



(12) 发明专利

(10) 授权公告号 CN 109790198 B

(45) 授权公告日 2022. 09. 09

(21) 申请号 201780053034.9
(22) 申请日 2017.09.01
(65) 同一申请的已公布的文献号
申请公布号 CN 109790198 A
(43) 申请公布日 2019.05.21
(30) 优先权数据
62/383,165 2016.09.02 US
(85) PCT国际申请进入国家阶段日
2019.02.27
(86) PCT国际申请的申请数据
PCT/US2017/049915 2017.09.01
(87) PCT国际申请的公布数据
W02018/045322 EN 2018.03.08
(73) 专利权人 梅约医学教育与研究基金会
地址 美国明尼苏达州
专利权人 精密科学发展有限责任公司
(72) 发明人 威廉·R·泰勒

大卫·A·阿尔奎斯特
道格拉斯·W·马奥尼
约翰·B·基谢尔
特蕾西·C·亚卜
哈特姆·T·阿拉维
玛丽亚·贾库莫普洛斯
格雷厄姆·P·利德加德

(74) 专利代理机构 成都超凡明远知识产权代理
有限公司 51258
专利代理师 魏彦 洪玉姬

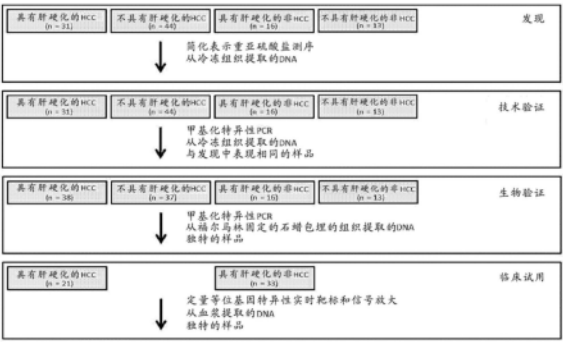
(51) Int.Cl.
C12Q 1/6886 (2018.01)
G16C 60/00 (2019.01)
G16C 20/00 (2019.01)
C07H 21/04 (2006.01)

审查员 王少华

权利要求书4页 说明书48页
序列表22页 附图47页

(54) 发明名称
检测肝细胞癌

(57) 摘要
本文提供了用于肝细胞癌筛查的技术并且特别地但不排他地涉及用于检测肝细胞癌的存在的方法、组合物以及相关用途。



1. 针对将肝细胞癌与正常对照区分开的两个或多个基因的引物制备用于表征生物样品的试剂盒的用途,其中所述两个或多个基因包括基因HOXA1和选自DAB2IP、TSPYL5、ACP1、BDH1、Chr12.133、CLEC11A、DBNL、EFNB2、LRRC4、SPINT2、CCNJ_3707、CCNJ_3124、PFKP、SCRN1以及ECE1中的一个或多个基因,所述用途包括:

如下测量人个体的生物样品中的所述两个或多个基因的CpG位点的甲基化水平:

用重亚硫酸盐处理所述生物样品中的基因组DNA;

使用针对所选择的两个或多个基因的引物组扩增所述重亚硫酸盐处理过的基因组DNA;并且

通过甲基化特异性PCR、定量甲基化特异性PCR、甲基化敏感性DNA限制酶分析、定量重亚硫酸盐焦磷酸测序或重亚硫酸盐基因组测序PCR测定所述CpG位点的甲基化水平,

其中,取决于所选择的两个或多个基因,使用以下引物扩增重亚硫酸盐处理的基因组DNA:

特异于HOXA1引物组,所述特异于HOXA1引物组特异性结合包括染色体7坐标27136145-27136425的遗传区域的至少一部分;

特异于DAB2IP引物组,所述特异于DAB2IP引物组特异性结合包括染色体9坐标124461305-124461420的遗传区域的至少一部分;

特异于TSPYL5引物组,所述特异于TSPYL5引物组特异性结合包括染色体8坐标98289858-98290220的遗传区域的至少一部分;

特异于ACP1引物组,所述特异于ACP1引物组特异性结合包括选自264204-264250和264087-264151的染色体2坐标的遗传区域的至少一部分;

特异于BDH1引物组,所述特异于BDH1引物组特异性结合包括选自197282831-197282922和197281722-197281827的染色体3坐标的遗传区域的至少一部分;

特异于Chr12.133引物组,所述特异于Chr12.133引物组特异性结合包括染色体12坐标133484978-133485739的遗传区域的至少一部分;

特异于CLEC11引物组,所述特异于CLEC11引物组特异性结合包括染色体19坐标51228217-51228732的遗传区域的至少一部分;

特异于DBNL引物组,所述特异于DBNL引物组特异性结合包括染色体7坐标44080227-44080310的遗传区域的至少一部分;

特异于EFNB2引物组,所述特异于EFNB2引物组特异性结合包括选自107188064-107188108、107188196-107188212、107187748-107187766、107187768-107187785、107188435-107188478、107188935-107189019、107188790-107188904、107187702-107187716和107188503-107188558的染色体13坐标的遗传区域的至少一部分;

特异于LRRC4引物组,所述特异于LRRC4引物组特异性结合包括选自127671917-127672169和127672388-127672445的染色体7坐标的遗传区域的至少一部分;

特异于SPINT2引物组,所述特异于SPINT2引物组特异性结合包括染色体19坐标38755130-38755164 的遗传区域的至少一部分;

特异于CCNJ_3707引物组,所述特异于CCNJ_3707引物组特异性结合包括染色体10坐标97803707-97803733的遗传区域的至少一部分;

特异于CCNJ_3124引物组,所述特异于CCNJ_3124引物组特异性结合包括染色体10坐标

97803124-97803152的遗传区域的至少一部分；

特异于PFKP引物组，所述特异于PFKP引物组特异性结合包括选自3110620-3110694、3110554-3110609、3111174-3111244和3110913-3110978的染色体10坐标的遗传区域的至少一部分；

特异于SCRN1引物组，所述特异于SCRN1引物组特异性结合包括选自30029222-30029267、30029115-30029160和30029187-30029209的染色体7坐标的遗传区域的至少一部分；和

特异于ECE1引物组，所述特异于ECE1引物组特异性结合包括染色体1坐标21616801-21616877的遗传区域的至少一部分。

2.如权利要求1所述的用途，其还包括

将所述甲基化水平与没有肝细胞癌的对照样品中的对应基因集合的甲基化水平进行比较；以及

当在所述两个或多个基因中所测量的甲基化水平高于在相应对照样品中所测量的甲基化水平时确定所述个体患有肝细胞癌。

3.如权利要求1所述的用途，其中如果所述生物样品为组织样品，那么使用以下针对所述所选择的两个或多个基因的引物组：

对于ACPI1，由SEQ ID NO： 10和11或SEQ ID NO： 58和59组成的引物组，

对于BDH1，由SEQ ID NO： 13和14或SEQ ID NO： 61和62组成的引物组，

对于Chr12.133，由SEQ ID NO： 25和26或SEQ ID NO： 49和50组成的引物组，

对于CLEC11A，由SEQ ID NO： 7和8或SEQ ID NO： 52和53组成的引物组，

对于DAB2IP，由SEQ ID NO： 4和5或SEQ ID NO： 46和47组成的引物组，

对于DBNL，由SEQ ID NO： 34和35或SEQ ID NO： 82和83组成的引物组，

对于EFNB2，由SEQ ID NO： 37和38或SEQ ID NO： 64和65组成的引物组，

对于HOXA1，由SEQ ID NO： 28和29或SEQ ID NO： 88和89组成的引物组，

对于LRRC4，由SEQ ID NO： 40和41或SEQ ID NO： 85和86组成的引物组，

对于SPINT2，由SEQ ID NO： 31和32或SEQ ID NO： 55和56组成的引物组，

对于TSPYL5，由SEQ ID NO： 1和2或SEQ ID NO： 43和44组成的引物组，

对于CCNJ_3707，由SEQ ID NO： 67和68组成的引物组，

对于CCNJ_3124，由SEQ ID NO： 70和71组成的引物组，

对于PFKP，由SEQ ID NO： 73和74组成的引物组，

对于SCRN1，由SEQ ID NO： 76和77组成的引物组，以及

对于ECE1，由SEQ ID NO： 79和80组成的引物组。

4.如权利要求1所述的用途，其中所述生物样品为血浆样品或组织样品。

5.如权利要求1所述的用途，其中所述CpG位点存在于编码区或调控区中。

6.如权利要求1所述的用途，其中所述测量两个或多个基因的CpG位点的甲基化水平包括选自由测定所述CpG位点的甲基化得分和测定所述CpG位点的甲基化频率组成的组的测定。

7.如权利要求1所述的用途，其中所述两个或多个基因为HOXA1和选自Chr12.133、CLEC11A、TSPYL5以及CCNJ_3707的一个或多个基因。

8. 针对将肝细胞癌与正常对照区分开的两个或更多个基因的引物制备用于表征血浆样品的试剂盒的用途,其中所述两个或多个基因包括基因H0XA1和选自DAB2IP、TSPYL5、ACP1、BDH1、Chr12.133、CLEC11A、DBNL、EFNB2、LRRC4、SPINT2、CCNJ_3707、CCNJ_3124、PFKP、SCRN1以及ECE1中的一个或多个基因,所述用途包括:

(a) 如下测量人个体的血浆样品中的H0XA1的CpG位点的甲基化水平以及选自以下的所述一个或更多个基因的CpG位点的甲基化水平:DAB2IP、TSPYL5、ACP1、BDH1、Chr12.133、CLEC11A、DBNL、EFNB2、LRRC4、SPINT2、CCNJ_3707、CCNJ_3124、PFKP、SCRN1以及ECE1:

用重亚硫酸盐处理所述生物样品中的基因组DNA;

使用以下针对H0XA1和所选择的一个或更多个基因的引物组扩增所述重亚硫酸盐处理过的基因组DNA:

对于ACP1,由SEQ ID NO: 10和11或SEQ ID NO: 58和59组成的引物组,

对于BDH1,由SEQ ID NO: 13和14或SEQ ID NO: 61和62组成的引物组,

对于Chr12.133,由SEQ ID NO: 25和26或SEQ ID NO: 49和50组成的引物组,

对于CLEC11A,由SEQ ID NO: 7和8或SEQ ID NO: 52和53组成的引物组,

对于DAB2IP,由SEQ ID NO: 4和5或SEQ ID NO: 46和47组成的引物组,

对于DBNL,由SEQ ID NO: 34和35或SEQ ID NO: 82和83组成的引物组,

对于EFNB2,由SEQ ID NO: 37和38或SEQ ID NO: 64和65组成的引物组,

对于H0XA1,由SEQ ID NO: 28和29或SEQ ID NO: 88和89组成的引物组,

对于LRRC4,由SEQ ID NO: 40和41或SEQ ID NO: 85和86组成的引物组,

对于SPINT2,由SEQ ID NO: 31和32或SEQ ID NO: 55和56组成的引物组,

对于TSPYL5,由SEQ ID NO: 1和2或SEQ ID NO: 43和44组成的引物组,

对于CCNJ_3707,由SEQ ID NO: 67和68组成的引物组,

对于CCNJ_3124,由SEQ ID NO: 70和71组成的引物组,

对于PFKP,由SEQ ID NO: 73和74组成的引物组,

对于SCRN1,由SEQ ID NO: 76和77组成的引物组,以及

对于ECE1,由SEQ ID NO: 79和80组成的引物组;并且

通过甲基化特异性PCR、定量甲基化特异性PCR、甲基化敏感性DNA限制酶分析、定量重亚硫酸盐焦磷酸测序或重亚硫酸盐基因组测序PCR测定所述CpG位点的甲基化水平;

(b) 将所述甲基化水平与没有胃癌的对照样品中的对应基因集合的甲基化水平进行比较;以及

(c) 当在H0XA1和所述两个或更多个基因中所测量的甲基化水平高于在相应对照样品中所测量的甲基化水平时确定所述个体患有肝细胞癌。

9. 如权利要求8所述的用途,其中所述CpG位点存在于编码区或调控区中。

10. 如权利要求8所述的用途,其中所述测量所述两个或更多个基因的CpG位点的甲基化水平包括选自由测定所述CpG位点的甲基化得分和测定所述CpG位点的甲基化频率组成的组的测定。

11. 如权利要求8所述的用途,其中所述两个或更多个基因为H0XA1和选自Chr12.133、CLEC11A、TYSPL5以及CCNJ_3707的一个或多个基因。

12. 一种试剂盒,其包含以下中的两者或更多者:

- 1) 重亚硫酸盐试剂;
- 2) 对照核酸,所述对照核酸包含HOXA1的DMR的序列,并且具有与未患癌症的受试者相关的甲基化状态;
- 3) 以下针对HOXA1的引物组:由SEQ ID NO: 28和29或SEQ ID NO: 88和89组成的引物组。

13. 根据权利要求12所述的试剂盒,

所述对照核酸还包含来自选自以下的DMR的序列:DAB2IP、TSPYL5、ACP1、BDH1、Chr12.133、CLEC11A、DBNL、EFNB2、LRRC4、SPINT2、CCNJ_3707、CCNJ_3124、PFKP、SCRN1以及ECE1,并且具有与未患癌症的受试者相关的甲基化状态;

所述试剂盒还包含以下针对EMX1、DAB2IP、TSPYL5、ACP1、BDH1、Chr12.133、CLEC11A、DBNL、EFNB2、LRRC4、SPINT2、CCNJ_3707、CCNJ_3124、PFKP、SCRN1以及ECE1的引物组中的一者或者多者:

对于ACP1,由SEQ ID NO: 10和11或SEQ ID NO: 58和59组成的引物组,

对于BDH1,由SEQ ID NO: 13和14或SEQ ID NO: 61和62组成的引物组,

对于Chr12.133,由SEQ ID NO: 25和26或SEQ ID NO: 49和50组成的引物组,

对于CLEC11A,由SEQ ID NO: 7和8或SEQ ID NO: 52和53组成的引物组,

对于DAB2IP,由SEQ ID NO: 4和5或SEQ ID NO: 46和47组成的引物组,

对于DBNL,由SEQ ID NO: 34和35或SEQ ID NO: 82和83组成的引物组,

对于EFNB2,由SEQ ID NO: 37和38或SEQ ID NO: 64和65组成的引物组,

对于LRRC4,由SEQ ID NO: 40和41或SEQ ID NO: 85和86组成的引物组,

对于SPINT2,由SEQ ID NO: 31和32或SEQ ID NO: 55和56组成的引物组,

对于TSPYL5,由SEQ ID NO: 1和2或SEQ ID NO: 43和44组成的引物组,

对于CCNJ_3707,由SEQ ID NO: 67和68组成的引物组,

对于CCNJ_3124,由SEQ ID NO: 70和71组成的引物组,

对于PFKP,由SEQ ID NO: 73和74组成的引物组,

对于SCRN1,由SEQ ID NO: 76和77组成的引物组,以及

对于ECE1,由SEQ ID NO: 79和80组成的引物组。

检测肝细胞癌

[0001] 相关申请的交叉引用

[0002] 本申请要求2016年9月2日提交的美国临时专利申请62/383,165的优先权益,该临时专利申请以全文引用的方式并入本文中。

发明领域

[0003] 本文提供了用于肝细胞癌筛查的技术并且特别地(但不排他地)涉及用于检测肝细胞癌的存在的方法、组合物以及相关用途。

背景技术

[0004] 肝细胞癌(HCC)为一种原发性肝脏恶性肿瘤且主要发生在患有潜在慢性肝病和肝硬化的患者中。起源细胞被认为是肝干细胞(参见Alison MR.Stem Cell Rev.2005.1(3):253-60)。肿瘤随局部扩张、肝内扩散以及远端转移而进展。

[0005] HCC目前为全世界癌症死亡的第三大主要原因,超过500,000人受到影响。在亚洲和非洲HCC发病率最高,在这些地方B型肝炎和C型肝炎的地方性高盛行率使得发生慢性肝病和后续发生HCC极其容易。

[0006] 在过去几十年间,HCC的表现已发生显著变化。虽然在过去,HCC通常在晚期表现为具有右上腹疼痛、体重减轻以及失代偿性肝病迹象,但目前由于使用横截面成像研究和血清甲胎蛋白(AFP)测量对患有已知肝硬化的患者的常规筛查已越来越多地在早得多的阶段被识别。

[0007] 预期在未来几年里HCC的威胁将持续增长(参见Llovet JM等,Liver Transpl.2004年2月10日(2增刊1):S115-20)。因此,极期需要对HCC的早期检测以改善这些患者的存活。

发明内容

[0008] 肝细胞癌(HCC)为全世界第2大致命癌症。在早期检测的情况下存活率提高,并且需要准确非侵入性筛查工具。急迫需要将递送准确、可承受并且安全的筛查工具以对最早阶段的HCC进行症状前检测的新方法。

[0009] 本发明解决了此需要。实际上,本发明提供新颖的甲基化DNA标记物,所述新颖的甲基化DNA标记物可辨别HCC与正常对照(具有和不具有肝硬化的对照)。

[0010] 已研究了在大多数肿瘤类型的组织中作为潜在类别的生物标记物的甲基化DNA。在许多情况情况下,DNA甲基转移酶在胞嘧啶-磷酸-鸟嘌呤(CpG)岛位点处向DNA添加甲基作为基因表达的表现遗传控制。在生物学上有吸引力的机制中,在肿瘤抑制基因的启动子区中获得的甲基化事件被认为使表达沉默,因此促进肿瘤形成。DNA甲基化可为比RNA或蛋白质表达更具化学和生物学稳定性的诊断工具(Laird(2010)Nat Rev Genet 11:191-203)。此外,在如散发性结肠癌等其他癌症中,甲基化标记物提供优良特异性且比个别DNA突变更广泛地提供信息并且更敏感(参见Zou等(2007)Cancer Epidemiol Biomarkers

Prev 16:2686-96)。

[0011] CpG岛的分析当应用于动物模型和人细胞系时产生重大发现。举例来说,Zhang和同事发现来自相同CpG岛的不同部分的扩增子可具有不同水平的甲基化(Zhang等(2009) PLoS Genet 5:e1000438)。另外,在高度甲基化与未甲基化序列之间甲基化水平呈双峰分布,这进一步支持DNA甲基转移酶活性的二元开关样模式(Zhang等(2009) PLoS Genet 5:e1000438)。体内鼠组织和体外细胞系的分析证实仅约0.3%的高CpG密度启动子(HCP,定义为在300碱基对区域内具有>7%的CpG序列)为甲基化的,而低CpG密度(定义为在300碱基对区域内具有<5%的CpG序列)的区域倾向于以动态组织特异性模式频繁甲基化(Meissner等(2008) Nature 454:766-70)。HCP包括普遍存在的看家基因和高度调控的发育基因的启动子。在HCP位点之中,以>50%甲基化的为若干确定的标记物,诸如Wnt 2、NDRG2、SFRP2以及BMP3(Meissner等(2008) Nature 454:766-70)。

[0012] 在本发明的研发实施方案过程中进行的实验将来自患有HCC的受试者的血浆的DNA标记物的甲基化状态与来自对照受试者(例如患有肝硬化的受试者或正常受试者)的相同DNA标记物的甲基化状态进行了比较。此类实验确定并验证了辨别HCC与此类对照组的甲基化DNA标记物候选物。

[0013] 因此,本文提供了用于HCC筛查的技术(例如监视)并且特别地(但不排他地)涉及用于检测HCC的存在的方法、组合物以及相关用途。

[0014] 鉴定了能够检测HCC的标记物和/或标记物组(参见实施例I、II以及III)(ACP1、BDH1、Chr12.133、CLEC11A、DAB2IP、DBNL、EMX1、EFNB2、HOXA1、LRRC4、SPINT2、TSPYL5、CCNJ_3707、CCNJ_3124、PFKP、SCRN1以及ECE1)。

[0015] 如本文所描述,所述技术提供了许多以高辨别力检测受试者中HCC的存在的甲基化DNA标记物及其子集(例如2、3、4、5、6、7、8、9、10、11、12种标记物的集合)。实验对候选标记物应用选择过滤器,以鉴定提供高信噪比和低背景水平的标记物,以例如当出于筛查或诊断(例如HCC筛查或诊断)的目的分析介质(例如血浆)时以提供高特异性。

[0016] 在一些实施方案中,所述技术涉及评估生物样品(例如血浆样品)中本文中所鉴定的标记物中的一者或多者的存在和甲基化状态。这些标记物包含如本文所论述的例如如表1和表4中所提供的一个或多个差异甲基化区(DMR)。在所述技术的实施方案中评估了甲基化状态。因此,本文提供的技术在测量基因的甲基化状态的方法方面不受限制。举例来说,在一些实施方案中,通过基因组扫描法来测量甲基化状态。举例来说,一种方法涉及限制性标志基因组扫描(参见Kawai等(1994) Mol. Cell. Biol. 14:7421-7427),并且另一实例涉及甲基化敏感性任意引物PCR(参见Gonzalzo等(1997) Cancer Res. 57:594-599)。在一些实施方案中,通过用甲基化敏感性限制酶消化基因组DNA随后对所关注的区域进行南方分析(Southern analysis)(消化-南方法)监测特定CpG位点处的甲基化模式的变化。在一些实施方案中,分析甲基化模式的变化涉及基于PCR的方法,所述基于PCR的方法涉及用甲基化敏感性限制酶消化基因组DNA然后进行PCR扩增(参见Singer-Sam等(1990) Nucl. Acids Res. 18:687)。此外,已报导利用对DNA的重亚硫酸盐处理作为甲基化分析的起始点的其他技术。这些技术包括甲基化特异性PCR(MSP)(参见Herman等(1992) Proc. Natl. Acad. Sci. USA 93:9821-9826)以及从重亚硫酸盐转化的DNA扩增的PCR产物的限制酶消化(参见Sadri和Hornsby(1996) Nucl. Acids Res. 24:5058-5059;以及Xiong和

Laird (1997) Nucl. Acids Res. 25:2532-2534)。已研发用于检测基因突变 (参见Kuppuswamy等 (1991) Proc. Natl. Acad. Sci. USA 88:1143-1147) 以及定量等位基因特异性表达 (Szabo和Mann (1995) Genes Dev. 9:3097-3108; 以及Singer-Sam等 (1992) PCR Methods增刊1:160-163) 的PCR技术。此类技术使用内部引物, 所述内部引物退火至PCR产生的模板并且在所要分析的单一核苷酸的5' 立即终止。在一些实施方案中使用如美国专利号7,037,650中所描述的使用“定量Ms-SNuPE分析”的方法。

[0017] 在评估甲基化状态后, 常常以在特定位置处 (例如在单一核苷酸处、在特定区域或基因座处、在所关注的较长序列 (例如DNA的长达约100-bp、200-bp、500-bp、1000-bp的子序列或更长序列) 处甲基化的个别DNA链相对于样品中包含所述特定位置的总DNA群体的分数或百分比来表示甲基化状态。传统上, 通过使用校正物的PCR测定未甲基化核酸的量。然后, 将已知量的DNA用重亚硫酸盐处理并且使用实时PCR或其他指数扩增 (例如QuARTS分析, 例如如美国专利号8,361,720、8,715,937以及8,916,344所提供) 测定所得甲基化特异性序列。

[0018] 举例来说, 在一些实施方案中方法包括通过使用外部标准物产生未甲基化靶标的标准曲线。由至少两个点构建标准曲线并且使未甲基化DNA的实时Ct值与已知定量标准物相关联。然后, 由至少两个点和外部标准物构建甲基化靶标的第二标准曲线。此第二标准曲线使甲基化DNA的Ct与已知定量标准物相关联。接着, 针对甲基化和未甲基化群体测定测试样品Ct值并且由通过头两个步骤制备的标准曲线计算DNA的基因组当量。由相对于群体中的总DNA量甲基化DNA的量 (例如 (甲基化DNA的数目) / (甲基化DNA的数目 + 未甲基化DNA的数目) × 100) 计算所关注位点处的甲基化百分比。

[0019] 本文还提供用于实践所述方法的组合物和试剂盒。举例来说, 在一些实施方案中, 单独或以集合形式 (例如用于扩增多种标记物的引物对集合) 提供对一种或多种标记物具特异性的试剂 (例如引物、探针)。还可提供用于进行检测分析的额外的试剂 (例如用于进行QuARTS、PCR、测序、重亚硫酸盐或其他分析的酶、缓冲液、阳性和阴性对照)。在一些实施方案中, 提供含有必需、足够或适用于执行方法的一种或多种试剂的试剂盒。还提供含有所述试剂的反应混合物。还提供含有可彼此添加和/或添加至测试样品中以完成反应混合物的多种试剂的主混合试剂集合。

[0020] 在一些实施方案中, 本文所描述的技术与被设计成进行如由本文所描述的方法所提供的一系列算术或逻辑操作的可编程机器相关。举例来说, 所述技术的一些实施方案与计算机软件和/或计算机硬件相关 (例如在其中实现)。在一个方面中, 所述技术涉及包括以下各项的计算机: 存储器形式、用于进行算术和逻辑操作的元件以及用于执行一系列指令 (例如如本文所提供的方法) 的处理元件 (例如微处理器), 以读取、操纵以及存储数据。在一些实施方案中, 微处理器为用于测定 (例如一个或多个DMR, 例如如表1和表4中所提供的DMR 1-400的) 甲基化状态; 比较 (例如一个或多个DMR, 例如如表1和表4中所提供的DMR 1-400的) 甲基化状态; 产生标准曲线; 测定Ct值; 计算 (例如一个或多个DMR, 例如如表1和表4中所提供DMR 1-400的) 甲基化的分数、频率或百分比; 鉴定CpG岛; 测定分析或标记物的特异性和/或灵敏度; 计算ROC曲线和相关AUC; 序列分析; 如本文所描述或如本领域已知的所有用途的系统的一部分。

[0021] 在一些实施方案中, 微处理器或计算机在算法中使用甲基化状态数据来预测癌症

部位。

[0022] 在一些实施方案中,软件或硬件组件接收多个分析(例如测定例如如表1和表4中所提供的多个DMR的甲基化状态)的结果并且基于多个分析的结果确定要向使用者报告的指示癌症危险的单一值结果。相关实施方案基于多个分析(例如测定多种标记物(诸如多个DMR,例如如表1和表4中所提供的多个DMR)的甲基化状态)的结果的数学组合(例如加权组合、线性组合)计算危险因子。在一些实施方案中,DMR的甲基化状态定义维度并且可在多维空间中具有值,并且由多个DMR的甲基化状态定义的坐标例如要向使用者报导的结果。

[0023] 一些实施方案包括存储介质和存储器组件。存储器组件(例如易失性和/或非易失性存储器)可用于存储指令(例如如本文所提供的方法的实施方案)和/或数据(例如被加工的数据,诸如甲基化测量、序列以及与此相关的统计描述)。一些实施方案涉及还包括CPU、图形卡以及用户界面(例如包括输出装置,诸如显示器;以及输入装置,诸如键盘)中的一者或多者的系统。

[0024] 与所述技术相关的可编程机器包括常规现存技术以及开发中或尚有待开发的技术(例如量子计算机、化学计算机、DNA计算机、光学计算机、基于自旋电子学的计算机等)。

[0025] 在一些实施方案中,所述技术包括用于传输数据的有线(例如金属电缆、光纤)或无线传输介质。举例来说,一些实施方案涉及经网络(例如局域网(LAN)、广域网(WAN)、自组网(ad-hoc network)、互联网等)进行数据传输。在一些实施方案中,可编程机器作为节点(peer)存在于此类网络上,而在一些实施方案中,可编程机器具有客户端/服务器关系。

[0026] 在一些实施方案中,数据存储诸如硬盘、闪速存储器、光学介质、软盘等计算机可读存储介质上。

[0027] 在一些实施方案中,本文提供的技术与多个可编程装置相关,所述多个可编程装置一起操作以进行如本文所描述的方法。举例来说,在一些实施方案中,多个计算机(例如由网络连接)可并行工作以收集和处理数据,例如在集群计算或网格计算的实现中或依赖于通过常规网络接口(诸如以太网、光纤)或通过无线网络技术连接到网络(私人、公共或互联网)的完整计算机(具有板载CPU、存储器、电源、网络接口等)的一些其他分布式计算机体系结构。

[0028] 举例来说,一些实施方案提供包括计算机可读介质的计算机。所述实施方案包括与处理器连接的随机存取存储器(RAM)。处理器执行存储器中存储的计算机可执行程序指令。此类处理器可包括微处理器、ASIC、状态机或其他处理器,并且可为许多计算机处理器中的任一种,诸如来自Intel Corporation, Santa Clara, California和Motorola Corporation, Schaumburg, Illinois的处理器。此类处理器包括存储指令的介质(例如计算机可读介质)或可与其连通,所述指令在由处理器执行时使得处理器进行本文所描述的步骤。

[0029] 计算机可读介质的实施方案包括但不限于能够为处理器提供计算机可读指令的电子、光学、磁性或其他存储或传输装置。适合的介质的其他实例包括但不限于软盘、CD-ROM、DVD、磁盘、存储器芯片、ROM、RAM、ASIC、已配置的处理器、所有光学介质、所有磁带或其他磁性介质或计算机处理器可从中读取指令的任何其他介质。另外,各种其他形式的计算机可读介质均可向计算机传输或运输指令,包括路由器、私人或公共网络或者其他有线与无线的传输装置或信道。指令可包括来自包括例如C、C++、C#、Visual Basic、Java、Python、

Perl以及JavaScript的任何适合的计算机编程语言的代码。

[0030] 在一些实施方案中,计算机连接至网络。计算机还可包括许多外部或内部装置,诸如鼠标、CD-ROM、DVD、键盘、显示器或其他输入或输出装置。计算机的实例为个人计算机、数字助理、个人数字助理、蜂窝电话、移动电话、智能电话、传呼机、数字平板、膝上型计算机、互联网设备以及其他基于处理器的装置。一般来说,与本文所提供的技术的诸多方面有关的计算机可为在能够支持包括本文所提供的技术的一个或多个程序的任何操作系统(诸如Microsoft Windows、Linux、UNIX、Mac OS X等)上操作的任何类型的基于处理器的平台。一些实施方案包括执行其他应用程序(application program)(例如应用程序(application))的个人计算机。所述应用程序可含于存储器中并且可包括例如文字处理应用程序、电子表格应用程序、电子邮件应用程序、即时消息应用程序、演示文稿应用程序、互联网浏览器应用程序、日历/管理器应用程序以及能够由客户端装置执行的任何其他应用程序。

[0031] 本文所描述的与所述技术相关的所有此类组件、计算机以及系统可为逻辑的或虚拟的。

[0032] 本文提供了与在从受试者获得的样品中筛查HCC的方法有关的技术,所述方法包括分析从受试者获得的样品中的标记物的甲基化状态;以及当标记物的甲基化状态不同于在未患HCC的受试者(例如未患HCC的受试者)(例如未患HCC但患有肝硬化的受试者)中所分析的标记物的甲基化状态时,将受试者鉴定为患有HCC,其中标记物包含选自以下的如表1和表4中所提供的差异甲基化区(DMR)中的一个或多个碱基:ACP1、BDH1、Chr12.133、CLEC11A、DAB2IP、DBNL、EMX1、EFNB2、HOXA1、LRRC4、SPINT2、TSPYL5、CCNJ_3707、CCNJ_3124、PFKP、SCRN1以及ECE1。

[0033] 所述技术在所评估的甲基化状态方面不受限制。在一些实施方案中,评估样品中标记物的甲基化状态包括测定一个碱基的甲基化状态。在一些实施方案中,分析样品中标记物的甲基化状态包括确定多个碱基处的甲基化程度。此外,在一些实施方案中,标记物的甲基化状态包括标记物的甲基化相对于标记物的正常甲基化状态有所增加。在一些实施方案中,标记物的甲基化状态包括标记物的甲基化相对于标记物的正常甲基化状态有所降低。在一些实施方案中,标记物的甲基化状态包括标记物的甲基化模式相对于标记物的正常甲基化状态有所不同。

[0034] 此外,在一些实施方案中,标记物为具有100或更少个碱基的区域,标记物为具有500或更少个碱基的区域,标记物为具有1000或更少个碱基的区域,标记物为具有5000或更少个碱基的区域,或者在一些实施方案中,标记物为一个碱基。在一些实施方案中,标记物位于高CpG密度启动子中。

[0035] 所述技术不受样品类型限制。举例来说,在一些实施方案中,样品为血液样品(例如血浆、血清、全血)、粪便样品、组织样品(例如胃组织、胰腺组织、胆管/肝组织、胰液以及结肠直肠组织)、排泄物或尿液样品。

[0036] 此外,所述技术在用于测定甲基化状态的方法方面不受限制。在一些实施方案中,所述分析包括使用甲基化特异性聚合酶链反应、核酸测序、质谱法、甲基化特异性核酸酶、基于质量的分离或靶标捕获。在一些实施方案中,所述分析包括使用甲基化特异性寡核苷酸。在一些实施方案中,所述技术使用大规模平行测序(例如下一代测序)来测定甲基化状

态,例如边合成边测序、实时(例如单分子)测序、珠粒乳液测序、纳米孔测序等。

[0037] 所述技术提供用于检测DMR的试剂,例如在一些实施方案中,提供包含由SEQ ID NO:1-94提供的序列的寡核苷酸(表2和表5)的集合。在一些实施方案中,提供一种包含与具有DMR中的碱基的染色体区互补的序列的寡核苷酸,例如对DMR的甲基化状态敏感的寡核苷酸。

[0038] 所述技术提供各种标记物组,例如在一些实施方案中,标记物包含具有表1或表3中所提供的注释的染色体区,并且其包含标记物。此外,实施方案提供一种分析来自表1和/或表4的DMR编号1-400中的一者或多者的DMR的方法。

[0039] 提供试剂盒实施方案,例如如下试剂盒,所述试剂盒包含:重亚硫酸盐试剂;以及对照核酸,所述对照核酸包含来自选自DMR1-400(来自表1和表4)组成的组的DMR的序列并且具有与未患HCC的受试者(例如未患HCC并且未患肝硬化的受试者)(例如未患HCC但患有肝硬化的受试者)相关的甲基化状态。提供试剂盒实施方案,例如如下试剂盒,所述试剂盒包含:重亚硫酸盐试剂;以及对照核酸,所述对照核酸包含来自选自DMR 1-400(来自表1和表4)组成的组的DMR的序列并且具有与未患HCC的受试者相关的甲基化状态。

[0040] 一些试剂盒实施方案包括用于从受试者获得样品(例如粪便样品)的样品收集器;用于从样品分离核酸的试剂;重亚硫酸盐试剂;以及如本文所描述的寡核苷酸。

[0041] 所述技术与组合物(例如反应混合物)的实施方案有关。在一些实施方案中,提供一种包含含有DMR的核酸和重亚硫酸盐试剂的组合物。一些实施方案提供一种包含含有DMR的核酸和如本文所描述的寡核苷酸的组合物。一些实施方案提供一种包含含有DMR的核酸和甲基化敏感性限制酶的组合物。一些实施方案提供一种包含含有DMR的核酸和聚合酶的组合物。

[0042] 提供用于在从受试者获得的样品(例如血浆样品)中筛查HCC的额外相关方法实施方案,例如如下方法,所述方法包括测定样品中包含作为DMR 1-400(来自表1和表4)中的一者或多者的DMR中的碱基的标记物的甲基化状态;将来自受试者样品的标记物的甲基化状态与来自未患HCC的受试者的正常对照样品的标记物的甲基化状态进行比较;以及确定受试者样品和正常对照样品的甲基化状态的差异的置信区间和/或p值。

[0043] 在一些实施方案中,置信区间为90%、95%、97.5%、98%、99%、99.5%、99.9%或99.99%,并且p值为0.1、0.05、0.025、0.02、0.01、0.005、0.001或0.0001。一些方法实施方案提供以下步骤:使包含DMR的核酸与重亚硫酸盐试剂反应以制备与重亚硫酸盐反应过的核酸;对与重亚硫酸盐反应过的核酸测序以提供与重亚硫酸盐反应过的核酸的核苷酸序列;将与重亚硫酸盐反应过的核酸的核苷酸序列与来自未患癌症的受试者的包含DMR的核酸的核苷酸序列进行比较,以确定两个序列的差异;以及当存在差异时将受试者鉴定为患有赘瘤。

[0044] 所述技术提供用于在从受试者获得的样品中筛查HCC的系统。系统的示例性实施方案包括例如用于在从受试者获得的样品中筛查HCC的系统,所述系统包括:分析组件,所述分析组件被配置成测定样品的甲基化状态;软件组件,所述软件组件被配置成将样品的甲基化状态与对照样品或数据库中所记录的参考样品甲基化状态进行比较;以及警报组件,所述警报组件被配置成警告使用者HCC相关甲基化状态(例如没有HCC的甲基化状态;有HCC的甲基化状态)。在一些实施方案中,通过软件组件来确定警报,所述软件组件接收多个

分析(例如测定多种标记物(例如DMR,例如如表1和表4中所提供的DMR)的甲基化状态)的结果并且基于多个结果计算要报导的值或结果。一些实施方案提供一种与本文所提供的每个DMR相关的加权参数的数据库,用于计算要报告给使用者(例如医师、护士、临床医师等)的值或结果和/或警报。在一些实施方案中,报导多个分析的所有结果,而在一些实施方案中,使用一个或多个结果来提供基于多个分析的一个或多个结果的复合结果的指示受试者中的HCC危险的得分、值或结果。

[0045] 在系统的一些实施方案中,样品包含含有DMR的核酸。在一些实施方案中,系统还包括用于分离核酸的组件、用于收集样品的组件,诸如用于收集血浆样品的组件。在一些实施方案中,系统包括包含DMR的核酸序列。在一些实施方案中,数据库包含来自未患HCC的受试者的核酸序列。还提供核酸,例如核酸集合,每种核酸具有包含DMR的序列。在一些实施方案中,为核酸集合,其中每种核酸具有来自未患HCC的受试者的序列。相关系统实施方案包括如所描述的核酸集合以及与核酸集合相关的核酸序列的数据库。一些实施方案还包括重亚硫酸盐试剂。并且,一些实施方案还包括核酸测序仪。

[0046] 在某些实施方案中,提供在从受试者获得的样品(例如血浆样品)中检测HCC的方法,所述方法包括a)从受试者获得包含DNA的样品;b)用选择性修饰所获得的DNA中的未甲基化胞嘧啶残基以产生修饰的残基但不修饰甲基化胞嘧啶残基的试剂处理所获得的DNA;c)测定已经历步骤b)的处理的DNA中一种或多种DNA甲基化标记物的甲基化水平,其中一种或多种DNA甲基化标记物包含如由DMR1-400(来自表1和表4)提供的差异甲基化区(DMR)中的碱基,d)将所测定的一种或多种DNA甲基化标记物的甲基化水平与未患HCC的受试者的一种或多种DNA甲基化标记物的甲基化水平参考值进行比较;以及e)当存在差异时将受试者鉴定为患有HCC。

[0047] 在一些实施方案中,确定DNA甲基化标记物中的一者或多者中的甲基化升高包括确定选自由CpG岛和CpG岛岸组成的组的区域内的甲基化改变。

[0048] 在一些实施方案中,确定CpG岛或CpG岸内的甲基化升高包括DNA甲基化标记物的编码区或调控区内的甲基化升高。

[0049] 在一些实施方案中,测定已经历步骤b)的处理的DNA中的一种或多种DNA甲基化标记物的甲基化水平包括测定一种或多种DNA甲基化标记物的甲基化得分和/或甲基化频率。在一些实施方案中,通过对所获得的DNA的重亚硫酸盐修饰实现步骤b)的处理。

[0050] 在一些实施方案中,测定已经历步骤b)的处理的DNA中的一种或多种DNA甲基化标记物的甲基化水平是通过选自由以下组成的组的技术来实现的:甲基化特异性PCR、定量甲基化特异性PCR、甲基化敏感性DNA限制酶分析、定量重亚硫酸盐焦磷酸测序以及重亚硫酸盐基因组测序PCR。

[0051] 基于本文中所含的教导,对于擅长相关技术的人来说额外的实施方案将为显而易见的。

附图说明

[0052] 图1:用于辨别HCC DNA样品与衍生自正常对照(例如患有或未患肝硬化的非HCC个体)的DNA的311个差异甲基化区(DMR)的鉴定中所涉及的四个阶段的描述。

[0053] 图2:互补3-标记物组合(EMX1、LRRC4以及BDH1)在血浆中鉴定出21个HCC中的20个

以及33个对照中32个;1个HCC具有低水平的BDH1,并且1个对照具有升高的LRRC4。对于HCC在97%特异性(95%CI,82%-100%)下此组为95%敏感的(95%CI,74%-100%)并且实现0.98的AUC(参见图2)。

[0054] 图3A至图3I:ACP1、Chr12.133、CLEC11A、DAB2IP、DBNL、EMX1、HOXA1、LRRC4、SPINT2以及TSPYL5的接受者操作特征曲线下面积信息。

[0055] 图4A至图4CC:来自生物组织验证数据的27种胃癌标记物(在增加的拟合点分析情况下为29种)的箱线图(对数刻度)。将正常肝脏情况下的样品安排在最左侧,随后为不具有肝硬化的HCC、具有肝硬化的HCC以及肝硬化对照(炎性)。竖直轴为分数甲基化(针对 β -肌动蛋白链归一化)。

[0056] 图5:在95%正常特异性下呈基质形式的于75个HCC组织样品和29个对照(16个肝硬化、13个正常肝脏)中的27种HCC癌症标记物的性能。标记物竖列列出,而样品水平列出。将正常肝脏(NI)情况下的样品安排在最左侧,随后为不具有肝硬化的HCC(HN)、具有肝硬化的HCC(HC)以及肝硬化对照(In)。阳性命中呈浅灰色,而丢失呈深灰色。此曲线允许以互补方式评估标记物。注意:对qMSP数据使用拟合点法第二次分析2种标记物TBX15和EGR2并且包括在本文中。

[0057] 图6提供通过QuARTs(定量等位基因特异性实时靶标和信号放大)分析检测甲基化DNA印迹时所用的FRET盒的寡核苷酸序列。每个FRET序列包括荧光团和淬灭剂,它们可一起复用至3个单独的分析中。

[0058] 图7A至图7D:A)具有针对最优3标记物组合(EMX1、BDH1、LRRC4)的截止值的rPart二叉树。通过以下方法构建该树:首先发现最佳地将数据分成两组的单变量(标记物)。将数据分离,然后分开地对每个子组应用此方法,递归地以此类推,直到子组达到最小尺寸或直到不再引起改善为止。满足或不满足截止值的对照样品数目呈现于分子位置并且病例样品呈现于分母位置。在此,所述组合在血浆中鉴定出20/21个HCC和32/33个对照。B)具有针对最优3标记物组合(EMX1、DAB2IP、TSPYL5)的截止值的rPart二叉树。C)具有针对最优3标记物组合(EMX1、HOXA1、ACP1)的截止值的rPart二叉树。D)具有针对最优3标记物组合(EMX1、EFNB2、SPINT2)的截止值的rPart二叉树。

[0059] 图8:展示阶段A-C在血浆中在100%特异性下使用甲基化标记物EMX1的HCC灵敏度的柱状图。归类为U的样品为阶段不确定的。

[0060] 图9:展示此分析中考虑的甲基化标记物中的每一者(ACP1、BDH1、Chr12.133、CLEC11A、DAB2IP、DBNL、EMX1、EFNB2、HOXA1、LRRC4、SPINT2、TSPYL5、CCNJ_3707、CCNJ_3124、PFKP、SCRN1以及ECE1)的相对重要性的图。整个标记物组的灵敏度和特异性的交叉验证估计值分别为75%和96%。

具体实施方式

[0061] 本文提供了用于肝细胞癌筛查的技术并且特别地(但不排他地)涉及用于检测肝细胞癌的存在的方法、组合物以及相关用途。

[0062] 在本文描述所述技术时,所用的部分标题仅用于组织目的,并且不应以任何方式解释为限制主题。

[0063] 在各个实施方案的此具体实施方式中,出于解释的目的,阐述了许多具体细节,以

提供对所公开的实施方案的透彻理解。然而,本领域技术人员将了解可在存在或不存在这些具体细节的情况下实践所述各个实施方案。在其他情况下,结构和装置以框图形式显示。此外,本领域技术人员可容易理解呈现和进行方法的特定顺序为说明性的,并且预期顺序可改变并且仍然在本文所公开的各个实施方案的精神和范围内。

[0064] 本申请中所引用的包括但不限于专利、专利申请、文章、书、专著以及互联网网页的所有文献和类似材料出于任何目的以全文引用的方式明确地并入本文中。除非另外定义,否则本文所用的所有技术和科学术语具有与本文所描述的各个实施方案所属领域的技术人员通常所理解相同的含义。当所并入的参考文献中的术语定义似乎不同于本发明教导中所提供的定义时,应以本发明教导中所提供的定义为准。

[0065] 定义

[0066] 为促进对本发明技术的理解,许多术语和短语定义如下。具体实施方式中阐述了额外的定义。

[0067] 除非上下文另外明确指示,否则在本说明书和权利要求书通篇中,以下术语采用本文中明确相关的含义。如本文所用的短语“在一个实施方案中”不一定指同一实施方案(虽然有可能)。此外,如本文所用的短语“在另一实施方案中”不一定指不同的实施方案(虽然有可能)。因此,如下文所描述,在不背离本发明的范围或精神的情况下,可轻易组合本发明的各个实施方案。

[0068] 此外,除非上下文另外明确指示,否则如本文所用,术语“或”为包括性的“或”运算符并且等效于术语“和/或”。除非上下文另外明确指示,否则术语“基于”不为排他性的,并且允许基于未描述的额外因素。此外,在本说明书通篇中,“一种/一个(a/an)”和“所述”的含义包括复数参考物。“在……中”的含义包括“在……中”和“在……上”。

[0069] 如本文所用,“核酸”或“核酸分子”通常指任何核糖核酸或脱氧核糖核酸,其可为未修饰或修饰的DNA或RNA。“核酸”包括但不限于单链和双链核酸。如本文所用,术语“核酸”还包括含有一个或多个修饰的碱基的如上文所描述的DNA。因此,具有出于稳定性或出于其他原因进行修饰的主链的DNA为“核酸”。术语“核酸”在本文中使用涵盖核酸的此类化学修饰、酶修饰或代谢修饰形式,以及病毒和细胞(包括例如简单和复杂的细胞)的DNA特征的化学形式。

[0070] 术语“寡核苷酸”或“多核苷酸”或“核苷酸”或“核酸”是指具有两个或更多个、优选超过三个并且通常超过十个脱氧核糖核苷酸或核糖核苷酸的分子。确切大小将取决于许多因素,这些因素又取决于寡核苷酸的最终功能或用途。可以任何方式产生寡核苷酸,包括化学合成、DNA复制、逆转录或其组合。对于DNA,典型脱氧核糖核苷酸为胸腺嘧啶、腺嘌呤、胞嘧啶以及鸟嘌呤。对于RNA,典型核糖核苷酸为尿嘧啶、腺嘌呤、胞嘧啶以及鸟嘌呤。

[0071] 如本文所用,术语核酸的“基因座”或“区域”是指核酸子区域,例如染色体上的基因、单一核苷酸、CpG岛等。

[0072] 术语“互补”和“互补性”是指根据碱基配对规则相关的核苷酸(例如1个核苷酸)或多核苷酸(例如核苷酸的序列)。举例来说,序列5'-A-G-T-3'与序列3'-T-C-A-5'互补。互补性可为“部分的”,其中仅一些核酸碱基根据碱基配对规则相匹配。或者,核酸之间可存在“完全”或“总体”互补。核酸链之间的互补程度影响核酸链之间杂交的效率和强度。这在扩增反应中和在依赖核酸之间的结合的检测方法中尤其重要。

[0073] 术语“基因”是指包含制备RNA或多肽或其前体所必需的编码序列的核酸(例如DNA或RNA)序列。功能性多肽可由全长编码序列或由编码序列的任何部分编码,只要保留多肽的所需活性或功能性质(例如酶活性、配体结合、信号转导等)即可。术语“部分”当用于提及基因时是指所述基因的片段。片段的大小可在几个核苷酸至整个基因序列减去一个核苷酸范围内。因此,“包含基因的至少一部分的核苷酸”可包含基因的片段或整个基因。

[0074] 术语“基因”还涵盖结构基因的编码区,并且包括在5'和3'末端例如在任一端约1kb的距离处与编码区相邻定位的序列,使得基因对应于全长mRNA的长度(例如包括编码序列、调控序列、结构序列以及其他序列)。定位于编码区的5'并且存在于mRNA上的序列被称为5'非翻译或未翻译序列。定位于编码区的3'或下游并且存在于mRNA上的序列被称为3'非翻译或3'未翻译序列。术语“基因”涵盖基因的cDNA与基因组形式两者。在一些有机体(例如真核生物)中,基因的基因组形式或克隆含有被称为“内含子”或“插入区”或“插入序列”的非编码序列中断的编码区。内含子为转录成核RNA(hnRNA)的基因区段;内含子可含有调控元件,诸如增强子。内含子从核或初级转录物中移除或“剪出”,因此内含子不存在于信使RNA(mRNA)转录物中。mRNA在翻译过程中起作用,以指定新生多肽中氨基酸的序列或顺序。

[0075] 除含有内含子之外,基因的基因组形式还可包括定位于存在于RNA转录物上的序列的5'与3'末端的序列。这些序列被称为“侧接”序列或区域(这些侧接序列定位于存在于mRNA转录物上的非翻译序列的5'或3'。5'侧接区域可含有控制或影响基因转录的调控序列,诸如启动子和增强子。3'侧接区域可含有指导转录终止、转录后裂解以及聚腺苷酸化的序列。

[0076] 术语“等位基因”是指基因的变异型式;所述变异型式包括但不限于变体和突变体、多态性基因座和单核苷酸多态性基因座、移码以及剪接突变。等位基因可天然存在于群体中,或者它可在群体的任何特定个体的寿命期间出现。

[0077] 因此,当用于提及核苷酸序列时,术语“变体”和“突变体”是指与另一通常相关的核苷酸序列相差一个或多个核苷酸的核酸序列。“变异”为两个不同核苷酸序列之间的差异;典型地,一个序列为参考序列。

[0078] “扩增”为涉及模板特异性的核酸复制的特例。它与非特异性模板复制(例如依赖于模板但不依赖于特异性模板的复制)形成对比。模板特异性在这里区别于复制的保真度(例如适当的多核苷酸序列的合成)和核苷酸(核糖或脱氧核糖核苷酸)特异性。经常用“靶标”特异性来描述模板特异性。靶序列在尝试将其从其他核酸中挑选出来的意义上为“靶标”。扩增技术主要被设计用于此挑选。

[0079] 核酸的扩增通常是指典型地从少量多核苷酸(例如单一多核苷酸分子、多核苷酸分子的10至100个拷贝,可能完全相同或不完全相同)开始,产生多核苷酸或多核苷酸的一部分的多个拷贝,其中扩增产物或扩增子通常为可检测的。多核苷酸的扩增涵盖多种化学处理和酶处理。在聚合酶链反应(PCR)或连接酶链反应(LCR;参见例如美国专利号5,494,810)期间从靶标或模板DNA分子的一个或几个拷贝产生多个DNA拷贝为扩增的形式。额外类型的扩增包括但不限于等位基因特异性PCR(参见例如美国专利号5,639,611)、组装PCR(参见例如美国专利号5,965,408)、解旋酶依赖性扩增(参见例如美国专利号7,662,594)、热启动PCR(参见例如美国专利号5,773,258和5,338,671)、序列间特异性PCR、反向PCR(参见例如Triglia等(1988)Nucleic Acids Res.,16:8186)、连结介导的PCR(参见例如Guilfoyle,

R.等,Nucleic Acids Research,25:1854-1858(1997);美国专利号5,508,169)、甲基化特异性PCR(参见例如Herman等,(1996)PNAS 93(13)9821-9826)、微引物PCR、多重连结依赖性探针扩增(参见例如Schouten等,(2002)Nucleic Acids Research 30(12):e57)、多重PCR(参见例如Chamberlain等,(1988)Nucleic Acids Research 16(23)11141-11156;Ballabio等,(1990)Human Genetics 84(6)571-573;Hayden等,(2008)BMC Genetics 9:80)、嵌套式PCR、重叠延伸PCR(参见例如Higuchi等,(1988)Nucleic Acids Research 16(15)7351-7367)、实时PCR(参见例如Higuchi等,(1992)Biotechnology 10:413-417;Higuchi等,(1993)Biotechnology 11:1026-1030)、逆转录PCR(参见例如Bustin,S.A.(2000)J.Molecular Endocrinology 25:169-193)、固相PCR、热不对称交错PCR以及降落PCR(参见例如Don等,Nucleic Acids Research(1991)19(14)4008;Roux,K.(1994)Biotechniques 16(5)812-814;Hecker等,(1996)Biotechniques 20(3)478-485)。多核苷酸扩增还可使用数字PCR来完成(参见例如Kalinina等,Nucleic Acids Research.25;1999-2004,(1997);Vogelstein和Kinzler,Proc Natl Acad Sci USA.96:9236-41,(1999);国际专利公布号W005023091A2;美国专利申请公布号20070202525)。

[0080] 术语“聚合酶链反应”(“PCR”)是指K.B.Mullis美国专利号4,683,195、4,683,202以及4,965,188的方法,这些专利描述了一种在不进行克隆或纯化的情况下增加基因组DNA混合物中靶序列区段的浓度的方法。此用于扩增靶序列的方法由以下步骤组成:将大量过量的两种寡核苷酸引物引入含有所需靶序列的DNA混合物中,随后在存在DNA聚合酶的情况下进行一系列精确的热循环。这两种引物与双链靶序列的相应链互补。为实现扩增,使混合物变性,然后使引物与其在靶分子内的互补序列退火。退火后,用聚合酶延伸引物,以便形成一对新的互补链。可将变性、引物退火以及聚合酶延伸的步骤多次重复(即,变性、退火以及延伸构成一个“循环”;可存在许多“循环”),以获得高浓度的所需靶序列的扩增区段。所需靶序列的扩增区段的长度由引物相对于彼此的相对位置决定,并且因此此长度为可控参数。由于所述方法的重复方面,所述方法被称为“聚合酶链反应”(“PCR”)。因为靶序列的所需扩增区段成为混合物中的主导序列(就浓度来说),所以称其被“PCR扩增”,并且为“PCR产物”或“扩增子”。

[0081] 在大多数扩增技术中通过酶的选择来实现模板特异性。扩增酶为在其使用条件下仅处理核酸的非均匀混合物中的特定核酸序列的酶。举例来说,在Q- β 复制酶的情况下,MDV-1 RNA为复制酶的特异性模板(Kacian等,Proc.Natl.Acad.Sci.USA,69:3038[1972])。其他核酸不会被此扩增酶复制。类似地,在T7RNA聚合酶的情况下,此扩增酶对于其本身的启动子具有严格特异性(Chamberlin等,Nature,228:227[1970])。在T4 DNA连接酶的情况下,酶不会连结两个寡核苷酸或多核苷酸,其中在连结接点处寡核苷酸或多核苷酸底物与模板之间存在错配(Wu和Wallace(1989)Genomics 4:560)。最后,发现热稳定性模板依赖性DNA聚合酶(例如Taq和Pfu DNA聚合酶)由于其在高温下起作用的能力而对引物结合并因此限定的序列显示高特异性;高温产生有利于引物与靶序列杂交而不与非靶序列杂交的热力学条件(H.A.Erlich(编),PCR Technology,Stockton Press[1989])。

[0082] 如本文所用,术语“核酸检测分析”是指测定所关注的核酸的核苷酸组成的任何方法。核酸检测分析包括但不限于DNA测序法、探针杂交法、结构特异性裂解分析(例如INVADER分析,Hologic,Inc.),并且描述于例如美国专利号5,846,717、5,985,557、5,994,

069,6,001,567、6,090,543以及6,872,816;Lyamichev等,Nat.Biotech.,17:292(1999);Hall等,PNAS,USA,97:8272(2000)以及US 2009/0253142中);酶错配裂解法(例如Variagenics,美国专利号6,110,684、5,958,692、5,851,770);聚合酶链反应;支化杂交法(例如Chiron,美国专利号5,849,481、5,710,264、5,124,246以及5,624,802);滚环复制(例如美国专利号6,210,884、6,183,960以及6,235,502);NASBA(例如美国专利号5,409,818);分子信标技术(例如美国专利号6,150,097);电子传感器技术(Motorola,美国专利号6,248,229、6,221,583、6,013,170以及6,063,573);循环探针技术(例如美国专利号5,403,711、5,011,769以及5,660,988);Dade Behring信号放大法(例如美国专利号6,121,001、6,110,677、5,914,230、5,882,867以及5,792,614);连接酶链反应(例如Barnay Proc.Natl.Acad.Sci USA 88,189-93(1991));以及夹心杂交法(例如美国专利号5,288,609)。

[0083] 术语“可扩增核酸”是指可通过任何扩增方法扩增的核酸。预期“可扩增核酸”通常将包括“样品模板”。

[0084] 术语“样品模板”是指来源于针对“靶标”(下文定义)的存在进行分析的样品的核酸。相比之下,“背景模板”用于提及可能存在于或可能不存在于样品中的不是样品模板的核酸。背景模板大多数常常为偶然的。它可能是携带的结果,或者它可能是归因于设法从样品中纯化掉的核酸污染物的存在。举例来说,来自有机体的不是要检测的核酸的核酸可作为背景存在于测试样品中。

[0085] 术语“引物”是指在纯化的限制性消化物中天然存在的或合成产生的寡核苷酸,当放置在其中诱导与核酸链互补的引物延伸产物合成的条件下(例如在存在核苷酸和诸如DNA聚合酶等诱导剂的情况下以及在适合的温度和PH下)时,所述寡核苷酸能够充当合成的起点。引物优选为单链的,以获得最大扩增效率,但或者也可作为双链的。如果是双链,则在用于制备延伸产物之前首先处理引物以将其链分离。优选地,引物为寡脱氧核苷酸。引物必须足够长以在存在诱导剂的情况下引发扩增产物的合成。引物的精确长度将取决于许多因素,包括温度、引物来源以及方法的使用。

[0086] 术语“探针”是指在纯化的限制性消化物中天然存在的或者合成、重组或通过PCR扩增产生的寡核苷酸(例如核苷酸序列),所述寡核苷酸能够与另一所关注的寡核苷酸杂交。探针可为单链或双链的。探针适用于检测、鉴定以及分离特定基因序列(例如“捕获探针”)。预期在一些实施方案中,本发明中所用的任何探针均可用任何“报告分子”进行标记,使得在包括但不限于以下的任何检测系统中可检测:酶(例如ELISA,以及基于酶的组织化学分析)、荧光、放射性以及发光系统。本发明不意图限制于任何特定检测系统或标记。

[0087] 如本文所用,“甲基化”是指在胞嘧啶的位置C5或N4处的胞嘧啶甲基化、腺嘌呤的N6位置或其他类型的核酸甲基化。体外扩增的DNA通常是未甲基化的,因为典型体外DNA扩增法不能保留扩增模板的甲基化模式。然而,“未甲基化DNA”或“甲基化DNA”还可分别指原始模板未发生甲基化或发生甲基化的扩增的DNA。

[0088] 因此,如本文所用,“甲基化核苷酸”或“甲基化核苷酸碱基”是指在核苷酸碱基上存在甲基部分,其中甲基部分不存在于认可的典型核苷酸碱基中。举例来说,胞嘧啶在其嘧啶环上不含甲基部分,但5-甲基胞嘧啶在其嘧啶环的位置5处含有甲基部分。因此,胞嘧啶不为甲基化核苷酸,而5-甲基胞嘧啶为甲基化核苷酸。在另一实例中,胸腺嘧啶在其嘧啶环

的位置5处含有甲基部分；然而，出于本文的目的，当存在于DNA中时不将胸腺嘧啶视为甲基化核苷酸，因为胸腺嘧啶为DNA的典型核苷酸碱基。

[0089] 如本文所用，“甲基化核酸分子”是指含有一个或多个甲基化核苷酸的核酸分子。

[0090] 如本文所用，核酸分子的“甲基化状态”、“甲基化型态”以及“甲基化状况”是指核酸分子中存在或不存在一个或多个甲基化核苷酸碱基。举例来说，将含有甲基化胞嘧啶的核酸分子视为甲基化的（例如核酸分子的甲基化状态为甲基化的）。将不含任何甲基化核苷酸的核酸分子视为未甲基化的。

[0091] 特定核酸序列（例如如本文所描述的基因标记物或DNA区域）的甲基化状态可指示序列中每个碱基的甲基化状态或可指示序列内碱基子集（例如一个或多个胞嘧啶）的甲基化状态，或者可指示关于序列内区域甲基化密度的信息，同时提供或不提供序列内甲基化发生的位置的精确信息。

[0092] 核酸分子中的核苷酸基因座的甲基化状态是指核酸分子中的特定基因座处存在或不存在甲基化核苷酸。举例来说，当核酸分子中第7个核苷酸处存在的核苷酸为5-甲基胞嘧啶时，核酸分子中第7个核苷酸处的胞嘧啶的甲基化状态为甲基化的。类似地，当核酸分子中第7个核苷酸处存在的核苷酸为胞嘧啶（而不是5-甲基胞嘧啶）时，核酸分子中第7个核苷酸处的胞嘧啶的甲基化状态为未甲基化的。

[0093] 甲基化状况可任选地由“甲基化值”表示或指示（例如表示甲基化频率、分数、比率、百分比等）。可例如通过在用甲基化依赖性限制酶限制性消化之后定量存在的完整核酸的量，或者通过比较重亚硫酸盐反应后的扩增型态，或者通过比较重亚硫酸盐处理过和未处理的核酸的序列来产生甲基化值。因此，值（例如甲基化值）代表甲基化状况，并且因此可用作基因座的多个拷贝中的甲基化状况的定量指示物。当需要将样品中的序列的甲基化状况与阈值或参考值进行比较时，这是特别有用的。

[0094] 如本文所用，“甲基化频率”或“甲基化百分比(%)”是指相对于分子或基因座未甲基化的情况的数目，分子或基因座甲基化的情况的数目。

[0095] 因此，甲基化状态描述核酸（例如基因组序列）的甲基化状态。此外，甲基化状态是指特定基因组基因座处的核酸区段的与甲基化相关的特征。此类特征包括但不限于DNA序列中的任何胞嘧啶(C)残基是否为甲基化的、甲基化C残基的位置、核酸的任何特定区域中甲基化C的频率或百分比以及甲基化中归因于例如等位基因来源差异的等位基因差异。术语“甲基化状态”、“甲基化型态”以及“甲基化状况”还指生物样品中核酸的任何特定区域中甲基化C或未甲基化C的相对浓度、绝对浓度或模式。举例来说，如果核酸序列内的胞嘧啶(C)残基为甲基化的，则可将其称为“高甲基化的”或具有“增加的甲基化”，而如果DNA序列内的胞嘧啶(C)残基序列不为甲基化的，则可将其称为“低甲基化的”或具有“降低的甲基化”。同样地，如果核酸序列内的胞嘧啶(C)残基与另一核酸序列（例如来自不同区域或不同个体等）相比为甲基化的，则认为所述序列与另一核酸序列相比为高甲基化的或具有增加的甲基化。或者，如果DNA序列内的胞嘧啶(C)残基与另一核酸序列（例如来自不同区域或不同个体等）相比不为甲基化的，则认为所述序列与另一核酸序列相比为低甲基化的或具有降低的甲基化。另外，如本文所用的术语“甲基化模式”是指核酸区域上的甲基化和未甲基化核苷酸的集合位点。当整个区域中甲基化和未甲基化核苷酸的数量相同或相似，但甲基化和未甲基化核苷酸的位置不同时，两个核酸可具有相同或相似的甲基化频率或甲基化百

分数,但具有不同的甲基化模式。当序列的甲基化的程度(例如一个相对于另一个具有增加的或降低的甲基化)、频率或模式不同时,称序列为“差异甲基化的”或具有“甲基化差异”或具有“不同的甲基化状态”。术语“差异甲基化”是指癌症阳性样品中的核酸甲基化水平或模式与癌症阴性样品中的核酸甲基化水平或模式相比的差异。它还可指手术后癌症复发的患者与未复发的患者之间的水平或模式的差异。差异甲基化和DNA甲基化的特定水平或模式例如在确定了正确的截止或预示特征后为预后和预示性生物标记物。

[0096] 可使用甲基化状态频率来描述个体的群体或来自单一个体的样品。举例来说,具有50%的甲基化状态频率的核苷酸基因座在50%的情况下为甲基化的,并且在50%的情况下为未甲基化的。可使用此类频率例如描述个体群体或核酸集合中核苷酸基因座或核酸区域的甲基化所达到的程度。因此,当第一群体或核酸分子汇集物中的甲基化与第二群体或核酸分子汇集物中的甲基化不同时,第一群体或汇集物的甲基化状态频率不同于第二群体或汇集物的甲基化状态频率。还可使用此类频率例如描述单一个体中核苷酸基因座或核酸区域的甲基化所达到的程度。举例来说,可使用此类频率来描述来自组织样品的细胞组在核苷酸基因座或核酸区域的甲基化或未甲基化所达到的程度。

[0097] 如本文所用,“核苷酸基因座”是指核酸分子中核苷酸的位置。甲基化核苷酸的核苷酸基因座是指核酸分子中甲基化核苷酸的位置。

[0098] 典型地,人DNA的甲基化发生在包括相邻鸟嘌呤和胞嘧啶的二核苷酸序列上,其中胞嘧啶定位于鸟嘌呤的5' (也称为CpG二核苷酸序列)。在人基因组中CpG二核苷酸中的大多数胞嘧啶为甲基化的,然而在特定的CpG二核苷酸富集的基因组区域(被称为CpG岛)中一些保持未甲基化(参见例如Antequera等(1990) Cell 62:503-514)。

[0099] 如本文所用,“CpG岛”是指基因组DNA中相对于总基因组DNA含有增加数目的CpG二核苷酸的富含G:C的区域。CpG岛的长度可为至少100、200或更多个碱基对,其中所述区域的G:C含量为至少50%,并且所观测到的CpG频率与预期频率的比率为0.6;在一些情况下,CpG岛的长度可为至少500个碱基对,其中所述区域的G:C含量为至少55%),并且观测到的CpG频率与预期频率的比率为0.65。可根据Gardiner-Garden等(1987) J. Mol. Biol. 196:261-281中所提供的方法来计算相对于预期频率所观测到的CpG频率。举例来说,可根据公式 $R = (A \times B) / (C \times D)$ 来计算相对于预期频率所观测到的CpG频率,其中R为所观测到的CpG频率与预期频率的比率,A为所分析序列中的CpG二核苷酸的数目,B为所分析序列中的核苷酸的总数目,C为所分析序列中的C核苷酸的总数目,并且D为所分析序列中的G核苷酸的总数目。典型地在CpG岛中,例如在启动子区确定甲基化状态。不过应理解,人基因组中的其他序列也倾向于DNA甲基化,诸如CpA和CpT(参见例如Ramsahoye (2000) Proc. Natl. Acad. Sci. USA 97:5237-5242; Salmon和Kaye (1970) Biochim. Biophys. Acta. 204:340-351; Grafstrom (1985) Nucleic Acids Res. 13:2827-2842; Nyce (1986) Nucleic Acids Res. 14:4353-4367; Woodcock (1987) Biochem. Biophys. Res. Commun. 145:888-894)。

[0100] 如本文所用,根据核酸分子的甲基化状态修饰核酸分子的核苷酸的试剂或者甲基化特异性试剂是指可以反映核酸分子的甲基化状态的方式改变核酸分子的核苷酸序列的化合物或组合物或其他剂。用此类试剂处理核酸分子的方法可包括结合额外步骤(如果需要)使核酸分子与试剂接触以实现所需核苷酸序列变化。核酸分子的核苷酸序列的此类变化可产生其中每个甲基化核苷酸被修饰成不同核苷酸的核酸分子。核酸核苷酸序列的此类

变化可产生其中每个未甲基化核苷酸被修饰成不同核苷酸的核酸分子。核酸核苷酸序列的此类变化可产生其中每个未甲基化的所选核苷酸(例如每个未甲基化的胞嘧啶)被修饰成不同的核苷酸的核酸分子。使用此类试剂改变核酸核苷酸序列可产生其中为甲基化核苷酸的每个核苷酸(例如每个甲基化胞嘧啶)被修饰成不同核苷酸的核酸分子。如本文所用,使用修饰所选核苷酸的试剂是指修饰核酸分子中四种典型存在的核苷酸(DNA的C、G、T以及A,以及RNA的C、G、U以及A)中的一种核苷酸的试剂,使得试剂修饰这一种核苷酸而不修饰另外三种核苷酸。在一个示例性实施方案中,此类试剂修饰未甲基化的所选核苷酸以产生不同的核苷酸。在另一示例性实施方案中,此类试剂可使未甲基化胞嘧啶核苷酸脱去氨基。示例性试剂为重亚硫酸盐。

[0101] 如本文所用,术语“重亚硫酸盐试剂”是指在一些实施方案中包括重亚硫酸盐(bisulfite)、二亚硫酸盐(disulfite)、亚硫酸氢盐(hydrogen sulfite)或其组合的试剂,以区分例如CpG二核苷酸序列中的甲基化和未甲基化胞苷。

[0102] 术语“甲基化分析”是指用于测定核酸序列内的一个或多个CpG二核苷酸序列的甲基化状态的任何分析。

[0103] 术语“MS AP-PCR”(甲基化敏感性任意引物聚合酶链反应)是指本领域公认的技术,所述技术允许使用富含CG的引物对基因组进行全局扫描,以集中于最可能含有CpG二核苷酸的区域,并且由Gonzalogo等(1997)Cancer Research 57:594-599描述。

[0104] 术语“MethyLight™”是指由Eads等(1999)Cancer Res.59:2302-2306描述的本领域公认的基于荧光的实时PCR技术。

[0105] 术语“HeavyMethyl™”是指如下分析,其中覆盖位于扩增引物之间或由扩增引物覆盖的CpG位置的甲基化特异性阻断探针(本文中也称为阻断剂)能够对核酸样品进行甲基化特异性选择性扩增。

[0106] 术语“HeavyMethyl™MethyLight™”分析是指HeavyMethyl™MethyLight™分析,所述分析为MethyLight™分析的变化型式,其中将MethyLight™分析与覆盖扩增引物之间的CpG位置的甲基化特异性阻断探针组合。

[0107] 术语“Ms-SNuPE”(甲基化敏感性单核苷酸引物延伸)是指由Gonzalogo和Jones(1997)Nucleic Acids Res.25:2529-2531描述的本领域公认的分析。

[0108] 术语“MSP”(甲基化特异性PCR)是指由Herman等(1996)Proc.Natl.Acad.Sci.USA 93:9821-9826和由美国专利号5,786,146描述的本领域公认的甲基化分析。

[0109] 术语“COBRA”(组合重亚硫酸盐限制分析)是指由Xiong和Laird(1997)Nucleic Acids Res.25:2532-2534描述的本领域公认的甲基化分析。

[0110] 术语“MCA”(甲基化CpG岛扩增)是指由Toyota等(1999)Cancer Res.59:2307-12描述以及WO 00/26401A1中的甲基化分析。

[0111] 如本文所用,“所选核苷酸”是指核酸分子中四种典型存在的核苷酸(DNA的C、G、T以及A,以及RNA的C、G、U以及A)中的一种核苷酸,并且可包括典型存在的核苷酸的甲基化衍生物(例如当C为所选核苷酸时,甲基化与未甲基化的C均包括在所选核苷酸的含义内),而甲基化的所选核苷酸尤其指甲基化的典型存在的核苷酸,未甲基化的所选核苷酸尤其指未甲基化的典型存在的核苷酸。

[0112] 术语“甲基化特异性限制酶”或“甲基化敏感性限制酶”是指根据识别位点的甲基

化状态选择性消化核酸的酶。在如果识别位点不为甲基化或为半甲基化的则进行特异性切割的限制酶的情况下,如果识别位点为甲基化的,则切割将不发生或者将以显著降低的效率发生。在如果识别位点为甲基化的则进行特异性切割的限制酶的情况下,如果识别位点不为甲基化的,则切割将不发生或者将以显著降低的效率发生。优选为如下甲基化特异性限制酶,其识别序列含有CG二核苷酸(例如诸如CGCG或CCCGGG等识别序列)。对于一些实施方案,进一步优选的为如果此二核苷酸中的胞嘧啶在碳原子C5处为甲基化的则不进行切割的限制酶。

[0113] 如本文所用,“不同核苷酸”是指如下核苷酸,所述核苷酸在化学上不同于所选核苷酸,典型地使得不同核苷酸具有与所选核苷酸不同的沃森-克里克碱基配对 (Watson-Crick base-pairing) 性质,借此与所选核苷酸互补的典型存在的核苷酸和与不同核苷酸互补的典型存在的核苷酸不相同。举例来说,当C为所选核苷酸时,U或T可为不同核苷酸,此通过C与G的互补性以及U或T与A性的互补性来例示。如本文所用,与所选核苷酸互补或与不同核苷酸互补的核苷酸是指在高严格条件下与所选核苷酸或不同核苷酸碱基配对的核苷酸,其亲和力高于互补核苷酸与四种典型存在的核苷酸中的三种核苷酸的碱基配对。互补性的实例为在DNA(例如A-T和C-G)和RNA(例如A-U和C-G)中的沃森-克里克碱基配对。因此,例如在高严格条件下,和G与G、A或T碱基配对相比,G以更高亲和力与C碱基配对,因此,当C为所选核苷酸时,G为与所选核苷酸互补的核苷酸。

[0114] 如本文所用,给定标记物的“灵敏度”是指报告高于区分赘生和非赘生样品的阈值的DNA甲基化值的样品的百分比。在一些实施方案中,阳性定义为报告高于阈值(例如与疾病相关的范围)的DNA甲基化值的组织学确认的赘瘤,并且假阴性定义为报告低于阈值(例如与无疾病相关的范围)的DNA甲基化值的组织学确认的赘瘤。因此,灵敏度的值反映了从已知患病样品获得的给定标记物的DNA甲基化测量将在疾病相关测量范围内的可能性。如在此所定义,计算的灵敏度值的临床相关性表示当应用于患有临床疾患的受试者时给定标记物将检测到所述疾患的存在的可能性的估计。

[0115] 如本文所用,给定标记物的“特异性”是指报告低于区分赘生和非赘生样品的阈值的DNA甲基化值的非赘生样品的百分比。在一些实施方案中,阴性定义为报告低于阈值(例如与无疾病相关的范围)的DNA甲基化值的组织学确认的非赘生样品,并且假阳性定义为报告高于阈值(例如与疾病相关的范围)的DNA甲基化值的组织学确认的非赘生样品。因此,特异性的值反映了从已知非赘生样品获得的给定标记物的DNA甲基化测量将在非疾病相关测量范围内的可能性。如在此所定义,所计算的特异性值的临床相关性表示当应用于未患临床疾患的患者时给定标记物将检测到所述疾患不存在的可能性的估计。

[0116] 如本文所用的术语“AUC”为“曲线下的面积”的缩写。特别地,它指接受者操作特征(ROC)曲线下的面积。ROC曲线为诊断测试的不同可能切割点的真阳性率对假阳性率的曲线。它示出了取决于所选择的切割点的灵敏度与特异性之间的折中(灵敏度的任何增加都将伴随着特异性的降低)。ROC曲线下面积(AUC)为诊断测试准确性的度量(面积越大越好;最优值为1;随机测试将具有位于对角线上的面积为0.5的ROC曲线;参考:J.P.Egan.(1975) Signal Detection Theory and ROC Analysis,Academic Press,New York)。

[0117] 如本文所用,术语“赘瘤”是指“生长超过了正常组织并且与正常组织不协调的异常组织团块”。参见例如Willis RA,“The Spread of Tumors in the Human Body”,

London, Butterworth & Co, 1952。

[0118] 如本文所用,术语“腺瘤”是指源于腺体的良性肿瘤。虽然这些生长为良性的,但随时间推移它们可发生进展从而变为恶性的。

[0119] 术语“癌前”或“赘生前”以及其等效物是指正在经历恶性转化的任何细胞增殖性病症。

[0120] 赘瘤、腺瘤、癌症等的“部位”或“区域”为赘瘤、腺瘤、癌症等定位的受试者体内的组织、器官、细胞类型、解剖区域、身体部位等。

[0121] 如本文所用,“诊断”测试应用包括检测或鉴定受试者的疾病状态或状况,确定受试者将感染给定疾病或疾患的可能性,确定患有疾病或疾患的受试者将对疗法作出反应的可能性,确定具有疾病或疾患的受试者的预后(或其可能的进展或消退),以及确定治疗对患有疾病或疾患的受试者的影响。举例来说,诊断法可用于检测受试者感染赘瘤的存在或可能性或此类受试者将有利地对化合物(例如医药,例如药物)或其他治疗作出反应的可能性。

[0122] 如本文所用,术语“标记物”是指能够通过例如基于甲基化状态区分疾病相关细胞(例如与疾病相关的非癌细胞)(与疾病相关的癌细胞)与正常细胞来诊断病症(例如非恶性疾病)(例如癌性疾病)的物质(例如核酸或核酸区域)。

[0123] 当用于指核酸时,如在“分离的寡核苷酸”中的术语“分离的”是指从在天然来源中通常与相关联的至少一种污染物核酸鉴定和分离的核酸序列。分离的核酸以与自然界中所发现的形式或设置不同的形式或设置存在。相比之下,非分离核酸(诸如DNA和RNA)以其在自然界中存在的状态被发现。非分离核酸的实例包括:在相邻基因附近的存在于宿主细胞染色体上的给定DNA序列(例如基因);以及与编码大量蛋白质的许多其他mRNA的混合物的形式存在于细胞中的RNA序列,诸如编码特定蛋白质的特定mRNA序列。然而,编码特定蛋白质的分离的核酸包括例如通常表达所述蛋白质的细胞中的此类核酸,其中所述核酸处于与天然细胞不同的染色体位置,或者侧接与在自然界中发现的不同的核酸序列。分离的核酸或寡核苷酸可以单链或双链形式存在。当利用分离的核酸或寡核苷酸来表达蛋白质时,寡核苷酸将至少含有正义链或编码链(即寡核苷酸可为单链的),但也可含有正义链与反义链两者(即,寡核苷酸可为双链的)。分离的核酸在从其天然或典型环境分离之后可与其他核酸或分子组合。举例来说,分离的核酸可存在于它被放置于其中的宿主细胞中,例如用于异源表达。

[0124] 术语“纯化的”是指从天然环境中移出、分离或分开的核酸或氨基酸序列的分子。因此,“分离的核酸序列”可为纯化的核酸序列。“基本上纯化的”分子至少60%不含、优选地至少75%不含或更优选地至少90%不含与其天然缔合的其他组分。如本文所用,术语“纯化的”或“以纯化”还指从样品中除去污染物。除去污染蛋白质使得样品中所关注的多肽或核酸的百分比增加。在另一实例中,重组多肽在植物、细菌、酵母或哺乳动物宿主细胞中表达,并且通过除去宿主细胞蛋白来纯化所述多肽;样品中重组多肽的百分比由此增加。

[0125] 术语“包含给定多核苷酸序列或多肽的组合物”广义上是指含有给定多核苷酸序列或多肽的任何组合物。组合物可包含含有盐(例如NaCl)、洗涤剂(例如SDS)以及其他组分(例如邓哈特溶液(Denhardt's solution)、干乳、鲑鱼精子DNA等)的水溶液。

[0126] 术语“样品”以其最广泛的意义使用。在一个意义上,它可指动物细胞或组织。在另

一意义上,它意在包括从任何来源获得的试样或培养物,以及生物和环境样品。生物样品可从植物或动物(包括人)获得并且涵盖流体、固体、组织以及气体。环境样品包括环境材料,诸如表面物质、土壤、水以及工业样品。这些实例不应被解释为限制适用于本发明的样品类型。

[0127] 如本文所用,如一些背景下所用的“远程样品”涉及从不是样品的细胞、组织或器官来源的部位间接收集的样品。举例来说,当在粪便样品(例如不是从胰腺直接取得的样品)中评估源自胰腺的样品材料时,样品为远程样品。

[0128] 如本文所用,术语“患者”或“受试者”是指要进行由所述技术提供的各种测试的有机体。术语“受试者”包括动物,优选为哺乳动物,包括人。在一个优选实施方案中,受试者为灵长类动物。在一个甚至更优选的实施方案中,受试者为人。

[0129] 如本文所用,术语“试剂盒”是指用于递送材料的任何递送系统。在反应分析的背景下,此类递送系统包括允许将反应试剂(例如适当容器中的寡核苷酸、酶等)和/或支持材料(例如缓冲液、用于进行分析的书面说明等)从一个位置储存、运输或递送到另一个位置的系统。举例来说,试剂盒包括一个或多个含有相关反应试剂和/或支持材料的外壳(例如盒子)。如本文所用,术语“分区试剂盒(fragmented kit)”是指包括两个或更多个分开的容器的递送系统,所述容器各自含有全部试剂盒组分的子部分。容器可一起或分开地递送给预期接受者。举例来说,第一容器可含有用于分析的酶,而第二容器含有寡核苷酸。术语“分区试剂盒”意图涵盖含有根据联邦食品、药品及化妆品法案(Federal Food, Drug, and Cosmetic Act)的章节520(e)规定的分析物特异性试剂(ASR)的试剂盒,但不限于此。实际上,包括两个或更多个分开的容器并且每个容器含有全部试剂盒组分的子部分的任何递送系统均包括在术语“分区试剂盒”内。相比之下,“组合试剂盒”是指在单一容器中(例如在容纳每种所需组分的单一盒外壳中)含有反应分析的所有组分的递送系统。术语“试剂盒”包括分区与组合试剂盒两者。

[0130] 所述技术的实施方案

[0131] 本文提供了用于HCC筛查的技术(例如监视)并且特别地(但不排他地)涉及用于检测受试者中HCC的存在的方法、组合物以及相关用途。

[0132] 鉴定了能够检测HCC的标记物和/或标记物组(例如具有表1和表4中所提供的注释的染色体区)(参见实施例I、II以及III)(例如ACP1、BDH1、Chr12.133、CLEC11A、DAB2IP、DBNL、EMX1、EFNB2、HOXA1、LRRC4、SPINT2、TSPYL5、CCNJ_3707、CCNJ_3124、PFKP、SCRN1以及ECE1)。

[0133] 虽然本文的公开内容提到某些所说明的实施方案,但应了解,这些实施方案是以举例的方式而不是以限制的方式呈现的。

[0134] 所述方法包括测定从受试者分离的生物样品中的至少一种甲基化标记物的甲基化状况,其中标记物的甲基化状态的变化指示HCC的存在或类别。特定实施方案涉及用于诊断(例如筛查)HCC的包含差异甲基化区(DMR,例如DMR 1-400(来自表1和表4))的标记物。

[0135] 除分析了包含本文提供并且列于表1和表3中的DMR(例如DMR 1-400)的至少一种标记物、标记物区域或标记物碱基的甲基化分析的实施方案之外,所述技术还提供包含可用于检测受试者中的HCC的包含DMR的至少一种标记物、标记物区域或标记物碱基的标记物组。

[0136] 所述技术的一些实施方案是基于对包含DMR的至少一种标记物、标记物区域或标记物碱基的CpG甲基化状况的分析。

[0137] 在一些实施方案中,本发明技术提供重亚硫酸盐技术与一种或多种甲基化分析组合用于测定包含DMR(例如如表1和表4中所提供的DMR(例如DMR 1-400))的至少一种标记物内的CpG二核苷酸序列的甲基化状况的用途。基因组CpG二核苷酸可为甲基化或未甲基化的(或者分别称为上甲基化和下甲基化)。然而,本发明的方法适合于分析非均质性性质的生物样品,例如远程样品(例如血液、器官流出物或粪便)的背景内的低浓度肿瘤细胞,或者由其产生的生物材料。因此,当分析此类样品内的CpG位置的甲基化状况时,可使用定量分析来测定特定CpG位置的甲基化水平(例如百分比、分数、比率、比例或程度)。

[0138] 根据本发明技术,测定包含DMR的标记物中的CpG二核苷酸序列的甲基化状况可用于诊断与表征受试者中的HCC。

[0139] 标记物的组合

[0140] 在一些实施方案中,所述技术涉及评估包含两个或更多个来自表1和/或表4的DMR(例如两个或更多个来自DMR编号1-400的DMR)的标记物的组合的甲基化状态。在一些实施方案中,评估超过一种标记物的甲基化状态使用于鉴定受试者中HCC的存在的筛查或诊断的特异性和/或灵敏度增加。

[0141] 例如如通过与预测的特异性和灵敏度有关的统计技术所鉴定,通过标记物的各种组合来预测各种癌症。所述技术提供鉴定一些癌症的预测组合和验证的预测组合的方法。

[0142] 举例来说,鉴定了能够检测HCC的标记物和/或标记物组(例如具有表1和表4中所提供的注释的染色体区)(参见实施例I、II以及III)(例如ACP1、BDH1、Chr12.133、CLEC11A、DAB2IP、DBNL、EMX1、EFNB2、H0XA1、LRRC4、SPINT2、TSPYL5、CCNJ_3707、CCNJ_3124、PFKP、SCRN1以及ECE1)。

[0143] 分析甲基化状态的方法

[0144] 最常使用的分析核酸中5-甲基胞嘧啶的存在的方法是基于由Frommer等描述的用于检测DNA中的5-甲基胞嘧啶的重亚硫酸盐法(Frommer等(1992) Proc. Natl. Acad. Sci. USA 89:1827-31)或其变化型式。定位5-甲基胞嘧啶的重亚硫酸盐法是基于以下观测结果,胞嘧啶与亚硫酸氢根离子(亦称为重亚硫酸根)反应,而5-甲基胞嘧啶不与亚硫酸氢根离子反应。所述反应通常根据以下步骤进行:首先,胞嘧啶与亚硫酸氢盐反应以形成磺化胞嘧啶。接着,磺化的反应中间物自发脱氨基产生磺化尿嘧啶。最后,磺化尿嘧啶在碱性条件下脱磺酸以形成尿嘧啶。检测是可能的,因为尿嘧啶与腺嘌呤形成碱基对(因此表现得像胸腺嘧啶),而5-甲基胞嘧啶与鸟嘌呤碱基配对(因此表现得像胞嘧啶)。这使得可通过例如重亚硫酸盐基因组测序(Grigg G和Clark S, Bioessays (1994) 16:431-36; Grigg G, DNA Seq. (1996) 6:189-98)或如例如美国专利号5,786,146中所公开的甲基化特异性PCR(MSP)来辨别甲基化胞嘧啶与非甲基化胞嘧啶。

[0145] 一些常规技术涉及包括将要分析的DNA封闭于琼脂糖基质中从而防止DNA扩散和复性(重亚硫酸盐仅与单链DNA反应)以及用快速渗析替代沉淀和纯化步骤的方法(Olek A, 等(1996) "A modified and improved method for bisulfite based cytosine methylation analysis" Nucleic Acids Res. 24:5064-6)。因此可分析个别细胞的甲基化状况,从而说明方法的效用和灵敏度。Rein, T.等(1998) Nucleic Acids Res. 26:2255提供

了检测5-甲基胞嘧啶的常规方法的综述。

[0146] 重亚硫酸盐技术典型地涉及在重亚硫酸盐处理后扩增已知核酸的短特异性片段,然后通过测序 (Olek和Walter (1997) Nat. Genet. 17:275-6) 或引物延伸反应 (Gonzalga和Jones (1997) Nucleic Acids Res. 25:2529-31; WO 95/00669; 美国专利号6,251,594) 分析产物,以分析个别胞嘧啶位置。一些方法使用酶消化 (Xiong和Laird (1997) Nucleic Acids Res. 25:2532-4)。本领域中还描述了通过杂交检测 (Olek等, WO 99/28498)。另外,已描述重亚硫酸盐技术用于个别基因的甲基化检测的用途 (Grigg和Clark (1994) Bioessays 16:431-6; Zeschnigk等 (1997) Hum Mol Genet. 6:387-95; Feil等 (1994) Nucleic Acids Res. 22:695; Martin等 (1995) Gene 157:261-4; WO 9746705; WO 9515373)。

[0147] 各种甲基化分析程序为本领域已知的并且可与根据本发明技术的重亚硫酸盐处理结合使用。这些分析允许测定核酸序列内的一个或多个CpG二核苷酸 (例如CpG岛) 的甲基化状态。此类分析涉及重亚硫酸盐处理过的核酸的测序、PCR (用于序列特异性扩增)、南方墨点分析 (Southern blot analysis) 以及甲基化敏感性限制酶的使用以及其他技术。

[0148] 举例来说,已通过使用重亚硫酸盐处理简化了基因组测序,以分析甲基化模式和5-甲基胞嘧啶分布 (Frommer等 (1992) Proc. Natl. Acad. Sci. USA 89:1827-1831)。另外,例如如由Sadri和Hornsby (1997) Nucl. Acids Res. 24:5058-5059所描述,或者如在称为COBRA (组合重亚硫酸盐限制性分析) 的方法所体现 (Xiong和Laird (1997) Nucleic Acids Res. 25:2532-2534), 从重亚硫酸盐转化的DNA扩增的PCR产物的限制酶消化可用于评估甲基化状态。

[0149] COBRATM分析为适用于确定少量基因组DNA中特定基因座处的DNA甲基化水平的定量甲基化分析 (Xiong和Laird, Nucleic Acids Res. 25:2532-2534, 1997)。简言之,使用限制酶消化来揭露重亚硫酸钠处理的DNA的PCR产物中的甲基化依赖性序列差异。首先,根据Frommer等 (Proc. Natl. Acad. Sci. USA 89:1827-1831, 1992) 描述的程序通过标准重亚硫酸盐处理将甲基化依赖性序列差异引入基因组DNA。然后使用对所关注的CpG岛具特异性的引物进行重亚硫酸盐转化的DNA的PCR扩增,随后进行限制性核酸内切酶消化、凝胶电泳以及使用特异性标记的杂交探针的检测。原始DNA样品中的甲基化水平在广泛范围的DNA甲基化水平内以线性定量方式由消化的和未消化的PCR产物的相对量表示。此外,此技术可以可靠地适用于从显微解剖的石蜡包埋的组织样品获得的DNA。

[0150] 用于COBRATM分析的典型试剂 (例如如可在典型的基于COBRATM的试剂盒中发现) 可包括但不限于: 用于特定基因座 (例如特定基因、标记物、DMR、基因区域、标记物区域、重亚硫酸盐处理过的DNA序列、CpG岛等) 的PCR引物; 限制酶和适当的缓冲液; 基因杂交寡核苷酸; 对照杂交寡核苷酸; 寡核苷酸探针的激酶标记试剂盒; 以及标记的核苷酸。另外,重亚硫酸盐转化试剂可包括: DNA变性缓冲液; 磺化缓冲液; DNA回收试剂或试剂盒 (例如沉淀、超滤、亲和柱); 脱磺酸缓冲液; 以及DNA回收组分。

[0151] 优选地,单独使用诸如“MethyLightTM” (基于荧光的实时PCR技术) (Eads等, Cancer Res. 59:2302-2306, 1999)、Ms-SNuPETM (甲基化敏感性单核苷酸引物延伸) 反应 (Gonzalga和Jones, Nucleic Acids Res. 25:2529-2531, 1997)、甲基化特异性PCR (“MSP”; Herman等, Proc. Natl. Acad. Sci. USA 93:9821-9826, 1996; 美国专利号5,786,146) 以及甲基化CpG岛扩增 (“MCA”; Toyota等, Cancer Res. 59:2307-12, 1999) 等分析或与这些方法中的一者或多

者组合使用。

[0152] “HeavyMethyl™”分析技术为基于重亚硫酸盐处理过的DNA的甲基化特异性扩增来评估甲基化差异的定量方法。覆盖位于扩增引物之间或由扩增引物覆盖的CpG位置的甲基化特异性阻断探针(“阻断剂”)使得能够对核酸样品进行甲基化特异性选择性扩增。

[0153] 术语“HeavyMethyl™MethyLight™”分析是指HeavyMethyl™MethyLight™分析,所述分析为MethyLight™分析的变化型式,其中将MethyLight™分析与覆盖扩增引物之间的CpG位置的甲基化特异性阻断探针组合。还可将HeavyMethyl™分析与甲基化特异性扩增引物组合使用。

[0154] 用于HeavyMethyl™分析的典型试剂(例如如可在典型的基于MethyLight™的试剂盒中发现)可包括但不限于:用于特定基因座(例如特定基因、标记物、DMR、基因区域、标记物区域、重亚硫酸盐处理过的DNA序列、CpG岛或重亚硫酸盐处理过的DNA序列或CpG岛等)的PCR引物;阻断性寡核苷酸;优化的PCR缓冲液和脱氧核苷酸;以及Taq聚合酶。

[0155] MSP(甲基化特异性PCR)允许评估CpG岛内几乎任何CpG位点组的甲基化状况,此与甲基化敏感性限制酶的使用无关(Herman等Proc.Natl.Acad.Sci.USA 93:9821-9826, 1996;美国专利号5,786,146)。简言之,DNA由重亚硫酸钠修饰,重亚硫酸钠将未甲基化胞嘧啶转化为尿嘧啶,但不将甲基化胞嘧啶转化为尿嘧啶,并且随后用相较于未甲基化DNA对甲基化DNA具特异性的引物对产物进行扩增。MSP仅需要少量的DNA,对给定CpG岛基因座的0.1%甲基化等位基因敏感,并且可对从石蜡包埋的样品提取的DNA进行。用于MSP分析的典型试剂(例如可在典型的基于MSP的试剂盒中发现)可包括但不限于:用于特定基因座(例如特定基因、标记物、DMR、基因区域、标记物区域、重亚硫酸盐处理过的DNA序列、CpG岛等)的甲基化和未甲基化PCR引物;优化的PCR缓冲液和脱氧核苷酸;以及特异性探针。

[0156] MethyLight™分析为利用基于荧光的实时PCR(例如**TaqMan®**)且在PCR步骤后不需要进一步操纵的高通量定量甲基化分析(Eads等,Cancer Res.59:2302-2306,1999)。简言之,MethyLight™方法从基因组DNA的混合样品开始,将其根据标准程序(重亚硫酸盐方法将未甲基化的胞嘧啶残基转化为尿嘧啶)在重亚硫酸钠反应中转化为具有甲基化依赖性序列差异的混合汇集物。然后例如使用与已知CpG二核苷酸重叠的PCR引物在“偏差”的反应中进行基于荧光的PCR。在扩增过程的层面上与荧光检测过程的层面上均发生序列辨别。

[0157] 使用MethyLight™分析作为核酸(例如基因组DNA样品)中的甲基化模式的定量测试,其中在探针杂交的层面上发生序列辨别。在定量型式中,PCR反应在存在与特定推定甲基化位点重叠的荧光探针的情况下提供甲基化特异性扩增。通过其中引物与探针均不覆盖任何CpG二核苷酸的反应来提供输入DNA的量的无偏对照。或者,通过用不覆盖已知甲基化位点的对照寡核苷酸(例如HeavyMethyl™和MSP技术的基于荧光的型式)或用覆盖潜在甲基化位点的寡核苷酸探测偏差PCR汇集物来实现基因组甲基化的定性测试。

[0158] 将MethyLight™方法与任何合适的探针(例如“**TaqMan®**”探针、**Lightcycler®**探针等)一起使用。举例来说,在一些应用中,将双链基因组DNA用重亚硫酸钠处理,并且进行使用**TaqMan®**探针例如加上MSP引物和/或HeavyMethyl阻断剂寡核苷酸和**TaqMan®**探针的两组PCR反应中的一组。将**TaqMan®**探针用荧光“报告”和“淬灭”分子双重标记,并且设计成对相对高GC含量的区域具特异性,使得它在PCR循环中在比正向引物或反向引物

高约10℃的温度下熔解。这允许 **TaqMan®** 探针在PCR退火/延伸步骤期间保持完全杂交。随着Taq聚合酶在PCR期间酶合成新的链,它最终将到达退火的 **TaqMan®** 探针。然后,Taq聚合酶5'至3'核酸内切酶活性将通过消化 **TaqMan®** 探针释放荧光报告分子来替换 **TaqMan®** 探针,以使用实时荧光检测系统定量检测其现在未淬灭的信号。

[0159] 用于MethyLight™分析的典型试剂(例如如可在典型的基于MethyLight™的试剂盒中发现)可包括但不限于:用于特定基因座(例如特定基因、标记物、DMR、基因区域、标记物区域、重亚硫酸盐处理过的DNA序列、CpG岛等)的PCR引物;**TaqMan®**或**Lightcycler®**探针;优化的PCR缓冲液和脱氧核苷酸;以及Taq聚合酶。

[0160] QM™(定量甲基化)分析为基因组DNA样品中的甲基化模式的替代定量测试,其中在探针杂交的层面上发生序列辨别。在此定量型式中,PCR反应在存在与特定推定甲基化位点重叠的荧光探针的情况下提供无偏扩增。通过其中引物与探针均不与覆盖任何CpG二核苷酸的反应来提供输入DNA的量的无偏对照。或者,通过用不覆盖已知甲基化位点的对照寡核苷酸(HeavyMethyl™和MSP技术的基于荧光的型式)或用覆盖潜在甲基化位点的寡核苷酸探测偏差PCR汇集物来实现基因组甲基化的定性测试。

[0161] 在扩增过程中,可将QM™方法与任何适合的探针(例如“**TaqMan®**”探针、**Lightcycler®**探针)一起使用。举例来说,将双链基因组DNA用重亚硫酸钠处理,并且经受无偏引物和**TaqMan®**探针。将**TaqMan®**探针用荧光“报告”和“淬灭”分子双重标记,并且设计成对相对高GC含量的区域具特异性,使得它在PCR循环中在比正向引物或反向引物高约10℃的温度下熔解。这允许 **TaqMan®** 探针在PCR退火/延伸步骤期间保持完全杂交。随着Taq聚合酶在PCR期间酶合成新的链,它最终将到达退火的 **TaqMan®** 探针。然后,Taq聚合酶5'至3'核酸内切酶活性将通过消化 **TaqMan®** 探针释放荧光报告分子来替换 **TaqMan®** 探针,以使用实时荧光检测系统定量检测其现在未淬灭的信号。用于QM™分析的典型试剂(例如如可在典型的基于QM™的试剂盒中发现)可包括但不限于:用于特定基因座(例如特定基因、标记物、DMR、基因区域、标记物区域、重亚硫酸盐处理过的DNA序列、CpG岛等)的PCR引物;**TaqMan®**或**Lightcycler®**探针;优化的PCR缓冲液和脱氧核苷酸;以及Taq聚合酶。

[0162] Ms-SNuPE™技术为基于对DNA的重亚硫酸盐处理随后单核苷酸引物延伸来评估特定CpG位点的甲基化差异的定量方法(Gonzalzo和Jones,Nucleic Acids Res.25:2529-2531,1997)。简言之,基因组DNA与重亚硫酸钠反应,将未甲基化的胞嘧啶转化为尿嘧啶,同时保持5-甲基胞嘧啶不变。然后使用对重亚硫酸盐转化的DNA具特异性的PCR引物进行所需靶序列的扩增,并且分离所得产物并用作所关注的CpG位点处的甲基化分析的模板。可分析少量的DNA(例如显微解剖的病理切片),并且这避免利用限制酶来测定CpG位点处的甲基化状况。

[0163] 用于Ms-SNuPE™分析的典型试剂(例如如可在典型的基于Ms-SNuPE™的试剂盒中发现)可包括但不限于:用于特定基因座(例如特定基因、标记物、DMR、基因区域、标记物区域、重亚硫酸盐处理过的DNA序列、CpG岛等)的PCR引物;优化的PCR缓冲液和脱氧核苷酸;凝胶提取试剂盒;阳性对照引物;用于特定基因座的Ms-SNuPE™引物;反应缓冲液(用于Ms-

SNuPE反应);以及标记的核苷酸。另外,重亚硫酸盐转化试剂可包括:DNA变性缓冲液;磺化缓冲液;DNA回收试剂或试剂盒(例如沉淀、超滤、亲和柱);脱磺酸缓冲液;以及DNA回收组分。

[0164] 简化表示重亚硫酸盐测序(Reduced Representation Bisulfite Sequencing, RRBS)从重亚硫酸盐处理核酸以将所有未甲基化的胞嘧啶转化为尿嘧啶开始,随后为限制酶消化(例如通过识别包括CG序列的位点的酶,诸如MspI)以及在连接至衔接子配体之后片段的完整测序。限制酶的选择富集了CpG致密区域的片段,从而使在分析期间可能定位到多个基因位置的冗余序列的数目减少。因此,RRBS通过选择用于测序的限制性片段的子集(例如通过使用制备型凝胶电泳的大小选择)降低核酸样品的复杂性。与全基因组重亚硫酸盐测序相反,通过限制酶消化产生的每个片段均含有至少一个CpG二核苷酸的DNA甲基化信息。因此,RRBS富集了样品的启动子、CpG岛以及在这些区域中具有高频率的限制酶切割位点的其他基因组特征,并且因此提供用于评估一个或多个基因组基因座的甲基化状态的分析。

[0165] RRBS的典型方案包括以下步骤:用诸如MspI等限制酶消化核酸样品,填充突出端和A尾,连结衔接子,重亚硫酸盐转化以及PCR。参见例如等(2005)“Genome-scale DNA methylation mapping of clinical samples at single-nucleotide resolution”*Nat Methods* 7:133-6;Meissner等(2005)“Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis”*Nucleic Acids Res.*33:5868-77。

[0166] 在一些实施方案中,使用定量等位基因特异性实时靶标和信号放大(QuARTS)分析来评估甲基化状态。在每次QuARTS分析中依次发生三个反应,包括在初级反应中的扩增(反应1)和靶探针裂解(反应2);以及在次级反应中的FRET裂解和荧光信号产生(反应3)。当用特异性引物扩增靶核酸时,具有侧翼序列(flank sequence)的特异性检测探针与扩增子松散结合。靶结合位点处特异性侵入性寡核苷酸的存在通过在检测探针与侧翼序列之间切割使得裂解酶释放侧翼序列。侧翼序列与相应FRET盒的非发夹部分互补。因此,侧翼序列在FRET盒上充当侵入性寡核苷酸,并且实现FRET盒荧光团与淬灭剂之间的裂解,这产生荧光信号。裂解反应可每个靶标切割多个探针,并且因此每个侧翼释放多个荧光团,从而提供指数信号放大。QuARTS可通过使用具有不同染料的FRET盒在单一反应孔中检测多个靶标。参见例如Zou等(2010)“Sensitive quantification of methylated markers with a novel methylation specific technology”*Clin Chem* 56:A199;美国专利申请序列号12/946,737、12/946,745、12/946,752以及61/548,639。

[0167] 术语“重亚硫酸盐试剂”是指包括重亚硫酸盐、二亚硫酸盐、亚硫酸氢盐或其组合的试剂,所述试剂如本文所公开适用于区分甲基化和未甲基化的CpG二核苷酸序列。所述处理的方法为本领域已知的(例如PCT/EP2004/011715)。优选的是,在存在诸如但不限于正烷基二醇或乙二醇二甲醚(DME)等变性溶剂的情况下或在存在二噁烷或二噁烷衍生物的情况下进行重亚硫酸盐处理。在一些实施方案中,以1%与35%(v/v)之间的浓度使用变性溶剂。在一些实施方案中,在存在诸如但不限于色满衍生物(例如6-羟基-2,5,7,8,-四甲基色满2-甲酸或三羟基苯甲酸及其衍生物,例如没食子酸)等清除剂的情况下进行重亚硫酸盐反应(参见PCT/EP2004/011715)。优选在30℃与70℃之间的反应温度下进行重亚硫酸盐转

化,据此在反应期间温度被短时间升高至超过85°C(参见PCT/EP2004/011715)。优选在定量之前将重亚硫酸盐处理过的DNA纯化。这可通过本领域已知的任何手段进行,诸如但不限于例如借助于Microcon™柱(由Millipore™制造)进行超滤。根据改良的制造商方案进行纯化(参见例如PCT/EP2004/011715)。

[0168] 在一些实施方案中,使用根据本发明的引物寡核苷酸组(例如参见表2)和扩增酶来扩增经处理的DNA的片段。若干DNA区段的扩增可在同一个反应容器中同时进行。典型地,使用聚合酶链反应(PCR)进行扩增。扩增子的长度典型地为100至2000个碱基对。

[0169] 在所述方法的另一实施方案中,可通过使用甲基化特异性引物寡核苷酸来检测包含DMR(例如DMR 1-400;表1和表4)的标记物内或附近的CpG位置的甲基化状况。在Herman的美国专利号6,265,171中已描述了此技术(MSP)。使用甲基化状况特异性引物扩增重亚硫酸盐处理过的DNA允许区分甲基化和未甲基化核酸。MSP引物对含有至少一种与重亚硫酸盐处理过的CpG二核苷酸杂交的引物。因此,所述引物的序列包含至少一个CpG二核苷酸。对非甲基化DNA具特异性的MSP引物在CpG中的C位置的位置处含有“T”。

[0170] 借助于扩增获得的片段可携带直接或间接可检测的标记。在一些实施方案中,标记为荧光标记、放射性核素或具有可在质谱仪中检测的典型质量的可分离分子片段。当所述标记为质量标记时,一些实施方案提供具有单一正净电荷或负净电荷的标记的扩增子,从而允许在质谱仪中具有更好的可检测性。可借助于例如基质辅助激光解吸/电离质谱法(MALDI)或使用电喷雾质谱法(ESI)来进行检测和可视化。

[0171] 用于分离适合于这些分析技术的DNA的方法为本领域已知的。特定而言,一些实施方案包括如美国专利申请序列号13/470,251(“Isolation of Nucleic Acids”)中所描述分离核酸。

[0172] 方法

[0173] 在所述技术的一些实施方案中,提供了包括以下步骤的方法:

[0174] 1) 使从受试者获得的核酸(例如基因组DNA,所述基因组DNA例如分离自体液,诸如血液样品(例如血浆样品)、粪便样品或组织样品)与区分包含DMR(例如DMR 1-400(来自表1和表4))的至少一种标记物内的甲基化和非甲基化CpG二核苷酸的至少一种试剂或一系列试剂接触,以及

[0175] 2) 检测HCC的缺乏(例如以大于或等于80%的灵敏度和大于或等于80%的特异性提供)。

[0176] 在所述技术的一些实施方案中,提供了包括以下步骤的方法:

[0177] 1) 使从受试者获得的核酸(例如基因组DNA,所述基因组DNA例如分离自体液,诸如血液样品(例如血浆样品)、粪便样品或组织样品)与区分包含DMR(例如DMR 1-400(来自表1和表4))的至少一种标记物内的甲基化和非甲基化CpG二核苷酸的至少一种试剂或一系列试剂接触,以及

[0178] 2) 将HCC归类(例如以大于或等于80%的灵敏度和大于或等于80%的特异性提供)。

[0179] 优选地,灵敏度为约70%至约100%,或约80%至约90%,或约80%至约85%。优选地,特异性为约70%至约100%,或约80%至约90%,或约80%至约85%。

[0180] 可通过任何手段来分离基因组DNA,包括使用可商购获得的试剂盒。简言之,在所

关注的DNA由细胞膜封装的情况下,必须通过酶、化学或机械手段破碎和溶解生物样品。然后可例如通过用蛋白酶K消化来从DNA溶液中清除蛋白质和其他污染物。然后从溶液中回收基因组DNA。这可借助于多种方法来进行,包括盐析、有机萃取或使DNA与固相载体结合。方法的选择将受若干因素影响,包括时间、费用以及所需DNA数量。包含赘生物质或赘生前物质的所有临床样品类型均适用于本发明方法,例如细胞系、组织学切片、活体组织切片、石蜡包埋的组织、体液、粪便、结肠流出物、尿液、血浆、血清、全血、分离的血细胞、从血液中分离的细胞以及其组合。

[0181] 所述技术在用于制备样品并提供用于测试的核酸的方法方面不受限制。举例来说,在一些实施方案中,使用例如如美国专利申请序列号61/485386中详述的直接基因捕获或通过相关方法从粪便样品或从血液或从血浆样品中分离DNA。

[0182] 然后用区分含有DMR(例如DMR 1-400,例如如由表1和表4所提供)的至少一种标记物中的甲基化和非甲基化CpG二核苷酸的至少一种试剂或一系列试剂处理基因组DNA样品。

[0183] 在一些实施方案中,试剂将在5'-位置未甲基化的胞嘧啶碱基转化为尿嘧啶、胸腺嘧啶或就杂交行为来说与胞嘧啶不相似的另一碱基。然而,在一些实施方案中,试剂可为甲基化敏感性限制酶。

[0184] 在一些实施方案中,以如下方式处理基因组DNA样品,所述方式使得在5'位置处未甲基化的胞嘧啶碱基转化为尿嘧啶、胸腺嘧啶或就杂交行为来说与胞嘧啶不相似的另一碱基。在一些实施方案中,用硫酸氢盐(亚硫酸氢盐、二亚硫酸盐)进行此处理,随后进行碱水解。

[0185] 然后,分析所处理的核酸以测定靶基因序列(来自包含DMR(例如选自例如如表1和表4中所提供的DMR 1-400的至少一个DMR)的标记物的至少一个基因、基因组序列或核苷酸)的甲基化状态。分析方法可选自本领域已知的那些分析方法,包括本文列举的那些分析方法,例如如本文所描述的QuARTS和MSP。

[0186] 所述技术涉及分析与HCC相关的任何样品。举例来说,在一些实施方案中,样品包括来自患者的血浆样品。在一些实施方案中,样品包括从患者获得的组织和/或生物流体。在一些实施方案中,样品包括肝组织。在一些实施方案中,样品包括分泌物。在一些实施方案中,样品包括血液、血浆和/或血清。在一些实施方案中,受试者为人。这些样品可来源于上部胃肠道、下部胃肠道或包括来自上部胃肠道与下部胃肠道两者的细胞、组织和/或分泌物。样品可包括来自肝脏、胆管、胰腺、胃、结肠、直肠、食管、小肠、阑尾、十二指肠、息肉、胆囊、肛门和/或腹膜的细胞、分泌物或组织。在一些实施方案中,样品包括细胞液、腹水、尿液、粪便、胰液、内窥镜检查期间获得的流体、血液、粘液或唾液。在一些实施方案中,样品为粪便样品。

[0187] 可通过本领域已知的诸如将对熟练人员来说显而易见的许多手段来获得此类样品。举例来说,尿液和粪便样品为可容易得到的,而血液、腹水、血清或胰液样品可通过使用例如针头和注射器经肠胃外获得。无细胞或基本上无细胞的样品可通过使样品经受本领域技术人员已知的包括但不限于离心和过滤的各种技术来获得。虽然通常优选的是不使用侵入性技术来获得样品,但获得诸如组织匀浆、组织切片以及活体组织切片试样等样品仍可能为优选的。

[0188] 在所述技术的一些实施方案中,提供一种诊断受试者中的HCC的方法。如本文所用

的术语“诊断(diagnosing/diagnosis)”是指使得熟练技工可估计并且甚至确定受试者是否罹患给定疾病或疾患或将来可能发展给定疾病或疾患的方法。熟练技工常常基于一种或多种诊断指示物作出诊断,诸如生物标记物(例如如本文所公开的DMR),其甲基化状态指示疾患的存在、严重程度或不存在。

[0189] 与诊断一起,临床癌症预后(例如对于HCC)涉及确定癌症的侵入性和肿瘤复发的可能性以计划出最有效的疗法。如果可作出更准确的预后或甚至可评估发展癌症的潜在危险,则可选择对于患者适当的疗法以及在某些情况下较不苛刻的疗法。评估癌症生物标记物(例如测定其甲基化状态)适用于将具有良好预后和/或低癌症发展危险的将不需要治疗或需要有限治疗的受试者与较可能发展癌症或经历癌症复发的可受益于更深入的治疗的受试者分开。

[0190] 因此,如本文所用,“作出诊断”或“诊断”还包括确定发展癌症的危险或确定预后,这些可基于对本文所公开的诊断性生物标记物(例如DMR)的测量提供以下各项:预测临床结果(在存在或不存在医学治疗的情况下)、选择适当治疗(或是否治疗将有效)或监测当前的治疗并潜在地改变所述治疗。另外,在目前所公开的主题的一些实施方案中,可随时间推移进行生物标记物的多次测定以有助于诊断和/或预后。生物标记物的时间变化可用于预测临床结果、监测HCC的进展和/或监测针对癌症的适当疗法的效力。在此类实施方案中,举例来说,在有效治疗的过程中,预期可观察到生物样品中本文所公开的一种或多种生物标记物(例如DMR)(以及潜在的一种或多种额外生物标记物(如果监测的话))的甲基化状态随时间的变化。

[0191] 在一些实施方案中,目前所公开的主题还提供一种确定是否开始或继续对受试者中的HCC的预防或治疗的方法。在一些实施方案中,所述方法包括在一段时间内从受试者提供一系列生物样品;分析这一系列生物样品以测定各个生物样品中的本文所公开的至少一种生物标记物的甲基化状态;以及比较各个生物样品中一种或多种生物标记物的甲基化状态的任何可测量的变化。生物标记物在这段时间内的甲基化状态的任何变化均可用于预测发展HCC的危险,预测临床结果,确定是否开始或继续对癌症的预防或治疗以及当前治疗是否有效治疗HCC。举例来说,可将第一时间点选择为在开始治疗之前,并且可将第二时间点选择为在开始治疗之后的一些时间。可在从不同时间点取得的各个样品中测量甲基化状态,并且记录定性和/或定量差异。可将来自不同样品的生物标记物水平的甲基化状态的变化与受试者中的病症的危险(例如HCC的危险)、预后、确定治疗效力和/或进展相关联。

[0192] 在优选实施方案中,本发明的方法和组合物是用于在早期阶段(例如在疾病症状出现之前)治疗或诊断疾病。在一些实施方案中,本发明的方法和组合物是用于在临床阶段治疗或诊断疾病。

[0193] 如所提到,在一些实施方案中,可进行一种或多种诊断或预后性生物标记物的多次测定,并且可使用标记物的时间变化来确定诊断或预后。举例来说,可在起始时间测定诊断性标记物,并且在第二时间再次测定。在此类实施方案中,从起始时间到第二时间标记物增加可指示病症的特定类型或严重程度,或给定预后。同样地,从初始时间到第二时间标记物减少可指示病症的特定类型或严重程度,或给定预后。此外,一种或多种标记物的变化程度可与病症的严重程度和将来的有害事件有关。熟练技工将了解,虽然在某些实施方案中,比较测量可由多个时间点的同一生物标记物构成,但也可在一个时间点测量给定生物标记

物,并且在第二时间点测量第二生物标记物,并且这些标记物的比较可提供诊断信息。

[0194] 如本文所用,短语“确定预后”是指使得熟练技工可预测受试者中的疾患的过程或结果的方法。术语“预后”不是指以100%准确性预测疾患的过程或结果的能力,或者甚至基于生物标记物(例如DMR)的甲基化状态可预测给定过程或结果或多或少可能发生。事实上,熟练技工将了解,术语“预后”是指某过程或结果将发生的可能性增加;换句话说,当与未展现给定疾患的个体相比时,在展现所述疾患的受试者中过程或结果更可能发生。举例来说,在未展现疾患(例如具有一种或多种DMR的正常甲基化状态)的个体中,给定结果的可能性可能非常低。

[0195] 在一些实施方案中,统计分析将预后指示物与有害结果的倾向相关联。举例来说,在一些实施方案中,与从未患病症的患者获得的正常对照样品中的甲基化状态不同的甲基化状态可表明,如通过统计显著性水平所确定,受试者同具有与对照样品中的甲基化状态更类似的水平的受试者相比更可能罹患病症。另外,甲基化状态相较于基线(例如“正常”)水平的变化可反映受试者的预后,并且甲基化状态的变化程度可与有害事件的严重程度相关。常常通过比较两个或更多个群体并且确定置信区间和/或p值来确定统计显著性(参见例如Dowdy和Wearden, *Statistics for Research*, John Wiley&Sons, New York, 1983)。本发明主题的示例性置信区间为90%、95%、97.5%、98%、99%、99.5%、99.9%以及99.99%,而示例性p值为0.1、0.05、0.025、0.02、0.01、0.005、0.001以及0.0001。

[0196] 在其他实施方案中,可建立本文所公开的预后或诊断性生物标记物(例如DMR)的甲基化状态的阈值变化程度,并且将生物样品中生物标记物的甲基化状态的变化程度与甲基化状态的阈值变化程度进行简单比较。本文提供的生物标记物的甲基化状态的优选阈值变化为约5%、约10%、约15%、约20%、约25%、约30%、约50%、约75%、约100%以及约150%。在其他实施方案中,可建立“列线图(nomogram)”,根据“列线图”,预后或诊断性指示物(生物标记物或生物标记物的组合)的甲基化状态与给定结果的相关倾向直接相关。熟练技工熟悉使用此类列线图来使两个数值相关联,并且了解此测量的不确定性与标记物浓度的不确定性相同,因为提到了个别样品测量,而不是群体平均值。

[0197] 在一些实施方案中,同时分析了对照样品与生物样品,使得可将生物样品获得的结果与从对照样品获得的结果相比较。另外,预期可提供标准曲线,使用所述标准曲线可将生物样品的分析结果进行比较。如果使用荧光标记,则此类标准曲线呈现作为分析单位(例如荧光信号强度)的函数的生物标记物的甲基化状态。使用从多个供体获取的样品,可提供针对正常组织中一种或多种生物标记物的对照甲基化状态以及针对从具有化生的供体或从具有病症(例如HCC)的供体取得的组织中一种或多种生物标记物的“处于危险中”的水平的标准曲线。在所述方法的某些实施方案中,在确定从受试者获得的生物样品中的本文所提供的一种或多种DMR的异常甲基化状态后将受试者鉴定为患有HCC。在所述方法的其他实施方案中,从受试者获得的生物样品中此类生物标记物中的一者或多者的异常甲基化状态的检测使得受试者被鉴定为患有HCC。

[0198] 可在一个测试样品内与额外标记物分开或同时进行标记物分析。举例来说,可将若干标记物组合在一个测试中用于有效处理多个样品并且潜在地提供更大的诊断和/或预后准确性。此外,本领域技术人员将认可测试来自相同受试者的多个样品(例如在连续时间点)的价值。系列样品的此类测试可允许鉴定标记物甲基化状态随时间的变化。甲基化状态

的变化以及不存在甲基化状态的变化可提供关于疾病状况的有用信息,包括但不限于确定距离事件开始的大致时间、可救治组织的存在和量、药物疗法的适当性、各种疗法的有效性以及鉴定受试者的结果,包括将来事件的危险。

[0199] 可以多种物理形式进行生物标记物的分析。举例来说,可使用微量滴定板或自动化来促进大量测试样品的处理。或者,可研发单一样品形式,以便于及时地进行立即治疗和诊断,例如在非卧床运输或急诊室环境中。

[0200] 在一些实施方案中,如果当与对照甲基化状态相比时,样品中至少一种生物标记物的甲基化状态存在可测量的差异,则将受试者诊断为患有HCC。相反地,当在生物样品中确定甲基化状态没有变化时,可将受试者鉴定为未患HCC,不处于HCC危险之中,或具有低HCC危险。在此方面,可将患有HCC或具有HCC危险的受试者与具有低HCC或HCC危险至基本上没有HCC或HCC危险的受试者区分开。可将具有发展HCC的危险的那些受试者放置于更深入和/或定期的筛查时间表中。

[0201] 如上文所提到,根据本发明技术的方法的实施方案,检测一种或多种生物标记物的甲基化状态的变化可为定性测定,或可为定量测定。因此,将受试者诊断为患有HCC或处于发展HCC的危险中的步骤指示进行某些阈值测量,例如生物样品中一种或多种生物标记物的甲基化状态与预定对照甲基化状态不同。在所述方法的一些实施方案中,对照甲基化状态为生物标记物的任何可检测的甲基化状态。在将对照样品与生物样品同时测试的方法的其他实施方案中,预定甲基化状态为对照样品中的甲基化状态。在所述方法的其他实施方案中,预定甲基化状态是基于标准曲线和/或通过标准曲线来确定。在所述方法的其他实施方案中,预定甲基化状态为特定状态或状态范围。因此,可部分基于所实践方法的实施方案和所需特异性等,在本领域技术人员将显而易见的可接受限值内选择预定甲基化状态。

[0202] 此外,关于诊断方法,优选受试者为脊椎动物受试者。优选脊椎动物为温血脊椎动物;优选温血脊椎动物为哺乳动物。优选哺乳动物最优选为人。如本文所用,术语“受试者”包括人与动物受试者两者。因此,本文提供了兽医学治疗用途。因此,本发明技术提供了对诸如人以及以下动物等哺乳动物的诊断:因濒危而具有重大意义的那些哺乳动物,诸如西伯利亚虎(Siberian tiger);具有经济意义的动物,诸如农场上饲养以供人消耗的动物;和/或对人具有社会意义的动物,诸如作为宠物或在动物园中饲养的动物。此类动物的实例包括但不限于:食肉动物,诸如猫和狗;猪类,包括猪、阉猪以及野猪;反刍动物和/或有蹄动物,诸如牛、公牛、绵羊、长颈鹿、鹿、山羊、野牛以及骆驼;以及马。因此,还提供了对牲畜,包括但不限于驯养的猪、反刍动物、有蹄动物、马(包括赛马)等的诊断和治疗。目前公开的主题还包括用于诊断受试者中的HCC的系统。所述系统可例如作为商业试剂盒提供,所述商业试剂盒可用于筛查已收集生物样品的受试者中的此类病症的危险。根据本发明技术提供的示例性系统包括评估如表1和/或表4中所提供的DMR的甲基化状态。

[0203] 实施例

[0204] 实施例I.

[0205] 此实施例描述用于辨别HCC DNA样品与衍生自正常对照(例如患有或未患肝硬化的非HCC个体)的DNA的311个差异甲基化区(DMR)的鉴定。

[0206] 以四个阶段进行实验。

[0207] 第一,使用RRBS对从冷冻肿瘤HCC组织(具有和不具有肝硬化)以及从冷冻正常肝

组织(具有和不具有肝硬化)以及来自健康志愿者的血沉棕黄层样品提取的DNA进行DNA甲基化标记物发现。通过严格过滤准则鉴定辨别性差异甲基化区(DMR)并且使用实时甲基化特异性定量PCR分析(qMSP)再次对相同或扩展的试样进行分析以确保结果的再现性(技术验证)。

[0208] 第二,通过用于生物验证的进一步排列准则通过对从独立档案病例和对照组织提取的DNA进行设盲qMSP分析来选择候选标记物。

[0209] 第三,在pan-GI RRBS测序集合中比较候选标记物的测序结果以确定甲基化特异性的水平。

[0210] 第四,应用单独的决策模型来选择将在大部分DNA来自非肝脏来源的基于血液的环境中表现最佳的小的HCC标记物集合。然后,在设盲独立血浆样品中测试所选标记物以评估临床介质中的HCC检测。

[0211] 图1汇总了这四个阶段。

[0212] 结果

[0213] 对从18个HCC和35个对照(9个肝硬化、26个正常肝脏)组织提取的DNA进行无偏差全甲基组测序。通过组织验证,使用甲基化特异性PCR在从来自75个HCC和29个对照(16个肝硬化、13个正常肝脏)的独立组织提取的DNA中确认了的最佳DMR。然后,对来自独立患者集合的血浆DNA进行靶向最优DMR的设盲定量等位基因特异性实时靶标和信号放大分析,所述独立患者集合包含21个HCC病例(9个BCLC[巴塞罗那诊所肝癌分期(Barcelona Clinic Liver Cancer staging)]阶段A、6个阶段B、6个阶段C)以及33个肝硬化对照。使用递归分割决策分析来鉴定最佳DMR组合。初始测序鉴定出311个AUC大于0.75的DMR。在生物验证之后,选择最优的12个DMR(ACPI、BDH1、Chr12.133、CLEC11A、DAB2IP、DBNL、EMX1、EFNB2、HOXA1、LRRC4、SPINT2、TSPYL5、CCNJ_3707、CCNJ_3124、PFKP、SCRN1以及ECE1)来进行血浆测试。单独的在血浆中最具辨别力的标记物EMX1具有0.89的AUC。互补3-标记物组合(EMX1、LRRC4以及BDH1)在血浆中鉴定出20/21个HCC和32/33个对照;1个HCC具有低水平的BDH1,并且1个对照具有升高的LRRC4。对于HCC在97%特异性(95%CI,82%-100%)下该组为95%敏感的(95%CI,74%-100%)并且实现0.98的AUC(参见图2)。图3A至图3I提供了生物组织验证阶段的ACPI、Chr12.133、CLEC11A、DAB2IP、DBNL、EMX1、HOXA1、LRRC4、SPINT2以及TSPYL5的接受者操作特征曲线下面积信息。

[0214] 图4A至图4CC提供了来自生物组织验证数据的27种胃癌标记物(在增加的拟合点分析情况下为29种)的箱线图(对数刻度)。将正常肝脏情况下的样品安排在最左侧,随后为不具有肝硬化的HCC、具有肝硬化的HCC以及肝硬化对照(炎性)。竖直轴为分数甲基化(针对 β -肌动蛋白链归一化)。

[0215] 图5示出了在95%正常特异性下呈基质形式的于75个HCC组织样品和29个对照(16个肝硬化、13个正常肝脏)中的27种HCC癌症标记物的性能。标记物竖直列出,而样品水平列出。将正常肝脏(NI)情况下的样品安排在最左侧,随后为不具有肝硬化的HCC(HN)、具有肝硬化的HCC(HC)以及肝硬化对照(In)。阳性命中呈浅灰色,而丢失呈深灰色。此曲线允许以互补方式评估标记物。注意:基于qMSP数据使用拟合点法第二次分析2种标记物TBX15和EGR2并且包括在本文中。

[0216] 表1提供区别HCC与正常对照的DMR的信息。

[0217] 表2提供用于来自表1的所选DMR的引物信息。

[0218] 表3提供在HCC与正常肝脏之间进行比较时特定DMR的AUC和倍数变化信息,其中倍数变化为病例的分数甲基化相较于对照的分数甲基化的比率。

[0219] 表1.

[0220]

DMR 编号	基因注释	染色体编号	DMR 起始和终止位置
1	Septin9	chr17	75368812-75368887
2	Septin9	chr17	75369228-75369313
3	Septin9	chr17	75368948-75369052
4	ACP1	chr2	264204-264250
5	ACP1	chr2	264087-264151
6	ACTG1	chr17	79481554-79481612
7	AKR1B1	chr7	134143298-134143341
8	AKR1B1	chr7	134143650-134143684
9	AKR1B1	chr7	134143461-134143500
10	ALDH1A3	chr15	101419689-101419707
11	ALDH1A3	chr15	101419988-101420013
12	ANKRD33B	chr5	10564710-10564793
13	ANKRD34B	chr5	79866119-79866169
14	ARHGEF19	chr1	16533241-16533318
15	ARL4C	chr2	235404686-235404738
16	ATL1	chr14	51027344-51027417
17	B3GNT4	chr12	122688498-122688504
18	B3GNT4	chr12	122688516-122688554
19	B3GNT5	chr3	182971710-182971817
20	B3GNT5	chr3	182971523-182971580
21	BAALC	chr8	104153072-104153134
22	BACE2	chr21	42539722-42539743
23	BACE2	chr21	42539915-42539974
24	BLVRB	chr19	40973164-40973202
25	BLVRB	chr19	40973019-40973041
26	BMP6	chr6	7727949-7728013
27	BMP6	chr6	7726324-7726384
28	BVES	chr6	105584689-105584756
29	BVES	chr6	105584565-105584590
30	BVES	chr6	105584763-105584789
31	C6orf126	chr6	35744417-35744467
32	C6orf174	chr6	127837271-127837331
33	C7orf51	chr7	100091317-100091392
34	C7orf57	chr7	48075209-48075261
35	CATSPERG	chr19	38852416-38852441
36	CATSPERG	chr19	38853253-38853336

[0221]

37	CCDC48	chr3	128745984-128746119
38	CCNI2	chr5	132082945-132083024
39	CCNJ	chr10	97803226-97803325
40	CCNJ	chr10	97803156-97803203
41	CCNJ	chr10	97803124-97803152
42	CCNJ	chr10	97803707-97803733
43	CDKN2A	chr9	21974872-21974890
44	CDKN2A	chr9	21974943-21975018
45	CELF2	chr10	11059845-11059861
46	CELF2	chr10	11059991-11060017
47	CELSR3	chr3	48693898-48693974
48	CHST2	chr3	142839074-142839107
49	CHST2	chr3	142839408-142839457
50	CHST2	chr3	142838525-142838580
51	CHST2	chr3	142838038-142838164
52	CHST2	chr3	142838701-142838753
53	CHST2	chr3	142838686-142838696
54	CHST2	chr3	142839245-142839275
55	CHST2	chr3	142838916-142838930
56	CLIP4	chr2	29338189-29338335
57	CORO2B	chr15	68871554-68871567
58	CORO2B	chr15	68871572-68871581
59	CORO2B	chr15	68871623-68871639
60	CORO2B	chr15	68871190-68871224
61	CRHR2	chr7	30722095-30722147
62	DCLK2	chr4	150999533-150999622
63	DCLK2	chr4	150999775-150999832
64	DGCR14	chr22	19137109-19137145
65	DGCR14	chr22	19137067-19137086
66	DKFZp686O24166	chr11	17373279-17373346
67	DNAJB6	chr7	157129265-157129325
68	DUSP4	chr8	29207072-29207112
69	DUSP4	chr8	29207421-29207440
70	ECE1	chr1	21616801-21616877
71	EFNB2	chr13	107188064-107188108
72	EFNB2	chr13	107188196-107188212
73	EFNB2	chr13	107187748-107187766
74	EFNB2	chr13	107187768-107187785
75	EFNB2	chr13	107188435-107188478
76	EFNB2	chr13	107188935-107189019
77	EFNB2	chr13	107188790-107188904
78	EFNB2	chr13	107187702-107187716
79	EFNB2	chr13	107188503-107188558
80	EGR2	chr10	64575092-64575141
81	EHD3	chr2	31456843-31456921
82	EHD3	chr2	31457043-31457061
83	EIF4E3	chr3	71803308-71803340
84	EIF4E3	chr3	71803236-71803290
85	EMILIN2	chr18	2847067-2847090
86	EMILIN2	chr18	2847092-2847100
87	EMILIN2	chr18	2847233-2847274

[0222]

88	FAHD2B	chr2	97760723-97760782
89	FAM105A	chr5	14582008-14582040
90	FAM105A	chr5	14582046-14582064
91	FAM105A	chr5	14582380-14582417
92	FAM55C	chr3	101498249-101498337
93	FBXL19	chr16	30935481-30935570
94	FCHSD1	chr5	141031071-141031122
95	FCHSD1	chr5	141031057-141031069
96	FHOD1	chr16	67281136-67281198
97	FHOD1	chr16	67281037-67281085
98	FIBCD1	chr9	133815197-133815238
99	FIBCD1	chr9	133814914-133814942
100	FOXD1	chr5	72743623-72743647
101	FOXD1	chr5	72743444-72743467
102	FOXD1	chr5	72743469-72743495
103	FUT4	chr11	94277618-94277669
104	FYN	chr6	112194124-112194158
105	FYN	chr6	112194611-112194633
106	GALNT12	chr9	101569940-101569975
107	GALNT12	chr9	101570002-101570029
108	GNG4	chr1	235813128-235813161
109	GNG4	chr1	235813683-235813724
110	GPX7	chr1	53068129-53068171
111	GPX7	chr1	53068016-53068063
112	GPX7	chr1	53068089-53068108
113	GRID2IP	chr7	6543224-6543246
114	GRID2IP	chr7	6570807-6570820
115	GRID2IP	chr7	6543285-6543348
116	GRID2IP	chr7	6570602-6570650
117	HDGFRP3	chr15	83876476-83876501
118	HDGFRP3	chr15	83876552-83876612
119	HDGFRP3	chr15	83876614-83876641
120	HK1	chr10	71078304-71078341
121	HK1	chr10	71078661-71078739
122	HOXA13	chr7	27239880-27239914
123	HOXA13	chr7	27239850-27239876
124	HPDL	chr1	45792443-45792512
125	IKZF1	chr7	50343562-50343603
126	IKZF1	chr7	50343493-50343512
127	IKZF1	chr7	50343859-50343873
128	IKZF1	chr7	50344046-50344106
129	IKZF1	chr7	50343990-50344044
130	ITPR3	chr6	33601477-33601537
131	KCNQ3	chr8	133493278-133493345
132	KCNS2	chr8	99439489-99439524
133	KCNS2	chr8	99439451-99439464
134	KCTD12	chr13	77460548-77460628
135	KCTD12	chr13	77460089-77460122
136	KCTD12	chr13	77459810-77459868
137	KIAA1614	chr1	180881686-180881721
138	KIAA1614	chr1	180881734-180881776
139	LCNL1	chr9	139872612-139872667

[0223]

140	LGALS3	chr14	55596202-55596277
141	LIMD2	chr17	61777597-61777647
142	LMO2	chr11	33891092-33891223
143	LOC440461	chr17	66195307-66195331
144	LOC440461	chr17	66195475-66195486
145	LOC440461	chr17	66195649-66195686
146	LOC441617	chr11	74953198-74953215
147	LOC441617	chr11	74952822-74952901
148	LOC441617	chr11	74953185-74953195
149	LPAR2	chr19	19739243-19739256
150	LPAR2	chr19	19739094-19739141
151	LRRC10B	chr11	61276689-61276765
152	LRRC10B	chr11	61276161-61276167
153	LRRC34	chr3	169530174-169530220
154	LRRC34	chr3	169530317-169530418
155	LRRK1	chr15	101459830-101459847
156	LRRK1	chr15	101459790-101459821
157	LRRN1	chr3	3841285-3841390
158	MATK	chr19	3785888-3785923
159	MATK	chr19	3785942-3785986
160	MAX.chr1.1535351-1535441	chr1	1535351-1535441
161	MAX.chr1.161582155-161582236	chr1	161582155-161582236
162	MAX.chr1.227976136-227976180	chr1	227976136-227976180
163	MAX.chr1.227976189-227976206	chr1	227976189-227976206
164	MAX.chr1.41847970-41848023	chr1	41847970-41848023
165	MAX.chr11.1357749-1357833	chr11	1357749-1357833
166	MAX.chr11.45377012-45377082	chr11	45377012-45377082
167	MAX.chr11.75947388-75947438	chr11	75947388-75947438
168	MAX.chr17.1132739-1132794	chr17	1132739-1132794
169	MAX.chr19.37288291-37288390	chr19	37288291-37288390
170	MAX.chr19.37464150-37464218	chr19	37464150-37464218
171	MAX.chr19.9896775-9896833	chr19	9896775-9896833
172	MAX.chr2.144694740-144694873	chr2	144694740-144694873
173	MAX.chr2.219773668-219773754	chr2	219773668-219773754
174	MAX.chr20.62461671-62461721	chr20	62461671-62461721
175	MAX.chr3.184243258-184243328	chr3	184243258-184243328
176	MAX.chr4.56915281-56915399	chr4	56915281-56915399
177	MAX.chr5.178957711-178957760	chr5	178957711-178957760
178	MAX.chr6.130686855-130686917	chr6	130686855-130686917
179	MAX.chr6.155316859-155316913	chr6	155316859-155316913
180	MAX.chr7.139930482-	chr7	139930482-139930532

[0224]

	139930532		
181	MAX.chr9.140024004-140024057	chr9	140024004-140024057
182	MAX.chr9.99983940-99983992	chr9	99983940-99983992
183	MECOM	chr3	169380257-169380286
184	MECOM	chr3	169380412-169380427
185	MEX3B	chr15	82339734-82339758
186	MEX3B	chr15	82339894-82339967
187	MGC16703	chr22	21368623-21368674
188	MIR3132	chr2	220417386-220417427
189	MIR3132	chr2	220417233-220417279
190	MIR92B	chr1	155164705-155164747
191	MIR92B	chr1	155164749-155164784
192	MNT	chr17	2297347-2297397
193	N4BP3	chr5	177540980-177541072
194	NCRNA00085	chr19	52207589-52207731
195	NCRNA00085	chr19	52207475-52207527
196	OBSCN	chr1	228401498-228401559
197	OBSCN	chr1	228463603-228463619
198	ODF2L	chr1	86861515-86861590
199	OVOL1	chr11	65554351-65554410
200	OXT	chr20	3052753-3052809
201	PAQR8	chr6	52227453-52227521
202	PDE4D	chr5	58335632-58335685
203	PFKP	chr10	3110620-3110694
204	PFKP	chr10	3110554-3110609
205	PFKP	chr10	3111174-3111244
206	PFKP	chr10	3110913-3110978
207	PLEKHO1	chr1	150122743-150122798
208	PMAIP1	chr18	57567061-57567089
209	PMAIP1	chr18	57567094-57567115
210	PPP2R5C	chr14	102248080-102248126
211	PPP2R5C	chr14	102248158-102248201
212	PRDM13	chr6	100061630-100061691
213	PRDM5	chr4	121843466-121843518
214	PSMG2	chr18	12658195-12658238
215	PSMG2	chr18	12658247-12658264
216	RANGRF	chr17	8192319-8192372
217	RANGRF	chr17	8192381-8192390
218	RASL11B	chr4	53728280-53728354
219	RASSF2	chr20	4803295-4803348
220	RASSF2	chr20	4803983-4804072
221	RECK	chr9	36037020-36037061
222	RECK	chr9	36037093-36037184
223	RGS10	chr10	121302301-121302309
224	RGS10	chr10	121302871-121302922
225	RGS10	chr10	121302655-121302672
226	RGS20	chr8	54793664-54793736
227	RPP25	chr15	75248658-75248693
228	RPP25	chr15	75249650-75249686
229	SCRN1	chr7	30029222-30029267
230	SCRN1	chr7	30029115-30029160

[0225]

231	SCRN1	chr7	30029187-30029209
232	SDK1	chr7	3340834-3340842
233	SDK1	chr7	3340543-3340635
234	SDK1	chr7	3340737-3340813
235	SDK1	chr7	3340869-3340900
236	SLC16A3	chr17	80186851-80186883
237	SLC16A3	chr17	80186820-80186835
238	SLC35F1	chr6	118228406-118228451
239	SLC35F1	chr6	118228566-118228578
240	SLC35F1	chr6	118228886-118228910
241	SLC6A20	chr3	45838034-45838084
242	SLC6A6	chr3	14443943-14444020
243	SLC7A5	chr16	87902822-87902863
244	SLC7A5	chr16	87903100-87903127
245	SPNS2	chr17	4403134-4403184
246	ST8SIA1	chr12	22487540-22487711
247	ST8SIA6	chr10	17496576-17496610
248	ST8SIA6	chr10	17496378-17496406
249	STK32C	chr10	134121022-134121044
250	STK32C	chr10	134121172-134121183
251	STK32C	chr10	134122104-134122151
252	STK32C	chr10	134122249-134122269
253	STXBP1	chr9	130370839-130370874
254	STXBP1	chr9	130370883-130370925
255	SYCE1L	chr16	77246378-77246432
256	TAF4B	chr18	23806638-23806671
257	TAF4B	chr18	23806718-23806723
258	TCF24	chr8	67875000-67875071
259	TMEM143	chr19	48837211-48837295
260	TMEM163	chr2	135476181-135476192
261	TMEM163	chr2	135476038-135476121
262	TMEM163	chr2	135476286-135476322
263	TRIM17	chr1	228604397-228604446
264	TRPC3	chr4	122872963-122873035
265	TSC22D1	chr13	45150818-45150847
266	TSC22D1	chr13	45150763-45150815
267	UAP1L1	chr9	139972155-139972218
268	VASH1	chr14	77228066-77228133
269	VASH2	chr1	213124546-213124598
270	WDR66	chr12	122356382-122356450
271	WNT7A	chr3	13921575-13921631
272	ZC3HAV1L	chr7	138720665-138720718
273	ZEB2	chr2	145274897-145275040
274	ZNF160	chr19	53606314-53606402
275	ZNF160	chr19	53606256-53606282
276	ZNF160	chr19	53606572-53606659
277	ZNF354C	chr5	178487210-178487287
278	ZNF468	chr19	53360825-53360874
279	ZNF506	chr19	19932506-19932566
280	ZNF510	chr9	99540238-99540295
281	ZNF549	chr19	58038905-58038971
282	ZNF568	chr19	37407242-37407355

[0226]

283	ZNF607	chr19	38210327-38210438
284	ZNF611	chr19	53238186-53238286
285	ZNF671	chr19	58238870-58238942
286	ZNF816	chr19	53466145-53466206
287	ZNF816	chr19	53466048-53466114
288	SPINT2	chr19	38755130-38755164
289	DBNL	chr7	44080227-44080310
290	PRKAG2	chr7	151329671-151329753
291	TDH	chr8	11205044-11205092
292	COTL1	chr16	84651245-84651314
293	PHF21B	chr22	45403120-45403172
294	PHF21B	chr22	45404843-45404858
295	PHF21B	chr22	45405759-45405766
296	MCF2L2	chr3	183146162-183146204
297	MCF2L2	chr3	183146290-183146355
298	MCF2L2	chr3	182897154-182897207
299	AMN	chr14	103394822-103394849
300	MARK1	chr1	220701604-220701619
301	MARK1	chr1	220701582-220701599
302	MARK1	chr1	220701657-220701686
303	EMX1	chr2	73147710-73147772
304	TSPYL5	chr8	98289858-98290220
305	DAB2IP	chr9	124461305-124461420
306	CLEC11A	chr19	51228217-51228732
307	HOXA1	chr7	27136145-27136425
308	ADCY1	chr7	45613877-45614572
309	DMRTA2	chr1	50884349-50884499
310	TBX15	chr1	119527066-119527655
311	AK055957	chr12	133484978-133485739

[0227] 表2.

[0228]

标记物	寡核苷酸 (OLIGO) 类型	序列
TSPYL5	引物	GCGCGGGAGGATTTTCG (SEQ ID NO: 1)
	引物	CCGCCACCATAAACGACC (SEQ ID NO: 2)
	探针	CCACGGACG CGAAATCGAAAT/3C6/ (SEQ ID NO: 3)
DAB2IP	引物	CGTTCGTTACGTCGTTTTCGT (SEQ ID NO: 4)
	引物	GATCGACGCGACTCGAC (SEQ ID NO: 5)
	探针	CCACGGACGCTCGACGTCGCC/3C6/ (SEQ ID NO: 6)
CLEC11A	引物	GCGGGAGTTTGGCGTAG (SEQ ID NO: 7)
	引物	CGCGCAAATACCGAATAAACG (SEQ ID NO: 8)
	探针	CCACGGACGGTCGGTAGATCG/3C6/ (SEQ ID NO: 9)
ACP1	引物	GCGCGTTGTTTCGTTTCG (SEQ ID NO: 10)
	引物	CGTCACCTACCGCAAATACG (SEQ ID NO: 11)
	探针	CCACGGACGGCGGATAAGGAG/3C6/ (SEQ ID NO: 12)
BDH1	引物	AGTACGTAAGTAGAGCGCG (SEQ ID NO: 13)
	引物	CTAAAATTAACACGCCGCCGT (SEQ ID NO: 14)

[0229]

	探针	CCACGGACGGAGAACGTTTCGA/3C6/ (SEQ ID NO: 15)
EMX1	引物	GGCGTCGCGTTTTTTAGAGAA (SEQ ID NO: 16)
	引物	TTCCTTTTCGTTTCGTATAAAATTTCGTT (SEQ ID NO: 17)
	探针	CCACGGACGATCGGGTTTTAG/3C6/ (SEQ ID NO: 18)
ZF_RASSF1	引物	TGCGTATGGTGGGCG AG (SEQ ID NO: 19)
	引物	CCTAATTTACACGTCAACCAATCGAA (SEQ ID NO: 20)
	探针	CCACGGACGGCGCGTTCGTTT/3C6/ (SEQ ID NO: 21)
BTACT	引物	TTTGTTTTTTTGATTAGGTGTTTAAGA (SEQ ID NO: 22)
	引物	CACCAACCTCATAACCTTATC (SEQ ID NO: 23)
	探针	GACGCGGAGATAGTGTGTGG /3C6/ (SEQ ID NO: 24)
AK055957 (chr12.133)	引物	GCGTTTTAGTTAGATAGGGCGG (SEQ ID NO: 25)
	引物	GAAAACCCCTTCCCCGAAAC (SEQ ID NO: 26)
	探针	CGCCGAGGCGCACGCCTAAA/3C6/ (SEQ ID NO: 27)
HOXA1	引物	AGTCGTTTTTTTAGGTAGTTTAGGCG (SEQ ID NO: 28)
	引物	CGACCTTTACAATCGCCGC (SEQ ID NO: 29)
	探针	CGCCGAGGGGCGGTAGTTGT/3C6/ (SEQ ID NO: 30)
SPINT2	引物	GGGAGCGGTCGCGTAG (SEQ ID NO: 31)
	引物	GCACCTAACTAAACAAAACGAACTAAAC (SEQ ID NO: 32)
	探针	CGCCGAGGCGCAAACGCAA/3C6/ (SEQ ID NO: 33)
DBNL	引物	AGGTGGCGCGTATTACG (SEQ ID NO: 34)
	引物	CCTACTAAACGCGCTCAACC (SEQ ID NO: 35)
	探针	CGCCGAGGCGCTCGATTCCC/3C6/ (SEQ ID NO: 36)
EFNB2	引物	TTCGATATTGGGTGTCGCG (SEQ ID NO: 37)
	引物	CGCGAAAACCAAAAACGAAAC (SEQ ID NO: 38)
	探针	CGCCGAGG GAGGCGGGGTTC/3C6/ (SEQ ID NO: 39)
LRRC4	引物	GCGTTAATTTTCGCGAGGTA (SEQ ID NO: 40)
	引物	ACAATACTCTTATATATTAACGCCGCTC (SEQ ID NO: 41)
	探针	CGCCGAGGAGGCGACGGAGG/3C6/ (SEQ ID NO: 42)

[0230] 表3.

[0231]

基因注释	曲线下面积	倍数变化 (相较于正常 肝脏)	倍数变化 (相较于正常血沉 棕黄层)
DAB2IP	0.94	55.04	403.81
TSPYL5	0.94	159.16	648.30
AK055957 (chr12.133)	0.94	166.22	483.57
EMX1	0.90	578.89	154.47
X57SPINT2	0.89	15.73	83.12
X61ACPI	0.89	378.07	17.96
TBX15.fit	0.86	10.02	不适用
X44SCRN1	0.85	5.68	12.94
TBX15	0.85	8.57	41.50
X53HDGFRP3	0.83	2.85	74.06
OPLAH	0.83	2.16	900.80

[0232]

X49ST8SIA6	0.82	5405378.50	90.32
X52HDGFRP3	0.82	285.68	42.86
CLEC11A	0.82	30.73	51.09
X63DBNL	0.81	3.04	32.20
X65EGR2.fit	0.80	9.83	不适用
HOXA1	0.80	9.13	2746.00
X50OVOL1	0.79	17.01	66.53
X65EGR2	0.78	23.71	12.57
X71COTL1	0.77	2407.68	10613.37
X72RASSF2	0.77	4.35	6960.58
X47TDH	0.76	8.55	13.52
X74GALNT12	0.75	16.31	42.20
X45PRKAG2	0.72	20.02	923.17
ST8SIA1	0.71	1.66	56.25
LRRC4	0.71	1.47	433.42
C13orf18	0.68	1.99	27.00
X39ECE1	0.65	3.36	163.46
X67chr4.56915281	0.64	6.39	102.50

[0233] 此类实验另外鉴定出89个肝脏上皮DMR,这89个肝脏上皮DMR在肝脏(癌症和正常)中为甲基化的,但在正常白细胞DNA样品中不为甲基化的。

[0234] 表4提供在肝脏(癌症和正常)中为甲基化的但在正常白细胞DNA样品中不为甲基化的肝脏上皮DMR的信息。

[0235] 表4.

[0236]

DMR 编号	基因注释	DMR 染色体编号以及起始终止位置
312	ABHD8	chr19:17403265-17403457
313	ABTB1	chr3:127391081-127391163
314	ADAM8	chr10:135090174-135090246
315	AGAP3	chr7:150812289-150812467
316	AGAP3	chr7:150812127-150812178
317	AMIGO3	chr3:49757070-49757161
318	ANXA2	chr15:60690852-60690949
319	APBB2	chr4:40859188-40859260
320	ATP2B4	chr1:203598701-203598782
321	B3GALT4	chr6:33245156-33245191
322	B3GALT4	chr6:33244921-33245067
323	BDH1	chr3:197282831-197282922
324	BDH1	chr3:197281722-197281827
325	BMP2	chr20:6750738-6750803
326	C17orf64	chr17:58499118-58499182
327	C6orf223	chr6:43970506-43970593
328	CELF6	chr15:72612688-72612773
329	DAP2IP	chr9:124461326-124461415
330	DLEC1	chr3:38080680-38080710

[0237]

331	EPS8L2	chr11:726256-726367
332	ESPNP	chr1:17027786-17027840
333	F12	chr5:176830943-176831003
334	F12	chr5:176830740-176830823
335	FIGNL2	chr12:52214703-52214931
336	FLJ45983	chr10:8097592-8097679
337	GAL3ST2	chr2:242743108-242743318
338	GRIN2D	chr19:48901811-48901852
339	GSTO2	chr10:106028898-106028963
340	IL17C	chr16:88701236-88701393
341	IL4I1	chr19:50393413-50493491
342	IRF4	chr6:393636-393768
343	JARID2	chr6:15244974-15245009
344	KCNQ4	chr1:41249959-41250042
345	KDM2B	chr12:121904281-121904436
346	LFNG	chr7:2559582-2559607
347	LGALS3	chr14:55595740-55595831
348	LIMD2	chr17:61778042-61778108
349	LOC389333	chr5:138728888-138728935
350	LOC389333	chr5:138728233-138728342
351	LRRC4	chr7:127671917-127672169
352	LRRC4	chr7:127672388-127672445
353	LRRFIP1	chr2:238600061-238600124
354	LTBP4	chr19:41119803-41119889
355	LYL1	chr19:13215369-13215437
356	LYL1	chr19:13210345-13210498
357	MACROD1	chr11:63828351-63828427
358	MARVELD1	chr10:99474326-99474382
359	MAX.chr10:22765154-22765214	chr10:22765154-22765214
360	MAX.chr11:518982-519057	chr11:518982-519057
361	MAX.chr14:107253126-107253203	chr14:107253126-107253203
362	MAX.chr20:1784481-1784547	chr20:1784481-1784547
363	MAX.chr6:167763903-167763975	chr6:167763903-167763975
364	MAX.chr8:142215988-142216025	chr8:142215988-142216025
365	MAX.chr8:145104176-145104352	chr8:145104176-145104352
366	MBOAT2	chr2:9144074-9144113
367	MTHFD1L	chr6:151187945-151188021
368	MTHFD2	chr2:74425838-74425898
369	MYPOP	chr19:46405010-46405058
370	NGEF	chr2:2337929980233793066
371	NR2F6	chr19:17346368-17346464
372	NR2F6	chr19:17346575-17346695
373	NTN1	chr17:9143556-9143609
374	NTRK1	chr1:156786617-156786674
375	PDZD7	chr10:102792180-102792249
376	PTK2B	chr8:27183159-27183229
377	PTPRE	chr10:129845681-129845740
378	RASSF1	chr3:50378497-50378540
379	RUNX3	chr1:25256236-25256294
380	S1PR4	chr19:3179884-3179960
381	SBNO2	chr19:1131812-1131869
382	SH3PXD2A	chr10:105453034-105453075

[0238]

383	SHH	chr7:155597905-155597937
384	SLC25A36	chr3:140661249-140661281
385	SOC3	chr17:76355477-76355512
386	SP9	chr2:175202178-175202322
387	SPDYA	chr2:29033684-29033774
388	TFR2	chr7:100230996-100231029
389	TIAM1	chr21:32930248-32930318
390	TIAM1	chr21:32931595-32931681
391	TIAM1	chr21:32931280-32931363
392	UCP2	chr11:73693845-73693912
393	VIM	chr10:17271919-17271971
394	WNT1	chr12:49375039-49375117
395	WNT11	chr11:75917494-75917626
396	ZFYVE28	chr4:2415181-2415265
397	ZMIZ1	chr10:81003086-81003157
398	ZMIZ1	chr10:81002084-81002169
399	ZMIZ1	chr10:81002889-81002992
400	ZNF703	chr8:37554906-37554971

[0239] 研究受试者和样品

[0240] 研究由梅奥诊所学会的审查委员会 (Mayo Clinic Institutional Review Board, Rochester, MN) 批准。新鲜冷冻 (FF) 组织、血浆以及血沉棕黄层样品由 IRB-批准的 患者生物库提供。由专家 GI 病理学家再次审查肿瘤组织切片以确认诊断和估计 增生细胞性。然后, 宏观解剖 (macro-dissect) 切片。将基因组 DNA 使用 QiaAmp 微型试剂盒 (Qiagen, Valencia CA) 纯化并且随后用 AMPure XP 试剂盒 (Beckman Coulter, Brea CA) 再纯化。

[0241] 简化表示重亚硫酸盐测序文库的制备

[0242] 使用先前公布的方法的改良型式制备测序文库。用 10U 的 MspI 将基因组 DNA (300ng) 消化过夜。除非另外指明, 否则用于此步骤和后续步骤的所有酶由 New England Biolabs (NEB) 提供。对片段进行末端修复并且用 5U 的 克列诺片段 (Klenow fragment) (3'-5' exo-) 加 A-尾, 并且连结至含有条形码序列的 TruSeq 衔接子 (Illumina) 过夜并且将胞嘧啶普遍甲基化。使用 SYBR Green qPCR (LightCycler 480-Roche) 来计量连结效率和片段质量。将样品用重亚硫酸盐处理并使用改良的 EpiTect 方案 (Qiagen) 纯化 (两次), 然后进行最终 AMPure XP 清除。使用 qPCR 来测定文库富集的最佳 PCR 循环。使用以下条件来进行富集 PCR: 每 50uL 反应物含有 5uL 的 10X 缓冲液、1.25uL 的 10mM 每种脱氧核苷三磷酸 (dNTP)、5uL 引物混合物 (约 5uM)、15uL 样品、1uL PfuTurbo Cx 热启动酶 (hotstart) 以及 22.75uL 水; 温度和时间分别为 95C-5 分钟; 98C-30 秒; 12 至 16 个循环的 98C-10 秒, 65C-30 秒, 72C-30 秒, 72C-5 分钟以及 4C 保持。通过 PicoGreen 分析 (Molecular Probes) 定量样品, 组合至随机化 4-丛 (plex) 文库中, 并且用生物分析仪 2100 (Agilent) 测试以进行大小验证。在根据经验确定的缓冲液浓度下再进行多轮 AMPure XP 纯化/大小选择以使衔接子二聚体污染最小化并且除去大于 350bp 的插入物。使用 phiX 对照标准物 (Illumina) 和衔接子特异性引物通过 qPCR 实现最终文库评估。

[0243] 大规模平行测序和生物信息学

[0244] 根据随机化泳道分配将样品加载至流动池上, 保留额外的通道用于内部分析对照。由下一代测序中心 (Next Generation Sequencing Core) 在梅奥诊所医学基因组研究中心 (Mayo Clinic Medical Genome Facility) 在 Illumina HiSeq 2000 上进行测序。101

次循环的读段是单向的。每个流动池泳道产生1亿-1.2亿个读段,这对于所比对序列的30-50倍测序深度的中值覆盖度为足够的。标准Illumina流水线软件访问碱基并以fastq格式产生读段。使用SAAP-RRBS(简化表示重亚硫酸盐测序的简单化分析和注释流水线)进行序列读段评估和清除、与参考基因组的比对、甲基化状况提取以及CpG报告和注释。排除覆盖度低(≤ 10)的CpG。三级分析由以下步骤组成:除去不提供信息的或样品覆盖度低的CpG,以及鉴定具有低背景的甲基化CpG区域以及滑动的100bp窗内的致密簇。读段深度准则是基于检测病例与对照之间的甲基化%的10%差异所需的统计功效。基于读段计数,通过每个DMR的甲基化百分比的罗吉斯回归(logistic regression)来确定统计显著性。为解释个别受试者间变化的读段深度,使用过度离散的罗吉斯回归模型,其中离散参数是使用拟合模型的残差的皮尔森卡方统计(Pearson Chi-square statistic)来估计。如果对照组中的甲基化% $\leq 1\%$ 并且癌症中的甲基化% $\geq 10\%$,则进一步考虑根据显著性水平排序的DMR。在大多数器官位点,这产生数百个潜在候选物。所利用的额外的过滤器为接受者操作特征曲线下面积、信号与背景甲基化%比率(倍数变化)以及样品间DMR中的CpG的共甲基化呈阳性(以及其在对照中的缺乏)。

[0245] 技术和生物组织验证

[0246] 对来自肝脏发现数据集的如通过上文所列的准则所确定最有前途的DMR中的30个展开甲基化特异性PCR(MSP)标记物分析。通过软件(Methprimer-University of California, San Francisco CA; MSP Primer-Johns Hopkins University, Baltimore, MD)或人工设计引物。对分析进行严格测试并且通过对重亚硫酸盐转化(甲基化和未甲基化基因组DNA)、未转化以及非模板对照进行SYBR Green qPCR加以优化。将与阴性对照交叉反应的分析重新设计或弃去。此外,进行熔解曲线分析以确保发生特异性扩增。对于技术验证阶段,通过qMSP对用于RRBS发现的相同样品再次进行测试。将设计成甲基化设盲的 β -肌动蛋白分析用作表示总DNA拷贝的分母。通过罗吉斯回归对数据进行分析并且将AUC和信号与背景结果与发现值相比较。略少于一半的标记物表现不佳并且将其去除。通过对扩展的104个独立组织样品的集合进行qMSP来测试其余标记物($N=16$)。此外,实验包括11种如下甲基化癌症标记物,这些甲基化癌症标记物是在对其他GI癌症(结肠、食管、胰腺、胆管)的早期测序研究中鉴定并验证的并且为强的多器官癌症标记物。结果度量为AUC和倍数变化比率(表3)。图4和图5中分别描绘了所分析的标记物的箱线图和互补矩阵。

[0247] 跨器官验证

[0248] 为评估最佳甲基化标记物在肝脏外部表现如何,实验使用在HCC样品以及早期测序的其他主要GI癌症(结肠、胰腺、食管以及胃)中验证DMR的测序读段构建了比较CpG甲基化%矩阵。选择最后一组标记物基于以下各项在血浆中进行测试:1)生物组织验证阶段中的总体表现,以及2)在其他癌症中标记物的位点特异性特征。为最好地检测血液中的HCC,鉴于非HCC DNA过量,选择将展现普遍与肝脏特异性癌症信号两者的稳健的12标记物组。所述标记物中有10种来自组织验证;在不经后续组织验证的情况下由RRBS数据直接设计展现非凡肝脏位点特异性的2种额外标记物EFNB2和BDH1并加以利用。

[0249] 血浆验证

[0250] 通过在Exact Sciences开发的自动二氧化硅珠粒法从2mL级分提取血浆DNA。

[0251]

1	2ml 的 Te 缓冲液(1mM Tris 0.1mM EDTA)
2	100μl 的含 120 cp/ul Zebrafish DNA 的 0.4 ng/μl Fish DNA 稀释剂
3	7ml 的血浆溶解缓冲液(4.3M GTC 10% IGEPAL)
4	2ml 的血浆
5	在 55℃下孵育 1 小时
6	添加 200μl 结合珠粒
7	添加 2.8ml 的 100%异丙醇
8	在 30℃下孵育 30 分钟。
9	将珠粒磁化并除去上清液
10	添加 750μl 3M GuHCl 56.8% EtOH 以再悬浮结合珠粒
11	在 400 RPM 下振荡 2 分钟。
12	结合珠粒并将上清液吸至废弃物中
13	1000μl 洗涤液 1 (80%ETOH), 在 30℃下孵育 3 分钟, 结合珠粒并将上清液吸至废弃物中
14	500μl 洗涤液 1 (80%ETOH), 在 30℃下孵育 3 分钟, 结合珠粒并将上清液吸至废弃物中
15	250μl 洗涤液 1 (80%ETOH), 在 30℃下孵育 3 分钟, 结合珠粒并将上清液吸至废弃物中
16	250μl 洗涤液 1 (80%ETOH), 在 30℃下孵育 3 分钟, 结合珠粒并将上清液吸至废弃物中
17	在振荡的同时在 70℃下干燥 15 分钟。
18	添加 125ul 的 Te 缓冲液(1mM Tris 0.1mM EDTA), 在振荡的同时 65℃孵育 25 分钟

[0252]

19	结合珠粒并将含有 DNA 的上清液转移至清洁的管子中
20	储存在-20℃下直至使用。

[0253]

然后,使用下文概述的专有方法将DNA用重亚硫酸盐转化并纯化。

[0254]

1	5ul 0.36%BSA
2	70ul的样品
3	5ul 1.6N NaOH
4	在42℃下孵育(变性)20分钟
5	冷却8分钟
6	添加120ul的重亚硫酸铵
7	在65℃下孵育(转化)75分钟(振荡3分钟)
8	添加750ul的7M GuHCl
9	添加50ul结合珠粒

10	在30℃下在振荡的同时孵育30分钟
11	结合珠粒
12	将上清液吸至废弃物中
13	添加1000ul 80%ETOH
14	在30℃下在振荡的同时孵育3分钟
15	结合珠粒
16	将上清液吸至废弃物中
17	分配200ul脱磺酸溶液
18	在30℃下在振荡的同时孵育7分钟
19	结合珠粒
20	将上清液吸至废弃物中
21	添加250ul 80%ETOH
22	在30℃下在振荡的同时孵育3分钟
23	结合珠粒
24	将上清液吸至废弃物中
25	在70℃下在振荡的同时将珠粒干燥15分钟
26	添加80ul的Te缓冲液(1mM Tris 0.1mM EDTA)
27	在振荡的同时65℃孵育25分钟
28	结合珠粒并将含有DNA的上清液转移至清洁的管子中
29	储存在-20℃下直至使用。

[0255] 在实时PCR仪器 (Roche LC480) 上以QuARTs格式运行样品 (参见美国专利号8,361,720), 使用由DMR序列产生的引物和探针 (参见表2)、GoTaq DNA聚合酶 (Promega)、裂解酶II (Hologic) 以及含有FAM、HEX以及Quasar 670染料的荧光共振能量转移报告盒 (FRET) (Biosearch Technologies)。

[0256] 图6提供通过QuARTs (定量等位基因特异性实时靶标和信号放大) 分析检测甲基化DNA印迹时所用的FRET盒的寡核苷酸序列。每个FRET序列包括荧光团和淬灭剂, 它们可一起复用到3个单独的分析中。

[0257] 从Genscript获得含有所关注的标记物序列的质粒并且在1XQuARTs试剂中稀释至每15ul反应物1个拷贝的标称浓度。将反应混合物分配至384个孔的每个孔中, 在LightCycler上循环进行45个循环, 并且收集数据。给予各孔含有或不含样品的称号。将泊松随机变量 (Poisson random variable) 设定为1并且通过尝试错误法输入平均成功率值并且用于计算所述值的累积概率。当累积概率等于具有信号的孔的百分比时, 就找到了正确平均成功率, 在此情况下为拷贝数。将这些质粒稀释并用作分析标准物。

[0258] 通过首先形成用引物对多达12个靶标进行经历11个扩增循环的预扩增的样品板来进行QuARTs-X (参见美国临时专利号62/249,097)。然后将此产物1:9稀释并用作在三重反应中仅含有三个靶标的后续QuARTs反应的模板。用于计算链计数的标准物不进行预扩增。通过对样品进行预扩增但不对标准物进行预扩增, 分析的灵敏度增加。

[0259] 通过回归分割 (rPart) 对结果进行分析。使用罗吉斯回归将多个甲基化标记物组合至单一危险得分中为标准技术。然而, 难以发现罗吉斯模型内标记物之间的高阶相互作用

用和/或将其模型化。当存在此类效应时,这限制了我们的标记物组的预测能力。回归分割树(rpart)为能够以使得标记物组的预测准确性最大化的方式发现标记物之间的高阶相互作用的决策树方法。使用rPart,最优的3标记物组合(EMX1、BDH1、LRRC4)的HCC血液样品的灵敏度和特异性分别为97%和95%(图7A)将三种替代标记物组合模型化:

[0260] 替代#1(EMX1、DAB2IP、TSPYL5):特异性=100%,灵敏度=90%(图7B)

[0261] 替代#2(EMX1、HOXA1、ACPI):特异性=88%,灵敏度=100%(图7C)

[0262] 替代#3(EMX1、EFNB2、SPINT2):特异性=100%,灵敏度=90%(图7D)

[0263] 在血浆中表现最佳的单一标记物EMX1在100%特异性下具有0.89的AUC以及77%的灵敏度。EMX1的信号随着阶段增加显示更高的 β 肌动蛋白归一化信号(图8)。

[0264] 实施例II.

[0265] 在实时PCR仪器(Roche LC480)上以QuARTs格式运行组织样品(75个HCC、20个肝硬化以及30个正常)(参见美国专利号8,361,720),使用由DMR序列产生的引物和探针(参见表5)、GoTaq DNA聚合酶(Promega)、裂解酶2.0(Hologic)以及含有FAM、HEX以及Quasar 670染料的荧光共振能量转移报告盒(FRET)(Biosearch Technologies)。表6示出了每种标记物以100%灵敏度辨别HCC与肝硬化和正常样品的能力。

[0266] 表5.

[0267]

基因	正向QuARTS引物(5'-3')	反向QuARTS引物(5'-3')	探针
TSPYL5	TTTGTTCGGTTTTTGGCG (SEQ ID NO: 43)	ACCATAAACGACCGAAATCGA (SEQ ID NO:44)	CCACGGACG GCGGGAGGAT TT/3C6/ (SEQ ID NO: 45)
DAB2IP	CGTTCGTTACGTCGTTTTCGT (SEQ ID NO: 46)	GATCGACGCGACTCGAC (SEQ ID NO: 47)	CCACGGACG CTCGACGTCGC C/3C6/ (SEQ ID NO: 48)
Chr12_133_V2	GCGTTTTAGTTAGATAGGCGG (SEQ ID NO: 49)	GAAAACCCCTTCCCGAAAC (SEQ ID NO: 50)	CGCCGAGG CGCACGCCTAA/3C6/ (SEQ ID NO: 51)
CLEC11A	GCGGGAGTTTGGCGTAG (SEQ ID NO: 52)	CGCGCAAATACCGAATAACG (SEQ ID NO: 53)	CCACGGACG-GTCGGTAGATCG/3C6/ (SEQ ID NO: 54)
SPINT2	GGGAGCGGTCGCGTAG (SEQ ID NO: 55)	GCACCTAACTAAACAAAA CGAACTAAAC (SEQ ID NO: 56)	CGCCGAGG CGCAAACGCAAA/3C6/ (SEQ ID NO: 57)
ACPI	GCGCGTTGTTTCGTTTCG (SEQ ID NO: 58)	CGTCACCTACCGCAAATACG (SEQ ID NO: 59)	CCACGGACG GCGGATAAGGAG/3C6/ (SEQ ID NO: 60)
BDH1	AGTACGTAAGTAGAG	CTAAAATTAACCTACGCCG	CCACGGACG

[0268]

	CGCG (SEQ ID NO: 61)	CCGT (SEQ ID NO: 62)	GAGAACGTC GA/3C6/ (SEQ ID NO: 63)
EFNB2	CGTTCGTTCTGTTATTTT TTTTCGA (SEQ ID NO: 64)	GCCCGCGAAAACCAAAA AC (SEQ ID NO: 65)	CGCCGAGG CGAAACTCAC CT/3C6/ (SEQ ID NO: 66)
CCNJ_3707	GCGTTTTTTTTTAGCG GGGTTA (SEQ ID NO: 67)	CCGAAACTAAAATTCTCC CGC (SEQ ID NO: 68)	CGCCGAGG ATGAGCGTGT A\3C6\ (SEQ ID NO: 69)
CCNJ_3124	CGTTTTTCGTTTGGGT ACG (SEQ ID NO: 70)	CCAACCCAAACCACGCC (SEQ ID NO: 71)	CCACGGACG CGCGCCGTAC GA\3C6\ (SEQ ID NO: 72)
PFKP	GGAGGTTGGCGGGGA G (SEQ ID NO: 73)	CATATACTATCGCCTTCC GACTC (SEQ ID NO: 74)	CGCCGAGG CGTAACAAAA AC\3C6\ (SEQ ID NO: 75)
SCRNI	TCGTTTTAGGTGAGTC GCG (SEQ ID NO: 76)	AAATAAAACCGCCGAAA AAACAAC (SEQ ID NO: 77)	CCACGGACG CGTACAAAC GC\3C6\ (SEQ ID NO: 78)
ECE1	GGAGGGGTTTCGTTCTG T (SEQ ID NO: 79)	CTACTATCGACGCTAAAA ATAAACGAAC (SEQ ID NO: 80)	CGCCGAGG CGCGACCTAA AA/3C6/ (SEQ ID NO: 81)
DBNL	AGGTGGCGCGTATTAC G (SEQ ID NO: 82)	CCTACTAAACGCGCTCAA CC (SEQ ID NO: 83)	CGCCGAGG CGCTCGATTCC C/3C6/ (SEQ ID NO: 84)
LRRC4	GCGTTAATTTTCGCGAG GTA (SEQ ID NO: 85)	ACAATACTCTTATATATT AACGCCGCTC (SEQ ID NO: 86)	CGCCGAGG AGGCGACGGA GG/3C6/ (SEQ ID NO: 87)
HOXA1	AGTCGTTTTTTTAGGT AGTTTAGGCG (SEQ ID NO: 88)	CGACCTTTACAATCGCCG C (SEQ ID NO: 89)	CGCCGAGG GGCGGTAGTT GT/3C6/ (SEQ ID NO: 90)
EMX1	GGCGTCGCGTTTTTTTA GAGAA (SEQ ID NO: 91)	TTCCTTTTCGTTCTGTATAA AATTCGT (SEQ ID NO: 92)	CCACGGACG ATCGGGTTTTA G/3C6/ (SEQ ID NO: 93) Or CGCCGAGG ATCGGGTTTTA G/3C6/ (SEQ ID NO: 94)

[0269] 表6.

[0270]

基因	在 100%特异性下辨别 HCC 与肝硬化或正常样品的灵敏度
TSPYL5	79%
DAB2IP	91%
Chr12_133_V2	91%
CLEC11A	69%
SPINT2	68%
ACP1	84%
BDH1	8%
EFNB2	45%
CCNJ_3707	53%
CCNJ_3124	49%
PFKP	57%
SCRN1	51%
ECE1	49%
EMX1	67%
LRRC4	20%
HOXA1	52%

[0271] 实施例III.

[0272] 此实施例的主要目的是确定预测肝细胞癌 (HCC) 的标记物组。将来自244个受试者 (95个肝细胞癌和149个对照) 的血浆调节至2mL并提取。149个对照由51个肝硬化患者和98个正常患者组成。

[0273] 图9展示了此分析中所考虑的甲基化标记物中的每一者的相对重要性。整个标记物组的灵敏度和特异性的交叉验证估计值分别为75%和96%。

[0274] 在对照相较于正常患者的88.6%的特异性下,以下标记物组 (Chr12.133、CLEC11A、EMX1、HOXA1、CCNJ_3707) 对于HCC产生85.3%的灵敏度(表7)。

[0275] 表7.

[0276]

	全部患者	预测		灵敏度
		阴性	阳性	
HCC	95	14	81	85.3%
对照	149	132	17	

[0277] 相同组 (Chr12.133、CLEC11A、EMX1、HOXA1、CCNJ_3707) 对于正常和肝硬化 (对照组) 患者 (表8) 产生以下特异性减低。

[0278] 表8.

	患者组	全部患者	预测		特异性
			阴性	阳性	
[0279]	肝硬化	51	41	10	80.4%
	正常	98	91	7	92.9%
	总计	149	132	17	88.6%

[0280] 用于从4mL血浆样品分离DNA的示例性程序将如下进行：

[0281] • 向2mL血浆样品中添加300μL的蛋白酶K (20mg/mL) 并混合。当样品为小于2mL的血浆时，通过添加10mM Tris-HCl、0.1mM EDTA溶液调节至2mL

[0282] • 将6mL的血浆溶解缓冲液1添加至血浆中并在室温下混合

[0283] 血浆溶解缓冲液为：

[0284] -4.3M硫氰酸胍

[0285] -10% IGEPAL CA-630 (辛基苯氧基聚(亚乙基氧基)乙醇, 分枝)

[0286] (5.3g IGEPAL CA-630与45mL 4.8M硫氰酸胍的组合)

[0287] • 添加200μL磁性二氧化硅结合珠粒[16μg珠粒/μL]并且再次混合。

[0288] • 将7mL的溶解缓冲液2添加至管中。

[0289] 通过混合60%溶解缓冲液1与40%异丙醇来制备血浆溶解缓冲液2。

[0290] • 将样品与溶解缓冲液2混合60分钟

[0291] • 将管放在磁体上并且让珠粒收集10分钟。吸出并弃去上清液。

[0292] • 将1000μL洗涤缓冲液 (10mM Tris HCl, 80%EtOH) 添加至珠粒中，并且在振荡的同时在30℃下孵育3分钟。

[0293] • 将管放在磁体上并且让珠粒进行收集。吸出并弃去上清液。

[0294] • 将500μL洗涤缓冲液添加至珠粒中并且在振荡的同时在30℃下孵育3分钟。

[0295] • 将管放在磁体上并且让珠粒进行收集。吸出并弃去上清液。

[0296] • 添加250μL洗涤缓冲液并且在振荡的同时在30℃下孵育3分钟。

[0297] • 将管放在磁体上并且让珠粒进行收集。吸出并弃去剩余的缓冲液。

[0298] • 添加250μL洗涤缓冲液并且在振荡的同时在30℃下孵育3分钟。

[0299] • 将管放在磁体上并且让珠粒进行收集。吸出并弃去剩余的缓冲液。

[0300] • 在振荡的同时，将珠粒在70℃下干燥15分钟。

[0301] • 将125μL洗脱缓冲液 (10mM Tris HCl, pH 8.0, 0.1mM EDTA) 添加至珠粒中并且在振荡的同时在65℃下孵育25分钟。

[0302] • 将管放在磁体上并且让珠粒收集10分钟。

[0303] • 吸出含有DNA的上清液并转移至新的容器或管中。

[0304] 重亚硫酸盐转化

[0305] I. 使用亚硫酸氢铵使DNA磺化

[0306] 1. 在每个管中，将64μL DNA、7μL 1N NaOH以及含有0.2mg/mL BSA和0.25mg/mL鱼DNA的9μL载体溶液组合。

[0307] 2. 在42℃下孵育20分钟。

[0308] 3. 添加120μL的45%亚硫酸氢铵并且在66°下孵育75分钟。

- [0309] 4.在4℃下孵育10分钟。
- [0310] II.使用磁性珠粒进行脱磺酸
- [0311] 材料
- [0312] 磁性珠粒(Promega MagneSil顺磁性粒子,Promega目录编号AS1050,16μg/μL)。
- [0313] 结合缓冲液:6.5-7M盐酸胍。
- [0314] 转化后洗涤缓冲液:80%乙醇加上10mM Tris HCl (pH 8.0)。
- [0315] 脱磺酸缓冲液:对于脱磺酸缓冲液,选择70%异丙醇、0.1N NaOH。
- [0316] 使用任何适当的装置或技术混合样品以在大体上如下文所描述的温度和混合速度下混合或孵育样品。举例来说,可使用恒温混匀仪(Thermomixer)(Eppendorf)来混合或孵育样品。示例性脱磺酸作用如下:
- [0317] 1.将珠粒储备物通过涡流瓶充分混合1分钟。
- [0318] 2.将50μL的珠粒等分至2.0mL管(例如来自USA Scientific)中。
- [0319] 3.将750μL的结合缓冲液添加至珠粒中。
- [0320] 4.添加150μL来自步骤1的磺化DNA。
- [0321] 5.混合(例如1000RPM,在30℃下30分钟)。
- [0322] 6.将管放在磁体架上并且留在原地5分钟。使管在架子上,移出并弃去上清液。
- [0323] 7.添加1,000μL的洗涤缓冲液。混合(例如1000RPM,在30℃下3分钟)。
- [0324] 8.将管放在磁体架上并且留在原地5分钟。使管在架子上,移出并弃去上清液。
- [0325] 9.添加250μL的洗涤缓冲液。混合(例如1000RPM,在30℃下3分钟)。
- [0326] 10.将管放在磁性支架上;在1分钟后移出并弃去上清液。
- [0327] 11.添加200μL的脱磺酸缓冲液。混合(例如1000RPM,在30℃下5分钟)。
- [0328] 12.将管放在磁性支架上;在1分钟后移出并弃去上清液。
- [0329] 13.添加250μL的洗涤缓冲液。混合(例如1000RPM,在30℃下3分钟)。
- [0330] 14.将管放在磁性支架上;在1分钟后移出并弃去上清液。
- [0331] 15.将250μL的洗涤缓冲液添加至管中。混合(例如1000RPM,在30℃下3分钟)。
- [0332] 16.将管放在磁性支架上;在1分钟后移出并弃去上清液。
- [0333] 17.打开盖子,将所有管在30℃下孵育15分钟。
- [0334] 18.将管从磁性支架移去并且将70μL的洗脱缓冲液直接添加至珠粒中。
- [0335] 19.在存在洗脱缓冲液的情况下孵育珠粒(例如1000RPM,在40℃下45分钟)。
- [0336] 20.将管放在磁性支架上约一分钟;移去并保存上清液。
- [0337] 然后将转化的DNA用于预扩增和/或侧翼核酸内切酶分析。
- [0338] 以上说明书中提到的所有出版物和专利出于所有目的以全文引用的方式并入本文中。在不背离如所描述的技术的范围和精神的情况下,对所描述的所述技术的组合物、方法以及用途的各种修改和变化对于本领域技术人员来说将为显而易见的。虽然已结合特定示例性实施方案描述了所述技术,但应了解如所要求的发明不应过度受限于此类特定实施方案。实际上,对所描述的用于执行本发明的模式的对药理学、生物化学、医学科学或相关领域的技术人员来说显而易见的各种修改意图在以下权利要求书的范围内。

- [0001] 序列表
- [0002] <110> 精密科学发展有限责任公司 (EXACT SCIENCES CORPORATION)
- [0003] 梅约医学教育与研究基金会 (MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH)
- [0004] <120> 检测肝细胞癌
- [0005] <130> PN19005040P
- [0006] <150> US 62/383,165
- [0007] <151> 2016-09-02
- [0008] <160> 104
- [0009] <170> PatentIn 3.5版
- [0010] <210> 1
- [0011] <211> 17
- [0012] <212> DNA
- [0013] <213> 人工序列
- [0014] <220>
- [0015] <223> 合成的
- [0016] <400> 1
- [0017] gcgcgggagg attttcg 17
- [0018] <210> 2
- [0019] <211> 18
- [0020] <212> DNA
- [0021] <213> 人工序列
- [0022] <220>
- [0023] <223> 合成的
- [0024] <400> 2
- [0025] ccgccaccat aaacgacc 18
- [0026] <210> 3
- [0027] <211> 21
- [0028] <212> DNA
- [0029] <213> 人工序列
- [0030] <220>
- [0031] <223> 合成的
- [0032] <400> 3
- [0033] ccacggacgc gaaatcgaaa t 21
- [0034] <210> 4
- [0035] <211> 21
- [0036] <212> DNA
- [0037] <213> 人工序列

[0038] <220>
[0039] <223> 合成的
[0040] <400> 4
[0041] cgttcgttac gtcgttttcg t 21
[0042] <210> 5
[0043] <211> 17
[0044] <212> DNA
[0045] <213> 人工序列
[0046] <220>
[0047] <223> 合成的
[0048] <400> 5
[0049] gatcgacgcg actcgac 17
[0050] <210> 6
[0051] <211> 21
[0052] <212> DNA
[0053] <213> 人工序列
[0054] <220>
[0055] <223> 合成的
[0056] <400> 6
[0057] ccacggacgc tcgacgtcgc c 21
[0058] <210> 7
[0059] <211> 17
[0060] <212> DNA
[0061] <213> 人工序列
[0062] <220>
[0063] <223> 合成的
[0064] <400> 7
[0065] gcgggagttt ggcgtag 17
[0066] <210> 8
[0067] <211> 21
[0068] <212> DNA
[0069] <213> 人工序列
[0070] <220>
[0071] <223> 合成的
[0072] <400> 8
[0073] cgcgcaaata ccgaataaac g 21
[0074] <210> 9
[0075] <211> 21
[0076] <212> DNA

[0077] <213> 人工序列
[0078] <220>
[0079] <223> 合成的
[0080] <400> 9
[0081] ccacggacgg tcggtagatc g 21
[0082] <210> 10
[0083] <211> 18
[0084] <212> DNA
[0085] <213> 人工序列
[0086] <220>
[0087] <223> 合成的
[0088] <400> 10
[0089] gcgcgttggtt tcgttttcg 18
[0090] <210> 11
[0091] <211> 20
[0092] <212> DNA
[0093] <213> 人工序列
[0094] <220>
[0095] <223> 合成的
[0096] <400> 11
[0097] cgtcacctac cgcaaatacg 20
[0098] <210> 12
[0099] <211> 21
[0100] <212> DNA
[0101] <213> 人工序列
[0102] <220>
[0103] <223> 合成的
[0104] <400> 12
[0105] ccacggacgg cggataagga g 21
[0106] <210> 13
[0107] <211> 19
[0108] <212> DNA
[0109] <213> 人工序列
[0110] <220>
[0111] <223> 合成的
[0112] <400> 13
[0113] agtacgtaag tagagcgcg 19
[0114] <210> 14
[0115] <211> 22

[0116] <212> DNA
[0117] <213> 人工序列
[0118] <220>
[0119] <223> 合成的
[0120] <400> 14
[0121] ctaaaattaa ctacgccgcc gt 22
[0122] <210> 15
[0123] <211> 21
[0124] <212> DNA
[0125] <213> 人工序列
[0126] <220>
[0127] <223> 合成的
[0128] <400> 15
[0129] ccacggacgg agaacgttcg a 21
[0130] <210> 16
[0131] <211> 21
[0132] <212> DNA
[0133] <213> 人工序列
[0134] <220>
[0135] <223> 合成的
[0136] <400> 16
[0137] ggcgtcgcgt tttttagaga a 21
[0138] <210> 17
[0139] <211> 28
[0140] <212> DNA
[0141] <213> 人工序列
[0142] <220>
[0143] <223> 合成的
[0144] <400> 17
[0145] ttccttttcg ttcgtataaa atttcggt 28
[0146] <210> 18
[0147] <211> 21
[0148] <212> DNA
[0149] <213> 人工序列
[0150] <220>
[0151] <223> 合成的
[0152] <400> 18
[0153] ccacggacga tcgggtttta g 21
[0154] <210> 19

[0155] <211> 17
[0156] <212> DNA
[0157] <213> 人工序列
[0158] <220>
[0159] <223> 合成的
[0160] <400> 19
[0161] tgcgtatggt gggcgag 17
[0162] <210> 20
[0163] <211> 26
[0164] <212> DNA
[0165] <213> 人工序列
[0166] <220>
[0167] <223> 合成的
[0168] <400> 20
[0169] cctaatttac acgtcaacca atcgaa 26
[0170] <210> 21
[0171] <211> 21
[0172] <212> DNA
[0173] <213> 人工序列
[0174] <220>
[0175] <223> 合成的
[0176] <400> 21
[0177] ccacggacgg cgcgtgcgtt t 21
[0178] <210> 22
[0179] <211> 27
[0180] <212> DNA
[0181] <213> 人工序列
[0182] <220>
[0183] <223> 合成的
[0184] <400> 22
[0185] tttgtttttt tgattaggtg tttaaga 27
[0186] <210> 23
[0187] <211> 21
[0188] <212> DNA
[0189] <213> 人工序列
[0190] <220>
[0191] <223> 合成的
[0192] <400> 23
[0193] caccaacctc ataaccttat c 21

[0194] <210> 24
[0195] <211> 21
[0196] <212> DNA
[0197] <213> 人工序列
[0198] <220>
[0199] <223> 合成的
[0200] <400> 24
[0201] gacgcggaga tagtgttgtg g 21
[0202] <210> 25
[0203] <211> 22
[0204] <212> DNA
[0205] <213> 人工序列
[0206] <220>
[0207] <223> 合成的
[0208] <400> 25
[0209] gcgttttagt tagatagggc gg 22
[0210] <210> 26
[0211] <211> 20
[0212] <212> DNA
[0213] <213> 人工序列
[0214] <220>
[0215] <223> 合成的
[0216] <400> 26
[0217] gaaaaccct tccccgaaac 20
[0218] <210> 27
[0219] <211> 20
[0220] <212> DNA
[0221] <213> 人工序列
[0222] <220>
[0223] <223> 合成的
[0224] <400> 27
[0225] cgccgaggcg cacgcctaaa 20
[0226] <210> 28
[0227] <211> 26
[0228] <212> DNA
[0229] <213> 人工序列
[0230] <220>
[0231] <223> 合成的
[0232] <400> 28

[0233] agtcgttttt ttaggtagtt taggcg 26
[0234] <210> 29
[0235] <211> 19
[0236] <212> DNA
[0237] <213> 人工序列
[0238] <220>
[0239] <223> 合成的
[0240] <400> 29
[0241] cgacctttac aatcgccgc 19
[0242] <210> 30
[0243] <211> 20
[0244] <212> DNA
[0245] <213> 人工序列
[0246] <220>
[0247] <223> 合成的
[0248] <400> 30
[0249] cgccgagggg cggtagttgt 20
[0250] <210> 31
[0251] <211> 16
[0252] <212> DNA
[0253] <213> 人工序列
[0254] <220>
[0255] <223> 合成的
[0256] <400> 31
[0257] gggagcggtc gcgtag 16
[0258] <210> 32
[0259] <211> 28
[0260] <212> DNA
[0261] <213> 人工序列
[0262] <220>
[0263] <223> 合成的
[0264] <400> 32
[0265] gcaccttaact aaacaaaacg aactaac 28
[0266] <210> 33
[0267] <211> 20
[0268] <212> DNA
[0269] <213> 人工序列
[0270] <220>
[0271] <223> 合成的

[0272] <400> 33
[0273] cgccgaggcg caaacgcaaa 20
[0274] <210> 34
[0275] <211> 17
[0276] <212> DNA
[0277] <213> 人工序列
[0278] <220>
[0279] <223> 合成的
[0280] <400> 34
[0281] aggtggcgcg tattacg 17
[0282] <210> 35
[0283] <211> 20
[0284] <212> DNA
[0285] <213> 人工序列
[0286] <220>
[0287] <223> 合成的
[0288] <400> 35
[0289] cctactaaac gcgctcaacc 20
[0290] <210> 36
[0291] <211> 20
[0292] <212> DNA
[0293] <213> 人工序列
[0294] <220>
[0295] <223> 合成的
[0296] <400> 36
[0297] cgccgaggcg ctcgattccc 20
[0298] <210> 37
[0299] <211> 19
[0300] <212> DNA
[0301] <213> 人工序列
[0302] <220>
[0303] <223> 合成的
[0304] <400> 37
[0305] ttcgatattg ggtgtcgcg 19
[0306] <210> 38
[0307] <211> 21
[0308] <212> DNA
[0309] <213> 人工序列
[0310] <220>

[0311] <223> 合成的
[0312] <400> 38
[0313] cgcgaaaacc aaaaacgaaa c 21
[0314] <210> 39
[0315] <211> 20
[0316] <212> DNA
[0317] <213> 人工序列
[0318] <220>
[0319] <223> 合成的
[0320] <400> 39
[0321] cgccgaggga ggcggggttc 20
[0322] <210> 40
[0323] <211> 19
[0324] <212> DNA
[0325] <213> 人工序列
[0326] <220>
[0327] <223> 合成的
[0328] <400> 40
[0329] gcgttaattt cgcgaggta 19
[0330] <210> 41
[0331] <211> 28
[0332] <212> DNA
[0333] <213> 人工序列
[0334] <220>
[0335] <223> 合成的
[0336] <400> 41
[0337] acaatactct tatatattaa cgccgctc 28
[0338] <210> 42
[0339] <211> 20
[0340] <212> DNA
[0341] <213> 人工序列
[0342] <220>
[0343] <223> 合成的
[0344] <400> 42
[0345] cgccgaggag gcgacggagg 20
[0346] <210> 43
[0347] <211> 19
[0348] <212> DNA
[0349] <213> 人工序列

[0350] <220>
[0351] <223> 合成的
[0352] <400> 43
[0353] tttgtttcgg tttttggcg 19
[0354] <210> 44
[0355] <211> 21
[0356] <212> DNA
[0357] <213> 人工序列
[0358] <220>
[0359] <223> 合成的
[0360] <400> 44
[0361] accataaacg accgaaatcg a 21
[0362] <210> 45
[0363] <211> 21
[0364] <212> DNA
[0365] <213> 人工序列
[0366] <220>
[0367] <223> 合成的
[0368] <400> 45
[0369] ccacggacgg cgggaggatt t 21
[0370] <210> 46
[0371] <211> 21
[0372] <212> DNA
[0373] <213> 人工序列
[0374] <220>
[0375] <223> 合成的
[0376] <400> 46
[0377] cgttcgttac gtcgttttcg t 21
[0378] <210> 47
[0379] <211> 17
[0380] <212> DNA
[0381] <213> 人工序列
[0382] <220>
[0383] <223> 合成的
[0384] <400> 47
[0385] gatcgacgcg actcgac 17
[0386] <210> 48
[0387] <211> 21
[0388] <212> DNA

- [0389] <213> 人工序列
[0390] <220>
[0391] <223> 合成的
[0392] <400> 48
[0393] ccacggacgc tcgacgtcgc c 21
[0394] <210> 49
[0395] <211> 22
[0396] <212> DNA
[0397] <213> 人工序列
[0398] <220>
[0399] <223> 合成的
[0400] <400> 49
[0401] gcgttttagt tagatagggc gg 22
[0402] <210> 50
[0403] <211> 20
[0404] <212> DNA
[0405] <213> 人工序列
[0406] <220>
[0407] <223> 合成的
[0408] <400> 50
[0409] gaaaaccct tccccgaaac 20
[0410] <210> 51
[0411] <211> 20
[0412] <212> DNA
[0413] <213> 人工序列
[0414] <220>
[0415] <223> 合成的
[0416] <400> 51
[0417] cgccgaggcg cacgcctaaa 20
[0418] <210> 52
[0419] <211> 17
[0420] <212> DNA
[0421] <213> 人工序列
[0422] <220>
[0423] <223> 合成的
[0424] <400> 52
[0425] gcgggagttt ggcgtag 17
[0426] <210> 53
[0427] <211> 21

[0428] <212> DNA
[0429] <213> 人工序列
[0430] <220>
[0431] <223> 合成的
[0432] <400> 53
[0433] cgcgcaaata ccgaataaac g 21
[0434] <210> 54
[0435] <211> 21
[0436] <212> DNA
[0437] <213> 人工序列
[0438] <220>
[0439] <223> 合成的
[0440] <400> 54
[0441] ccacggacgg tcggtagatc g 21
[0442] <210> 55
[0443] <211> 16
[0444] <212> DNA
[0445] <213> 人工序列
[0446] <220>
[0447] <223> 合成的
[0448] <400> 55
[0449] gggagcggtc gcgtag 16
[0450] <210> 56
[0451] <211> 28
[0452] <212> DNA
[0453] <213> 人工序列
[0454] <220>
[0455] <223> 合成的
[0456] <400> 56
[0457] gcacctaact aaacaaaacg aactaac 28
[0458] <210> 57
[0459] <211> 20
[0460] <212> DNA
[0461] <213> 人工序列
[0462] <220>
[0463] <223> 合成的
[0464] <400> 57
[0465] cgccgaggcg caaacgcaa 20
[0466] <210> 58

[0467] <211> 18
[0468] <212> DNA
[0469] <213> 人工序列
[0470] <220>
[0471] <223> 合成的
[0472] <400> 58
[0473] gcgcgttggt tcgtttcg 18
[0474] <210> 59
[0475] <211> 20
[0476] <212> DNA
[0477] <213> 人工序列
[0478] <220>
[0479] <223> 合成的
[0480] <400> 59
[0481] cgtcacctac cgcaaatacg 20
[0482] <210> 60
[0483] <211> 21
[0484] <212> DNA
[0485] <213> 人工序列
[0486] <220>
[0487] <223> 合成的
[0488] <400> 60
[0489] ccacggacgg cggataagga g 21
[0490] <210> 61
[0491] <211> 19
[0492] <212> DNA
[0493] <213> 人工序列
[0494] <220>
[0495] <223> 合成的
[0496] <400> 61
[0497] agtacgtaag tagagcgcg 19
[0498] <210> 62
[0499] <211> 22
[0500] <212> DNA
[0501] <213> 人工序列
[0502] <220>
[0503] <223> 合成的
[0504] <400> 62
[0505] ctaaaattaa ctacgccgcc gt 22

[0506] <210> 63
[0507] <211> 21
[0508] <212> DNA
[0509] <213> 人工序列
[0510] <220>
[0511] <223> 合成的
[0512] <400> 63
[0513] ccacggacgg agaacgttcg a 21
[0514] <210> 64
[0515] <211> 24
[0516] <212> DNA
[0517] <213> 人工序列
[0518] <220>
[0519] <223> 合成的
[0520] <400> 64
[0521] cgttcgttcg ttatTTTTTT tcga 24
[0522] <210> 65
[0523] <211> 19
[0524] <212> DNA
[0525] <213> 人工序列
[0526] <220>
[0527] <223> 合成的
[0528] <400> 65
[0529] gcccgcgaaa accaaaaac 19
[0530] <210> 66
[0531] <211> 20
[0532] <212> DNA
[0533] <213> 人工序列
[0534] <220>
[0535] <223> 合成的
[0536] <400> 66
[0537] cgccgaggcg aaactcacct 20
[0538] <210> 67
[0539] <211> 22
[0540] <212> DNA
[0541] <213> 人工序列
[0542] <220>
[0543] <223> 合成的
[0544] <400> 67

[0545] gcgttttttt ttagcggggt ta 22
[0546] <210> 68
[0547] <211> 21
[0548] <212> DNA
[0549] <213> 人工序列
[0550] <220>
[0551] <223> 合成的
[0552] <400> 68
[0553] ccgaaactaa aattctcccg c 21
[0554] <210> 69
[0555] <211> 20
[0556] <212> DNA
[0557] <213> 人工序列
[0558] <220>
[0559] <223> 合成的
[0560] <400> 69
[0561] cgccgaggat gagcgtgtta 20
[0562] <210> 70
[0563] <211> 19
[0564] <212> DNA
[0565] <213> 人工序列
[0566] <220>
[0567] <223> 合成的
[0568] <400> 70
[0569] cggttttcgt ttgggtacg 19
[0570] <210> 71
[0571] <211> 17
[0572] <212> DNA
[0573] <213> 人工序列
[0574] <220>
[0575] <223> 合成的
[0576] <400> 71
[0577] ccaaccctaaa ccacgcc 17
[0578] <210> 72
[0579] <211> 21
[0580] <212> DNA
[0581] <213> 人工序列
[0582] <220>
[0583] <223> 合成的

[0584] <400> 72
[0585] ccacggacgc gcgccgtacg a 21
[0586] <210> 73
[0587] <211> 16
[0588] <212> DNA
[0589] <213> 人工序列
[0590] <220>
[0591] <223> 合成的
[0592] <400> 73
[0593] ggaggttggc ggggag 16
[0594] <210> 74
[0595] <211> 23
[0596] <212> DNA
[0597] <213> 人工序列
[0598] <220>
[0599] <223> 合成的
[0600] <400> 74
[0601] catatactat cgccttccga ctc 23
[0602] <210> 75
[0603] <211> 20
[0604] <212> DNA
[0605] <213> 人工序列
[0606] <220>
[0607] <223> 合成的
[0608] <400> 75
[0609] cgccgaggcg taacaaaaac 20
[0610] <210> 76
[0611] <211> 19
[0612] <212> DNA
[0613] <213> 人工序列
[0614] <220>
[0615] <223> 合成的
[0616] <400> 76
[0617] tcgttttagg tgagtcgcg 19
[0618] <210> 77
[0619] <211> 24
[0620] <212> DNA
[0621] <213> 人工序列
[0622] <220>

[0623]	<223> 合成的
[0624]	<400> 77
[0625]	aaataaaacc gccgaaaaa caac 24
[0626]	<210> 78
[0627]	<211> 21
[0628]	<212> DNA
[0629]	<213> 人工序列
[0630]	<220>
[0631]	<223> 合成的
[0632]	<400> 78
[0633]	ccacggacgc gctacaaacg c 21
[0634]	<210> 79
[0635]	<211> 17
[0636]	<212> DNA
[0637]	<213> 人工序列
[0638]	<220>
[0639]	<223> 合成的
[0640]	<400> 79
[0641]	ggaggggttt cgttcgt 17
[0642]	<210> 80
[0643]	<211> 29
[0644]	<212> DNA
[0645]	<213> 人工序列
[0646]	<220>
[0647]	<223> 合成的
[0648]	<400> 80
[0649]	ctactatcga cgctaaaaat aaaacgaac 29
[0650]	<210> 81
[0651]	<211> 20
[0652]	<212> DNA
[0653]	<213> 人工序列
[0654]	<220>
[0655]	<223> 合成的
[0656]	<400> 81
[0657]	cgccgaggcg cgacctaaa 20
[0658]	<210> 82
[0659]	<211> 17
[0660]	<212> DNA
[0661]	<213> 人工序列

[0662] <220>
[0663] <223> 合成的
[0664] <400> 82
[0665] aggtggcgcgg tattacg 17
[0666] <210> 83
[0667] <211> 20
[0668] <212> DNA
[0669] <213> 人工序列
[0670] <220>
[0671] <223> 合成的
[0672] <400> 83
[0673] cctactaaac gcgctcaacc 20
[0674] <210> 84
[0675] <211> 20
[0676] <212> DNA
[0677] <213> 人工序列
[0678] <220>
[0679] <223> 合成的
[0680] <400> 84
[0681] cgccgaggcg ctcgattccc 20
[0682] <210> 85
[0683] <211> 19
[0684] <212> DNA
[0685] <213> 人工序列
[0686] <220>
[0687] <223> 合成的
[0688] <400> 85
[0689] gcgttaattt cgcgaggta 19
[0690] <210> 86
[0691] <211> 28
[0692] <212> DNA
[0693] <213> 人工序列
[0694] <220>
[0695] <223> 合成的
[0696] <400> 86
[0697] acaatactct tatatattaa cgccgctc 28
[0698] <210> 87
[0699] <211> 20
[0700] <212> DNA

[0701] <213> 人工序列
[0702] <220>
[0703] <223> 合成的
[0704] <400> 87
[0705] cgccgaggag gcgacggagg 20
[0706] <210> 88
[0707] <211> 26
[0708] <212> DNA
[0709] <213> 人工序列
[0710] <220>
[0711] <223> 合成的
[0712] <400> 88
[0713] agtcgtttttt ttaggtagtt taggcg 26
[0714] <210> 89
[0715] <211> 19
[0716] <212> DNA
[0717] <213> 人工序列
[0718] <220>
[0719] <223> 合成的
[0720] <400> 89
[0721] cgacctttac aatcgccgc 19
[0722] <210> 90
[0723] <211> 20
[0724] <212> DNA
[0725] <213> 人工序列
[0726] <220>
[0727] <223> 合成的
[0728] <400> 90
[0729] cgccgagggg cggtagttgt 20
[0730] <210> 91
[0731] <211> 21
[0732] <212> DNA
[0733] <213> 人工序列
[0734] <220>
[0735] <223> 合成的
[0736] <400> 91
[0737] ggcgtcgcgt tttttagaga a 21
[0738] <210> 92
[0739] <211> 27

[0740] <212> DNA
[0741] <213> 人工序列
[0742] <220>
[0743] <223> 合成的
[0744] <400> 92
[0745] ttccttttcg ttcgtataaa atttcgt 27
[0746] <210> 93
[0747] <211> 21
[0748] <212> DNA
[0749] <213> 人工序列
[0750] <220>
[0751] <223> 合成的
[0752] <400> 93
[0753] ccacggacga tcgggtttta g 21
[0754] <210> 94
[0755] <211> 20
[0756] <212> DNA
[0757] <213> 人工序列
[0758] <220>
[0759] <223> 合成的
[0760] <400> 94
[0761] cgccgaggat cgggttttag 20
[0762] <210> 95
[0763] <211> 29
[0764] <212> DNA
[0765] <213> 人工序列
[0766] <220>
[0767] <223> 合成的
[0768] <400> 95
[0769] agccggtttt cggctgaga ctccgcgtc 29
[0770] <210> 96
[0771] <211> 32
[0772] <212> DNA
[0773] <213> 人工序列
[0774] <220>
[0775] <223> 合成的
[0776] <220>
[0777] <221> 混杂特征
[0778] <222> (3) .. (3)

[0779] <223> 用淬灭剂修饰的
[0780] <400> 96
[0781] tctagccggt tttccggctg agactccgcg tc 32
[0782] <210> 97
[0783] <211> 29
[0784] <212> DNA
[0785] <213> 人工序列
[0786] <220>
[0787] <223> 合成的
[0788] <400> 97
[0789] agccggttttt ccggctgaga cgtccgtgg 29
[0790] <210> 98
[0791] <211> 31
[0792] <212> DNA
[0793] <213> 人工序列
[0794] <220>
[0795] <223> 合成的
[0796] <220>
[0797] <221> 混杂特征
[0798] <222> (3) .. (3)
[0799] <223> 用淬灭剂修饰的
[0800] <400> 98
[0801] tctagccggtt ttccggctga gacgtccgtg g 31
[0802] <210> 99
[0803] <211> 28
[0804] <212> DNA
[0805] <213> 人工序列
[0806] <220>
[0807] <223> 合成的
[0808] <400> 99
[0809] agccggttttt ccggctgaga cctcggcg 28
[0810] <210> 100
[0811] <211> 31
[0812] <212> DNA
[0813] <213> 人工序列
[0814] <220>
[0815] <223> 合成的
[0816] <220>
[0817] <221> 混杂特征

[0818]	<222> (3) .. (3)
[0819]	<223> 用淬灭剂修饰的
[0820]	<400> 100
[0821]	tctagccggt tttccggctg agacctcggc g 31
[0822]	<210> 101
[0823]	<211> 8
[0824]	<212> DNA
[0825]	<213> 人工序列
[0826]	<220>
[0827]	<223> 合成的
[0828]	<400> 101
[0829]	cgccgagg 8
[0830]	<210> 102
[0831]	<211> 9
[0832]	<212> DNA
[0833]	<213> 人工序列
[0834]	<220>
[0835]	<223> 合成的
[0836]	<400> 102
[0837]	gacgcggag 9
[0838]	<210> 103
[0839]	<211> 9
[0840]	<212> DNA
[0841]	<213> 人工序列
[0842]	<220>
[0843]	<223> 合成的
[0844]	<400> 103
[0845]	ccacggacg 9
[0846]	<210> 104
[0847]	<211> 8
[0848]	<212> DNA
[0849]	<213> 人工序列
[0850]	<220>
[0851]	<223> 合成的
[0852]	<400> 104
[0853]	gcgcgtcc 8

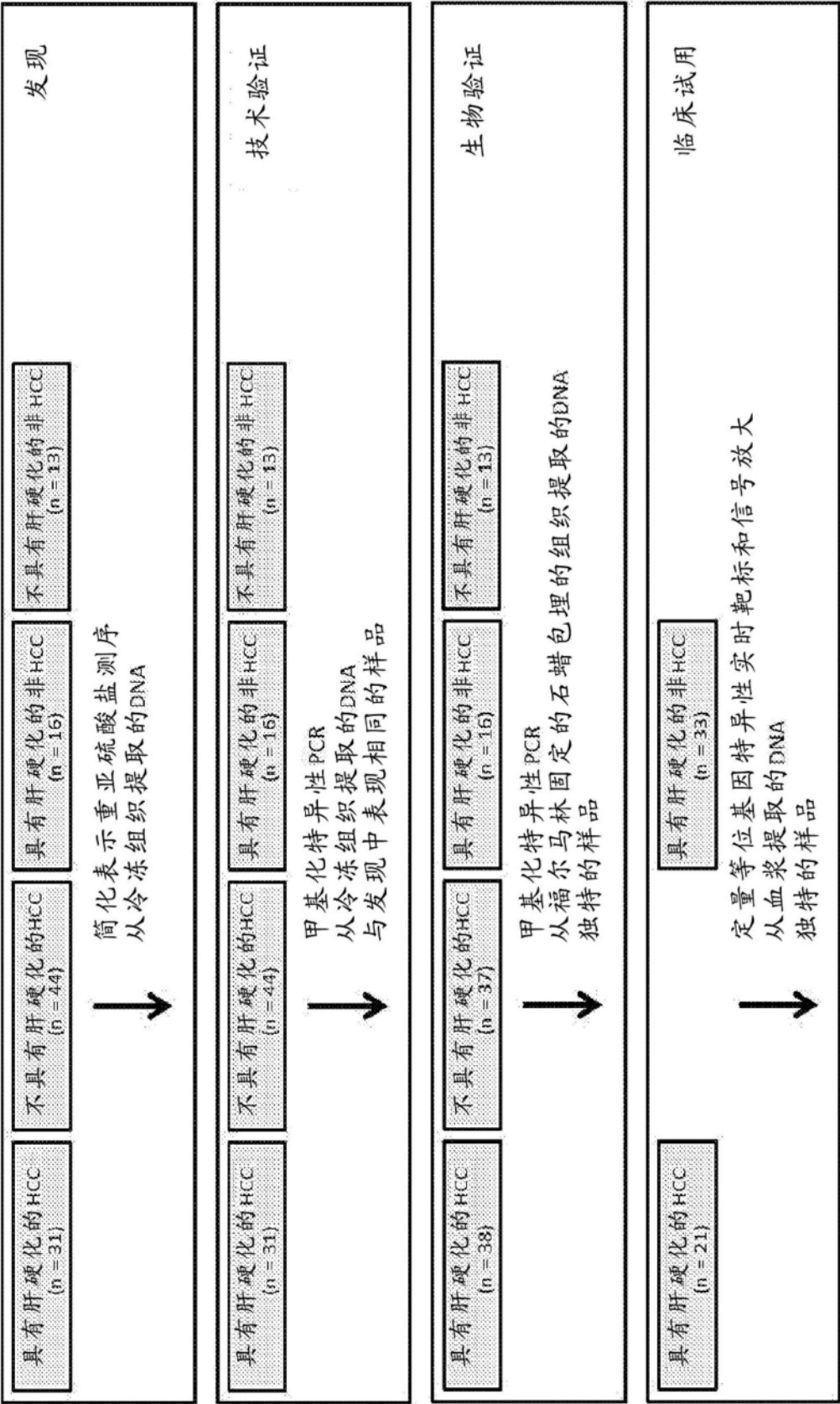


图1

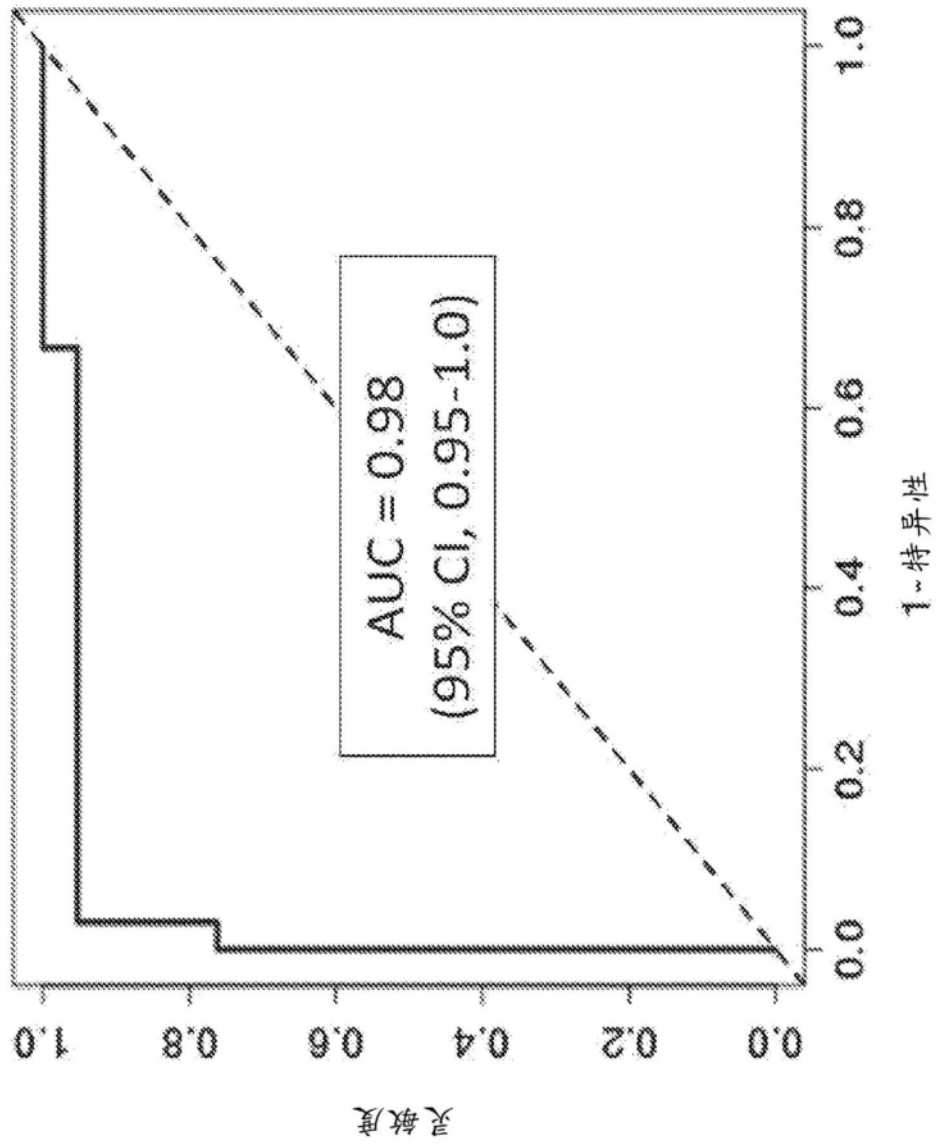


图2

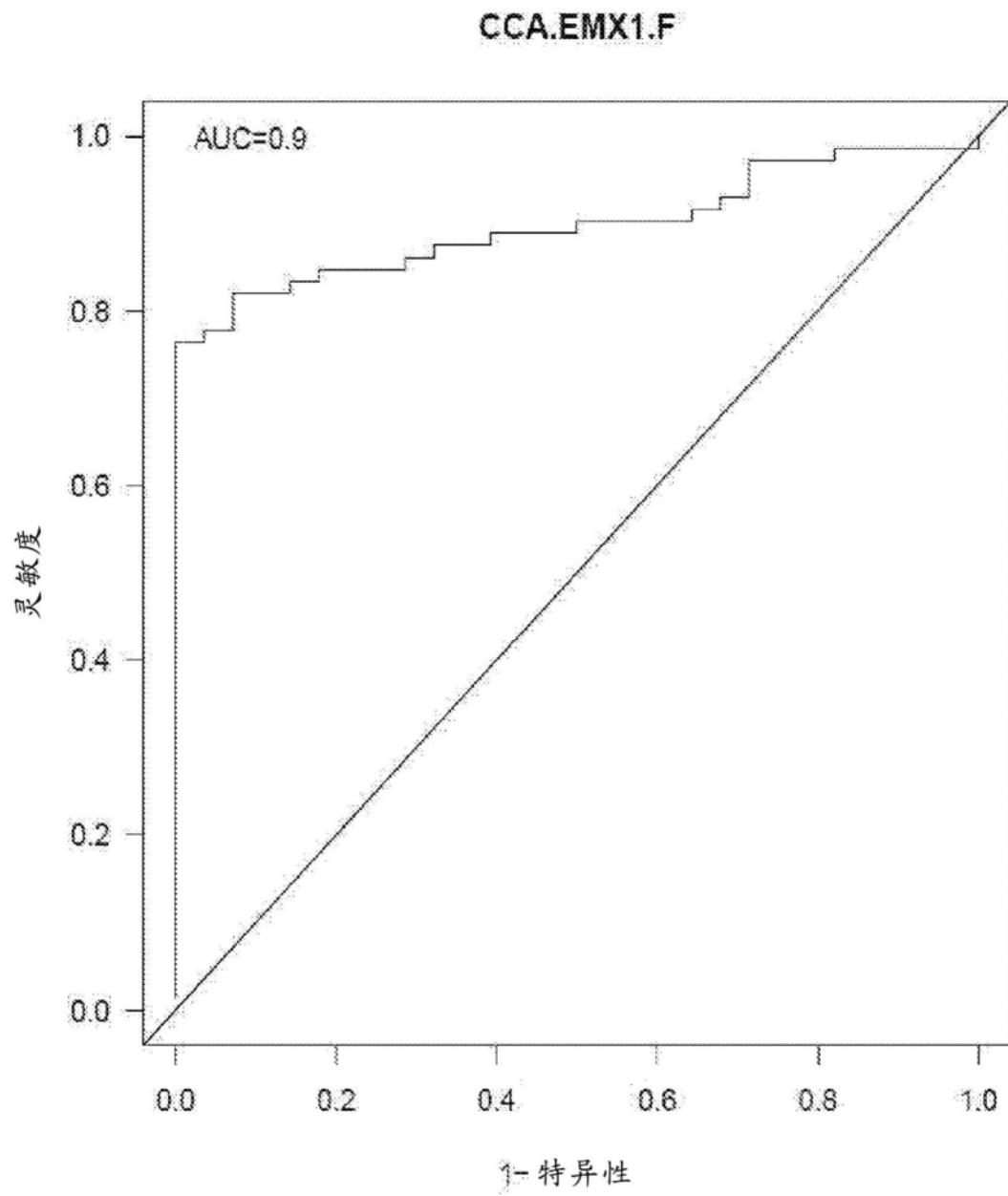


图3A

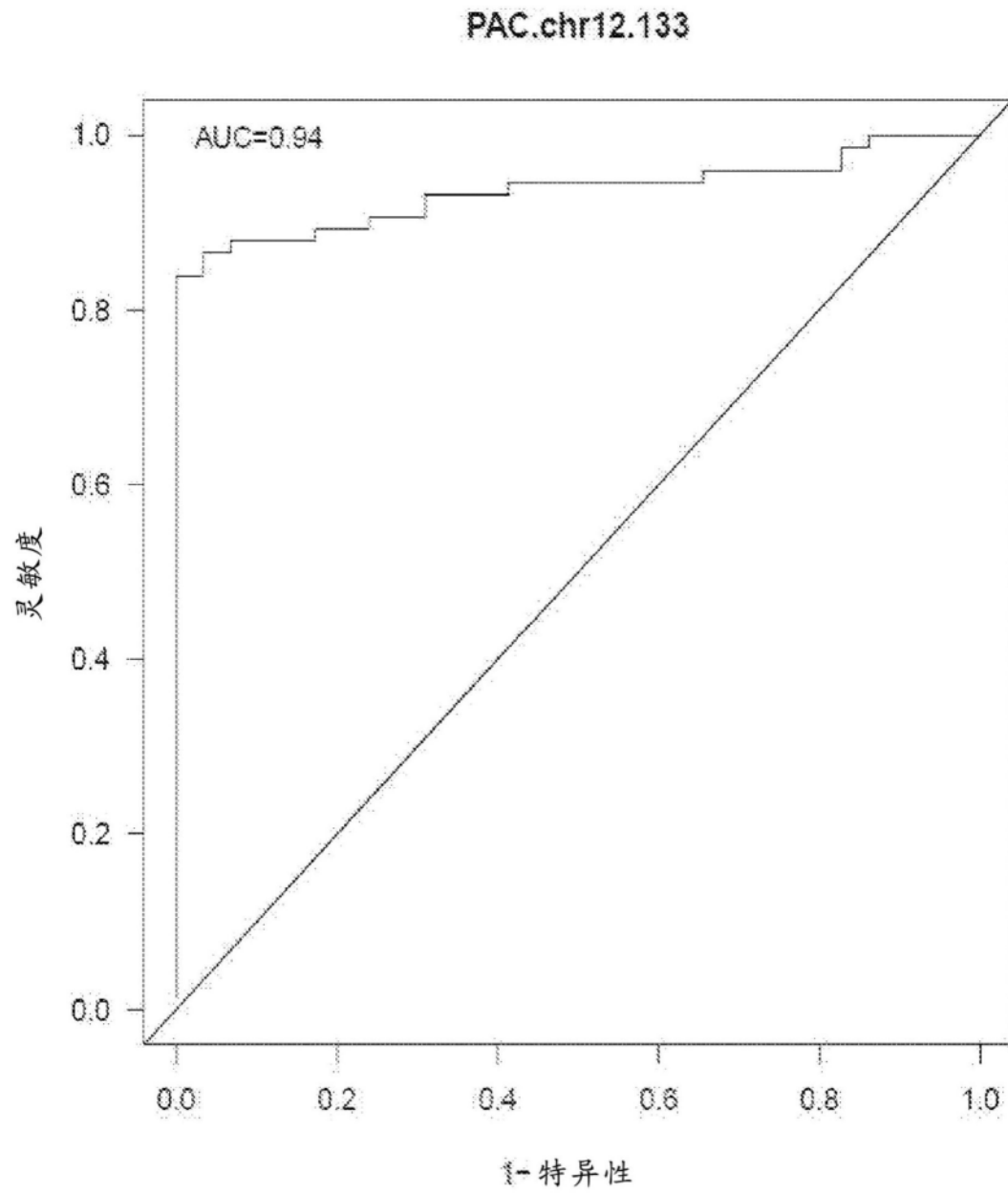


图3B

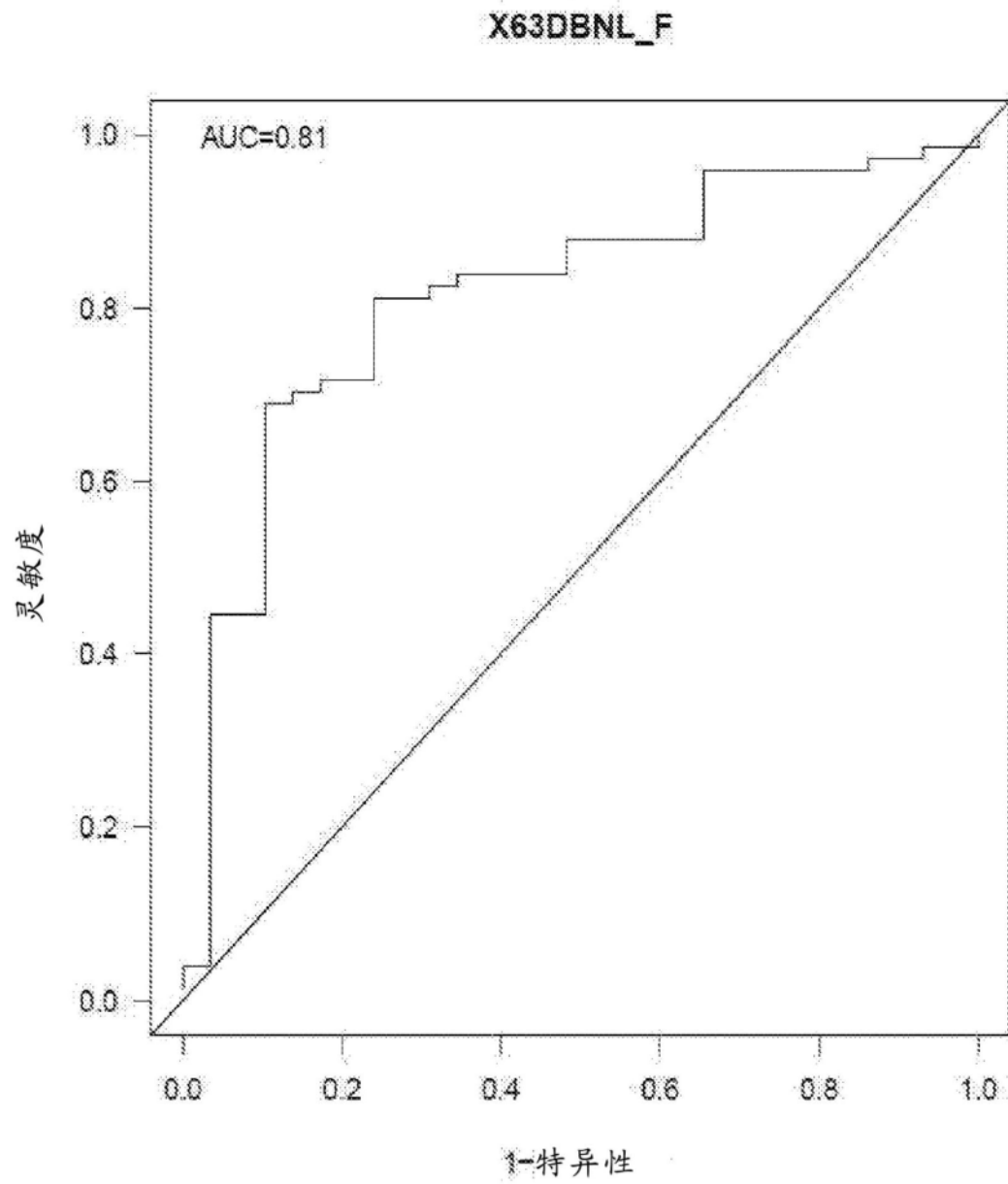


图3C

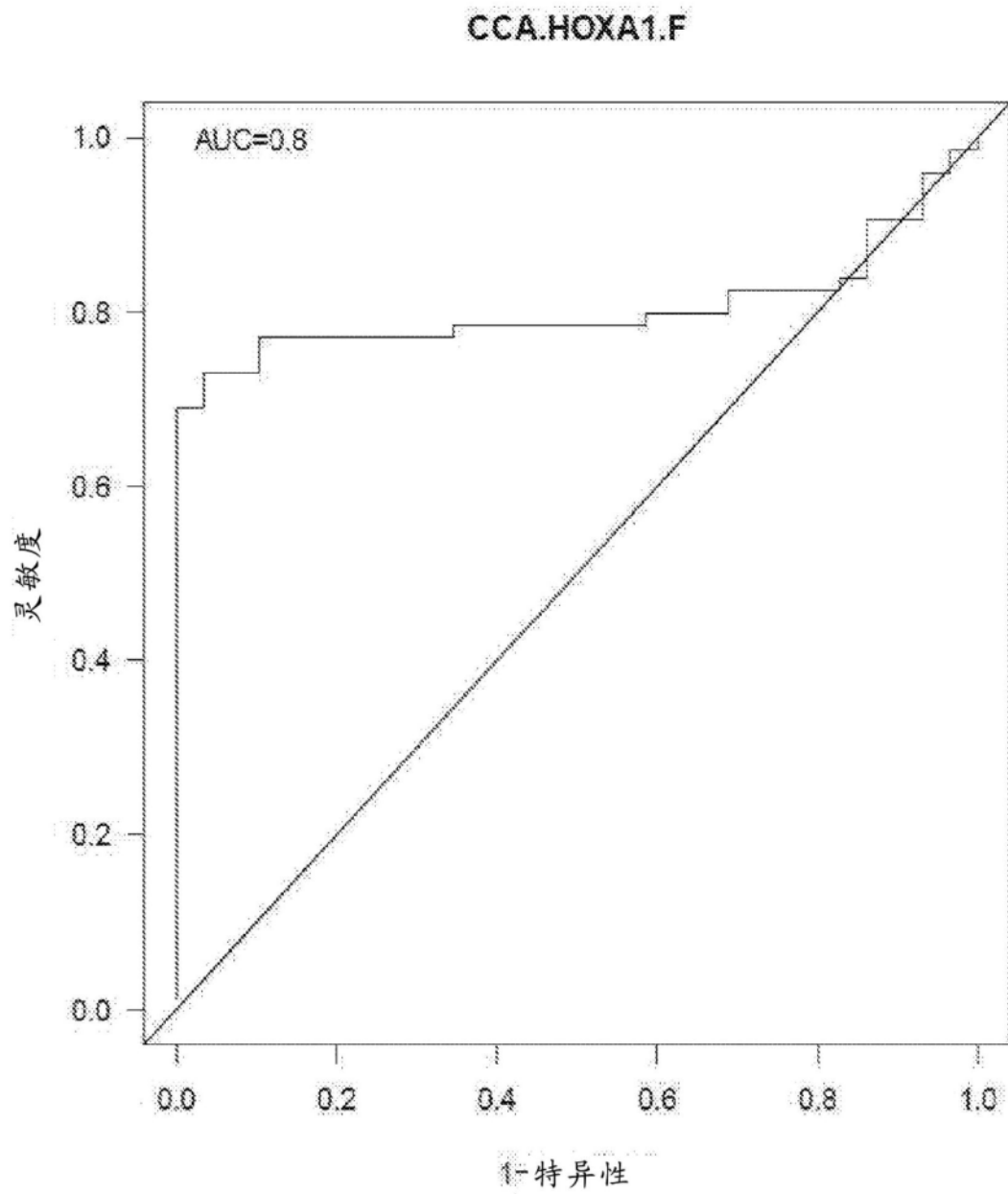


图3D

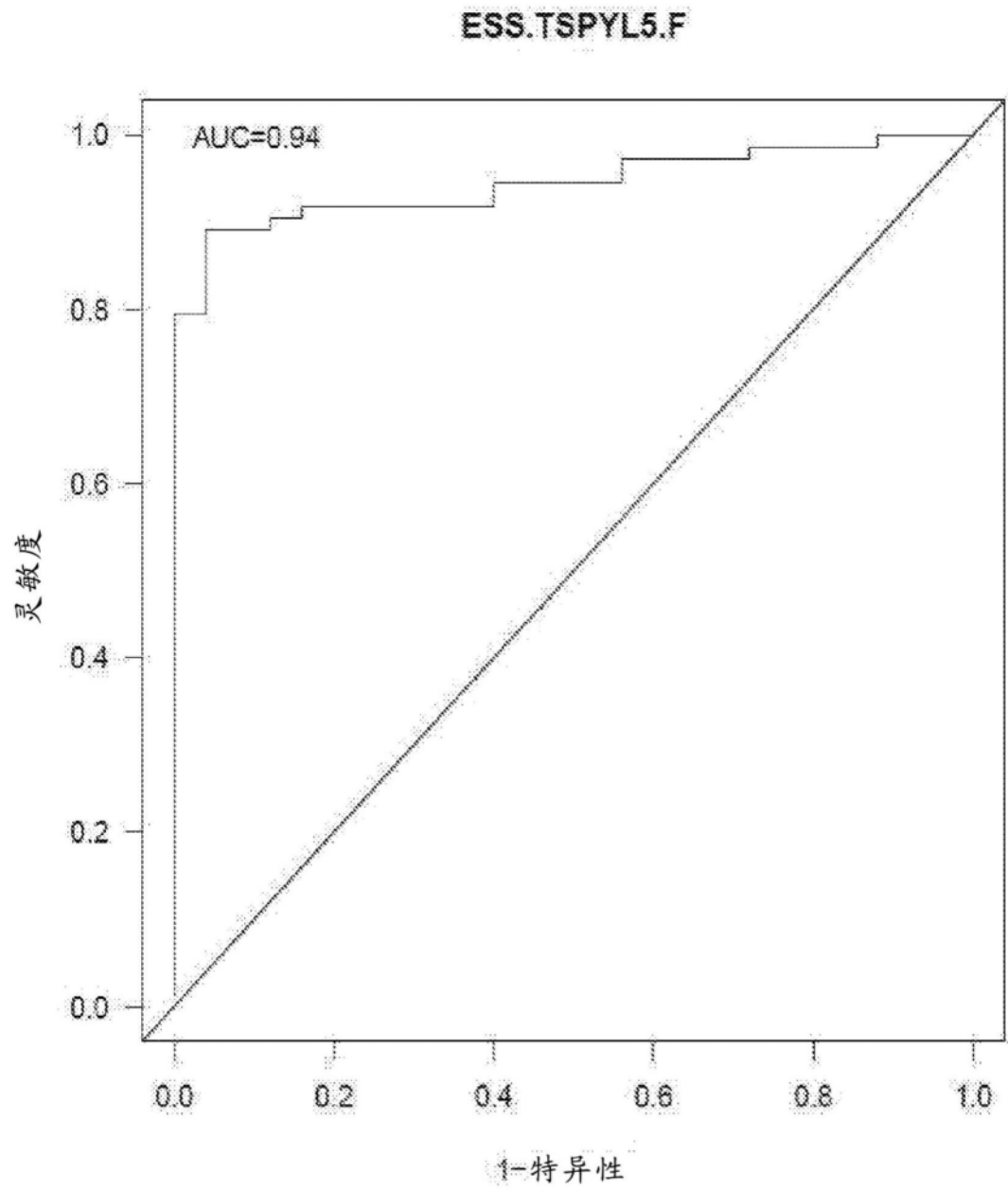


图3E

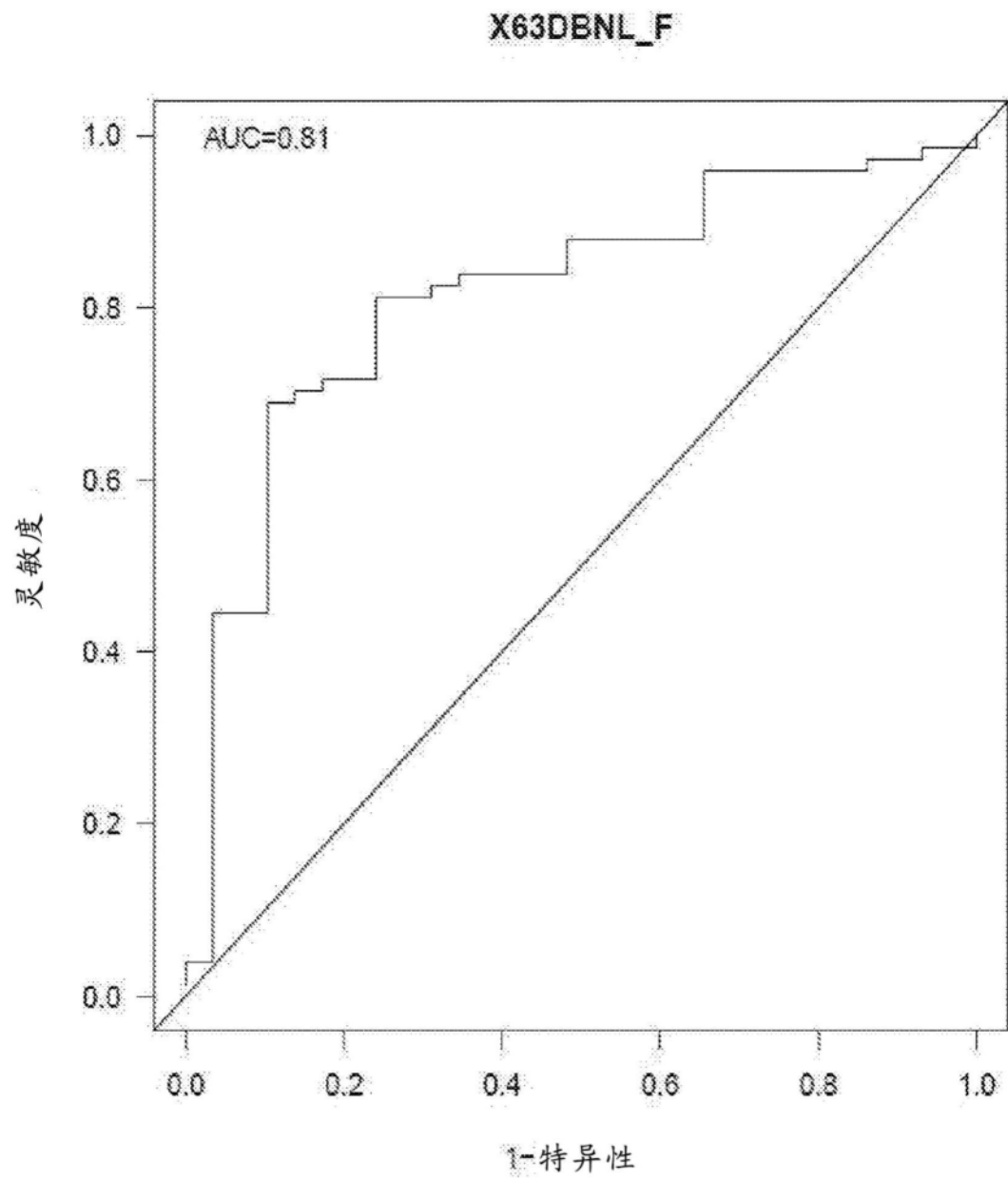


图3F

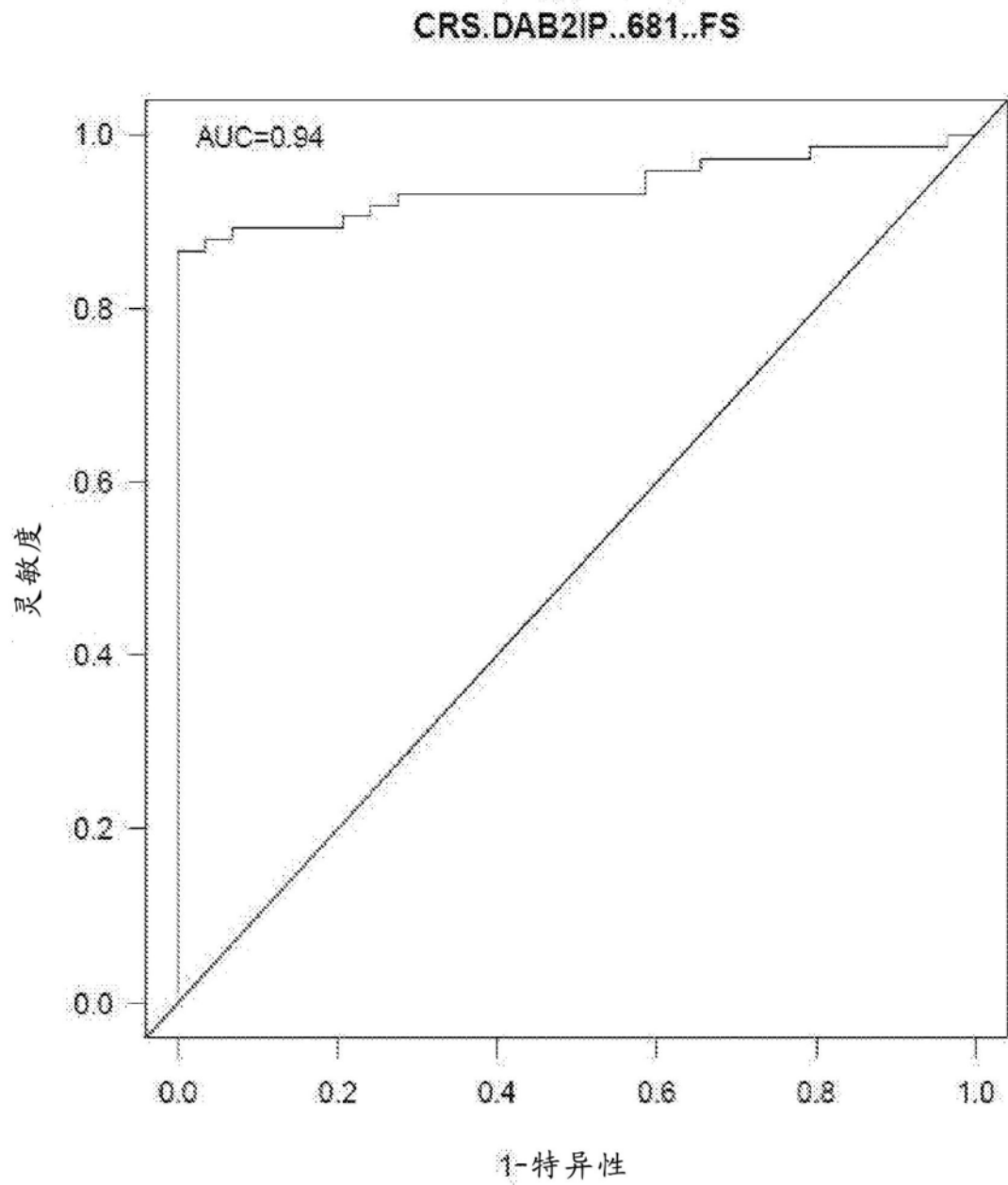


图3G

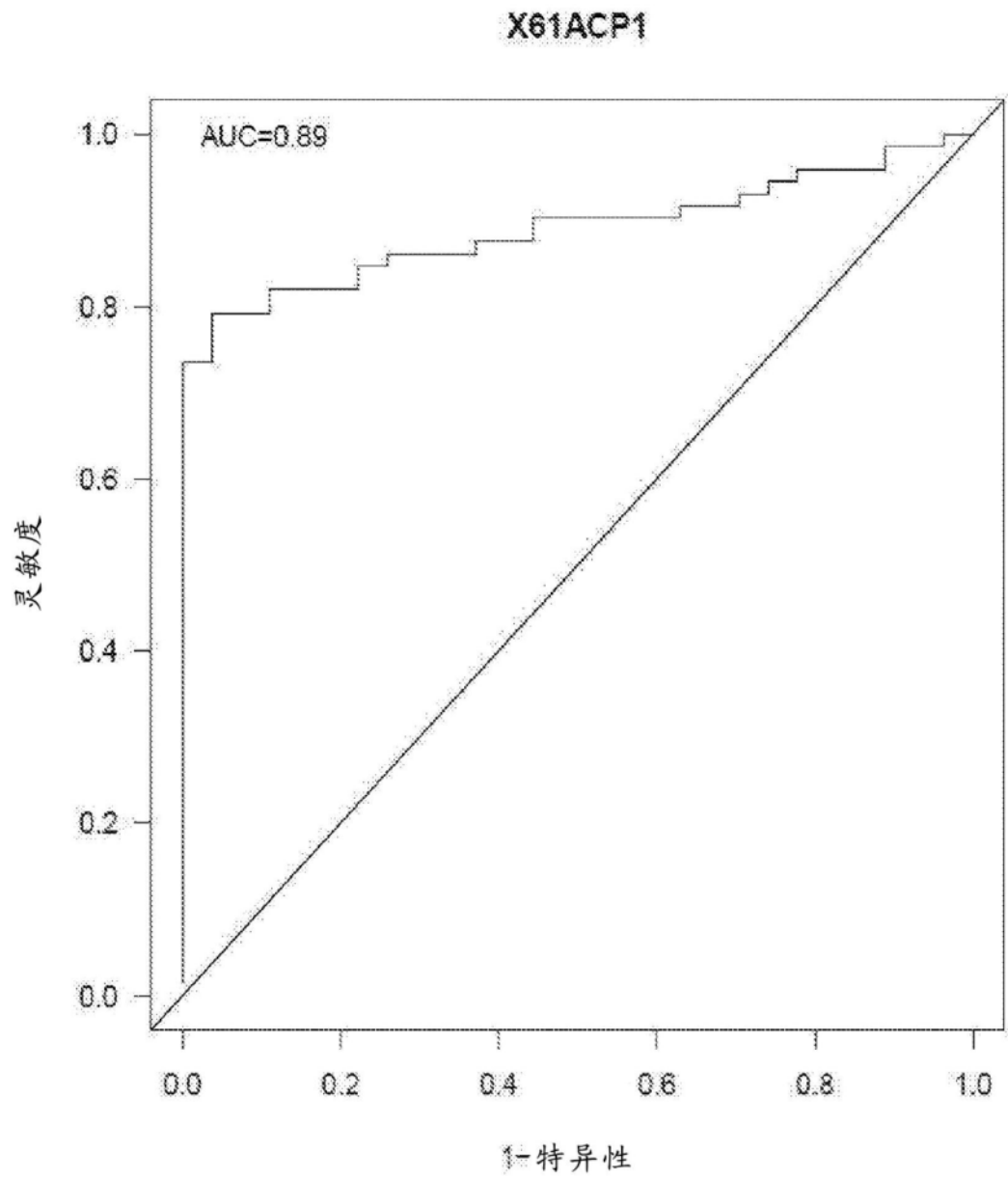


图3G

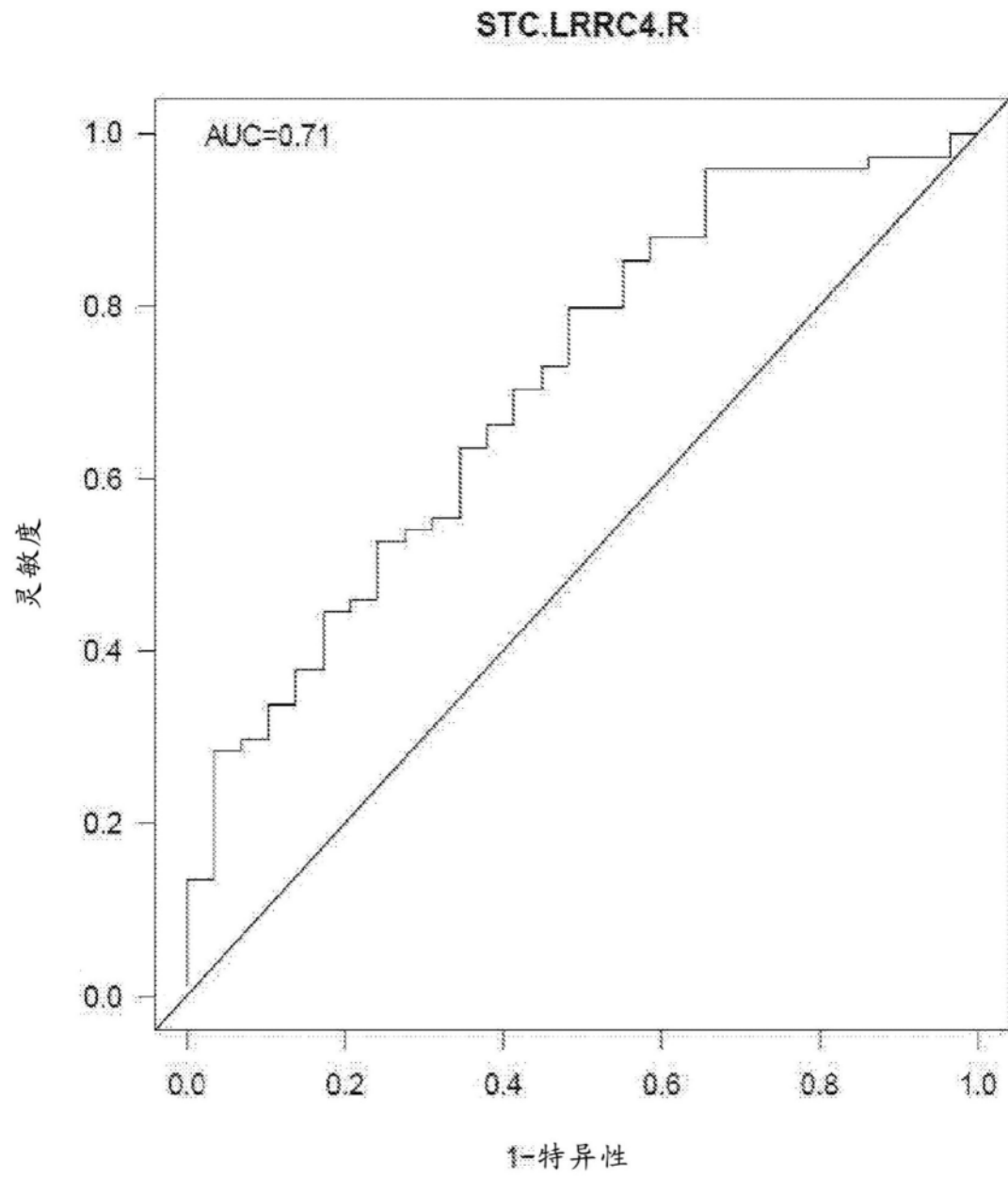


图3H

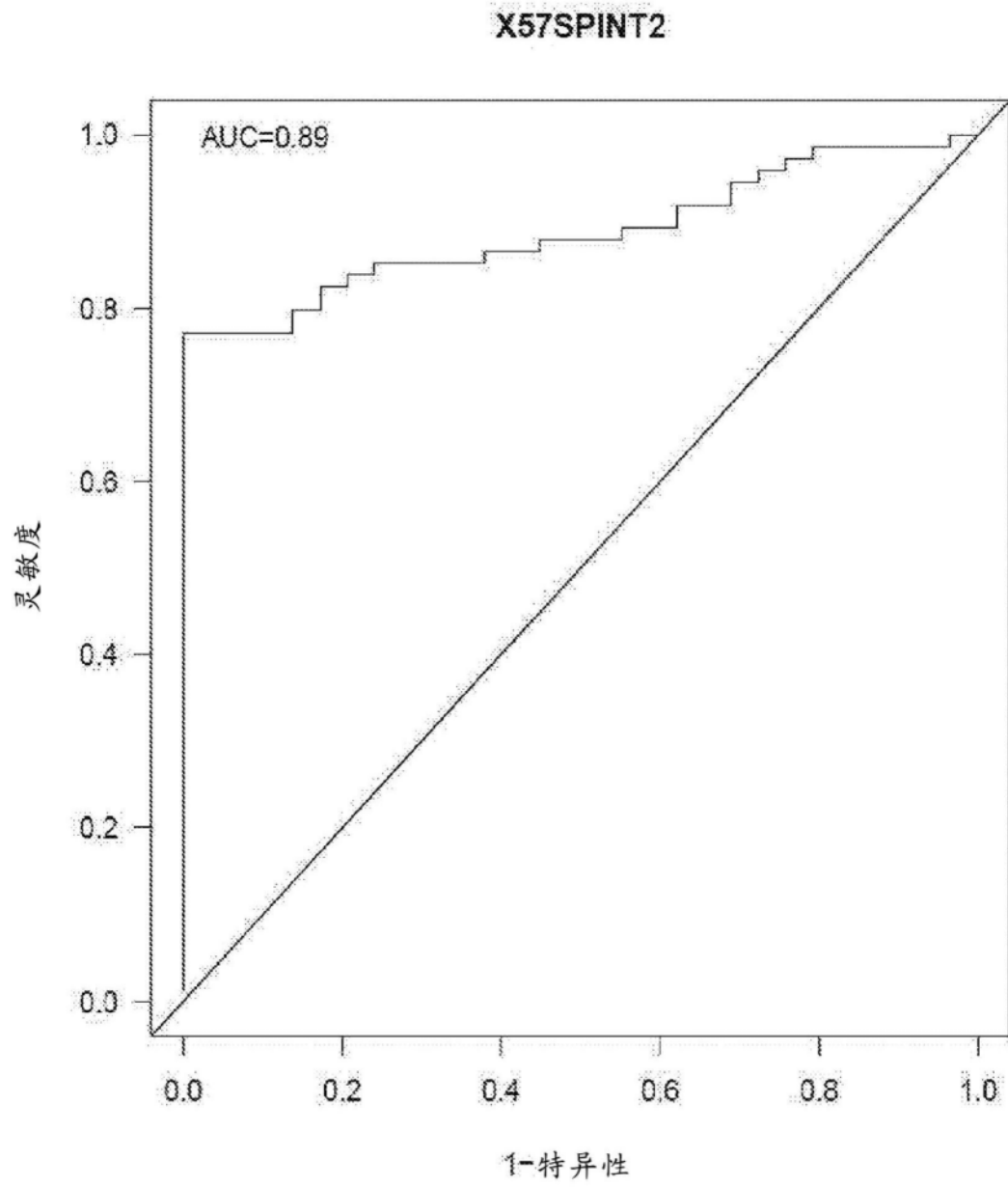


图3I

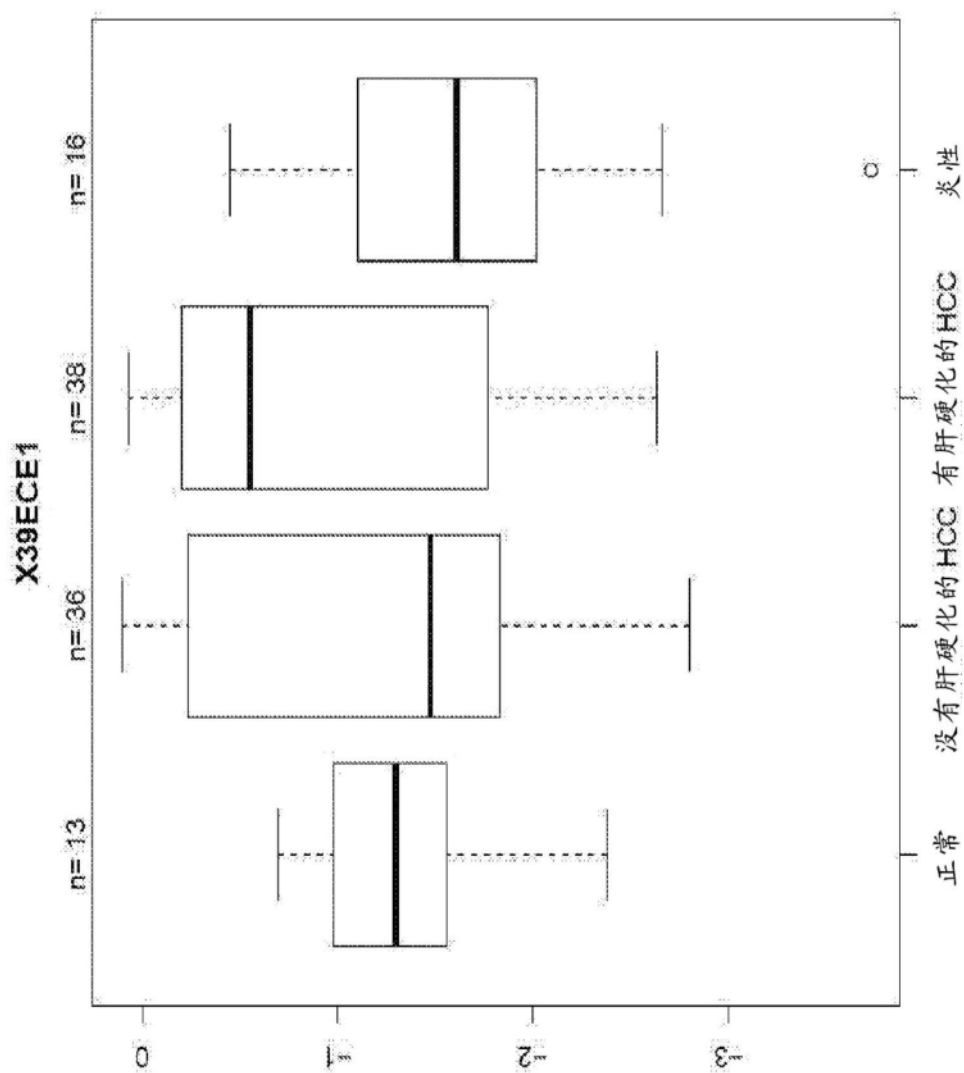


图4A

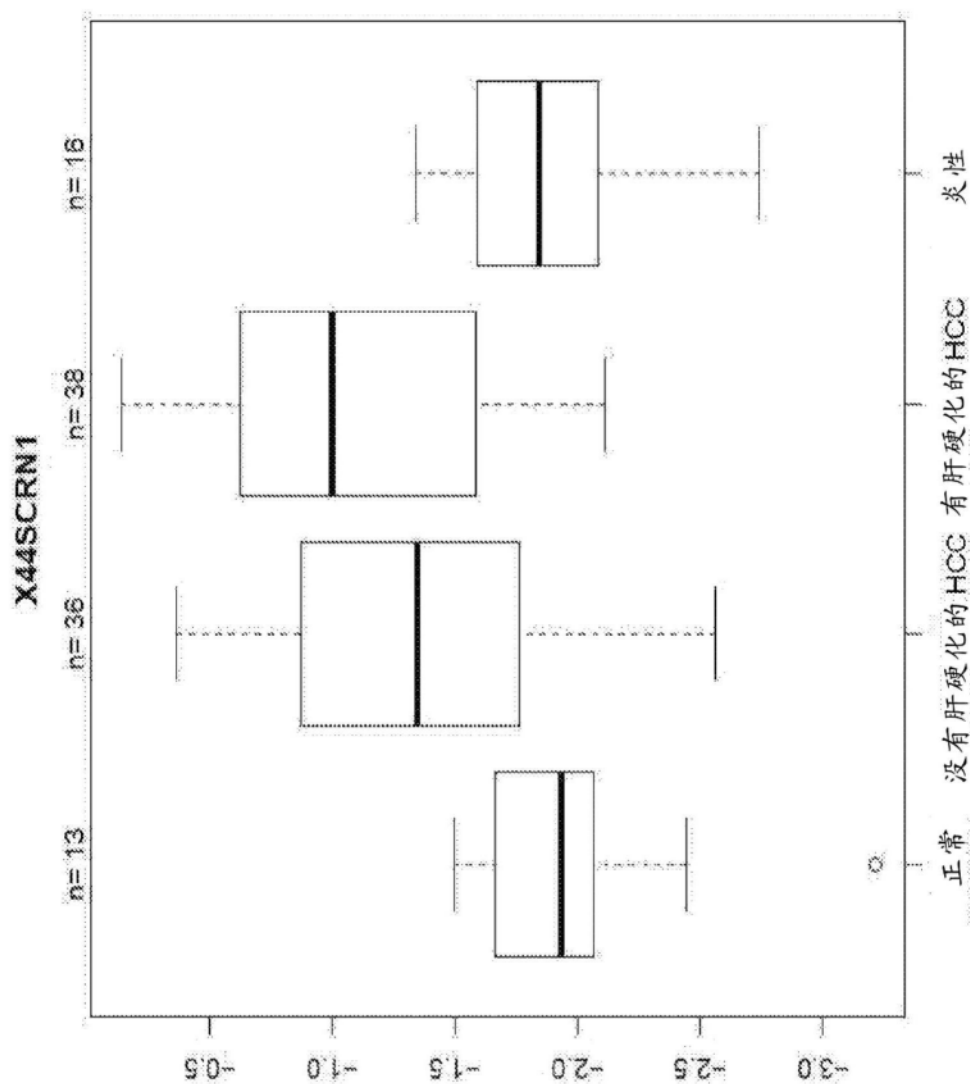


图4B

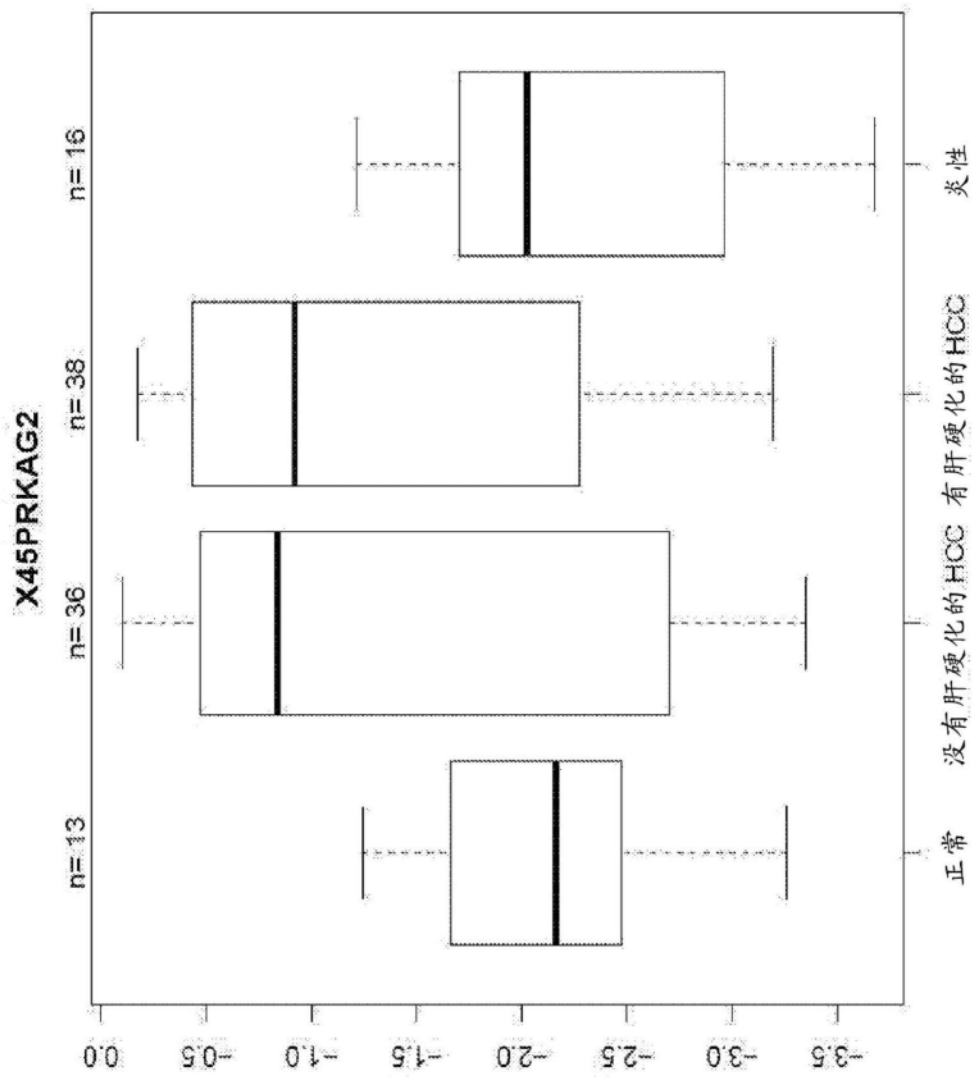


图4C

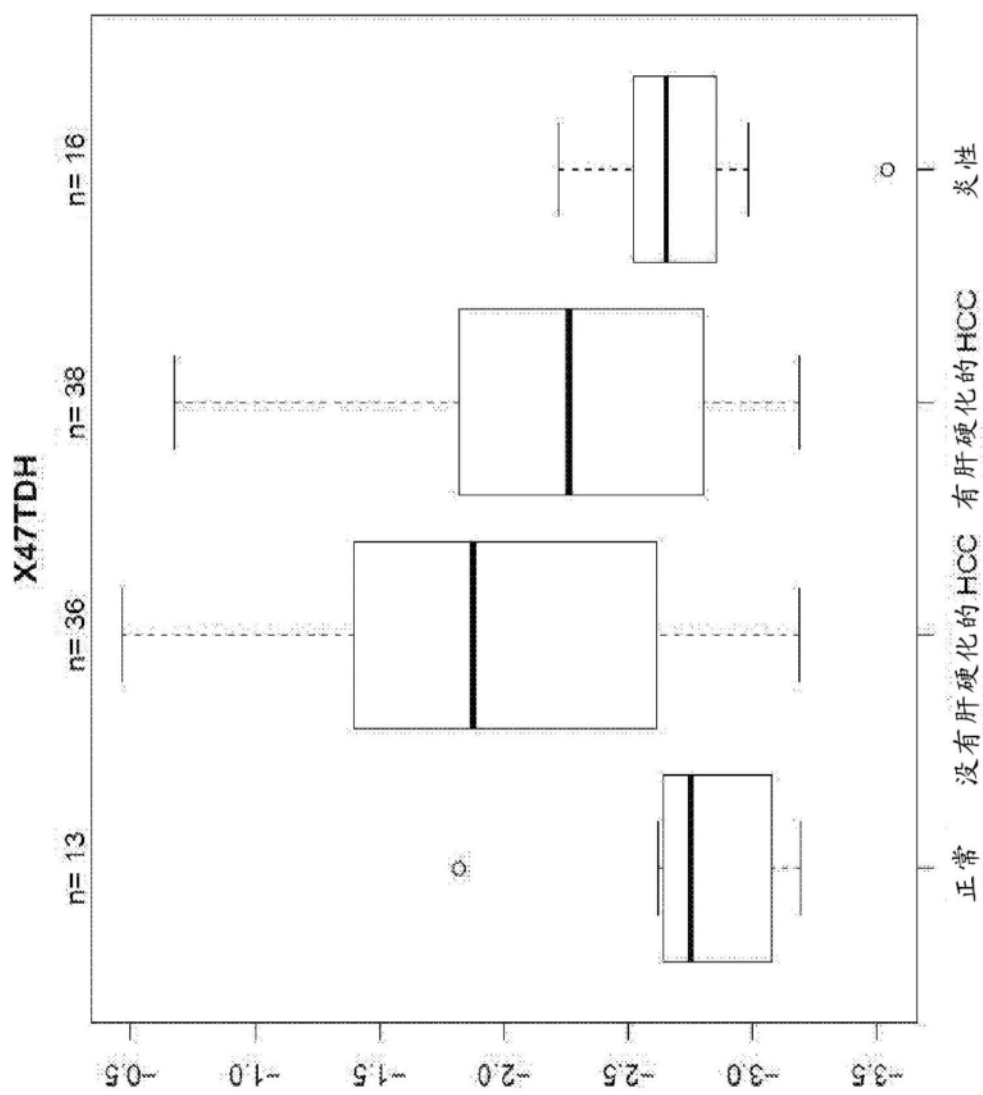


图4D

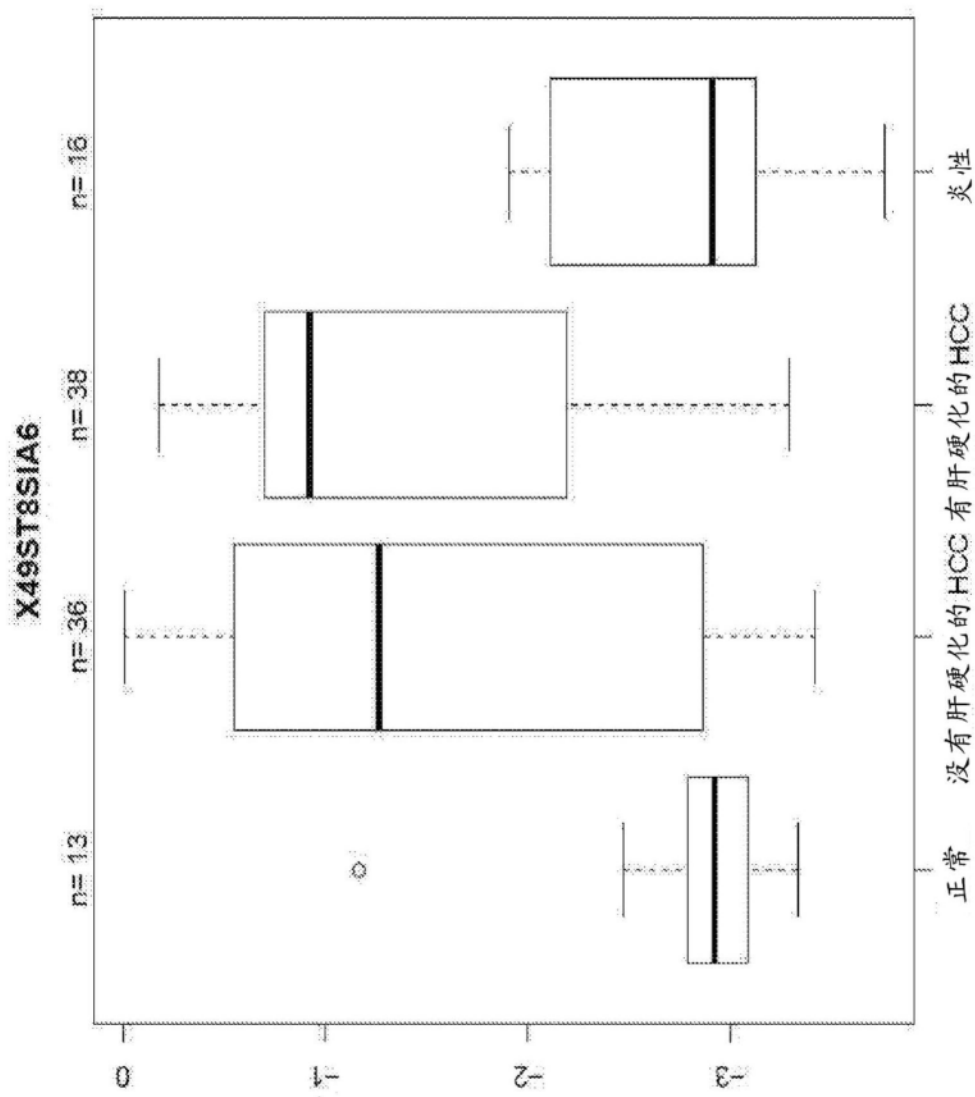


图4E

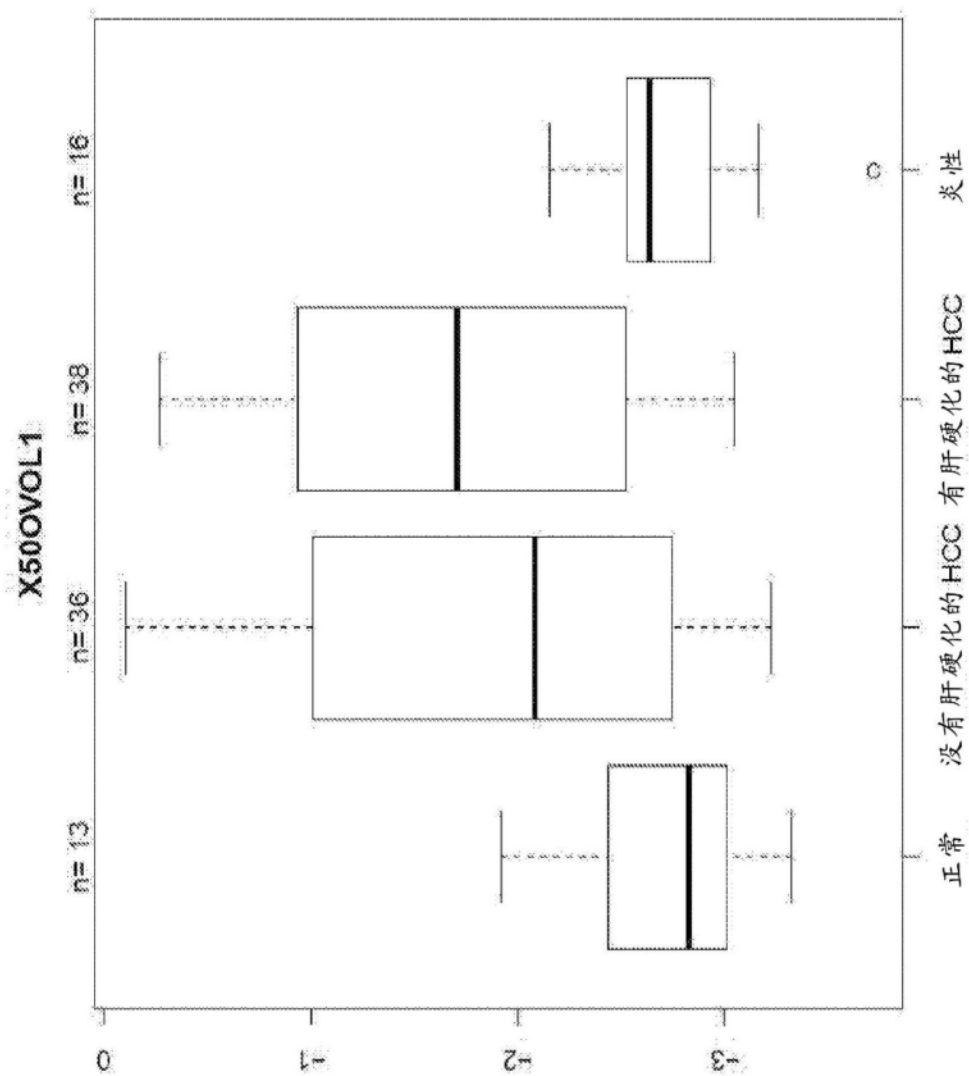


图4F

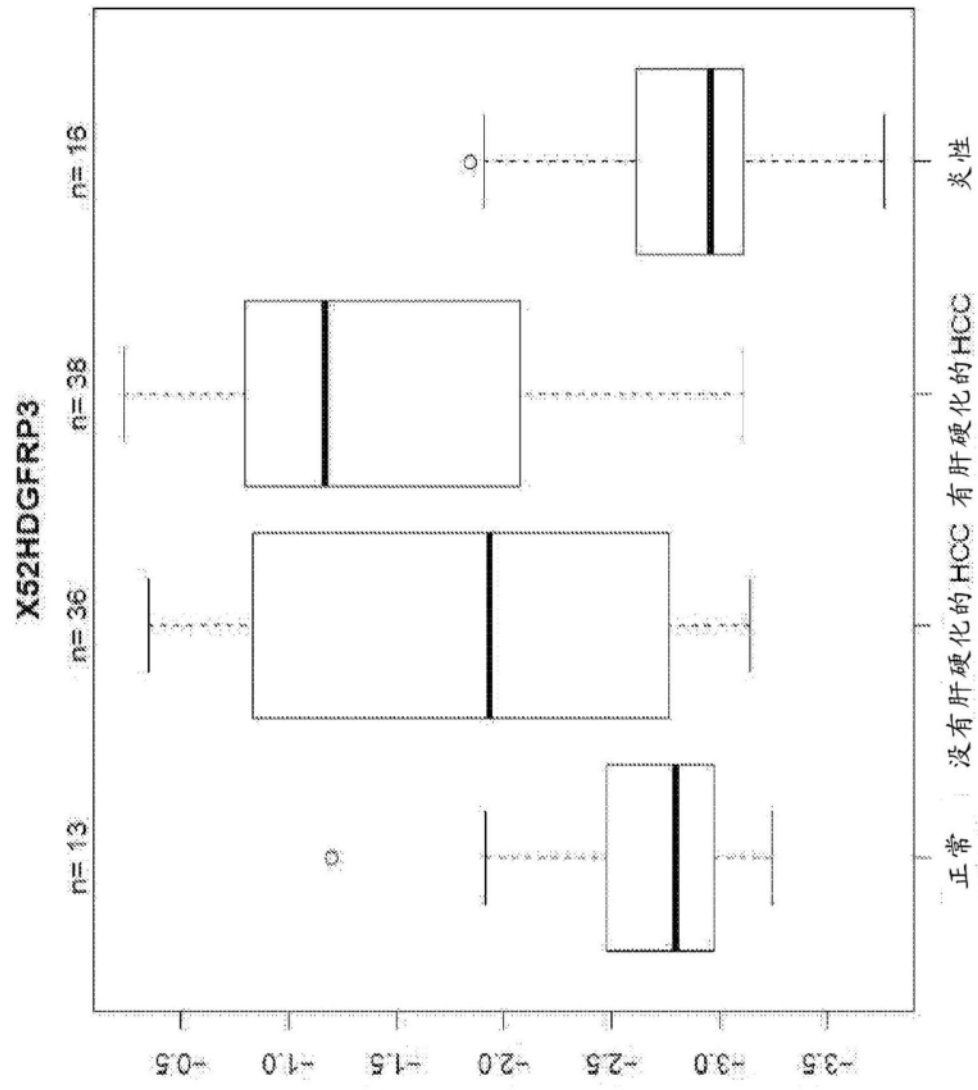


图4G

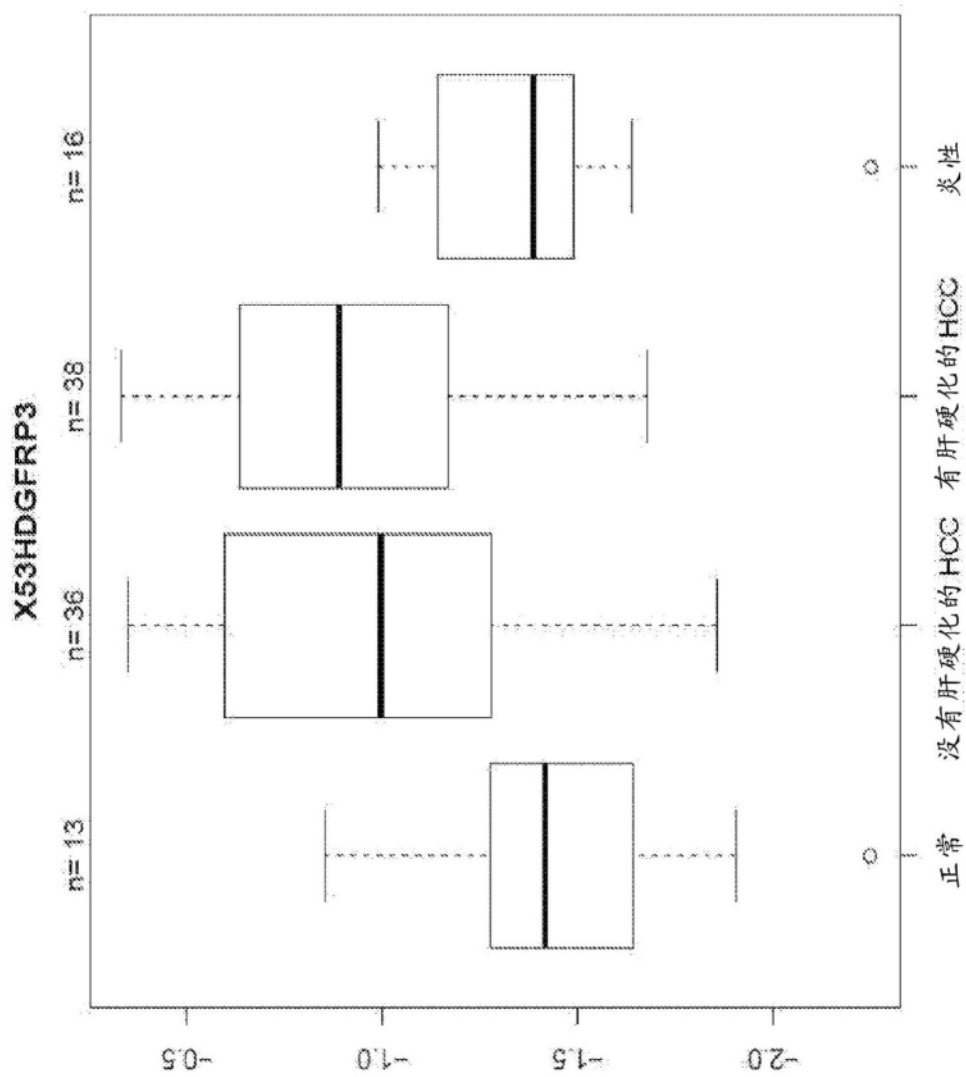


图4H

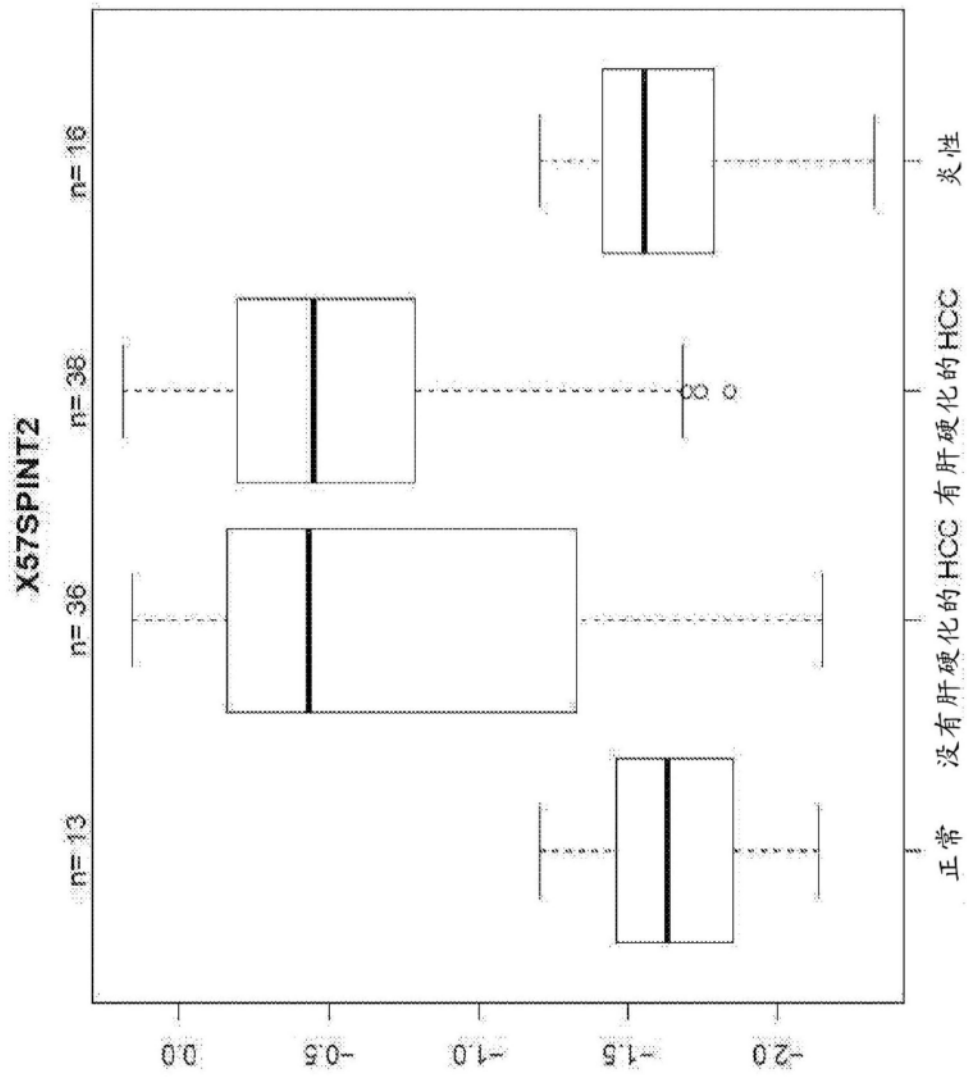


图4I

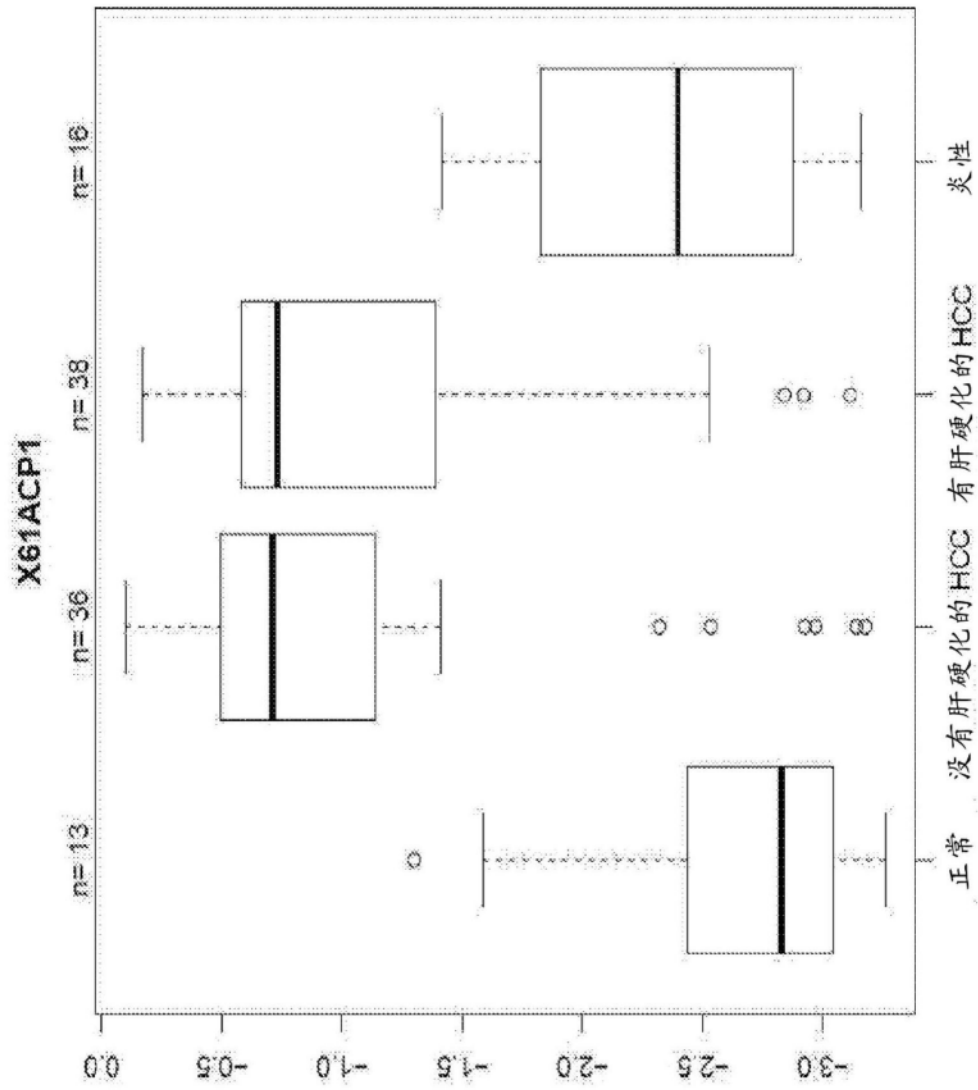


图4J

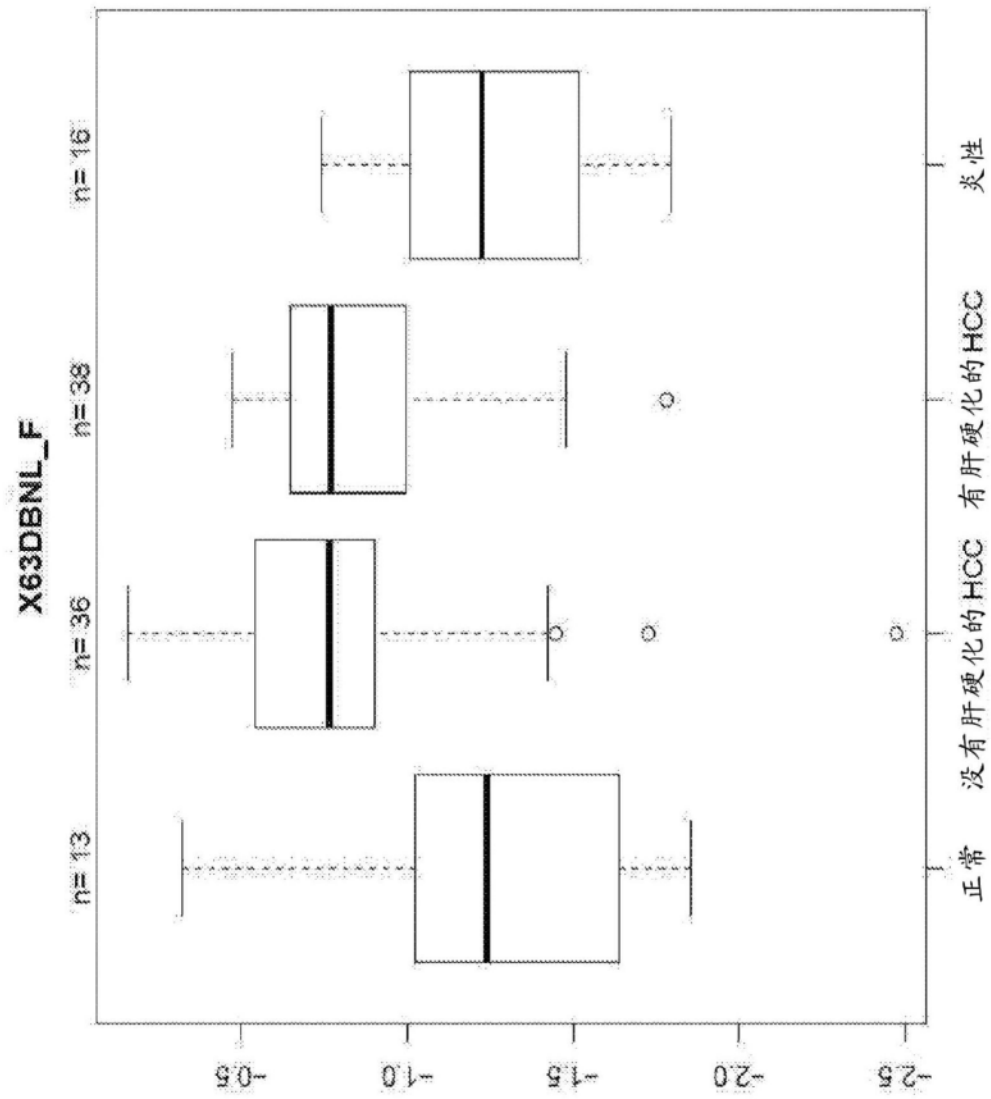


图4K

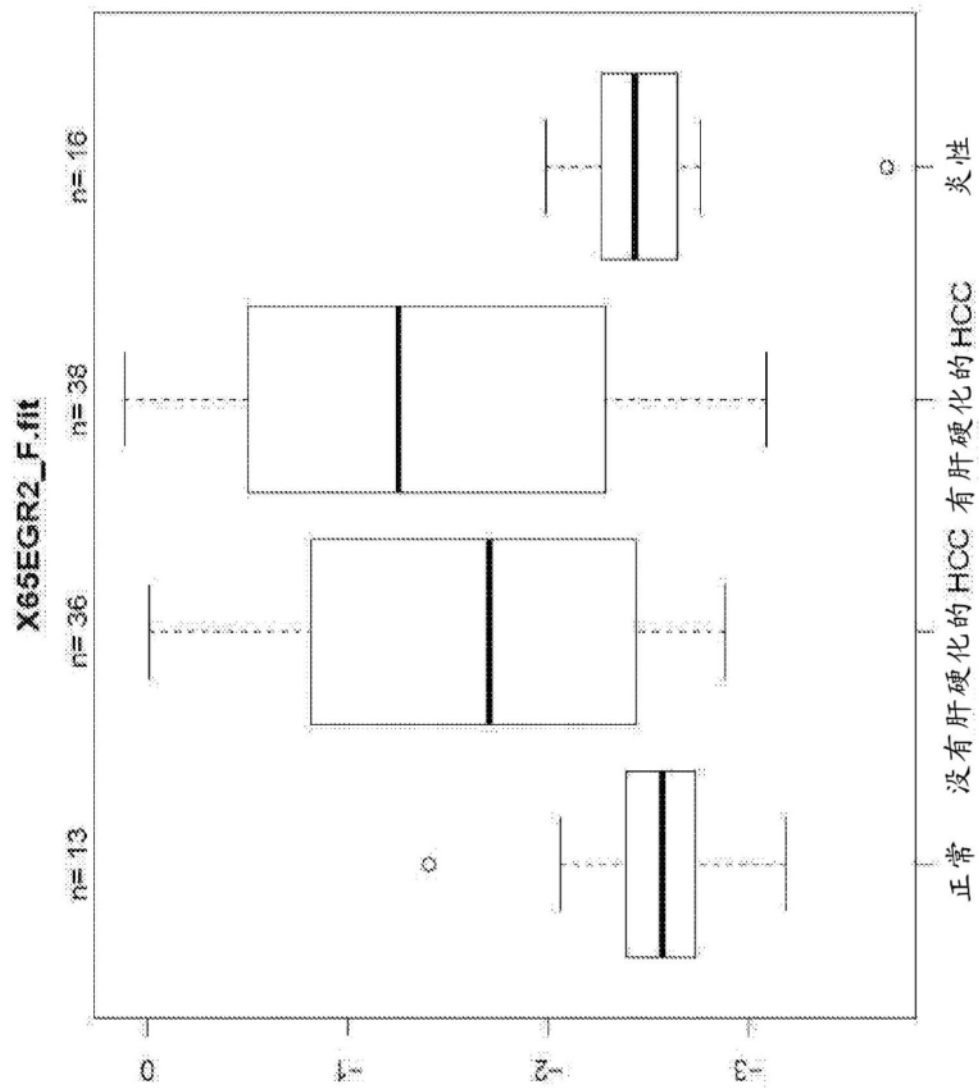


图4L

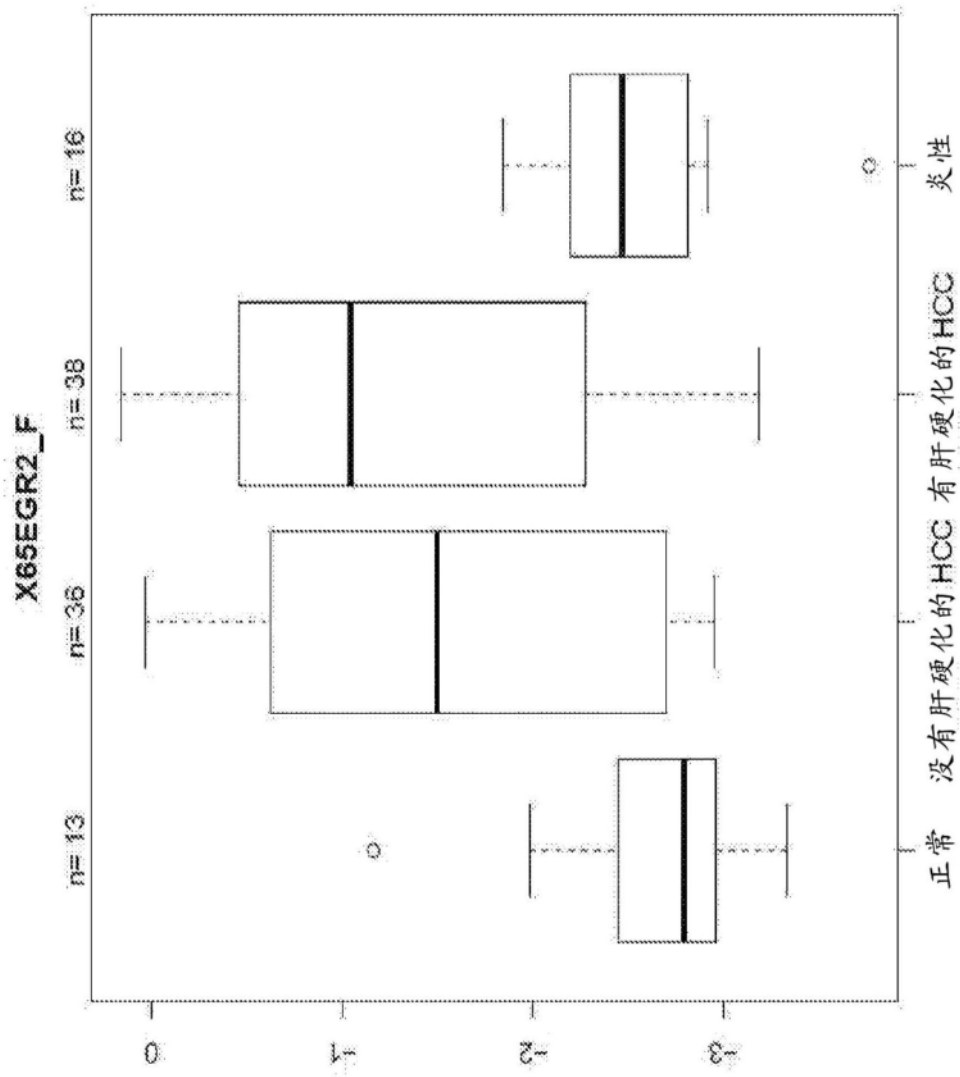


图4M

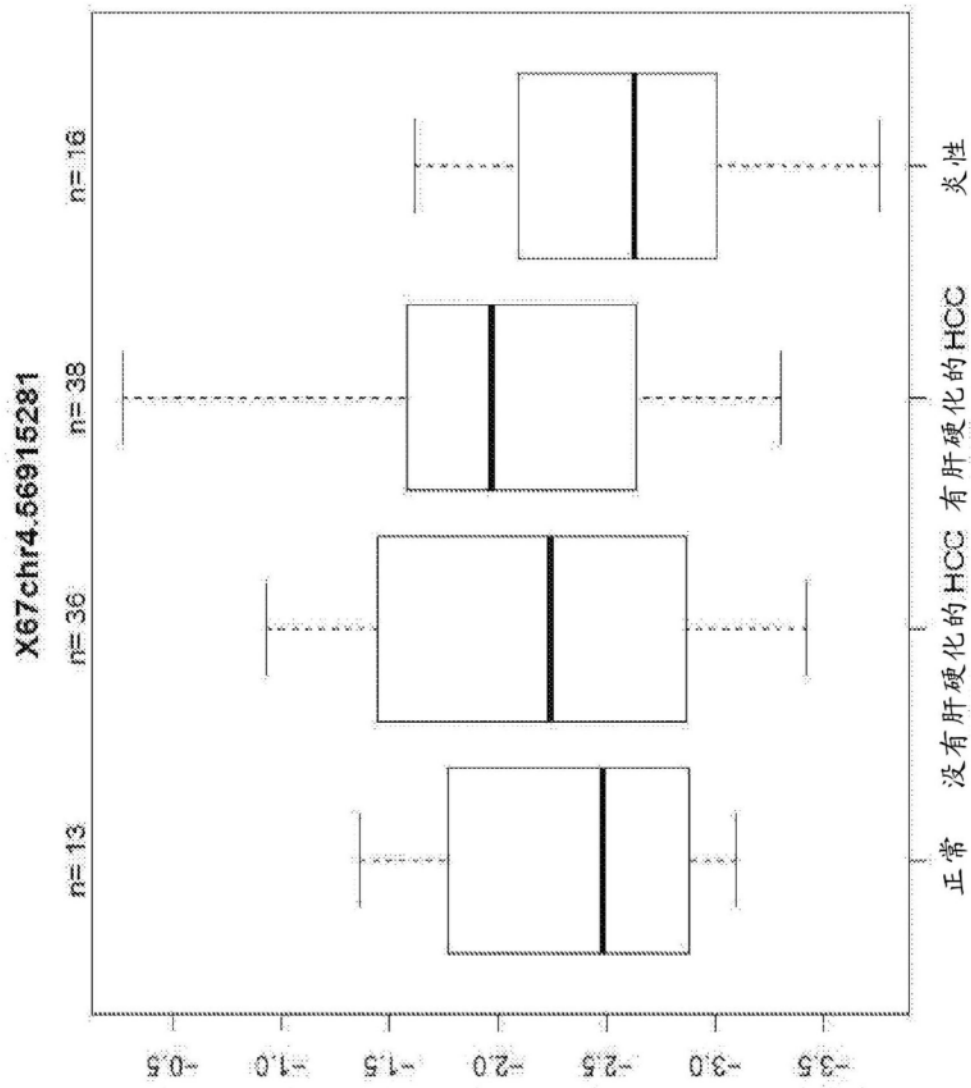


图4N

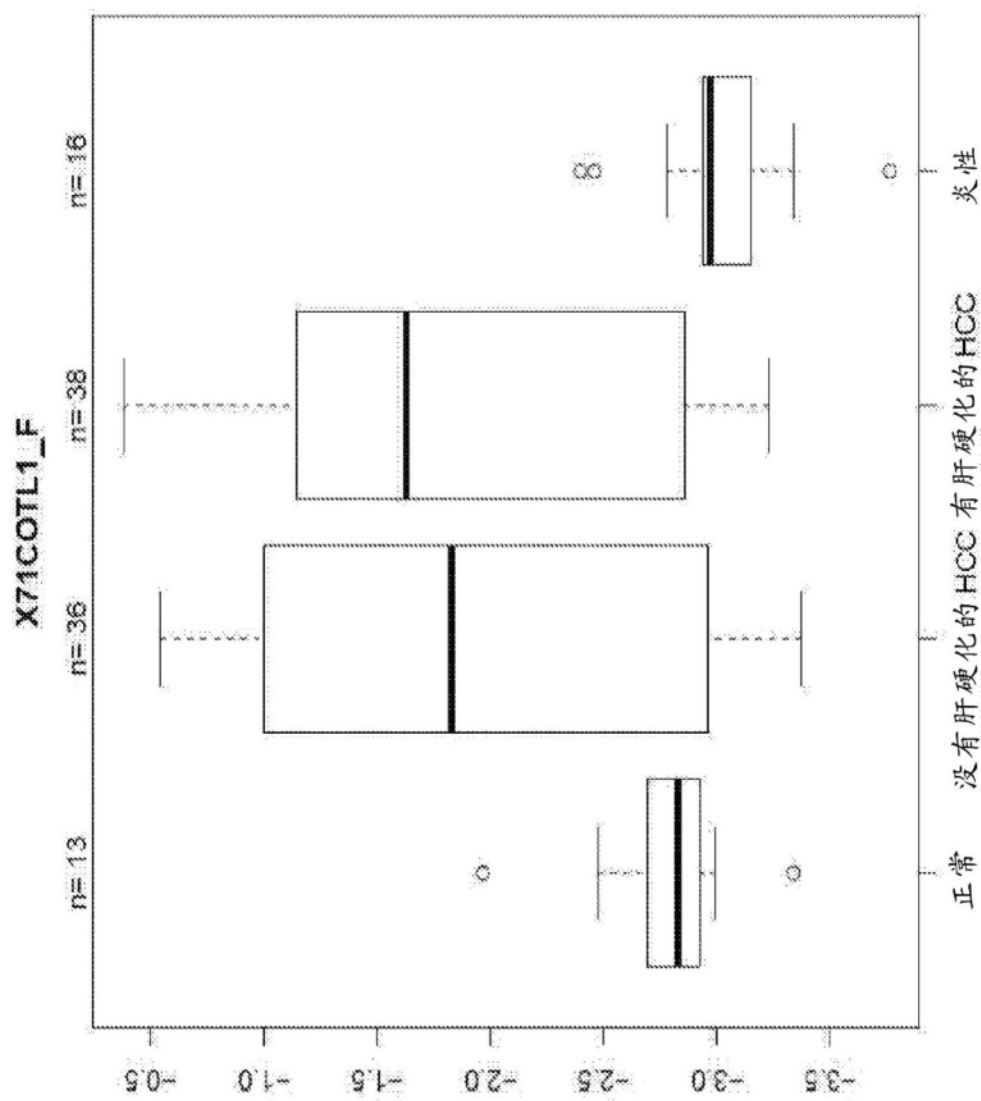


图40

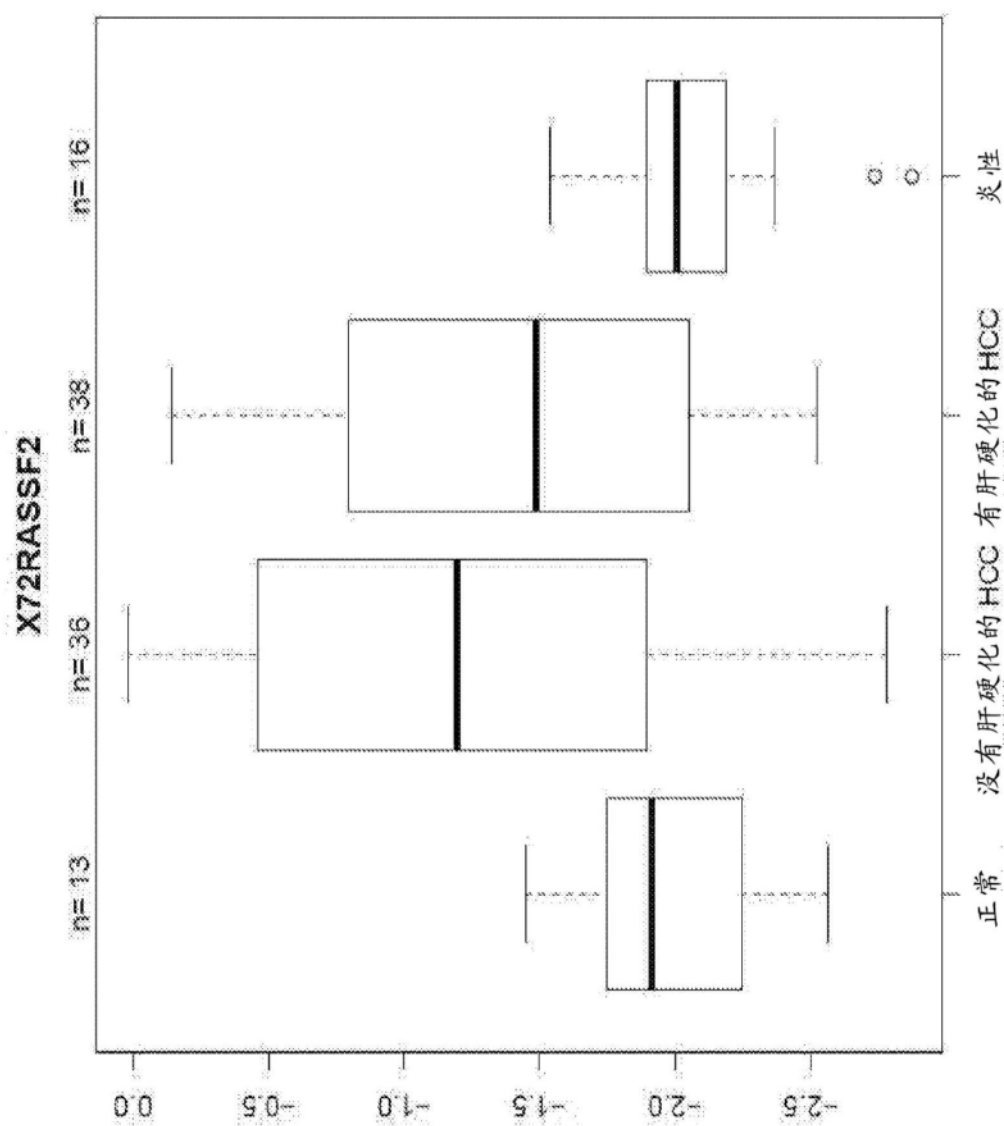


图4P

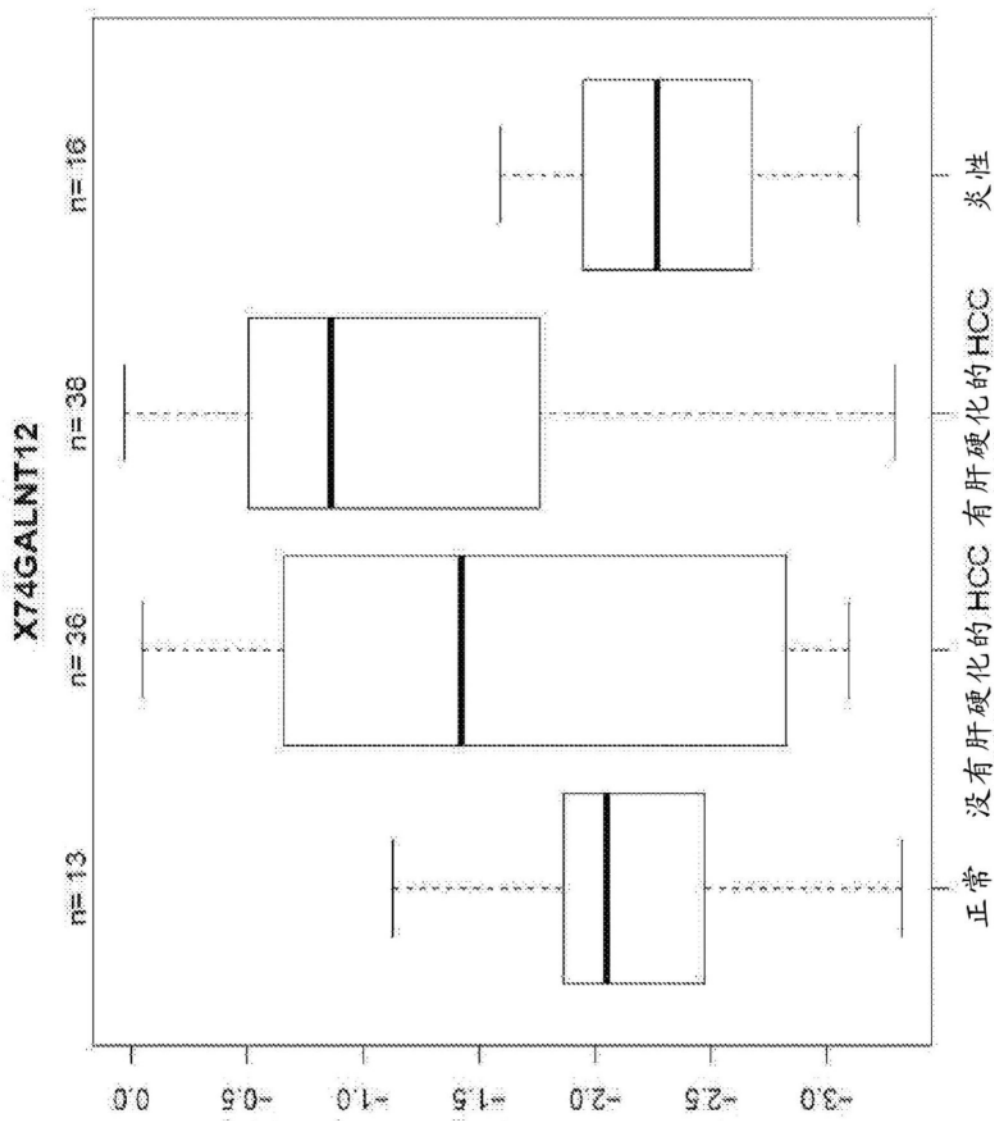


图4Q

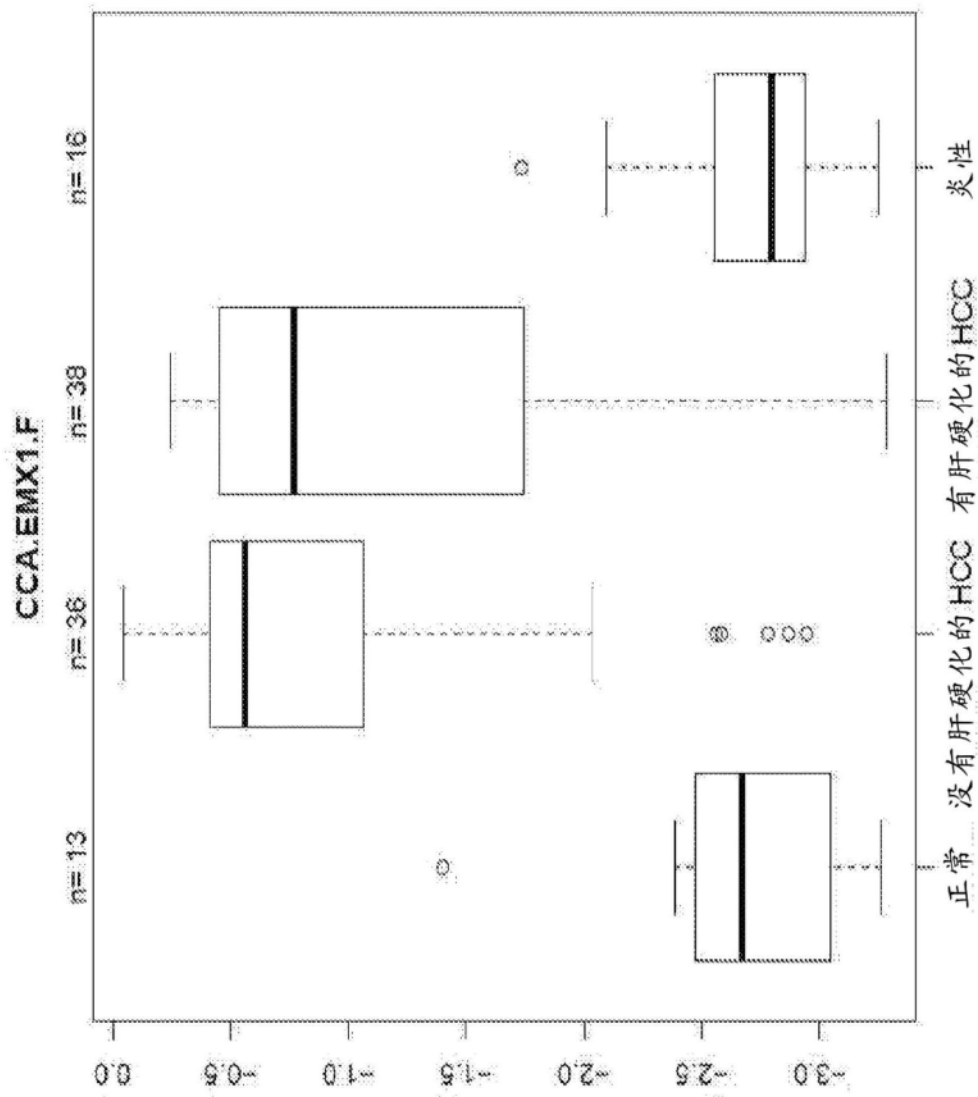


图4R

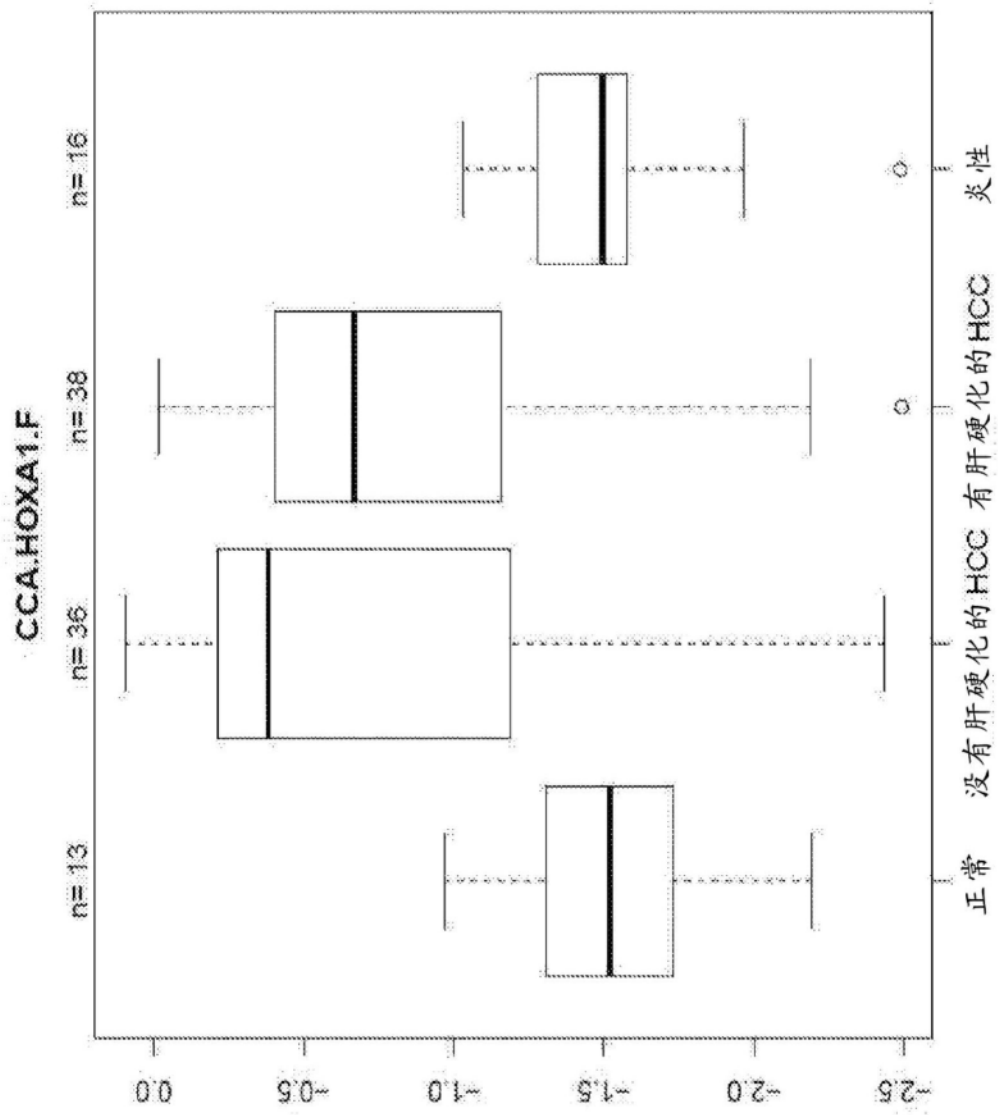


图4S

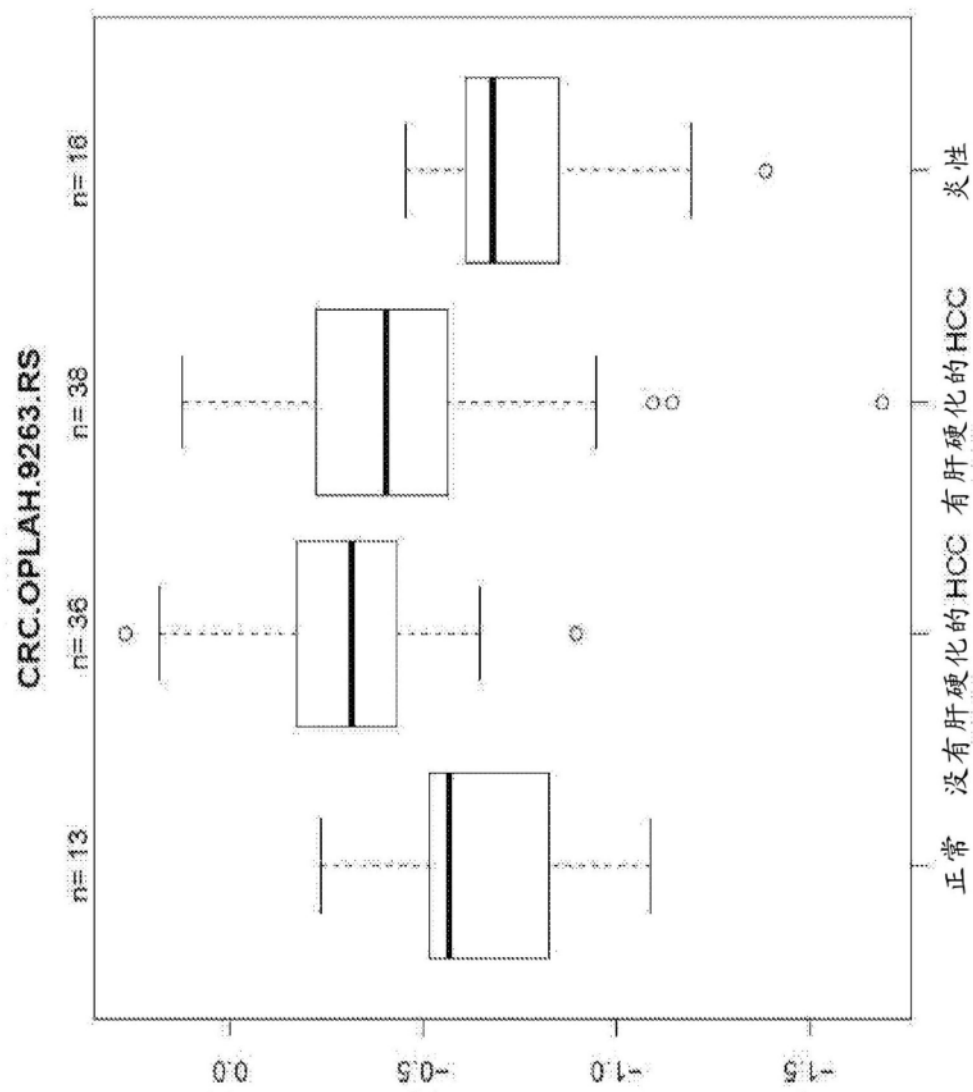


图4T

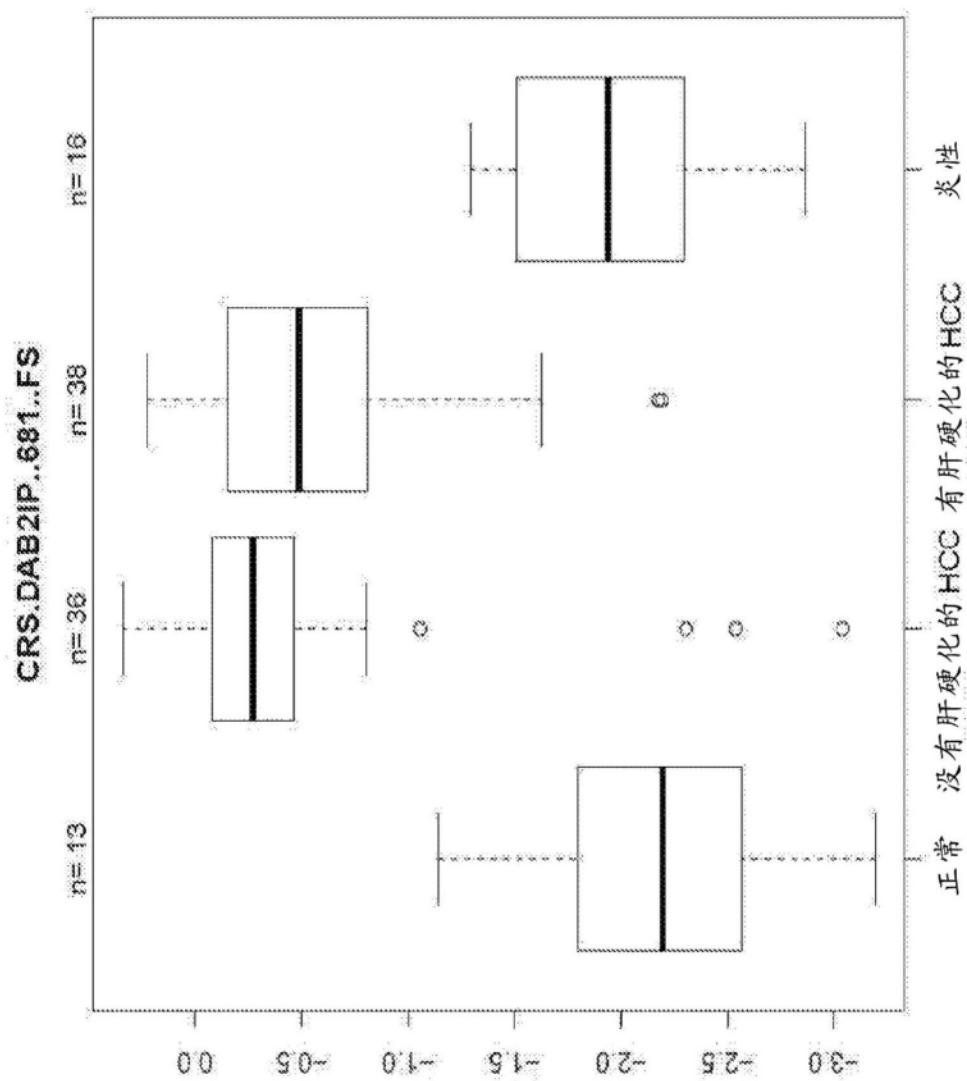


图4U

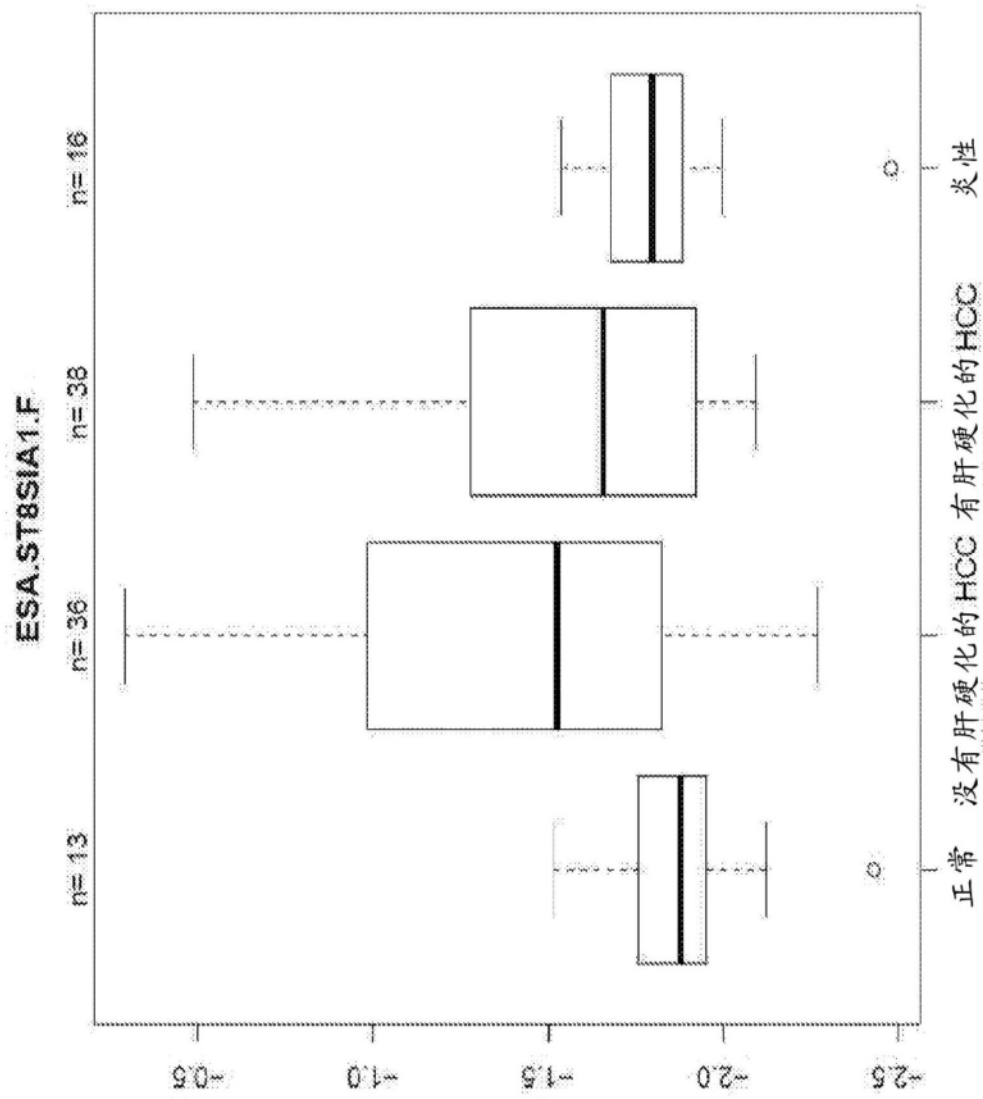


图4V

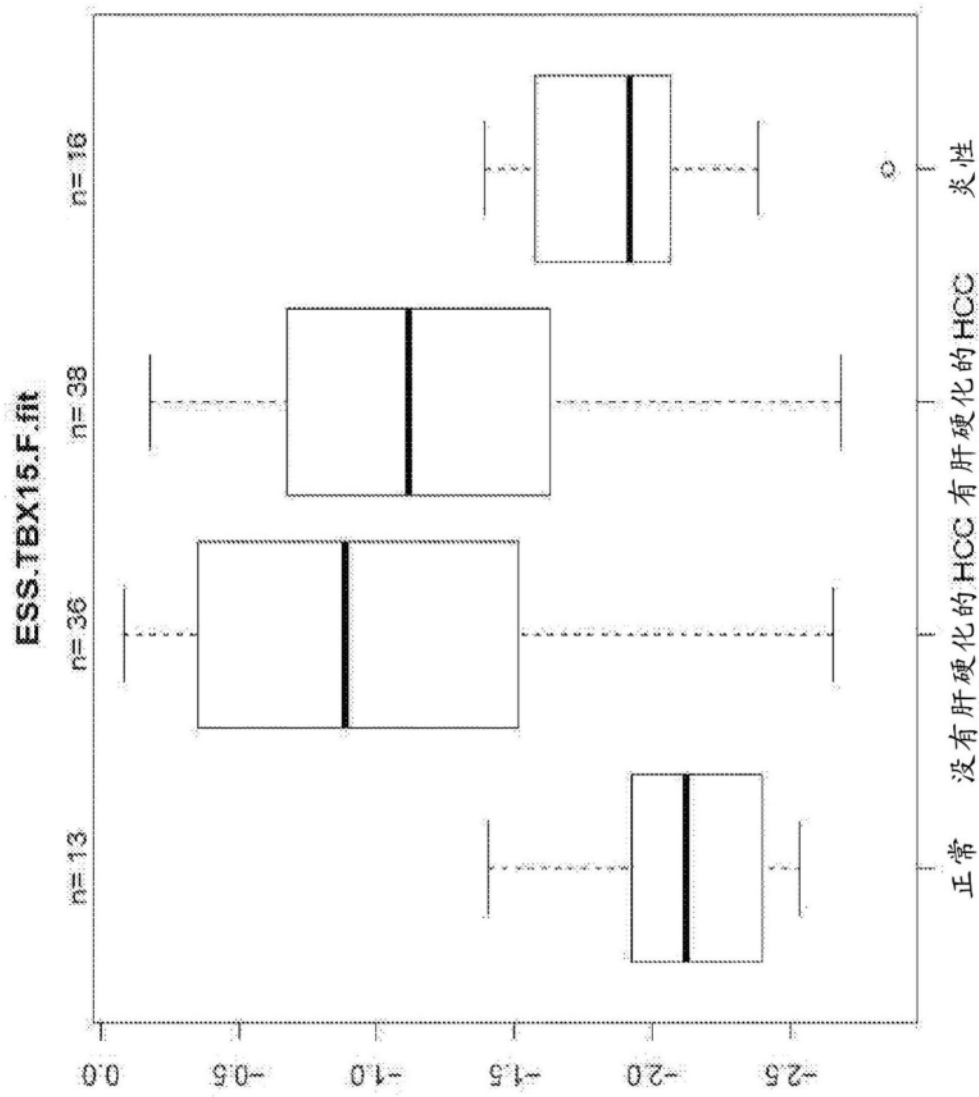


图4W

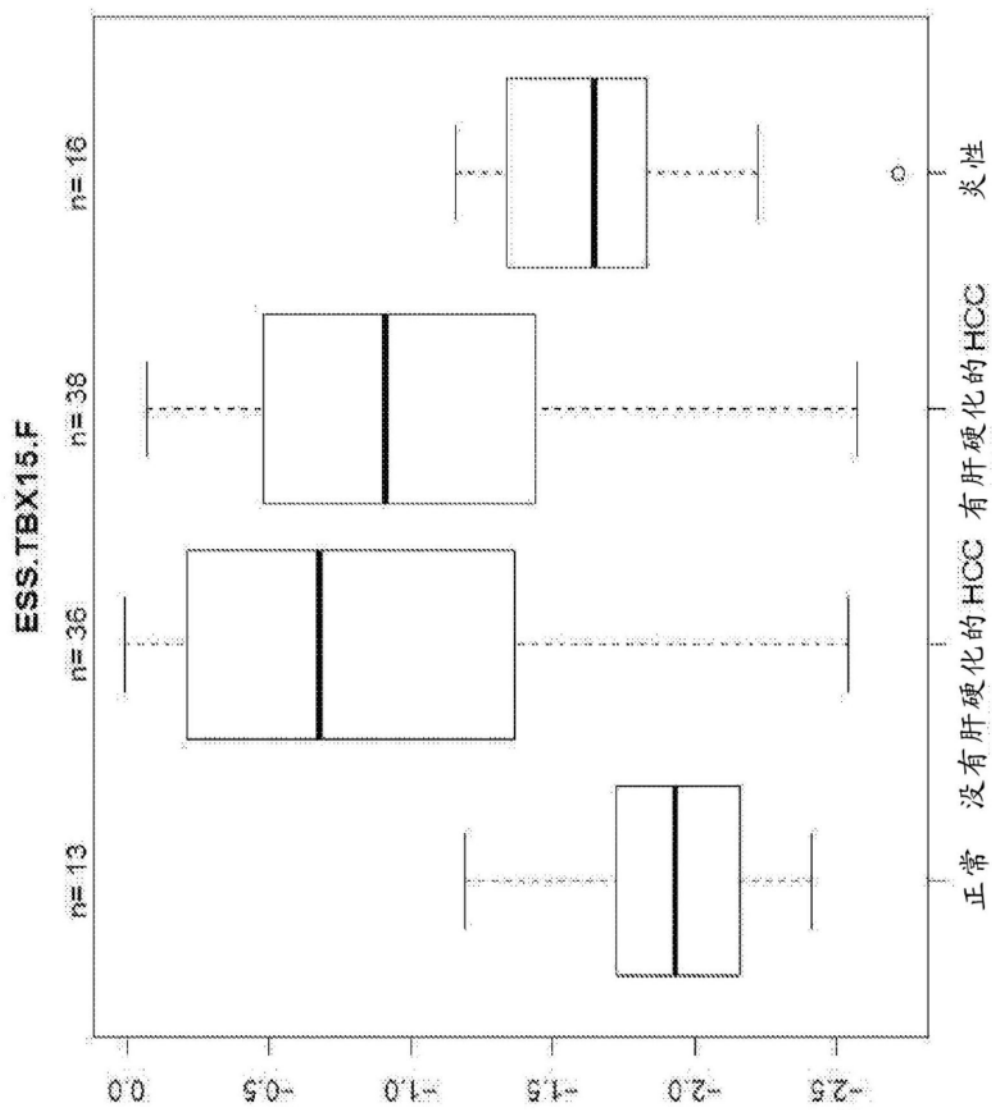


图4X

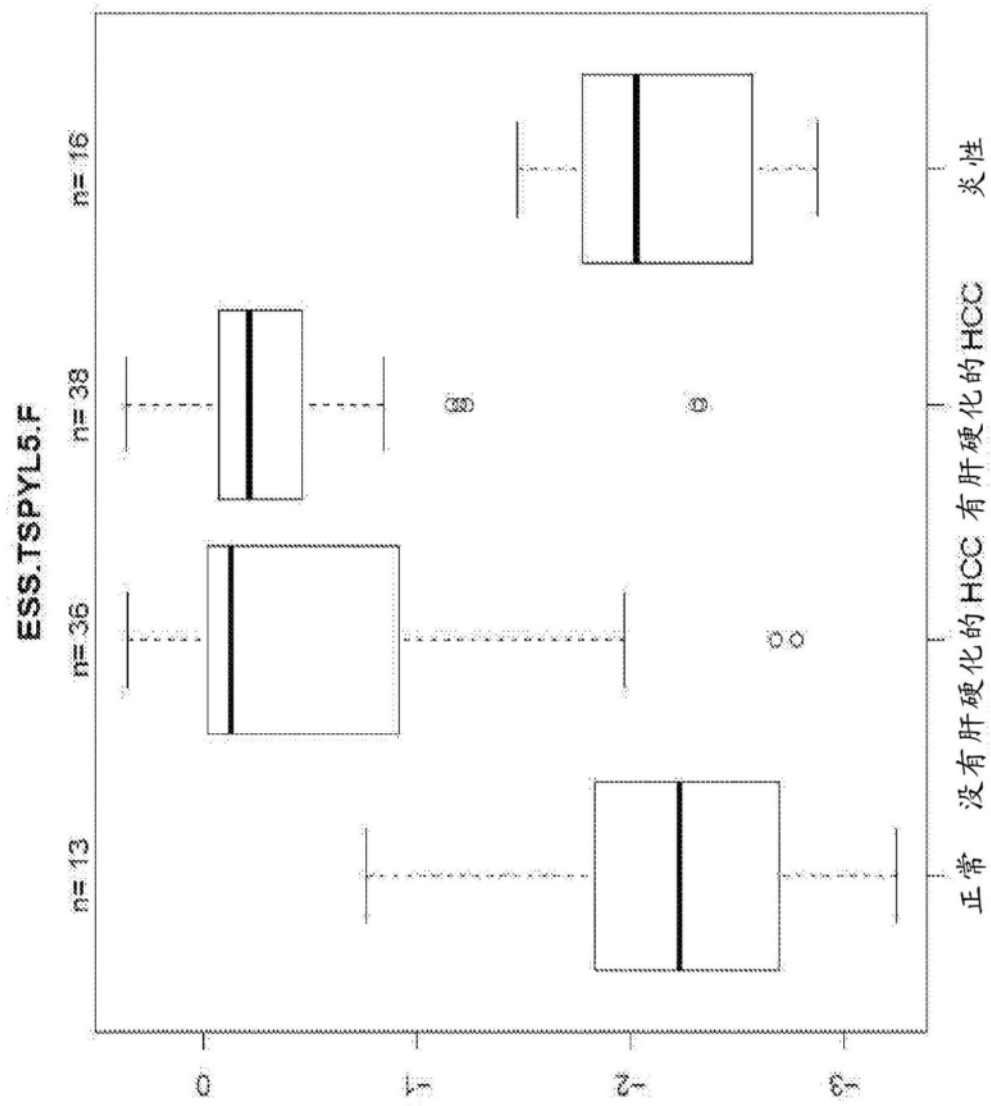


图4Y

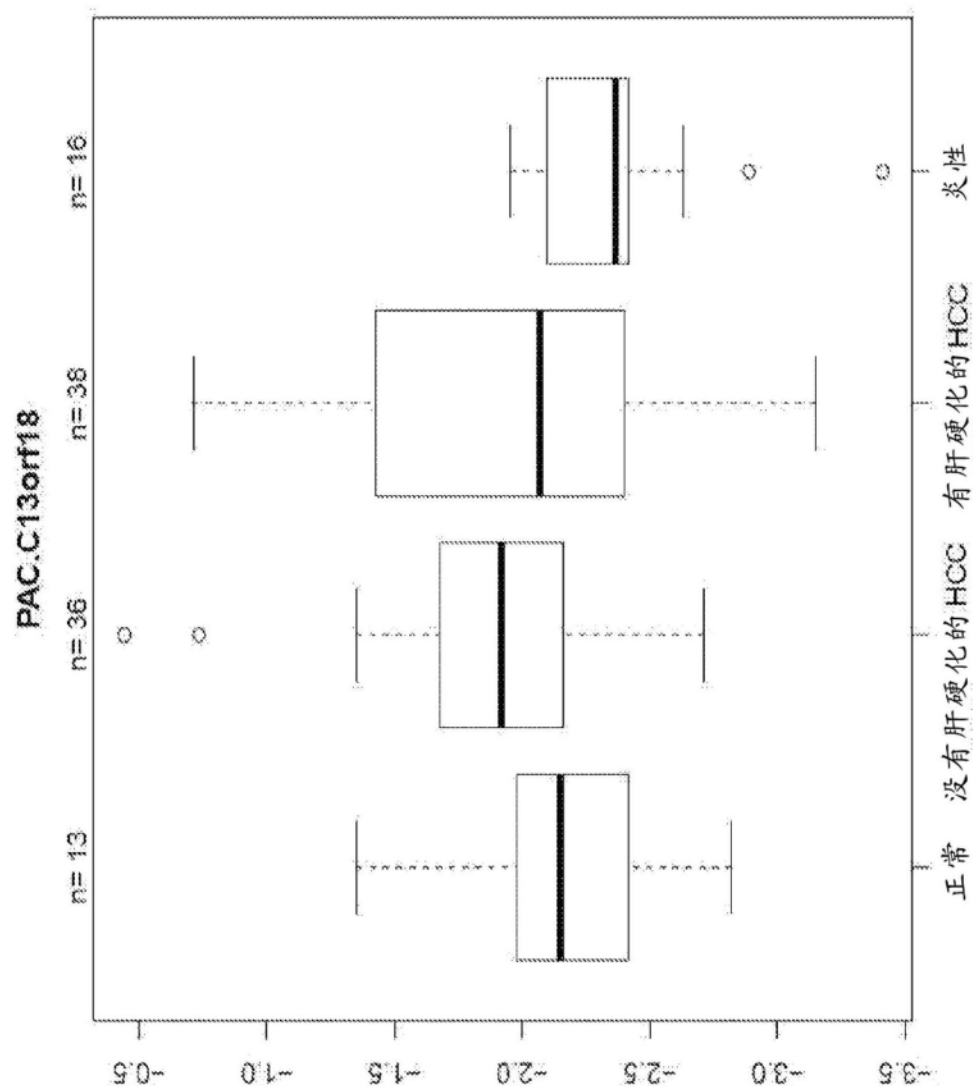


图4Z

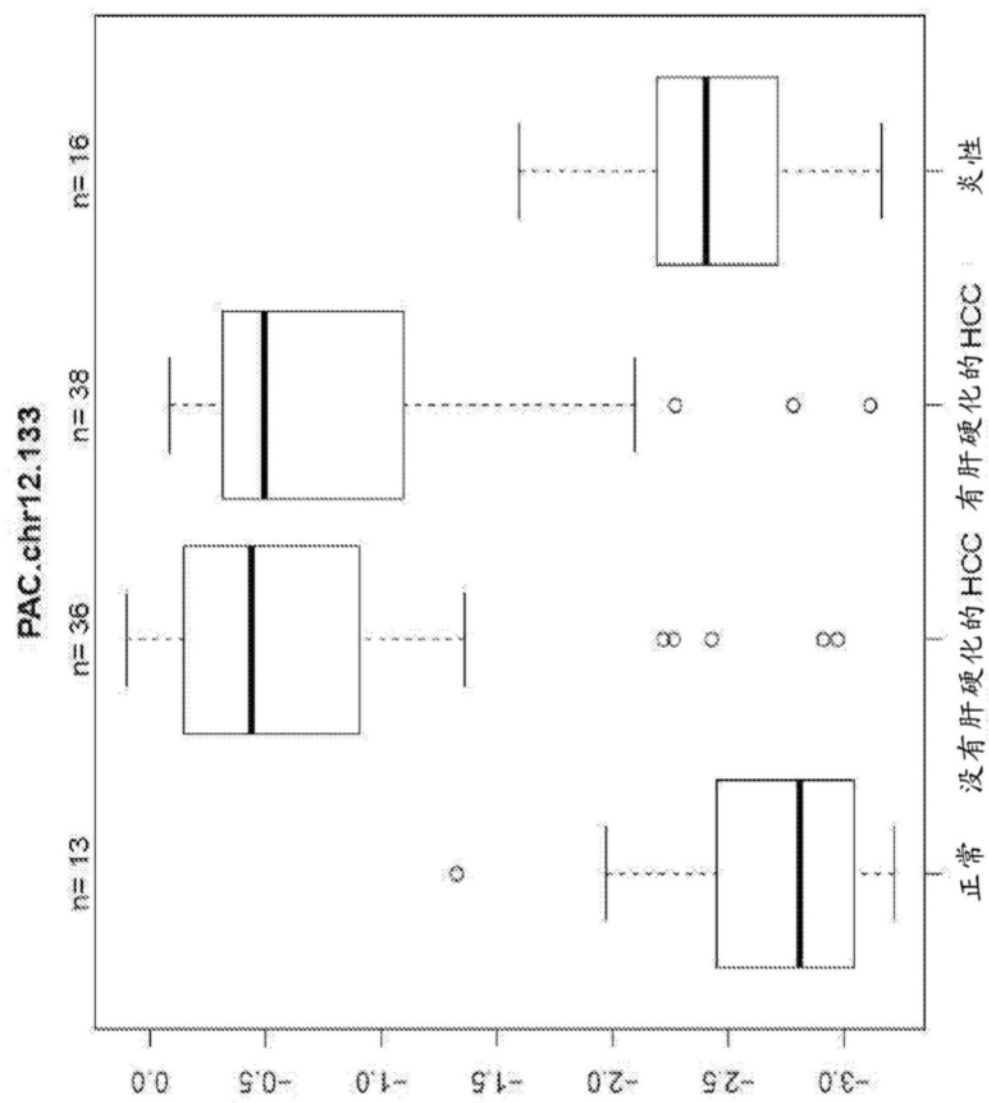


图4AA

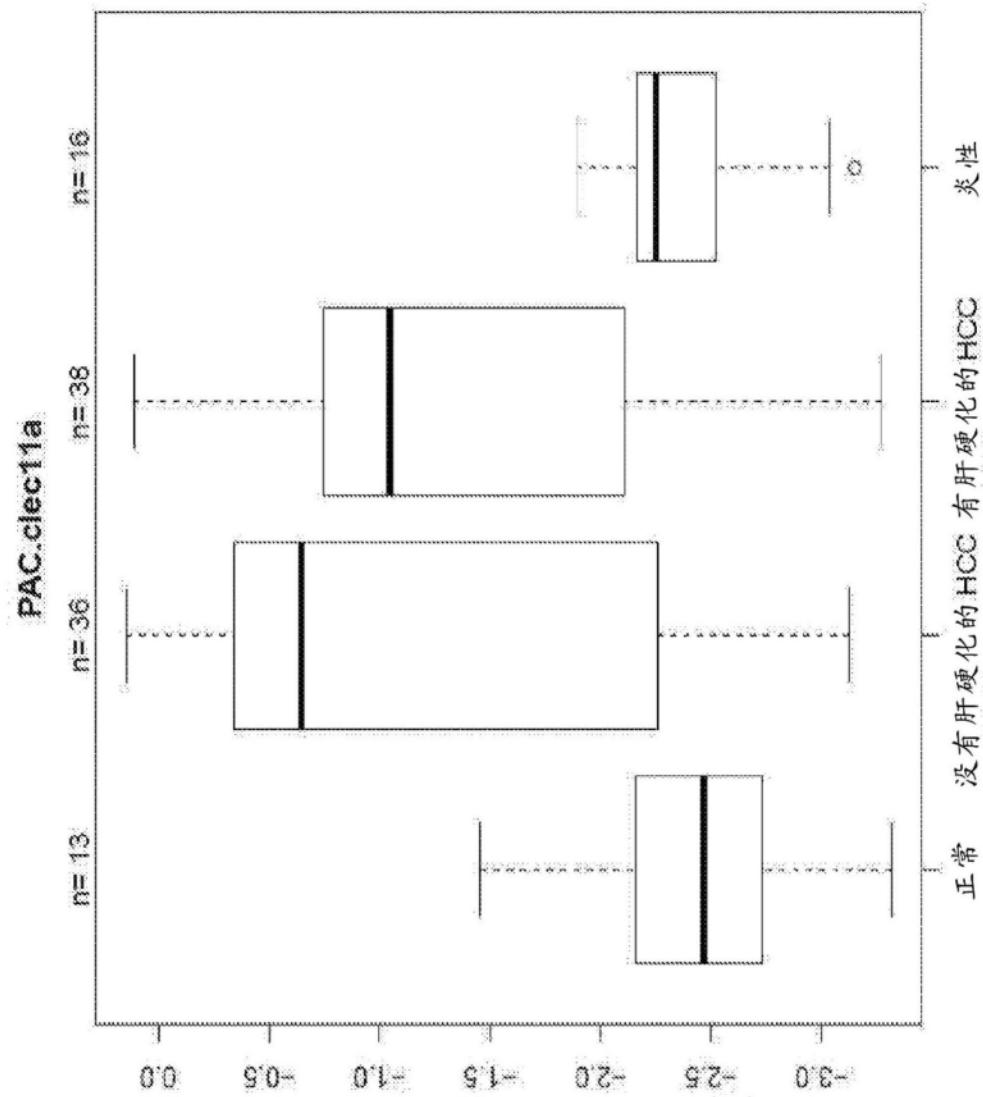


图4BB

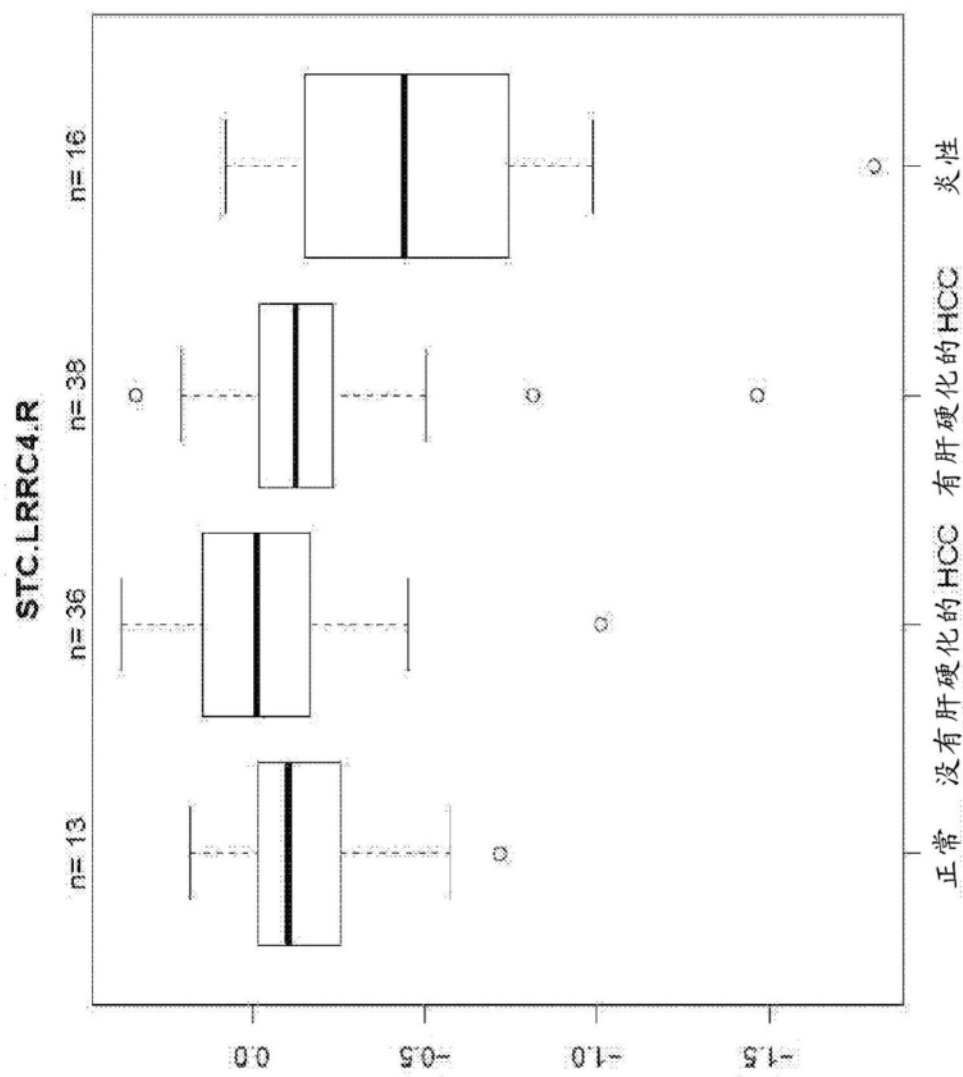


图4CC

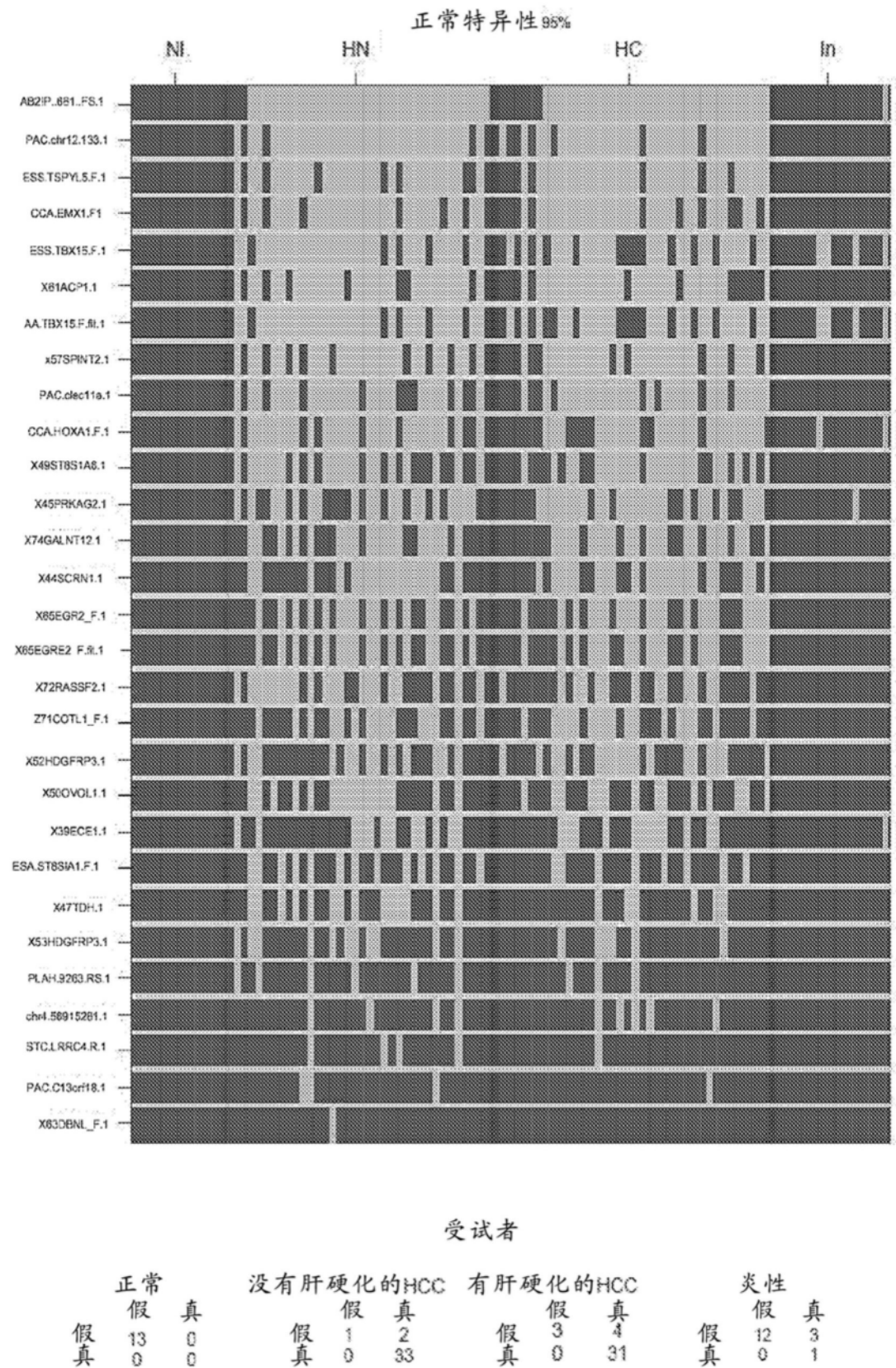
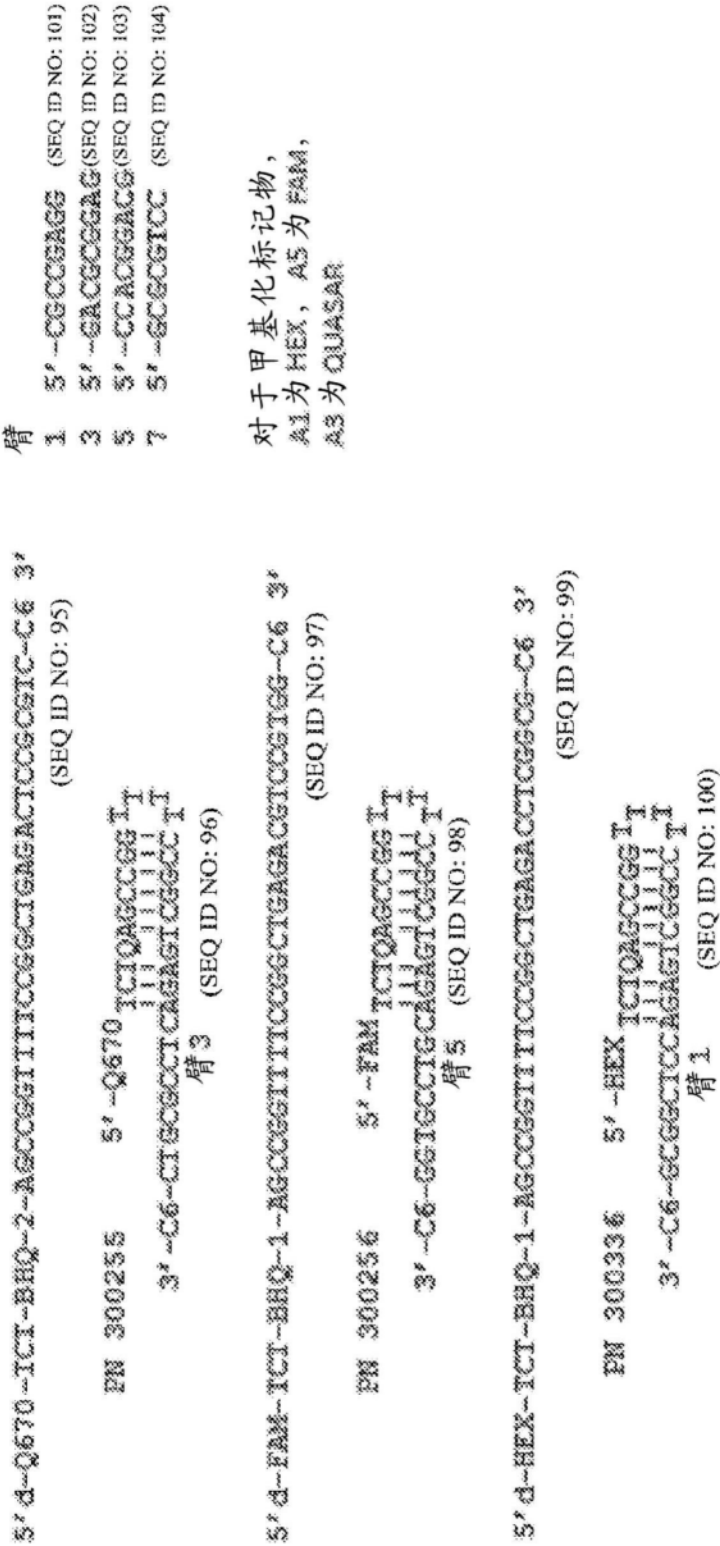


图5



对于甲基化标记物，
A1为 HEX，A5为 FAM，
A3为 QUASAR

图6

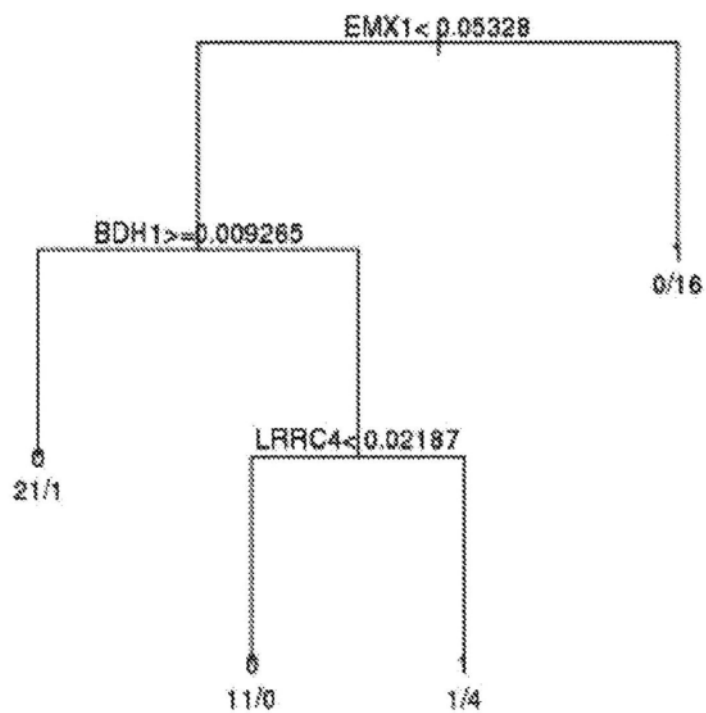


图7A

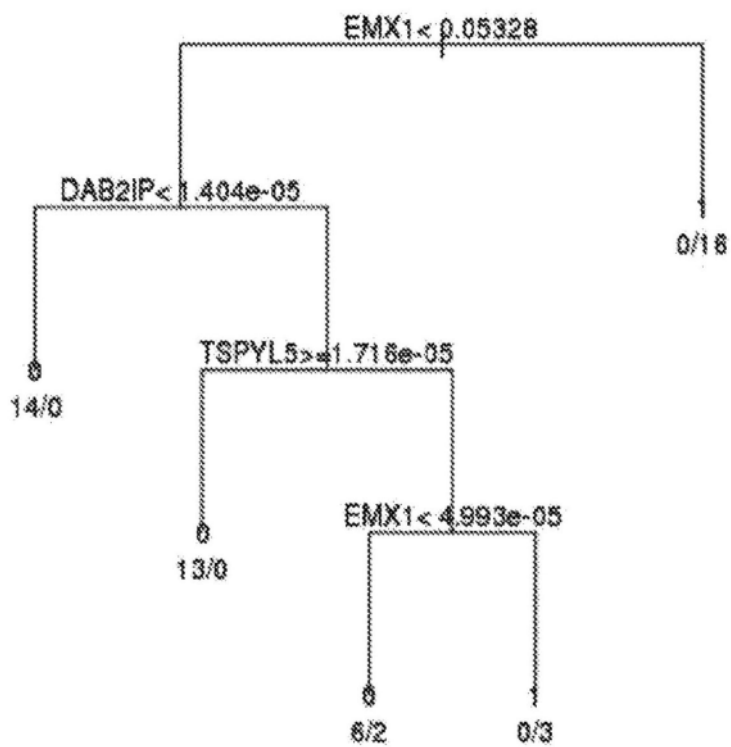


图7B

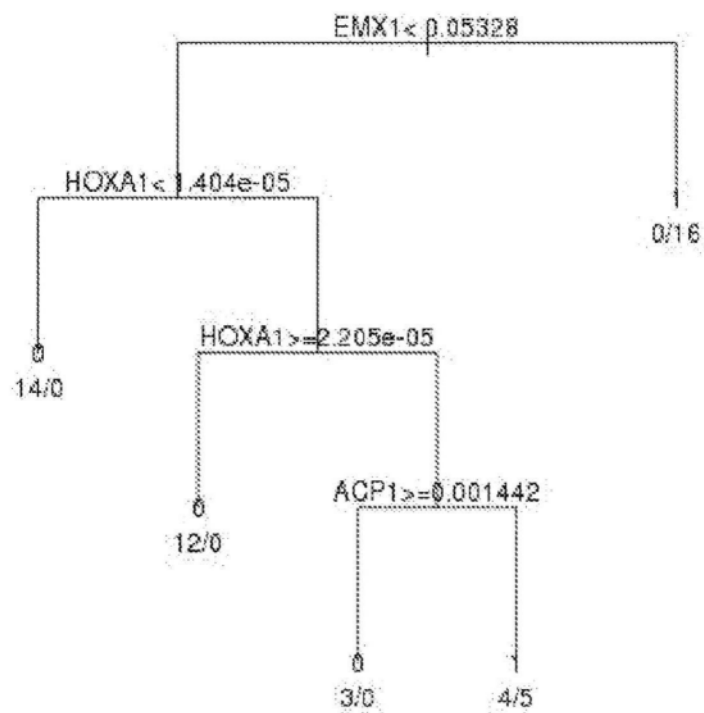


图7C

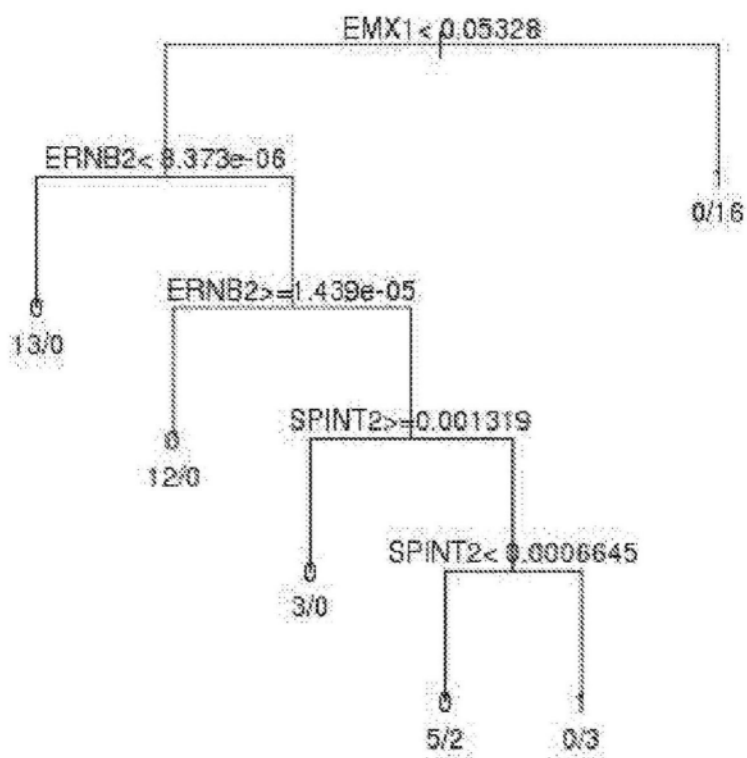


图7D

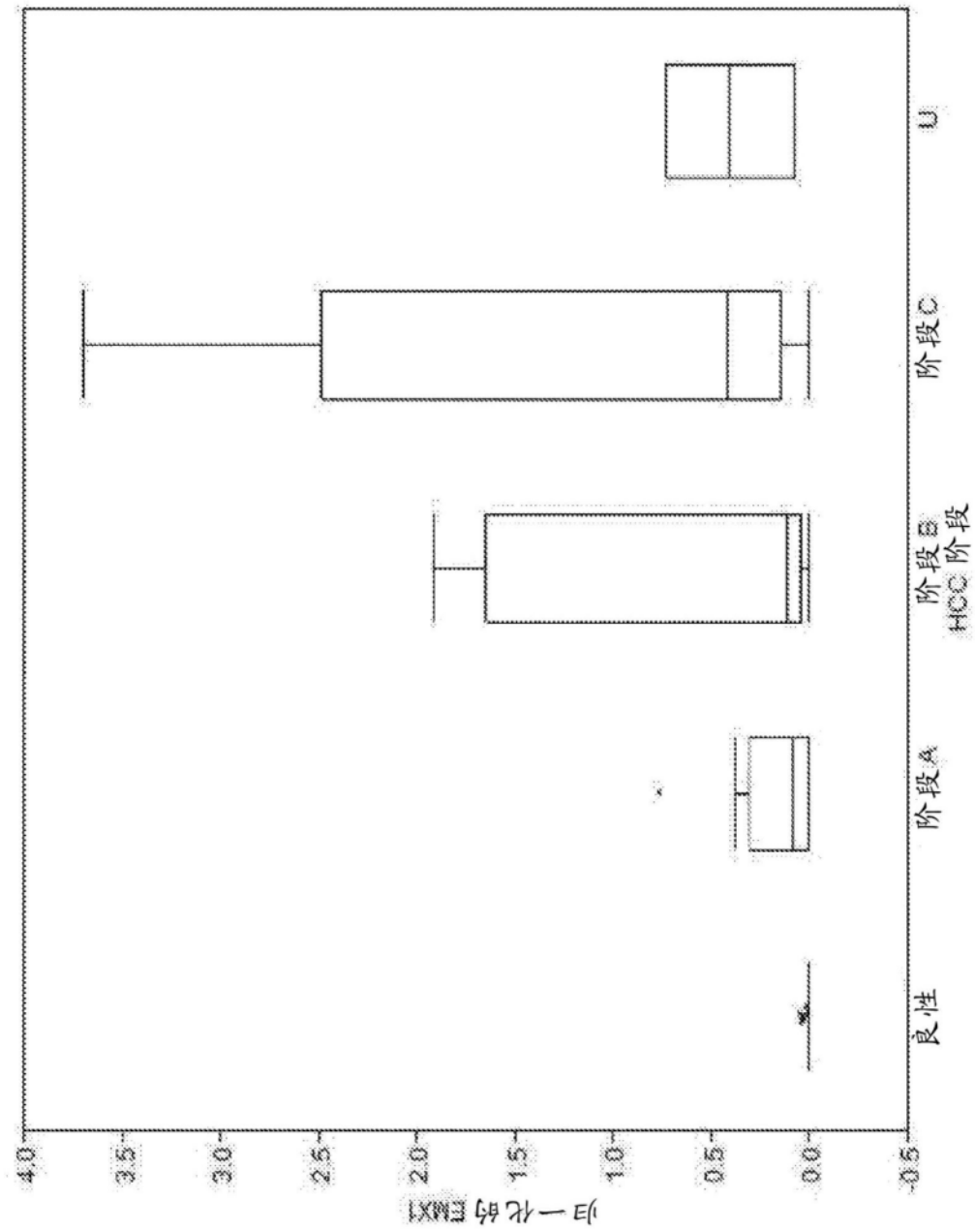


图8

相对重要性排序

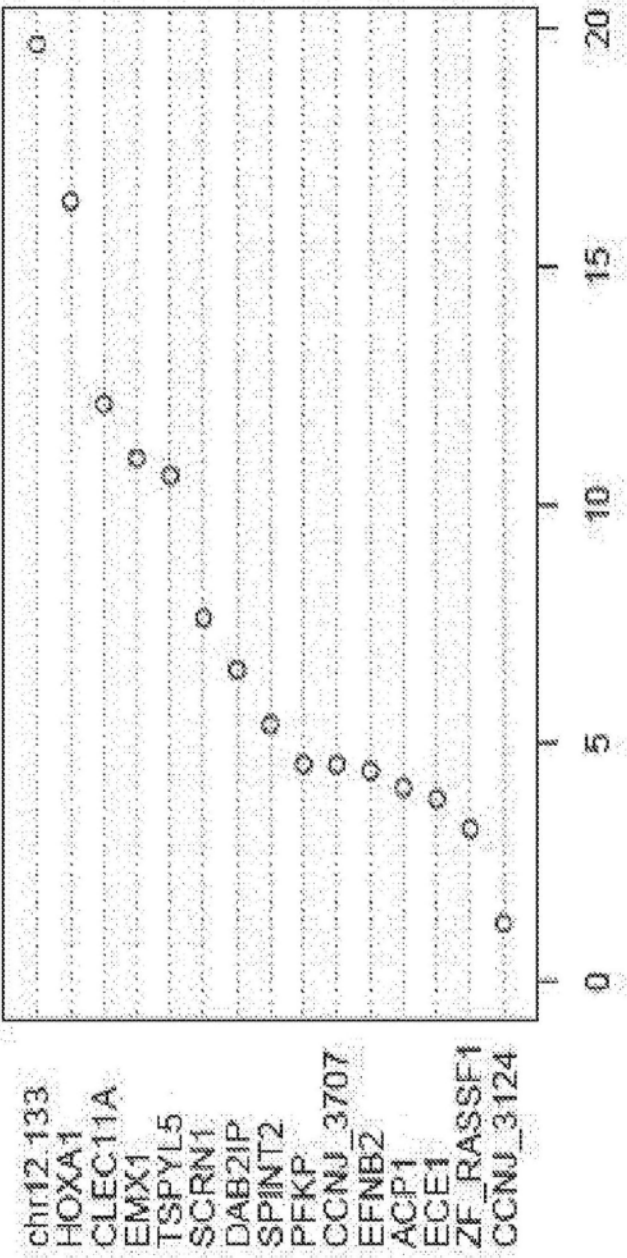


图9