1	Fs
-3.802493578	0.906151904	-1.103567731	-1	Sw
-2.223740309	1.304754929	1.304754929	1	Fs
-2.785028975	1.172485811	1.172485811	1	Fs
-3.046315984	1.266780276	1.266780276	1	Fs
-3.440610909	1.236784796	1.236784796	1	Fs
-2.750635647	1.219396451	1.219396451	1	Fs
-3.654353119	0.857350501	-1.166384109	-1	Sw
-2.770141569	1.328203808	1.328203808	1	Fs

[0392] 表5d

探针序列		探针位置		
60 mer		Chr	起始 1	末端 1
TATATTTAAAAATACATACTGGTATACATCGATCTCATGACTTTGCTATTATGCATAGTG		8	104973104	104973133
CCCCAGCCCAGCAACCTGGCTCACCTGATCGAGTACATCTCAAGCCATCCTGTGTGCC		6	43737710	43737739
TATAAATAATACAGCTCTATTTGCTACTCGATTAAGAATCATATTATATCCTTAATTC		2	212294784	212294813
CACTATGCATAATAGCAAAGTCATGAGATCGAAAATGTTTGTCAAGCAGTAGGTTTGGG		8	105218019	105218048
GATTTTAGAATCTCAACAAGGCTGCAATCGAGGTTAGCTGCTGCAGAAAAGAGAAAA		X	151608938	151608967
GGTGACTAAAAATGAGATTGCATTTTCTTCGACCATTGGCCAGCATGCCAAACACTGGT		3	7259140	7259169
TAACCTCTCCTTTCTTAGGTTCTCCATATCGATAGAAAATTGTCTGCAGCCCTTAATGCC		6	102055310	102055339
AGCTCACGGTCAGTGCCGTTCCGTTTCTCGAATAAAGAACAAGGACCTTAAAAAATAGA		4	7226426	7226455
AAAGTCATACAACACTACTATGTAAGATATTCGAATACCTGTTAGAAATAGGTGAAGGTTTAT		9	28333746	28333775

[0394] 表5e

探针位置		4 kb 探针位置				
起始 2	末端 2	Chr	起始 1	末端 1	起始 2	末端 2
105218019	105218048	8	104969134	104973133	105214049	105218048
43777559	43777588	6	43733740	43737739	43777559	43781558
212317331	212317360	2	212290814	212294813	212317331	212321330
105310107	105310136	8	105214049	105218048	105306137	105310136
151687205	151687234	X	151604968	151608967	151687205	151691204
7401196	7401225	3	7259140	7263139	7397226	7401225
102078833	102078862	6	102051340	102055339	102078833	102082832
7248642	7248671	4	7222456	7226455	7244672	7248671
28371889	28371918	9	28329776	28333775	28371889	28375888

[0396] 表6a

[0397] 6个月样品的阵列数据

[0398]

探针	位点	探针计数总数	显著探针计数	HyperG_Stats
ZFPM2_8_104964347_104973135_105209447_105218050_FF	ZFPM2	187	8	0.007868339
VEGFA_6_43733863_43737741_43777557_43780533_FR	VEGFA	42	1	0.472787206

ERBB4_2_212287088_212294815_212317329_212325591_FR	ERBB4	193	13	8.61E-06
ZFPM2_8_105209447_105218050_105302264_105310138_FF	ZFPM2	187	8	0.007868339
PASD1_X_151600201_151608969_151687203_151692271_FR	PASD1	41	18	1.87E-22
GRM7_3_7259138_7267165_7394377_7401227_RF	GRM7	174	6	0.049498361
GRIK2_6_102047545_102055341_102078831_102089392_FR	GRIK2	188	6	0.066833261
SORCS2_4_7219417_7226457_7241158_7248673_FF	SORCS2	84	3	0.134527864
LING02_9_28314155_28333777_28371887_28374860_FR	LING02	194	4	0.337532174

[0399] 表6b

FDR_HyperG	% 特征	logFC	AveExpr	t	P.值	adj.P.值
0.204139683	4.28	0.214103447	7.772607282	2.884747708	0.007463232	0.999935724
1	2.38	-0.142760759	14.62826333	-2.062959216	0.04851942	0.999935724
0.001005104	6.74	0.365781024	7.597586859	3.027136877	0.005258488	0.999935724
0.204139683	4.28	0.23990173	9.577318093	2.927455517	0.006723676	0.999935724

[0400]

8.73E-20	43.9	0.331193838	9.148011765	2.543709927	0.016785386	0.999935724
0.825561956	3.45	0.354666043	9.301524803	2.701032396	0.011610123	0.999935724
1	3.19	0.294782642	9.345860096	2.909894328	0.007018965	0.999935724
1	3.57	-0.229277936	10.03535881	-2.242733507	0.033029594	0.999935724
1	2.06	0.440106238	6.9839738	3.021861688	0.005327738	0.999935724

[0401]

[0402] 表6c

[0403]

B	FC	FC_1	LS	检测到的环
-2.662323695	1.159982827	1.159982827	1	Fs
-3.791998096	0.905784176	-1.104015754	-1	Sw
-2.449689974	1.288579031	1.288579031	1	Fs
-2.598930246	1.18091222	1.18091222	1	Fs
-3.154405805	1.25805399	1.25805399	1	Fs
-2.930776901	1.278689554	1.278689554	1	Fs
-2.625038885	1.226700147	1.226700147	1	Fs
-3.562656794	0.853061739	-1.172248096	-1	Sw
-2.457630227	1.35670423	1.35670423	1	Fs

[0404] 表6d

探针序列 60 mer	探针位置		
	Chr	起始 1	末端 1
TATATTTAAAAATACATACTGGTATACATCGATCTCATGACTTTGCTATTATGCATAGTG	8	104973104	104973133
CCCCAGCCAGCAACCTGGCTCACCTGATCGAGTACATCTCAAGCCATCCTGTGTGCC	6	43737710	43737739
TATAAATAATACAGCTCTATTGCTACTCGATTAAGAATCATATTATATCCTTAATTC	2	212294784	212294813
CACTATGCATAATAGCAAAGTCATGAGATCGAAAATGTTTGCAAGCAGTAGGTTTGGG	8	105218019	105218048
GATTTTAGAATCTCTAACAAAGGCTGCAATCGAGGTTAGCTGCTGCAGAAAGAAGAGAAAA	X	151608938	151608967
GGTGACTAAAATGAGATTGCATTTTCTTCGACCATTTGGCCAGCATGCCAAACACTGGT	3	7259140	7259169
TAACCTCTCCTTTCTTAGGTTCTCCATATCGATAGAAAATTGCTGCAGCCCTTAATGCC	6	102055310	102055339
AGCTCACGGTCAGTGCCGTTCCGTTTCTCGAATAAAGAACAAGGACCTTAAAAAATAGA	4	7226426	7226455
AAAGTCATACAACACTACTATGTAAGATATTCGAATACCTGTTAGAATAGGTGAAGGTTTAT	9	28333746	28333775

[0405]

[0406] 表6e

探针位置		4 kb 序列位置				
起始 2	末端 2	Chr	起始 1	末端 1	起始 2	末端 2
105218019	105218048	8	104969134	104973133	105214049	105218048
43777559	43777588	6	43733740	43737739	43777559	43781558
212317331	212317360	2	212290814	212294813	212317331	212321330
105310107	105310136	8	105214049	105218048	105306137	105310136
151687205	151687234	X	151604968	151608967	151687205	151691204
7401196	7401225	3	7259140	7263139	7397226	7401225
102078833	102078862	6	102051340	102055339	102078833	102082832
7248642	7248671	4	7222456	7226455	7244672	7248671
28371889	28371918	9	28329776	28333775	28371889	28375888

[0408] 表7a

探针	内部引物	
	PCR-引物 1_ID	PCR 引物 1
ZFPM2_8_104964347_104973135_105209447_105218050_FF	OBD112_773	GCAGCACACAGGGAAGTCTCTT
VEGFA_6_43733863_43737741_43777557_43780533_FR	OBD112_765	GCTAGGAAGGCCTGGGATG
ERBB4_2_212287088_212294815_212317329_212325591_FR	OBD112_477	TCCAAAATGTTAATACTGCCTAGA
ZFPM2_8_105209447_105218050_105302264_105310138_FF	OBD112_777	CAGCCACTGTAGAGAGCAGT
PASD1_X_151600201_151608969_151687203_151692271_FR	OBD112_637	ACTTCTCCCAAGTCACTTTTTC
GRM7_3_7259138_7267165_7394377_7401227_RF	OBD112_517	TGACGAAGAAGCAATCCCTGGT
GRIK2_6_102047545_102055341_102078831_102089392_FR	OBD112_501	AATCTCTGCCCTCTCATCTTG
SORCS2_4_7219417_7226457_7241158_7248673_FF	OBD112_725	AGGTATGCAGCCAGCCTGAG
LINGO2_9_28314155_28333777_28371887_28374860_FR	OBD112_561	CCGTGCCATATCTCTGATTTATGC

[0410] 表7b

内部引物		
PCR-引物 2_ID	PCR_引物 2	GLMNET
OBD112_775	TTGTTGAGCCAGCAATTCCTTT	-0.18669073
OBD112_767	CTCAGTGGGCACACTCCA	0.15278372
OBD112_479	AGCCATGTGGTCTGGAATCT	-0.24015518
OBD112_779	TAACCACAGCAGCAAGGT	0.22911033
OBD112_647	TGGCCATCTTGCTTGCCTC	0.52468015
OBD112_519	GGACCTACCTCCACTGGGTTG	0.10503793
OBD112_503	CATGTTCCACAGCAAGGAAGTTA	-0.13060979
OBD112_727	TTCCGTGCCAGTGCCTGT	-0.1217346
OBD112_563	GGCTGACCTCAACAGATTCGC	0.01500745

[0412] 表8

基因	标志物	GLMNET
ERBB4_2	OBD112_477.479_0.4ng	-0.24015518
ZFPM2_8	OBD112_773.775_6.5ng	-0.18669073
GRIK2_6	OBD112_501.503_0.8ng	-0.13060979
SORCS2_4	OBD112_725.727_3.25ng	-0.1217346
LINGO2_9	OBD112_561.563_13ng	0.01500745
GRM7_3	OBD112_517.519_13ng	0.10503793
VEGFA_6	OBD112_765.767_13ng	0.15278372
ZFPM2_8	OBD112_777.779_6.5ng	0.22911033
PASD1_X	OBD112_637.647_3.25ng	0.52468015

[0414] 表9

基因	全名	活性
ZFPM2	锌指蛋白, FOG 家庭成员 2	转录因子, 调节 GATA 家族蛋白的活性, 调节 GATA 靶基因的表达
VEGFA	血管末端内皮生长因子 A	刺激血管内皮细胞的增殖和迁移, 刺激生理和病理性血管生成, 调节肿瘤阶段和进展
ERBB4	Erb-B2 受体酪氨酸激酶 4	刺激各种细胞反应, 包括有丝分裂和分化,
[0415] PASD1	含 PAS 域阻遏子 1	编码与癌症相关的抗原, 诱导自体 T 细胞反应, 是治疗各种造血系统恶性肿瘤的免疫治疗靶标
GRM7	谷氨酸代谢型受体 7	激活离子型和代谢型谷氨酸受体, 参与正常脑功能大多数方面, 并在神经病理学状况下受到干扰
GRIK2	谷氨酸离子型受体红藻氨酸型亚基 2	调节神经生理过程
SORCS2	含排序素相关的 VPS10 域的受体 2	蛋白质编码基因, 在中枢神经系统中过度表达
LINGO2	富亮氨酸重复和 Ig 域包含 2	蛋白编码基因

[0416] 表10a

N	阵列_探针	引物 1_ID
1	DCLK1_13_35749195_35757694_36000600_36005542_RR	OBD112_453
2	GRM7_3_7259138_7267165_7394377_7401227_RF	OBD112_517
3	LINGO2_9_28314155_28333777_28371887_28374860_FR	OBD112_561
[0417] 4	PASD1_X_151600201_151608969_151687203_151692271_FR	OBD112_637
5	UBQLN2_X_56536168_56538402_56570114_56575112_RR	OBD112_757
6	ZFPM2_8_104964347_104973135_105209447_105218050_FF	OBD112_773
7	ZFPM2_8_105209447_105218050_105302264_105310138_FF	OBD112_777
8	ZNF804B_7_89108281_89114839_89127203_89146680_RF	OBD112_797

[0418] 表10b

引物 1_序列	引物 2_ID	引物 2_序列
TCTTGACACGGTTGGTGGT	OBD112_455	TGTCACCTATGTGCTGAGTACTGG
TGACGAAGAAGCAATCCCTGGT	OBD112_519	GGACCTACCTCCACTGGGTTG
CCGTGCCATATCCTCTGATTATGCG	OBD112_563	GGCTGACCTTCAACAGATTCGCG
[0419] ACTTCTTCCCAAGTCACTTTTTGCG	OBD112_647	TGGCCATCTTGCTTTGCGCTC
GGATATGCAGTTTTCTGGCACTAC	OBD112_759	CATGCTAGGGCCGAGTAATCATCT
GCAGCACACAGGGAAGTCTCTT	OBD112_775	TTGTTGAGCCCAGCAATTCCTTT
CAGCCACTGTAGAGAGCAGT	OBD112_779	TAACCCACCAGCAGCAAGGT
AGTAGCTTCCCTGTTAGAGTCTTG	OBD112_799	AGCCAGTGACTCCACAAGTCTT

[0420] 表10c

水解_探针_ID	水解_探针_序列
OBD112_88_FAM	AGGTTGCTTTGAAAGTACAGATATCACT
OBD112_57_FAM	ATTTTCTTTGACCATTTGGCCAGCA
OBD112_90_FAM	TTCTAACAGGTATTCGAATATCTTACATAGT
[0421] OBD112_26_FAM	AAGTAACATTCGACAAACCTCTGGTAA
OBD112_95_FAM	TATTCTCTTTGATCGAGGCATATAGCT
OBD112_80_FAM	AAAGTCATGAGATCGATGTATAACAGTAT
OBD112_32_FAM	TTGACAAACATTTTCGATCTCATGACTT
OBD112_33_FAM	TGTGGCTCAGTCGACAGAAAGTACA

[0422] 表11a(i)

探针	位点	探针计数总数	显著探针计数
----	----	--------	--------

DCLK1_13_35749195_35757694_36000600_36005542_RR	DCLK1	182	3
UBQLN2_X_56536168_56538402_56570114_56575112_RR	UBQLN2	10	7
ZNF804B_7_89108281_89114839_89127203_89146680_RF	ZNF804B	191	10

[0424] 表11a(ii)

HyperG_Stats	FDR_HyperG	%_特征	logFC	AveExpr	t
0.254316537	1	1.65	0.199442299	7.402197005	2.87151
0.000000000001	0.000000000386	70	0.283366605	9.419823948	2.114955
1.74E-05	0.002050325	5.24	0.228756351	9.389157739	2.52348

[0426] 表11a(iii)

P.值	adj.P.值	B	FC	FC_1	LS	检测到的环
0.007707652	0.999953189	-2.681902147	1.14825439	1.14825439	0	较快
0.043482517	0.999953189	-3.726899057	1.217031581	1.217031581	0	较快
0.017587875	0.999953189	-3.182692934	1.171824364	1.171824364	0	较快

[0428] 表11b(i)

探针	PCR-引物1_ID	PCR_引物1
DCLK1_13_35749195_35757694_36000600_36005542_RR	OBD112_453	TCTTGACACGGTTGGTGGT
UBQLN2_X_56536168_56538402_56570114_56575112_RR	OBD112_757	GGATATGCAGTTTCTGGCACTAC
ZNF804B_7_89108281_89114839_89127203_89146680_RF	OBD112_797	AGTAGCTCCCTGTTAGAGGTCTTG

[0430] 表11b(ii)

PCR-引物2_ID	PCR_引物2	GLMNET
OBD112_455	TGTCACCTATGTGCTGAGTACTGG	0.047800994
OBD112_759	CATGCTAGGGCCGAGTAATCATCT	-0.013191036
OBD112_799	AGCCAGTGACTCCACAACCTTCTT	-0.025426843

[0432] 表11c

标志物	GLMNET
OBD112_453.455_3.25ng	0.047801
OBD112_757.759_13ng	-0.013191
OBD112_797.799_1.6ng	-0.025427

[0434]

表 12a

探针		位点	检测到的环	探针序列
探针		60 mer		
HTT_4_3088056_3090591_3182426_3183765_FR	HTT	HD_阴性	AGATCTAGTTCACAGTAGCACCAATATATCGACAGATAGCTGACATCATCTCCCAATGT	
HTT_4_3076247_3078445_3182426_3183765_RR	HTT	HD_阴性	AGAGTACTTCCTAACTCTACTGTACACTCGACAGATAGCTGACATCATCTCCCAATGT	
HTT_4_3062575_3066389_3251623_3253432_RF	HTT	HD_阳性	TATAACCAGTGTCTACGAAGCGGCTCGAAGTCTCAAACCTCACCTTCTCTGTGCGC	
HTT_4_3076247_3078445_3253432_3258164_RR	HTT	HD_阳性	AGAGTACTTCCTAACTCTACTGTACACTCGAGGATGATCGTCCGACAGCTCCTCCAGC	
HTT_4_3070353_3072847_3212506_3214871_FR	HTT	HD_阴性	GGGTTTCGCCATGTTGGCCAGGCTGGTCTCGAAGTTGATGCATCTGTGCTCAGGTTTGCA	
HTT_4_3088056_3090591_3212506_3214871_FF	HTT	HD_阳性	AGATCTAGTTCACAGTAGCACCAATATATCGACTGTCTCCTGTTGGCCATCTCTCACCCCT	
HTT_4_3088056_3090591_3167292_3170536_FR	HTT	HD_阴性	AGATCTAGTTCACAGTAGCACCAATATATCGAACTCCTGACCTTGTGATCCACCACCTC	

表 12b

Chr	探针位置				探针	PCR-引物 1_ID	PCR_引物 1
	起始 1	末端 1	起始 2	末端 2			
4	3090560	3090591	3182426	3182457	HTT_4_3088056_3090591_3182426_3183765_FR	RD031_037	GAATACCCAGGAATGCTTACTTGAGC
4	3076247	3076278	3182426	3182457	HTT_4_3076247_3078445_3182426_3183765_RR	RD031_061	ACCAAATGCCATCTGGGACACATCCA
4	3062575	3062606	3253401	3253432	HTT_4_3062575_3066389_3251623_3253432_RF	RD031_185	CCCGCTAAGTCCACCCTCTGTGA
4	3076247	3076278	3253432	3253463	HTT_4_3076247_3078445_3253432_3258164_RR	RD031_237	ACCAAATGCCATCTGGGACACATCCA
4	3072816	3072847	3212506	3212537	HTT_4_3070353_3072847_3212506_3214871_FR	RD031_241	AGTCTGCCACTGAGGTAACATAACA
4	3090560	3090591	3214840	3214871	HTT_4_3088056_3090591_3212506_3214871_FF	RD031_329	GAATACCCAGGAATGCTTACTTGAGC
4	3090560	3090591	3167292	3167323	HTT_4_3088056_3090591_3167292_3170536_FR	RD031_333	GAATACCCAGGAATGCTTACTTGAGC

[0435]

表 12c

PCR-引物 2_ID	PCR_引物 2
RD031_039	GATTCAGCCACCACCTTTCACAAG
RD031_063	TGATTCAGCCACCACCTTTCACAA
RD031_187	GAACCGCACTTACCCTCAGCAGT
RD031_239	GGACAAGCAGACACACTACTGAACT
RD031_243	ATCCCTGAAACAGAGAGGACCTCGTG
RD031_331	GAGCCGTCTCATAATAACCTCAGGGT
RD031_335	CAGTGGTTTAGGGCAAAAGAGAGGGAG

表 13a

探针		位点	检测到的环	探针序列 60 mer
HTT_4_3088056_3090591_3112545_3114867_FR		HTT	无差异	AGATCTAGTTCACAGTAGCACCACAATATATCGACATTAAGTTTCATGAGAGTCTTCTATATT
HTT_4_3062575_3066389_3119060_3125597_RR		HTT	无差异	GCGCACAGGAGAAGTGAAGTTTGAGACTTCGAAAAAGGAATAAAATTGAAAAATAGAGGAA
HTT_4_3076247_3078445_3241592_3247577_RR		HTT	无差异	AGAGTACTTCTAAACTCCTACTGTACACTCGATGTTTTGGGTATTGAATGTGGTAAAGTGG
HTT_4_3080953_3084487_3253432_3258164_RR		HTT	无差异	ATCAAGACTGTATGGTACTGGCACAGGATCGAGGATGATCGCTCCGACAGCTCCTCCAGC
HTT_4_3080953_3084487_3241592_3247577_RR		HTT	无差异	ATCAAGACTGTATGGTACTGGCACAGGATCGATGTTTTGGGTATTGAATGTGGTAAAGTGG
HTT_4_3080953_3084487_3112545_3114867_RR		HTT	无差异	ATCAAGACTGTATGGTACTGGCACAGGATCGACATTAAGTTTCATGAGAGTCTTCTATATT
HTT_4_3076247_3078445_3112545_3114867_RF		HTT	无差异	TGTCAGTTAAATTGCAGATATTCGGATCGAGTACAGTAGGAGTTAGGAAGTACTCT
HTT_4_3080953_3084487_3167292_3170536_RR		HTT	无差异	ATCAAGACTGTATGGTACTGGCACAGGATCGAACTCCTGACCTTGATCCACCCACCTC
HTT_4_3038381_3040238_3088056_3090591_RF		HTT	无差异	AGATCTAGTTCACAGTAGCACCACAATATATCGAATCTGGGTGTTTCAAAGACATCAACAAA
HTT_4_3088056_3090591_3253432_3258164_FR		HTT	无差异	AGATCTAGTTCACAGTAGCACCACAATATATCGAGGATGATCGCTCCGACAGCTCCTCCAGC
HTT_4_3062575_3066389_3167292_3170536_FR		HTT	无差异	GGGTTTACCACATGTTGGCCAGGATGGTCTCGAACTCCTGACCTTGTGATCCACCCACCTC
HTT_4_3088056_3090591_3251623_3253432_FF		HTT	无差异	AGATCTAGTTCACAGTAGCACCACAATATATCGAAGCGGCCCTTCGTAGGAGCACTGGTTATA
HTT_4_3088056_3090591_3241592_3247577_FF		HTT	无差异	AGATCTAGTTCACAGTAGCACCACAATATATCGACTAAGGTACGCCCTCCGTGGCTGGGGGCT

[0436]

表 13b

Chr	探针位置				探针	PCR-引物 1_ID	PCR_引物 1
	起始 1	末端 1	起始 2	末端 2			
4	3090560	3090591	3112545	3112576	HTT_4_3088056_3090591_3112545_3114867_FR	RD031_001	GAATACCCAGGAATGCTTACTTGAGC
4	3062575	3062606	3119060	3119091	HTT_4_3062575_3066389_3119060_3125597_RR	RD031_009	GGTCCAGAGAACCAGCACTTACCC
4	3076247	3076278	3241592	3241623	HTT_4_3076247_3078445_3241592_3247577_RR	RD031_025	ACCAAATGCCATCTGGGACACATCCA
4	3080953	3080984	3253432	3253463	HTT_4_3080953_3084487_3253432_3258164_RR	RD031_041	TCCCTCTCAGAAATCCTACTTGGCTTC
4	3080953	3080984	3241592	3241623	HTT_4_3080953_3084487_3241592_3247577_RR	RD031_057	TCCCTCTCAGAAATCCTACTTGGCTTC
4	3080953	3080984	3112545	3112576	HTT_4_3080953_3084487_3112545_3114867_RR	RD031_065	TCCCTCTCAGAAATCCTACTTGGCTTC
4	3076247	3076278	3114836	3114867	HTT_4_3076247_3078445_3112545_3114867_RF	RD031_097	GAACAGTTTGGGGTAGTATGCGGTC
4	3080953	3080984	3167292	3167323	HTT_4_3080953_3084487_3167292_3170536_RR	RD031_149	TCCCTCTCAGAAATCCTACTTGGCTTC
4	3038381	3038412	3090560	3090591	HTT_4_3038381_3040238_3088056_3090591_RF	RD031_157	GAATACCCAGGAATGCTTACTTGAGC
4	3090560	3090591	3253432	3253463	HTT_4_3088056_3090591_3253432_3258164_FR	RD031_205	GAATACCCAGGAATGCTTACTTGAGC
4	3066358	3066389	3167292	3167323	HTT_4_3062575_3066389_3167292_3170536_FR	RD031_261	TCTGGACAAGGTATCTGGAAGGCTGA
4	3090560	3090591	3253401	3253432	HTT_4_3088056_3090591_3251623_3253432_FF	RD031_301	GAATACCCAGGAATGCTTACTTGAGC
4	3090560	3090591	3247546	3247577	HTT_4_3088056_3090591_3241592_3247577_FF	RD031_337	GAATACCCAGGAATGCTTACTTGAGC

[0437]

表 13c

PCR-引物 2_ID	PCR_引物 2
RD031_003	GCACTGCTCGCAATAGCCAAGAAGCTA
RD031_011	CTTGCTGACCCCTCCTATGGATGG
RD031_027	GCACAGAGTTCCAACATTTCTCCAC
RD031_043	GGACAAGCAGACACACTACCTGAACT
RD031_059	GCACAGAGTTCCAACATTTCTCCAC
RD031_067	GCACTGCTCGCAATAGCCAAGAAGCTA
RD031_099	ACCAAATGCCATCTGGGACACATCCA
RD031_151	ACAGTGGTTTAGGGCAAAGAGAGGGA
RD031_159	GAAGAAAGCCTCTGTTTGGTCTGGAAA
RD031_207	GGAAGGACAAGCAGACACACTACCTG
RD031_263	GTGGTTTAGGGCAAAGAGAGGGAGAT
RD031_303	TCCCCGAGTCTGTGATGGCAAACT
RD031_339	ATGGCTACGGAAAGGGCATTCCGGAC

[0438]

表 14

受试者 ID(每位受试者的独特 ID)	疾病或对照	生物样本类型 (例如, FFPE 原代转移, 血清, 血液, DNA 等)	年龄 (在收集时的时间)	性别 (男或女)	采集日期 (TBD)	种族	诊断年龄	诊断方法	CAG 重复的基因检测结果	记录的症状	采集时间时的医药	如果有, 任何其他基因检测结果	HIV 检测结果	HBV 检测结果	HCV 测试结果	梅毒检查结果
Control01	对照	全血 9mL	32	F	2017-11-07	高加索人 (白人)	-	PCR	n1=17 n2=28	无	无	无	阴性	阴性	阴性	阴性
对照 02	对照	全血 9mL	65	F	2017-11-10	高加索人 (白人)	-	PCR	n1<35 n2<35	无	无	无	阴性	阴性	阴性	阴性
对照 03	对照	全血 9mL	41	M	2017-11-13	高加索人 (白人)	-	PCR	n1<35 n2<35	无	无	无	阴性	阴性	阴性	阴性
对照 04	对照	全血 9mL	19	M	2017-11-10	高加索人 (白人)	-	PCR	n1=17 n2=21	无	无	无	阴性	阴性	阴性	阴性
对照 05	对照	全血 9mL	46	M	2017-11-10	高加索人 (白人)	-	PCR	n1=15 n2=17	无	无	无	阴性	阴性	阴性	阴性
对照 06	对照	全血 9mL	29	F	2017-11-13	高加索人 (白人)	-	PCR	n1<35 n2<35	无	无	无	阴性	阴性	阴性	阴性
对照 07	对照	全血 9mL	47	F	2017-11-14	高加索人 (白人)	-	PCR	n1<35 n2<35	无	无	无	阴性	阴性	阴性	阴性
对照 08	对照	全血 9mL	22	M	2017-11-14	高加索人 (白人)	-	PCR	n1=19 n2=22	无	无	无	阴性	阴性	阴性	阴性
对照 09	对照	全血 9mL	35	F	2017-11-15	高加索人 (白人)	-	PCR	n1<35 n2<35	无	无	无	阴性	阴性	阴性	阴性
对照 11	对照	全血 9mL	33	M	2017-11-16	高加索人 (白人)	-	PCR	n1<35 n2<35	无	无	无	阴性	阴性	阴性	阴性
HD01	亨廷顿病	全血 9mL	32	M	2017-10-30	高加索人 (白人)	27	PCR	n1<35 n2=48	易怒, 舞蹈病	无	无	阴性	阴性	阴性	阴性
HD02	亨廷顿病	全血 9mL	32	F	2017-10-30	高加索人 (白人)	29	PCR	n1<35 n2=42	无	无	无	阴性	阴性	阴性	阴性
HD03	亨廷顿病	全血 9mL	28	F	2017-11-01	高加索人 (白人)	25	PCR	n1=203 n2=46	易怒	无	无	阴性	阴性	阴性	阴性
HD12	亨廷顿病	全血 9mL	50	M	2017-11-02	高加索人 (白人)	46	PCR	n1=203 n2=48	易怒, 舞蹈病	无	无	阴性	阴性	阴性	阴性
HD05	亨廷顿病	全血 9mL	52	M	2017-11-02	高加索人 (白人)	49	PCR	n1<35 n2=45	易怒, 舞蹈病	无	无	阴性	阴性	阴性	阴性
HD06	亨廷顿病	全血 9mL	43	M	2017-11-02	高加索人 (白人)	35	PCR	n1=22 n2=40	易怒, 舞蹈病	丁苯那唑	无	阴性	阴性	阴性	阴性
HD08	亨廷顿病	全血 9mL	23	F	2017-11-02	高加索人 (白人)	-	PCR	n1<35 n2=44	无	无	无	阴性	阴性	阴性	阴性
HD09	亨廷顿病	全血 9mL	21	F	2017-11-02	高加索人 (白人)	-	PCR	n1=35 n2=41	无	无	无	阴性	阴性	阴性	阴性
HD10	亨廷顿病	全血 9mL	38	F	2017-11-02	高加索人 (白人)	36	PCR	n1<35 n2=45	易怒	无	无	阴性	阴性	阴性	阴性
HD11	亨廷顿病	全血 9mL	34	M	2017-11-15	高加索人 (白人)	32	PCR	n1=24 n2=43	易怒, 舞蹈病	舍曲林	无	阴性	阴性	阴性	阴性

[0439]

表 15

区域	hg19 坐标	hg38 坐标
锚定点 CAG 重复位点	chr4:3055890-3097657	chr4:3054163-3095930
区域 1	chr4:3232680-3259897	chr4:3230953-3258170
区域 2	chr4:3107637-3130235	chr4:3105908-3128508
区域 3	chr4:3121999-3218545	chr4:3211272-3216818
区域 4	chr4:3161885-3185826	chr4:3160158-3184009
区域 5	chr4:3033588-3050187	chr4:3031861-3048460

表 16

	健康对照 (N=10)	亨廷顿氏病 (N=10)
性别		
男性 (N, (%))	5 (50)	5 (50)
女性 (N, (%))	5 (50)	5 (50)
种族		
白人 (N, (%))	10 (100)	10 (100)
亨廷顿病类型		
有症状 (N, (%))	N/A	7 (70)
无症状 (N, (%))	N/A	3 (30)
CAG 重复长度 (平均, (SD))	25.7 (5.4)	44.2 (2.6)
诊断时年龄 (平均, (SD))	N/A	34.9* (8.1)
采样时年龄 (平均, (SD))	36.9 (12.8)	35.3 (10.0)
疾病时长** (年数) (平均, (SD))	N/A	3.8 (1.9)
%报告易怒	N/A	70
%报告舞蹈症	N/A	50
* 诊断时年龄对于 10 个 HD 患者中的 2 个不可获得		
N/A =不适用		

[0440]

表 17

相互作用ID	区域5	锚定	区域2	区域4	区域3	区域1
RD031_025/27		✓				✓
RD031_041/43		✓				✓
RD031_057/59		✓				✓
RD031_061/63		✓		✓		
RD031_065/67		✓	✓			
RD031_097/99		✓	✓			
RD031_009/11		✓	✓			
RD031_149/151		✓		✓		
RD031_237/239		✓		✓		✓
RD031_261/263		✓		✓		
RD031_157/159	✓	✓				
RD031_001/3		✓	✓			
RD031_329/331		✓			✓	
RD031_301/303		✓				✓
RD031_337/339		✓				✓
RD031_037/39		✓		✓		
RD031_185/187		✓				✓
RD031_205/207		✓				✓
RD031_241/243		✓			✓	
RD031_333/335		✓		✓		

[0441]

表 18a

标志物	HD01	HD03	HD05	HD06	HD10	HD11	HD12	检出率
RD031_237.239	0	0	0	1	1	1	1	57.14
RD031_329.331	0	1	1	1	1	1	1	85.71
RD031_185.187	0	0	1	1	0	1	0	42.86

表 18b

237.239	症状	无症状	风险率
环存在	4	1	0.8
环不存在	3	12	4
			优势比 16

表 18c

329.331	症状	无症状	风险率
环存在	6	1	0.85714
环不存在	1	12	11.1429
			优势比 72

表 18d

185.187	症状	无症状	风险率
环存在	3	1	0.75
环不存在	4	12	3
			优势比 9

[0442]

表 19

```

>GRM7_3_7259138_7267165_7394377_7401227_RF
TTGGCAAACTACTTATTCGTATGAGTTTCAGTGTGCACATCCACCAATGTGGATAATAATATGG
TTAACTCCTAGGATGTTGTGAGAAATTGAATAATAATATCATCTTAGCCAGTATAATCATTAAATT
TTTTTTTAGTAAGCATTAGCTGGAAATCATTGGTATTGTTTTTATTCTTAICTACCCCTCAAACATTGA
AAGTGTCTGGAGCAGAAGTTGTATATTACATACTTTCAGAAATCTCTACCATGGGGCCCTTTTGCTT
TAATTAATTCCTGTATAAATTAATTCCTCTATTATGCCATTTTCIGATTTTCATCCAGACACACTTAA
TGTATATGACGAAGAAGCAATCCCTGGTGACTAAAATGAGATTGCAATTTCTTTGGACCAATTTGGCC
AGCATGCCAAACACTGGTAAAATGGCCATTCCAAAGCTCACATGGCTGATTGATAAATTCACACTTT
GGGCTCAGTTTGGCCATCTGTAAAGTGTAAAGTGTGATCAAAATGAGCAAGTGTATTTATTTAATTTAA
TACTTTTATTTTAGATTTGGGATACATGTGTAAAGTTTGTCTATATAGGTAATTCATGTCATGGGAGT
TTATTGTAAGAATTATTTTCATCACCAGGTACTAAGCCTAGTACACAATAGTTATTTTTTTCTGATTC
TTTTCCCTCCCAACTTAAACCAAGTGGAGSSTAGGTTCCAGTGTCTGTTGTCCTCCCTCTTTGTGTC
TGTGTTCTCATCATTAGCCCCCACTTAAAGTGTATAATATGTTGGT

```

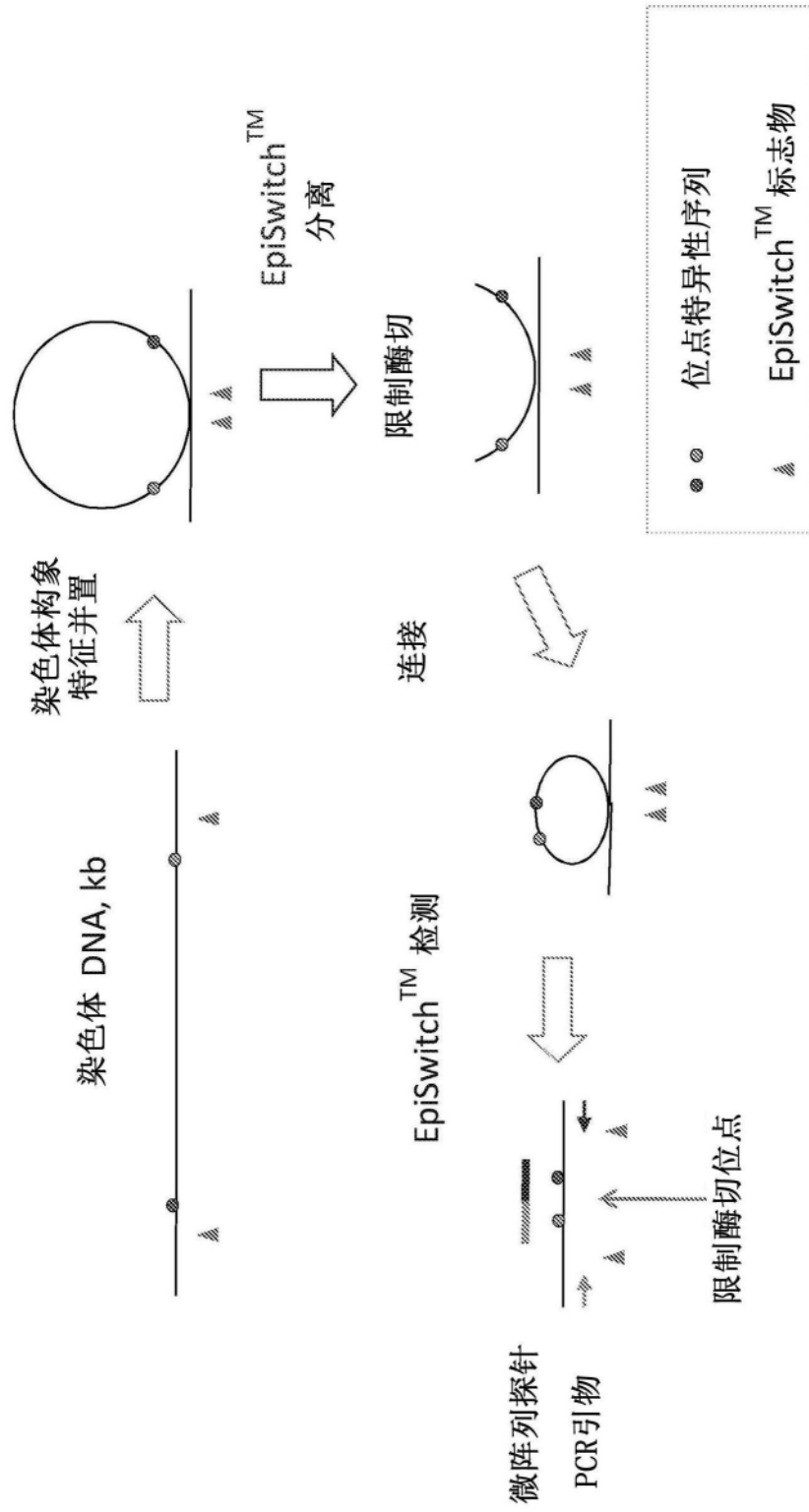


图1

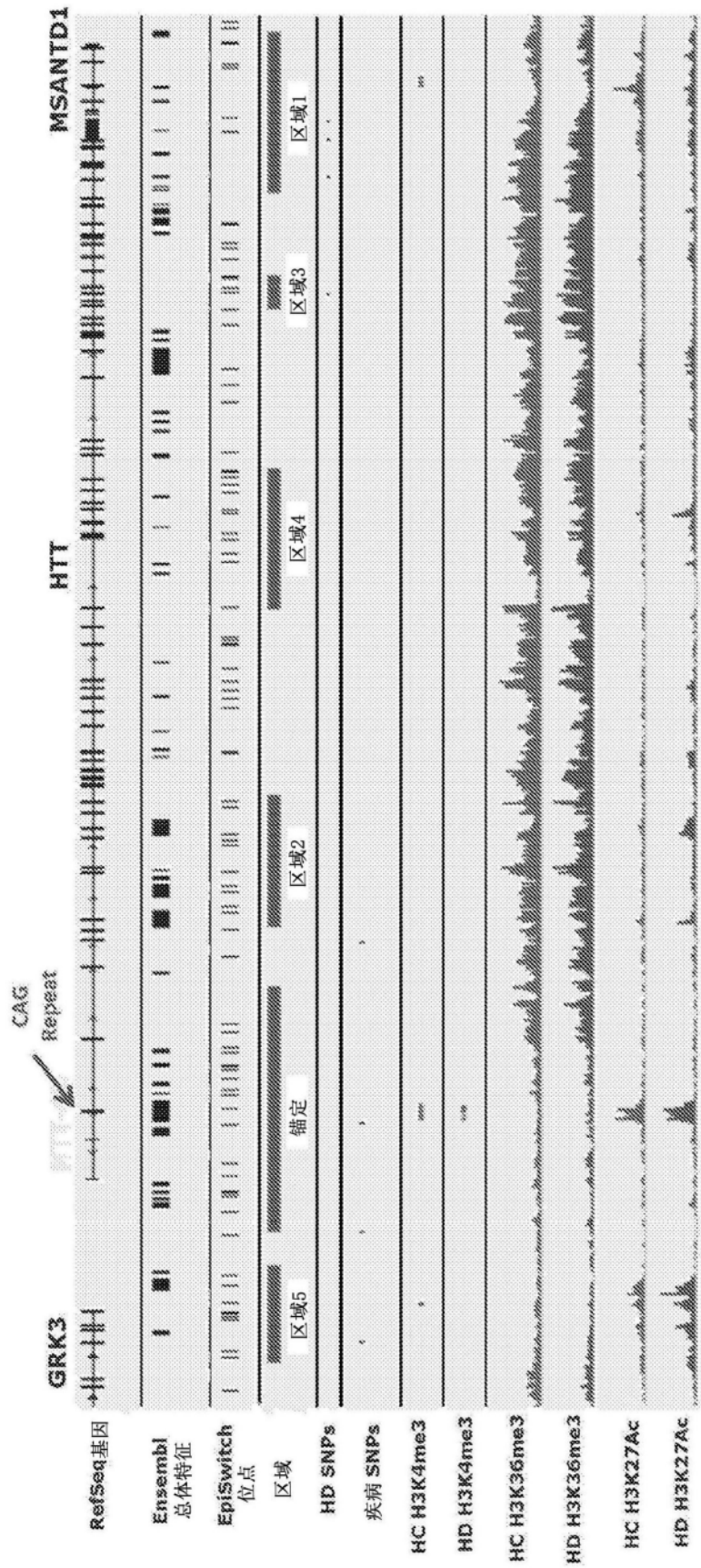


图2

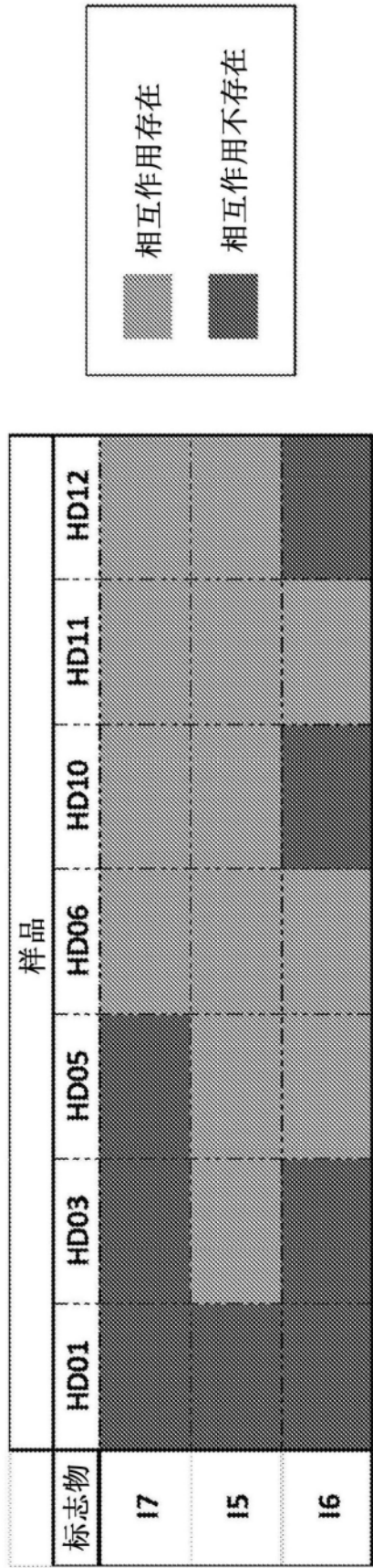


图3

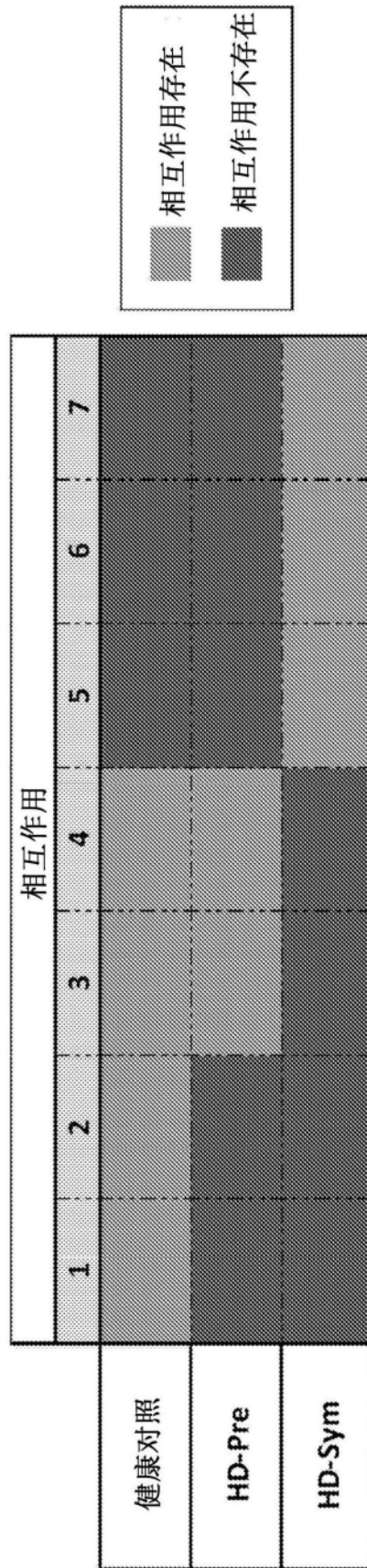


图4

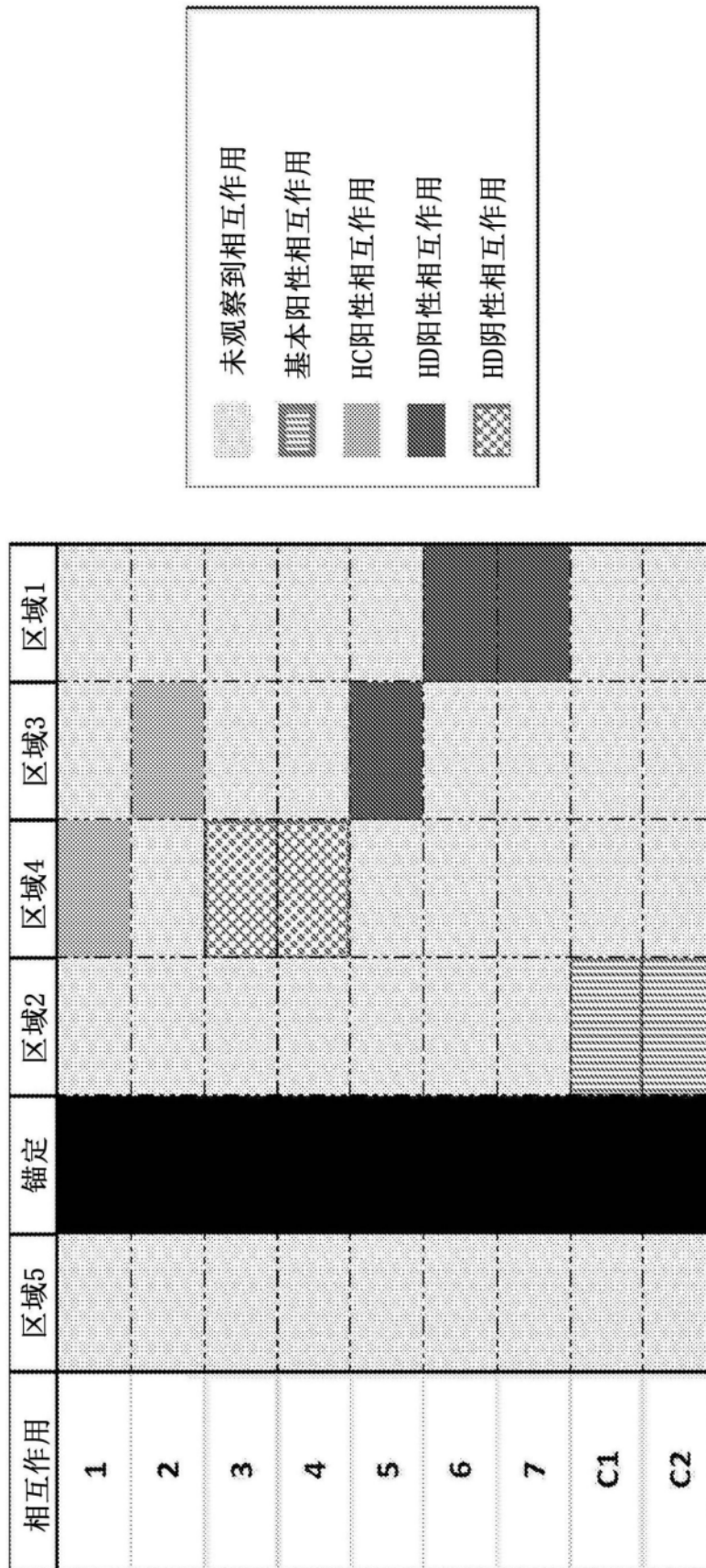


图5