



## (12)发明专利申请

(10)申请公布号 CN 106755464 A

(43)申请公布日 2017.05.31

---

(21)申请号 201710020252.4

(22)申请日 2017.01.11

(71)申请人 上海易毕恩基因科技有限公司

地址 201203 上海市浦东新区爱迪生路328  
号2楼201-2室

(72)发明人 陆星宇 宋艳群

(74)专利代理机构 北京市金杜律师事务所  
11256

代理人 孟凡宏

(51)Int.Cl.

C12Q 1/68(2006.01)

权利要求书2页 说明书9页 附图2页

---

### (54)发明名称

用于筛选肠癌和/或胃癌的基因标志物的方法、用该方法筛选的基因标志物及其用途

### (57)摘要

本发明涉及用于检测肠癌或胃癌的基因标志物及其用途。本发明还涉及使用所述基因标志物对肠癌或胃癌进行检测的方法。

1. 用于检测肠癌的基因标志物,包括一个或多个选自以下的基因:ADAM金属肽酶域20 (ADAM20)、F盒和富亮氨酸重复蛋白7 (FBXL7)、卵泡抑素 (FST)、TP53凋亡效应器 (PERP)、普列克底物蛋白同源相似域家族A成员3 (PHLDA3)、Runt相关转录因子1移动伴侣1 (RUNX1T1)、互养蛋白  $\gamma$  2 (SNTG2)、精子相关抗原4 (SPAG4)、硫酸酯酶1 (SULF1) 和NME/NM23核苷二磷酸激酶 (NME3)。

2. 权利要求1所述的基因标志物,包括ADAM20、FBXL7、FST、PERP、PHLDA3、RUNX1T1、SNTG2、SPAG4、SULF1和NME3。

3. 权利要求1或2所述的基因标志物在用于检测肠癌的方法中的用途。

4. 用于检测胃癌的基因标志物,包括一个或多个选自以下的基因:Rho GTP酶激活蛋白28 (ARHGAP28)、BMP结合内皮调控者 (BMPER)、染色体9开放读码框92 (C9orf92)、钙依赖分泌激活者2 (CADPS2)、钙粘蛋白11 (CDH11)、F盒和富亮氨酸重复蛋白7 (FBXL7)、间质同源框2 (MEOX2)、氧化一氮合成酶1 (NOS1)、抑瘤素M受体 (OSMR)、细胞膜调控蛋白paralemmin2 (PALM2)、磷酸二酯酶10A (PDE10A)、RNA结合模体单链相互作用蛋白3 (RBMS3)、硫酸酯酶1 (SULF1)、wntless Wnt配体分泌调节者 (WLS)、Wilms瘤1作用蛋白 (WTIP)、锌指蛋白518B (ZNF518B)、锌指蛋白714 (ZNF714) 和RadS2模体包含1 (RDM1)。

5. 权利要求4所述的基因标志物,包括ARHGAP28、BMPER、C9orf92、CADPS2、CDH11、FBXL7、MEOX2、NOS1、OSMR、PALM2、PDE10A、RBMS3、SULF1、WLS、WTIP、ANF518B、ZNF714和RDM1。

6. 权利要求4或5所述的基因标志物在用于检测胃癌的方法中的用途。

7. 一种用于检测肠癌的方法,包括以下步骤:

(a) 测定正常样品和受试者样品中权利要求1或2所述的基因标志物的5-羟甲基胞嘧啶 (5-hmC) 的含量;

(b) 用正常样品中所述基因标志物的5-hmC含量作为参照,将受试者样品中对应的基因标志物的5-hmC含量标准化;

(c) 对步骤(b)中经标准化的所述基因标志物的5-hmC含量进行数学关联,并获得评分;和

(d) 根据所述评分获得检测结果。

8. 一种用于检测胃癌的方法,包括以下步骤:

(a) 测定正常样品和受试者样品中权利要求4或5所述的基因标志物的5-hmC含量;

(b) 用正常品中所述基因标志物的5-hmC含量作为参照,将受试者样品中对应的基因标志物的5-hmC含量标准化;

(c) 对经标准化的所述基因标志物的5-hmC含量进行数学关联,并获得评分;和

(d) 根据所述评分获得检测结果。

9. 权利要求7或8所述的方法,其中步骤(a)是测定所述基因标志物全长或其片段上的5-hmC的含量。

10. 权利要求7或8所述的方法,其中所述样品是来自正常人或受试者体液中游离的DNA片段,或来源于细胞器、细胞以及组织中的完整基因组DNA。

11. 权利要求10所述的方法,其中所述体液是血液、尿液、汗液、痰液、粪便、脑脊液、腹水、胸水、胆汁或胰腺液。

12. 一种用于检测肠癌的试剂盒,包括:

- (a) 用于测定权利要求1或2所述的基因标志物的5-hmC含量的试剂;和
- (b) 说明书。

13. 一种用于检测胃癌的试剂盒,包括:

- (a) 用于测定权利要求4或5所述的基因标志物的5-hmC含量的试剂;和
- (b) 说明书。

## 用于筛选肠癌和/或胃癌的基因标志物的方法、用该方法筛选的基因标志物及其用途

### 技术领域

[0001] 本发明涉及通过高通量测序筛选基因标志物的领域。具体地,本发明涉及通过高通量测序对肠癌和/或胃癌的基因标志物进行筛选的方法,以及利用该方法筛选出的基因标志物及其用途。

### 背景技术

[0002] 随着生活环境和习惯的改变,近些年来癌症的发病率在世界范围内逐年升高,人们对其关注也随之增加。胃癌和肠癌分别是我国第二大和第三大高发癌,属于发病率和病死率都非常高的恶性肿瘤,它们的发生和发展是一个多因素参与和多阶段累积致癌的复杂过程。本着现有通行的癌症早发现早治疗方案,越早期发现,越有可能早期控制甚至治愈癌症。如何对胃癌和肠癌进行早期诊断,及时治疗,正确地判断预后,逐渐受到人们重视。

[0003] 传统筛查方法有胃肠镜检查和大便隐血试验,但都有自身不可克服的缺点。胃肠镜加病理活检被认为是胃肠癌症筛查和诊断的金标准,但是因为镜检的侵入性和肠道准备的不适,很多病人不愿意接受肠胃镜检查。最近有研究表明肠镜对右侧结肠内常见的扁平息肉的诊断效果不好,肠镜筛查并不能降低右侧结肠癌的发病率。大便隐血试验是另一种临床上常用的大肠癌筛查方法,该法有完全无创和廉价等优点,但是其检测肿瘤的准确性低,通过免疫化学的改进方法也仅能检测到50-60%的大肠癌和30%左右的癌前腺瘤。并且,该检测方法假阳性率较高,故很难得到推广和普及。

[0004] 除以上传统筛查方法以外,人们还在不断探索其他方法,以提高早期诊断率,肿瘤标志物是其中之一。例如,通过DHPLC分析发现T1151A作为错配修复基因hMLH1上的一个多态位点,可作为胃及大肠肿瘤,尤其是低龄胃及大肠肿瘤高危人群筛选的候选指标(参见张晓梅等,《肿瘤》,2005,25(1):62-65)。采用CA19-9(糖类胃肠癌相关抗原)和SA(唾液酸)的肿瘤标志物联合检测对结/直肠癌和胃/贲门癌的检测阳性率分别可以达到76.7%和82.5%(参见蔡发成等,《实用医技杂志》,2005,12(3B):722-723)。此外,我国还开发了C12蛋白芯片检测系统,通过分析血清中CA19-9、神经元特异性烯醇化酶(NSE)、癌胚抗原(CEA)、糖原242(CA242)、糖原125(CA125)、糖原153(CA15-3)、甲胎蛋白(AFP)、铁蛋白、人绒毛膜促性腺激素(HCG)、人生长激素(HGH)中这十种肿瘤标志物的表达水平,以实现肿瘤的早期诊断和监测。然而,大多数发现的肿瘤标志物还停留在研究阶段,并没有发展到临床诊断阶段;或者已经应用在临床诊断上,但检测准确度或灵敏度不高。例如,研究发现,上述C12系统对胃肠癌患者的总体诊断率为39.21%,I、II、III、IV期患者的诊断率分别为13.73%、33.33%、38.30%、58.03%,表明C12检测系统对晚期胃肠癌的诊断有一定的价值,但对早期的敏感性不高(参见杨雪琴等,《南京医科大学学报(自然科学版)》,2008(10):1285-1289)。

[0005] 近期,在现有常规基因标记物的基础上又发展了另外两种肠癌的体外诊断方法:来自美国Exact Sciences公司的Cologuard技术和来自Epigenomics公司的Epi proColon

技术。前者主要检测粪便中的异常DNA变化及人类粪便中存在的潜血红蛋白；后者检测血液游离DNA中的SEPT9甲基化标记物。大规模临床实验表明，CoLoguard的检出率效果较好，但需要操作粪便，用户体验较差；而SEPT9甲基化属于液体活检，虽然用户体验较好，但是特异性和灵敏度都较差（分别为80%和72%）。

[0006] 因此迫切需要具有高特异性和高灵敏度并可临床检测使用的胃肠癌的基因标志物，使能够以无创或微创、快捷的方法有效诊断胃肠癌，以改善病人接受长期监测的意愿。

## 发明内容

[0007] 发明人通过对正常样品和肠癌或胃癌样品进行高通量测序，并对其中各基因上的5-羟甲基胞嘧啶(5-hmC)含量进行分析，出乎意料地发现了多个极具信息的可用于检测肠癌或胃癌的基因标志物。

[0008] 因此，本发明的第一个方面涉及用于检测肠癌的基因标志物，包括一个或多个选自以下的基因：ADAM金属肽酶域20(ADAM20)、F盒和富亮氨酸重复蛋白7(FBXL7)、卵泡抑素(FST)、TP53凋亡效应器(PERP)、普列克底物蛋白同源相似域家族A成员3(PHLDA3)、Runt相关转录因子1移动伴侣1(RUNX1T1)、互养蛋白 $\gamma$ 2(SNTG2)、精子相关抗原4(SPAG4)、硫酸酯酶1(SULF1)和NME/NM23核昔二磷酸激酶(NME3)。优选的，所述基因标志物包括至少两个、至少三个、至少四个、至少五个、至少六个、至少七个、至少八个、至少九个或十个选自以下的基因：ADAM20、FBXL7、FST、PERP、PHLDA3、RUNX1T1、SNTG2、SPAG4、SULF1和NME3。更优选的，所述基因标志物包括ADAM20、FBXL7、FS孔PERP、PHLDA3、RUNX1T1、SNTG2、SPAG4、SULF1和NME3。

[0009] 本发明的第二个方面涉及用于检测胃癌的基因标志物，包括一个或多个选自以下的基因：Rho GTP酶激活蛋白28(ARHGAP28)、BMP结合内皮调控者(BMPER)、染色体9开放读码框92(C9orf92)、钙依赖分泌激活者2(CADPS2)、钙粘蛋白11(CDH11)、F盒和富亮氨酸重复蛋白7(FBXL7)、间质同源框2(MEOX2)、氧化一氮合成酶1(NOS1)、抑瘤素M受体(OSMR)、细胞膜调控蛋白paralemmin2(PALM2)、磷酸二酯酶10A(PDE10A)、RNA结合模体单链相互作用蛋白3(RBMS3)、硫酸酯酶1(SULF1)、wntless Wnt配体分泌调节者(WLS)、Wilms瘤1作用蛋白(WTIP)、锌指蛋白518B(ZNF518B)、锌指蛋白714(ZNF714)和Rad52模体包含1(RDM1)。优选的，所述基因标志物包括至少两个、至少三个、至少四个、至少五个、至少六个、至少七个、至少八个、至少九个、至少十个、至少十一个、至少十二个、至少十三个、至少十四个、至少十五个、至少十六个、至少十七个或十八个选自以下的基因：ARHGAP28、BMPER、C9orf92、CADPS2、CDH11、FBXL7、MEOX2、NOS1、OSMR、PALM2、PDE10A、RBMS3、SULF1、WLS、WTIP、ANF518B、ZNF714和RDM1。更优选的，所述基因标志物包括ARHGAP28、BMPER、C9orf92、CADPS2、CDH11、FBXL7、MEOX2、NOS1、OSMR、PALM2、PDE10A、RBMS3、SULF1、WLS、WTIP、ANF518B、ZNF714和RDM1。

[0010] 本发明还涉及上述基因标志物在检测肠癌或胃癌中的用途。本发明还涉及利用上述基因标志物进行肠癌或胃癌检测的试剂盒，其包括用于测定上述基因标志物的5-hmC含量的试剂和说明书。

[0011] 本发明的第三个方面涉及用于检测胃癌或肠癌的方法，包括以下步骤：

[0012] (a) 测定正常样品和受试者样品中本发明所述的基因标志物的5-hmC的含量；

[0013] (b) 用正常样品中所述基因标志物的5-hmC含量作为参照，将受试者样品中对应的

基因标志物的5-hmC含量标准化；

[0014] (c) 对经标准化的所述基因标志物的5-hmC含量进行数学关联,并获得评分;和

[0015] (d) 根据所述评分获得检测结果。

[0016] 在一个实施方案中,所述样品是受试者或正常人体液中游离的DNA片段,或来源于细胞器、细胞以及组织中的完整基因组DNA。其中,体液是血液、尿液、汗液、痰液、粪便、脑脊液、腹水、胸水、胆汁、胰腺液等。

[0017] 在一个实施方案中,本发明所述的基因标志物的5-hmC含量可通过本领域技术人员已知的任何方法进行测定,例如包括但不限于,葡糖基化法、限制性内切酶法、化学标记法、与高通量测序方法联用的沉淀法、单分子实时测序法(SMRT)、氧化重亚硫酸盐测序法(OxBS-Seq)等。葡糖基化法的原理是采用T4噬菌体 $\beta$ -葡萄糖转移酶( $\beta$ -GT),在葡萄糖供体底物尿核苷二磷酸葡萄糖(UDP-Glu)存在下,将葡萄糖转移至羟基位置,从而生成 $\beta$ -葡萄糖基-5-羟甲基胞嘧啶(5-ghmC)。同时可采用同位素标记底物进行定量。在葡糖基化法基础上进一步发展出限制性内切酶法和化学标记法。限制性内切酶法的原理是:葡糖基化反应改变了一些限制性内切酶的酶切特性。甲基化依赖的限制性内切酶MspI和HpaII可识别同样的序列(CCGG),但它们对甲基化状态的敏感性是不同:MspI识别并切割5-甲基胞嘧啶(5-mC)和5-hmC,但不能切割5-ghmC;HpaII只切割完全未修饰的位点,胞嘧啶上的任何修饰(5-mC、5-hmC、5-ghmC)均阻碍切割。若CpG位点含有5-hmC,那么糖基化、酶解之后能检测到条带,未糖基化对照反应中没有条带;同时可采用qPCR进行定量分析。另外,其他限制性内切酶也同样存在阻碍5-ghmC酶切的情况,可应用于5-hmC检测(如:GmrSD,MspJI,PvuRts1I,TaqI等)。化学标记法的原理是:将酶反应底物上的葡萄糖进行化学修饰转变成UDP-6-N3-glucose,将6-N3-glucose转移到羟甲基位置,生成N3-5ghmC。随后,通过点击化学方法在每个5-hmC上添加一分子生物素,结合下一代高通量DNA测序技术或单分子测序技术,可分析5-hmC在基因组DNA中的分布情况。沉淀法是将5-hmC用特殊方式修饰后再将其特异性地从基因组DNA中捕获下来,并进行测序分析。氧化重亚硫酸盐测序法是首个以单碱基分辨率对5-hmC进行定量测序的方法。首先将5-hmC进行K<sub>2</sub>Cr<sub>2</sub>O<sub>7</sub>氧化处理,生成5-甲酰胞嘧啶(5fC),然后采用重亚硫酸盐测序。在此过程中,5-hmC先氧化为5fC,而后脱氨形成U。通常,同时采用多种检测方法对5-hmC进行定量检测。

[0018] 在本发明的一个实施方案中,利用化学标记法结合高通量测序来测定本发明的基因标志物的5-hmC含量。在该具体的实施方案中,测定本发明的基因标志物的5-hmC含量的方法包括以下步骤:将来自肠癌或胃癌患者和正常人的样品的DNA片段化;将所述片段化的DNA末端修复并末端补齐;将末端补齐的DNA与测序接头连接,获得连接产物;通过标记反应对连接产物中的5-羟甲基胞嘧啶进行标记;富集含有5-羟甲基胞嘧啶标记的DNA片段,获得富集产物;对富集产物进行PCR扩增,获得测序文库;对测序文库进行高通量测序,获得测序结果;根据测序结果确定5-羟甲基胞嘧啶在基因上的含量。其中,标记反应包括:i)利用糖基转移酶将带有修饰基团的糖共价连接到5-羟甲基胞嘧啶的羟甲基上,和ii)将直接或间接连有生物素的点击化学底物与带有修饰基团的5-羟甲基胞嘧啶反应。其中,步骤i)和步骤ii)可以按顺序进行,也可以在一个反应中同时进行。这种标记方法减少了测序所需的样本量,且5-羟甲基胞嘧啶上的生物素标签使其在测序中显示出更高的动力学信号,提高了核苷酸识别的准确性。在该实施方案中,所述糖基转移酶包括但不限于:T4噬菌体 $\beta$ -葡糖基

转移酶( $\beta$ -GT)、T4噬菌体 $\alpha$ -葡糖基转移酶( $\alpha$ -GT)及其具有相同或相似活性的衍生物、类似物、或重组酶;所述带有修饰基团的糖包括但不限于:带有叠氮修饰的糖类(例如6-N3-葡萄糖)或带有其他化学修饰(例如羰基、巯基、羟基、羧基、碳-碳双键、碳-碳三键、二硫键、胺基、酰胺基、双烯等)的糖类,其中优选带有叠氮修饰的糖类;所述用于间接连接生物素和点击化学底物的化学基团包括但不限于:羰基、巯基、羟基、羧基、碳-碳双键、碳-碳三键、二硫键、胺基、酰胺基、双烯。在该实施方案中,优选通过固相材料来富集含有5-hmC标记的DNA片段。具体地,可以通过固相亲和反应或其他特异性结合反应将含有5-羟甲基胞嘧啶标记的DNA片段结合在固相材料上,然后通过多次洗涤去除未结合的DNA片段。固相材料包括但不限于带有表面修饰的硅片或其他芯片,例如人工高分子小球(优选直径为1nm-100um)、磁性小球(优选直径为1nm-100um)、琼脂糖小球等(优选直径为1nm-100um)。固相富集中所用的洗涤液是本领域技术人员熟知的缓冲液,包括但不限于:含有Tris-HCl、MOPS、HEPES(pH=6.0-10.0,浓度在1mM到1M之间)、NaCl(0-2M)或表面活性剂如Tween20(0.01%-5%)的缓冲液。在该实施方案中,优选直接在固相上进行PCR扩增从而制备测序文库。如有需要,在固相上进行PCR扩增后,可以回收扩增产物后进行第二轮PCR扩增来制备测序文库。所述第二轮PCR扩增可用本领域技术人员已知的常规方法进行。任选地,在制备测序文库的过程中可进一步包括一个或多个纯化步骤。本领域技术人员知晓的或可商购的任何纯化试剂盒均可用于本发明。纯化方法包括但不限于:凝胶电泳切胶回收、硅胶膜离心柱法、磁珠法、乙醇或异丙醇沉淀法或其组合。任选地,在高通量测序之前,对测序文库进行质量检查。例如,对文库进行片段大小分析并使用qPCR方法对文库的浓度进行绝对定量。通过质量检查的测序文库可用于高通量测序。然后将一定数量(1-96个)含有不同barcode的文库按相同浓度混匀并根据二代测序仪的标准上机方法上机测序,获得测序结果。本领域已知的各种二代测序平台及其相关的试剂可用于本发明。

[0019] 在本发明的一个实施方案中,优选将测序结果与标准人类基因组参考序列进行比对,挑选出其中比对到本发明基因标志物上的序列,即选择比对位点与基因特征(如组蛋白修饰位点、转录因子结合位点、基因外显子内含子区域以及基因启动子等)重合区域的读段数量,以代表5-hmC在该基因上的修饰水平,从而测定5-hmC在该基因标志物上的含量。优选在进行比对前,首先将测序结果清除低质量测序位点,其中衡量测序位点质量的因素包括但不限于:碱基质量、reads质量、GC含量、重复序列和Overrepresented序列数量等。该步骤中涉及的各种比对软件和分析方法是本领域已知的。

[0020] 在本发明的一个实施方案中,测定基因标志物的5-hmC含量是指测定该基因标志物全长上的5-hmC含量或测定该基因标志物上某一片段的5-hmC含量或其组合。

[0021] 根据本发明,在测定各基因标志物上5-hmC含量之后,用正常样品中所述基因标志物的5-hmC含量作为参照,将受试者样品中对应的基因标志物的5-hmC含量标准化。举例而言,正常样品和受试者样品中同一基因标志物的5-hmC含量分别为X和Y,则受试者样品中该基因标志物的标准化5-hmC含量为Y/X。

[0022] 根据本发明,在数据标准化后,对各基因标志物的标准化5-hmC含量进行数学关联以获得评分,从而根据所述评分获得检测结果。如本文所用,“数学关联”是指将来自生物样品的基因标志物的5-hmC含量与肠癌或胃癌诊断结果相关联的任何计算方法或机器学习方法。本领域普通技术人员理解,可选择不同的计算方法或工具用于提供本发明的数学关联,

例如弹性网络正则化、决策树、广义线性模型、逻辑回归、最高分值对、神经网络、线性和二次判别式分析(LQA和QDA)、朴素贝叶斯、随机森林和支持向量机。

[0023] 在本发明的一个实施方案中,对各基因标志物的标准化5-hmC含量进行数学关联并获得评分的具体步骤如下:将各基因标志物的标准化5-hmC含量乘以加权系数,获得该基因标志物的预测因子t;将各基因标志物的预测因子t相加,获得总预测因子T;将总预测因子T经过Logistic转换获得评分P;若 $P > 0.5$ ,则该受试者样品患有肠癌或胃癌;若 $P < 0.5$ ,则该受试者样品为正常。本文所述的加权系数是指在考虑可能影响5-hmC含量的因素(例如受试者地域、年龄、性别、低于、吸烟史、饮酒史、家族史等)的情况下,通过各种高级统计分析方法获得的系数。

[0024] 与现有技术相比,本发明中用于检测肠癌和/或胃癌的方法是基于基因标志物上的5-hmC含量,因此可以使用更为广泛的DNA样品来源。因此,本发明中用于检测肠癌和/或胃癌的方法具有以下几个优点:(1)安全无创,即使无症状人群也对该检测接受度高;(2)DNA来源广泛,不存在影像学中的检测盲区;(3)准确性高,有较高的灵敏度和特异性;(4)操作方便,用户体验好,容易进行疾病动态监测。

#### 附图说明

[0025] 图1:用本发明的肠癌基因标志物区分第一批样品的结果。

[0026] 图2:用本发明的肠癌基因标志物区分第二批样品的结果。

[0027] 图3:用本发明的胃癌基因标志物区分第三批样品的结果。

[0028] 图4:用本发明的胃癌基因标志物区分第四批样品的结果。

#### 具体实施方式

[0029] 下面将参考附图并结合实施例来详细说明本发明,以使本领域的技术人员可以更好的理解本发明并能予以实施。需要说明的是,本领域的技术人员应该理解本发明的附图及其实施例仅仅是为了说明的目的,并不能对本发明构成任何限制。在不矛盾的情况下,本申请中的实施例及实施例中的特征可以相互组合。

[0030] 实施例1. 肠癌基因标志物的筛选

[0031] (1) 抽提血浆DNA:

[0032] 从来自15位肠癌患者和18位正常人的样品中分别抽提10ng血浆DNA。可利用本领域技术人员所熟知的任何适用于抽提血浆DNA的方法、和试剂进行此步骤。

[0033] (2) 将血浆DNA进行末端补齐、悬A并与测序接头连接:

[0034] 根据Kapa Hyper PerpKit说明书制备含有50uL血浆DNA、7uLEnd Repair&A-Tailing Buffer和3uL End Repair&A-Tailing Enzyme mix的反应混合液(总体积为60uL),在20°C温浴30分钟,然后在65°C温浴30分钟。在1.5mL低吸附EP管中配置以下连接反应混合物:5uL Nuclease free water,30uL Ligation Buffer以及10uL DNA Ligase。向45uL连接反应混合物中加入5uL的测序接头,混合,于20°C加热20分钟,然后保持于4°C。使用AmpureXP beads对反应产物进行纯化,用20uL含Tris-HCl(10mM,pH=8.0)及EDTA(0.1mM)的缓冲液进行洗脱获得最终的DNA连接样品。

[0035] (3) 标记5-羟甲基胞嘧啶:



[0036] 制备总体积为26uL的标记反应混合液：叠氮修饰的二磷酸尿苷葡萄糖（即UDP-N3-Glu,终浓度为50uM）、 $\beta$ -GT（终浓度为1uM）、 $Mg^{2+}$ （终浓度为25mM）、HEPES（pH=8.0,终浓度为50mM）和来自上述步骤的20uL DNA。将混合液在37℃温浴1小时。取出混合液，用AmpureXP beads纯化，获得纯化的20uL DNA。

[0037] 然后在上述纯化的20uL DNA中加入1uL连接有生物素的二苯基环辛炔（DBC0-Biotin），于37℃反应2小时，接着用AmpureXP beads纯化，获得纯化的标记产物。

[0038] (4) 固相富集含有标记的5-羟甲基胞嘧啶的DNA片段：

[0039] 首先，按以下步骤准备磁珠：取出0.5uL Clstreptadvin beads (life technology) 并加入100uL缓冲液（5mM Tris, pH=7.5, 1M NaCl, 0.02% Tween20），涡旋混合30秒，然后用100uL洗涤液（5mM Tris, pH=7.5, 1M NaCl, 0.02% Tween20）洗涤磁珠3次，最后加入25uL结合缓冲液（10mM Tris, pH=7.5, 2M NaCl, 0.04% Tween20或其他表面活性剂），并混合均匀。

[0040] 然后，在磁珠混合液中加入上述步骤获得的纯化的标记产物，并在旋转混合器中混合15min使其充分结合。

[0041] 最后，用100uL洗涤液（5mM Tris, pH=7.5, 1M NaCl, 0.02% Tween20）洗涤磁珠3次，离心去掉上清液，加入23.75uL不含核酸酶的水。

[0042] (5) PCR扩增：

[0043] 向上述步骤的最终体系中加入25uL的2X PCR master mix和1.25uL PCR引物（总体积为50uL），按照下述PCR反应循环的温度和条件进行扩增：

[0044]	98℃	45s	1 个循环
	98℃	15s	14 个循环
	60℃	30s	
	72℃	30s	
	72℃	1min	1 个循环

[0045] 将扩增产物用AmpureXP beads纯化，得到最终测序文库。

[0046] (6) 对测序文库进行质检后进行高通量测序：

[0047] 将获得的测序文库通过qPCR进行浓度测定，并用Agilent2100对文库中DNA片段大小含量进行确定。将通过质检的测序文库以相同浓度混合，用Illumina HiSeq 4000进行测序。

[0048] (7) 确定各基因标志物的5-hmC含量和加权系数

[0049] 将获得的测序结果进行初步质控评估，清除低质量测序位点后，将达到测序质量标准的读段利用Bowtie2工具与人类标准基因组参考序列进行比较。然后利用featureCounts和HtSeq-Count工具来统计读段数量以确定各基因标志物的5-hmC含量。同时利用高通量测序结果，将可能影响5-hmC含量的因素作为共变量，通过逻辑回归和弹性网络正则化获得各基因标志物的加权系数。结果如表1所示。

[0050] 表1：本发明的肠癌基因标志物的平均标准化5-hmC含量和加权系数

[0051]

正式符号	平均标准化 5-hmC 含量	p 值 (FDR)	加权系数	基因 ID	基因名称
ADAM20	1.34	6.52E-03	0.07	8747	ADAM 金属肽酶域 20
FBXL7	1.69	7.02E-08	0.53	23194	F 盒和富亮氨酸重复蛋白 7
FST	1.38	1.60E-03	1.37	10468	卵泡抑素
PERP	1.40	1.11E-05	4.11	64065	TP53 凋亡效应器
PHLDA3	1.38	1.75E-03	0.59	23612	普列克底物蛋白同源相似域家族 A 成员 3
RUNX1T1	1.68	3.15E-08	1.45	862	Runt 相关转录因子 1 移动伴侣 1
SNTG2	1.62	1.78E-09	0.080	54221	互养蛋白 $\gamma$ 2
SPAG4	1.30	2.21E-03	0.52	6676	精子相关抗原 4
SULF1	1.74	1.78E-09	2.63	23213	硫酸酯酶 1
NME3	0.75	1.91E-03	-0.25	4832	NME/NM23 核苷二磷酸激酶

[0052] 如上所述,平均标准化5-hmC含量是指肠癌样品中该基因标志物的平均5-hmC含量与正常样品中同一基因标志物的平均5-hmC含量之比。从表1可以看出,本发明的肠癌基因标志物的5-hmC含量在正常样品中和肠癌样品中存在显著差异,并且除NME3之外,其余基因标志物的5-hmC含量相对于正常人均显著增加。

[0053] 实施例2. 肠癌基因标志物的有效性

[0054] 本实施例验证本发明的肠癌基因标志物用于检测肠癌的有效性。

[0055] 根据实施例1的方法测定第一批59个样品(24例肠癌和35例对照)中本发明所述的10个肠癌基因标志物的5-hmC含量,并确定各基因标志物的加权系数。

[0056] 将各基因标志物的标准化5-hmC含量乘以其对应的加权系数,获得该基因标志物的预测因子 $t$ 后,将各基因标志物的预测因子 $t$ 相加,获得总预测因子 $T$ ,然后将总预测因子 $T$ 根据以下公式经过Logistic转换获得评分 $P$ :

$$[0057] \quad P = \frac{1}{1 + e^{-T}}$$

[0058] 若 $P > 0.5$ ,则该受试者样品患有肠癌;若 $P < 0.5$ ,则该受试者样品为正常。

[0059] 图1示出了根据本发明的方法区分该批样品的结果。如图1所示,本发明的方法能够达到83%的灵敏度和94%的特异性。发明人进一步用本发明的方法区分第二批69个样品(32例肠癌和37例对照),其结果表明该方法能够达到88%的灵敏度和89%的特异性(图2)。

[0060] 这些结果表明,与现有技术相比,根据本发明的方法能够以更高的灵敏度和特异性检测肠癌。

[0061] 实施例3. 胃癌基因标志物的筛选

[0062] 根据实施例1所述的方法筛选胃癌的基因标志物,唯一区别在于所用样品是来自7位胃癌患者和18位正常人的血浆游离DNA。筛选到的胃癌标志物如表2所示。

[0063] 表2: 本发明的胃癌基因标志物的平均标准化5-hmC含量和加权系数

[0064]

正式符号	平均标准化 5-hmC 含量	p 值 (FDR)	加权系数	基因 ID	基因名称
ARHGAP28	1.47	2.68E-04	0.14	79822	Rho GTP 酶激活蛋白 28
BMPER	1.45	2.31E-04	0.13	168667	BMP 结合内皮调控者
C9orf92	1.46	2.36E-04	0.09	1001293 85	染色体 9 开放读码框 92
CADPS2	1.62	2.68E-04	0.05	93664	钙依赖分泌激活者 2
CDH11	1.59	5.73E-04	0.07	1009	钙粘蛋白 11
FBXL7	1.63	5.17E-04	0.12	23194	F 盒和富亮氨酸重复蛋白 7
MEOX2	1.67	4.04E-04	0.06	4223	间质同源框 2
NOS1	1.42	2.36E-04	0.06	4842	氧化一氮合成酶 1
OSMR	1.53	2.38E-04	0.11	9180	抑瘤素 M 受体
PALM2	1.36	6.60E-04	0.01	114299	细胞膜调控蛋白 paralemmin2
PDE10A	1.44	2.31E-04	0.14	10846	磷酸二酯酶 10A
RBMS3	1.52	7.71E-04	0.12	27303	RNA 结合模体单链相互作用蛋

[0065]

					白 3
SULF1	1.62	9.68E-06	0.22	23213	硫酸酯酶 1
WLS	1.35	3.09E-04	0.03	79971	wntless Wnt 配体分泌调节者
WTIP	1.39	7.08E-04	0.06	126374	Wilms 瘤 1 作用蛋白
ZNF518B	0.70	7.71E-04	-0.03	85460	锌指蛋白 518B
ZNF714	0.68	6.77E-04	-0.10	148206	锌指蛋白 714
RDM1	0.68	1.97E-04	-0.19	201299	Rad52 模体包含 1

[0066] 如上所述,平均标准化5-hmC含量是指胃癌样品中该基因标志物的平均5-hmC含量与正常样品中同一基因标志物的平均5-hmC含量之比。从表2可以看出,本发明的胃癌基因标志物的5-hmC含量在正常样品中和肠癌样品中存在显著差异,其中在胃癌样品中,3种基因标志物显示5-hmC含量降低:RDM1、ZNF714和ZNF518B,15种基因标志物显示5-hmC含量升高:ARHGAP28、BMPER、C9orf92、CADPS2、CDH11、FBXL7、MEOX2、NOS1、OSMR、PALM2、PDE10A、RBMS3、SULF1、WLS和WTIP。

[0067] 实施例4. 胃癌基因标志物的有效性测定

[0068] 本实施例验证本发明的胃癌基因标志物用于检测胃癌的有效性。

[0069] 根据实施例1的方法测定第三批60个样品(25例胃癌和35例对照)中本发明所述的18个胃癌基因标志物的5-hmC含量,并确定各基因标志物的加权系数。

[0070] 将各基因标志物的标准化5-hmC含量乘以其对应的加权系数,获得该基因标志物的预测因子t后,将各基因标志物的预测因子t相加,获得总预测因子T,然后将总预测因子T根据以下公式经过Logistic转换获得评分P:

[0071] 
$$P = \frac{1}{1 + e^{-T}}$$

[0072] 若 $P > 0.5$ , 则该受试者样品患有胃癌; 若 $P < 0.5$ , 则该受试者样品为正常。

[0073] 图3示出了根据本发明的方法区分该批样品的结果。如图3所示, 本发明的方法能够达到92%的灵敏度和91%的特异性。发明人进一步用本发明的方法区分第四批63个样品(29例胃癌和37例对照), 其结果表明该方法能够达到90%的灵敏度和97%的特异性(图4)。

[0074] 这些结果表明, 与现有技术相比, 根据本发明的方法能够以更高的灵敏度和特异性检测肠癌。

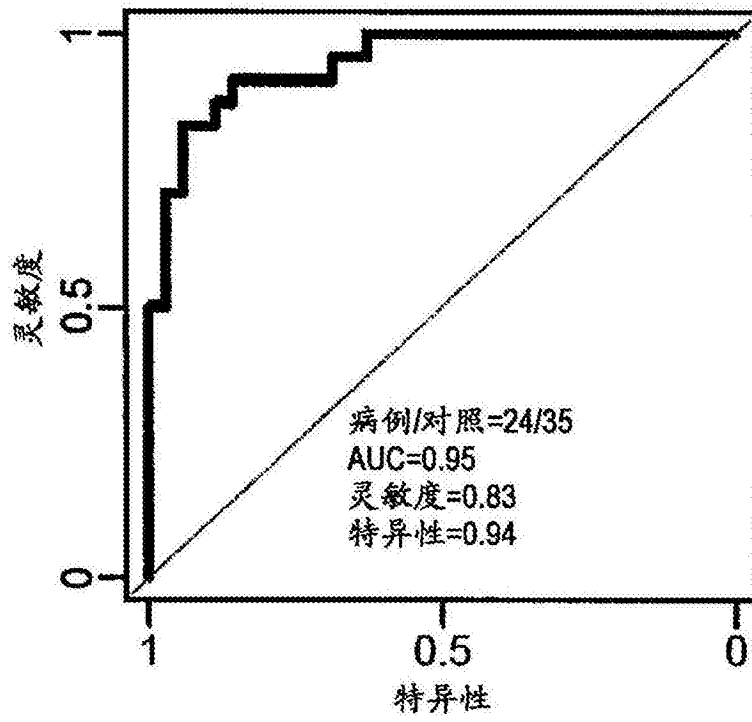


图1

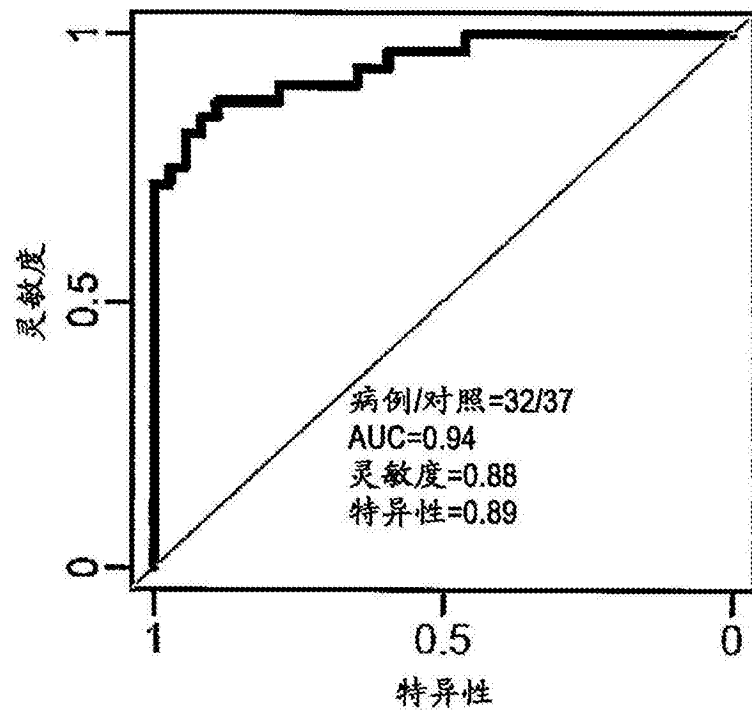


图2

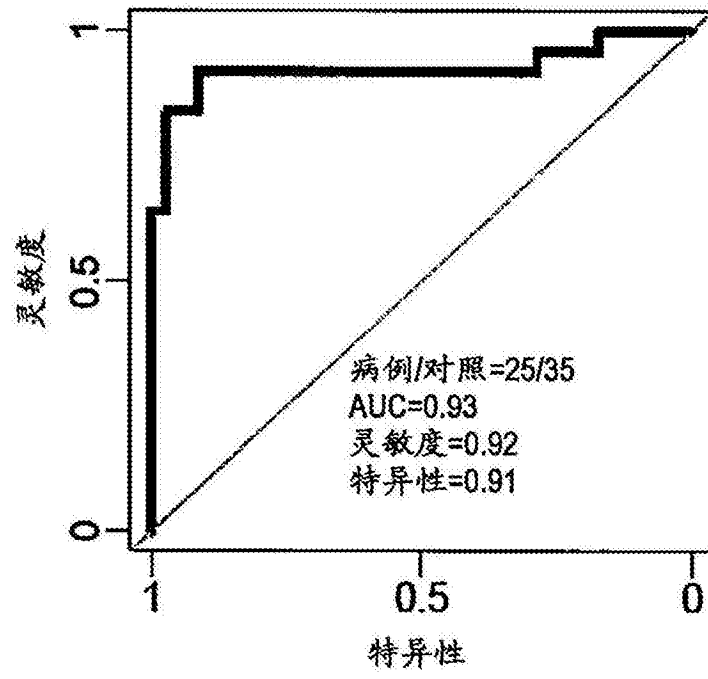


图3

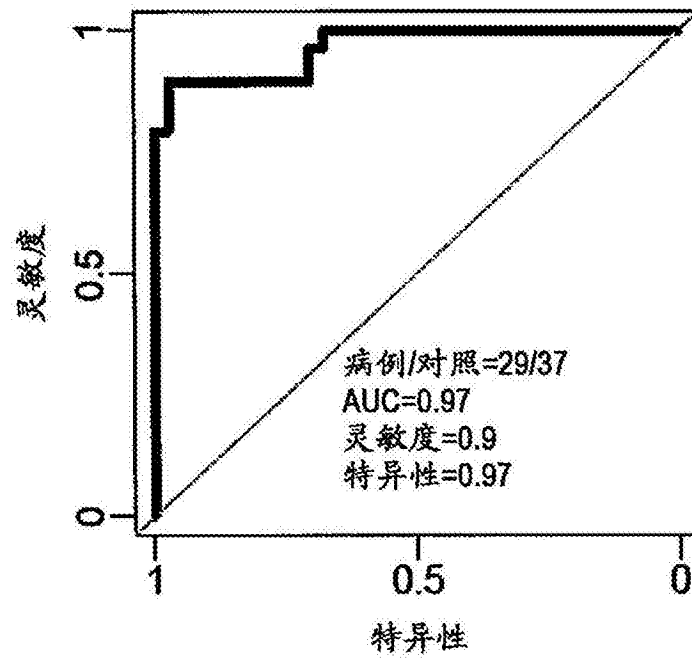


图4