



(12) 发明专利申请



(10) 申请公布号 CN 117202939 A

(43) 申请公布日 2023. 12. 08

(21) 申请号 202280023043.4

(22) 申请日 2022.03.04

(30) 优先权数据

63/157,437 2021.03.05 US

(85) PCT国际申请进入国家阶段日

2023.09.20

(86) PCT国际申请的申请数据

PCT/US2022/019010 2022.03.04

(87) PCT国际申请的公布数据

W02022/187695 EN 2022.09.09

(71) 申请人 梅奥医疗教育与研究基金会

地址 美国明尼苏达州

(72) 发明人 W·R·泰勒 D·W·马奥尼

J·B·基西尔 D·A·阿尔奎斯特

(74) 专利代理机构 北京市柳沈律师事务所

11105

专利代理师 秦剑

(51) Int.Cl.

A61K 48/00 (2006.01)

权利要求书5页 说明书76页
序列表16页

(54) 发明名称

检测宫颈癌

(57) 摘要

本文提供了用于宫颈癌筛查的技术。特别地,本文提供了用于检测宫颈癌、宫颈癌前病变(例如宫颈相关原位腺癌、宫颈上皮内瘤变)和宫颈癌亚型(例如宫颈腺癌、鳞状细胞宫颈癌)的存在或不存在,或用于从生物样品(例如组织样品(例如宫颈组织)、血液样品、血浆样品、血清样品、全血样品、分泌物样品(例如宫颈分泌物、阴道分泌物)、器官分泌物样品、CSF样品、唾液样品、尿液样品,或粪便样品)中区分宫颈癌与其他类型的妇科癌症(例如子宫内膜癌和卵巢癌)的方法、组合物和相关用途。

1. 一种用于表征生物样品的方法,所述方法包括:

测量所述生物样品中选自表I、III和X的一种或多种甲基化标记物的甲基化水平,其中测量一种或多种甲基化标记物的甲基化水平包括用以甲基化特异性方式修饰DNA的试剂处理来自生物样品的DNA。

2. 如权利要求1所述的方法,其中所述生物样品选自组织样品、血液样品、血浆样品、血清样品、全血样品、分泌物样品、器官分泌物样品、脑脊液(CSF)样品、唾液样品、尿液样品和粪便样品。

3. 如权利要求2所述的方法,其中所述组织样品是宫颈组织样品。

4. 如权利要求3所述的方法,其中所述宫颈组织样品还包括阴道组织、阴道细胞、子宫内膜组织、子宫内膜细胞、卵巢组织和卵巢细胞中的一种或多种。

5. 如权利要求2所述的方法,其中所述分泌物样品是宫颈分泌物样品。

6. 如权利要求5所述的方法,其中所述宫颈分泌物样品进一步包括阴道分泌物、子宫内膜分泌物和卵巢分泌物中的一种或多种。

7. 如权利要求1所述的方法,其中将测量的一种或多种甲基化标记物的甲基化水平与无宫颈癌的对照样品中相应的一种或多种甲基化标记物的甲基化水平进行比较。

8. 如权利要求7所述的方法,其还包括当在一种或多种甲基化标记物中测量的甲基化水平高于在相应对照样品中测量的甲基化水平时确定所述个体患有宫颈癌。

9. 如权利要求8所述的方法,其中所述一种或多种甲基化标记物选自以下组之一:

●表I和/或III中列举的甲基化标记物;

●MAX.chr6.58147682-58147771、C10RF114、ASCL1、ARHGAP12、ZNF773、TTYH1、NEUROG3、ZNF781、NXPH1、MAX.chr9.36739811-36739868、NID2、TMEM200C、CRHR2、ABCB1、ZNF69、ATP10A、MAX.chr18.73167725-73167817、MAX.chr2.127783183-127783403、CACNA1C、ZNF382、BARHL1、MAX.chr4.8859853-8859939、ST8SIA1、MAX.chr1.98510958-98511049、C20RF40、SLC9A3、PRDM12、HOPX和KCNQ5;

●C10rf114、MAX.chr6.58147682-58147771、ZNF773、NEUROG3、ASCL1、NID2、ZNF781、CRHR2和MAX.chr9.36739811-36739868;和

●ABCB1、c1orf95、CACNA1C、CACNG8、CHST2、ELM01、EMID2、FBN1_B、FLT3_A、FLT3_B、GLIS1、GPC6、GREM2、JAM2、KCNK12_A、LOC100129620、MAX.chr15.78112404-78112692、MAX.chr19.4584907-4585088、MAX.chr3.69591689-69591784、NCAM1、NT5C1A、ST8SIA3、ZNF382、ZNF419、ZNF69和ZSCAN18。

10. 如权利要求5所述的方法,其进一步包括确定所述个体患有宫颈癌亚型。

11. 如权利要求10所述的方法,其中所述宫颈癌亚型选自宫颈腺癌和鳞状细胞宫颈癌。

12. 如权利要求10所述的方法,其中所述一种或多种甲基化标记物选自以下组之一:

●ABCB1、ARHGAP12、ASCL1、ATP10A、BARHL1、C10rf114、CACNA1C、CRHR2、MAX.chr1.98510968-98511049、MAX.chr18.73167751-73167791、MAX.chr2.127783183-127783403、MAX.chr4.8859853-8859939、MAX.chr6.58147682-58147771、MAX.chr9.36739811-36739868、NEUROG3、NID2、NXPH1、TMEM200C、TTYH1、ZNF382、ZNF69、ZNF773和ZNF781。

13. 如权利要求7所述的方法,其进一步包括确定所述个体患有宫颈癌前病变。

14. 如权利要求13所述的方法,其中所述宫颈癌前病变选自宫颈相关原位腺癌和宫颈上皮内瘤变。

15. 如权利要求13所述的方法,其中所述一种或多种甲基化标记物选自以下组之一:

●MAX.chr6.58147682-58147771、C10RF114、ASCL1、ZNF773、TTYH1、NEUROG3、ZNF781、MAX.chr9.36739811-36739868、CRHR2和NID2;和

ABCB1、ARHGAP12、ASCL1、ATP10A、BARHL1、C1orf114、CACNA1C、CRHR2、MAX.chr1.98510968-98511049、MAX.chr18.73167751-73167791、MAX.chr2.127783183-127783403、MAX.chr4.8859853-8859939、MAX.chr6.58147682-58147771、MAX.chr9.36739811-36739868、NEUROG3、NID2、NXPH1、TMEM200C、TTYH1、ZNF 382、ZNF69、ZNF773和ZNF781。

16. 如权利要求1所述的方法,其中将测量的所述一种或多种甲基化标记物的甲基化水平与子宫内膜癌样品和/或卵巢癌样品中相应的一种或多种甲基化标记物的甲基化水平进行比较。

17. 如权利要求16所述的方法,进一步包括区分宫颈癌与子宫内膜癌和/或卵巢癌。

18. 如权利要求16所述的方法,其中所述一种或多种甲基化标记物选自以下组之一:

●表X中列举的标记物;

●ABCB1、c1orf95、CACNA1C、CACNG8、CHST2、ELM01、EMID2、FBN1_B、FLT3_A、FLT3_B、GLIS1、GPC6、GREM2、JAM2、KCNK12_A、LOC100129620、MAX.chr15.78112404-78112692、MAX.chr19.4584907-4585088、MAX.chr3.69591689-69591784、NCAM1、NT5C1A、ST8SIA3、ZNF382、ZNF419、ZNF69和ZSCAN 18;和

●AK5、RABC3、ZNF491、ZNF610、ZNF91、ZNF480、TRPC3_B和ELMOD1。

19. 如权利要求1所述的方法,其中以甲基化特异性方式修饰DNA的试剂是硼烷还原剂。

20. 如权利要求1所述的方法,其中所述硼烷还原剂是2-甲基吡啶硼烷。

21. 如权利要求1所述的方法,其中以甲基化特异性方式修饰DNA的试剂包括甲基化敏感性限制酶、甲基化依赖性限制酶和亚硫酸氢盐试剂中的一种或多种。

22. 如权利要求1所述的方法,其中以甲基化特异性方式修饰DNA的试剂是亚硫酸氢盐试剂,并且所述处理产生亚硫酸氢盐处理的DNA。

23. 如权利要求1所述的方法,其中用对所述一种或多种甲基化标记物具有特异性的一组引物扩增所处理的DNA。

24. 如权利要求23所述的方法,其中对所述一种或多种甲基化标记物具有特异性的所述引物组选自表V和XII中列举的组。

25. 如权利要求23所述的方法,其中对所述一种或多种甲基化标记物具有特异性的所述引物组能够结合由表V和XII中列举的特定甲基化标记基因的引物序列结合的扩增子,其中由表V和XII中列举的甲基化标记基因的引物序列结合的所述扩增子是表I、III和X中列举的甲基化标记物的基因区的至少一部分。

26. 如权利要求23所述的方法,其中对所述一种或多种甲基化标记物具有特异性的所述引物组是特异性结合包含表I、III和X中列举的甲基化标记物的染色体坐标的基因区的至少一部分的引物组。

27. 如权利要求1所述的方法,其中测量一种或多种甲基化标记物的甲基化水平包括多

重扩增。

28. 如权利要求1所述的方法,其中测量一种或多种甲基化标记物的甲基化水平包括使用选自自由甲基化特异性PCR、定量甲基化特异性PCR、甲基化特异性DNA限制酶分析、定量亚硫酸氢盐焦磷酸测序、flap核酸内切酶测定、PCR-flap测定和亚硫酸氢盐基因组测序PCR组成的组的一种或多种方法。

29. 如权利要求1所述的方法,其中测量一种或多种甲基化标记物的甲基化水平包括测量所述一种或多种甲基化标记物的CpG位点的甲基化。

30. 如权利要求29所述的方法,其中所述CpG位点存在于编码区或调控区中。

31. 如权利要求1所述的方法,其中所述一种或多种甲基化标记物通过表I、III和X中所示的基因组坐标来描述。

32. 如权利要求1所述的方法,其中所述生物样品来自人类受试者。

33. 如权利要求32所述的方法,其中所述人类受试者患有或疑似患有宫颈癌、宫颈癌亚型或宫颈癌前病变。

34. 如权利要求2所述的方法,其中所述生物样品用具有吸收构件的收集装置收集,所述吸收构件能够在与身体区域接触后收集所述生物样品。

35. 如权利要求34所述的方法,其中所述吸收构件是具有适合插入身体孔口的形状和尺寸的海绵。

36. 如权利要求34所述的方法,其中所述收集装置选自卫生棉条、将液体释放到阴道中并重新收集液体的灌洗器、宫颈刷、Fournier宫颈自采样装置和拭子。

37. 一种从生物样品中制备脱氧核糖核酸(DNA)级分的方法,所述脱氧核糖核酸级分可用于分析参与一种或多种染色体畸变的一个或多个遗传基因座,所述方法包括:

(a) 从生物样品中提取基因组DNA;

(b) 通过以下方式产生经提取的基因组DNA的级分:

(i) 用以甲基化特异性方式修饰DNA的试剂处理经提取的基因组DNA;

(ii) 使用对表I、III和X中列举的一种或多种甲基化标记物具有特异性的单独引物扩增经处理的基因组DNA;

(c) 通过测量所述一种或多种甲基化标记物中每一种的甲基化水平来分析经提取的基因组DNA的产生级分中的一个或多个遗传基因座。

38. 如权利要求37所述的方法,其中以甲基化特异性方式修饰DNA的试剂是硼烷还原剂。

39. 如权利要求38所述的方法,其中所述硼烷还原剂是2-甲基吡啶硼烷。

40. 如权利要求37所述的方法,其中以甲基化特异性方式修饰DNA的试剂包括甲基化敏感性限制酶、甲基化依赖性限制酶和亚硫酸氢盐试剂中的一种或多种。

41. 如权利要求37所述的方法,其中以甲基化特异性方式修饰DNA的试剂是亚硫酸氢盐试剂,并且所述处理产生亚硫酸氢盐处理的DNA。

42. 如权利要求37所述的方法,其中对所述一种或多种甲基化标记物具有特异性的所述引物组选自表V和XII中列举的组。

43. 如权利要求37所述的方法,其中对所述一种或多种甲基化标记物具有特异性的所述引物组能够结合由表V和XII中列举的特定甲基化标记基因的引物序列结合的扩增子,其

中由表V和XII中列举的甲基化标记基因的引物序列结合的所述扩增子是表I、III和X中列举的甲基化标记物的基因区的至少一部分。

44. 如权利要求37所述的方法,其中对所述一种或多种甲基化标记物具有特异性的所述引物组是特异性结合包含表I、III和X中列举的甲基化标记物的染色体坐标的基因区的至少一部分的引物组。

45. 如权利要求37所述的方法,其中测量一种或多种甲基化标记物的甲基化水平包括多重扩增。

46. 如权利要求37所述的方法,其中测量一种或多种甲基化标记物的甲基化水平包括使用选自自由甲基化特异性PCR、定量甲基化特异性PCR、甲基化特异性DNA限制酶分析、定量亚硫酸氢盐焦磷酸测序、flap核酸内切酶测定、PCR-flap测定和亚硫酸氢盐基因组测序PCR组成的组的一种或多种方法。

47. 如权利要求37所述的方法,其中测量一种或多种甲基化标记物的甲基化水平包括测量所述一种或多种甲基化标记物的CpG位点的甲基化。

48. 如权利要求47所述的方法,其中所述CpG位点存在于编码区或调控区中。

49. 如权利要求37所述的方法,其中所述一种或多种甲基化标记物通过表I、III和X中所示的基因组坐标来描述。

50. 如权利要求37所述的方法,其中所述生物样品选自组织样品、血液样品、血浆样品、血清样品、全血样品、分泌物样品、器官分泌物样品、脑脊液(CSF)样品、唾液样品、尿液样品和粪便样品。

51. 如权利要求50所述的方法,其中所述组织样品是宫颈组织样品。

52. 如权利要求51所述的方法,其中所述宫颈组织样品还包括阴道组织、阴道细胞、子宫内膜组织、子宫内膜细胞、卵巢组织和卵巢细胞中的一种或多种。

53. 如权利要求50所述的方法,其中所述分泌物样品是宫颈分泌物样品。

54. 如权利要求53所述的方法,其中所述宫颈分泌物样品进一步包括阴道分泌物、子宫内膜分泌物和卵巢分泌物中的一种或多种。

55. 如权利要求37所述的方法,其中所述生物样品用具有吸收构件的收集装置收集,所述吸收构件能够在与身体区域接触后收集所述生物样品。

56. 如权利要求55所述的方法,其中所述吸收构件是具有适合插入身体孔口的形状和尺寸的海绵。

57. 如权利要求55所述的方法,其中所述收集装置选自卫生棉条、将液体释放到阴道中并重新收集液体的灌洗器、宫颈刷、Fournier宫颈自采样装置和拭子。

58. 如权利要求37所述的方法,其中所述生物样品来自人类受试者。

59. 如权利要求58所述的方法,其中所述人类受试者患有或疑似患有宫颈癌、宫颈癌亚型或宫颈癌前病变。

60. 如权利要求37所述的方法,其中所述一种或多种甲基化标记物选自以下组之一:

●MAX.chr6.58147682-58147771、C10RF114、ASCL1、ARHG AP12、ZNF773、TTYH1、NEUROG3、ZNF781、NXPH1、MAX.ch r9.36739811-36739868、NID2、TMEM200C、CRHR2、ABCB1、ZNF69、ATP10A、MAX.chr18.73167725-73167817、MAX.chr2.127783183-127783403、CACNA1C、ZNF382、BARHL1、MAX.chr4.8859853-8859939、ST8SIA1、MAX.chr1.98510958-98511049、

C2ORF40、SLC9A3、PRDM12、HOPX_C和KCNQ5;

●C1orf114、MAX.chr6.58147682-58147771、ZNF773、NEUROG3、ASCL1、NID2、ZNF781、CRHR2和MAX.chr9.36739811-36739868;

●ABCB1、ARHGAP12、ASCL1、BARHL1、C1orf114、C2orf40、CACNA1C、CRHR2、HOPX_C、KCNQ5、MAX.chr1.98510968-98511049、MAX.chr18.73167751-73167791、MAX.chr2.127783183-127783403、MAX.chr4.8859853-8859939、MAX.chr6.58147682-58147771、MAX.chr9.36739811-36739868、NEUROG3、NID2、NXP1、PRDM12、SLC9A3、TMEM200C、TTYH1、ZNF382、ZNF773和ZNF781;

●ABCB1、c1orf95、CACNA1C、CACNG8、CHST2、ELMO1、EMID2、FBN1_B、FLT3_A、FLT3_B、GLIS1、GPC6、GREM2、JAM2、KCNK12_A、LOC100129620、MAX.chr15.78112404-78112692、MAX.chr19.4584907-4585088、MAX.chr3.69591689-69591784、NCAM1、NT5C1A、ST8SIA3、ZNF382、ZNF419、ZNF69和ZSCAN18;和

●AK5、RAB3、ZNF491、ZNF610、ZNF91、ZNF480、TRPC3_B和ELMOD1。

检测宫颈癌

[0001] 相关申请的交叉引用

[0002] 本申请要求2021年3月5日提交的美国临时专利申请No. 63/157,437的优先权,其在此以引用的方式整体并入。

技术领域

[0003] 本文提供了用于宫颈癌筛查的技术。特别地,本文提供了用于检测宫颈癌、宫颈癌前病变(例如宫颈相关原位腺癌、宫颈上皮内瘤变)和宫颈癌亚型(例如宫颈腺癌、鳞状细胞宫颈癌)存在或不存在,或用于从生物样品(例如组织样品(例如宫颈组织)、血液样品、血浆样品、血清样品、全血样品、分泌物样品(例如宫颈分泌物、阴道分泌物)、器官分泌物样品、CSF样品、唾液样品、尿液样品,或粪便样品)中区分宫颈癌与其他类型的妇科癌症(例如子宫内膜癌和卵巢癌)的方法、组合物和相关用途。

背景技术

[0004] 宫颈癌(CC)筛查方法不断发展,并且对宫颈发育不良和CC的存在高度敏感。然而,鉴于高风险HPV感染的普遍存在(无需肿瘤转化而被清除),当前基于分子的检测(即仅高风险HPV)的特异性有限。因此,迫切需要改进的诊断工具来从单个生物样品中检测宫颈癌。

[0005] 本发明满足了这种需要。

发明内容

[0006] 本文提供了用于宫颈癌筛查的技术,特别是但不排他地,涉及用于检测宫颈癌、宫颈癌前病变(例如宫颈相关原位腺癌、宫颈上皮内瘤变)和宫颈癌亚型(例如宫颈腺癌、鳞状细胞宫颈癌)是否存在,以及用于区分宫颈癌与其他类型的妇科癌症(例如子宫内膜癌和卵巢癌)的方法、组合物和相关用途。

[0007] 实际上,如实施例I和II中所述,在鉴定本发明实施方案的过程中进行的实验鉴定了一组新的差异甲基化区域(DMR),其用于将宫颈来源的DNA的癌症和癌前病变与非肿瘤对照DNA和其他类型的妇科癌症(例如子宫内膜癌和卵巢癌)区分开。

[0008] 此类实验列出并描述了423种新型DNA甲基化标记物,其用于区分宫颈癌(宫颈癌亚型)和癌前病变与良性宫颈样品(参见表I-IV、VI-VIII、实施例I),并用于区分宫颈癌与其他类型妇科癌症(例如子宫内膜癌和卵巢癌)。(参见表XI和XII,实施例II)。

[0009] 从这423种新型DNA甲基化标记物中,进一步的实验鉴定了以下能够区分宫颈癌与良性宫颈样品的标记物和/或标记物组:

[0010] • 表I中列举的任何标记物(参见实施例I);

[0011] • 表III中列举的任何标记物(参见实施例I);

[0012] • MAX.chr6.58147682-58147771、C10RF114、ASCL1、ARHG AP12、ZNF773、TTYH1、NEUROG3、ZNF781、NXPH1、MAX.ch r9.36739811-36739868、NID2、TMEM200C、CRHR2、ABCB1、ZN F69、ATP10A、MAX.chr18.73167725-73167817、MAX.chr2.127783183-127783403、CACNA1C、

ZNF382、BARHL1、MAX.chr4.8859853-8859939、ST8SIA1、MAX.chr1.98510958-98511049、C2ORF40、SLC9A3、PRDM12、HOPX_C和KCNQ5(实施例I)；

[0013] • C1orf114、MAX.chr6.58147682-58147771、ZNF773、NEUROG3、ASCL1、NID2、ZNF781、CRHR2和MAX.chr9.36739811-36739868(参见表VI和实施例I)；和

[0014] • ABCB1、c1orf95、CACNA1C、CACNG8、CHST2、ELM01、EMID2、FBN1_B、FLT3_A、FLT3_B、GLIS1、GPC6、GREM2、JAM2、KCNK12_A、LOC100129620、MAX.chr15.78112404-78112692、MAX.chr19.4584907-4585088、MAX.chr3.69591689-69591784、NCAM1、NT5C1A、ST8SIA3、ZNF382、ZNF419、ZNF69和ZSCAN18(参见表VIII,实施例I)。

[0015] 从这423种新型DNA甲基化标记物中,进一步的实验鉴定了以下能够区分宫颈腺癌与良性宫颈样品的标记物和/或标记物组:

[0016] • ABCB1、ARHGAP12、ASCL1、BARHL1、C1orf114、C2orf40、CACNA1C、CRHR2、HOPX_C、KCNQ5、MAX.chr1.98510968-98511049、MAX.chr18.73167751-73167791、MAX.chr2.127783183-127783403、MAX.chr4.8859853-8859939、MAX.chr6.58147682-58147771、MAX.chr9.36739811-36739868、NEUROG3、NID2、NXPH1、PRDM12、SLC9A3、TMEM200C、TTYH1、ZNF382、ZNF773和ZNF781(参见表VII,实施例I)。

[0017] 从这423种新型DNA甲基化标记物中,进一步的实验鉴定了以下能够区分宫颈鳞状细胞癌与良性宫颈样品的标记物和/或标记物组:

[0018] • ABCB1、ARHGAP12、ASCL1、ATP10A、BARHL1、C1orf114、CACNA1C、CRHR2、MAX.chr1.98510968-98511049、MAX.chr18.73167751-73167791、MAX.chr2.127783183-127783403、MAX.chr4.8859853-8859939、MAX.chr6.58147682-58147771、MAX.chr9.36739811-36739868、NEUROG3、NID2、NXPH1、TMEM200C、TTYH1、ZNF382、ZNF69、ZNF773和ZNF781(参见表VII,实施例I)。

[0019] 从这423种新型DNA甲基化标记物中,进一步的实验鉴定了以下能够区分宫颈相关原位腺癌(AIS)与良性宫颈样品的标记物和/或标记物组:

[0020] • MAX.chr6.58147682-58147771、C1ORF114、ASCL1、ZNF773、TTYH1、NEUROG3、ZNF781、MAX.chr9.36739811-36739868、CRHR2和NID2(参见表VI和实施例I)；和

[0021] • ABCB1、ARHGAP12、ASCL1、BARHL1、C1orf114、C2orf40、CACNA1C、CRHR2、HOPX_C、KCNQ5、MAX.chr1.98510968-98511049、MAX.chr18.73167751-73167791、MAX.chr2.127783183-127783403、MAX.chr4.8859853-8859939、MAX.chr6.58147682-58147771、MAX.chr9.36739811-36739868、NEUROG3、NID2、NXPH1、PRDM12、SLC9A3、TMEM200C、TTYH1、ZNF382、ZNF773和ZNF781(参见表VII,实施例I)。

[0022] 从这423种新型DNA甲基化标记物中,进一步的实验鉴定了以下能够区分宫颈上皮内瘤变(CIN)与良性宫颈样品的标记物和/或标记物组:

[0023] • MAX.chr6.58147682-58147771、C1ORF114、ASCL1、ZNF773、TTYH1、NEUROG3、ZNF781、MAX.chr9.36739811-36739868、CRHR2和NID2(参见表VI和实施例I)；和

[0024] • ABCB1、ARHGAP12、ASCL1、ATP10A、BARHL1、C1orf114、CACNA1C、CRHR2、MAX.chr1.98510968-98511049、MAX.chr18.73167751-73167791、MAX.chr2.127783183-127783403、MAX.chr4.8859853-8859939、MAX.chr6.58147682-58147771、MAX.chr9.36739811-36739868、NEUROG3、NID2、NXPH1、TMEM200C、TTYH1、ZNF382、ZNF69、

ZNF773和ZNF781(参见表VII,实施例I)。

[0025] 从这423种新型DNA甲基化标记物中,进一步的实验鉴定了以下能够区分宫颈癌与子宫内膜癌和/或卵巢癌的标记物和/或标记物组:

[0026] • ABCB1、c1orf95、CACNA1C、CACNG8、CHST2、ELM01、EMID2、FBN1_B、FLT3_A、FLT3_B、GLIS1、GPC6、GREM2、JAM2、KCNK12_A、LOC100129620、MAX.chr15.78112404-78112692、MAX.chr19.4584907-4585088、MAX.chr3.69591689-69591784、NCAM1、NT5C1A、ST8SIA3、ZNF382、ZNF419、ZNF69和ZSCAN 18(参见表VIII,实施例I);

[0027] • 表X中列举的任何标记物(参见实施例I);以及

[0028] • AK5、RABC3、ZNF491、ZNF610、ZNF91、ZNF480、TRPC3_B和ELMOD1(参见表X和XI,实施例II)。

[0029] 如本文所述,该技术提供了许多甲基化DNA标记物(MDM)及其子集(例如,2、3、4、5、6、7、8、20、50、100、150、200、300、400、423个标记物的集合),其对宫颈癌、各种类型的宫颈癌(例如宫颈腺癌、鳞状细胞宫颈癌)、各种类型的宫颈癌前病变(如宫颈相关原位腺癌、宫颈上皮内瘤变),以及对于区分宫颈癌与其他类型的妇科癌症(例如子宫内膜癌和卵巢癌)具有较高的区分度。宫颈上皮内瘤变(CIN)的特征是CIN 1、CIN 2和CIN 3,CIN 1是指影响约三分之一的宫颈上皮厚度的异常细胞,CIN 2是指影响约三分之二至三分之二的宫颈上皮的异常细胞,并且CIN 3是指影响超过三分之二的宫颈上皮的异常细胞。

[0030] 在某些实施方案中,本公开提供了用于表征生物样品的方法,该方法包括测量生物样品中选自表I、III和X的一种或多种甲基化标记物的甲基化水平,其中测量一种或多种甲基化标记物的甲基化水平包括用以甲基化特异性方式修饰DNA的试剂处理来自生物样品的DNA。

[0031] 在一些实施方案中,生物样品来自人类受试者。在一些实施方案中,人类受试者患有或疑似患有宫颈癌、宫颈癌亚型和/或宫颈癌前病变。

[0032] 在一些实施方案中,生物样品选自组织样品、血液样品、血浆样品、血清样品、全血样品、分泌物样品(例如宫颈分泌物、阴道分泌物)、器官分泌物样品、脑脊液(CSF)样品、唾液样品、尿液样品和粪便样品。在一些实施方案中,组织样品是宫颈组织样品。在一些实施方案中,宫颈组织样品还包括阴道组织、阴道细胞、子宫内膜组织、子宫内膜细胞、卵巢组织和卵巢细胞中的一种或多种。在一些实施方案中,分泌物样品是宫颈分泌物样品。在一些实施方案中,宫颈分泌物样品进一步包括阴道分泌物、子宫内膜分泌物和卵巢分泌物中的一种或多种。在一些实施方案中,用具有吸收构件的收集装置收集生物样品,该吸收构件能够在与身体区域接触后收集生物样品。在一些实施方案中,吸收构件是具有适合插入身体孔口的形状和尺寸的海绵。在一些实施方案中,收集装置是卫生棉条(例如,标准卫生棉条)、将液体释放到阴道中并重新收集液体的灌洗器(例如,Pantarhei筛查器)、宫颈刷(例如,插入阴道并转动收集细胞的刷子)、Fournier宫颈自采样装置(卫生棉条状塑料棒)或拭子。

[0033] 在一些实施方案中,将测量的该一种或多种甲基化标记物的甲基化水平与无宫颈癌的对照样品中相应的一种或多种甲基化标记物的甲基化水平进行比较。

[0034] 在一些实施方案中,该方法还包括当在该一种或多种甲基化标记物中测量的甲基化水平高于在相应对照样品中测量的甲基化水平时确定个体患有宫颈癌。在一些实施方案中,其中该方法还包括确定个体患有宫颈癌,该一种或多种甲基化标记物选自以下组之一:

[0035] • 表I和/或表III中列举的甲基化标记物;

[0036] • MAX.chr6.58147682-58147771、C10RF114、ASCL1、ARHGAP12、ZNF773、TTYH1、NEUROG3、ZNF781、NXPH1、MAX.chr9.36739811-36739868、NID2、TMEM200C、CRHR2、ABCB1、ZNF69、ATP10A、MAX.chr18.73167725-73167817、MAX.chr2.127783183-127783403、CACNA1C、ZNF382、BARHL1、MAX.chr4.8859853-8859939、ST8SIA1、MAX.chr1.98510958-98511049、C2ORF40、SLC9A3、PRDM12、HOPX_C和KCNQ5(实施例I);

[0037] • C1orf114、MAX.chr6.58147682-58147771、ZNF773、NEUROG3、ASCL1、NID2、ZNF781、CRHR2和MAX.chr9.36739811-36739868(参见表VI和实施例I);和

[0038] • ABCB1、c1orf95、CACNA1C、CACNG8、CHST2、ELM01、EMID2、FBN1_B、FLT3_A、FLT3_B、GLIS1、GPC6、GREM2、JAM2、KCNK12_A、LOC100129620、MAX.chr15.78112404-78112692、MAX.chr19.4584907-4585088、MAX.chr3.69591689-69591784、NCAM1、NT5C1A、ST8SIA3、ZNF382、ZNF419、ZNF69和ZSCAN18(参见表VIII,实施例I)。

[0039] 在一些实施方案中,该方法还包括确定个体患有宫颈癌亚型。在一些实施方案中,宫颈癌的亚型选自宫颈腺癌和鳞状细胞宫颈癌。在一些实施方案中,其中该方法还包括确定个体患有宫颈癌亚型,该一种或多种甲基化标记物选自以下组之一:

[0040] • ABCB1、ARHGAP12、ASCL1、BARHL1、C1orf114、C2orf40、CACNA1C、CRHR2、HOPX_C、KCNQ5、MAX.chr1.98510968-98511049、MAX.chr18.73167751-73167791、MAX.chr2.127783183-127783403、MAX.chr4.8859853-8859939、MAX.chr6.58147682-58147771、MAX.chr9.36739811-36739868、NEUROG3、NID2、NXPH1、PRDM12、SLC9A3、TMEM200C、TTYH1、ZNF382、ZNF773和ZNF781(参见表VII,实施例I)。

[0041] 在一些实施方案中,该方法还包括确定个体患有宫颈癌前病变。在一些实施方案中,宫颈癌前病变选自宫颈相关原位腺癌和宫颈上皮内瘤变。在一些实施方案中,其中该方法还包括确定个体患有宫颈癌前病变,该一种或多种甲基化标记物选自以下组之一:

[0042] • MAX.chr6.58147682-58147771、C10RF114、ASCL1、ZNF773、TTYH1、NEUROG3、ZNF781、MAX.chr9.36739811-36739868、CRHR2和NID2(参见表VI和实施例I);和

[0043] • ABCB1、ARHGAP12、ASCL1、BARHL1、C1orf114、C2orf40、CACNA1C、CRHR2、HOPX_C、KCNQ5、MAX.chr1.98510968-98511049、MAX.chr18.73167751-73167791、MAX.chr2.127783183-127783403、MAX.chr4.8859853-8859939、MAX.chr6.58147682-58147771、MAX.chr9.36739811-36739868、NEUROG3、NID2、NXPH1、PRDM12、SLC9A3、TMEM200C、TTYH1、ZNF382、ZNF773和ZNF781(参见表VII,实施例I)。

[0044] 在一些实施方案中,将测量的该一种或多种甲基化标记物的甲基化水平与子宫内膜癌样品和/或卵巢癌样品中相应的一种或多种甲基化标记物的甲基化水平进行比较。在一些实施方案中,该方法还包括区分宫颈癌与子宫内膜癌和/或卵巢癌。在一些实施方案中,其中该方法还包括区分宫颈癌与子宫内膜癌和/或卵巢癌,该一种或多种甲基化标记物选自以下组之一:

[0045] • 表X中列举的标记物;

[0046] • ABCB1、c1orf95、CACNA1C、CACNG8、CHST2、ELM01、EMID2、FBN1_B、FLT3_A、FLT3_B、GLIS1、GPC6、GREM2、JAM2、KCNK12_A、LOC100129620、MAX.chr15.78112404-78112692、MAX.chr19.4584907-4585088、MAX.chr3.69591689-69591784、NCAM1、NT5C1A、ST8SIA3、

ZNF382、ZNF419、ZNF69和ZSCAN18(参见表VIII,实施例I);和

[0047] • AK5、RABC3、ZNF491、ZNF610、ZNF91、ZNF480、TRPC3_B和ELMOD1(参见表X和XI,实施例II)。

[0048] 在一些实施方案中,以甲基化特异性方式修饰DNA的试剂是硼烷还原剂。在一些实施方案中,硼烷还原剂是2-甲基吡啶硼烷。在一些实施方案中,以甲基化特异性方式修饰DNA的试剂包含甲基化敏感性限制酶、甲基化依赖性限制酶和亚硫酸氢盐试剂中的一种或多种。在一些实施方案中,以甲基化特异性方式修饰DNA的试剂是亚硫酸氢盐试剂,并且该处理产生亚硫酸氢盐处理的DNA。

[0049] 在一些实施方案中,用对该一种或多种甲基化标记物具有特异性的一组引物扩增处理的DNA。在一些实施方案中,对该一种或多种甲基化标记物具有特异性的引物组选自表V和XII中列举的组。在一些实施方案中,对该一种或多种甲基化标记物具有特异性的引物组能够结合由表V和XII中列举的特定甲基化标记基因的引物序列结合的扩增子,其中由表V和XII中列举的甲基化标记基因的引物序列结合的扩增子是表I、III和X中列举的甲基化标记物的基因区的至少一部分。在一些实施方案中,对该一种或多种甲基化标记物具有特异性的引物组是特异性结合包含表I、III和X中列举的甲基化标记物的染色体坐标的基因区的至少一部分的引物组。

[0050] 在一些实施方案中,测量一种或多种甲基化标记物的甲基化水平包括多重扩增。在一些实施方案中,测量一种或多种甲基化标记物的甲基化水平包括使用选自由甲基化特异性PCR、定量甲基化特异性PCR、甲基化特异性DNA限制酶分析、定量亚硫酸氢盐焦磷酸测序、flap核酸内切酶测定、PCR-flap测定和亚硫酸氢盐基因组测序PCR组成的组的一种或多种方法。在一些实施方案中,测量一种或多种甲基化标记物的甲基化水平包括测量该一种或多种甲基化标记物的CpG位点的甲基化。在一些实施方案中,CpG位点存在于编码区或调控区中。在一些实施方案中,该一种或多种甲基化标记物通过表I、III和X中所示的基因组坐标来描述。

[0051] 在某些实施方案中,本公开提供了从生物样品中制备脱氧核糖核酸(DNA)级分的方法,该脱氧核糖核酸级分可用于分析涉及一种或多种染色体畸变的一个或多个遗传基因座,该方法包括:

[0052] (a) 从生物样品中提取基因组DNA;

[0053] (b) 通过以下方式产生经提取的基因组DNA的级分:

[0054] (i) 用以甲基化特异性方式修饰DNA的试剂处理经提取的基因组DNA;

[0055] (ii) 使用对表I、III和X中列举的一种或多种甲基化标记物具有特异性的单独引物扩增经处理的基因组DNA;

[0056] (c) 通过测量该一种或多种甲基化标记物中每一种的甲基化水平来分析经提取的基因组DNA的产生级分中的一个或多个遗传基因座。

[0057] 在一些实施方案中,以甲基化特异性方式修饰DNA的试剂是硼烷还原剂。在一些实施方案中,硼烷还原剂是2-甲基吡啶硼烷。在一些实施方案中,其中以甲基化特异性方式修饰DNA的试剂包含甲基化敏感性限制酶、甲基化依赖性限制酶和亚硫酸氢盐试剂中的一种或多种。在一些实施方案中,以甲基化特异性方式修饰DNA的试剂是亚硫酸氢盐试剂,并且该处理产生亚硫酸氢盐处理的DNA。

[0058] 在一些实施方案中,对该一种或多种甲基化标记物具有特异性的引物组选自表V和XII中列举的组。在一些实施方案中,对该一种或多种甲基化标记物具有特异性的引物组能够结合由表V和XII中列举的特定甲基化标记基因的引物序列结合的扩增子,其中由表V和XII中列举的甲基化标记基因的引物序列结合的扩增子是表I、III和X中列举的甲基化标记物的基因区的至少一部分。在一些实施方案中,对该一种或多种甲基化标记物具有特异性的引物组是特异性结合包含表I、III和X中列举的甲基化标记物的染色体坐标的基因区的至少一部分的引物组。

[0059] 在一些实施方案中,测量一种或多种甲基化标记物的甲基化水平包括多重扩增。在一些实施方案中,测量一种或多种甲基化标记物的甲基化水平包括使用选自由甲基化特异性PCR、定量甲基化特异性PCR、甲基化特异性DNA限制酶分析、定量亚硫酸氢盐焦磷酸测序、flap核酸内切酶测定、PCR-flap测定和亚硫酸氢盐基因组测序PCR组成的组的一种或多种方法。在一些实施方案中,测量一种或多种甲基化标记物的甲基化水平包括测量该一种或多种甲基化标记物的CpG位点的甲基化。在一些实施方案中,CpG位点存在于编码区或调控区中。在一些实施方案中,该一种或多种甲基化标记物通过表I、III和X中所示的基因组坐标来描述。

[0060] 在一些实施方案中,生物样品选自组织样品、血液样品、血浆样品、血清样品、全血样品、分泌物样品(例如,宫颈分泌物、阴道分泌物)、器官分泌物样品、脑脊液(CSF)样品、唾液样品、尿液样品和粪便样品。在一些实施方案中,组织样品是宫颈组织样品。在一些实施方案中,宫颈组织样品还包括阴道组织、阴道细胞、子宫内膜组织、子宫内膜细胞、卵巢组织和卵巢细胞中的一种或多种。在一些实施方案中,分泌物样品是宫颈分泌物样品。在一些实施方案中,宫颈分泌物样品还包含阴道分泌物、子宫内膜分泌物和卵巢分泌物中的一种或多种。在一些实施方案中,用具有吸收构件的收集装置收集生物样品,该吸收构件能够在与身体区域接触后收集组织和/或细胞。在一些实施方案中,吸收构件是具有适合插入身体孔口的形状和尺寸的海绵。在一些实施方案中,收集装置是卫生棉条(例如,标准卫生棉条)、将液体释放到阴道中并重新收集液体的灌洗器(例如,Pantarhei筛查器)、宫颈刷(例如,妇女插入阴道并转动收集细胞的刷子)、Fournier宫颈自采样装置(卫生棉条状塑料棒)或拭子。

[0061] 在一些实施方案中,生物样品来自人类受试者。在一些实施方案中,人类受试者患有或疑似患有宫颈癌、宫颈癌亚型和/或宫颈癌前病变。

[0062] 在一些实施方案中,该一种或多种甲基化标记物选自以下组之

[0063] • MAX.chr6.58147682-58147771、C10RF114、ASCL1、ARHG AP12、ZNF773、TTYH1、NEUROG3、ZNF781、NXPH1、MAX.ch r9.36739811-36739868、NID2、TMEM200C、CRHR2、ABCB1、ZNF69、ATP10A、MAX.chr18.73167725-73167817、MAX.chr2.127783183-127783403、CACNA1C、ZNF382、BARHL1、MAX.chr4.8859853-8859939、ST8SIA1、MAX.chr1.98510958-98511049、C2ORF40、SL C9A3、PRDM12、HOPX_C和KCNQ5(实施例I);

[0064] • C1orf114、MAX.chr6.58147682-58147771、ZNF773、NEURO G3、ASCL1、NID2、ZNF781、CRHR2和MAX.chr9.36739811-36739868(参见表VI和实施例I);

[0065] • ABCB1、ARHGAP12、ASCL1、BARHL1、C1orf114、C2orf40、CACNA1C、CRHR2、HOPX_C、KCNQ5、MAX.chr1.98510968-98511049、MAX.chr18.73167751-73167791、

MAX.chr2.127783183-127783403、MAX.chr4.8859853-8859939、MAX.chr6.58147682-58147771、MAX.chr9.36739811-36739868、NEUROG3、NID2、NXPH1、PRDM12、SLC9A3、TMEM200C、TTYH1、ZNF382、ZNF773和ZN F781(参见表VII,实施例I);

[0066] • ABCB1、c1orf95、CACNA1C、CACNG8、CHST2、ELM01、EMID2、FBN1_B、FLT3_A、FLT3_B、GLIS1、GPC6、GREM2、JAM2、KCNK12_A、LOC100129620、MAX.chr15.78112404-78112692、MAX.chr19.4584907-4585088、MAX.chr3.69591689-69591784、NCAM1、NT5C1A、ST8SIA3、ZNF382、ZNF419、ZNF69和ZSCAN18(参见表VIII,实施例I);和

[0067] • AK5、RABC3、ZNF491、ZNF610、ZNF91、ZNF480、TRPC3_B和ELMOD1(参见表X和XI,实施例II)。

[0068] 在某些实施方案中,该技术涉及评估生物样品(例如,组织样品(例如,宫颈组织)、血液样品、血浆样品、血清样品、全血样品、分泌物样品(宫颈分泌物、阴道分泌物)、器官分泌物样品、CSF样品、唾液样品、尿液样品或粪便样品)中一种或多种本文所述的MDM的存在和甲基化状态。这些MDM包括一种或多种如本文所讨论的差异甲基化区域(DMR),例如,如表I、III和X中提供的。在该技术的实施方案中评估甲基化状态。因此,本文提供的技术不限于测量基因的甲基化状态的方法,因此可以通过本领域已知的任何方法测量基因的甲基化状态。

[0069] 本文还提供了用于实施所述方法的组合物和试剂盒。例如,在一些实施方案中,对一种或多种MDM具有特异性的试剂(例如,引物、探针)单独或成组提供(例如,用于扩增多个标记物的引物对组)。还可以提供用于进行检测测定的额外试剂(例如,用于进行QuARTS、PCR、测序的酶、缓冲液、阳性和阴性对照、亚硫酸氢盐、10-11易位(TET)酶(例如,人TET1、人TET2、人TET3、鼠TET1、鼠TET2、鼠TET3、Naegleria TET(NgTET)、灰拟鬼伞(CcTET))或其变体)、有机硼烷或其他测定)。在一些实施方案中,试剂盒含有能够以甲基化特异性方式修饰DNA的试剂(例如,甲基化敏感的限制酶、甲基化依赖性限制酶和亚硫酸氢盐试剂)(例如,甲基化敏感的限制酶、甲基化依赖性限制酶、10-11易位(TET)酶(例如人TET1、人TET2、人TET3、鼠TET1、鼠TET2、鼠TET3、Naegleria TET(NgTET)、灰拟鬼伞(CcTET))或其变体)、有机硼烷)。在一些实施方案中,提供了含有进行方法所需的、足够的或有用的一种或多种试剂的试剂盒。还提供了含有试剂的反应混合物。还提供了含有多种试剂的预混试剂组,这些试剂可以相互添加和/或添加至测试样品中以完善反应混合物。在一些实施方案中,试剂盒包含对照核酸,该对照核酸包含来自DMR 1-423(来自表I、III和X)的一个或多个序列并且具有与患有宫颈癌、宫颈癌亚型(例如,宫颈腺癌、鳞状细胞宫颈癌)和/或宫颈癌前病变(例如,原位腺癌、宫颈上皮内瘤变)的受试者相关的甲基化状态。在一些实施方案中,试剂盒包含用于从受试者获得样品(例如,组织样品(例如,宫颈组织)、血液样品、血浆样品、血清样品、全血样品、分泌物样品(例如,宫颈分泌物、阴道分泌物)、器官分泌物样品、CSF样品、唾液样品、尿液样品或粪便样品)的样品收集器。在一些实施方案中,试剂盒包括如本文所述的寡核苷酸。

[0070] 定义

[0071] 为促进对本技术的理解,以下定义多个术语和短语。附加的定义在整个详述中阐述。

[0072] 在整个说明书和权利要求书中,下列术语采用本文明确关联的含义,除非上下文

另有明确说明。如本文所使用的短语“在一个实施方案中”不一定指代相同的实施方案,尽管它可以指代相同的实施方案。此外,本文所使用的短语“在另一实施方案中”不一定指代不同的实施方案,尽管它可以指代不同的实施方案。因此,如下所述,在不脱离本发明的范围或精神的情况下,可以容易地组合本发明的各种实施方案。

[0073] 另外,如本文所使用,术语“或”是包含性“或”运算符并且等同于术语“和/或”,除非上下文另外明确指出。术语“基于”不是排他性的,并且允许基于未描述的附加因素,除非上下文另外明确指出。另外,在整个说明书中,“一(a)”、“一(an)”和“所述”的含义包括复数引用物。“在……中”的含义包括“在……中”和“在……上”。

[0074] 本申请的权利要求中使用的过渡短语“基本上由……组成”将权利要求的范围限制为所要求保护的发明的指定材料或步骤“以及那些不会实质上影响基本和新型特征的材料或步骤”,如In re Herz, 537F.2d549, 551-52, 190USPQ 461, 463 (CCPA 1976) 中所述。例如,“基本上由”所列举的元素组成的组合物可以含有一定水平的未列举的污染物,使得尽管存在该污染物,但与纯组合物相比,该污染物不会改变所列举的组合物的功能,即“由所列举的组分组成”的组合物。

[0075] 如本文所用,术语“一个或多个”是指大于一的数字。例如,术语“一个或多个”涵盖以下任何一种:两个或更多、三个或更多、四个或更多、五个或更多、六个或更多、七个或更多、八个或更多、九个或更多、十个或更多、十二个或更多、十三个或更多、十四个或更多、十五个或更多、二十个或更多、五十个或更多、100个或更多、或甚至更多数量。

[0076] 术语“一个或多个但少于一个更大的数目”、“两个或多个但少于一个更大的数目”、“三个或更多但少于一个更大的数目”、“四个或更多但少于一个更大的数目”、“五个或更多但少于一个更大的数目”、“六个或更多但少于一个更大的数目”、“七个或更多但少于一个更大的数目”、“八个或更多但少于一个更大的数目”、“九个或更多但少于一个更大的数目”、“十个或更多但少于一个更大的数目”、“十一个或更多但少于一个更大的数目”、“十二个或更多但少于一个更大的数目”、“十三个或更多但少于一个更大的数目”、“十四个或更多但少于一个更大的数目”或“十五个或更多但少于一个更大的数目”不限于一个更大的数目。例如,更大的数目可以是10,000、1,000、100、50等。例如,更大的数目可以是约64 (例如,64、63、62、61、60、59、58、57、56、55、54、53、52、51、50、49、48、47、46、45、44、43、42、41、40、39、38、37、36、35、34、33、32、31、30、29、28、27、26、25、24、23、22、21、20、19、18、17、16、15、14、13、12、11、10、9、8、7、6、5、4、3或2)。例如,更大的数目可以为约423。

[0077] 术语“一个或多个甲基化标记物”或“一个或多个DMR”或“一个或多个基因”或“一个或多个标记物”或“多个甲基化标记物”或“多个标记物”或“多个基因”或“多个DMR”同样不限于特定的数字组合。实际上,考虑了甲基化标记物的任何数值组合 (例如,1-2个甲基化标记物、1-3、1-4、1-5、1-6、1-7、1-8、1-9、1-10、1-11、1-12、1-13、1-14、1-15、1-16、1-17、1-18、1-19、1-20、1-21、1-22、1-23、1-24、1-25、1-26、1-27、1-28、1-29、1-30、1-31、1-32、1-33、1-34、1-35、1-36、1-37、1-38、1-39、1-40、1-41、1-42、1-43、1-44、1-45、1-46、1-47、1-48、1-49、1-50、1-51、1-52、1-53、1-54、1-55、1-56、1-57、1-58、1-59、1-60、1-61、1-62、1-63、1-64) (例如,2-3、2-4、2-5、2-6、2-7、2-8、2-9、2-10、2-11、2-12、2-13、2-14、2-15、2-16、2-17、2-18、2-19、2-20、2-21、2-22、2-23、2-24、2-25、2-26、2-27、2-28、2-29、2-30、2-31、2-32、2-33、2-34、2-35、2-36、2-37、2-38、2-39、2-40、2-41、2-42、2-43、2-44、2-45、2-46、2-

47、2-48、2-49、2-50、2-51、2-52、2-53、2-54、2-55、2-56、2-57、2-58、2-59、2-60、2-61、2-62、2-63、2-64) (例如,3-4、3-5、3-6、3-7、3-8、3-9、3-10、3-11、3-12、3-13、3-14、3-15、3-16、3-17、3-18、3-19、3-20、3-21、3-22、3-23、3-24、3-25、3-26、3-27、3-28、3-29、3-30、3-31、3-32、3-33、3-34、3-35、3-36、3-37、3-38、3-39、3-40、3-41、3-42、3-43、3-44、3-45、3-46、3-47、3-48、3-49、3-50、3-51、3-52、3-53、3-54、3-55、3-56、3-57、3-58、3-59、3-60、3-61、3-62、3-63、3-64) (例如,4-5、4-6、4-7、4-8、4-9、4-10、4-11、4-12、4-13、4-14、4-15、4-16、4-17、4-18、4-19、4-20、4-21、4-22、4-23、4-24、4-25、4-26、4-27、4-28、4-29、4-30、4-31、4-32、4-33、4-34、4-35、4-36、4-37、4-38、4-39、4-40、4-41、4-42、4-43、4-44、4-45、4-46、4-47、4-48、4-49、4-50、4-51、4-52、4-53、4-54、4-55、4-56、4-57、4-58、4-59、4-60、4-61、4-62、4-63、4-64) (例如,5-6、5-7、5-8、5-9、5-10、5-11、5-12、5-13、5-14、5-15、5-16、5-17、5-18、5-19、5-20、5-21、5-22、5-23、5-24、5-25、5-26、5-27、5-28、5-29、5-30、5-31、5-32、5-33、5-34、5-35、5-36、5-37、5-38、5-39、5-40、5-41、5-42、5-43、5-44、5-45、5-46、5-47、5-48、5-49、5-50、5-51、5-52、5-53、5-54、5-55、5-56、5-57、5-58、5-59、5-60、5-61、5-62、5-63、5-64) (例如,6-7、6-8、6-9、6-10、6-11、6-12、6-13、6-14、6-15、6-16、6-17、6-18、6-19、6-20、6-21、6-22、6-23、6-24、6-25、6-26、6-27、6-28、6-29、6-30、6-31、6-32、6-33、6-34、6-35、6-36、6-37、6-38、6-39、6-40、6-41、6-42、6-43、6-44、6-45、6-46、6-47、6-48、6-49、6-50、6-51、6-52、6-53、6-54、6-55、6-56、6-57、6-58、6-59、6-60、6-61、6-62、6-63、6-64) (例如,7-8、7-9、7-10、7-11、7-12、7-13、7-14、7-15、7-16、7-17、7-18、7-19、7-20、7-21、7-22、7-23、7-24、7-25、7-26、7-27、7-28、7-29、7-30、7-31、7-32、7-33、7-34、7-35、7-36、7-37、7-38、7-39、7-40、7-41、7-42、7-43、7-44、7-45、7-46、7-47、7-48、7-49、7-50、7-51、7-52、7-53、7-54、7-55、7-56、7-57、7-58、7-59、7-60、7-61、7-62、7-63、7-64) (例如,8-9、8-10、8-11、8-12、8-13、8-14、8-15、8-16、8-17、8-18、8-19、8-20、8-21、8-22、8-23、8-24、8-25、8-26、8-27、8-28、8-29、8-30、8-31、8-32、8-33、8-34、8-35、8-36、8-37、8-38、8-39、8-40、8-41、8-42、8-43、8-44、8-45、8-46、8-47、8-48、8-49、8-50、8-51、8-52、8-53、8-54、8-55、8-56、8-57、8-58、8-59、8-60、8-61、8-62、8-63、8-64) (例如,9-10、9-11、9-12、9-13、9-14、9-15、9-16、9-17、9-18、9-19、9-20、9-21、9-22、9-23、9-24、9-25、9-26、9-27、9-28、9-29、9-30、9-31、9-32、9-33、9-34、9-35、9-36、9-37、9-38、9-39、9-40、9-41、9-42、9-43、9-44、9-45、9-46、9-47、9-48、9-49、9-50、9-51、9-52、9-53、9-54、9-55、9-56、9-57、9-58、9-59、9-60、9-61、9-62、9-63、9-64) (例如,10-11、10-12、10-13、10-14、10-15、10-16、10-17、10-18、10-19、10-20、10-21、10-22、10-23、10-24、10-25、10-26、10-27、10-28、10-29、10-30、10-31、10-32、10-33、10-34、10-35、10-36、10-37、10-38、10-39、10-40、10-41、10-42、10-43、10-44、10-45、10-46、10-47、10-48、10-49、10-50、10-51、10-52、10-53、10-54、10-55、10-56、10-57、10-58、10-59、10-60、10-61、10-62、10-63、10-64) (例如,11-12、11-13、11-14、11-15、11-16、11-17、11-18、11-19、11-20、11-21、11-22、11-23、11-24、11-25、11-26、11-27、11-28、11-29、11-30、11-31、11-32、11-33、11-34、11-35、11-36、11-37、11-38、11-39、11-40、11-41、11-42、11-43、11-44、11-45、11-46、11-47、11-48、11-49、11-50、11-51、11-52、11-53、11-54、11-55、11-56、11-57、11-58、11-59、11-60、11-61、11-62、11-63、11-64) (例如,12-13、12-14、12-15、12-16、12-17、12-18、12-19、12-20、12-21、12-22、12-23、12-24、12-25、12-26、12-27、12-28、12-29、12-30、12-31、12-32、12-33、12-34、12-35、

12-36、12-37、12-38、12-39、12-40、12-41、12-42、12-43、12-44、12-45、12-46、12-47、12-48、12-49、12-50、12-51、12-52、12-53、12-54、12-55、12-56、12-57、12-58、12-59、12-60、12-61、12-62、12-63、12-64) (例如, 13-14、13-15、13-16、13-17、13-18、13-19、13-20、13-21、13-22、13-23、13-24、13-25、13-26、13-27、13-28、13-29、13-30、13-31、13-32、13-33、13-34、13-35、13-36、13-37、13-38、13-39、13-40、13-41、13-42、13-43、13-44、13-45、13-46、13-47、13-48、13-49、13-50、13-51、13-52、13-53、13-54、13-55、13-56、13-57、13-58、13-59、13-60、13-61、13-62、13-63、13-64) (例如, 14-15、14-16、14-17、14-18、14-19、14-20、14-21、14-22、14-23、14-24、14-25、14-26、14-27、14-28、14-29、14-30、14-31、14-32、14-33、14-34、14-35、14-36、14-37、14-38、14-39、14-40、14-41、14-42、14-43、14-44、14-45、14-46、14-47、14-48、14-49、14-50、14-51、14-52、14-53、14-54、14-55、14-56、14-57、14-58、14-59、14-60、14-61、14-62、14-63、14-64) (例如, 15-16、15-17、15-18、15-19、15-20、15-21、15-22、15-23、15-24、15-25、15-26、15-27、15-28、15-29、15-30、15-31、15-32、15-33、15-34、15-35、15-36、15-37、15-38、15-39、15-40、15-41、15-42、15-43、15-44、15-45、15-46、15-47、15-48、15-49、15-50、15-51、15-52、15-53、15-54、15-55、15-56、15-57、15-58、15-59、15-60、15-61、15-62、15-63、15-64) (例如, 16-17、16-18、16-19、16-20、16-21、16-22、16-23、16-24、16-25、16-26、16-27、16-28、16-29、16-30、16-31、16-32、16-33、16-34、16-35、16-36、16-37、16-38、16-39、16-40、16-41、16-42、16-43、16-44、16-45、16-46、16-47、16-48、16-49、16-50、16-51、16-52、16-53、16-54、16-55、16-56、16-57、16-58、16-59、16-60、16-61、16-62、16-63、16-64) (例如, 17-18、17-19、17-20、17-21、17-22、17-23、17-24、17-25、17-26、17-27、17-28、17-29、17-30、17-31、17-32、17-33、17-34、17-35、17-36、17-37、17-38、17-39、17-40、17-41、17-42、17-43、17-44、17-45、17-46、17-47、17-48、17-49、17-50、17-51、17-52、17-53、17-54、17-55、17-56、17-57、17-58、17-59、17-60、17-61、17-62、17-63、17-64) (例如, 18-19、18-20、18-21、18-22、18-23、18-24、18-25、18-26、18-27、18-28、18-29、18-30、18-31、18-32、18-33、18-34、18-35、18-36、18-37、18-38、18-39、18-40、18-41、18-42、18-43、18-44、18-45、18-46、18-47、18-48、18-49、18-50、18-51、18-52、18-53、18-54、18-55、18-56、18-57、18-58、18-59、18-60、18-61、18-62、18-63、18-64) (例如, 19-20、19-21、19-22、19-23、19-24、19-25、19-26、19-27、19-28、19-29、19-30、19-31、19-32、19-33、19-34、19-35、19-36、19-37、19-38、19-39、19-40、19-41、19-42、19-43、19-44、19-45、19-46、19-47、19-48、19-49、19-50、19-51、19-52、19-53、19-54、19-55、19-56、19-57、19-58、19-59、19-60、19-61、19-62、19-63、19-64) (例如, 20-21、20-22、20-23、20-24、20-25、20-26、20-27、20-28、20-29、20-30、20-31、20-32、20-33、20-34、20-35、20-36、20-37、20-38、20-39、20-40、20-41、20-42、20-43、20-44、20-45、20-46、20-47、20-48、20-49、20-50、20-51、20-52、20-53、20-54、20-55、20-56、20-57、20-58、20-59、20-60、20-61、20-62、20-63、20-64) (例如, 21-22、21-23、21-24、21-25、21-26、21-27、21-28、21-29、21-30、21-31、21-32、21-33、21-34、21-35、21-36、21-37、21-38、21-39、21-40、21-41、21-42、21-43、21-44、21-45、21-46、21-47、21-48、21-49、21-50、21-51、21-52、21-53、21-54、21-55、21-56、21-57、21-58、21-59、21-60、21-61、21-62、21-63、21-64) (例如, 22-23、22-24、22-25、22-26、22-27、22-28、22-29、22-30、22-31、22-32、22-33、22-34、22-35、22-36、22-37、22-38、22-39、22-40、22-41、22-42、22-43、22-44、22-45、22-46、22-47、22-

48、22-49、22-50、22-51、22-52、22-53、22-54、22-55、22-56、22-57、22-58、22-59、22-60、22-61、22-62、22-63、22-64) (例如, 23-24、23-25、23-26、23-27、23-28、23-29、23-30、23-31、23-32、23-33、23-34、23-35、23-36、23-37、23-38、23-39、23-40、23-41、23-42、23-43、23-44、23-45、23-46、23-47、23-48、23-49、23-50、23-51、23-52、23-53、23-54、23-55、23-56、23-57、23-58、23-59、23-60、23-61、23-62、23-63、23-64) (例如, 24-25、24-26、24-27、24-28、24-29、24-30、24-31、24-32、24-33、24-34、24-35、24-36、24-37、24-38、24-39、24-40、24-41、24-42、24-43、24-44、24-45、24-46、24-47、24-48、24-49、24-50、24-51、24-52、24-53、24-54、24-55、24-56、24-57、24-58、24-59、24-60、24-61、24-62、24-63、24-64) (例如, 25-26、25-27、25-28、25-29、25-30、25-31、25-32、25-33、25-34、25-35、25-36、25-37、25-38、25-39、25-40、25-41、25-42、25-43、25-44、25-45、25-46、25-47、25-48、25-49、25-50、25-51、25-52、25-53、25-54、25-55、25-56、25-57、25-58、25-59、25-60、25-61、25-62、25-63、25-64) (例如, 26-27、26-28、26-29、26-30、26-31、26-32、26-33、26-34、26-35、26-36、26-37、26-38、26-39、26-40、26-41、26-42、26-43、26-44、26-45、26-46、26-47、26-48、26-49、26-50、26-51、26-52、26-53、26-54、26-55、26-56、26-57、26-58、26-59、26-60、26-61、26-62、26-63、26-64) (例如, 27-28、27-29、27-30、27-31、27-32、27-33、27-34、27-35、27-36、27-37、27-38、27-39、27-40、27-41、27-42、27-43、27-44、27-45、27-46、27-47、27-48、27-49、27-50、27-51、27-52、27-53、27-54、27-55、27-56、27-57、27-58、27-59、27-60、27-61、27-62、27-63、27-64) (例如, 28-29、28-30、28-31、28-32、28-33、28-34、28-35、28-36、28-37、28-38、28-39、28-40、28-41、28-42、28-43、28-44、28-45、28-46、28-47、28-48、28-49、28-50、28-51、28-52、28-53、28-54、28-55、28-56、28-57、28-58、28-59、28-60、28-61、28-62、28-63、28-64) (例如, 29-30、29-31、29-32、29-33、29-34、29-35、29-36、29-37、29-38、29-39、29-40、29-41、29-42、29-43、29-44、29-45、29-46、29-47、29-48、29-49、29-50、29-51、29-52、29-53、29-54、29-55、29-56、29-57、29-58、29-59、29-60、29-61、29-62、29-63、29-64) (例如, 30-31、30-32、30-33、30-34、30-35、30-36、30-37、30-38、30-39、30-40、30-41、30-42、30-43、30-44、30-45、30-46、30-47、30-48、30-49、30-50、30-51、30-52、30-53、30-54、30-55、30-56、30-57、30-58、30-59、30-60、30-61、30-62、30-63、30-64) (例如, 31-32、31-33、31-34、31-35、31-36、31-37、31-38、31-39、31-40、31-41、31-42、31-43、31-44、31-45、31-46、31-47、31-48、31-49、31-50、31-51、31-52、31-53、31-54、31-55、31-56、31-57、31-58、31-59、31-60、31-61、31-62、31-63、31-64) (例如, 32-33、32-34、32-35、32-36、32-37、32-38、32-39、32-40、32-41、32-42、32-43、32-44、32-45、32-46、32-47、32-48、32-49、32-50、32-51、32-52、32-53、32-54、32-55、32-56、32-57、32-58、32-59、32-60、32-61、32-62、32-63、32-64) (例如, 33-34、33-35、33-36、33-37、33-38、33-39、33-40、33-41、33-42、33-43、33-44、33-45、33-46、33-47、33-48、33-49、33-50、33-51、33-52、33-53、33-54、33-55、33-56、33-57、33-58、33-59、33-60、33-61、33-62、33-63、33-64) (例如, 34-35、34-36、34-37、34-38、34-39、34-40、34-41、34-42、34-43、34-44、34-45、34-46、34-47、34-48、34-49、34-50、34-51、34-52、34-53、34-54、34-55、34-56、34-57、34-58、34-59、34-60、34-61、34-62、34-63、34-64) (例如, 35-36、35-37、35-38、35-39、35-40、35-41、35-42、35-43、35-44、35-45、35-46、35-47、35-48、35-49、35-50、35-51、35-52、35-53、35-54、35-55、35-56、35-57、35-58、35-59、35-60、35-61、35-62、35-63、35-64) (例如, 36-37、36-38、

36-39、36-40、36-41、36-42、36-43、36-44、36-45、36-46、36-47、36-48、36-49、36-50、36-51、36-52、36-53、36-54、36-55、36-56、36-57、36-58、36-59、36-60、36-61、36-62、36-63、36-64) (例如, 37-38、37-39、37-40、37-41、37-42、37-43、37-44、37-45、37-46、37-47、37-48、37-49、37-50、37-51、37-52、37-53、37-54、37-55、37-56、37-57、37-58、37-59、37-60、37-61、37-62、37-63、37-64) (例如, 38-39、38-40、38-41、38-42、38-43、38-44、38-45、38-46、38-47、38-48、38-49、38-50、38-51、38-52、38-53、38-54、38-55、38-56、38-57、38-58、38-59、38-60、38-61、38-62、38-63、38-64)、(例如, 39-40、39-41、39-42、39-43、39-44、39-45、39-46、39-47、39-48、39-49、39-50、39-51、39-52、39-53、39-54、39-55、39-56、39-57、39-58、39-59、39-60、39-61、39-62、39-63、39-64)、(例如, 40-41、40-42、40-43、40-44、40-45、40-46、40-47、40-48、40-49、40-50、40-51、40-52、40-53、40-54、40-55、40-56、40-57、40-58、40-59、40-60、40-61、40-62、40-63、40-64)、(例如, 41-42、41-43、41-44、41-45、41-46、41-47、41-48、41-49、41-50、41-51、41-52、41-53、41-54、41-55、41-56、41-57、41-58、41-59、41-60、41-61、41-62、41-63、41-64)、(例如, 42-43、42-44、42-45、42-46、42-47、42-48、42-49、42-50、42-51、42-52、42-53、42-54、42-55、42-56、42-57、42-58、42-59、42-60、42-61、42-62、42-63、42-64) (例如, 43-44、43-45、43-46、43-47、43-48、43-49、43-50、43-51、43-52、43-53、43-54、43-55、43-56、43-57、43-58、43-59、43-60、43-61、43-62、43-63、43-64) (例如, 44-45、44-46、44-47、44-48、44-49、44-50、44-51、44-52、44-53、44-54、44-55、44-56、44-57、44-58、44-59、44-60、44-61、44-62、44-63、44-64) (例如, 45-46、45-47、45-48、45-49、45-50、45-51、45-52、45-53、45-54、45-55、45-56、45-57、45-58、45-59、45-60、45-61、45-62、45-63、45-64) (例如, 46-47、46-48、46-49、46-50、46-51、46-52、46-53、46-54、46-55、46-56、46-57、46-58、46-59、46-60、46-61、46-62、46-63、46-64) (例如, 47-48、47-49、47-50、47-51、47-52、47-53、47-54、47-55、47-56、47-57、47-58、47-59、47-60、47-61、47-62、47-63、47-64) (例如, 48-49、48-50、48-51、48-52、48-53、48-54、48-55、48-56、48-57、48-58、48-59、48-60、48-61、48-62、48-63、48-64) (例如, 49-50、49-51、49-52、49-53、49-54、49-55、49-56、49-57、49-58、49-59、49-60、49-61、49-62、49-63、49-64) (例如, 50-51、50-52、50-53、50-54、50-55、50-56、50-57、50-58、50-59、50-60、50-61、50-62、50-63、50-64) (例如, 51-52、51-53、51-54、51-55、51-56、51-57、51-58、51-59、51-60、51-61、51-62、51-63、51-64) (例如, 52-53、52-54、52-55、52-56、52-57、52-58、52-59、52-60、52-61、52-62、52-63、52-64) (例如, 53-54、53-55、53-56、53-57、53-58、53-59、53-60、53-61、53-62、53-63、53-64) (例如, 54-55、54-56、54-57、54-58、54-59、54-60、54-61、54-62、54-63、54-64) (例如, 55-56、55-57、55-58、55-59、55-60、55-61、55-62、55-63、55-64) (例如, 56-57、56-58、56-59、56-60、56-61、56-62、56-63、56-64) (例如, 57-58、57-59、57-60、57-61、57-62、57-63、57-64) (例如, 58-59、58-60、58-61、58-62、58-63、58-64) (例如, 59-60、59-61、59-62、59-63、59-64) (例如, 60-61、60-62、60-63、60-64) (例如, 61-62、61-63、61-64) (例如, 62-63、62-64) (例如, 63-64) (例如, 1-423、10-423、20-423、41-423、50-423、62-423、64-423、100-423、150-423、200-423、300-423、320-423、361-423、400-423、410-423、415-423、420-423、421-423、422-423) (例如, 423或更少、400或更少、350或更少、361或更少、320或更少、300或更少、250或更少、200或更少、150或更少、100或更少、75或更少、70或更少、65或更少、64或更少、62或更少; 61或更少; 60或更少; 59或更少; 58或更少; 57或更少;

56或更少;55或更少;54或更少;53或更少;52或更少;51或更少;50或更少;49或更少;48或更少;47或更少;46或更少;45或更少;44或更少;43或更少;42或更少;41或更少;40或更少;39或更少;38或更少;37或更少;36或更少;35或更少;34或更少;33或更少;32或更少;31或更少;30或更少;29或更少;28或更少;27或更少;26或更少;25或更少;24或更少;23或更少;22或更少;21或更少;20或更少;19或更少;18或更少;17或更少;16或更少;15或更少;14或更少;13或更少;12或更少;11或更少;10或更少;9或更少;8或更少;7或更少;6或更少;5或更少;4或更少;3或更少;2或1)。

[0078] 如本文所用,“核酸”或“核酸分子”一般是指任何核糖核酸或脱氧核糖核酸,其可以是未修饰的或修饰的DNA或RNA。“核酸”包括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碱基的如上所述的DNA。因此,为了稳定性或其他原因而对骨架进行修饰的DNA是“核酸”。如本文所用,术语“核酸”涵盖这样的化学、酶促或代谢修饰形式的核酸,以及病毒和细胞(包括例如简单和复杂细胞)特征性的DNA化学形式。

[0079] 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,优选地超过三个和通常超过十个脱氧核糖核苷酸或核糖核苷酸的分子。确切的大小将取决于许多因素,而这些因素又取决于寡核苷酸的最终功能或用途。寡核苷酸可以通过任何方式产生,包括化学合成、DNA复制、逆转录或它们的组合。DNA的典型脱氧核糖核苷酸是胸腺嘧啶、腺嘌呤、胞嘧啶和鸟嘌呤。RNA的典型核糖核苷酸是尿嘧啶、腺嘌呤、胞嘧啶和鸟嘌呤。

[0080] 如本文所用,术语核酸的“基因座”或“区域”是指核酸的亚区域,例如染色体上的基因、单个核苷酸、CpG岛等。

[0081] 术语“互补”和“互补性”是指通过碱基配对规则相关的核苷酸(例如1个核苷酸)或多核苷酸(例如核苷酸序列)。例如,序列5'-A-G-T-3'与序列3'-T-C-A-5'互补。互补性可以是“部分的”,其中仅一些核酸碱基根据碱基配对规则匹配。或者,核酸之间可能存在“完全”或“全部”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。这在依赖于核酸之间的结合的扩增反应和检测方法中特别重要。

[0082] 术语“基因”是指包含产生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特性(例如酶活性、配体结合、信号转导等)即可。当用于指代基因时,术语“部分”是指该基因的片段。片段的大小可以在几个核苷酸至整个基因序列减去一个核苷酸的范围内变化。因此,“包含基因至少一部分的核苷酸”可以包含基因片段或整个基因。

[0083] 术语“基因”还涵盖结构基因的编码区,并且包括在5'和3'末端均邻近编码区定位的序列,例如在任一端距离约1kb,使得基因对应于全长mRNA的长度(例如,包含编码序列、调控序列、结构序列和其他序列)。位于编码区5'并且存在于mRNA上的序列称为5'非翻译或未翻译序列。位于编码区3'或下游并且存在于mRNA上的序列称为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。在一些生物体(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。内含子是转录成核RNA(hnRNA)的基因区段;内含子可以含有调控元件,例如增强子。内含子从核转录本或初级转录本中移出或“剪除”;因此,在信使RNA(mRNA)转录本中不存在内含子。mRNA在翻译过程中发挥作用,以指定新生多肽中氨基酸的序列或顺序。

[0084] 除含有内含子外,基因的基因组形式还可以包括位于RNA转录本上存在的序列的

5'和3'末端的序列。这些序列被称为“侧接”序列或区域(这些侧接序列位于mRNA转录本上存在的非翻译序列的5'或3')。5'侧接区域可以含有调控序列,例如启动子和强化子,其控制或影响基因的转录。3'侧接区域可以含有指导转录终止、转录后裂解和多腺苷酸化的序列。

[0085] 在提及基因时术语“野生型”是指具有从天然存在的来源分离的基因特征的基因。在提及基因产物时术语“野生型”是指具有从天然存在的来源分离的基因产物特征的基因产物。在提及蛋白时术语“野生型”是指具有天然存在的蛋白的特征的蛋白。如应用于物体的术语“天然存在”是指可以在自然界中发现物体的事实。例如,存在于生物体(包括病毒)中的可以从自然界来源分离并且未经实验室人员有意修饰的多肽或多核苷酸序列是天然存在的。野生型基因通常是在群体中最常观察到的基因或等位基因,因此被任意指定为基因的“正常”或“野生型”形式。相比之下,在提及基因或基因产物时术语“修饰的”或“突变的”分别是指与野生型基因或基因产物相比显示序列和/或功能特性修饰(例如特征改变)的基因或基因产物。注意,可以分离天然存在的突变体;这些通过它们与野生型基因或基因产物相比特征改变的事实来鉴定。

[0086] 术语“等位基因”是指基因的变异;所述变异包括但不限于变体和突变体、多态基因座和单核苷酸多态基因座、移码和剪接突变。等位基因可能自然存在于群体中,或可能在群体中任何特定个体的一生中出现。

[0087] 因此,在提及核苷酸序列时使用术语“变体”和“突变体”是指与另一个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”是两个不同核苷酸序列之间的差异;典型地,一个序列是参考序列。

[0088] 术语“引物”指寡核苷酸,无论是天然存在的,例如来自限制性消化的核酸片段,还是合成产生的,当置于诱导合成与核酸模板链互补的引物延伸产物的条件下时,其能够充当合成的起始点,(例如,在核苷酸和诱导剂例如DNA聚合酶的存在下,并且在合适的温度和pH下)。为了最大的扩增效率,引物优选是单链的,但可以替代地是双链的。如果是双链,则在用于制备延伸产物之前首先对引物进行处理以分离其链。优选地,引物是寡脱氧核糖核苷酸。引物必须足够长以在诱导剂存在下引发延伸产物的合成。引物的确切长度取决于许多因素,包括温度、引物来源和方法的使用。在一些实施方案中,引物对特定MDM(例如,表I、III和X中的MDM)具有特异性,并且特异性结合包含MDM的基因区域的至少一部分(例如,表I、III和/或X中的染色体坐标)。

[0089] 术语“探针”是指能够与另一感兴趣的寡核苷酸杂交的寡核苷酸(例如,核苷酸序列),无论是天然存在的(如纯化的限制性消化物)还是合成、重组或通过PCR扩增产生的。探针可以是单链或双链的。探针可用于检测、鉴定和分离特定基因序列(例如“捕获探针”)。预期在一些实施方案中,本发明中使用的任何探针可以用任何“报告分子”标记,以便在任何检测系统中可检测,包括但不限于酶(例如,ELISA,以及基于酶的组织化学测定)、荧光、放射性和发光系统。本发明无意限于任何特定的检测系统或标记。

[0090] 如本文所用,术语“靶标”是指寻求例如通过探针结合、扩增、分离、捕获等从其他核酸中分选出来的核酸。例如,当用于提及聚合酶链式反应时,“靶标”是指由用于聚合酶链式反应的引物结合的核酸区域,而当用于其中靶DNA不被扩增的测定时,例如,在侵入性切割测定的一些实施方案中,靶标包含探针和侵入性寡核苷酸(例如INVADER寡核苷酸)结合

以形成侵入性切割结构的位点,使得可以检测靶核酸的存在。“区段”定义为靶序列内的核酸区域。

[0091] 因此,如本文所用,“非靶标”,例如,当其用于描述核酸如DNA时,是指可能存在于反应中但不是通过反应检测或表征的对象的核酸。在一些实施方案中,非靶核酸可以指例如不含有靶序列的样品中存在的核酸,而在一些实施方案中,非靶核酸可以指外源核酸,即不是源自含有或疑似含有靶核酸的样品的核酸,并且被添加到反应中,例如以使酶(例如聚合酶)的活性标准化以减少反应中酶的性能的变异性。

[0092] 如本文所用,“甲基化”是指胞嘧啶的C5或N4位置的胞嘧啶甲基化、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,因为典型的体外DNA扩增方法不保留扩增模板的甲基化模式。然而,“未甲基化DNA”或“甲基化DNA”也可以分别指原始模板未甲基化或甲基化的扩增DNA。

[0093] 本文所用的术语“扩增试剂”是指除引物、核酸模板和扩增酶之外的扩增所需的那些试剂(脱氧核糖核苷三磷酸、缓冲液等)。通常,扩增试剂与其他反应组分一起放置并包含在反应容器中。

[0094] 如本文所用,术语“对照”当在提及核酸检测或分析时被使用时是指具有已知特征(例如,已知序列、已知每个细胞的拷贝数)的核酸,用于与实验靶标(例如,未知浓度的核酸)进行比较。对照可以是内源性、优选不变的基因,可以针对其标准化测定中的测试或靶核酸。这种标准化控制可能发生在例如样品处理、测定效率等中的样品间变化,并且允许准确的样品间数据比较。发现可用于对人类样品进行标准化核酸检测测定的基因包括例如b-肌动蛋白、ZDHHC1和B3GALT6(参见例如美国专利申请序列号14/966,617和62/364,082,各自通过引用并入本文)。如本文所用,“ZDHHC1”指编码特征为锌指的蛋白质的基因,含DHHC型1,位于Chr 16(16q22.1)上的人DNA并且属于DHHC棕榈酰转移酶家族。

[0095] 对照也可以是外部的。例如,在定量测定如qPCR、QuARTS等中,“校准品”或“校准对照”是已知序列的核酸,例如具有与实验靶核酸的一部分相同的序列,以及已知浓度或系列浓度(例如,用于在定量PCR中生成校准曲线的连续稀释的对照靶标)。通常,使用与实验DNA相同的试剂和反应条件来分析校准对照。在某些实施方案中,校准品的测量与实验测定同时进行,例如在同一热循环仪中进行。在优选的实施方案中,多个校准品可以包含在单个质粒中,使得可以容易地以等摩尔量提供不同的校准品序列。在特别优选的实施方案中,例如用一种或多种限制酶消化质粒校准品,以从质粒载体释放校准品部分。参见例如WO 2015/066695,其通过引用并入本文。

[0096] 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公认的典型核苷酸碱基中。例如,胞嘧啶在其嘧啶环上不含有甲基部分,但5-甲基胞嘧啶在其嘧啶环的5位上含有甲基部分。因此,胞嘧啶不是甲基化核苷酸,5-甲基胞嘧啶是甲基化核苷酸。在另一个实例中,胸腺嘧啶在其嘧啶环的5位上含有甲基部分;然而,出于本文的目的,当存在于DNA中时,胸腺嘧啶不被认为是甲基化核苷酸,因为胸腺嘧啶是DNA的典型核苷酸碱基。

[0097] 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。

[0098] 如本文所用,核酸分子的“甲基化状态”、“甲基化概况”和“甲基化状态”是指核酸分子中是否存在一个或多个甲基化核苷酸碱基。例如,含有甲基化胞嘧啶的核酸分子被认

为是甲基化的(例如,核酸分子的甲基化状态是甲基化的)。不含任何甲基化核苷酸的核酸分子被认为是未甲基化的。

[0099] 如本文所用,应用于甲基化标记物的术语“甲基化水平”是指特定甲基化标记物内的甲基化量。甲基化水平还可以指与已建立的标准或对照相比的特定甲基化标记物内的甲基化量。甲基化水平还可以指CpG背景中存在的一个或多个胞嘧啶残基是否具有或不具有甲基化基团。甲基化水平还可以指样品中在此类胞嘧啶上具有或不具有甲基化基团的细胞的分数。替代地,甲基化水平还可以描述单个CpG二核苷酸是否被甲基化。

[0100] 特定核酸序列(例如,如本文所述的基因标记物或DNA区域)的甲基化状态可以指示序列中每个碱基的甲基化状态,或者可以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,或者可以指示关于序列内区域甲基化密度的信息,提供或不提供序列内发生甲基化的位置的精确信息。

[0101] 核酸分子中核苷酸基因座的甲基化状态是指核酸分子中特定基因座处甲基化核苷酸的存在或不存在。例如,当核酸分子中第7个核苷酸处存在的核苷酸是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被甲基化。同样,当核酸分子中第7个核苷酸处存在的核苷酸是胞嘧啶(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未被甲基化。

[0102] 甲基化状态可以任选地由“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)。甲基化值可以例如通过定量用甲基化依赖性限制性酶进行限制性消化后存在的完整核酸的量或通过比较亚硫酸氢盐反应后的扩增谱或通过比较经亚硫酸氢盐处理和未处理的核酸的序列或通过比较经TET处理和未处理的核酸来产生。因此,值,例如甲基化值,代表甲基化状况并且因此可以用作跨基因座的多个拷贝的甲基化状况的定量指标。当需要将样品中序列的甲基化状态与阈值或参考值进行比较时,这特别有用。

[0103] 如本文所用,“甲基化频率”或“甲基化百分比(%)”是指分子或基因座甲基化的实例数相对于分子或基因座未甲基化的实例数。

[0104] 本文使用的术语“甲基化得分”是指示标记物或标记物组中检测到的甲基化事件的得分,与来自没有特定感兴趣肿瘤的随机哺乳动物群体(例如10、20、30、40、50、100或500只哺乳动物的随机群体)的标记物或标记物组的中位甲基化事件进行比较。标记物或标记物组中升高的甲基化得分可以是任何得分,只要该得分大于相应的参考得分即可。例如,标记物或标记物组中升高的甲基化得分可以比参考甲基化得分高0.5、1、2、3、4、5、6、7、8、9、10或更多倍。

[0105] 因此,甲基化状态描述了核酸(例如,基因组序列)的甲基化状态。此外,甲基化状态是指与甲基化相关的特定基因组基因座处的核酸区段的特征。这些特征包括但不限于该DNA序列内的任何胞嘧啶(C)残基是否被甲基化、甲基化C残基的位置、在DNA序列的任何特定区域中甲基化C的频率或百分比。核酸,以及由于等位基因起源的差异而导致的等位基因甲基化差异。术语“甲基化状态”、“甲基化谱”和“甲基化状态”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。例如,如果核酸序列内的一个或多个胞嘧啶(C)残基被甲基化,则可以被称为“高甲基化”或具有“增加的甲基化”,而如果DNA序列内的一个或多个胞嘧啶(C)残基序列未被甲基化,可以被称为“低甲基化”或具有“减少的甲基化”。同样地,如果核酸序列内的一个或多个胞嘧啶(C)残基与另一个核酸序列

(例如,来自不同区域或来自不同个体等)相比被甲基化,则该序列被认为与其他核酸序列相比是高甲基化的或具有增加的甲基化。替代地,如果DNA序列内的一个或多个胞嘧啶(C)残基与另一个核酸序列(例如,来自不同区域或来自不同个体等)相比未被甲基化,则该序列被认为与其他核酸序列相比是低甲基化的或具有减少的甲基化。此外,如本文所用的术语“甲基化模式”是指核酸区域上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,两个核酸可能具有相同或相似的甲基化频率或甲基化百分比但具有不同的甲基化模式。当序列在甲基化程度(例如,一个序列相对于另一个序列具有增加或减少的甲基化)、频率或模式不同时,这些序列被称为“差异甲基化”或具有“甲基化差异”或具有“不同甲基化状态”。术语“差异甲基化”是指与癌症阴性样品中的核酸甲基化水平或模式相比,癌症阳性样品中的核酸甲基化水平或模式的差异。它还可能指手术后癌症复发的患者与未复发的患者之间水平或模式的差异。差异甲基化和DNA甲基化的特定水平或模式是预后和预测性生物标记物,例如,一旦确定了正确的截止或预测特征。

[0106] 甲基化状态频率可用于描述个体群体或来自单个个体的样品。例如,甲基化状态频率为50%的核苷酸基因座在50%的情况下是甲基化的,在50%的情况下是未甲基化的。这样的频率可用于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被甲基化的程度。因此,当核酸分子的第一群体或池中的甲基化不同于核酸分子的第二群体或池中的甲基化时,第一群体或池的甲基化状态频率将不同于第二群体或池的甲基化状态频率。例如,这样的频率还可以用于描述单个个体中核苷酸基因座或核酸区域被甲基化的程度。例如,这样的频率可用于描述来自组织样品的一组细胞在核苷酸基因座或核酸区域处被甲基化或未甲基化的程度。

[0107] 通常,人类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,其中胞嘧啶位于鸟嘌呤的5' (也称为CpG二核苷酸序列)。CpG二核苷酸内的大多数胞嘧啶在人类基因组中被甲基化,然而,一些在特定的富含CpG二核苷酸的基因组区域中保持未甲基化,称为CpG岛(参见,例如,Antequera等人(1990)Cell 62:503-514)。

[0108] 如本文所用,“CpG岛”或“胞嘧啶-磷酸-鸟嘌呤岛”是指基因组DNA的富含G:C的区域,其含有相对于总基因组DNA数量增加的CpG二核苷酸。CpG岛的长度可以是至少100、200或更多个碱基对,其中该区域的G:C含量为至少50%,并且观察到的CpG频率与预期频率的比率为0.6;在一些情况下,CpG岛的长度可以是至少500个碱基对,其中该区域的G:C含量是至少55%,并且观察到的CpG频率与预期频率的比率是0.65。观察到的CpG频率相对于预期频率可以根据Gardiner-Garden等人(1987)J.Mol.Biol.196:261-281中提供的方法计算。例如,观察到的CpG频率相对于预期频率可以根据公式 $R = (A \times B) / (C \times D)$ 计算,其中R是观察到的CpG频率与预期频率的比率,A是分析序列中CpG二核苷酸的数量,B是分析序列中的核苷酸总数,C是分析序列中C核苷酸的总数,D是分析序列中G核苷酸的总数。典型地在CpG岛中,例如在启动子区域确定甲基化状态。应当理解,人类基因组中的其他序列例如CpA和CpT容易发生DNA甲基化(参见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)。

[0109] 如本文所用,“甲基化特异性试剂”是指根据核酸分子的甲基化状态来修饰核酸分子的核苷酸的试剂,或甲基化特异性试剂是指可以以反映核酸分子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试剂。用此类试剂处理核酸分子的方法可包括使核酸分子与试剂接触,如果需要的话,结合额外的步骤,以实现核苷酸序列的所需改变。此类方法可以以将未甲基化的核苷酸(例如,每个未甲基化的胞嘧啶)修饰成不同的核苷酸的方式应用。例如,在一些实施方案中,这种试剂可以使未甲基化的胞嘧啶核苷酸脱氨基,产生脱氧尿嘧啶残基。此类试剂的实例包括但不限于甲基化敏感性限制酶、甲基化依赖性限制酶、亚硫酸氢盐试剂、TET酶和硼烷还原剂。

[0110] 甲基化特异性试剂对核酸核苷酸序列的改变也可以产生核酸分子,其中每个甲基化核苷酸被修饰为不同的核苷酸。

[0111] 术语“甲基化测定”是指用于确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测定。

[0112] 术语“MS AP-PCR”(甲基化敏感任意引物聚合酶链式反应)是指本领域公认的技术,该技术允许使用富含CG的引物对基因组进行全局扫描,以重点关注最可能含有CpG二核苷酸的区域,如Gonzalgo等人(1997)Cancer Research 57:594-599所述。

[0113] 术语“MethyLight™”是指Eads等人(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。

[0114] 术语“HeavyMethyl™”是指其中覆盖扩增引物之间或被扩增引物覆盖的CpG位置的甲基化特异性阻断探针(在本文中也称为阻断剂)能够实现核酸样品的甲基化特异性选择性扩增的测定。

[0115] 术语“HeavyMethyl™MethyLight™”测定是指HeavyMethyl™MethyLight™测定,其是MethyLight™测定的变体,其中MethyLight™测定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。

[0116] 术语“Ms-SNuPE”(甲基化敏感单核苷酸引物延伸)是指Gonzalgo&Jones(1997)Nucleic Acids Res.25:2529-2531描述的本领域公认的测定。

[0117] 术语“MSP”(甲基化特异性PCR)是指由Herman等人(1996)Proc.Natl.Acad.Sci.USA 93:9821-9826和美国专利号5,786,146描述的本领域公认的甲基化测定。

[0118] 术语“COBRA”(组合亚硫酸氢盐限制性分析)是指由Xiong&Laird(1997)Nucleic Acids Res.25:2532-2534描述的本领域公认的甲基化测定。

[0119] 术语“MCA”(甲基化CpG岛扩增)是指Toyota等人(1999)Cancer Res.59:2307-12和WO 00/26401A1描述的甲基化测定。

[0120] 如本文所用,“选定核苷酸”是指核酸分子中四种典型存在的核苷酸中的一种核苷酸(DNA为C、G、T和A, RNA为C、G、U和A),并且可以包括典型存在的核苷酸的甲基化衍生物(例如,当C是选定核苷酸时,甲基化和未甲基化的C都包括在选定核苷酸的含义内),而甲基化的选定核苷酸特指甲基化的典型存在的核苷酸并且未甲基化的选定核苷酸特指未甲基化的典型存在的核苷酸。

[0121] 术语“甲基化特异性限制酶”是指根据核酸识别位点的甲基化状态选择性消化核酸的限制酶。在识别位点未甲基化或半甲基化(甲基化敏感性酶)时特异性切割的限制酶的

情况下,如果识别位点在一条或两条链上是甲基化的,那么切割不会发生(或会发生,但效率显著降低)。在只有识别位点甲基化时才特异性切割的限制酶的情况下(甲基化依赖性酶),如果识别位点未甲基化,那么切割不会发生(或会发生,但效率显著降低)。优选甲基化特异性限制酶,其识别序列含有CG二核苷酸(例如识别序列如CGCG或CCCGGG)。对于一些实施方案进一步优选的是如果该二核苷酸中的胞嘧啶在碳原子C5处被甲基化则不会切割的限制酶。

[0122] 如本文所用,给定标记物(或一起使用的标记物组)的“敏感性”是指报告DNA甲基化值高于区分肿瘤和非肿瘤样品的阈值的样品的百分比。在一些实施方案中,阳性被定义为报告DNA甲基化值高于阈值(例如,与疾病相关的范围)的经组织学证实的瘤形成,并且假阴性被定义为报告DNA甲基化值低于阈值(例如,与无疾病相关的范围)的经组织学证实的瘤形成。因此,敏感度值反映了从已知患病样品中获得的给定标记物的DNA甲基化测量值处于疾病相关测量范围内的概率。如此处所定义的,所计算的敏感度值的临床相关性表示当将给定标记物应用于患有病症的受试者时将检测到该临床病症的存在的概率的估计。

[0123] 如本文所用,给定标记物(或一起使用的标记物组)的“特异性”是指报告DNA甲基化值低于区分肿瘤和非肿瘤样品的阈值的非肿瘤样品的百分比。在一些实施方案中,阴性被定义为报告DNA甲基化值低于阈值(例如,与无疾病相关的范围)的经组织学证实的非肿瘤样品,并且假阳性被定义为报告DNA甲基化值高于阈值(例如,与疾病相关的范围)的经组织学证实的非肿瘤样品。因此,特异性值反映了从已知非肿瘤样品中获得的给定标记物的DNA甲基化测量值处于非疾病相关测量范围内的概率。如此处所定义的,所计算的特异性值的临床相关性表示当将给定标记物应用于未患有病症的患者时将检测到该临床病症的不存在的概率的估计。

[0124] 如本文所用,术语“AUC”是“曲线下面积”的缩写。其尤其是指接受者操作特征(ROC)曲线下面积。ROC曲线是针对诊断测试的不同可能切点的真阳性率与假阳性率的关系图。它显示了敏感度和特异性之间的权衡,具体取决于所选的切割点(敏感度的任何增加都将伴随着特异性的降低)。ROC曲线下面积(AUC)是诊断测试准确性的量度(面积越大越好;最佳值为1;随机测试的ROC曲线位于对角线上,面积为0.5;供参考:J.P.Egan.(1975) Signal Detection Theory and ROC Analysis,Academic Press,New York)。

[0125] 如本文所用,术语“赘生物”是指组织的任何新的异常生长。因此,赘生物可以是癌前赘生物或恶性赘生物。

[0126] 本文所用的术语“肿瘤特异性标记物”是指可用于指示肿瘤存在的任何生物材料或元件。生物材料的实例包括但不限于核酸、多肽、碳水化合物、脂肪酸、细胞组分(例如细胞膜和线粒体)和全细胞。在一些情况下,标记物是特定核酸区域(例如,基因、基因内区域、特定基因座等)。作为标记物的核酸区域可以称为例如“标记基因”、“标记区域”、“标记序列”、“标记基因座”等。

[0127] 如本文所用,术语“腺瘤”是指腺来源的良性肿瘤。尽管这些生长是良性的,但随着时间的推移,它们可能会发展为恶性。

[0128] 术语“癌前期”或“肿瘤前期”及其等同物是指正在经历恶性转化的任何细胞增殖性疾病。

[0129] 赘生物、腺瘤、癌症等的“部位”是受试者体内赘生物、腺瘤、癌症等所在的组织、器

官、细胞类型、解剖区域、身体部位等。

[0130] 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疾患,确定受试者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,确定患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。例如,诊断可以用于检测受试者感染肿瘤的存在或可能性,或者这样的受试者对化合物(例如药物,例如药物)或其他治疗产生有利反应的可能性。

[0131] 当涉及核酸使用时,术语“分离的”,如在“分离的寡核苷酸”中,是指从至少一种在其天然来源中通常与其相关的污染物核酸中鉴定和分离的核酸序列。分离的核酸以不同于在自然界中发现的形式或设置存在。相比之下,未分离的核酸,如DNA和RNA,在它们存在于自然界中的状态下被发现。非分离核酸的实例包括:在宿主细胞染色体上发现的与邻近基因邻近的给定DNA序列(例如,基因);在细胞中发现的RNA序列,例如编码特定蛋白质的特定mRNA序列,作为与编码多种蛋白质的许多其他mRNA混合物。然而,编码特定蛋白质的分离的核酸包括例如在通常表达蛋白质的细胞中的这种核酸,其中该核酸在与天然细胞不同的染色体位置中,或者另外由不同于自然界中发现的核酸侧接。分离的核酸或寡核苷酸可以单链或双链形式存在。当分离的核酸或寡核苷酸用于表达蛋白质时,该寡核苷酸将至少包含有义链或编码链(即寡核苷酸可以是单链的),但可以同时含有有义链和反义链(即,寡核苷酸可以是双链的)。分离的核酸在从其天然或典型环境分离后,可以与其他核酸或分子组合。例如,分离的核酸可以存在于其已被置于其中的宿主细胞中,例如用于异源表达。

[0132] 术语“纯化的”是指从其天然环境中移出、分离或分离出的分子,核酸或氨基酸序列。因此,“分离的核酸序列”可以是纯化的核酸序列。“基本上纯化的”分子至少60%不含,优选至少75%不含,更优选至少90%不含与其天然相关的其他组分。如本文所用,术语“纯化的”或“进行纯化”还指从样品中去除污染物。污染蛋白的去除导致样品中感兴趣的多肽或核酸的百分比增加。在另一个实例中,重组多肽在植物、细菌、酵母或哺乳动物宿主细胞中表达,并且通过去除宿主细胞蛋白质来纯化多肽;从而增加了样品中重组多肽的百分比。

[0133] 术语“包含”给定多核苷酸序列或多肽的组合物广泛地指含有给定多核苷酸序列或多肽的任何组合物。该组合物可以包含含有盐(例如,NaCl)、洗涤剂(例如,SDS)和其他组分(例如,Denhardt溶液、奶粉、鲑鱼精子DNA等)的水溶液。

[0134] 术语“样品”以其最广泛的含义使用。在某种意义上,它可以指动物细胞或组织。在另一种意义上,它是指从任何来源获得的样品或培养物,以及生物和环境样品。生物样品可以从植物或动物(包括人类)获得并且涵盖流体、固体、组织和气体。环境样品包括环境材料,例如表面物质、土壤、水和工业样品。这些实例不应被解释为限制适用于本发明的样品类型。

[0135] 如本文所用,在一些情况下使用的“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。例如,当评估粪便样品中源自宫颈的样品材料时,该样品是远程样品。

[0136] 如本文所用,术语“患者”或“受试者”是指要经受这项技术提供的各种测试的生物体。术语“受试者”包括动物,优选哺乳动物,包括人类。在一个优选的实施方案中,所述受试者是灵长类动物。在一个甚至更优选的实施方案中,所述受试者是人类。进一步关于诊断方法,优选的受试者是脊椎动物受试者。优选的脊椎动物是温血的;优选的温血脊椎动物是哺

乳动物。优选的哺乳动物最优选是人类。如本文所用,术语“受试者”包括人类和动物受试者。因此,本文提供了兽医治疗用途。因此,本技术提供了对哺乳动物(例如人)以及由于濒危而重要的(例如西伯利亚虎)或具有经济重要性(例如在农场饲养供人类消费的动物;和/或对人类具有社会重要性的动物,例如作为宠物或在动物园中饲养的动物)的哺乳动物的诊断。此类动物的实例包括但不限于:诸如猫和狗的食肉动物;猪,包括猪、野猪和野猪;反刍动物和/或有蹄类动物,诸如牛、公牛、绵阳、长颈鹿、鹿、山羊、野牛和骆驼;鳍足类动物;和马。因此,还提供了家畜的诊断和治疗,包括但不限于家养猪、反刍动物、有蹄类动物、马(包括赛马)等。目前公开的主题进一步包括用于诊断受试者中宫颈癌、宫颈癌亚型(例如宫颈腺癌、鳞状细胞宫颈癌)和/或宫颈癌前病变(例如宫颈相关原位腺癌、宫颈上皮内瘤变)的系统。该系统可以例如作为商业试剂盒提供,其可以用于筛查宫颈癌、宫颈癌亚型(例如宫颈腺癌、鳞状细胞宫颈癌)和/或宫颈癌前病变(例如,宫颈相关原位腺癌、宫颈上皮内瘤变)的风险,或诊断已从其收集生物样品的受试者的宫颈癌、子宫颈癌亚型和/或宫颈癌前病变。根据本发明的技术提供的示例性系统包括评估本文所述的标记物的甲基化状态。

[0137] 如本文所用,术语“试剂盒”是指递送材料的任何递送系统。在反应测定的上下文中,此类递送系统包括允许将反应试剂(例如,在适当的容器中的寡核苷酸、酶等)和/或支持材料(例如,缓冲液、对于进行测定的书面说明等)从一个位置储存、运输或递送至另一个位置的系统。例如,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)。如本文使用,术语“零散试剂盒”是指包括两个或更多个单独容器的递送系统,每个容器含有全部试剂盒组分的子部分。这些容器可共同或单独递送至预定接受者。例如,第一容器可以包含用于测定的酶,而第二容器包含寡核苷酸。术语“片段化试剂盒”旨在涵盖含有受联邦食品、药品和化妆品法案第520(e)条监管的分析物特异性试剂(ASR)的试剂盒,但不限于此。事实上,包含各自含有全部试剂盒组分的子部分的两个或更多个单独容器的任何输送系统包括在术语“零散试剂盒”中。相比之下,“组合试剂盒”是指在单个容器中(例如,在容纳每个所需组分的单个盒中)包含反应测定的所有组分的递送系统。术语“试剂盒”包括零散试剂盒和组合试剂盒。

[0138] 如本文所用,术语“信息”是指任何事实或数据的集合。关于使用计算机系统(包括但不限于互联网)存储或处理的信息,该术语指以任何格式(例如,模拟、数字、光学等)存储的任何数据。如本文所用,术语“与受试者有关的信息”是指与受试者(例如,人类、植物或动物)有关的事实或数据。术语“基因组信息”是指与基因组有关的信息,包括但不限于核酸序列、基因、甲基化百分比、等位基因频率、RNA表达水平、蛋白质表达、与基因型相关的表型等。“等位基因频率信息”是指与等位基因频率有关的事实或数据,包括但不限于等位基因身份、等位基因的存在与受试者(例如人类受试者)的一种特征之间的统计相关性、个体或群体中等位基因的存在或不存在、在具有一种或多种特定特征的个体中存在等位基因的可能性百分比等。

具体实施方式

[0139] 本文提供了用于宫颈癌筛查的技术,特别是(但不限于)用于检测宫颈癌和/或特定形式的宫颈癌(例如宫颈腺癌、鳞状细胞宫颈癌、宫颈相关原位腺癌、宫颈上皮内瘤变)的存在或用于从生物样品(例如组织样品(例如宫颈组织)、血液样品、血浆样品、血清样品、全

血样品、分泌物样品(例如宫颈分泌物、阴道分泌物)、器官分泌物样品、CSF样品、唾液样品、尿液样品,或粪便样品)中区分宫颈癌与其他类型的妇科癌症(例如子宫内膜癌和卵巢癌)的方法、组合物和相关用途。

[0140] 实际上,如实施例I和II中所述,在鉴定本发明实施方案的过程中进行的实验鉴定了一组新的差异甲基化区域(DMR),其用于将宫颈来源DNA的癌症与非肿瘤对照DNA区分开,以及用于将宫颈癌组织与子宫内膜和卵巢癌组织区分开。此类实验列出并描述了区分宫颈癌、宫颈癌亚型和宫颈癌前病变组织与良性宫颈组织(参见表I-IV、VI-VIII、实施例I)以及宫颈癌组织与子宫内膜癌组织和卵巢癌组织(参见表XI和XII,实施例II)的423种新DNA甲基化标记物。

[0141] 在特定方面,本技术提供了用于鉴定、确定和/或分类癌症例如宫颈癌、宫颈癌亚型(例如宫颈腺癌、鳞状细胞宫颈癌)和/或宫颈癌前病变(例如,宫颈相关原位腺癌、宫颈上皮内瘤变)的组合物和方法。该方法包括测定从受试者分离的生物样品(例如组织样品(例如宫颈组织)、血液样品、血浆样品、血清样品、全血样品、分泌物样品(例如,宫颈分泌物、阴道分泌物)、器官分泌物样品、CSF样品、唾液样品、尿液样品或粪便样品)中至少一种甲基化标记物的甲基化状态,其中标记物的甲基化状态的变化指示宫颈癌、宫颈癌亚型(例如宫颈腺癌、鳞状细胞宫颈癌)和/或宫颈癌前病变(例如宫颈相关原位腺癌、宫颈上皮内瘤变)的存在、类别或部位。特定实施方案涉及包含差异甲基化区域(DMR,例如DMR 1-423,参见表I、III和X)的标记物,其用于诊断(例如,筛查)宫颈癌、宫颈癌亚型、以及/或宫颈癌前病变,或将宫颈癌与其他类型的妇科癌症(例如子宫内膜癌和卵巢癌)区分开。

[0142] 在这项技术的某些实施方案中,提供包括以下步骤的方法:

[0143] 1) 将从受试者获得的生物样品(例如组织样品(例如宫颈组织)、血液样品、血浆样品、血清样品、全血样品、分泌物(例如,宫颈分泌物、阴道分泌物)、器官分泌物样品、CSF样品、唾液样品、尿液样品或粪便样品)中的核酸(例如基因组DNA)与区分一种或多种甲基化标记物内的甲基化和非甲基化二核苷酸(例如,CpG二核苷酸)的至少一种试剂或一系列试剂接触;以及

[0144] 2) 检测宫颈癌、宫颈癌亚型(例如宫颈腺癌、鳞状细胞宫颈癌)、宫颈癌前病变(例如宫颈相关原位腺癌、宫颈上皮内瘤变),或将宫颈癌与其他类型的妇科癌症(例如,子宫内膜癌和卵巢癌)区分开(例如,具有大于或等于80%敏感性和大于或等于80%特异性)。

[0145] 在这项技术的某些实施方案中,提供包括以下步骤的方法:

[0146] 1) 通过用以甲基化特异性方式修饰DNA的试剂处理生物样品中的基因组DNA,测量人类个体的生物样品(例如组织样品(例如宫颈组织)、血液样品、血浆样品、血清样品、全血样品、分泌物样品(例如,宫颈分泌物、阴道分泌物)、器官分泌物样品、CSF样品、唾液样品、尿液样品或粪便样品)中的一种或多种基因或甲基化标记物的甲基化水平;

[0147] 2) 使用一组针对所选择的一种或多种基因或甲基化标记物的引物扩增经处理的基因组DNA;以及

[0148] 3) 确定该一种或多种基因或甲基化标记物的甲基化水平。

[0149] 在这项技术的某些实施方案中,提供包括以下步骤的方法:

[0150] 1) 测量来自生物样品(例如组织样品(例如宫颈组织)、血液样品、血浆样品、血清样品、全血样品、分泌物样品(例如,宫颈分泌物、阴道分泌物)、器官分泌物样品、CSF样品、

唾液样品、尿液样品或粪便样品)的DNA中一种或多种甲基化标记基因的量;

[0151] 2) 测量DNA中的至少一种参考标记物的量;以及

[0152] 3) 计算DNA中测量的该一种或多种甲基化标记基因的量的值,作为DNA中测量的参考标记基因的量的百分比,其中该值表示生物样品中测量的该一种或多种甲基化标记DNA的量。

[0153] 在这项技术的某些实施方案中,提供包括以下步骤的方法:

[0154] 1) 通过用能够以甲基化特异性方式修饰DNA的亚硫酸氢盐试剂处理生物样品中的基因组DNA,测量人类个体的生物样品(例如组织样品(例如宫颈组织)、血液样品、血浆样品、血清样品、全血样品、分泌物样品(例如,宫颈分泌物、阴道分泌物)、器官分泌物样品、CSF样品、唾液样品、尿液样品或粪便样品)中的一种或多种基因的CpG位点的甲基化水平;

[0155] 2) 使用一组针对所选择的一种或多种基因的引物扩增经过修饰的基因组DNA;以及

[0156] 3) 确定选定的一种或多种基因的CpG位点的甲基化水平。

[0157] 在某些实施方案中,该技术提供了用于表征生物样品的方法,其包括:

[0158] (a) 通过用亚硫酸氢盐处理生物样品中的基因组DNA来测量人类个体的生物样品(例如组织样品(例如宫颈组织)、血液样品、血浆样品、血清样品、全血样品、分泌物样品(例如,宫颈分泌物、阴道分泌物)、器官分泌物样品、CSF样品、唾液样品、尿液样品或粪便样品)中一种或多种基因的CpG位点的甲基化水平;使用针对选定的一种或多种基因的一组引物扩增亚硫酸氢盐处理的基因组DNA;以及测定CpG位点的甲基化水平;和

[0159] (b) 将该一种或多种基因的甲基化水平与无宫颈癌的对照样品中的一组相应基因的甲基化水平进行比较;以及

[0160] (c) 当该一种或多种基因中测量的甲基化水平高于相应对照样品中测量的甲基化水平时,确定个体患有宫颈癌、宫颈癌亚型(例如宫颈腺癌、鳞状细胞宫颈癌)和/或宫颈癌前病变(例如宫颈相关原位腺癌、宫颈上皮内瘤变)。

[0161] 在某些实施方案中,该技术提供了用于在生物样品中区分宫颈癌与其他类型的妇科癌症(例如,子宫内膜癌和卵巢癌)的方法,该方法包括:

[0162] (a) 通过用亚硫酸氢盐处理生物样品中的基因组DNA来测量人类个体的生物样品(例如组织样品(例如宫颈组织)、血液样品、血浆样品、血清样品、全血样品、分泌物样品(例如,宫颈分泌物、阴道分泌物)、器官分泌物样品、CSF样品、唾液样品、尿液样品或粪便样品)中一种或多种基因的CpG位点的甲基化水平;使用针对选定的一种或多种基因的一组引物扩增亚硫酸氢盐处理的基因组DNA;以及测定CpG位点的甲基化水平;和

[0163] (b) 将该一种或多种基因的甲基化水平与子宫内膜癌样品和/或卵巢癌样品中的一组相应基因的甲基化水平进行比较;以及

[0164] (c) 当在该一种或多种基因中测量的甲基化水平高于在相应子宫内膜癌和/或卵巢癌样品中测量的甲基化水平时确定个体患有宫颈癌。

[0165] 在某些实施方案中,该技术提供了筛查从受试者获得的生物样品(例如组织样品(例如宫颈组织)、血液样品、血浆样品、血清样品、全血样品、分泌物样品(例如,宫颈分泌物、阴道分泌物)、器官分泌物样品、CSF样品、唾液样品、尿液样品或粪便样品)中宫颈癌、宫颈癌亚型(例如宫颈腺癌、鳞状细胞宫颈癌)和/或宫颈癌前病变(例如宫颈相关原位腺癌、

宫颈上皮内瘤变)的方法,该方法包括

[0166] 1) 测定一种或多种DNA甲基化标记物的甲基化状态;和

[0167] 2) 当标记物的甲基化状态与在未患有宫颈癌的受试者中测定的标记物的甲基化状态不同时,鉴定受试者患有宫颈癌、宫颈癌亚型(例如宫颈腺癌、鳞状细胞宫颈癌)和/或宫颈癌前病变(例如宫颈相关原位腺癌、宫颈上皮内瘤变)。

[0168] 在某些实施方案中,该技术提供了用于表征生物样品(例如组织样品(例如宫颈组织)、血液样品、血浆样品、血清样品、全血样品、分泌物样品(例如,宫颈分泌物、阴道分泌物)、器官分泌物样品、CSF样品、唾液样品、尿液样品或粪便样品)的方法。分泌物、阴道分泌物、器官分泌物样品、脑脊液样品、唾液样品、尿液样品或粪便样品),该方法包括测量从生物样品提取的DNA中一种或多种甲基化标记基因的量;用亚硫酸氢盐处理生物样品中的基因组DNA;使用对每个标记基因的CpG位点具有特异性的引物扩增经亚硫酸氢盐处理的基因组DNA,其中对每个标记基因具有特异性的引物能够结合标记基因的引物序列(例如,表V和/或XII中列举的引物)结合的扩增子,其中标记基因的引物序列结合的扩增子是表I、III和/或X中所述的甲基化标记基因的基因区域的至少一部分;确定一种或多种基因的CpG位点的甲基化水平。

[0169] 在某些实施方案中,该技术提供了方法,该方法包括通过从疑似患有或患有癌症的人类个体的生物样品提取基因组DNA,测量从生物样品(例如组织样品(例如宫颈组织)、血液样品、血浆样品、血清样品、全血样品、分泌物样品(例如,宫颈分泌物、阴道分泌物)、器官分泌物样品、CSF样品、唾液样品、尿液样品或粪便样品)提取的DNA中一种或多种甲基化标记基因的甲基化水平;用亚硫酸氢盐处理提取的基因组DNA,用对该一种或多种基因具有特异性的引物扩增亚硫酸氢盐处理的基因组DNA,其中对该一种或多种基因具有特异性的引物能够结合表I、III和X中列举的标记物的染色体区域的经亚硫酸氢盐处理的基因组DNA的至少一部分;并测量该一种或多种甲基化标记基因的甲基化水平。

[0170] 在某些实施方案中,该技术提供了从人类个体的生物样品(例如组织样品(例如宫颈组织)、血液样品、血浆样品、血清样品、全血样品、分泌物样品(例如,宫颈分泌物、阴道分泌物)、器官分泌物样品、CSF样品、唾液样品、尿液样品或粪便样品)制备DNA级分的方法,可用于分析涉及一种或多种染色体畸变的一个或多个遗传基因座,该方法包括:

[0171] (a) 从人类个体的生物样品中提取基因组DNA;

[0172] (b) 通过以下方式产生经提取的基因组DNA的级分:

[0173] (i) 用以甲基化特异性方式修饰DNA的试剂处理经提取的基因组DNA;

[0174] (ii) 使用对一种或多种甲基化标记物具有特异性的单独引物扩增经亚硫酸氢盐处理的基因组DNA;

[0175] (c) 通过测量该一种或多种甲基化标记物中每一种的CpG位点的甲基化水平来分析经提取的基因组DNA的产生部分中的一个或多个遗传基因座。

[0176] 在某些实施方案中,该技术提供了从人类个体的生物样品(例如组织样品(例如宫颈组织)、血液样品、血浆样品、血清样品、全血样品、分泌物样品(例如,宫颈分泌物、阴道分泌物)、器官分泌物样品、CSF样品、唾液样品、尿液样品或粪便样品)制备DNA级分的方法,可用于分析涉及一种或多种染色体畸变的一个或多个DNA片段,该方法包括:

[0177] (a) 从人类个体的生物样品中提取基因组DNA;

[0178] (b) 通过以下方式产生经提取的基因组DNA的级分:

[0179] (i) 用以甲基化特异性方式修饰DNA的试剂处理经提取的基因组DNA;

[0180] (ii) 使用对一种或多种甲基化标记物具有特异性的单独引物扩增经亚硫酸氢盐处理的基因组DNA;以及

[0181] (c) 通过测量该一种或多种甲基化标记物中每一种的CpG位点的甲基化水平来分析经提取的基因组DNA的产生部分中的一个或多个DNA片段。

[0182] 优选地,这类方法的灵敏度为约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%。

[0183] 此类方法不限于特定甲基化标记物、甲基化标记基因、基因、DMR和/或DNA甲基化标记物。在一些实施方案中,该一种或多种甲基化标记物、甲基化标记基因、基因、DMR和/或DNA甲基化标记物包含选自表I、III和X中提供的DMR 1-423组成的组的DMR中的碱基。

[0184] 在一些实施方案中,该一种或多种甲基化标记物、甲基化标记基因、基因、DMR和/或DNA甲基化标记物选自表I和/或III。

[0185] 在一些实施方案中,该一种或多种甲基化标记物、甲基化标记基因、基因、DMR和/或DNA甲基化标记物选自MAX.chr6.58147682-58147771、C10RF114、ASCL1、ARHGAP12、ZNF773、TTYH1、NEUROG3、ZNF781、NXPH1、MAX.chr9.36739811-36739868、NID2、TMEM200C、CRHR2、ABCB1、ZNF69、ATP10A、MAX.chr18.73167725-73167817、MAX.chr2.127783183-127783403、CACNA1C、ZNF382、BARHL1、MAX.chr4.8859853-8859939、ST8SIA1、MAX.chr1.98510958-98511049、C2ORF40、SLC9A3、PRDM12、HOPX_C和KCNQ5(实施例I)。

[0186] 在一些实施方案中,该一种或多种甲基化标记物、甲基化标记基因、基因、DMR和/或DNA甲基化标记物选自C1orf114、MAX.chr6.58147682-58147771、ZNF773、NEUROG3、ASCL1、NID2、ZNF781、CRHR2、和MAX.chr9.36739811-36739868(参见表VI和实施例I)。

[0187] 在一些实施方案中,该一种或多种甲基化标记物、甲基化标记基因、基因、DMR和/或DNA甲基化标记物选自ABCB1、ARHGAP12、ASCL1、BARHL1、C1orf114、C2orf40、CACNA1C、CRHR2、HOPX_C、KCNQ5、MAX.chr1.98510968-98511049、MAX.chr18.73167751-73167791、MAX.chr2.127783183-127783403、MAX.chr4.8859853-8859939、MAX.chr6.58147682-58147771、MAX.chr9.36739811-36739868、NEUROG3、NID2、NXPH1、PRDM12、SLC9A3、TMEM200C、TTYH1、ZNF382、ZNF773和ZNF781(参见表VII实施例I)。

[0188] 在一些实施方案中,该一种或多种甲基化标记物、甲基化标记基因、基因、DMR和/或DNA甲基化标记物选自ABCB1、c1orf95、CACNA1C、CACNG8、CHST2、ELM01、EMID2、FBN1_B、FLT3_A、FLT3_B、GLIS1、GPC6、GREM2、JAM2、KCNK12_A、LOC100129620、MAX.chr15.78112404-78112692、MAX.chr19.4584907-4585088、MAX.chr3.69591689-69591784、NCAM1、NT5C1A、ST8SIA3、ZNF382、ZNF419、ZNF69和ZSCAN18(参见表VIII,实施例I)。

[0189] 在一些实施方案中,该一种或多种甲基化标记物、甲基化标记基因、基因、DMR和/或DNA甲基化标记物选自表X。

[0190] 在一些实施方案中,该一种或多种甲基化标记物、甲基化标记基因、基因、DMR和/或DNA甲基化标记物选自AK5、RABC3、ZNF491、ZNF610、ZNF91、ZNF480、TRPC3_B和ELMOD1(参见表X和XI,实施例II)。

[0191] 此类方法不限于特定样品或生物样品类型。例如,在一些实施方案中,生物样品是

组织样品(例如,宫颈组织)、血液样品、血浆样品、血清样品、全血样品、分泌物样品(例如,宫颈分泌物、阴道分泌物)、器官分泌物样品、CSF样品、唾液样品、尿液样品或粪便样品。

[0192] 利用例如如通过与预测特异性和灵敏度有关的统计技术鉴定的标记物的多种组合预测多种癌症(例如宫颈癌和宫颈癌亚型)和癌前病变(例如宫颈癌前病变)。这项技术进一步提供了鉴定一些癌症的预测组合和验证预测组合的方法。

[0193] 此类方法不限于受试者类型。在一些实施方案中,受试者是哺乳动物。在一些实施方案中,受试者是人。

[0194] 此类方法不限于用于确定表征、测量或分析一种或多种甲基化标记物、甲基化标记基因、基因、DMR和/或DNA甲基化标记物的甲基化的特定方式或技术。在一些实施方案中,这类技术基于对包含DMR的至少一种标记物、标记物区域或标记物的碱基的甲基化状态(例如,CpG甲基化状态)的分析。

[0195] 在一些实施方案中,测量样品中甲基化标记物的甲基化状态包括确定一种碱基的甲基化状态。在一些实施方案中,测量样品中标记物的甲基化状态包括确定多个碱基的甲基化程度。此外,在一些实施方案中,甲基化标记物的甲基化状态包含相对于标记物的正常甲基化状态增加的标记物甲基化。在一些实施方案中,标记物的甲基化状态包含相对于标记物的正常甲基化状态减少的标记物甲基化。在一些实施方案中,标记物的甲基化状态包含相对于标记物的正常甲基化状态不同的标记物甲基化模式。

[0196] 此外,在一些实施方案中,标记物是100个或更少碱基的区域,标记物是500个或更少碱基的区域,标记物是1000个或更少碱基的区域,标记物是5000个或更少碱基的区域,或在一些实施方案中,标记物是一个碱基。在一些实施方案中,标记物在高CpG密度启动子中。

[0197] 在某些实施方案中,用于分析核酸中5-甲基胞嘧啶的存在的方法涉及用以甲基化特异性方式修饰DNA的试剂处理DNA。此类试剂的实例包括但不限于甲基化敏感性限制酶、甲基化依赖性限制酶、亚硫酸氢盐试剂、TET酶和硼烷还原剂。

[0198] 分析核酸中是否存在5-甲基胞嘧啶的常用方法是基于Frommer等人描述的亚硫酸氢盐方法,其用于检测DNA中的5-甲基胞嘧啶(Frommer等人(1992) Proc.Natl.Acad.Sci.USA 89:1827-31,出于所有目的通过引用将其全部内容明确并入本文)或其变体。绘制5-甲基胞嘧啶的亚硫酸氢盐法基于以下观察结果:胞嘧啶(而非5-甲基胞嘧啶)与亚硫酸氢离子(也称为亚硫酸氢盐)发生反应。该反应通常按以下步骤进行:首先,胞嘧啶与亚硫酸氢反应形成磺化胞嘧啶。接下来,磺化反应中间体的自发脱氨基作用产生磺化尿嘧啶。最后,磺化尿嘧啶在碱性条件下脱磺化,形成尿嘧啶。检测是可能的,因为尿嘧啶碱基与腺嘌呤配对(因此表现得像胸腺嘧啶),而5-甲基胞嘧啶碱基与鸟嘌呤配对(因此表现得像胞嘧啶)。这使得可以将甲基化胞嘧啶与非甲基化胞嘧啶区分开,通过例如亚硫酸氢盐基因组测序(Grigg G,&Clark S,Bioessays(1994)16:431-36;Grigg G,DNA Seq.(1996)6:189-98),如美国专利号5,786,146中公开的甲基化特异性PCR(MSP),或使用包含序列特异性探针切割的测定法,例如QuARTS flap核酸内切酶测定法(参见例如Zou等人(2010)“Sensitive quantification of methylated markers with a novel methylation specific technology”Clin Chem 56:A199;以及美国专利号8,361,720;8,715,937;8,916,344;和9,212,392。

[0199] 一些常规技术涉及如下方法,所述方法包括将待分析的DNA封装在琼脂糖基质中,

从而防止DNA的扩散和复性(亚硫酸氢盐仅与单链DNA反应),并用快速透析代替沉淀和纯化步骤(Olek A等人,(1996)“A modified and improved method for bisulfite based cytosine methylation analysis”*Nucleic Acids Res.*24:5064-6)。因此可以分析单个细胞的甲基化状态,说明该方法的实用性和敏感度。Rein,T.等人,(1998)*Nucleic Acids Res.*26:2255概述了检测5-甲基胞嘧啶的常规方法。

[0200] 亚硫酸氢盐技术通常涉及在亚硫酸氢盐处理后扩增已知核酸的短特定片段,然后通过测序(Olek&Walter (1997)*Nat. Genet.*17:275-6)或引物延伸反应(Gonzalga&Jones (1997)*Nucleic Acids Res.*25:2529-31;WO 95/00669;美国专利号6,251,594)测定产物来分析单个胞嘧啶位置。一些方法使用酶消化(Xiong和Laird (1997)*Nucleic Acids Res.*25:2532-4)。本领域还描述了通过杂交进行的检测(Olek等人,WO 99/28498)。此外,已经描述了使用亚硫酸氢盐技术检测单个基因的甲基化(Grigg和Clark (1994)*Bioessays* 16:431-6;Zeschnigk等人(1997)*Hum Mol Genet.*6:387-95;Feil等人(1994)*Nucleic Acids Res.*22:695;Martin等人(1995)*Gene* 157:261-4;WO 9746705;WO 9515373)。

[0201] 各种甲基化测定程序可以与根据本技术的亚硫酸氢盐处理结合使用。这些测定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。除其他技术外,此类测定涉及亚硫酸氢盐处理的核酸的测序、PCR(用于序列特异性扩增)、Southern印迹分析以及甲基化特异性限制酶(例如甲基化敏感的或甲基化依赖性酶)的使用。

[0202] 例如,通过使用亚硫酸氢盐处理,基因组测序已被简化用于甲基化模式和5-甲基胞嘧啶分布的分析(Frommer等人(1992)*Proc. Natl. Acad. Sci. USA* 89:1827-1831)。另外,对从亚硫酸氢盐转化的DNA扩增的PCR产物进行限制酶消化可用于评估甲基化状态,例如,如Sadri&Hornsby (1997)*Nucl. Acids Res.*24:5058-5059所述或如称为COBRA(组合亚硫酸氢盐限制分析)的方法所体现(Xiong&Laird (1997)*Nucleic Acids Res.*25:2532-2534)。

[0203] COBRA™分析是一种定量甲基化测定,可用于确定少量基因组DNA中特定基因座的DNA甲基化水平(Xiong&Laird,*Nucleic Acids Res.*25:2532-2534,1997)。简而言之,限制酶消化用于揭示亚硫酸氢钠处理的DNA的PCR产物中的甲基化依赖性序列差异。首先根据Frommer等人(*Proc. Natl. Acad. Sci. USA* 89:1827-1831,1992)描述的程序,通过标准亚硫酸氢盐处理将甲基化依赖性序列差异引入基因组DNA中。然后使用对感兴趣的CpG岛具有特异性的引物对亚硫酸氢盐转化的DNA进行PCR扩增,然后进行限制性核酸内切酶消化、凝胶电泳并使用特异性标记杂交探针进行检测。原始DNA样品中的甲基化水平在广泛的DNA甲基化水平范围内以线性定量方式由消化和未消化PCR产物的相对量表示。此外,这项技术可以可靠地应用于从显微解剖的石蜡包埋组织样品中获得的DNA。

[0204] 用于COBRA™分析的典型试剂(例如,可以在典型的基于COBRA™的试剂盒中找到)可以包括但不限于: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、亚硫酸氢盐处理的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;用于寡核苷酸探针的激酶标记试剂盒;和标记的核苷酸。此外,亚硫酸氢盐转化试剂可能包括:DNA变性缓冲液;磺化缓冲液;DNA回收试剂或试剂盒(如沉淀、超滤、亲和柱);脱磺酸缓冲液;和DNA回收组分。诸如“MethyLight™”(基于荧光的实时PCR技术)(Eads等人,*Cancer Res.*59:2302-2306,1999)、Ms-SNuPE™(甲基化敏感单核苷酸引物延伸)反应(Gonzalga&Jones,*Nucleic Acids Res.*25:2529-2531,1997)、甲基化特异性PCR

(“MSP”;Herman等人,Proc.Natl.Acad.Sci.USA 93:9821-9826,1996;美国专利号5,786,146)和甲基化CpG岛扩增(“MCA”;Toyota等人,Cancer Res.59:2307-12,1999)测定单独使用或与这些方法中的一种或多种组合使用。

[0205] “HeavyMethylTM”测定技术是一种基于亚硫酸氢盐处理的DNA的甲基化特异性扩增来评估甲基化差异的定量方法。覆盖扩增引物之间或被扩增引物覆盖的CpG位置的甲基化特异性阻断探针(“阻断剂”)能够实现核酸样品的甲基化特异性选择性扩增。

[0206] 术语“HeavyMethylTMMethyLightTM”测定是指HeavyMethylTMMethyLightTM测定,其是MethyLightTM测定的变体,其中MethyLightTM测定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。HeavyMethylTM测定还可与甲基化特异性扩增引物结合使用。

[0207] 用于HeavyMethylTM分析的典型试剂(例如,可能在典型的基于MethyLightTM的试剂盒中找到)可以包括但不限于:用于特定基因座(例如特定基因、标记物、基因区域、标记物区域、亚硫酸氢盐处理的DNA序列、CpG岛、或亚硫酸氢盐处理的DNA序列或CpG岛等)的PCR引物;阻断寡核苷酸;优化的PCR缓冲液和脱氧核苷酸;和Taq聚合酶。MSP(甲基化特异性PCR)允许评估CpG岛内几乎任何CpG位点组的甲基化状态,与甲基化敏感限制酶的使用无关(Herman等人Proc.Natl.Acad.Sci.USA 93:9821-9826,1996;美国专利号5,786,146)。简而言之,DNA用亚硫酸氢钠修饰,将未甲基化而非甲基化的胞嘧啶转化为尿嘧啶,随后用对甲基化相对于未甲基化DNA具有特异性的引物扩增产物。MSP仅需要少量DNA,对给定CpG岛基因座的0.1%甲基化等位基因敏感,并且可以对从石蜡包埋样品中提取的DNA进行。用于MSP分析的典型试剂(例如,可能在典型的基于MSP的试剂盒中找到)可以包括但不限于:用于特定基因座(例如特定基因、标记物、基因区域、标记物区域、亚硫酸氢盐处理的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定探针。

[0208] MethyLightTM测定是一种高通量定量甲基化测定,其利用基于荧光的实时PCR(例如TaqMan[®]),在PCR步骤后无需进一步操作(Eads等人,Cancer Res.59:2302-2306,1999)。简而言之,MethyLightTM方法从基因组DNA的混合样品开始,根据标准程序,在亚硫酸氢钠反应中将其转化为甲基化依赖性序列差异的混合池(亚硫酸氢盐过程将未甲基化的胞嘧啶残基转化为尿嘧啶)。然后在“偏倚”反应中进行基于荧光的PCR,例如使用与已知CpG二核苷酸重叠的PCR引物。序列区分既发生在扩增过程的水平上,也发生在荧光检测过程的水平上。

[0209] MethyLightTM测定被用作核酸例如基因组DNA样品中甲基化模式的定量测试,其中序列区分发生在探针杂交的水平上。在定量版本中,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。通过反应提供对输入DNA量的无偏控制,其中引物和探针均不覆盖任何CpG二核苷酸。替代地,通过使用不覆盖已知甲基化位点的对照寡核苷酸(例如,基于荧光的HeavyMethylTM和MSP技术版本)或覆盖潜在甲基化位点的寡核苷酸探测有偏倚的PCR池,来实现基因组甲基化的定性测试。

[0210] MethyLightTM方法可与任何合适的探针(例如“TaqMan[®]”探针、Lightcycler[®]探针等)一起使用。例如,在一些应用中,双链基因组DNA用亚硫酸氢钠处理,并使用TaqMan[®]探针进行两组PCR反应中的一组,例如使用MSP引物和/或HeavyMethyl阻断寡核苷酸和TaqMan[®]探针。TaqMan[®]探针采用荧光“报告分子”和“猝灭剂”分子双重标记,

专为GC含量相对较高的区域而设计,因此在PCR循环中的熔解温度比正向或反向引物高约10℃。这使得**TaqMan®**探针在PCR退火/延伸步骤中保持完全杂交。当Taq聚合酶在PCR过程中酶促合成一条新链时,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来取代**TaqMan®**探针,释放荧光报告分子,以便使用实时荧光检测系统定量检测其目前未猝灭的信号。

[0211] 用于MethyLight™分析的典型试剂(例如,可能在典型的基于MethyLight™的试剂盒中找到)可以包括但不限于:用于特定基因座(例如特定基因、标记物、基因区域、标记物区域、亚硫酸氢盐处理的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。

[0212] QM™(定量甲基化)测定是基因组DNA样品中甲基化模式的替代定量测试,其中序列区分发生在探针杂交水平上。在此定量版本中,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未偏倚扩增。通过反应提供对输入DNA量的无偏控制,其中引物和探针均不覆盖任何CpG二核苷酸。替代地,通过使用不覆盖已知甲基化位点的对照寡核苷酸(基于荧光的HeavyMethyl™和MSP技术版本)或覆盖潜在甲基化位点的寡核苷酸探测有偏倚的PCR池,来实现基因组甲基化的定性测试。

[0213] QM™方法可以在扩增过程中与任何合适的探针一起使用,例如“**TaqMan®**”探针、**Lightcycler®**探针。例如,双链基因组DNA用亚硫酸氢钠处理,并用无偏引物和**TaqMan®**探针进行处理。**TaqMan®**探针采用荧光“报告分子”和“猝灭剂”分子双重标记,专为GC含量相对较高的区域而设计,因此在PCR循环中的熔解温度比正向或反向引物高约10℃。这使得**TaqMan®**探针在PCR退火/延伸步骤中保持完全杂交。当Taq聚合酶在PCR过程中酶促合成一条新链时,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来取代**TaqMan®**探针,释放荧光报告分子,以便使用实时荧光检测系统定量检测其目前未猝灭的信号。用于QM™分析的典型试剂(例如,可能在典型的基于QM™的试剂盒中找到)可以包括但不限于:用于特定基因座(例如特定基因、标记物、基因区域、标记物区域、亚硫酸氢盐处理的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。

[0214] Ms-SNuPE™技术是一种定量方法,其用于基于DNA的亚硫酸氢盐处理以及随后的单核苷酸引物延伸来评估特定CpG位点处的甲基化差异(Gonzalzo&Jones,Nucleic Acids Res.25:2529-2531,1997)。简而言之,基因组DNA与亚硫酸氢钠反应,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。然后使用对亚硫酸氢盐转化的DNA具有特异性的PCR引物扩增所需的靶序列,分离所得产物并用作感兴趣的CpG位点甲基化分析的模板。可以分析少量的DNA(例如,显微解剖的病理切片)并且避免使用限制酶来确定CpG位点的甲基化状态。

[0215] 用于Ms-SNuPE™分析的典型试剂(例如,可能在典型的基于Ms-SNuPE™的试剂盒中找到)可以包括但不限于:用于特定基因座(例如特定基因、标记物、基因区域、标记物区域、亚硫酸氢盐处理的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);和标记的核苷酸。此外,亚硫酸氢盐转化试剂可能包括:DNA变性缓冲液;磺化缓冲液;DNA回收试剂或试剂盒(如沉淀、超滤、亲和柱);脱磺酸缓冲液;和DNA回收组分。

[0216] 简化代表性亚硫酸氢盐测序(RRBS)从对核酸进行亚硫酸氢盐处理开始,将所有未甲基化的胞嘧啶转化为尿嘧啶,然后进行限制酶消化(例如,通过识别包括CG序列的位点的酶,例如MspI)以及与衔接子配体偶联后片段的完整测序。限制酶的选择丰富了CpG密集区域的片段,减少了分析过程中可能映射到多个基因位置的冗余序列的数量。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进行大小选择)来降低核酸样品的复杂性。与全基因组亚硫酸氢盐测序相反,限制酶消化产生的每个片段都包含至少一个CpG二核苷酸的DNA甲基化信息。因此,RRBS通过这些区域中的高频率限制酶切位点富集了样品的启动子、CpG岛和其他基因组特征,从而提供了一种评估一个或多个基因组基因座的甲基化状态的测定。

[0217] RRBS的典型方案包括用限制酶(例如MspI)消化核酸样品、填充突出端和A尾部、连接衔接子、亚硫酸氢盐转化和PCR的步骤。参见,例如,等人(2005)“Genome-scale DNA methylation mapping of clinical samples at single-nucleotide resolution”*Nat Methods* 7:133-6;Meissner等人(2005)“Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis”*Nucleic Acids Res.*33:5868-77。

[0218] 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)测定来评估甲基化状态。每个QuARTS测定中依次发生三个反应,包括初级反应中的扩增(反应1)和靶探针切割(反应2);以及二次反应中的FRET切割和荧光信号产生(反应3)。当用特定引物扩增靶核酸时,具有flap序列的特定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引起5'核酸酶,例如FEN-1核酸内切酶,通过在检测探针和flap序列之间切割而释放flap序列。flap序列与相应FRET盒的非发夹部分互补。因此,flap序列充当FRET盒上的侵入性寡核苷酸,并实现FRET盒荧光团和猝灭剂之间的切割,从而产生荧光信号。切割反应可以切割每个靶标多个探针,从而每个flap释放多个荧光团,从而提供指数信号放大。QuARTS可以通过使用具有不同染料的FRET盒在单个反应孔中检测多个靶标。参见,例如,Zou等人(2010)“Sensitive quantification of methylated markers with a novel methylation specific technology”*Clin Chem* 56:A199),和美国专利号8,361,720;8,715,937;8,916,344;和9,212,392,出于所有目的将其每一个通过引用并入本文。

[0219] 术语“亚硫酸氢盐试剂”是指包含亚硫酸氢盐、焦亚硫酸盐、亚硫酸氢盐或其组合的试剂,如本文所公开的用于区分甲基化和未甲基化的CpG二核苷酸序列。所述治疗的方法是本领域已知的(例如,PCT/EP2004/011715和WO 2013/116375,其各自通过引用整体并入)。在一些实施方案中,亚硫酸氢盐处理在变性溶剂例如但不限于正亚烷基二醇或二甘醇二甲醚(DME)的存在下或在二噁烷或二噁烷衍生物的存在下进行。在一些实施方案中,变性溶剂以介于1%与35%(v/v)之间的浓度使用。在一些实施方案中,亚硫酸氢盐反应在清除剂存在下进行,所述清除剂例如但不限于色烷衍生物,例如6-羟基-2,5,7,8,-四甲基色烷2-羧酸或三羟基苯甲酸及其衍生物,例如没食子酸(参见:PCT/EP2004/011715,其通过引用整体并入)。在某些优选的实施方案中,亚硫酸氢盐反应包括用亚硫酸氢铵处理,例如,如WO

2013/116375中所述。

[0220] 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V和XII)和扩增酶来扩增经处理的DNA的片段。几个DNA区段的扩增可以在同一个反应容器中同时进行。通常,使用聚合酶链式反应(PCR)进行扩增。扩增子的长度通常为100至2000个碱基对。

[0221] 在该方法的另一个实施方案中,可以通过使用甲基化特异性引物寡核苷酸来检测在包含DMR(例如DMR 1-423,表I、III和X)的标记物内或附近的CpG位置的甲基化状态。这种技术(MSP)已在Herman的美国专利No.6,265,171中描述。使用甲基化状态特异性引物扩增经亚硫酸氢盐处理的DNA允许区分甲基化和未甲基化核酸。MSP引物对含有至少一个与亚硫酸氢盐处理的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。对未甲基化的DNA具有特异性的MSP引物在CpG的C位置处含有“T”。

[0222] 此类方法不限于与该一种或多种甲基化标记物、甲基化标记基因、基因、DMR和/或DNA甲基化标记物相关的特定类型或种类的引物或引物对。在一些实施方案中,引物或引物对在表V(SEQ ID No:1-76)中列出。在一些实施方案中,对每个甲基化标记基因具有特异性的引物或引物对能够结合由表V和/或XII中列举的标记基因的引物序列结合的扩增子,其中由表V和/或XII中列举的标记基因的引物序列结合的扩增子是表I、III和X中列举的甲基化标记基因的基因区的至少一部分。在一些实施方案中,用于甲基化标记物的引物或引物对是特异性结合包含表I、III和X中列举的特定甲基化标记物的染色体坐标的基因区的至少一部分的引物组。

[0223] 在另一个实施方案中,本发明提供了将细胞游离DNA中氧化的5-甲基胞嘧啶残基转化为二氢尿嘧啶残基的方法(参见Liu等人,2019,Nat Biotechnol.37,第424-429页;美国申请公开号202000370114)。该方法涉及选自5-甲酰胞嘧啶(5fC)、5-羧甲基胞嘧啶(5caC)及其组合的氧化5mC残基与硼烷还原剂的反应。氧化的5mC残基可以是天然存在的,或更通常地,是5mC或5hmC残基先前氧化的结果,例如用TET家族酶(例如TET1、TET2或TET3)氧化5mC或5hmC,或5mC或5hmC的化学氧化,例如用高钌酸钾(KRuO₄)或无机过氧化合物或组合物例如过氧钨酸盐(参见例如Okamoto等人(2011)Chem. Commun. 47:11231-33)和高氯酸铜(II)/2,2,6,6-四甲基哌啶-1-氧基(TEMPO)组合(参见Matsushita等人(2017)Chem. Commun. 53:5756-59)。

[0224] 硼烷还原剂的特征可以是硼烷和选自氮杂环和叔胺的含氮化合物的络合物。氮杂环可以是单环、双环或多环,但通常是单环,5元环或6元环形式,其包含氮杂原子和任选的一个或多个选自N、O和S的另外的杂原子。氮杂环可以是芳香族或脂环族。本文优选的氮杂环包括2-吡咯啉、2H-吡咯、1H-吡咯、吡唑烷、咪唑烷、2-吡唑啉、2-咪唑啉、吡唑、咪唑、1,2,4-三唑、1,2,4-三唑、哒嗪、嘧啶、吡嗪、1,2,4-三嗪和1,3,5-三嗪,它们中的任一种可以是未取代的或被一个或多个非氢取代基取代。典型的非氢取代基是烷基,特别是低级烷基,例如甲基、乙基、正丙基、异丙基、正丁基、异丁基、叔丁基等。示例性化合物包括吡啶硼烷、2-甲基吡啶硼烷(也称为2-甲基吡啶硼烷)和5-乙基-2-吡啶。

[0225] 硼烷还原剂与细胞游离DNA中氧化的5mC残基的反应是有利的,因为可以采用无毒试剂和温和的反应条件;不需要任何硫酸氢盐,也不需要任何其他潜在的DNA降解试剂。此外,用硼烷还原剂将氧化的5mC残基转化为二氢尿嘧啶可以在“一锅”或“单管”反应中进行,无需分离任何中间体。这是非常重要的,因为转化涉及多个步骤,即(1)氧化5mC中连接C-4

和C-5的烯烃键的还原, (2) 脱氨, 以及 (3) 脱羧, 如果氧化5mC是5caC, 或者去甲酰化, 如果氧化的5mC是5fC。

[0226] 除了将细胞游离DNA中氧化的5-甲基胞嘧啶残基转化为二氢尿嘧啶残基的方法之外, 本发明还提供了与上述方法相关的反应混合物。该反应混合物包含含有至少一个选自5caC、5fC及其组合的氧化的5-甲基胞嘧啶残基的细胞游离DNA样品, 以及硼烷还原剂, 该硼烷还原剂有效地还原、脱氨、以及脱羧或去甲酰化至少一个氧化的5-甲基胞嘧啶残基。硼烷还原剂是硼烷和选自氮杂环和叔胺的含氮化合物的络合物, 如上所述。在一个优选的实施方案中, 反应混合物基本上不含亚硫酸氢盐, 意思是基本上不含亚硫酸氢根离子和亚硫酸氢盐。理想地, 反应混合物不含亚硫酸氢盐。

[0227] 在本发明的一个相关方面, 提供了用于将细胞游离DNA中的5mC残基转化为二氢尿嘧啶残基的试剂盒, 其中该试剂盒包括用于阻断5hmC残基的试剂、用于氧化羟甲基化以外的5mC残基以提供氧化的5mC残基的试剂, 和硼烷还原剂, 该硼烷还原剂有效地还原、脱氨、以及脱羧或去甲酰化氧化的5mC残基。该试剂盒还可以包括使用各组分来执行上述方法的说明书。

[0228] 在另一个实施方案中, 提供了利用上述氧化反应的方法。该方法能够检测细胞游离DNA中5-甲基胞嘧啶残基的存在和位置, 并且包括以下步骤:

[0229] (a) 修饰片段化的、衔接子连接的细胞游离DNA中的5hmC残基以在其上提供亲和和标签, 其中所述亲和和标签能够从细胞游离DNA中去除含有修饰的5hmC的DNA;

[0230] (b) 从细胞游离DNA中去除含有修饰的5hmC的DNA, 留下含有未修饰的5mC残基的DNA;

[0231] (c) 氧化未修饰的5mC残基以得到含有选自5caC、5fC及其组合的氧化5mC残基的DNA;

[0232] (d) 使含有氧化的5mC残基的DNA与有效还原、脱氨、以及脱羧或去甲酰化氧化的5mC残基的硼烷还原剂接触, 从而提供含有二氢尿嘧啶残基代替氧化的5mC残基的DNA;

[0233] (e) 对含有二氢尿嘧啶残基的DNA进行扩增和测序;

[0234] (f) 根据(e)中的测序结果确定5-甲基化模式。

[0235] 在另一个实施方案中, 提供了用于鉴定靶核酸中的5-甲基胞嘧啶(5mC)或5-羟甲基胞嘧啶(5hmC)的方法, 其包括以下步骤:

[0236] 提供包含靶核酸的生物样品;

[0237] 修饰靶核酸, 包括以下步骤:

[0238] 通过使核酸样品与TET酶接触, 将核酸样品中的5mC和5hmC转化为5-羧基胞嘧啶(5caC)和/或5-甲酰基胞嘧啶(5fC), 从而产生一个或多个5caC或5fC残基; 和

[0239] 通过用硼烷还原剂处理靶核酸, 将5caC和/或5fC转化为二氢尿嘧啶(DHU), 以提供包含修饰的靶核酸的修饰的核酸样品; 和

[0240] 检测修饰的靶核酸的序列; 其中与靶核酸相比, 修饰的靶核酸的序列中的胞嘧啶(C)至胸腺嘧啶(T)转变或胞嘧啶(C)至DHU转变提供了靶核酸中5mC或5hmC的位置。

[0241] 在一些实施方案中, 硼烷还原剂是2-甲基吡啶硼烷。

[0242] 在一些实施方案中, 检测修饰的靶核酸的序列的步骤包括链终止测序、微阵列、高通量测序和限制酶分析中的一种或多种。

[0243] 在一些实施方案中,TET酶选自由人TET1、TET2和TET3;鼠Tet1、Tet2和Tet3;Naegleria TET(NgTET);和灰拟鬼伞(CcTET)组成的组。

[0244] 在一些实施方案中,该方法还包括阻断一个或多个修饰的胞嘧啶的步骤。在一些实施方案中,阻断步骤包括向5hmC添加糖。

[0245] 在一些实施方案中,该方法还包括扩增一个或多个核酸序列的拷贝数的步骤。

[0246] 在一些实施方案中,氧化剂是高钨酸钾或Cu(II)/TEMPO(2,2,6,6-四甲基哌啶-1-氧基)。

[0247] 细胞游离DNA是从受试者的身体样品中提取的,其中身体样品通常是全血、血浆或血清,最通常是血浆,但样品也可以是组织(例如宫颈组织)、分泌物(例如,宫颈分泌物、阴道分泌物)、器官分泌物、CSF、尿液、唾液、粘膜排泄物、痰液、粪便或眼泪。在一些实施方案中,细胞游离DNA源自肿瘤。在其他实施方案中,细胞游离DNA来自患有疾病或其他致病疾患的患者。细胞游离DNA可能源自或可能不源自肿瘤。在步骤(a)中,应当注意,其中5hmC残基要修饰的细胞游离DNA是纯化的、片段化形式并且是衔接子连接的。在这种情况下,DNA纯化可以使用本领域普通技术人员已知的和/或相关文献中描述的任何合适的方法进行,并且,虽然细胞游离DNA本身可以高度片段化,但有时可能需要进一步片段化,例如,如美国专利公开号2017/0253924中所描述的。细胞游离DNA片段的大小通常在约20个核苷酸至约500个核苷酸的范围内,更典型地在约20个核苷酸至约250个核苷酸的范围内。步骤(a)中修饰的纯化的细胞游离DNA片段已使用常规手段(例如限制酶)进行末端修复,使得片段在每个3'和5'末端具有平末端。在优选的方法中,如WO 2017/176630中所述,还使用聚合酶如Taq聚合酶,提供具有包含单个腺嘌呤残基的3'突出端的平端片段。这有利于随后连接所选择的通用衔接子,即连接至细胞游离DNA片段两端并含有至少一个分子条形码的衔接子,例如Y-衔接子或发夹衔接子。使用衔接子还可以对衔接子连接的DNA片段进行选择PCR富集。

[0248] 那么,在步骤(a)中,“纯化的、片段化的细胞游离DNA”包含衔接子连接的DNA片段。如步骤(a)中指定的,用亲和标签对这些细胞游离DNA片段中的5hmC残基进行修饰,以便能够随后从细胞游离DNA中去除含有修饰的5hmC的DNA。在一个实施方案中,亲和标签包含生物素部分,例如生物素、脱硫生物素、氧化生物素、2-亚氨基生物素、二氨基生物素、生物素亚砷、生物胞素等。使用生物素部分作为亲和标签允许用链霉亲和素例如链霉亲和素珠、磁性链霉亲和素珠等轻松去除。

[0249] 用生物素部分或其他亲和标签标记5hmC残基是通过化学选择性基团与DNA片段中的5hmC残基共价连接来完成的,其中化学选择性基团能够与功能化亲和标签发生反应,从而将亲和标签连接到5hmC残基。在一个实施方案中,化学选择性基团是UDP葡萄糖-6-叠氮化物,其与炔官能化的生物素部分发生自发的1,3-环加成反应,如Robertson等人(2011) Biochem.Biophys.Res.Comm.411(1):40-3,美国专利号8,741,567和WO 2017/176630中所述。因此,添加炔官能化的生物素部分导致生物素部分共价连接至每个5hmC残基。

[0250] 然后,在一个实施方案中,可以使用链霉亲和素珠、磁性链霉亲和素珠等形式的链霉亲和素在步骤(b)中拉下亲和标记的DNA片段,并且如果需要的话,将其搁置用于以后的分析。去除亲和标记的片段后剩余的上清液含有具有未修饰的5mC残基且不含5hmC残基的DNA。

[0251] 在步骤(c)中,使用任何合适的方法将未修饰的5mC残基氧化以提供5caC残基和/

或5fC残基。选择氧化剂以氧化羟甲基化以外的5mC残基,即提供5caC和/或5fC残基。氧化可以使用具有催化活性的TET家族酶以酶促方式进行。本文使用的术语“TET家族酶”或“TET酶”是指具有催化活性的“TET家族蛋白”或“TET催化活性片段”,如美国专利号9,115,386中所定义,其公开内容通过引用并入本文。在本文中优选的TET酶是TET2;参见Ito等人(2011) *Science* 333 (6047):1300-1303。氧化也可以通过化学方式进行,如前一节所述,使用化学氧化剂。合适的氧化剂的实例包括但不限于:无机或有机高钌酸盐形式的高钌酸阴离子,包括金属高钌酸盐如高钌酸钾(KRuO₄)、四烷基高钌酸铵如四丙基高钌酸铵(TPAP)和四丁基高钌酸铵(TBAP))和聚合物支撑的高钌酸盐(PSP);以及无机过氧化合物和组合物,例如过氧钨酸盐或高氯酸铜(II)/TEMPO组合。此时没有必要将含有5fC的片段与含有5caC的片段分开,因为在该方法的下一步骤中,步骤(e)将5fC残基和5caC残基两者转化为二氢尿嘧啶(DHU)。

[0252] 在一些实施方案中,用 β -葡萄糖基转移酶(β 3GT)阻断5-羟甲基胞嘧啶残基,同时用TET酶氧化5-甲基胞嘧啶残基,有效提供5-甲酰基胞嘧啶和5-羧甲基胞嘧啶的混合物。含有这两种氧化物质的混合物可以与2-甲基吡啶硼烷或另一种硼烷还原剂反应,得到二氢尿嘧啶。在该实施方案的变体中,在步骤(b)中不去除含有5hmC的片段。相反,“TET辅助甲基吡啶硼烷测序(TAPS)”将包含5mC的片段和包含5hmC的片段一起酶氧化以提供包含5fC和5caC的片段。与2-甲基吡啶硼烷反应会在最初存在5mC和5hmC残基的地方产生DHU残基。“化学辅助甲基吡啶硼烷测序(CAPS)”涉及用高钌酸钾选择性氧化含有5hmC的片段,使5mC残基保持不变。

[0253] 本实施方案的方法有很多优点:不需要亚硫酸氢盐,使用无毒试剂和反应物;该过程在温和的条件下进行。此外,整个过程可以在单个管中进行,无需分离任何中间体。

[0254] 在相关的实施方案中,上述方法还包括步骤:(g)鉴定在步骤(b)中从细胞游离DNA中去除的含有5hmC的DNA中的羟甲基化模式。这可以使用WO 2017/176630中详细描述的技术来进行。该方法可以在不去除或分离中间体的情况下在单管方法中进行。例如,首先,用 β GT催化的尿苷二磷酸葡萄糖6-叠氮化物对细胞游离DNA片段,优选衔接子连接的DNA片段进行功能化,然后通过化学选择性叠氮化物基团进行生物素化。此过程会在每个5hmC位点上产生共价连接的生物素。在下一步中,生物素化链和含有未修饰(天然)5mC的链同时被拉下以进行进一步处理。如本领域已知的,使用抗5mC抗体或甲基-CpG结合结构域(MBD)蛋白拉下天然的含有5mC的链。然后,在5hmC残基被阻断的情况下,使用任何合适的技术将未修饰的5mC残基选择性氧化以将5mC转化为5fC和/或5caC,如本文别处所述。

[0255] 通过扩增获得的片段可以带有直接或间接的可检测标记。在一些实施方案中,标记是荧光标记、放射性核素或具有可在质谱仪中检测到的典型质量的可分离分子片段。当所述标记是质量标记时,一些实施方案提供标记的扩增子具有单个正或负净电荷,从而允许质谱仪中更好的可检测性。可以通过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。

[0256] 分离适用于这些测定技术的DNA的方法是本领域已知的。特别地,一些实施方案包括核酸分离,如美国专利申请序列号13/470,251(“核酸的分离”)中所述,其通过引用整体并入本文。

[0257] 在一些实施方案中,本文所述的标记物可用于对生物样品(例如,组织样品(例如,

宫颈组织)、血液样品、血浆样品、血清样品、全血样品、分泌物样品(例如,宫颈分泌物、阴道分泌物)、器官分泌物样品、CSF样品、唾液样品、尿液样品或粪便样品)进行的QUARTS测定。在一些实施方案中,用于产生DNA样品的方法,并且特别是用于产生包含小体积(例如,小于100、小于60微升)的高度纯化、低丰度核酸并且基本上和/或有效地不含有抑制用于测试DNA样品的测定(例如PCR、INVADER、QuARTS测定等)的物质的DNA样品的方法。此类DNA样品可用于定性检测从患者采集的样品中存在的基因、基因变体(例如,等位基因)或基因修饰(例如,甲基化)的存在或定量测量其活性、表达或量的诊断测定。例如,一些癌症与特定突变等位基因或特定甲基化状态的存在相关,因此检测和/或量化此类突变等位基因或甲基化状态在癌症的诊断和治疗中具有预测价值。

[0258] 许多有价值的遗传标记物在样品中的量极低,并且产生此类标记物的许多事件都很罕见。因此,即使是PCR等灵敏的检测方法也需要大量DNA来提供足够的低丰度靶标,以满足或取代测定的检测阈值。此外,至少量抑制物质的存在也会损害这些旨在检测如此少量靶标的测定的准确性和精密度。因此,本文提供了对体积和浓度提供必要管理以产生这类DNA样品的方法。

[0259] 在一些实施方案中,样品包括组织样品(例如宫颈组织)、血液、血浆、血清、全血、分泌物(例如宫颈分泌物、阴道分泌物)、器官分泌物、CSF、唾液、尿液或粪便。在一些实施方案中,受试者是人。这类样品可以通过本领域已知的,例如熟练技术人员显而易见的多种方法获得。例如,尿液和粪便样品很容易获得,而血液、腹水、血清或胰液样品可以通过使用例如针头和注射器以肠胃外方式获得。可以通过使样品经受本领域技术人员已知的各种技术,包括但不限于离心和过滤来获得无细胞或基本上无细胞的样品。尽管通常优选不使用侵入性技术来获得样品,但获得样品例如组织匀浆、组织切片和活检标本仍可能是优选的。

[0260] 在一些实施方案中,用能够获得所需样品类型的任何类型或种类的收集装置来获得样品。例如,收集装置可以是能够获得宫颈组织样品的装置。在某些实施方案中,收集装置是能够从宫颈或宫颈附近获得组织或细胞的装置。在一些实施方案中,宫颈组织样品包括例如包含任何宫颈组织或宫颈细胞的样品,并且可以包含来自解剖学上宫颈附近区域的组织或细胞(例如,阴道组织、阴道细胞、子宫内膜组织、子宫内膜细胞、卵巢组织、卵巢细胞等)以及宫颈组织或细胞。在另一个实施方案中,宫颈分泌物样品包括例如包含任何宫颈分泌物或来自解剖学上宫颈附近区域的分泌物(例如阴道分泌物、子宫内膜分泌物和卵巢分泌物等)的样品。在一些实施方案中,收集装置具有吸收构件,该吸收构件能够在与身体区域(例如宫颈、阴道管)接触时收集样品(例如组织、分泌物和/或细胞)。在一些实施方案中,吸收构件是具有适合插入身体孔口(例如宫颈、阴道管)的形状和尺寸的海绵,例如具有圆柱形形状。在一些实施方案中,收集装置是卫生棉条(例如,标准卫生棉条)、将液体释放到阴道中并重新收集液体的灌洗器(例如,Pantarhei筛查器)、宫颈刷(例如,插入阴道并转动收集细胞的刷子)、Fournier宫颈自采样装置(卫生棉条状塑料棒)或拭子。在一些实施方案中,吸收构件由能够收集所需样品的材料制成。在一些实施方案中,吸收构件是海绵材料,诸如人造丝和/或棉。

[0261] 可以通过使样品经受本领域技术人员已知的各种技术,包括但不限于离心和过滤来获得无细胞或基本上无细胞的样品。尽管通常优选不使用侵入性技术来获得样品,但获得样品例如组织匀浆、组织切片和活检标本仍可能是优选的。该技术不限于用于制备样品

和提供用于测试的核酸的方法。例如,在一些实施方案中,使用直接基因捕获从样品(例如,组织样品(例如,宫颈组织)、血液样品、血浆样品、血清样品、全血样品、分泌物样品(例如,宫颈分泌物、阴道分泌物)、器官分泌物样品、CSF样品、唾液样品、尿液样品或粪便样品)中分离DNA,例如,美国专利号8,808,990和9,169,511,以及WO 2012/155072中详述,或通过相关方法。

[0262] 标记物的分析可以单独进行,或与一个测试样品中的其他标记物同时进行。例如,可以将多种标记物组合到一个测试中,以有效处理多个样品并潜在地提供更高的诊断和/或预后准确性。另外,本领域技术人员将认识到测试来自同一受试者的多个样品(例如,在连续的时间点)的价值。对系列样品的此类测试可以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甲基化状态没有变化可以提供有关疾病状态的有用信息,包括但不限于确定事件发生后的大致时间、可挽救组织的存在和量、药物治疗的适当性、各种治疗的有效性以及受试者结果的鉴定,包括未来事件的风险。生物标记物的分析可以多种物理形式进行。例如,可以使用微量滴定板或自动化来促进大量测试样品的处理。可替代地,可以开发单一样品格式以促进以及及时方式进行即刻治疗和诊断,例如在非卧床运输或急诊室环境中。

[0263] 基因组DNA可以通过任何方式分离,包括使用市售试剂盒。简单地说,当感兴趣的DNA被细胞膜包裹时,生物样品必须通过酶促、化学或机械方式进行破坏和溶解。然后可以例如通过用蛋白酶K消化来清除DNA溶液中的蛋白质和其他污染物。然后从溶液中回收基因组DNA。这可以通过多种方法来进行,包括盐析、有机提取或将DNA结合到固相支持物上。方法的选择将受到几种因素的影响,包括时间、费用和所需的DNA数量。包含赘生物或肿瘤发生前物质的所有临床样品类型都适用于本发明的方法,例如组织(例如宫颈组织)、细胞系、组织切片、活检、石蜡包埋组织、分泌物(例如宫颈分泌物、阴道分泌物)、体液、粪便、组织、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它们的组合。

[0264] 该技术不限于用于制备样品和提供用于测试的核酸的方法。例如,在一些实施方案中,使用直接基因捕获从生物样品(例如,组织样品(例如,宫颈组织)、血液样品、血浆样品、血清样品、全血样品、分泌物样品(例如,宫颈分泌物、阴道分泌物)、器官分泌物样品、CSF样品、唾液样品、尿液样品或粪便样品)中分离DNA,例如,美国专利申请序列号61/485386中所详述,或通过相关方法。

[0265] 然后用至少一种试剂或一系列试剂处理基因组DNA样品,所述试剂区分包含DMR(例如DMR 1-423,如表I、III和X所提供)的至少一种标记物内的甲基化与未甲基化的CpG二核苷酸。

[0266] 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、胸腺嘧啶或在杂交行为方面与胞嘧啶不同的另一碱基。然而,在一些实施方案中,试剂可以是甲基化敏感性限制酶。

[0267] 在一些实施方案中,基因组DNA样品通过如下方式进行处理,使得将在5'-位置未甲基化的胞嘧啶碱基转化为尿嘧啶、胸腺嘧啶或在杂交行为方面与胞嘧啶不同的另一碱基。在一些实施方案中,这种处理是用亚硫酸氢盐(亚硫酸氢盐、亚硫酸氢盐),然后碱水解进行的。

[0268] 然后分析处理过的核酸以确定靶基因序列(来自包含DMR,例如至少一种选自DMR

1-423 (例如,如表I、III和X中提供)的DMR的标记物的至少一种基因、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,包括本文所列的那些,例如本文所述的QuARTS和MSP。

[0269] 异常甲基化,更具体地包含DMR (例如,DMR 1-423,例如,如表I、III和X所提供)的标记物的高甲基化与宫颈癌、宫颈癌亚型和宫颈癌前病变相关。

[0270] 在一些实施方案中,这项技术涉及一种治疗患者 (例如,患有任何宫颈癌和/或宫颈癌亚型的患者)的方法,该方法包括确定1)如本文提供的一个或多个甲基化标记物的甲基化状态之一或两个,并基于确定甲基化状态的结果对患者施用治疗。治疗可以是施用药物化合物、疫苗、进行手术、对患者进行成像、进行另一测试。优选地,所述使用是在临床筛查方法、预后评估方法、监测疗法结果的方法、鉴定最可能对特定治疗性治疗有反应的患者、对患者或受试者进行成像的方法以及药物筛查和开发的方法中。

[0271] 在本技术的一些实施方案中,提供了一种用于诊断受试者中特定类型的癌症 (例如,宫颈癌或宫颈癌亚型)和/或癌前病变 (例如,宫颈癌前病变)的方法。如本文所用,术语“诊断 (diagnosing)”和“诊断 (diagnosis)”是指熟练技术人员可以估计甚至确定受试者是否患有给定疾病或疾患或将来是否可能显现给定疾病或疾患的方法。熟练技术人员通常基于一种或多种诊断指标做出诊断,所述诊断指标例如一种或多种生物标记物 (例如,本文公开的一种或多种甲基化标记物、甲基化标记基因、基因、DMR和/或DNA甲基化标记物),其甲基化状态指示疾患存在、严重性或不存在。

[0272] 连同诊断一起,临床癌症预后涉及确定癌症的侵袭性和肿瘤复发的可能性,以规划最有效的疗法。如果可以做出更准确的预后,或者甚至可以评估发生癌症的潜在风险,则可以为患者选择适当的治疗,并且在某些情况下可以选择不太严重的治疗。癌症生物标记物的评估 (例如,确定甲基化状态)可用于将具有良好预后和/或低癌症 (例如,宫颈癌或宫颈癌亚型)发生风险的不需要治疗或有限治疗的受试者与那些更可能发生癌症或癌症复发的可能会受益于更强化治疗的受试者分开。

[0273] 因此,如本文所用,“做出诊断”或“诊断”进一步包括确定发生癌症 (例如,宫颈癌或宫颈癌亚型)的风险或确定预后,其可以基于本文公开的诊断生物标记物 (例如,DMR)的测量,预测临床结果 (使用或不使用医学治疗)、选择适当的治疗 (或治疗是否有效)、或监测当前治疗并可能改变治疗。此外,在本公开主题的一些实施方案中,可以随时间对生物标记物进行多次测定以促进诊断和/或预后。生物标记物的时间变化可用于预测临床结果、监测癌症或癌症亚型 (例如宫颈癌或宫颈癌亚型)的进展和/或监测针对癌症的适当疗法的功效。例如,在这样的实施方案中,预期看到生物样品 (例如,组织样品 (例如,宫颈组织)、血液样品、血浆样品、血清样品、全血样品、分泌物样品 (例如,宫颈分泌物、阴道分泌物)、器官分泌物样品、CSF样品、唾液样品、尿液样品或粪便样品)中本文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另外的生物标记物,如果监测的话)的甲基化状态在有效治疗过程中随时间的变化。

[0274] 在一些实施方案中,本发明公开的主题还提供了一种用于确定是否在受试者中开始或继续癌症 (例如,宫颈癌或宫颈癌亚型)的预防或治疗的方法。一段时间内的任何变化都可用于预测发生癌症 (例如,宫颈癌或宫颈癌亚型)的风险,预测临床结果,确定是否开始或继续癌症的预防或治疗,以及当前治疗是否有效治疗癌症。例如,可以在开始治疗之前选

择第一时间点,并且可以在开始治疗之后的某个时间选择第二时间点。甲基化状态可以在取自不同时间点的每个样品中测量,并记录定性和/或定量差异。来自不同样品的生物标记物水平的甲基化状态的变化可与受试者中的特定癌症(例如,宫颈癌或宫颈癌亚型)风险、预后、确定治疗功效和/或癌症进展相关。

[0275] 在优选的实施方案中,本发明的方法和组合物用于在早期阶段,例如在疾病症状出现之前治疗或诊断疾病(例如,宫颈癌或宫颈癌亚型)。在一些实施方案中,本发明的方法和组合物用于在临床阶段治疗或诊断疾病。

[0276] 如所指出的,在一些实施方案中,可以对一种或多种诊断或预后生物标记物进行多次测定,并且标记物的时间变化可用于确定诊断或预后。例如,可以在初始时间确定诊断标记物,并在第二时间再次确定。在此类实施方案中,标记物从初始时间到第二时间的增加可以诊断特定类型或严重程度的癌症(例如,宫颈癌或宫颈癌的亚型),或给定的预后。同样,标记物从初始时间到第二时间的减少可以指示特定类型或严重程度的癌症,或给定的预后。此外,一种或多种标记物的变化程度可以与癌症的严重程度和未来的不良事件相关。技术人员将理解,虽然在某些实施方案中可以在多个时间点对相同的生物标记物进行比较测量,但也可以在一个时间点测量给定的生物标记物,并在第二时间点测量第二生物标记物,并且比较这些标记物可以提供诊断信息。

[0277] 如本文所用,短语“确定预后”是指技术人员可以通过其预测受试者疾患的进程或结果的方法。术语“预后”并不是指以100%准确度预测疾患进程或结果的能力,甚至不是指根据生物标记物(例如,DMR)的甲基化状态可预测给定进程或结果发生的可能性或多或少。相反,熟练技术人员会理解,术语“预后”是指某个过程或结果发生的可能性增加;也就是说,与未表现出给定疾患的那些个体相比,表现出疾患的受试者更可能发生过程或结果。例如,在未表现出疾患(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(例如,患有特定类型癌症(例如,宫颈癌或宫颈癌亚型))的机会可能非常低。

[0278] 在一些实施方案中,统计分析将预后指标与不利结果的倾向相关联。例如,在一些实施方案中,与从未患有癌症(例如,宫颈癌或宫颈癌亚型)的患者获得的正常对照样品不同的甲基化状态可以表明受试者比水平与对照样品中的甲基化状态更相似的受试者相比更有可能患癌症,这由统计显著性水平确定。此外,甲基化状态相对于基线(例如,“正常”)水平的变化可以反映受试者的预后,并且甲基化状态的变化程度可以与不良事件的严重程度相关。统计显著性通常通过比较两个或多个群体并确定置信区间和/或p值来确定。参见例如Dowdy和Wearden, *Statistics for Research*, John Wiley&Sons, New York 1983,其通过引用整体并入本文。本发明主题的示例性置信区间为90%、95%、97.5%、98%、99%、99.5%、99.9%和99.99%,而示例性p值为0.1、0.05、0.025、0.02、0.01、0.005、0.001和0.0001。

[0279] 在其他实施方案中,可以建立本文公开的预后或诊断生物标记物(例如,DMR)的甲基化状态变化的阈值程度,并且简单地比较生物样品中生物标记物的甲基化状态变化程度与甲基化状态变化的阈值程度。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%和约150%。在又其他实施方案中,可以建立“列线图”,通过该“列线图”,预后或诊断指标(生物标记物或生物标记物的组合)的甲基化状态与针对给定结果的相关处置直接相关。本领域技术人员熟悉

使用这样的列线图来关联两个数值,并理解该测量的不确定度与标记物浓度的不确定度相同,因为参考的是个体样品测量而不是群体平均值。

[0280] 在一些实施方案中,对照样品与生物样品(例如,组织样品(例如,宫颈组织)、血液样品、血浆样品、血清样品、全血样品、分泌物样品(例如,宫颈分泌物、阴道分泌物)、器官分泌物样品、CSF样品、唾液样品、尿液样品或粪便样品)同时进行分析,使得从生物样品获得的结果可以与从对照样品获得的结果进行比较。另外,预期可以提供标准曲线,可以与生物样品的测定结果进行比较。如果使用荧光标记,则这类标准曲线将生物标记物的甲基化状态呈现为测定单位的函数,例如荧光信号强度。使用取自多个供体的样品,可以提供正常组织中该一种或多种生物标记物的对照甲基化状态,以及取自患有特定类型癌症(例如宫颈癌或宫颈癌亚型)或癌前病变(例如宫颈癌前病变)的供体的血浆中该一种或多种生物标记物的“风险”水平的标准曲线。在该方法的某些实施方案中,在从受试者获得的生物样品中鉴定本文提供的一种或多种DMR的异常甲基化状态后,受试者被鉴定为患有癌症(例如宫颈癌或宫颈癌亚型)或癌前病变(例如宫颈癌前病变)。在该方法的其他实施方案中,在从受试者获得的生物样品中检测到一种或多种此类生物标记物的异常甲基化状态导致受试者被鉴定为患有癌症(例如宫颈癌或宫颈癌亚型)或癌前病变(例如宫颈癌前病变)。

[0281] 标记物的分析可以单独进行,或一个测试样品中的其他标记物同时进行。例如,可以将多种标记物组合到一个测试中,以有效处理多个样品并潜在地提供更高的诊断和/或预后准确性。另外,本领域技术人员将认识到测试来自同一受试者的多个样品(例如,在连续的时间点)的价值。对系列样品的此类测试可以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甲基化状态没有变化可以提供有关疾病状态的有用信息,包括但不限于确定事件发生后的大致时间、可挽救组织的存在和量、药物治疗的适当性、各种治疗的有效性以及受试者结果的鉴定,包括未来事件的风险。

[0282] 生物标记物的分析可以多种物理形式进行。例如,可以使用微量滴定板或自动化来促进大量测试样品的处理。可替代地,可以开发单一样品格式以促进以及时方式进行即刻治疗和诊断,例如在非卧床运输或急诊室环境中。

[0283] 在一些实施方案中,如果与对照甲基化状态相比,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,则受试者被诊断为患有宫颈癌、宫颈癌亚型和/或宫颈癌前病变。相反,当在生物样品中未鉴定出甲基化状态的变化时,可以鉴定受试者为未患宫颈癌、宫颈癌亚型和/或宫颈癌前病变、没有患癌症或癌前病变的风险或具有低的患癌症或癌前病变的风险。在这点上,患有癌症或其风险的受试者可以与患有低至基本上没有癌症或其风险的受试者相区分。那些具有发生宫颈癌、宫颈癌亚型和/或宫颈癌前病变风险的受试者可以接受更强化和/或定期的筛查计划。另一方面,那些具有低至基本上没有风险的受试者可以避免接受额外的癌症风险测试(例如,侵入性程序),直到例如未来的筛查,例如根据本发明的技术进行的筛查表明在那些受试者中出现了癌症风险的时间。

[0284] 如上所述,根据本发明的技术方法的实施方案,检测该一种或多种生物标记物的甲基化状态的变化可以是定性测定,或者可以是定量测定。因此,将受试者诊断为患有宫颈癌或宫颈癌亚型或有发生宫颈癌或宫颈癌亚型风险的步骤指示进行某些阈值测量,例如生物样品(例如,组织样品(例如,宫颈组织)、血液样品、血浆样品、血清样品、全血样品、分泌物样品(例如,宫颈分泌物、阴道分泌物)、器官分泌物样品、CSF样品、唾液样品、尿液样品或

粪便样品)中的该一种或多种生物标记物的甲基化状态与预定的对照甲基化状态不同。在该方法的一些实施方案中,对照甲基化状态是生物标记物的任何可检测的甲基化状态。在其中对照样品与生物样品同时测试的方法的其他实施方案中,预定的甲基化状态是对照样品中的甲基化状态。在该方法的其他实施方案中,预定的甲基化状态基于标准曲线和/或通过标准曲线鉴定。在该方法的其他实施方案中,预定的甲基化是特定状态或状态范围。因此,可以在本领域技术人员显而易见的可接受限度内,部分地基于所实践的方法的实施方案和期望的特异性等来选择预定甲基化状态。

[0285] 在某些实施方案中,本技术提供了使包含DMR的核酸与能够以甲基化特异性方式修饰核酸的试剂(例如,甲基化敏感性限制酶、甲基化依赖性限制酶和亚硫酸氢盐试剂)(例如,甲基化敏感的限制酶、甲基化依赖性限制酶、10-11易位(TET)酶(例如人TET1、人TET2、人TET3、鼠TET1、鼠TET2、鼠TET3、Naegleria TET(NgTET)、灰拟鬼伞(CcTET))或其变体)、有机硼烷)反应的步骤,以产生例如以甲基化特异性方式修饰的核酸;对以甲基化特异性方式修饰的核酸进行测序,以提供以甲基化特异性方式修饰的核酸的核苷酸序列;将以甲基化特异性方式修饰的核酸的核苷酸序列与来自未患宫颈癌或宫颈癌亚型的受试者的包含DMR的核酸的核苷酸序列进行比较,以鉴定这两个序列的差异;以及当存在差异时,将受试者鉴定为患有宫颈癌或宫颈癌亚型。

[0286] 该技术还提供了组合物。在某些实施方案中,本技术提供了包含含有DMR的核酸和亚硫酸氢盐试剂的组合物。在某些实施方案中,提供了包含含有DMR的核酸和一个或多个根据SEQ ID NO 1-76的寡核苷酸的组合物。在某些实施方案中,提供了包含含有DMR的核酸和甲基化敏感性限制酶的组合物。在某些实施方案中,提供了包含含有DMR的核酸和聚合酶的组合物。

[0287] 该技术还提供了试剂盒。试剂盒包括本文描述的组合物、装置、设备等的实施方案,以及试剂盒的使用说明书。此类说明书描述了用于从样品制备分析物的适当方法,例如用于收集样品(例如,组织样品(例如,宫颈组织)、血液样品、血浆样品、血清样品、全血样品、分泌物样品(例如,宫颈分泌物、阴道分泌物)、器官分泌物样品、CSF样品、唾液样品、尿液样品或粪便样品)并从样品制备核酸。在一些实施方案中,试剂盒包含一个或多个能够从宫颈或宫颈附近获取样品(例如组织、分泌物和/或细胞)的收集装置(例如,卫生棉条(例如,标准卫生棉条)、将液体释放到阴道中并重新收集液体的灌洗器(例如,Pantarhei筛查器)、宫颈刷(例如,插入阴道并转动收集细胞的刷子)、Fournier宫颈自采样装置(卫生棉条状塑料棒)或拭子)。将试剂盒的各个组分包装在适当的容器和包装中(例如,小瓶、盒子、泡罩包装、安瓿、罐子、瓶子、管等),并且将组分一起包装在适当的容器(例如,一个或多个盒子)中,以便于试剂盒用户的储存、运输和/或使用。应当理解,液体组分(例如,缓冲液)可以以冻干形式提供以由用户重构。试剂盒可包括用于评估、验证和/或确保试剂盒性能的对照或参考。例如,用于测定样品中存在的核酸的量的试剂盒可以包括对照,所述对照包含已知浓度的用于比较的相同或另一种核酸,并且在一些实施方案中,包括对对照核酸具有特异性的检测试剂(例如引物)。试剂盒适合在临床环境中使用,并且在一些实施方案中,适合在用户家中使用。在一些实施方案中,试剂盒的组分提供用于从样品制备核酸溶液的系统的功能。在一些实施方案中,系统的某些组件由用户提供。

[0288] 在某些实施方案中,这项技术涉及组合物(例如,反应混合物)的实施方案。在一些

实施方案中,提供的组合物包含含有DMR的核酸和能够以甲基化特异性方式修饰DNA的试剂(例如,甲基化敏感的限制酶、甲基化依赖性限制酶和亚硫酸氢盐试剂)(例如,甲基化敏感的限制酶、甲基化依赖性限制酶、10-11易位(TET)酶(例如人TET1、人TET2、人TET3、鼠TET1、鼠TET2、鼠TET3、Naegleria TET(NgTET)、灰拟鬼伞(CcTET))或其变体)、硼烷还原剂)。一些实施方案提供包含核酸的组合物,所述核酸包含DMR和如本文所述的寡核苷酸。一些实施方案提供一种组合物,所述组合物包含含有DMR的核酸和甲基化敏感性限制酶。一些实施方案提供包含核酸的组合物,所述核酸包含DMR和聚合酶。

[0289] 在一些实施方案中,本文所述的技术与可编程机器相关联,该可编程机器被设计用于执行由本文所述的方法提供的一系列算术或逻辑运算。例如,这项技术的一些实施方案与计算机软件和/或计算机硬件相关联(例如,在其中执行)。在一方面,这项技术涉及一种计算机,该计算机包括一种形式的存储器、用于执行算术和逻辑运算的元件以及用于执行一系列指令(例如,本文提供的方法)以读取、操作和存储数据的处理元件(例如,微处理器)。在一些实施方案中,微处理器是系统的一部分,其用于:确定甲基化状态(例如,一种或多种DMR,例如表I、III和/或X中提供的DMR 1-423);比较甲基化状态;生成标准曲线;确定Ct值;计算甲基化的分数、频率或百分比;鉴定CpG岛;确定测定或标记物的特异性和/或灵敏度;计算ROC曲线和相关的AUC;序列分析;皆如本文所述或本领域已知。在一些实施方案中,微处理器是系统的一部分,其用于:确定甲基化状态(例如,一种或多种DMR,例如表I、III和/或X中提供的DMR 1-423);比较甲基化状态;生成标准曲线;确定Ct值;计算甲基化的分数、频率或百分比;鉴定CpG岛;确定测定或标记物的特异性和/或灵敏度;计算ROC曲线和相关的AUC;序列分析;皆如本文所述或本领域已知。

[0290] 在一些实施方案中,软件或硬件组件接收多个测定的结果并基于多个测定的结果(例如,确定如表I、III和X中提供的多个DMR的甲基化状态)确定单个值结果以向用户报告,指示癌症风险(例如,宫颈癌或宫颈癌亚型)或癌前病变风险(例如,宫颈癌前病变)。相关实施方案基于来自多个测定的结果(例如,确定如表I、III和X中提供的多个DMR的甲基化状态)的数学组合(例如加权组合、线性组合)计算风险因子。在一些实施方案中,DMR的甲基化状态定义了一个维度并且可以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是例如向用户报告的结果,例如与癌症相关风险。

[0291] 在一些实施方案中,本文提供的技术与协同操作以执行如本文所述的方法的多个可编程装置相关联。例如,在一些实施方案中,多个计算机(例如,通过网络连接)可以并行工作以收集和处理数据,例如,在集群计算或网格计算或依赖于通过传统网络接口(例如以太网、光纤或无线网络技术)连接到网络(专用、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架构的实现中。

[0292] 例如,一些实施方案提供一种包括计算机可读介质的计算机。该实施方案包括耦合到处理器的随机存取存储器(RAM)。处理器执行被存储在存储器中的计算机可执行程序指令。这样的处理器可以包括微处理器、ASIC、状态机或其他处理器,并且可以是多种计算机处理器中的任一个,例如来自加利福尼亚州圣克拉拉的英特尔公司和伊利诺伊州绍姆堡的摩托罗拉公司的处理器。这样的处理器包括或者可以与介质,例如计算机可读介质通信,所述介质存储指令,当由处理器执行时,使处理器执行本文描述的步骤。

[0293] 在一些实施方案中,计算机连接到网络。计算机还可以包括多个外部或内部装置,

例如鼠标、CD-ROM、DVD、键盘、显示器或其他输入或输出装置。计算机的实例是个人计算机、数字助理、个人数字助理、蜂窝电话、移动电话、智能电话、寻呼机、数字平板电脑、膝上型计算机、互联网器具和其他基于处理器的装置。一般而言,与本文提供的技术的各方面相关的计算机可以是在任何操作系统上操作的任何类型的基于处理器的平台,例如Microsoft Windows、Linux、UNIX、Mac OS X等,能够支持一种或多种包含本文提供的技术的程序。一些实施方案包括执行其他应用程序(例如,应用程序)的个人计算机。应用程序可以包含在存储器中并且可以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。

[0294] 本文所述的与这项技术相关联的所有这类组件、计算机和系统可以是逻辑的或虚拟的。

[0295] 在某些实施方案中,该技术提供了用于在由该技术提供的从受试者获得的样品中筛查宫颈癌、宫颈癌亚型和/或宫颈癌前病变的系统。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(例如,组织样品(例如,宫颈组织)、血液样品、血浆样品、血清样品、全血样品、分泌物样品(例如,宫颈分泌物、阴道分泌物)、器官分泌物样品、CSF样品、唾液样品、尿液样品或粪便样品)中筛查宫颈癌或宫颈癌亚型的系统,该系统包括:

[0296] 分析组件,被配置为确定样品中一个或多个甲基化标记物的甲基化状态中的一个或两个,

[0297] 软件组件,被配置为将样品中的一个或多个甲基化标记物的甲基化状态与数据库中记录的对照样品或参考样品进行比较,以及

[0298] 警报组件,被配置为警告用户癌症相关状态。

[0299] 在一些实施方案中,警报由软件组件确定,所述软件组件接收多个测定的结果(例如,确定一个或多个甲基化标记物的甲基化状态)并基于多个结果计算值或结果以进行报告。

[0300] 一些实施例提供与本文提供的每个甲基化标记物相关联的加权参数的数据库,其用于计算值或结果和/或警报以向用户(例如,如医师、护士、临床医生等)报告。在一些实施方案中,报告来自多个测定的所有结果。在一些实施方案中,一个或多个结果用于提供评分、值或结果,所述评分、值或结果基于来自多个测定的一个或多个结果的复合物,其指示受试者中的癌症风险。此类方法不限于特定的甲基化标记物。

[0301] 在此类方法和系统中,一个或多个甲基化标记物包含选自由表I、III和X中提供的DMR 1-423组成的组的DMR中的碱基。

[0302] 在对各种实施方案的详细描述中,为了解释的目的,阐述了许多特定细节以提供对所公开的实施方案的透彻理解。然而,本领域技术人员将理解,这些不同的实施方案可以在有或没有这些特定细节的情况下进行实践。在其他情况下,结构和装置以框图形式示出。此外,本领域技术人员可以容易地理解,其中呈现和执行方法的特定顺序是例示性的,并且预期这些顺序可以变化并且仍然保持在本文公开的各种实施方案的精神和范围内。

[0303] 实施例

[0304] 实施例I.

[0305] 本实施例描述了为评估一组甲基化DNA标记物(MDM)的靶向测定用于检测宫颈癌

的可行性而进行的实验。

[0306] 使用样品制备、测序、分析管道和过滤器的专有方法来鉴定差异甲基化区域 (DMR) 并将其缩小到那些可以精确定位宫颈癌并在临床测试环境中表现出色的DMR。从组织到组织分析,鉴定了320个高甲基化CC DMR (表I)。表II显示表I中所述标记物的宫颈癌组织对比良性宫颈组织的曲线下面积 (AUC)、倍数变化和p值。已鉴定的320个高甲基化CC DMR包括CC 特异性区域和CC亚型特异性区域。

[0307] 表I. 区分宫颈癌组织和良性宫颈组织的已鉴定的甲基化区域 (所示区域的基因组坐标基于人类2009年2月 (GRCh37/hg19) 组装体)

[0308]

DMR	基因注释	染色体 编号	染色体坐标
1	A1BG	19	58858561-58858983
2	ABCB1	7	87257074-87257239
3	ABCG5	2	44059004-44059094
4	ACCN1	17	31619175-31619259
5	ADAMTS3	4	73433788-73434198

[0309]

6	AFF3	2	100721515-100721845
7	AK5	1	77747491-77747536
8	ALX4	11	44331425-44331467
9	ARHGAP12	10	32218791-32218902
10	ARHGAP20	11	110583038-110583345
11	ARL5C_A	17	37321102-37321336
12	ASCL1	12	103351856-103352059
13	ATP10A	15	26108318-26108663
14	AVPR1A	12	63545292-63545422
15	B4GALNT1_A	12	58026043-58026232
16	B4GALNT1_B	12	58026291-58026477
17	BARHL1	9	135462629-135462711
18	BCAT1	12	25101413-25101483
19	BEGAIN	14	101034327-101035263
20	BHLHE23	20	61638294-61638471
21	BMPER	7	33943944-33944081
22	BNIP3_A	10	133795098-133795395
23	BNIP3_B	10	133795980-133796111
24	C1orf114	1	169396696-169396795
25	C1orf200	1	9712256-9712761
26	C1orf95	1	226736897-226737246
27	C2orf40	2	106682149-106682286
28	CACNA1C	12	2162008-2162843
29	CACNA1G	17	48636452-48636851
30	CACNA2D3	3	54156377-54156417
31	CACNG8	19	54485957-54486134
32	CBFA2T3	16	89007193-89007582
33	CCDC39	3	180397115-180397545
34	CCDC63	12	111284678-111284903
35	CCND2_A	12	4378126-4378403
36	CCND2_B	12	4380227-4380491
37	CD1D	1	158150624-158151087
38	CD200	3	112052145-112052405
39	CD70	19	6590492-6590700
40	CDO1	5	115152309-115152413
41	CELF2	10	11207361-11207907
42	CHAT	10	50822420-50822511
43	CHST2	3	142839245-142839555
44	CLSTN2	3	139653864-139654264
45	CNRIP1	2	68546511-68546679
46	COLEC12	18	500692-500954
47	CPEB1_A	15	83316356-83316578
48	CPEB1_B	15	83316610-83316711
49	CPEB1_C	15	83316761-83316889
50	CRHR2	7	30721989-30722129
51	CTNND2	5	11903651-11903727
52	CYTH2	19	48984028-48984223
53	DFNA5	7	24796494-24796719

[0310]

54	DPF3	14	73358910-73359002
55	DPP4	2	162930241-162930531
56	DPY19L2	12	64062412-64063072
57	DSCR6	21	38378602-38378731
58	DTX1	12	113494423-113494759
59	DYSF	2	71693487-71693650
60	ECEL1	2	233352371-233352699
61	EFNA2	19	1295824-1295929
62	ELMO1	7	37488818-37488937
63	EMID2	7	101006343-101006606
64	EOMES	3	27763266-27763325
65	EVC	4	5710271-5710311
66	EVI5L	19	7927499-7927549
67	FAIM2	12	50297738-50297953
68	FAM110B	8	59058124-59058186
69	FAM150A	8	53478223-53478377
70	FAM155A_A	13	108520148-108520265
71	FAM181B	11	82444358-82444592
72	FAM89A	1	231176282-231176575
73	FBN1_A	15	48936889-48936984
74	FBN1_B	15	48937320-48937731
75	FBN1_C	15	48938100-48938405
76	FEV	2	219849133-219849257
77	FGF3_A	11	69632118-69632678
78	FGF3_B	11	69633425-69633515
79	FLI1	11	128563956-128564293
80	FLJ32063	2	200328573-200328645
81	FLT1_A	13	29068764-29068896
82	FLT1_B	13	29069073-29069455
83	FLT3_A	13	28674335-28674415
84	FLT3_B	13	28674451-28674770
85	FOXP2	7	113725336-113725431
86	GALR1_A	18	74962194-74962252
87	GALR1_B	18	74962418-74962484
88	GBGT1	9	136038966-136039283
89	GDF6	8	97172702-97172754
90	GDF7	2	20866000-20866539
91	GLIS1	1	54204427-54205225
92	GLIS2	16	4377932-4378133
93	GPC6	13	93880086-93880188
94	GPM6A	4	176923218-176923302
95	GPR88	1	101004724-101004953
96	GREM2	1	240775358-240775401
97	GRIK3	1	37500263-37500335
98	GSX1	13	28363637-28363972
99	GYPE_A	2	127413591-127413703
100	GYPE_B	2	127413823-127413901
101	HLF	17	53343530-53343591

[0311]

102	HOPX_A	4	57521687-57521903
103	HOPX_B	4	57522083-57522309
104	HOPX_C	4	57522385-57522479
105	HOPX_D	4	57522507-57522621
106	HOXA11	7	27227694-27227736
107	HRH2	5	175085354-175085467
108	IGSF9B_A	11	133826275-133826414
109	IGSF9B_B	11	133826934-133827316
110	IKZF1	7	50343331-50343552
111	IQUB	7	123172894-123173074
112	IRF4_A	6	391781-391900
113	IRF4_B	6	392442-392483
114	IRF4_C	6	393540-393626
115	ITGA4	2	182321830-182322266
116	ITGA5	12	54812290-54812533
117	ITPKA	15	41793563-41793621
118	ITPKB	1	226924925-226925000
119	JAM2	21	27011766-27012116
120	JAM3_A	11	133938788-133938917
121	JAM3_B	11	133939006-133939134
122	JSRP1_A	19	2253228-2253345
123	KATNAL2	18	44526733-44526773
124	KCNA1	12	5019366-5019433
125	KCNA3_A	1	111217011-111217092
126	KCNA3_B	1	111217621-111217859
127	KCNK12_A	2	47796802-47796931
128	KCNK12_B	2	47797362-47797417
129	KCNK17	6	39281389-39281585
130	KCNK9_A	8	140715057-140715148
131	KCNK9_B	8	140716165-140716252
132	KCNQ5	6	73331613-73331787
133	KIAA1383	1	232941228-232941421
134	LBH	2	30453651-30454103
135	LOC100129620	1	99470477-99470580
136	LOC100132891	8	72756111-72756258
137	LOC100192379	4	122686140-122686432
138	LOC157627_A	8	9763927-9764066
139	LOC157627_B	8	9764330-9764450
140	LOC157627_C	8	9764461-9764505
141	LOC220930	10	31608483-31609348
142	LOC642345	13	88323596-88323783
143	LOC644189	19	36909224-36909659
144	LOC648809	15	84748985-84749166
145	LPHN1	19	14260386-14260651
146	LPPR3	19	821418-821789
147	LY6H	8	144241411-144241518
148	MATK	19	3785879-3786193
149	MAX.chr1.161591532-16159	1	161591532-161591616

[0312]

	1616		
150	MAX.chr1.228652332-228652455	1	228652332-228652455
151	MAX.chr1.241587394-241587493	1	241587394-241587493
152	MAX.chr1.35394602-35395059	1	35394602-35395059
153	MAX.chr1.98510958-98511049	1	98510958-98511049
154	MAX.chr10.102497254-102497366	10	102497254-102497366
155	MAX.chr10.131769903-131770042	10	131769903-131770042
156	MAX.chr11.14926602-14926647	11	14926602-14926647
157	MAX.chr11.57250204-57250611	11	57250204-57250611
158	MAX.chr11.58903539-58903592	11	58903539-58903592
159	MAX.chr11.59323785-59323833	11	59323785-59323833
160	MAX.chr12.4273874-4274123	12	4273874-4274123
161	MAX.chr12.52652294-52652357	12	52652294-52652357
162	MAX.chr12.53108215-53108272	12	53108215-53108272
163	MAX.chr13.25116339-25116387	13	25116339-25116387
164	MAX.chr13.29106835-29106997	13	29106835-29106997
165	MAX.chr14.96342482-96342588	14	96342482-96342588
166	MAX.chr15.28351832-28352241	15	28351832-28352241
167	MAX.chr15.78112404-78112692	15	78112404-78112692
168	MAX.chr17.45867384-45867662	17	45867384-45867662
169	MAX.chr17.8230314-8230459	17	8230314-8230459
170	MAX.chr18.73167725-73167817	18	73167725-73167817
171	MAX.chr19.17501437-17501524	19	17501437-17501524
172	MAX.chr19.20959229-20959643	19	20959229-20959643
173	MAX.chr19.30718424-30718720	19	30718424-30718720

[0313]

174	MAX.chr19.4580599-4580736	19	4580599-4580736
175	MAX.chr19.4584907-4585088	19	4584907-4585088
176	MAX.chr2.105488688-105488830	2	105488688-105488830
177	MAX.chr2.127783183-127783403	2	127783183-127783403
178	MAX.chr2.173099703-173099999	2	173099703-173099999
179	MAX.chr2.45162181-45162420	2	45162181-45162420
180	MAX.chr2.97193223-97193287	2	97193223-97193287
181	MAX.chr2.97193452-97193624	2	97193452-97193624
182	MAX.chr20.21491441-21491503	20	21491441-21491503
183	MAX.chr20.34893992-34894061	20	34893992-34894061
184	MAX.chr20.58146884-58146954	20	58146884-58146954
185	MAX.chr20.62733800-62733905	20	62733800-62733905
186	MAX.chr22.42679482-42679979	22	42679482-42679979
187	MAX.chr22.50064415-50064560	22	50064415-50064560
188	MAX.chr22.50118517-50118677	22	50118517-50118677
189	MAX.chr3.14852716-14852812	3	14852716-14852812
190	MAX.chr3.28616834-28616874	3	28616834-28616874
191	MAX.chr3.69591689-69591784	3	69591689-69591784
192	MAX.chr4.41884120-41884180	4	41884120-41884180
193	MAX.chr4.8859853-8859939	4	8859853-8859939
194	MAX.chr5.42952182-42952307	5	42952182-42952307
195	MAX.chr5.77148578-77148655	5	77148578-77148655
196	MAX.chr5.77268554-77268725	5	77268554-77268725
197	MAX.chr5.87437130-87437457	5	87437130-87437457
198	MAX.chr6.130686783-130687268	6	130686783-130687268

[0314]

199	MAX.chr6.58147682-58147771	6	58147682-58147771
200	MAX.chr7.121956750-121956806	7	121956750-121956806
201	MAX.chr7.155259633-155259737	7	155259633-155259737
202	MAX.chr7.1704248-1704556	7	1704248-1704556
203	MAX.chr7.1706132-1706343	7	1706132-1706343
204	MAX.chr7.63767351-63767404	7	63767351-63767404
205	MAX.chr8.30769438-30769726	8	30769438-30769726
206	MAX.chr8.688331-688393	8	688331-688393
207	MAX.chr9.99983863-99983910	9	99983863-99983910
208	MIAT	22	27053248-27053559
209	MUC12	7	100609556-100609627
210	NALCN	13	102068469-102068567
211	NCAM1	11	112833899-112834016
212	NEGR1	1	72747781-72747847
213	NEUROG3	10	71332209-71332333
214	NID2	14	52535717-52536126
215	NPAS1	19	47523654-47523718
216	NT5C1A	1	40137279-40137914
217	NTRK3_A	15	88799876-88800380
218	NTRK3_B	15	88800388-88800700
219	NXPH1	7	8473513-8473626
220	OLIG1	21	34442361-34442433
221	PARVB	22	44420554-44420623
222	PAX2	10	102588392-102588514
223	PAX5	9	37034459-37034563
224	PCDH20	13	61987998-61988042
225	PCDH9	13	67804510-67804731
226	PDE4B	1	66798105-66798275
227	PDGFD	11	104034774-104034920
228	PDGFRA	4	55099048-55099343
229	PIF1	15	65116475-65116558
230	PLCL1	2	198669886-198670352
231	PLEKHO1	1	150122962-150123146
232	POMC	2	25391018-25391218
233	PRDM12	9	133536476-133536577
234	PRKCG	19	54410112-54410197
235	PTENP1	9	33676793-33676938
236	PTGDR_A	14	52734526-52734775
237	PTGDR_B	14	52735213-52735395
238	PTPRM	18	7568044-7568115
239	PTPRU	1	29586341-29586390
240	RFX4	12	106979881-106979929
241	RORB	9	77111792-77112062

[0315]

242	RYR3	15	33603624-33603787
243	SALL3	18	76739229-76739404
244	Septin9_A	17	75370102-75370194
245	Septin9_B	17	75370525-75370663
246	SFMBT2_A	10	7450246-7450327
247	SFMBT2_B	10	7452029-7452478
248	SHANK3_A	22	51112191-51112399
249	SHANK3_B	22	51112422-51112586
250	SLC24A4	14	92790534-92790575
251	SLC26A10	12	58015554-58015696
252	SLC35F1	6	118228394-118228489
253	SLC6A3	5	1445473-1445666
254	SLC9A3	5	528714-528778
255	SNX32	11	65601128-65601301
256	SPOCK2	10	73847889-73848052
257	ST6GALNAC5	1	77334046-77334125
258	ST8SIA1	12	22486883-22487168
259	ST8SIA3	18	55021390-55021467
260	STX1B	16	31021723-31022164
261	SUSD5	3	33260131-33260258
262	SYNE1	6	152958097-152958463
263	SYT15	10	46970720-46970775
264	SYT6	1	114695532-114695720
265	TET1	10	70320226-70320873
266	TJP2	9	71789424-71789541
267	TLX2	2	74741203-74741284
268	TMC2	20	2539517-2539610
269	TMEFF2	2	193059999-193060210
270	TMEM178	2	39892822-39893199
271	TMEM200C	18	5890757-5890849
272	TRIM15	6	30139641-30139766
273	TRIM58	1	248020399-248020450
274	TRIM71	3	32859744-32859793
275	TRIM9	14	51561868-51562422
276	TRPC3_A	4	122872067-122872241
277	TRPC3_B	4	122872703-122873038
278	TSHZ3_A	19	31839734-31840137
279	TSHZ3_B	19	31841190-31841535
280	TSPAN11	12	31079362-31079640
281	TTYH1	19	54926696-54926845
282	ULBP1	6	150285465-150285546
283	UTF1	10	135043803-135043880
284	VILL_A	3	38035507-38035743
285	VILL_B	3	38035975-38036061
286	VSTM2B_A	19	30016244-30016358
287	VSTM2B_B	19	30017444-30017485
288	VSX1	20	25065255-25065331
289	WDR17	4	176987103-176987198

[0316]

290	WNT3	17	44896147-44896284
291	XKR6	8	10872819-10873457
292	ZBTB16	11	113929791-113930273
293	ZFP41	8	144328582-144328648
294	ZIK1	19	58095616-58095660
295	ZMIZ1_A	10	81002818-81002953
296	ZMIZ1_B	10	81003082-81003162
297	ZNF132	19	58951402-58951453
298	ZNF134	19	58125542-58125779
299	ZNF304	19	57862463-57863095
300	ZNF382	19	37095959-37096132
301	ZNF419	19	57999105-57999506
302	ZNF43	19	22018452-22018947
303	ZNF470	19	57078657-57078833
304	ZNF530	19	58111265-58111624
305	ZNF549	19	58038983-58039279
306	ZNF568	19	37407214-37407365
307	ZNF583	19	56915358-56915921
308	ZNF586_A	19	58280987-58281064
309	ZNF586_B	19	58281140-58281369
310	ZNF69	19	11998671-11998972
311	ZNF701	19	53073536-53073713
312	ZNF737	19	20748263-20748472
313	ZNF763	19	12075727-12076043
314	ZNF773	19	58011327-58011598
316	ZNF776	19	58258169-58258561
317	ZNF781	19	38182950-38183230
318	ZNF844	19	12175522-12175663
319	ZNF85	19	21106043-21106387
320	ZSCAN18	19	58609761-58609888

[0317] 表II.表I中所述的标记物的宫颈癌组织与良性宫颈组织(正常)的曲线下面积(AUC)、倍数变化(FC)和p值。

[0318]

DMR	基因注释	AUC 宫颈癌 对比正常	FC 宫颈癌对 比正常	宫颈癌对比正 常的 p 值
1	A1BG	0.8681	163.9	0.0007041
2	ABCB1	0.9062	225.3	4.30E-05
3	ABCG5	0.8715	216.2	0.04206
4	ACCN1	0.905	100.1	0.005458
5	ADAMTS3	0.8513	150.1	0.0339
6	AFF3	0.9007	332.5	0.005802
7	AK5	0.875	121.8	0.03173
8	ALX4	0.8556	128.8	0.01156
9	ARHGAP12	0.9187	29.39	0.001022
10	ARHGAP20	0.8824	90.6	0.01308
11	ARL5C_A	0.9575	110.5	0.001267

[0319]

12	ASCL1	0.9127	145.7	0.00005461
13	ATP10A	0.923	178.3	0.001411
14	AVPR1A	0.8873	24.75	0.001443
15	B4GALNT1_A	0.9073	79.99	9.17E-05
16	B4GALNT1_B	0.9173	66.23	0.02039
17	BARHL1	0.9982	87.43	2.42E-07
18	BCAT1	0.8859	51.82	0.01044
19	BEGAIN	0.9297	147	0.03445
20	BHLHE23	0.8929	149.8	2.77E-06
21	BMPER	0.8705	588	0.04749
22	BNIP3_A	0.8599	53.82	0.04754
23	BNIP3_B	0.891	83.44	0.005236
24	C1orf114	0.9464	191.9	0.002913
25	C1orf200	0.8858	237.3	0.01615
26	C1orf95	0.9308	189.9	0.02143
27	C2orf40	0.9065	58.72	0.002932
28	CACNA1C	0.9755	238.7	0.007272
29	CACNA1G	0.9377	54.43	0.003661
30	CACNA2D3	0.8616	288.6	0.03224
31	CACNG8	0.9273	62.21	0.04407
32	CBFA2T3	0.9014	183.9	0.008917
33	CCDC39	0.9048	106.5	0.04371
34	CCDC63	0.8798	66.55	0.007124
35	CCND2_A	0.9135	52.33	0.0007264
36	CCND2_B	0.8711	45.64	0.006238
37	CD1D	0.8919	92.98	0.0002146
38	CD200	0.8685	113.4	0.02155
39	CD70	0.9127	37.58	0.02947
40	CDO1	0.9608	248.3	0.0002878
41	CELF2	0.8905	49.93	6.38E-05
42	CHAT	0.8503	95.68	0.002803
43	CHST2	0.9022	280.8	0.04599
44	CLSTN2	0.8806	206.3	0.006455
45	CNRIP1	0.9287	180.2	0.004064
46	COLEC12	0.8922	79.93	0.0001851
47	CPEB1_A	0.8708	51.97	0.001492
48	CPEB1_B	0.8595	71.87	0.04946
49	CPEB1_C	0.8937	37.77	0.01035
50	CRHR2	0.9299	127.5	0.009806
51	CTNND2	0.9053	47.93	3.55E-05
52	CYTH2	0.8975	25.22	1.69E-05
53	DFNA5	0.8564	88.33	0.02238
54	DPF3	0.8505	153.8	0.03496
55	DPP4	0.9273	69.34	0.008132
56	DPY19L2	0.8933	94.9	0.01712
57	DSCR6	0.8859	83.2	2.49E-05
58	DTX1	0.8633	293.7	0.02987
59	DYSF	0.8508	211.1	0.00192

[0320]

60	ECEL1	0.8927	115.9	0.006306
61	EFNA2	0.8824	24.26	0.00152
62	ELMO1	0.9031	85.56	0.01097
63	EMID2	0.9073	95.76	0.006303
64	EOMES	0.8972	301.1	0.03415
65	EVC	0.8857	153	0.03326
66	EVI5L	0.8529	26.61	2.49E-06
67	FAIM2	0.954	97.37	9.08E-07
68	FAM110B	0.8253	30.67	0.006014
69	FAM150A	0.858	84.76	0.0004955
70	FAM155A_A	0.8672	135.3	1.64E-05
71	FAM181B	0.8654	255.1	0.01558
72	FAM89A	0.9213	36.37	0.005412
73	FBN1_A	0.8634	95.51	2.54E-05
74	FBN1_B	0.9299	266.2	0.001173
75	FBN1_C	0.891	253.6	0.007426
76	FEV	0.8859	93.65	0.001174
77	FGF3_A	0.934	116.1	0.004175
78	FGF3_B	0.8653	174.8	0.02558
79	FLI1	0.918	80.1	0.004199
80	FLJ32063	0.8851	57.12	7.70E-05
81	FLT1_A	0.8672	73.17	0.0006431
82	FLT1_B	0.8601	257.8	0.002949
83	FLT3_A	0.9118	45.13	0.0002446
84	FLT3_B	0.9152	75.26	0.0003903
85	FOXP2	0.8512	39.81	0.0004998
86	GALR1_A	0.8599	155.1	0.01261
87	GALR1_B	0.8872	99.43	0.006009
88	GBGT1	0.8927	87.98	0.0002011
89	GDF6	0.8814	97.61	9.11E-07
90	GDF7	0.8529	177.4	0.003736
91	GLIS1	0.904	110.8	0.02295
92	GLIS2	0.8503	146.3	0.03235
93	GPC6	0.9007	154.1	1.03E-08
94	GPM6A	0.8627	89.7	9.47E-05
95	GPR88	0.8966	108	9.18E-05
96	GREM2	0.9047	117.7	0.02786
97	GRIK3	0.869	120.2	1.08E-06
98	GSX1	0.8681	34.82	0.008509
99	GYPC_A	0.9549	46.74	4.47E-08
100	GYPC_B	0.9101	213.8	0.0006732
101	HLF	0.8501	34.39	0.01946
102	HOPX_A	0.891	83.98	0.0009785
103	HOPX_B	0.9069	178.4	0.005854
104	HOPX_C	0.9108	138.4	0.006714
105	HOPX_D	0.8655	57.32	1.56E-05
106	HOXA11	0.875	41.78	0.00165
107	HRH2	0.8997	76.84	0.0001794

[0321]

108	IGSF9B_A	0.917	82.34	0.007127
109	IGSF9B_B	0.8979	158	0.001897
110	IKZF1	0.92	78.83	0.03592
111	IQUB	0.893	72.54	0.03801
112	IRF4_A	0.9245	138.3	0.0003516
113	IRF4_B	0.9003	184.6	0.000418
114	IRF4_C	0.8807	92.83	0.02831
115	ITGA4	0.9013	222.2	0.01869
116	ITGA5	0.8704	35.52	2.58E-05
117	ITPKA	0.8585	27.23	0.0001397
118	ITPKB	0.9057	54.73	0.02796
119	JAM2	0.9268	162	0.0143
120	JAM3_A	0.8594	205.6	0.01462
121	JAM3_B	0.8733	129.6	0.004121
122	JSRP1_A	0.8737	69.72	0.0001102
123	KATNAL2	0.875	68.06	0.01101
124	KCNA1	0.9066	169.7	0.000422
125	KCNA3_A	0.872	155.8	0.0002929
126	KCNA3_B	0.8548	95.19	6.61E-05
127	KCNK12_A	0.9048	202.2	0.02022
128	KCNK12_B	0.8729	47.09	0.001777
129	KCNK17	0.9171	108	0.002709
130	KCNK9_A	0.8841	75.1	0.001264
131	KCNK9_B	0.8567	69.72	0.003352
132	KCNQ5	0.9031	101.9	0.003513
133	KIAA1383	0.8607	88.92	0.0007306
134	LBH	0.9343	97.12	0.0002605
135	LOC100129620	0.9062	290.7	0.02926
136	LOC100132891	0.8729	120.4	0.0007793
137	LOC100192379	0.8529	130.7	4.65E-05
138	LOC157627_A	0.8644	109.4	0.0001663
139	LOC157627_B	0.8676	81.49	0.005897
140	LOC157627_C	0.8656	218.6	0.0007851
141	LOC220930	0.924	120.9	0.04857
142	LOC642345	0.8946	111.8	0.001642
143	LOC644189	0.9125	106.2	0.01375
144	LOC648809	0.8966	478.2	0.04067
145	LPHN1	0.9446	109.9	0.03741
146	LPPR3	0.9412	106.1	0.01488
147	LY6H	0.8931	42.29	0.0003825
148	MATK	0.9134	279.5	0.003352
149	MAX.chr1.161591 532-161591616	0.9248	146.6	0.02991
150	MAX.chr1.228652 332-228652455	0.8895	34.63	0.001126
151	MAX.chr1.241587 394-241587493	0.8676	39.12	0.005924
152	MAX.chr1.353946	0.8824	104.6	0.002789

[0322]

	02-35395059			
153	MAX.chr1.985109 58-98511049	0.9472	74.07	2.52E-06
154	MAX.chr10.10249 7254-102497366	0.8752	27.42	0.001174
155	MAX.chr10.13176 9903-131770042	0.8913	109.3	2.56E-06
156	MAX.chr11.14926 602-14926647	0.9747	222.3	0.01192
157	MAX.chr11.57250 204-57250611	0.8962	30.1	6.59E-05
158	MAX.chr11.58903 539-58903592	0.8859	488.4	0.04909
159	MAX.chr11.59323 785-59323833	0.8581	32.36	0.001669
160	MAX.chr12.42738 74-4274123	0.8686	61.99	0.00166
161	MAX.chr12.52652 294-52652357	0.9148	120.2	0.0001635
162	MAX.chr12.53108 215-53108272	0.8879	79.99	0.02578
163	MAX.chr13.25116 339-25116387	0.852	20.66	0.0229
164	MAX.chr13.29106 835-29106997	0.9255	360.4	0.00597
165	MAX.chr14.96342 482-96342588	0.8844	58.34	0.02044
166	MAX.chr15.28351 832-28352241	0.861	215.6	0.03226
167	MAX.chr15.78112 404-78112692	0.9055	24.62	0.0008102
168	MAX.chr17.45867 384-45867662	0.9013	143.9	0.003555
169	MAX.chr17.82303 14-8230459	0.8958	39.54	0.003105
170	MAX.chr18.73167 725-73167817	0.8613	46	0.01328
171	MAX.chr19.17501 437-17501524	0.8824	23.07	0.002121
172	MAX.chr19.20959 229-20959643	0.9066	248	0.02103
173	MAX.chr19.30718 424-30718720	0.8806	50.54	0.01945
174	MAX.chr19.45805 99-4580736	0.9228	123.2	0.0269
175	MAX.chr19.45849 07-4585088	0.9733	213.8	0.004035
176	MAX.chr2.105488 688-105488830	0.8685	26.75	0.0004967

[0323]

177	MAX.chr2.127783 183-127783403	0.8868	133.3	0.007291
178	MAX.chr2.173099 703-173099999	0.8633	48.69	0.006586
179	MAX.chr2.451621 81-45162420	0.866	52.28	0.0005311
180	MAX.chr2.971932 23-97193287	0.8667	43.5	0.003816
181	MAX.chr2.971934 52-97193624	0.8708	65.02	4.31E-05
182	MAX.chr20.21491 441-21491503	0.9103	115.9	0.001712
183	MAX.chr20.34893 992-34894061	0.8698	113.1	0.04005
184	MAX.chr20.58146 884-58146954	0.8841	71.37	0.0002409
185	MAX.chr20.62733 800-62733905	0.9011	89.44	0.02824
186	MAX.chr22.42679 482-42679979	0.9083	43.53	0.0004212
187	MAX.chr22.50064 415-50064560	0.888	51.26	0.001201
188	MAX.chr22.50118 517-50118677	0.8753	49.43	0.03994
189	MAX.chr3.148527 16-14852812	0.9066	58.55	0.01096
190	MAX.chr3.286168 34-28616874	0.8985	78.99	0.04249
191	MAX.chr3.695916 89-69591784	0.9256	85.32	0.03526
192	MAX.chr4.418841 20-41884180	0.8877	59.23	0.004574
193	MAX.chr4.885985 3-8859939	0.9246	58.29	5.48E-05
194	MAX.chr5.429521 82-42952307	0.8824	42.08	3.48E-07
195	MAX.chr5.771485 78-77148655	0.8695	38.05	0.0007677
196	MAX.chr5.772685 54-77268725	0.8879	28.07	3.68E-05
197	MAX.chr5.874371 30-87437457	0.8519	53.85	0.0004652
198	MAX.chr6.130686 783-130687268	0.8824	121.3	0.007539
199	MAX.chr6.581476 82-58147771	0.8845	331.6	9.05E-06
200	MAX.chr7.121956 750-121956806	0.8754	34.26	2.83E-06
201	MAX.chr7.155259	0.8628	48.14	0.003016

[0324]

	633-155259737			
202	MAX.chr7.170424 8-1704556	0.8592	127	0.004151
203	MAX.chr7.170613 2-1706343	0.8824	108.9	0.0008021
204	MAX.chr7.637673 51-63767404	0.8507	97.4	0.005447
205	MAX.chr8.307694 38-30769726	0.8702	146.2	0.01308
206	MAX.chr8.688331- 688393	0.8695	41.94	0.000996
207	MAX.chr9.999838 63-99983910	0.9188	162.4	0.002767
208	MIAT	0.9118	45.21	0.005443
209	MUC12	0.8578	209.5	0.02915
210	NALCN	0.857	51.5	0.004903
211	NCAM1	0.9044	188.9	0.03646
212	NEGR1	0.8529	171.2	0.04721
213	NEUROG3	0.9165	158.2	0.003677
214	NID2	0.8449	98.22	0.0004152
215	NPAS1	0.8591	84.86	0.01148
216	NT5C1A	0.9671	124.2	0.006341
217	NTRK3 A	0.8945	77.83	1.95E-05
218	NTRK3 B	0.9002	66.48	4.92E-06
219	NXPH1	0.8351	81.58	0.000002667
220	OLIG1	0.8538	88.52	0.008082
221	PARVB	0.9444	418	0.02117
222	PAX2	0.8754	36.95	1.24E-05
223	PAX5	0.8585	66.76	0.02582
224	PCDH20	0.9173	161.7	0.001513
225	PCDH9	0.893	328.9	0.02864
226	PDE4B	0.862	27.99	0.0004054
227	PDGFD	0.8706	443.5	0.006209
228	PDGFRA	0.8973	137.5	0.001659
229	PIF1	0.8529	55.36	0.004562
230	PLCL1	0.8547	115.3	0.0007377
231	PLEKHO1	0.9055	39.62	0.0383
232	POMC	0.8746	38.48	0.001239
233	PRDM12	0.9146	68.76	0.002603
234	PRKCG	0.8655	43.59	0.02049
235	PTENP1	0.8763	455	0.01436
236	PTGDR A	0.8613	67.78	3.93E-05
237	PTGDR B	0.9231	45.63	3.22E-06
238	PTPRM	0.869	131.4	0.002421
239	PTPRU	0.8883	120.6	0.003687
240	RFX4	0.8815	318.9	0.008332
241	RORB	0.8971	141	0.004515
242	RYR3	0.8754	85.35	0.0003078

[0325]

243	SALL3	0.8952	42.15	1.83E-07
244	Septin9_A	0.8609	127.4	0.02107
245	Septin9_B	0.9608	140.1	0.008477
246	SFMBT2_A	0.9474	68.95	0.005568
247	SFMBT2_B	0.8717	128.1	0.001783
248	SHANK3_A	0.8975	120.4	0.001461
249	SHANK3_B	0.8824	85.32	0.002408
250	SLC24A4	0.8539	107.5	0.002063
251	SLC26A10	0.8979	73.4	0.0265
252	SLC35F1	0.8529	322.5	0.01256
253	SLC6A3	0.8558	136.1	1.67E-06
254	SLC9A3	0.9479	26.3	0.003185
255	SNX32	0.9483	270.9	0.04352
256	SPOCK2	0.8851	128.9	0.02648
257	ST6GALNAC5	0.9195	169.3	2.43E-05
258	ST8SIA1	0.9585	143.1	1.96E-05
259	ST8SIA3	0.9273	88.4	0.0004173
260	STX1B	0.8616	55.36	3.55E-05
261	SUSD5	0.8768	90.17	0.01131
262	SYNE1	0.8788	94.57	0.01375
263	SYT15	0.9064	42.78	0.0494
264	SYT6	0.8868	136.1	0.01021
265	TET1	0.8997	179.4	0.02966
266	TJP2	0.8663	61.93	0.009941
267	TLX2	0.8979	70.83	0.01152
268	TMC2	0.9772	130.4	0.001039
269	TMEFF2	0.8755	58.54	1.57E-06
270	TMEM178	0.8841	47.63	0.001044
271	TMEM200C	0.9655	216.8	0.005405
272	TRIM15	0.943	185.6	0.03524
273	TRIM58	0.9725	95.12	0.0003784
274	TRIM71	0.8599	347.2	0.003817
275	TRIM9	0.8503	173.3	0.04287
276	TRPC3_A	0.8624	306.4	0.0004037
277	TRPC3_B	0.8642	389.8	0.04104
278	TSHZ3_A	0.9167	64.37	0.0007652
279	TSHZ3_B	0.8663	49.2	0.02169
280	TSPAN11	0.8877	211.8	0.03463
281	TTYH1	0.9421	122.1	0.02287
282	ULBP1	0.8859	66.25	1.53E-05
283	UTF1	0.9194	118.3	0.02302
284	VILL_A	0.9314	222.6	0.0003675
285	VILL_B	0.926	133.9	0.001691
286	VSTM2B_A	0.8704	106	0.000169
287	VSTM2B_B	0.9933	134	0.01423
288	VSX1	0.8786	58.31	3.46E-08
289	WDR17	0.8994	132.6	0.01159
290	WNT3	0.8636	114.5	0.02764

[0326]

291	XKR6	0.8578	78.88	0.0005239
292	ZBTB16	0.9074	48.02	8.33E-06
293	ZFP41	0.9492	29.94	0.001962
294	ZIK1	0.8545	71.77	0.00014
295	ZMIZ1_A	0.8702	89.3	0.02504
296	ZMIZ1_B	0.9221	192.4	0.02884
297	ZNF132	0.8917	233.1	0.03695
298	ZNF134	0.8502	628.8	0.008461
299	ZNF304	0.917	231.5	0.004078
300	ZNF382	0.9152	126.8	0.003379
301	ZNF419	0.9002	301	0.04685
302	ZNF43	0.8849	161.7	0.01134
303	ZNF470	0.8734	578.3	0.02413
304	ZNF530	0.8962	113.4	0.04597
305	ZNF549	0.8636	92.32	0.0001819
306	ZNF568	0.8722	68.63	0.001613
307	ZNF583	0.8676	270.6	0.04555
308	ZNF586_A	0.9073	78.46	0.01952
309	ZNF586_B	0.8619	269.6	0.02721
310	ZNF69	0.9377	56.69	0.000363
311	ZNF701	0.8574	233.8	0.03808
312	ZNF737	0.9066	310	0.02542
313	ZNF763	0.8824	89.22	0.007205
314	ZNF773	0.9376	134.5	0.0003857
316	ZNF776	0.8676	43.03	0.01003
317	ZNF781	0.8841	386.9	0.001581
318	ZNF844	0.8832	122.8	0.04092
319	ZNF85	0.8503	120.1	0.01831
320	ZSCAN18	0.9239	81	0.04061

[0327] 组织对白细胞(血沉棕黄层)分析产生了41个高甲基化宫颈组织DMR,其中WBC中的噪音小于1%(表III,表IV)。

[0328] 表III.区分宫颈癌组织和白细胞(血沉棕黄层)的已鉴定的甲基化区域(所示区域的基因组坐标基于人类2009年2月(GRCh37/hg19)组装体)

[0329]

DMR	基因注释	染色体编号	染色体坐标
321	AGPAT3	21	45336891-45337191
322	AGRN	1	969276-969322
323	ARL5C_B	17	37321559-37321723
324	BCL2L11	2	111876440-111876822
325	BZRAP1	17	56406236-56406457
326	CA3	8	86350671-86350862

[0330]

327	CALCA	11	14995338-14995473
328	CCDC88B	11	64108152-64108329
329	CD93	20	23066944-23067181
330	DLX5	7	96650568-96650655
331	EMX1	2	73147439-73147538
332	EMX1	2	73147887-73147940
333	FAM155A_B	13	108519903-108520040
334	GATA4	8	11565295-11565473
335	HLX	1	221050491-221050533
336	HOXD4	2	177017193-177017267
337	JSRP1_B	19	2252447-2252627
338	MAX.chr10.102900288-102900393	10	102900288-102900393
339	MAX.chr18.44780825-44780880	18	44780825-44780880
340	MAX.chr19.16394457-16394536	19	16394457-16394536
341	MAX.chr2.235355101-235355212	2	235355101-235355212
342	MAX.chr3.138658547-138658804	3	138658547-138658804
343	MAX.chr5.145725411-145725510	5	145725411-145725510
344	MAX.chr6.108440666-108440885	6	108440666-108440885
345	MAX.chr9.36739811-36739868	9	36739811-36739868
346	MAX.chr9.87904996-87905325	9	87904996-87905325
347	MIR196A1	17	46711093-46711156
348	MYF6	12	81102123-81102258
349	NCOR2	12	124950720-124950803
350	PKN1	19	14551093-14551303
351	PRIC285	20	62199258-62199703
352	PRMT7	16	68390231-68390556
353	psiTPTE22	22	17090731-17090824
354	SIX2	2	45232418-45232481
355	SORCS1	10	108924547-108924660
356	TBX5	12	114847371-114847489
357	TMEM132E	17	32964694-32964757
358	TNIK	3	171175939-171176064
359	TNK1	17	7287567-7287652
360	VSTM2B_C	19	30017026-30017110
361	ZC3H12D	6	149803483-149803682

[0331] 表IV.表III中所述的标记物的宫颈癌组织对比白细胞(血沉棕黄层)的曲线下面积(AUC)、倍数变化(FC)和p值。

[0332]

DMR	基因注释	AUC 宫颈癌 对比血沉棕黄 层	FC 宫颈癌对 比血沉棕黄 层	p 值宫颈癌对 比血沉棕黄层
321	AGPAT3	1	267.8	5.21E-06
322	AGRN	0.9596	195.9	0.000978
323	ARL5C_B	0.9753	125.5	0.000231
324	BCL2L11	0.9745	203.1	0.000338
325	BZRAP1	0.9552	161.1	0.000548
326	CA3	0.9562	264.9	0.000311
327	CALCA	0.9542	136.1	0.000885
328	CCDC88B	0.9696	125	0.000115
329	CD93	1	196	0.000108
330	DLX5	1	293.1	0.000771
331	EMX1	0.9562	252.4	0.000281
332	EMX1	0.9534	221.9	1.26E-08
333	FAM155A_B	0.9583	176.8	0.000519
334	GATA4	0.967	65.29	0.000473
335	HLX	1	252.8	0.000155
336	HOXD4	0.9615	124.3	1.12E-06
337	JSRP1_B	0.9526	78.12	3.43E-06
338	MAX.chr10.10290028 8-102900393	0.9512	62.76	0.000973
339	MAX.chr18.44780825 -44780880	0.9606	62.88	0.000485
340	MAX.chr19.16394457 -16394536	0.9545	290.9	0.000159
341	MAX.chr2.235355101 -235355212	1	267.3	0.000389
342	MAX.chr3.138658547 -138658804	0.9747	134	0.000255
343	MAX.chr5.145725411 -145725510	0.9688	113.5	0.000119
344	MAX.chr6.108440666 -108440885	0.9983	192.6	0.000562
345	MAX.chr9.36739811- 36739868	0.9966	128	0.000668
346	MAX.chr9.87904996- 87905325	0.9714	363	5.03E-09
347	MIR196A1	1	191.2	0.000149
348	MYF6	0.9869	50.03	1.38E-07
349	NCOR2	1	9719	3.27E-08
350	PKN1	0.9618	276.1	0.000263
351	PRIC285	0.9815	695	0.000469
352	PRMT7	1	370.5	1.57E-06
353	psiTPTE22	0.9545	97.69	0.000368
354	SIX2	1	497.3	0.000122
355	SORCS1	0.9755	175.1	0.000343
356	TBX5	0.9983	134.2	0.000536

[0333]	357	TMEM132E	1	423	6.84E-07
	358	TNIK	0.9531	303.1	7.77E-05
	359	TNK1	0.9825	778.4	0.000116
	360	VSTM2B_C	0.9545	134	0.000204
	361	ZC3H12D	0.9512	91.08	8.80E-05

[0334] 从这些标记物组中,选择以下29个候选者进行初始试点:MAX.chr6.58147682-58147771、C10RF114、ASCL1、ARHGAP12、ZNF773、TTYH1、NEUROG3、ZNF781、NXPH1、MAX.chr9.36739811-36739868、NID2、TMEM200C、CRHR2、ABCB1、ZNF69、ATP10A、MAX.chr18.73167725-73167817、MAX.chr2.127783183-127783403、CACNA1C、ZNF382、BARHL1、MAX.chr4.8859853-8859939、ST8SIA1、MAX.chr1.98510958-98511049、C20RF40、SLC9A3、PRDM12、HOPX_C和KCNQ5;开发了定量甲基化特异性PCR测定并在独立样品上进行了测试。短扩增子引物(<150bp)被设计成靶向DMR中最具辨别力的CpG,并在对照上进行测试以确保完全甲基化的片段以线性方式稳健地扩增,而未甲基化和/或未转化的片段不会扩增。60条引物序列列于表V中。

[0335] 表V.选择用于初始分析的29个候选者的引物序列。

DMR#	名称	SEQ ID NO:	正向引物 5'-3'序列 (hg19)	SEQ ID NO:	反向引物 5'-3'序列 (hg19)
[0336]	2 ABCB1	1	CGCGAAAAGATT TTATATTGGTATT ACGT	2	AACCCGCACC CGAAACTAAC GAC
	9 ARHGAP12	3	TCG TAT TTT GTA AGG TAT AAT TCG G	4	AAC TCT TTC CTA TAC TTC ACT TCG TA
	12 ASCL1	5	TAGAGTTTTTTAT GCGTAGCGGCG G	6	ATTCTCTCTAT ATCCCCCTCGC GAA
	13 ATP10A	7	TCG TGG TTA TCG CGG CGG AAAC	8	TAC CGC ACA AAA CCC CCT TAT CCT CG
	13 ATP10A	9	TCG TGG TTA TCG CGG CGG AAAC	10	AAA CGC AAA CTA AAC GAA TAC CGC A
	17 BARHL1	11	ATA TAA ACG TTA CGG GTA	12	CAA CGA CAT ATC AAA ACC

[0337]

			GGA GCG G		CGA CCT CG
24	C1orf114	13	TTT TAA TTA CGA GAG CGA TAA AAA TTT GCG T	14	CCG AAA TAA ATA CCG AAA AAA ATC GAT
27	C2orf40	15	CGG TTA GGG TTA GGA TAG TAG GTC GCG C	16	ATT CTC CCT CGC AAC ACC TCG AAA TAC G
28	CACNA1 C	17	TTT CGA ATT TTG CGC GAA AGT CGT C	18	CGA CGA ATC TAA ATT AAT CCC TCC TCC GAA
50	CRHR2	19	GTT TTT GGG CGT TAT TTT CGG TCG T	20	TAC ACT CGA CGA CTC CTC TCC GAA
104	HOPX_C	21	GGA TAT AGT TTT TGT AAG GGG TTT CGG	22	TAT TAC GAA CTA ACG ACA AAA CCT AAC GTC
132	KCNQ5	23	AAG AGA AAT TTT TTT AAA GTG ACG T	24	AAA AAA AAC TAC AAA AAA AAC CGAA
153	MAX.chr1 .98510968 -98511049	25	ATT TTT TTT CGG AGA ATT CGA AAA GAA AAT ATG C	26	AAC TTC CAA ATC GAA AAT ATA ACG AA
170	MAX.chr1 8.7316775 1-7316779 1	27	TAG GAG GGG ACG TAG AGT TTA CGG CGA	28	GCG CAA CCC GAA CGAAC GA
177	MAX.chr2 .12778318 3-1277834 03	29	TTA GGT AGG ATT CGG ATG GCG AGG C	30	TCT ACA ACC GCA ACC AAT AAC GCC G
193	MAX.chr4 .8859853- 8859939	31	TTT TCG TTT ATT TGG GGG AAA TGG ATT TTC	32	ATC AAC ACA TCC GAA CCT CGC T
199	MAX.chr6 .58147682 -58147771	33	TGT TTA TGG ATT TAG GTG AGG ACG G	34	AAA AAA TAC AAC GTT TAA CCG CGA A
345	MAX.chr9 .36739811 -36739868	35	TGATAGGATGTT CGTTTAGTCGCG G	36	AAAAAACTAC GCCGATCCCC GAA
213	NEUROG 3	37	GGT CGT TTT TTT AGC GAC GCG GC	38	CGA AAA ACT AAA CAA CCC AAA CGA T
214	NID2	39	AGG AGT TTT TAT TTC GTT AGT TTC GT	40	ATA ACC ACC ACA TCT AAT TCT CGT T

[0338]

219	NXPH1	41	TAGTTCGCGAGA GTTTGAGAGTCG G	42	AACACGCCTA CCTTCCTAACA CGAA
233	PRDM12	43	AGG TTA TTT TCG TCG TTC GGG	44	ATA TCC CTC CGT AAA ACC GTC GAC C
254	SLC9A3	45	GCG TTT ACG GGG GAG GTC GT	46	TCT ACT ACG ATA TCT TAA CGA TCC GCG AAC
258	ST8SIA1	47	CGG TTG TTT AAC GAG AAA GAG ATC GT	48	GAT CTA ATT CCT CCT CCA CGC CGT A
271	TMEM20 0C	49	GAA CGC GGT TTT TAG GAG ATT TCG A	50	CAT ACT ACC CTC TAC GCC GCG AA
281	TTYH1	51	TTT TTG TTT TTG GGG CGC GAA AAC	52	ATA CAT CTC CTC CAC CAA CTA CCC CGC
300	ZNF382	53	GGG AGT CGG GGT TTT GGT AGA AGC G	54	ACA CCC ACG ACC GCC CTA TTA CGA C
310	ZNF69	55	GGGGTTGTTTGG ATCGATCGGAAT C	56	CCCGCAAAAA AACC AAAATC TCGC
314	ZNF773	57	TAT AGG TTT CGA TGG CGG CGG TTA C	58	ATT CGA AAA AAA CCG AAA AAA CGC GAC
317	ZNF781	59	TAG TTT ATA AAC GCG GCG GAA TCG G	60	CCA ACG ACT ACT AAA TCA AAA AAC GCA

[0339] 29个MDM中的每一个 (MAX.chr6.58147682-58147771、C10RF114、ASCL1、ARHGAP12、ZNF773、TTYH1、NEUROG3、ZNF781、NXPH1、MAX.chr9.36739811-36739868、NID2、TMEM200C、CRHR2、ABCB1、ZNF69、ATP10A、MAX.chr18.73167725-73167817、MAX.chr2.127783183-127783403、CACNA1C、ZNF382、BARHL1、MAX.chr4.8859853-8859939、ST8SIA1、MAX.chr1.98510958-98511049、C20RF40、SLC9A3、PRDM12、HOPX_C和KCNQ5) 高度区分CC与良性宫颈阴道 (BCV) 组织,其中10个MDM的曲线下面积 (AUC) >0.90 (表VI)。CC MDM还高度区分原位腺癌 (AIS) 与BCV,但在宫颈上皮内瘤变 (CIN) 2/3和CIN 1中表现不佳 (表VI)。

[0340] 表VI. 用于区分CC和BCV的标记物AUC值,以及用于区分CC和BCV、AIS和BCV、CIN 2/3和BCV、以及CIN 1和BCV在92.5%特异性下的阳性率。

[0341]

甲基化 DNA 标记物 (MDM)	AUC	对照中在 92.5%特异性截止值下的阳性率			
	CC 对比 BCV	CC 对比 BCV	AIS 对比 BCV	CIN 2/3 对比 BCV	CIN 1 对比 BCV
C1orf114	0.97 (0.94 - 1)	0.94 (0.86-0.98)	0.72 (0.55-0.86)	0.66 (0.47-0.81)	0.64 (0.31-0.89)
MAX.chr6.58147682-58147771	0.97 (0.93 - 1)	0.9 (0.81-0.96)	0.72 (0.55-0.86)	0.41 (0.24-0.59)	0.64 (0.31-0.89)
TTYH1	0.95 (0.91 - 1)	0.9 (0.81-0.96)	0.69 (0.52-0.84)	0.34 (0.19-0.53)	0.45 (0.17-0.77)
ZNF773	0.95(0.91 - 0.99)	0.88 (0.78-0.94)	0.72 (0.55-0.86)	0.25 (0.11-0.43)	0.27 (0.06-0.61)
NEUROG3	0.95 (0.9 - 0.99)	0.88 (0.78-0.94)	0.72 (0.55-0.86)	0.25 (0.11-0.43)	0.45 (0.17-0.77)
ASCL1	0.94(0.9 - 0.99)	0.91 (0.83-0.96)	0.69 (0.52-0.84)	0.34 (0.19-0.53)	0.18 (0.02-0.52)
NID2	0.92 (0.86 - 0.98)	0.8 (0.7-0.88)	0.56 (0.38-0.72)	0.22 (0.09-0.4)	0.27 (0.06-0.61)
ZNF781	0.91 (0.86 - 0.97)	0.85 (0.76-0.92)	0.64 (0.46-0.79)	0.31 (0.16-0.5)	0.27 (0.06-0.61)
CRHR2	0.91 (0.85 - 0.97)	0.85 (0.76-0.92)	0.67 (0.49-0.81)	0.38 (0.21-0.56)	0.27 (0.06-0.61)
MAX.chr9.36739811-36739868	0.9(0.84 - 0.96)	0.79 (0.69-0.87)	0.58 (0.41-0.74)	0.38 (0.21-0.56)	0.45 (0.17-0.77)

[0342] 从以前的工作中认识到,器官内癌症亚型的表观遗传学不同,并且最好的组来自亚型标记物的组合。结果在表VII中突出显示。每个标记物测定的区别强度均以数字表示。许多测定已接近完美地区分腺癌和鳞状细胞癌与良性组织。通过多种标记物测定检测到腺癌原位样品,其AUC在高80,并且CIN2/3在低80。癌症对比对照的甲基化FC%范围为2至340。

[0343] 表VII.标记物AUC和倍数变化值区分腺癌与BCV、鳞状细胞癌与BCV以及AIS与BCV。

[0344]

标记物	AUC 腺癌 对良 宫阴 癌 (BCV)	倍 数 变 化	AUC 鳞状 细胞 癌对 比良 性宫 颈阴 道癌	倍 数 变 化	AUC 原位 腺癌 对比 良性 宫阴 道癌	AUC CIN 2/3 对比 良性 宫阴 道癌
ABCB1	0.85	84	0.92	44	0.84	0.65
ARHGAP12	0.80	5	0.93	9	0.79	0.71
ASCL1	0.94	27	0.95	17	0.80	0.71
ATP10A(其中引物序列 ID 号 7 和 8)	0.73	9	0.88	13	0.57	0.54
ATP10A(引物序列 ID 号 9 和 10)	0.72	6	0.86	8	0.60	0.62
BARHL1	0.89	44	0.79	21	0.75	0.56
C1orf114	0.95	129	0.97	66	0.86	0.80
C2orf40	0.82	12	0.74	4	0.56	0.60
CACNA1C	0.93	27	0.84	12	0.83	0.38
CRHR2	0.91	46	0.88	20	0.82	0.57
HOPX_C	0.84	43	0.64	5	0.76	0.55
KCNQ5	0.82	10	0.56	2	0.70	0.54
MAX.chr1.98510968-98511049	0.91	81	0.79	23	0.82	0.72
MAX.chr18.73167751-73167791	0.85	18	0.86	12	0.73	0.71
MAX.chr2.127783183-127783403	0.90	34	0.85	13	0.76	0.52
MAX.chr4.8859853-8859939	0.91	15	0.80	8	0.78	0.60
MAX.chr6.58147682-58147771	0.97	24	0.97	16	0.85	0.83
MAX.chr9.36739811-36739868	0.90	15	0.90	17	0.74	0.65
NEUROG3	0.95	26	0.93	14	0.80	0.70
NID2	0.93	29	0.92	25	0.80	0.66
NXPH1	0.84	15	0.88	13	0.72	0.64
PRDM12	0.82	15	0.66	5	0.69	0.47
SLC9A3	0.86	26	0.71	7	0.68	0.59
ST8SIA1	0.93	42	0.81	18	0.81	0.54
TMEM200C	0.86	74	0.91	46	0.78	0.55
TTYH1	0.96	17	0.94	12	0.81	0.68
ZNF382	0.87	15	0.83	9	0.69	0.59
ZNF69	0.75	12	0.91	59	0.62	0.56
ZNF773	0.96	340	0.94	247	0.87	0.68
ZNF781	0.92	8	0.91	6	0.76	0.66

[0345] 还进行了一些实验,探讨CC MDM测定分离妇科恶性肿瘤的潜力。评估了测定的器官部位特异性,这对于基于卫生棉条的临床检验至关重要。表VIII突出显示了不同测试队列中27个选定MDM的甲基化百分比。27个MDM在宫颈癌队列中高度甲基化,但在子宫内膜癌中通常<5%。

[0346] 表VIII.标记物甲基化AUC和倍数变化值区分宫颈癌和子宫内膜癌。

[0347]

基因	子宫内 膜癌甲 基化	宫颈癌 甲基化	auc 宫 颈癌对 比 BCV	FC 宫 颈癌对 比 BCV	FC 宫 颈癌对 比子宫 内膜癌
ABCB1	0.0170	0.7881	0.9062	225.30	46.49
C1orf95	0.0209	0.5044	0.9308	189.90	24.18
CACNA1C	0.0420	0.5008	0.9755	238.70	11.92
CACNG8	0.0177	0.3506	0.9273	62.21	19.78
CHST2	0.0050	0.5196	0.9022	280.80	104.44
ELMO1	0.0241	0.3677	0.9031	85.56	15.24
EMID2	0.0252	0.3303	0.9073	95.76	13.11
FBN1_B	0.0340	0.5594	0.9299	266.20	16.47
FLT3_A	0.0109	0.5418	0.9152	75.26	49.55
FLT3_B	0.0190	0.5145	0.9118	45.13	27.02
GLIS1	0.0196	0.4107	0.904	110.80	20.98
GPC6	0.0226	0.5852	0.9007	154.10	25.93
GREM2	0.0178	0.5089	0.9047	117.70	28.64
JAM2	0.0136	0.2773	0.9268	162.00	20.34
KCNK12_A	0.0365	0.5892	0.9048	202.20	16.12
LOC100129620	0.0463	0.8123	0.9062	290.70	17.55
LOC220930	0.0059	0.2983	0.924	120.90	50.49
MAX.chr15.78112404-78112692	0.0053	0.2422	0.9055	24.62	45.40
MAX.chr19.4584907-4585088	0.0298	0.57	0.9733	213.80	19.10
MAX.chr3.69591689-69591784	0.0260	0.5668	0.9256	85.32	21.80
NCAM1	0.0514	0.7085	0.9044	188.90	13.78
NT5C1A	0.0259	0.3969	0.9671	124.20	15.33
ST8SIA3	0.0239	0.5227	0.9273	88.40	21.85
ZNF382	0.0424	0.573	0.9152	126.80	13.51
ZNF419	0.0218	0.4464	0.9002	301.00	20.45
ZNF69	0.0618	0.4054	0.9377	56.69	6.56
ZSCAN18	0.0076	0.3602	0.9239	81.00	47.59

[0348] 全甲基化组测序、严格的过滤标准和生物学验证得出宫颈癌的出色候选MDM。一些MDM以相对较高的灵敏度区分CC组织学和良性宫颈,而其他MDM则表现出亚型偏好。

[0349] 样品:

[0350] 组织和血液从Mayo Clinic生物样品库获得,并接受机构IRB监督。样品的选择严格遵守受试者研究授权和纳入/排除标准。宫颈亚型包括1) 腺癌和2) 鳞状细胞癌。对照包括良性宫颈阴道 (BCV) 组织和全血来源的白细胞。还进行了子宫内膜癌和对照。对组织进行宏观解剖并由专门的GI病理学家审查组织学。样品经过年龄性别匹配,随机化并且是盲样。使用QIAamp DNA Tissue Mini试剂盒(冷冻组织)、QIAamp DNA FFPE Tissue试剂盒(FFPE组织)和QIAamp DNA Blood Mini试剂盒(血沉棕黄层样品)(Qiagen,Valencia CA)纯化来自113个冷冻组织(16个1/2级子宫内膜样(G1/2E)、16个3级子宫内膜样(G3E)、11个浆液性、11个透明细胞EC、15个子宫癌肉瘤、44个良性子宫内膜(BE)组织(14个增生、12个萎缩、18个增生紊乱)、70个福尔马林固定石蜡包埋(FFPE)宫颈癌(CC)(36个鳞状细胞癌、34个腺癌)和18个无癌女性的血沉棕黄层)的DNA。DNA用AMPure XP珠(Beckman-Coulter,Brea CA)重新纯

化,并通过PicoGreen(Thermo-Fisher,Waltham MA)定量。使用qPCR评估DNA完整性。

[0351] 测序:

[0352] RRBS测序文库是按照Meissner协议(Gu等人Nature Protocols2011)并经过修改制备的。样品以4重形式合并,并通过Mayo Genomics Facility在Illumina HiSeq 2500仪器(Illumina, San Diego CA)上进行测序。读数由Illumina管道模块处理以用于图像分析和碱基判读。使用Mayo开发的生物信息学套件SAAP-RRBS进行二次分析。简单地说,使用Trim-Galore清理读数,并与使用BSMAP构建的GRCh37/hg19参考基因组比对。对于覆盖率 $\geq 10\times$ 并且碱基质量得分 ≥ 20 的CpG,甲基化比率通过计算 $C/(C+T)$ 或对于映射到反向链的读数,相反地,计算 $G/(G+A)$ 来确定。

[0353] 生物标记物选择:

[0354] 单个CpG按高甲基化比率排序,即给定基因座的甲基化胞嘧啶数量占该基因座总胞嘧啶计数的比率。对于病例,要求比率 ≥ 0.20 (20%);对于BCV组织对照, ≤ 0.05 (5%);对于血沉棕黄层对照, ≤ 0.01 (1%)。不符合这些标准的CpG被丢弃。随后,候选CpG通过基因组定位被分入DMR(差异甲基化区域),范围在大约60-200bp,每个区域的最小截止值为5个CpG。CpG密度过高($>30\%$)的DMR被排除在外,以避免在验证阶段出现GC相关的扩增问题。对于每个候选区域,创建了一个2D矩阵,将样品中的单个CpG与病例和对照的样品方式进行比较。分析总体CC对比所有良性子宫内膜和/或无癌血沉棕黄层,以及亚型比较。然后将这些CpG矩阵与参考序列进行比较,以评估在初始过滤过程中是否丢弃了基因组连续的甲基化位点。从这个区域子集中,最终选择要求在每个样品水平上跨DMR序列的单个CpG协调和连续的高甲基化(在病例下)。相反,对照样品的甲基化程度必须至少比病例低10倍,并且CpG模式必须更加随机且不那么协调。要求亚型队列中至少10%的癌症样品对DMR内的每个CpG位点具有至少50%的高甲基化比率。

[0355] 在单独的分析中,利用专有的DMR鉴定管道和回归包,根据CpG的平均甲基化值推导出DMR。比较CC病例、组织对照和血沉棕黄层对照之间平均甲基化百分比的差异;每个映射的CpG的100个碱基对内的平铺阅读框用于鉴定对照甲基化 $<5\%$ 的DMR;仅当覆盖的总深度为每个受试者平均10个读段并且亚组间的差异 >0 时才分析DMR。假设比值比的生物学相关增加 $>3\times$ 并且覆盖深度为10个读段,在双侧检验中,在显著性水平为5%并且假设二项式方差膨胀因子为1时,每组需要 ≥ 18 个样品才能达到80%的功效。

[0356] 回归后,DMR按p值、接受者操作特征曲线下面积(AUC)和病例与所有对照之间的倍数变化差异排序。由于事先计划了独立验证,因此在此阶段没有对错误发现进行调整。

[0357] 生物标记物验证:

[0358] 选择宫颈癌DMR的一个子集进行进一步开发。标准主要是逻辑推导的ROC曲线下面积度量,它提供了对该区域辨别潜力的性能评估。选择0.85的AUC作为病例与对照组织比较的截止值。0.95是病例与血液比较的截止值。此外,计算甲基化倍数变化比(平均癌症高甲基化比率/平均对照高甲基化比率),并且组织与组织比较采用下限20,并且组织与血沉棕黄层比较采用下限50。要求p值分别小于0.05和0.001。DMR必须在平均和单个CpG选择过程中列出。使用MethPrimer(参见Li LC和Dahiya R.Bioinformatics 2002年11月;18(11):1427-31)为候选区域设计定量甲基化特异性PCR(qMSP)引物,并对20ng(6250当量)阳性和阴性基因组甲基化对照进行QC检查。测试多个退火温度以获得最佳辨别。通过qMSP对一组

可比较的独立组织样品进行验证。其他队列包括原位腺癌 (AIS) 和宫颈上皮内瘤变 (CIN1-3)。患者人口统计数据显示在表IX中。

[0359] 表IX. 患者人口统计数据

	良性 (N=40)	CIN 1 (N=11)	CIN 2/3 (N=32)	原位腺癌 (N=36)	腺鳞癌 (N=1)	鳞状细胞 (N=38)	腺癌 (N=43)
年龄							
平均值 (SD)	47.250 (14.143)	37.455 (8.251)	43.406 (14.771)	42.861 (10.139)	39.000 (NA)	48.053 (13.511)	42.000 (13.069)
范围	18.000 -83.000	24.000 - 51.000	23.000 - 81.000	27.000 - 64.000	39.000 - 39.000	27.000 - 82.000	22.000 - 70.000
BMI							
N-Miss	0	1	1	4	0	6	13
平均值 (SD)	28.295 (7.177)	26.581 (4.413)	27.138 (6.175)	28.718 (7.628)	37.090 (NA)	29.555 (11.873)	26.825 (6.387)
范围	17.640 - 53.600	22.030 - 35.520	19.230 - 43.170	17.720 - 45.200	37.090 - 37.090	19.530 - 86.760	18.980 - 41.980
烟草使用							
N-Miss	1	0	0	0	0	0	1
绝不	23	7	18	16 (44.4%)	0 (0.0%)	19	24

[0360]

[0361]

	(59.0%)	(63.6%)	(56.2%)			(50.0%)	(57.1%)
曾经	16 (41.0 %)	4 (36.4%)	14 (43.8%)	20 (55.6%)	1 (100.0%)	19 (50.0%)	18 (42.9%)
怀孕次数							
N-Miss	0	0	0	0	0	0	1
平均值 (SD)	2.650 (1.494)	2.636 (2.461)	2.156 (2.302)	2.694 (1.600)	1.000 (NA)	2.447 (1.796)	1.857 (1.676)
范围	0.000 – 5.000	0.000 – 7.000	0.000 – 13.000	0.000 – 7.000	1.000 – 1.000	0.000 – 7.000	0.000 – 8.000
活产数							
平均值 (SD)	2.200 (1.265)	2.000 (1.673)	1.750 (2.095)	2.167 (1.404)	1.000 (NA)	2.237 (1.715)	1.465 (1.202)
范围	0.000 – 5.000	0.000 – 4.000	0.000 – 12.000	0.000 – 5.000	1.000 – 1.000	0.000 – 7.000	0.000 – 5.000
绝经状态							
未知	2 (5.0%)	1 (9.1%)	1 (3.1%)	6 (16.7%)	0 (0.0%)	8 (21.1%)	2 (4.7%)
前	23 (57.5%)	10 (90.9%)	22 (68.8%)	23 (63.9%)	1 (100.0%)	17 (44.7%)	30 (69.8%)
围	2 (5.0%)	0 (0.0%)	0 (0.0%)	2 (5.6%)	0 (0.0%)	2 (5.3%)	3 (7.0%)
后	13 (32.5%)	0 (0.0%)	9 (28.1%)	5 (13.9%)	0 (0.0%)	11 (28.9%)	8 (18.6%)
HPV 状态							
N-Miss	19	1	14	19	0	37	34
未知	2 (9.5%)	2 (20.0%)	6 (33.3%)	6 (35.3%)	1 (100.0%)	0 (0.0%)	5 (55.6%)
阴性	19 (90.5%)	1 (10.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
阳性	0 (0.0%)	7 (70.0%)	12 (66.7%)	11 (64.7%)	0 (0.0%)	1 (100.0%)	4 (44.4%)
不足	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
未完成	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

[0362] 这些组织像以前一样通过专家临床和病理审查进行鉴定。如前所述进行DNA纯化。EZ-96DNA甲基化试剂盒 (Zymo Research, Irvine CA) 用于亚硫酸氢盐转化步骤。在Roche 480LightCyclers (Roche, Basel Switzerland) 上使用SYBR Green检测扩增10ng转化的DNA (每个标记物)。连续稀释的通用甲基化基因组DNA (Zymo Research) 用作定量标准。使用CpG不可知的ACTB (β -肌动蛋白) 测定作为输入参考和标准化对照。结果表示为甲基化拷贝 (特定标记物) / ACTB拷贝。

[0363] 统计数据:

[0364] 对单个MDM (甲基化DNA标记物) 性能的结果进行逻辑分析。对于标记物组合, 随机森林回归 (rForest) 用于生成500个单独的模型, 这些模型适合原始数据的自举样品 (约2/3数据用于训练), 并用于估计整个MDM组的交叉验证误差 (1/3数据用于测试) 并重复500次以避免低估或高估真实交叉验证指标的虚假分割。然后将结果在500次迭代中取平均值。

[0365] 实施例II.

[0366] 本实施例描述了能够区分宫颈癌与子宫内膜癌和卵巢癌的甲基化标记物的鉴定。

[0367] 为了揭示三种妇科癌症(例如宫颈癌、子宫内膜癌和卵巢癌)及其各自亚型中的器官特异性高甲基化区域,进行了组合或合并这些癌症的先前生成、验证和公开的独特组织RRBS(简化表示亚硫酸氢盐测序)数据集的实验。仅分析所有三项研究共有的那些CpG。区域必须在有义链或反义链上包含至少6个连续的CpG。发现样品包括34例宫颈腺癌、36例宫颈鳞癌、15例1级或2级子宫内膜样癌、16例3级子宫内膜样癌、11例浆液性子宫癌和11例透明细胞子宫癌、15例子宫癌肉瘤以及18例浆液性、15例透明细胞、6例粘液性和18种子宫内膜样卵巢癌。良性对照包括18例宫颈阴道样品、44例子宫内膜组织(14例增殖型、12例萎缩型、18例增殖障碍型、10例分泌型)、20例输卵管样品和36例非癌血沉棕黄层或外周血白细胞样品。

[0368] 对于该应用,进行的实验集中于宫颈癌(腺癌和鳞状细胞品种两者)的特异性标记物。第一步,要求良性宫颈阴道细胞和组织中所有差异甲基化区域的背景或噪音都非常低($<1\%$)。这种组织类型的DNA代表了迄今为止在卫生棉条上发现的最高比例的核酸。其次,为了规避炎症产生的任何潜在信号,排除了白细胞DNA中甲基化 $>1\%$ 的区域。剩余的CpG用于比较宫颈癌与子宫内膜癌和卵巢癌的总体情况。分别分析腺癌和鳞状细胞癌。评估了三个指标:1)宫颈癌甲基化与子宫内膜癌和卵巢癌甲基化的比率;2)在宫颈癌样品中观察到的高甲基化的强度、频率和连续(读段水平)性质;和3)宫颈癌的绝对甲基化。首先,使用CC/EC >20 和CC/OC >50 的截止值。这些是根据经验选择的,以将DMR候选物数量从数千减少到数百。对于第三个,使用 $>20\%$ 的截止值。这使候选物数量减少到100以下。至于第二个指标,利用评分系统来鉴定整个DMR中一致的高甲基化。丢弃具有随机不一致CpG甲基化的区域。该分析鉴定了表X中所示的64个DMR组。表XI显示了表X中所述的标记物的宫颈癌甲基化与子宫内膜癌、卵巢癌和白细胞(血沉棕黄层)甲基化的比率。

[0369] 表X.组织样品中区分宫颈癌与子宫内膜癌和卵巢癌的已鉴定的甲基化区域(所示区域的基因组坐标基于人类2009年2月(GRCh37/hg19)组装体)

[0370]

DMR	基因注释	染色体 编号	染色体坐标
7	AK5	1	77747411-77747907
362	BMP6	6	7726181-7726473
363	C12orf68	12	48577334-48577492
364	C13orf18	13	46961163-46961464
365	C1orf61	1	156391403-156391670
366	CHST10	2	101033659-101033898
367	COL13A1	10	71562414-71562559
368	COL19A1	6	70577142-70577563
369	DGKZ	11	46354636-46354919
370	EBF1	5	158526311-158526385
371	ELMOD1	11	107461821-107462200
372	EML6	2	54951792-54951889
373	FAM126A	7	23053556-23054025
374	FYN	6	112194444-112194670
375	GLT25D2	1	184006028-184006497
376	KCNA2	1	111149195-111149409
377	LOC100287216	2	109745183-109745250
378	LOC255130	4	58029916-58030017
379	MAST4	5	65892420-65892687
380	MAX.chr1.161582152-161582620	1	161582152-161582620
381	MAX.chr1.42501008-42501128	1	42501008-42501128
382	MAX.chr11.133920394-133920591	11	133920394-133920591
383	MAX.chr13.29394378-29394547	13	29394378-29394547
384	MAX.chr19.11805552-11805639	19	11805552-11805639
385	MAX.chr19.12098868-12099059	19	12098868-12099059

[0371]

386	MAX.chr19.24216166-24216321	19	24216166-24216321
387	MAX.chr2.168149321-168149609	2	168149321-168149609
388	MAX.chr2.60808918-60809065	2	60808918-60809065
389	MAX.chr5.174220882-174220905	5	174220882-174220905
390	MAX.chr6.114663564-114663647	6	114663564-114663647
391	MAX.chr6.34113111-34113344	6	34113111-34113344
392	MAX.chr9.2242025-2242102	9	2242025-2242102
393	MAX.chr9.74061774-74061839	9	74061774-74061839
394	MED12L	3	150804540-150804760
395	MMP16	8	89339593-89339662
396	MYH10	17	8534612-8534792
397	NEGR1_B	1	72748492-72748610
398	NTNG1	1	107682964-107683280
399	PAQR9	3	142682557-142682820
400	PDE3B	11	14665532-14666436
401	PDE4A	19	10531731-10531820
402	PPM1E	17	56833450-56833580
403	PPP1R9A	7	94537822-94537957
404	RAB3C	5	57878796-57878920
405	SAMD5	6	147829302-147829357
406	SDC2	8	97506020-97506341
407	SDK2	17	71641381-71641605
408	SPINK2	4	57687751-57687933
409	TAF4B	18	23806366-23806832
410	TAF7	5	140700301-140700442
411	TRIM9	14	51561896-51562422
277	TRPC3_B	4	122872891-122873038
412	TSPAN5	4	99579208-99579526
413	ZNF14	19	19843701-19843789
414	ZNF211	19	58144584-58144700
415	ZNF280B	22	22862713-22862908
416	ZNF480	19	52800087-52800525
417	ZNF491	19	11909290-11909484
418	ZNF569	19	37960066-37960403
419	ZNF610	19	52839768-52839937
420	ZNF702P	19	53496769-53496864
421	ZNF709	19	12595655-12595874
422	ZNF845	19	53836815-53837093
423	ZNF91	19	23577939-23578063

[0372] 表XI.表X中所述的标记物的宫颈癌甲基化与子宫内膜癌(EC)、卵巢癌(OC)和白细胞(血沉棕黄层)甲基化的比率。

[0373]

DMR	基因注释	CC/EC 比率	CC/OC 比率	CC/血沉 棕黄比率
7	AK5	80	1178	277
362	BMP6	100	105	96

[0374]

363	C12orf68	61	159	595
364	C13orf18	63	59	464
365	C1orf61	64	102	205
366	CHST10	173	193	859
367	COL13A1	55	90	177
368	COL19A1	82	97	169
369	DGKZ	42	53	136
370	EBF1	58	53	176
371	ELMOD1	105	115	266
372	EML6	52	99	374
373	FAM126A	126	483	234
374	FYN	118	683	562
375	GLT25D2	75	102	194
376	KCNA2	45	231	346
377	LOC100287216	156	204	645
378	LOC255130	220	>1000	116
379	MAST4	119	325	383
380	MAX.chr1.161582152-161582620	41	128	112
381	MAX.chr1.42501008-42501128	164	107	178
382	MAX.chr11.133920394-133920591	66	90	61
383	MAX.chr13.29394378-29394547	159	67	172
384	MAX.chr19.11805552-11805639	152	64	74
385	MAX.chr19.12098868-12099059	105	122	197
386	MAX.chr19.24216166-24216321	185	>1000	79
387	MAX.chr2.168149321-168149609	73	>1000	113
388	MAX.chr2.60808918-60809065	44	138	103
389	MAX.chr5.174220882-174220905	196	529	466
390	MAX.chr6.114663564-114663647	62	133	130
391	MAX.chr6.34113111-34113344	109	70	127
392	MAX.chr9.2242025-2242102	179	396	104
393	MAX.chr9.74061774-74061839	36	104	59
394	MED12L	113	>1000	367
395	MMP16	68	>1000	58
396	MYH10	112	367	229
397	NEGR1_B	64	62	184
398	NTNG1	50	108	225
399	PAQR9	81	180	190
400	PDE3B	120	174	711
401	PDE4A	90	59	268
402	PPM1E	80	923	279
403	PPP1R9A	37	102	38
404	RAB3C	70	302	238
405	SAMD5	93	76	110
406	SDC2	123	166	164
407	SDK2	46	>1000	217
408	SPINK2	99	69	758
409	TAF4B	244	1223	984
410	TAF7	68	193	58

[0375]

411	TRIM9	145	74	175
277	TRPC3_B	156	813	95
412	TSPAN5	73	68	430
413	ZNF14	53	>1000	157
414	ZNF211	48	348	1776
415	ZNF280B	100	126	212
416	ZNF480	53	96	116
417	ZNF491	22	428	116
418	ZNF569	46	145	1418
419	ZNF610	40	278	52
420	ZNF702P	22	52	25
421	ZNF709	91	>1000	101
422	ZNF845	91	88	129
423	ZNF91	81	268	378

[0376] 根据这些,由于DNA数量限制,选择以下8个DMR转换为qMSP(定量甲基化特异性PCR)测定并在独立样品集中进行验证:AK5、RABC3、ZNF491、ZNF610、ZNF91、ZNF480、TRPC3_B和ELMOD1。表XII中提供了8个DMR(AK5、RABC3、ZNF491、ZNF610、ZNF91、ZNF480、TRPC3_B和ELMOD1)的引物序列。

[0377] 表XII.AK5、RABC3、ZNF491、ZNF610、ZNF91、ZNF480、TRPC3_B和ELMOD1的引物序列

[0378]

DMR #	名称	SEQ ID NO:	正向引物 5'-3'序列 (hg19)	SEQ ID NO:	反向引物 5'-3'序列 (hg19)
7	AK5	61	CGA CGT TTT ATT GCG TGC GTC GT	62	TTC CCT TAA CCA CCT AAT CCC CGA T
404	RABC3	63	GTT TGT CGG GAA TAT TCG GAG GGC	64	ACT ATC CTC TCC TAA CGC CGC ACA CG
417	ZNF491	65	TTA ATT CGG GGA AGT AGA AGG TCG T	66	AAA ACT AAA TAC AAA ACG CAA CGA A
419	ZNF610	67	ATT GAT TTA ACG TTT TGT TTC GCG T	68	AAA CGA AAT TAA AAA ACT CCC CGA A
423	ZNF91	69	CGG AGT TCG TTT GTT AAC GTA GTC GT	70	CCG AAT TCT CCT TAC CCA ACT CGA C
416	ZNF480	71	TAG ATT TCG GGT ATA GAA	72	ACA AAC CCG AAA ACG AAT

[0379]

			GCG CGG		CGC GTA
277	TRPC3_B	73	TTT CGC GGC GTT TTT TTA TTA TTT TTC GC	74	CTC CTA CCT TCC CGC CCT AAA CCG
371	ELMOD1	75	GCG GTT GTC GTA TTG GTT GC	76	GAA TAC ATC CCG ACT TAC TCC GCT

[0380] 样品包括38例宫颈腺癌、36例宫颈鳞状细胞癌、18例1级或2级子宫内膜样癌、24例3级子宫内膜样癌、16例浆液性子宫癌和7例透明细胞子宫癌、18例子宫癌肉瘤以及36例浆液性、21例透明细胞、4例粘液性和21种子子宫内膜样卵巢癌。良性对照包括29例宫颈阴道样品、14例子宫内膜组织(8例增殖型、2例萎缩型、3例增殖障碍型、1例分泌型)和29例输卵管样品。将宫颈癌总数与子宫内膜癌和卵巢癌总数进行逻辑比较。

[0381] 单个MDM(甲基化DNA标记物)的性能范围从30%灵敏度(在98%特异性下)到73%灵敏度(在99%特异性下)不等。两到三个MDM组是互补的。例如,AK5和RABC3一起;80%灵敏度,在98%特异性下。因此,标记物组合在4/5秒的时间内检测到特定的宫颈癌甲基化,假阳性率为2%。

[0382] 总之,这8个MDM(AK5、RABC3、ZNF491、ZNF610、ZNF91、ZNF480、TRPC3_B和ELMOD1)以及其余56个DMR有可能表明宫颈癌的存在,无论是腺癌还是鳞状细胞亚型,与其他两种妇科器官癌症区分的准确率很高。

[0383] 出于所有目的,以上说明书中提到的所有公布和专利通过引用整体并入本文。本技术的所描述组合物、方法和用途的各种改进和变化对于本领域技术人员是显而易知的,而不背离所描述的技术的范围和精神。尽管已结合具体示例性实施方案对本技术进行了描述,但是应当理解所要求保护的本发明不应当不适当地限定于此类具体实施方案。实际上,所描述的用于进行本发明的方式的各种修改对于药理学、生物化学、医学或相关领域技术人员而言是明显的,并且旨在包含在以下权利要求的范围内。

序列表

<110> 梅奥医学教育与研究基金会 (Mayo Foundation for Medical Education and Research)

<120> 检测宫颈癌

<130> EXCTM-39298-601

<150> US 63/157,437

<151> 2021-03-05

<160> 76

<170> PatentIn 3.5版

<210> 1

<211> 29

<212> DNA

<213> 人工序列 (Artificial Sequence)

<220>

<223> 合成

<400> 1

cgcgaaaaga ttttatattg gtattacgt 29

<210> 2

<211> 23

<212> DNA

<213> 人工序列 (Artificial Sequence)

<220>

<223> 合成

<400> 2

aacccgcacc cgaaactaac gac 23

<210> 3

<211> 25

<212> DNA

<213> 人工序列 (Artificial Sequence)

<220>

<223> 合成

<400> 3

tcgtattttg taaggataaa ttcgg 25

<210> 4

<211> 26

<212> DNA

<213> 人工序列 (Artificial Sequence)

<220>

<223> 合成
<400> 4
aactctttcc tatacttcac ttcgta 26
<210> 5
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 5
tagagttttt tatgcgtagc ggcgg 25
<210> 6
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 6
attctctcta tatccccctc gcgaa 25
<210> 7
<211> 22
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 7
tcgtggttat cgcgcggaac ac 22
<210> 8
<211> 26
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 8
taccgcacaa aaccccccta tcctcg 26
<210> 9
<211> 22
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 合成
<400> 9
tcgtggttat cgcggcggaa ac 22
<210> 10
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 10
aaacgcaaac taaacgaata ccgca 25
<210> 11
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 11
atataaacgt tacgggtagg agcgg 25
<210> 12
<211> 26
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 12
caacgacata tcaaaacccg acctcg 26
<210> 13
<211> 31
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 13
ttttaattac gagagcgata aaaatttgcg t 31
<210> 14
<211> 27
<212> DNA

<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 14
ccgaaataaa taccgaaaaa aatcgat 27
<210> 15
<211> 28
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 15
cggtagggt taggatagta ggtcgcgc 28
<210> 16
<211> 28
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 16
attctccctc gcaacacctc gaaatacg 28
<210> 17
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 17
tttcgaattt tgcgcgaaag tcgtc 25
<210> 18
<211> 30
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 18
cgacgaatct aaattaatcc ctctccgaa 30
<210> 19
<211> 25

<212> DNA
<213> 人工序列 (Artificial Sequence)
<220>
<223> 合成
<400> 19
gttttttgggc gttattttcg gtcgt 25
<210> 20
<211> 24
<212> DNA
<213> 人工序列 (Artificial Sequence)
<220>
<223> 合成
<400> 20
tacactcgac gactcctctc cgaa 24
<210> 21
<211> 27
<212> DNA
<213> 人工序列 (Artificial Sequence)
<220>
<223> 合成
<400> 21
ggatatagtt tttgtaagg gtttcgg 27
<210> 22
<211> 30
<212> DNA
<213> 人工序列 (Artificial Sequence)
<220>
<223> 合成
<400> 22
tattacgaac taacgacaaa acctaacgtc 30
<210> 23
<211> 25
<212> DNA
<213> 人工序列 (Artificial Sequence)
<220>
<223> 合成
<400> 23
aagagaaatt tttttaagt gacgt 25
<210> 24

<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 24
aaaaaaaaact acaaaaaaaaaa ccgaa 25
<210> 25
<211> 34
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 25
atttttttttc ggagaattcg aaaagaaaat atgc 34
<210> 26
<211> 26
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 26
aacttccaaa tcgaaaatat aacgaa 26
<210> 27
<211> 27
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 27
taggagggga cgtagagttt acggcga 27
<210> 28
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 28
gcgcaacccg aacgaaacga 20

<210> 29
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 29
ttaggtagga ttcggatggc gaggc 25
<210> 30
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 30
tctacaaccg caaccaataa cgccg 25
<210> 31
<211> 30
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 31
ttttcgttta tttgggggaa atggattttc 30
<210> 32
<211> 22
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 32
atcaacacat ccgaacctcg ct 22
<210> 33
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 33

tgtttatgga tttaggtgag gacgg 25
<210> 34
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 34
aaaaaataca acgtttaacc gcgaa 25
<210> 35
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 35
tgataggatg ttcgtttagt cgcgg 25
<210> 36
<211> 23
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 36
aaaaaactac gccgatcccc gaa 23
<210> 37
<211> 23
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 37
ggtcgttttt ttagcgacgc ggc 23
<210> 38
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成

<400> 38
cgaaaaacta aacaacccaa acgat 25
<210> 39
<211> 26
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 39
aggagttttt atttcgtag tttcgt 26
<210> 40
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 40
ataaccacca catctaattc tcgtt 25
<210> 41
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 41
tagttcgcca gagtttgaga gtcgg 25
<210> 42
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 42
aacacgccta ccttcctaac acgaa 25
<210> 43
<211> 21
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>

<223> 合成
<400> 43
aggttattttt cgtcgttcgg g 21
<210> 44
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 44
atatccctcc gtaaaaccgt cgacc 25
<210> 45
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 45
gcgttttacgg gggaggtcgt 20
<210> 46
<211> 30
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 46
tctactacga tatcttaacg atccgcgaac 30
<210> 47
<211> 26
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 47
cggttgttta acgagaaaga gatcgt 26
<210> 48
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 合成
<400> 48
gatctaattc ctcctccacg ccgta 25
<210> 49
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 49
gaacgcggtt tttaggagat ttcga 25
<210> 50
<211> 23
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 50
catactaccc tctacgccgc gaa 23
<210> 51
<211> 24
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 51
tttttgTTTT tggggcgCGa aaac 24
<210> 52
<211> 27
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 52
atacatctcc tccaccaact accccgc 27
<210> 53
<211> 25
<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<223> 合成

<400> 53

gggagtcggg gttttggtag aagcg 25

<210> 54

<211> 25

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<223> 合成

<400> 54

acacccacga ccgccctatt acgac 25

<210> 55

<211> 25

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<223> 合成

<400> 55

ggggttgttt ggatcgatcg gaatc 25

<210> 56

<211> 24

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<223> 合成

<400> 56

cccgcaaaaa aacaaaaatc tcgc 24

<210> 57

<211> 25

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<223> 合成

<400> 57

tataggtttc gatggcggcg gttac 25

<210> 58

<211> 27

<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 58
attcgaaaaa aaccgaaaaa acgcgac 27
<210> 59
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 59
tagtttataa acgcggcgga atcgg 25
<210> 60
<211> 27
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 60
ccaacgacta ctaaatacaa aaacgca 27
<210> 61
<211> 23
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 61
cgacgtttta ttgcgtgcgt cgt 23
<210> 62
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 62
ttcccttaac cacctaatacc ccgat 25
<210> 63

<211> 24
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 63
gtttgtcggg aatattcgga gggc 24
<210> 64
<211> 26
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 64
actatcctct cctaacgccg cacacg 26
<210> 65
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 65
ttaattcggg gaagtagaag gtcgt 25
<210> 66
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 66
aaaactaaat acaaaacgca acgaa 25
<210> 67
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 67
attgatttaa cgttttgttt cgcgt 25

<210> 68
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 68
aaacgaaatt aaaaaactcc ccgaa 25
<210> 69
<211> 26
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 69
cggagttcgt ttgttaacgt agtcgt 26
<210> 70
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 70
ccgaattctc cttacccaac tcgac 25
<210> 71
<211> 24
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 71
tagatttcgg gtatagaagc gcgg 24
<210> 72
<211> 24
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 72

acaaacccga aaacgaatcg cgta 24
<210> 73
<211> 29
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 73
tttcgcggcg tttttttatt atttttcgc 29
<210> 74
<211> 24
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 74
ctcctacctt cccgccctaa accg 24
<210> 75
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 75
gcggttgctg tattggttgc 20
<210> 76
<211> 24
<212> DNA
<213> 人工序列(Artificial Sequence)
<220>
<223> 合成
<400> 76
gaatacatcc cgacttactc cgct 24